Spring 2022

Results on Select Combinatorial Problems With an Extremal Nature

Stephen Smith

Results on select Combinatorial Problems with an Extremal Nature

by

Stephen Smith

Bachelor of Science
University of Delaware 2014

Master of Science
University of Delaware 2016

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Mathematics

College of Arts and Sciences

University of South Carolina

2022

Accepted by:

Éva Czabarka, Major Professor

László Székely, Major Professor

Joshua Cooper, Committee Member

Frank Thorne, Committee Member

Jason O'Kane, Committee Member

Tracey Weldon, Interim Dean and Vice Provost for Graduate Education

Dedication

To my father, Edward, and our many tableside philosophical discussions. Your impact on me was greater than you could've ever known.

Acknowledgments

I would like to express my sincerest gratitude to my advisors, Éva Czabarka and László Székely, for welcoming me into their research family. This dissertation would not have been possible without their continuous support and unique insights. I will be forever grateful for the opportunity to work with and learn from both of them. Their guidance throughout this process has been invaluable and arguably one of the most pivotal parts of my experience at the University of South Carolina.

Writing a dissertation is far from an individual effort. I am extremely grateful for the support and assistance of my committee members: Joshua Cooper, Frank Thorne, and Jason O'Kane.

I owe a debt of gratitude to the many professors that supported and encouraged me through the years. It would be impossible to list each by name, but I would like to give special thanks to Felix Lazebnik, Robert Coulter, and Sebastian Cioabă, for introducing me to the wonderful world of discrete math and for being amazing mentors to me during my time at the University of Delaware.

I would also like to thank my friends and family for the various roles that they played. A huge thank you to my mom and stepfather, Linda and Michael Doto, and to my mother- and grandmother-in-law, Laurie and Laura Petrella, for always cheering me on. Another huge thank you to my good friend, James Wisniewski, for helping to keep me balanced with our sporadic late-night adventures in Faerûn, Hydaelyn, and Sanctuary. Finally, it is my pleasure to thank my wife, Annamarie, for her incredible support. Always there to listen and give insightful, honest, and heartfelt advice, she has been my center throughout this demanding experience.

Abstract

This dissertation is split into three sections, each containing new results on a particular combinatorial problem. In the first section, we consider the set of 3-connected quadrangulations on n vertices and the set of 5-connected triangulations on n vertices. In each case, we find the minimum Wiener index of any graph in the given class, and identify graphs that obtain this minimum value. Moreover, we prove that these graphs are unique up to isomorphism.

In the second section, we work with structures emerging from the biological sciences called tanglegrams. In particular, our work pertains to an invariant of tanglegrams called the tangle crossing number, an invariant which is NP-hard to compute. Czabarka, Székely, and Wagner found a finite characterization of tanglegrams with tangle crossing number equal to 0, which motivated the work here. In particular, our aim was to find a similar finite (and minimal) characterization of tanglegrams with tangle crossing number at least k, for any fixed $k \geq 2$. We set out to prove this using an elegant order-theoretic argument, but came to another surprising result instead; we proved that the set of tanglegrams with the induced subtanglegram relation is not a well partial order.

In the final section, we work on the problem of finding an upper bound on the diameter of graphs with particular properties. It was proven independently by several groups that for fixed minimum degree $\delta \geq 2$, every connected graph G of order n satisfies $\operatorname{diam}(G) \leq \dfrac{3n}{\delta+1} + O(1)$ as $n \to \infty$. Erdős, Pach, Pollack, and Tuza noticed that the graphs which achieve the aforementioned bound all contain complete subgraphs whose order increases with n, and conjectured that if we disallowed complete

subgraphs of a given fixed size, then we could improve the bound. Czabarka, Singgih, and Székely recently found a counterexample to part of the conjecture of Erdős *et al.* and formulated a new conjecture. Under a stronger assumption, we verify two cases of this new conjecture using a novel unified duality approach.

Table of Contents

List of Figures

Chapter 1

Introduction

Extremal stuctures and values are important for many collections of objects that are ordered or have a well-defined parameter that belongs to an ordered set. For these objects, answers to questions like "What is the best-case scenario?", "What is the worst-case scenario?", "How large can it be?", or "How small can it be?" often provide invaluable information to both theory and application. In this work, we will explore select combinatorial results that are all extremal in nature.

1.1 Definitions and Notation

For the many standard terms and definitions related to graph theory, I refer the reader to the classic text by Reinhard Diestel ([14]). I include only a few definitions and notational conventions that are absolutely central to the work presented here. Throughout this work, all graphs are simple graphs.

Definition 1. Let $G = (V, E)$ be a graph. The *order* of G is equal to $|V|$ and is denoted by $|G|$. The *size* of G is equal to $|E|$ and is denoted by $||G||$.

Notation 1. Let $G = (V, E)$ be a graph. For any vertex $x \in V$, we write $N(x)$ to denote the set of neighbors of x in G.

Definition 2. Let $G = (V, E)$ be a graph, and let $x \in V$. The *degree* of x is equal to $|N(x)|$ and is denoted by $d(x)$.

Definition 3. Let $G = (V, E)$ be a connected graph. For any two vertices $x, y \in V$, the *distance* from x to y is the length of the shortest $x - y$ path in G. This is denoted by $d(x, y)$, or $d_G(x, y)$ if it is important to specify the underlying graph.

Definition 4. Let $G = (V, E)$ be a connected graph, and let $x \in V$. The *eccentricity* of x, denoted by $\mathrm{ecc}(x)$, is the maximum distance from x to any vertex in the graph.

Definition 5. Let $G = (V, E)$ be a connected graph. The *diameter* of G is equal to the maximum distance between any two vertices, and is denoted by $\mathrm{diam}(G)$.

Definition 6. Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two graphs. A function $\phi : V_1 \to V_2$ is called a *graph isomorphism* if ϕ is bijective and satisifies the property that $\{u, v\} \in E_1$ if and only if $\{\phi(u), \phi(v)\} \in E_2$. If such a function exists, we say that G_1 and G_2 are *isomorphic* and write $G_1 \simeq G_2$.

Definition 7. A graph $G = (V, E)$ is called *planar* if it can be embedded in the plane, i.e., if it is isomorphic to a plane graph.

Definition 8. Let $G = (V, E)$ be a graph. For a positive integer, c, we say that G is c*-connected* if $|G| > c$ and $G - X$ is connected for every set $X \subseteq V$ with $|X| < c$. The *connectivity* of G is the largest c for which G is c-connected, and is denoted by $\kappa(G)$.

Notation 2. We denote the complete graph on n vertices by K_n.

Notation 3. We denote the complete bipartite graph with partite classes of size m and n by $K_{m,n}$.

Notation 4. We denote the path of length n by P_n.

Notation 5. For the most part, C_n denotes the cycle of length n. In Sections 3.2 and 3.3, the same notation is used to refer to a special kind of tree. The exact meaning of C_n should be clear from context.

Notation 6. For a graph $G = (V, E)$, we denote by $\delta(G)$ and $\Delta(G)$ the minimum and maximum degree of G, respectively.

Chapter 3 includes order-theoretic ideas and results. Some definitions relevant to these concepts are included below.

Definition 9. A *partially ordered set* (also called a *poset* or *partial order*) is a set equipped with a binary relation that is reflexive, antisymmetric, and transitive. If the set is denoted by X and the relation is denoted by $\leq$, then we will denote the partial order by $(X, \leq)$.

Definition 10. Let $(X, \leq)$ be a partially ordered set. Then a subset $A \subseteq X$ is called a *chain* if any two elements of A are comparable. In other words, for all $a, b \in A$, either $a \leq b$ or $b \leq a$.

Definition 11. Let $(X, \leq)$ be a partially ordered set. We say that $(X, \leq)$ is *well-founded* if it does not contain any infinite strictly decreasing chains.

Definition 12. Let $(X, \leq)$ be a partially ordered set. Then a subset $A \subseteq X$ is called an *antichain* if no two distinct elements of A are comparable. In other words, for all $a, b \in A$, if $a \neq b$, then $a \not\leq b$ and $b \not\leq a$.

Definition 13. Let $(X, \leq)$ be a partially ordered set. We say that $(X, \leq)$ is a *well partial order* if it is well-founded and has no infinite antichains.

Definition 14. Let $(X, \leq)$ be a partially ordered set. We say that a subset $U \subseteq X$ is an *upward closed set* (also called an *upset*) if whenever $u \in U$, $x \in X$, and $u \leq x$, it is also true that $x \in U$.

Definition 15. Let $(X, \leq)$ be a partially ordered set and let $U \subseteq X$ be an upward closed set. If $A \subseteq U$ such that for all $u \in U$, there exists $a \in A$ with $a \leq u$, then we say that A *generates* U, and we call A a *generating set*.

Chapter 2

Minimum Wiener Index

2.1 History and Background

The primary graph invariant of concern in this chapter is the so-called *Wiener index*:

Definition 16. Given a connected graph, G, the *Wiener index* of G is the sum of the distances between all unordered pairs of vertices and is denoted by $W(G)$. This is more succinctly written in formula as

$$W(G) = \sum_{\{u,v\}\subseteq V(G)} d_G(u,v). \tag{2.1}$$

This graph invariant was introduced in 1947 by Harry Wiener ([40]) to predict the boiling point of alkanes. It is related to the average distance between vertices of a given graph and provides important information for scientists working on graph-theoretic models in fields such as chemical graph theory and nanotechnology (see for example [15], [16], [19], [24], [34], [35], [37], [39], [41], [42], [43]). It is natural to ask the following: What is the maximum or minimum possible Wiener index in a given collection of graphs and which graph(s) achieve these bounds? Answers to these questions can provide upper or lower bounds to quantities defined in terms the Wiener index. For example, one could ask the question: What is the minimum possible Wiener index among all (connected) graphs on n vertices and which graph(s) achieve this minimum value?

The next few results are rather trivial, but I include them to convey the spirit of the arguments to come.

Theorem 1. *Let $n \geq 1$ be fixed. Up to isomorphism, the graph K_n is the unique minimizer of the Wiener index among all graphs of order n and has Wiener index $\binom{n}{2} = \frac{n(n-1)}{2}$.*

Proof. By minimizing each term of the right-hand side of Eq. 2.1 individually, we minimize $W(G)$. The smallest that the distance can be between two distinct vertices is 1, which occurs if and only if the two vertices in question are connected by an edge. Let G be a graph on n vertices that minimizes the Wiener index. Since $W(G)$ is minimized when each term of the right-hand side of Eq. 2.1 is 1, there must be an edge between every pair of (distinct) vertices of G. This forces G to be isomorphic to K_n, which has Wiener index $\binom{n}{2} = \frac{n(n-1)}{2}$. □

We can ask for a similar result about the maximum Wiener index among all connected graphs on n vertices:

Theorem 2. *Let $n \geq 1$ be fixed. Up to isomorphism, the graph P_{n-1} is the unique maximizer of the Wiener index among all graphs of order n and has Wiener index $\dfrac{n^3 - n}{6}$.*

Proof. We prove by induction on n that the path P_{n-1} is the unique maximizer of the Wiener index among graphs of order n. For $n \in \{1, 2\}$, this is clearly true as the paths P_0 and P_1 are the only connected graphs on 1 and 2 vertices, respectively.

Suppose that the statement is true for some $n \geq 2$, and let G be a connected graph of order $n + 1$. There must be some vertex $v \in V(G)$ such that $G - v$ is connected, and by induction, $W(G - v) \leq W(P_{n-1})$. Next, we note that

$$\begin{aligned} W(G) &\leq W(G - v) + \sum_{u \in V(G-v)} d_G(u, v) \\ &\leq W(P_{n-1}) + \sum_{i=1}^{n} i \\ &= W(P_n). \end{aligned}$$

To see that P_n is the unique maximizer, note that equality holds in the equation above if and only if $G - v$ is equal to P_{n-1} and $\{d_G(v,u) : u \in V(G) \setminus \{v\}\} = [n]$. This can only occur if $G = P_n$.

Finally, to see that $W(P_{n-1}) = \dfrac{n^3 - n}{6}$, note that

$$\begin{aligned} W(P_{n-1}) &= \sum_{i=1}^{n-1} \sum_{j=1}^{i} j \\ &= \sum_{i=1}^{n-1} \frac{i(i+1)}{2} \\ &= \frac{1}{2} \left(\sum_{i=1}^{n-1} i^2 + \sum_{i=1}^{n-1} i \right) \\ &= \frac{1}{2} \left(\frac{(n-1)(n)(2n-1)}{6} + \frac{n(n-1)}{2} \right). \end{aligned}$$

The result follows. $\square$

Since the path also happens to be a tree, it follows that the path P_{n-1} also uniquely maximizes the Wiener index among all trees of order n. In the next result, we consider which graph obtains the minimum Wiener index among all trees of order n:

Theorem 3. *Let $n \geq 2$ be fixed. Up to isomorphism, the star $K_{1,n-1}$ is the unique minimizer of the Wiener index among all trees of order n and has Wiener index $n^2 - 2n + 1$.*

Proof. Let $n \geq 2$ be fixed. The statement is clearly true for $n = 2$ since there is exactly one tree of order 2, namely $K_{1,1}$. Assume that T is a tree of order $n \geq 3$. As with every tree of order n, T has precisely $n-1$ edges, i.e. T has $n-1$ pairs of vertices which are distance 1 apart. Futhermore, any non-adjacent pair of vertices are distance at least 2 apart, and there are exactly $\binom{n}{2} - (n-1)$ of these pairs. Therefore, $W(T) \geq 1 \cdot (n-1) + 2 \cdot \left(\binom{n}{2} - (n-1)\right) = n^2 - 2n + 1$, with equality if and only if $\operatorname{diam}(T) = 2$. Since $|T| \geq 3$, we have that $\operatorname{diam}(T) \geq 2$ with equality if and only if $T = K_{1,n-1}$. $\square$

There are many classes of graphs for which one might be interested in results similar to those of Theorems 1, 2, and 3. Important classes of graphs that we will be considering in this work are the set of *triangulations* on n vertices and the set of *quadrangulations* on n vertices.

Definition 17. A *triangulation* is a simple graph drawn in the sphere in which every face is a triangle.

Definition 18. A *quadrangulation* is a simple graph drawn in the sphere in which every face is a quadrangle.

Triangulations and quadrangulations are edge-maximal planar and edge-maximal planar bipartite graphs, respectively. We will be repeatedly using the following classical facts about trianglulations and quadrangulations:

Fact 1. Because triangulations and quadrangulations are planar graphs which we are considering to be drawn on the sphere, Euler's formula applies to them: For any finite, connected planar graph with n vertices, e edges, and f faces,

$$n - e + f = 2. \tag{2.2}$$

Fact 2. Let G be a triangulation on n vertices. Euler's Formula implies the following:

i. G has $3n - 6$ edges.

ii. G has $2n - 4$ faces.

iii. G is either 3-, 4-, or 5-connected.

iv. If G is 5-connected, then $|V(G)| \geq 12$.

Fact 3. Let G be a quadrangulation on n vertices. Euler's Formula implies the following:

i. G has $2n - 4$ edges.

ii. G has $n-2$ faces.

iii. G is either 2- or 3-connected.

iv. If G is 3-connected, then $|V(G)| \geq 8$.

There have been numerous results recently on the maximum Wiener index of triangulations and quadrangulations (see [7], [8], [20], and [22]). Lower bounds for the Wiener index of these classes of graphs were also stated in [7] and [8] without consideration for the connectivity:

Theorem 4 ([8])**.** *Assume $n \geq 6$. The triangulation T_n^4 defined in* Figure 2.1 *minimizes the Wiener index among all triangulations of order n. The triangulation T_n^4 is* 4*-connected. Consequently, the triangulation T_n^4 minimizes the Wiener index among all* 4*-connected n-vertex triangulations as well.*

Remark: T_5^4 is the only triangulation of order 5, but it is not 4-connected. Gray vertices and dashed edges in the figures indicate the pattern to be repeated as n increases.

Proof. A triangulation contains $3n-6$ edges, thus there are exactly $3n-6$ pairs of vertices at distance 1 apart. If we can make sure that every remaining pair of vertices are at distance 2 apart, then we have a triangulation whose Wiener index is $2\left(\binom{n}{2}-(3n-6)\right)+(3n-6)=n^2-4n+6$, and this is clearly the minimum possible Wiener index. This is the case with T_n^4. Furthermore, it is easy to see that T_n^4 is 4-connected for all $n \geq 6$. $\square$

Theorem 5 ([7], [8])**.** *Assume $n \geq 4$. The complete bipartite graph $K_{2,n-2}$ minimizes the Wiener index among all quadrangulations.*

Proof. A quadrangulation contains $2n-4$ edges, thus exactly $2n-4$ pairs of vertices are at distance 1 apart. If we can make sure that every remaining pair of vertices are

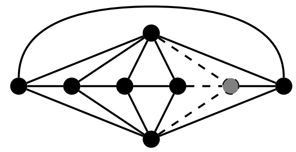


Figure 2.1: The triangulation T_n^4, which is the join of the cycle C_{n-2} with the edgeless graph on two vertices, minimizes the Wiener index among all triangulations of order $n \geq 5$ and are 4-connected for $n \geq 6$.

at distance 2 apart, we have a quadrangulation of Wiener index $2\left(\binom{n}{2} - (2n-4)\right) + (2n-4) = n^2 - 3n + 4$. This is the case with the quadrangulation $K_{2,n-2}$. Clearly this is the least possible Wiener index of a quadrangulation. $\square$

Theorem 6. *Assume $n \geq 4$. Up to isomorphism, the graph $K_{2,n-2}$ is the unique minimizer of the Wiener index among all quadrangulations of order n.*

Proof. Let Q be a quadrangulation of order n that has the same Wiener index as $K_{2,n-2}$, i.e. every non-adjacent pair of vertices are at distance 2. As quadrangulations are 2-connected, the minimum degree $\delta := \delta(Q) \geq 2$. Let v be a vertex of Q with $d(v) = \delta$, and let $u_1, \ldots, u_\delta$ be the neighbors of v. The remaining $n - \delta - 1$ vertices are at distance 2 from v. As quadrangulations are bipartite, these $n - \delta - 1$ vertices can only be adjacent to $u_1, \ldots, u_\delta$, and have degree at least δ. Thus we get that $Q \simeq K_{\delta,n-\delta}$. Since δ is the minimum degree, $\delta \leq n - \delta$, therefore Q contains $K_{\delta,\delta}$ as a subgraph. Since Q is planar, we get $\delta = 2$ and $Q \simeq K_{2,n-2}$. $\square$

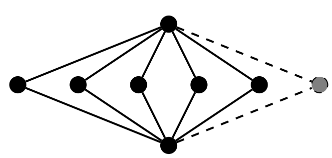


Figure 2.2: The graph $K_{2,n-2}$ which minimizes the Wiener index among all quadrangulations of order n.

The results of Theorems 4, 5, and 6 are obtained when the class of graphs in considerations is the class of *all* triangulations on n vertices and *all* quadrangulations on n vertices. In the work presented here, we will be resricting ourselves to the class of 5-connected triangulations on n vertices and the class of 3-connected quadrangulations on n vertices. The results that follow were originally presented in [30], but were proven there with extensive use of the output of an elaborate program written by Olsen. In this work, we have removed all computer aid from the proofs.

The results presented in the remainder of this chapter are joint work with Éva Czabarka, Trevor Olsen, and László Székely.

2.2 Minimum Wiener Index of 3-connected Quadrangulations

In this section, we prove the following:

Theorem 7. *Assume that $n \geq 8$, $n \neq 9$. The minimum Wiener index of 3-connected quadrangulations of order n is*

$$4\left\lceil \frac{n}{2} \right\rceil^2 + \left(\left\lceil \frac{n}{2} \right\rceil + 21\right)\left(\left\lfloor \frac{n}{2} \right\rfloor - 21\right) - 5n + 449 = \begin{cases} \frac{5n^2}{4} - 5n + 8, & \text{if } n \text{ is even,} \\ \frac{5n^2}{4} - 3n - \frac{49}{4}, & \text{if } n \text{ is odd.} \end{cases}$$

The unique minimizer of the Wiener index among 3-connected quadrangulations of order n is Q^3_n, defined in Figure 2.5.

A combination of Lemma 12 (e) and Theorems 14 and 16 prove the above theorem. To that end, we first define an auxiliary drawn graph, which we will use extensively in this section. Let v be a vertex of a 3-connected quadrangulation G. We define the *sunflower graph* S_v around v (in a planar drawing of G), as v connected to its neighbors $u_1, \ldots, u_d$ (listed in the cyclic order of the drawing, $d = d(v)$), and different vertices $w_1, \ldots, w_d$ where w_i is connected to u_i and u_{i+1} (indices taken modulo d, see Figure 2.3). We understand S_v as a part of the drawing of G.

We need to show that such a graph, with distinct vertices, exists in the drawing. We will also need some special properties of the sunflower graph, which will be shown in Lemma 8 below.

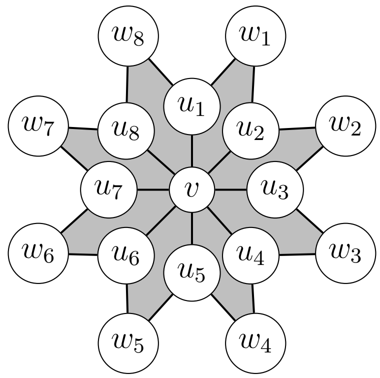


Figure 2.3: The sunflower graph S_v around v with $d(v) = 8$. The region $\mathcal{R}_v$ is shaded.

Lemma 8. *Assume that Q is a drawing of a 3-connected quadrangulation. Then, for any vertex v, Q contains a sunflower graph S_v with $2d(v) + 1$ distinct vertices. Furthermore, the region $\mathcal{R}_v$ that contains v and is bounded by the cycle $C_v = u_1 w_1 \dots u_{d(v)} w_{d(v)}$ contains no vertices or edges that are not in S_v.*

Proof. We know $\delta(Q) \geq 3$ by the 3-connectedness. Label the neighbors of v by $u_1, \dots, u_d$, in their planar cyclic order around v. For each pair of successive neighbors u_i and u_{i+1} (indices taken modulo d), let $w_i \neq v$ be their common neighbor that completes the face f_i that has u_i, v, u_{i+1} on its boundary. This means, in particular, that the interior of f_i has no vertices or edges. If y is a neighbor of u_i and $y \notin \{v, w_{i-1}, w_i\}$ then y must lie between w_{i-1} and w_i in the planar cyclic order around u_i, in particular, $w_{i-1} \neq w_i$ as $d(u_i) \geq 3$. As Q is bipartite, $u_i \neq w_j$ for all $1 \leq i, j \leq d$. We will show that each of the w_i's must be distinct. As $\mathcal{R}_v$ is the union of the faces f_i, this finishes the proof. Assume that $w_i = w_j$ for some $j \neq i$. We already know that $j \notin \{i-1, i+1\}$ and the vertices $u_i, u_{i+1}, u_j, u_{j+1}$ are all different. We consider two regions of the planar drawing of Q: $\mathcal{R}_1$ is bounded by the 4-cycle $u_{i+1} v u_j w_i$ and does not contain the vertex u_i, and $\mathcal{R}_2$ is bounded by the 4-cycle $u_i v u_{j+1} w_i$ and does not contain the vertex u_{i+1}. Thus the faces bounded by $u_i v u_{i+1} w_i$ and $u_j v u_{j+1} w_j$ are disjoint from $\mathcal{R}_1$ and $\mathcal{R}_2$. The neighbors of u_i that differ from v and w_i must lie in $\mathcal{R}_2$ and the neighbors of u_j that differ from v and w_i must lie in $\mathcal{R}_1$.

Hence $\{v, w_i\}$ separates u_i from u_j (See Figure 2.4), contradicting the fact that Q is 3-connected. $\square$

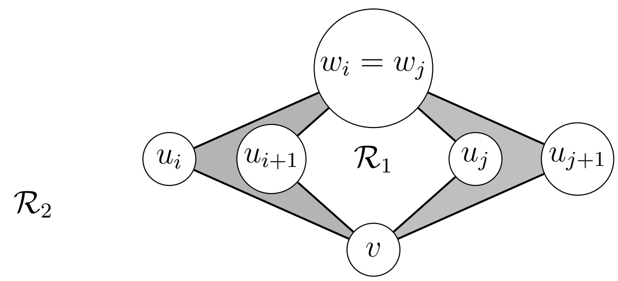


Figure 2.4: 2-element cutset appears in S_v, when $w_i = w_j$. The two faces $vu_iw_iu_{i+1}$ and $vu_jw_ju_{j+1}$ are shaded.

Lemma 9. *Assume that G is a 3-connected quadrangulation with partite sets A, B. Then*

(a) $\Delta(G) \leq \min\{|A| - 1, |B| - 1\}$;

(b) *If $|B| < |A|$, then for all $x \in A$ we have $d(x) \leq |B| - 2$; and*

(c) $|V(G)| \neq 9$, *i.e., no 3-connected quadrangulations exist on 9 vertices.*

Proof. (a) Let v be a vertex with degree $\Delta = \Delta(G)$, we may assume $v \in B$. As the sunflower S_v is a subgraph of the planar drawing of G, $\Delta \leq \min(|B| - 1, |A|)$. We are done unless $|B| > |A| = \Delta$, so assume that is the case. As $A = N(v)$, all neighbors of the vertices of B lie in $N(v)$, in particular, every w_i has at least 3 neighbors in A. For each i let k_i be the largest positive integer such that w_i has no neighbors in the set $\{u_{i-t} : 1 \leq t \leq k_i - 1\} \cup \{u_{i+1+t} : 1 \leq t \leq k_i - 1\}$. Since for $k = 1$ the sets $\{u_{i-t} : 1 \leq t \leq k - 1\}$ and $\{u_{i+1+t} : 1 \leq t \leq k - 1\}$ are empty, such positive integers exist, they have an upper bound from the fact that w_i has at least 3 neighbors in A, and for the largest such integer k_i we have that at least one of u_{i-k_i}, u_{i+1+k_i} is a neighbor of w_i that is different from u_i, u_{i+1}. Choose i_0 such that $k = k_{i_0}$ is minimal amongst the k_i. By renumbering the u_i if necessary and changing the direction of the cyclic order we can assume that $i_0 = 1$ and w_1 is connected to u_1, u_2, u_{2+k} but none of u_{1-t}, u_{2+t} for all $1 \leq t \leq k-1$. Let $\mathcal{R}$ be the region of the sphere bounded by the 4-cycle $u_2 v u_{2+k} w_1$ that does not contain u_1. Consider w_2. By the definition of w_2 and the minimality of k, w_2 lies in $\mathcal{R}$ and it has at least one neighbor u_j that does not lie in $\mathcal{R}$. The edge $w_2 u_j$ must cross the boundary of $\mathcal{R}$, which contradicts the planarity of G. Thus, we have $\Delta(G) \leq \min\{|A|-1, |B|-1\}$, as claimed.

To prove the case (b), assume $|B| < |A|$, i.e. $n > 2|B|$. We already know that $\Delta(G) \leq |B|-1$. Assume that A contains a vertex v of degree $|B|-1$. Since $|A| = n - |B|$ and all other vertices of A have degree at least 3, we have that $2n-4 \geq (|B|-1) + 3(n-|B|-1) = 3n - 2|B| - 4$, so $n \leq 2|B|$, a contradiction.

To prove the case (c), assume to the contrary that G has 9 vertices and partite sets A, B. We may assume $|B| < |A|$, and therefore $|B| \leq 4$. Then every vertex in A has degree at most 2, a contradiction. $\square$

Lemma 10. *In a 3-connected quadrangulation G of order n, the number of unordered pairs of vertices at distance 2 is at most*

$$\frac{1}{2}\sum_v d^2(v) - 4(n-2).$$

This estimate is exact precisely when G has no non-facial 4-cycles.

Proof. Euler's Formula gives us that any quadrangulation on n vertices has $2(n-2)$ edges and $n-2$ faces. The number of 2-paths in G is equal to $\sum_v \binom{d(v)}{2} = \frac{1}{2}\sum_v d^2(v) - 2(n-2)$. This sum, however, overcounts the number of pairs of vertices distance 2 apart. In a 3-connected quadrangulation, two faces cannot share two consecutive edges from their boundaries. Thus, for each face, we are double counting the two pairs of vertices distance 2 apart, and so we may safely subtract $2(n-2)$. There are pairs of vertices which we have double counted even after the substraction precisely when there are non-facial 4-cycles. $\square$

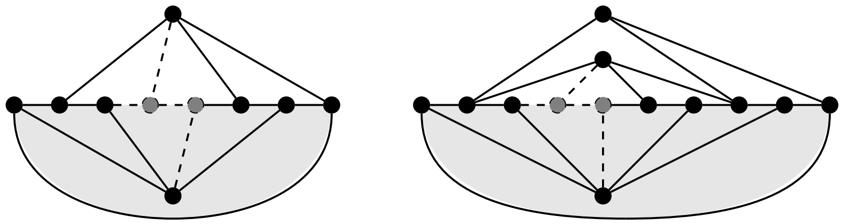


Figure 2.5: The quadrangulation Q_n^3 of order $n = 2k \geq 8$ (left) and $n = 2k+1 \geq 11$ (right), which minimizes the Wiener index among all 3-connected quadrangulations of order n. Gray vertices and dashed edges indicate the pattern to be repeated. The light gray regions are the sunflower graphs around a maximum degree vertex.

Lemma 11. *Let Q be a 3-connected quadrangulation of order n, with partite sets A, B. Then*

$$W(Q) \geq 2n^2 + 2n - 8 - |A||B| - \sum_v d^2(v). \tag{2.3}$$

Equality holds in (2.3) *precisely when the diameter of Q is at most* 4 *and Q has no non-facial* 4*-cycles.*

Proof. Let Q be an arbitrary 3-connected quadrangulation on n vertices, with partite sets A, B. Let D_i denote the number of unordered pairs of vertices at distance i in Q. Clearly $W(Q) = \sum_i i \cdot D_i$. Observe that $D_1 = 2n - 4$, the number of edges; $D_2 \leq \frac{1}{2}\sum_v d^2(v) - 4(n-2)$ by Lemma 10; $D_2 + D_4 + D_6 + D_8 \cdots = \binom{|A|}{2} + \binom{|B|}{2}$, as pairs of vertices are at even distance precisely when they are from the same partite set; and finally, $D_1 + D_3 + D_5 + D_7 + \cdots = |A| \cdot |B|$, as pairs of vertices are at odd distance precisely when they are from different partite sets.

Combining all this information with the identity $|A| + |B| = n$, we obtain that

$$\begin{aligned} W(Q) &\geq (2n-4) + 2D_2 + 3\Big[|A| \cdot |B| - (2n-4)\Big] + 4\left[\binom{|A|}{2} + \binom{|B|}{2} - D_2\right] \\ &= 2n^2 - 2n - |A||B| - 2(D_2 + 2n - 4) \\ &\geq 2n^2 + 2n - 8 - |A||B| - \sum_v d^2(v). \end{aligned}$$

The first inequality in the displayed formula is an equality precisely when the diameter of Q is at most 4, and the second inequality is an equality precisely when Q has no non-facial 4-cycles. $\square$

The 3-connected quadrangulation Q^3_n of order $n \geq 8$, $n \neq 9$ is defined in Figure 2.5. The following lemma is easy to verify, and we leave the details to the reader.

Lemma 12. *Assume that $n \geq 8$, $n \neq 9$.*

(a) *Q^3_n is a 3-connected quadrangulation.*

(b) *Q^3_n has no non-facial 4-cycle.*

(c) *If n is even, Q^3_n has diameter 3 and degree sequence $\frac{n}{2}-1, \frac{n}{2}-1, 3, \ldots, 3$.*

(d) *If n is odd, Q^3_n has diameter 4 and degree sequence $\lfloor \frac{n}{2} \rfloor - 1, \lfloor \frac{n}{2} \rfloor - 2, 4, 4, 3, \ldots, 3$. (For $n = 11$, the terms in this sequence are* not *in decreasing order.)*

(e)

$$W(Q^3_n) = \begin{cases} \frac{5n^2}{4} - 5n + 8, & \textit{if } n \textit{ is even,} \\ \frac{5n^2}{4} - 3n - \frac{49}{4}, & \textit{if } n \textit{ is odd.} \end{cases}$$

The following is obvious, and we make use of it frequently

Lemma 13. *Assume that $\sum_{i=1}^n x_i = a > 0$ is given, where the x_i's are required to be integers from the interval $[b, c]$ with $0 \leq b$, and we have to maximize $\sum_{i=1}^n x_i^2$. As long as for some $(i \neq j)$ we have $b + 1 \leq x_j \leq x_i \leq c - 1$, we can increase the sum of squares while keeping the conditions by changing x_i to $x_i + 1$ and x_j to $x_j - 1$.*

Theorem 14. *Assume that the number $n \geq 8$ is even. The quadrangulation Q^3_n defined in* Figure 2.5 *minimizes the Wiener index among all 3-connected quadrangulations of order n. Moreover, up to isomorphism, this minimizer is unique.*

Proof. Let Q be an arbitrary 3-connected quadrangulation on $n = 2k$ vertices, with partite sets A, B. Since Q is 3-connected, for all v, we have $d(v) \geq 3$, and by Lemma 9, $d(v) \leq \Delta(Q) \leq \min(|A| - 1, |B| - 1) \leq \frac{n}{2} - 1$. By Lemma 13 and Lemma 12 (c), $\sum_{v \in V(Q)} d^2(v) \leq \sum_{v \in V(Q^3_n)} d^2(v)$ with equality precisely when Q has the same degree sequence as Q_n^3. Also, $|A| \cdot |B| \leq \frac{n^2}{4}$ with equality precicely when $|A| = |B| = \frac{n}{2}$. Lemma 11 gives that $W(Q) \geq W(Q_n^3)$ with equality precisely when Q has the same degree sequence as Q_n^3, $|A| = |B| = \frac{n}{2}$, Q has diameter at most 4 and no nonfacial 4-cycles. In particular, Q_n^3 minimizes the Wiener index among n-vertex 3-connected quadrangulations.

We will show that the extremal quadrangulation is in fact unique. Assume that $W(Q) = W(Q_n^3)$, so Q has the same degree sequence as Q_n^3 and $|A| = |B| = k$ vertices. Then in both A and B we have $k-1$ vertices of degree 3, and the remaining one vertex must have degree $k-1$ ($k-1 \geq 3$).

As before, let v be a vertex of maximum degree $k-1$, and construct the sunflower graph S_v around v. Since S_v has exactly $n-1$ vertices, Q has one additional vertex v'. This vertex v' is in the same partite class as the u_i vertices, and differs from v. Each of the w_i has one edge not in S_v incident upon it, connecting them to either v' or one of the u_j. If all vertices u_i have degree 3, then v' has degree $k-1 \geq 3$, and it is adjacent to all w_i (in which case we have Q_n^3). Otherwise the degree of v' is 3 and exactly one of the u_i (say u_2) has degree $k-1 > 3$ in Q. Assume that the latter is the case. As w_1 and w_2 have an edge not in $E(S_v)$ incident upon them, and $w_1u_2, w_2u_2 \in E(S_v)$, both w_1 and w_2 are adjacent to v'. Thus, $v'w_1u_2w_2$ bounds a facial region $\mathcal{R}$. As $w_1u_2w_2$, of which u_2 is an internal vertex, is the common boundary of $\mathcal{R}_v$ and $\mathcal{R}$, u_2 cannot have any edge outside of S_v incident upon it, a contradiction. □

Lemma 15. *Assume $n = 2k+1$, and let Q be a 3-connected quadrangulation of order n, with partite sets A, B. If*

$$\sum_{v \in V(Q)} d^2(v) < 2k^2 + 12k + 10, \tag{2.4}$$

then $W(Q) > W(Q_n^3)$. If $\Delta(Q) \leq k-2$, then $W(Q) > W(Q_n^3)$.

Proof. First note that by Lemma 12 (d)

$$\sum_{x \in V(Q_n^3)} d^2(x) \;=\; (k-1)^2 + (k-2)^2 + 2 \cdot 4^2 + 3^2(2k-3) = 2k^2 + 12k + 10,$$

and also $|A||B| \le k(k+1)$ (note that $k, k+1$ are the sizes of the partite classes in Q_n^3), so if (2.4) holds, then by Lemma 11 and Lemma 12 (d) we have $W(Q) > W(Q_n)$.

Since Q has odd number of vertices and minimum degree 3, the Handshaking Lemma implies $\Delta(Q) \ge 4$. Assume now that $\Delta(Q) \le k-2$, so $k \ge 6$. Let $x_1, \dots, x_{2k+1}$ be a sequence of integers that maximizes $\sum x_i^2$ subject to the conditions that that $\sum x_i = 4(n-2)$ and $3 \le x_i \le k-2$. If $k = 6$, the maximizing sequence is $4, 4, 4, 4, 4, 3, \dots, 3$ of length 13, and if $k = 7$ the maximizing sequence by Lemma 13 is $5, 5, 5, 4, 3, \dots 3$ of length 15. In both of these cases, we have $\sum x_i^2 < 2k^2 + 12k + 10$. For $k \ge 8$, Lemma 13 gives that the maximizing sequence is $k-2, k-2, 6, 3, 3, \dots, 3$, so

$$\sum_{x \in V(Q)} d^2(x) \le \sum_{i=1}^{2k+1} x_i^2 = 2(k-2)^2 + 6^2 + 3^2(2k-2) = 2k^2 + 10k + 26 \le 2k^2 + 12k + 10.$$

Therefore $W(Q) > W(Q_n^3)$ unless $k = 8$ and the degree sequence of Q is $6, 6, 6, 3, 3, \dots, 3$.

So for the rest of this proof $k = 8$, the degree sequence of Q is $6, 6, 6, 3, 3, \dots, 3$ and is of length 17. By Lemma 11 if $W(Q) \le W(Q_n^3)$, then $W(Q) = W(Q_{17}^3)$, the diameter of Q is 4, Q has no nonfacial 4-cycles, $|A| = 9$ and $|B| = 8$. We will show that such a Q does not exist, which finishes the proof.

Because the sum of the degrees of the vertices in each partite class must be the same (in this case, 30), B contains exactly two of the degree 6 vertices. Let $v \in B$ with degree 6, consider the sunflower S_v, and label the 4 vertices outside S_v by x, y_1, y_2, y_3 such that $B = \{v, w_1, \dots, w_6, x\}$ and $A = \{u_1, \dots, u_6, y_1, y_2, y_3\}$. Since $d(x) \in \{3, 6\}$, $N(x) \subseteq A$ and at most one of the u_i has an edge not from S_v incident upon it, we have $d(x) = 3$ and without loss of generality $y_1, y_2 \in N(x)$.

Assume first that $N(x) = \{y_1, y_2, y_3\}$ and consider the sunflower S_x. Let j_i be chosen such that w_{j_i} is the common neighbor of y_i and y_{i+1} (indices taken modulo

3) in S_x. Then each of the w_{j_i} are different and have degree at least 4 in Q, a contradiction.

So we can assume without loss of generality that $N(x) = \{u_1, y_1, y_2\}$. Then the unique degree 6 vertex in A is u_1, so there are two different indices t_1 and t_2 such that $w_{t_i} \in N(u_1) \setminus \{w_1, w_6\}$. For $i \in \{1, 2\}$ let z_i be the common neighbor of u_1 and y_i in the sunflower S_x, and let z_3 be the common neighbor of y_1 and y_2 in S_x. Then $\{z_1, z_2\} \subseteq \{w_1, w_2, w_{t_1}, w_{t_2}\}$ and $z_3 \in \{w_1, \ldots, w_6\} \setminus \{z_1, z_2\}$. In particular, the degree of z_3 in Q is at least 4, therefore z_1 and z_2 must have degree 3 in Q. If $w_{t_i} \in \{z_1, z_2\}$, then w_{t_i} has degree at least 4 and consequently degree 6. This gives $\{z_1, z_2\} \cap \{w_{t_1}, w_{t_2}\} = \emptyset$. Therefore without loss of generality $w_1 = z_1$, $w_6 = z_2$ and the $u_1w_{t_i}$ edges cannot run inside $\mathcal{R}_v$ or any of the faces bounded the 4-cycles $u_1w_1y_1xu_1$ and $u_1w_6y_2xu_1$, which leaves them no place to be, a contradiction. $\square$

Theorem 16. *Assume that the number $n \geq 11$ is odd. The quadrangulation Q_n^3 in* Figure 2.5 *minimizes the Wiener index among all* 3*-connected quadrangulations of order n. Moreover, up to isomorphism, this minimizer is unique.*

Proof. Let $n = 2k + 1$ and assume that Q is a 3-connected quadrangulation on n vertices of minimum Wiener index, and with partite sets A, B, such that $|A| > |B|$.

First we want to show that $|A| = k + 1$, $|B| = k$, and the degree sequence of Q restricted to the partite sets is the same as the degree sequence of Q_n^3 restricted to its partite sets.

We have $|A| \geq k + 1$ and $|B| \leq k$. Lemma 9 (a) gives $\Delta(Q) \leq |B| - 1 \leq k - 1$. Lemma 15 gives $\Delta(Q) = k - 1$, which in turn shows $|B| = k$ and $|A| = k + 1$. In addition, if $d(v) = k - 1$, Lemma 9 (b) gives $v \in B$.

As the degree sequence of quadrangulations is unique for $n = 11$ under the condition that every degree is 3 or 4, we may assume now that $n \geq 13$, i.e., $k \geq 6$. As the sum of the k degrees in B is the number of edges $2n - 4 = 4k - 2$, and every degree is at least 3, only two degree sequences are possible for B: $(k-1, 5, 3, 3, \ldots, 3)$ or $(k-1, 4, 4, 3, \ldots, 3)$. We claim that $\Delta(A)$, the maximum degree of a vertex in A is $k-2$. Lemma 9 (b) showed $\Delta(A) \leq |B| - 2 = k-2$.

Assume for contradiction that $\Delta(A) \leq k-3$. Since the minimum degree is at least 3 and $\sum_{x \in A} d(x) = 4k-2$, for $k = 6$ we get that $3 \cdot 7 = 4 \cdot 6 - 2$, a contradiction. Therefore we have that $k \geq 7$,

$$\sum_{x \in A} d^2(x) \leq (k-3)^2 + 4^2 + 3^2(k-1) = k^2 + 3k + 16,$$

and

$$\sum_{x \in V(Q)} d^2(x) \quad \leq \quad k^2 + 3k + 16 + (k-1)^2 + 5^2 + 3^2(k-2) = 2k^2 + 10k + 24.$$

By Lemma 15 $W(Q) > W(Q_n^3)$ when $k \geq 8$ so we may assume that $k = 7$. In particular, for $k \geq 8$ the degree sequence of A is $(k-2, 3, \ldots, 3)$.

If $k = 7$, the degree sequence of A is $(4, 4, 3, 3, 3, 3, 3, 3)$ and the degree sequence of B is $(6, 4, 4, 3, 3, 3, 3)$ then $\sum_{x \in V(Q)} d^2(x) = 190 < 192 = 2 \cdot 7^2 + 12 \cdot 7 + 10$, and Lemma 15 contradicts the minimality of the Wiener index of Q.

Hence the only case that remains to be checked is when $k = 7$, the degree sequence of A is $(4, 4, 3, 3, 3, 3, 3, 3)$ and the degree sequence of B is $(6, 5, 3, 3, 3, 3, 3)$. In this case $\sum_{x \in V(Q)} d^2(x) = 192 = \sum_{x \in V(Q_{15}^3)} d^2(x)$, so by Lemma 11 and Lemma 12 (a), (d) the minimality of $W(Q)$ implies that Q has no nonfacial 4-cycles. Let $v \in B$, $d(v) = 6$ and consider the sunflower S_v. Let x, y be the vertices outside S_v. Then $B = \{v, w_1, \ldots, w_6\}$ and $A = \{u_1, \ldots, u_6, x, y\}$, without loss of generality $d(w_1) = 5$, and the rest of the w_i have degree 3. Therefore there is an $i \in \{3, 4, 5, 6\}$ such that w_1 is adjacent to u_i. Since w_1 and u_i cuts C_v into two paths, one contains w_2 and the other w_6, the vertices w_2 and w_6 lie inside two different regions bounded by the 4-cycle $w_1 u_j v u_1 w_1$. As this cycle is nonfacial, we have a contradiction.

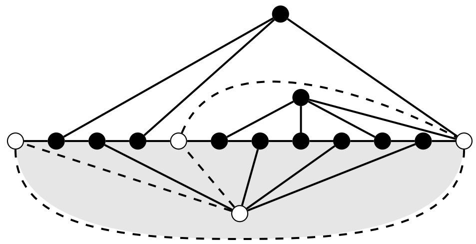


Figure 2.6: The quadrangulation Y of order 13 with Wiener index 164. The gray region shows the sunflower around a maximal degree vertex. The white vertices and the dotted edges form one of the non-facial 4-cycles.

So we have that the degree sequence of Q restricted to A is the same as the degree sequence of Q_n^3 restricted to its A, i.e. $(k-2, 3, \dots, 3)$. We need to figure out what the degree sequence of Q restricted to B is. Assume $k \geq 6$, and B has degree sequence $(k-1, 5, 3, 3, \dots, 3)$. Referring to the sunflower graph S_v at vertex v, where $d(v) = k-1$, we have $B = \{v, w_1, \dots, w_{k-1}\}$ and $A = \{u_1, \dots, u_{k-1}, x, y\}$. We can assume without loss of generality that $w_1 \in B$ has degree 5, and for $i : 2 \leq i \leq k-1$ set z_i be the unique vertex in $N(w_i) \setminus \{u_i, u_{i+1}\}$. Since w_1 is adjacent to 3 vertices of $A \setminus \{u_1, u_2\}$, it is adjacent to at least one (and at most three) vertices in $\{u_i : 3 \leq i \leq k-1\}$, and consequently these vertices have degree at least 4. As A has a single vertex with degree more than 3, we conclude that there is a unique $j : 3 \leq j \leq k-1$ that w_1 is adjacent to u_j, $d(u_j) = k-2$ and $d(x) = d(y) = 3$, and w_1 is adjacent to x and y. In addition, for $i : 2 \leq i \leq k-1$ we have that $z_i \in \{x, y, u_j\}$; in particular $z_{j-1}, z_j \in \{x, y\}$. We may assume without loss of generality that $z_{j-1} = x$.

Let $\mathcal{P}$ be the region bounded by C_v that is different from $\mathcal{R}_v$, and let $\mathcal{R}_1$ and $\mathcal{R}_2$ be the two subregions that the edge w_1u_j cuts $\mathcal{P}$ into; without loss of generality the boundary of $\mathcal{R}_1$ is the cycle $w_1u_2w_2u_3 \dots u_j$. Now $\mathcal{R}_1, \mathcal{R}_2$ and $\mathcal{R}_v$ share only vertices on the boundary, and the common boundary of $\mathcal{R}_1$ and $\mathcal{R}_2$ is the edge u_jw_1. $\mathcal{R}_1$ has $j-2 \geq 1$ vertices $w_2, \dots w_{j-1}$ from $B \setminus \{w_1\}$ on its common boundary with $\mathcal{R}_v$, and for $i : 2 \leq i \leq j-1$ the vertex z_i lies in $\mathcal{R}_1$ (inside or on the boundary). Since $z_{j-1} = x$, x is inside $\mathcal{R}_1$. Let $\mathcal{Q}$ be the subregion of $\mathcal{R}_1$ bounded by the cycle $w_1u_2w_2\cdots u_{j-1}w_{j-1}xw_1$. Then for $i : 2 \leq i \leq j-2$ the vertex z_i lies in $\mathcal{Q}$, so $z_i \in \{x, y\}$. $\mathcal{R}_2$ has $k-j \geq 1$ vertices $w_j, w_{j+1}, \ldots, w_{k-1}$ from $B \setminus \{w_1\}$ on its common boundary with $\mathcal{R}_v$ and for $i : j \leq i \leq k-1$, z_i lies in $\mathcal{R}_2$ (inside or on the boundary). Since $z_j \in \{x, y\}$ and x is inside $\mathcal{R}_1$, this implies that $z_j = y$, y is inside $\mathcal{R}_2$, for all $i : j \leq i \leq k-1$ we have $z_i \in \{u_j, y\}$ and for all $i : 2 \leq i \leq j-1$ we have $z_i = x$. Similar logic as before gives that for all $i : j \leq i \leq k-1$ we have $z_i = y$ Since $d(x) = d(y) = 3$, this means $3 = j-1 = k-j+1$, so $j = 4$ and $k = 6$. We have $Q \simeq Y$ (see Figure 2.6) and $W(Q) = 164 > 160 = W(Q_{13}^3)$, a contradiction.

For the rest of the proof we assume that $n \geq 11$, so $k \geq 5$. The integer sequence that maximizes the sum of squares, and satisfies the conditions we have for the degree sequence of G in A (respectively B) is $k-2, 3, \ldots, 3$ (respectively $k-1, 4, 4, 3, \ldots, 3$), the degree sequence of Q_n^3. Since Q_n^3 is a 3-connected quadrangulation with diameter at most 4, this shows that $W(Q_n^3)$ is minimal, and the degree sequence of Q is the same as the degree sequence of Q_n^3, and furthermore, the degree sequences of their respective partite sets are the same. Last, we need to show that $Q \simeq Q_n^3$.

Let $v \in Q$ with $d(v) = k-1$, and consider the sunflower S_v around v. Let x, y be the vertices of Q not in S_v. Then $B = \{v, w_1, w_2, \ldots, w_{k-1}\}$, $A = \{u_1, \ldots, u_{k-1}, x, y\}$, and without loss of generality the two vertices of degree 4 in B are w_1 and w_j.

If every u_i has degree 3 (this must happen in particular when $k = 5$ and vertices of A all have degree 3), then none of the u_i has a neighbor outside of S_v. In this case w_1 and w_j must both be adjacent to x and y. Without loss of generality the region $\mathcal{R}$ bounded by the cycle $xw_1u_2w_2\ldots u_jw_jx$ that does not contain v contains y. (Otherwise we exchange the name of x and y). If $j = k-1$, then x is on the interior of the 4-cycle $yw_{k-1}u_1w_1y$ that does not contain v, and the degree of x can only be 2, which is a contradiction. If $j = 2$, then $\mathcal{R}$ is bounded by a 4-cycle and y can have only degree 2, a contradiction. So $3 \leq j \leq k-2$, the $k-1-j \geq 1$ vertices $w_{j+1}, w_{j+2}, \ldots, w_{k-1}$ must have x as their third neighbor, and the the $j-2 \geq 1$ vertices

$w_2, w_3, \ldots, w_{j-1}$ must have y as their third neighbor. So $\{d(x), d(y)\} = \{k+1-j, j\} = \{3, k-2\}$, which gives $j \in \{3, k-2\}$. This is precisely the graph Q_n^3.

Now let i be chosen so u_i have degree greater than 3. As all but one of the vertices of A have degree 3, for $j \neq i$ we have $d(u_j) = 3$, u_j has the same neighbors in Q and S_v, $d(u_i) = k-2$, $d(x) = d(y) = 3$, and $k \geq 6$. This means that u_i is adjacent to precisely $k-5 \geq 1$ of the vertices in $\{w_s : s \notin \{i-1, i\}, 1 \leq s \leq k-1\}$, so it is adjacent to at least one w_ℓ such that $\ell \notin \{i-1, i\}$. If $i < \ell \leq k-1$ then $u_i v u_\ell w_\ell u_i$ is a non-facial 4-cycle (as $u_i w_i u_{i+1} \ldots w_{\ell-1} u_\ell$ lies in one of the regions bounded by this cycle while $w_\ell u_{\ell+1} w_{\ell+1} \ldots w_{i-1} u_i$ lies in the other region). If $1 \leq \ell < i-1$ then $u_i v u_{\ell+1} w_\ell u_i$ is a non-facial 4-cycle. Since Q can not have non-facial 4 cycles by Lemma 11, this is a contradiction. $\square$

2.3 Minimum Wiener Index of 5-connected Triangulations

In this section, we prove the following:

Theorem 17. *Assume that $n \geq 12$, $n \neq 13$. The minimum Wiener index of 5-connected triangulations of order n is*

$$2n\left\lceil\frac{n}{2}\right\rceil + \left(\left\lceil\frac{n}{2}\right\rceil + 14\right)\left(\left\lfloor\frac{n}{2}\right\rfloor - 14\right) - 7n + 208 = \begin{cases} \frac{5n^2}{4} - 7n + 12, & \text{if } n \text{ is even,} \\ \frac{5n^2}{4} - 6n - \frac{9}{4}, & \text{if } n \text{ is odd.} \end{cases}$$

The unique minimizer of the Wiener index among 5-connected triangulations of order $n \neq 19$ is T_n^5, defined on Figure 2.9*, while for $n = 19$, exactly two minimizers exist, namely T_{19}^5, and the 5-connected triangulation X of order 19, defined on* Figure 2.10.

The techniques used to do so are similar to Section 2.2 and the proof is a combination of Lemma 21 (e) and Theorems 23 and 27.

First we state some facts about triangulations of a simple n-gon not using additional vertices. Triangulations of an n-gon can be viewed as planar graphs, where the outer face is bounded by an n cycle and all other faces are bounded by a 3-cycle (we will refer to such faces as *triangles*).

Lemma 18. *Let $n \geq 4$. Any triangulation of a simple n-gon uses $n-3$ additional edges (i.e. edges which are not edges of the n-gon), and has at least 2 triangles with exactly two of their boundary edges on the n-gon.*

Proof. The fact that the triangulation has $n-3$ edges (and consequently $n-2$ triangles) is easy to prove by induction on n. When $n \geq 4$, all these triangles have at most 2 boundary edges on the n-gon. As there are $n-2$ triangles inside and n edges on the n-gon itself, by the pigeonhole principle some two triangles must have two edges from edges of the n-gon. □

We need the following basic facts about 5-connected triangulations:

Lemma 19. *Let T be a 5-connected triangulation of order n. The following are true:*

(a) *Every 3-cycle is the boundary of a face and every 4-cycle is the boundary of a region whose interior does not contain vertices of the graph, and contains exactly one edge.*

(b) *Every edge lies on exactly two triangles. If abc and bcd are triangles of T, then ad is not an edge of T.*

(c) *For every edge xy of T, there is precisely one 4-cycle in T that goes through its vertices, but does not use the xy edge; hence the number of 4-cycles in T is $3(n-2)$.*

(d) *If x, y are non-adjacent vertices in T, then there is at most one 4-cycle that contains them.*

(e) *Let D_i denote the number of unordered pairs of vertices at distance i in T. We have $D_1 = 3(n-2)$ and*

$$D_2 = \frac{1}{2} \sum_{x \in V(T)} d^2(x) - 12(n-2).$$

(f)

$$W(T) \geq 3\binom{n}{2} + 6(n-2) - \frac{1}{2} \sum_{x \in V(T)} d^2(x),$$

with equality if and only if T has diameter at most 3.

Proof. (a): If C is a cycle that separates two regions that both contain vertices in their interior, then the vertices of C form a cutset, therefore C has at least 5 vertices.

(b): An edge bounds two faces that are triangles, and if there is a third 3-cycle using the edge, the other two edges of two of these 3-cycles give a 4-cycle that has vertices in both of its regions. If *abc* and *bcd* are triangles such that *ad* is an edge, then one of *abd*, *acd* would be a non-facial triangle unless $n = 4$. Both of these contradict (a), and (b) follows.

(c): Since every 4-cycle *abcd* bounds a region that has no vertices but has an edge (say *ac*), and if *ac* is an edge then *bd* cannot be an edge by (b), for every 4-cycle there is a unique edge that is not part of the cycle and connects two of its vertices. So we can map 4-cycles to edges by assigning this edge to each cycle. This map is injective. If two different 4-cycles would map to the same edge, this edge is part of three triangles, contradicting (b).

As each edge lies on two triangles which together form a 4 cycle, every edge is assigned to precisely one of these 4-cycles, so the map is surjective as well. Thus, the number of 4-cycles is the same as the number of edges, which is $3(n-2)$ in any planar triangulation. (c) follows.

(d): Assume x, y are non-adjacent vertices that appear on two 4-cycles. As each 4-cycle containing x, y has two x-y paths of length 2, we have at least three x-y paths of length 2, say xa_1y, xa_2y, xa_3y. By (a), for each $i, j \in \{1, 2, 3\}$, $i \neq j$ the cycle

$C_{ij} = xa_iya_jx$ bounds a region $\mathcal{R}_{ij}$ that contains no vertices in its interior. But the three regions $\mathcal{R}_{12}, \mathcal{R}_{13}\mathcal{R}_{23}$ together with their boundaries cover the entire plane, so T has no other vertices besides x, y, a_1, a_2, a_3. This is a contradiction, as 5-connected triangulations must have at least 12 vertices.

(e): Observe that D_1 is exactly the number of edges of T, $3(n-2)$. The formula $\sum \binom{d(x)}{2} = \frac{1}{2}\sum d^2(x) - 3(n-2)$ counts the number of paths of length 2 between unordered pairs of vertices. If an unordered pair of vertices has more than 1 such path, it appears on a 4-cycle, and by (c) and (d) this 4-cycle is unique. As each 4-cycle contains exactly two such unordered pairs of vertices, the number of unordered pairs of vertices that have a path of length 2 between them is $\frac{1}{2}\sum d^2(x) - 9(n-2)$ by (c). As every edge is contained in exactly one 4-cycle, this equals $D_1 + D_2$, proving (e).

(f): As $\sum_i D_i = \binom{n}{2}$, we get

$$\begin{aligned} W(T) &= \sum_i iD_i \geq D_1 + 2D_2 + 3\left(\binom{n}{2} - D_1 - D_2\right) = 3\binom{n}{2} - 2D_1 - D_2 \\ &= 3\binom{n}{2} + 6(n-2) - \frac{1}{2}\sum_{x \in V(T)} d^2(x), \end{aligned}$$

and equality holds precisely when the diameter is at most 3. □

Analogously to Section 2.2, we define an auxiliary drawn graph, which we will use extensively. Let T be a 5-connected triangulation, and let $v \in V(T)$ have degree d. We define the *mosaic graph* M_v at vertex v, together with its planar drawing, in the following way. M_v contains the neighbors of v in G, $u_1, u_2, \ldots, u_d$, with the edges vu_i, such that vertices u_i are labeled according the clockwise cyclic order of the edges. We include the edges $u_iu_{i+1} \in E(T)$ for every $1 \leq i \leq d$ (indices are taken modulo d) in M_v, following the drawing of T. We also add a vertex $w_i \neq v$, which is a common neighbor of u_i and u_{i+1}, together with edges joining them to u_i and u_{i+1} in T, following the drawing of T, for every i. We understand M_v as a part of the drawing of T. We will show that M_v has $2d+1$ distinct vertices.

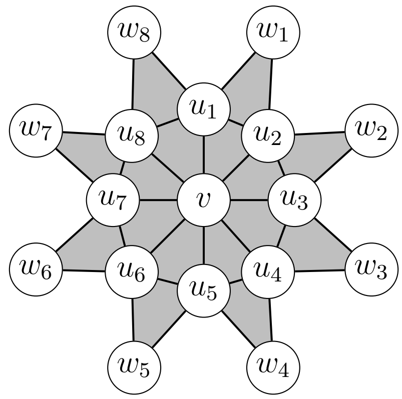


Figure 2.7: The mosaic graph M_v around v, with $d = 8$. The grey region is $\mathcal{R}_v$.

Lemma 20. *If T is a drawing of a 5-connected triangulation of order n, and v is any vertex of T with degree $d(v) = d$, the mosaic graph M_v in T has $2d+1$ distinct vertices. Furthermore, the region $\mathcal{R}_v$ that is bounded by the cycle $u_1w_1u_2w_2\dots u_dw_d$ and contains the vertex v, contains edges and vertices from T if and only if they are edges and vertices in the mosaic graph M_v. In addition, T contains at least one vertex not in M_v, consequently $\Delta(T) \leq \lfloor \frac{n}{2} \rfloor - 1$. Moreover, $n \neq 13$.*

Proof. Since T is 5-connected, $\delta(T) \geq 5$. As before, label the neighbors of v by $u_1, \dots, u_d$, in their planar clockwise cyclic order around v. We get for free that u_iu_{i+1} is an edge in T, since we have a triangulation. For each pair of successive neighbors u_i and u_{i+1} (indices taken modulo d), let $w_i \neq v$ be their common neighbor that completes the face that has u_iu_{i+1} on its boundary, but not v. This means, in particular, that $\mathcal{R}_v$ will satisfy the required property, so we just need to show that the vertices listed in M_v are all distinct.

If y is a neighbor of u_i and $y \notin \{v, w_{i-1}, w_i, u_{i-1}, u_{i+1}\}$ then y must lie between w_{i-1} and w_i in the planar cyclic order around u_i, In particular, as $d(u_i) \geq 5$, we have that $w_{i-1} \neq w_i$.

Also, $u_i \neq w_j$ for all $1 \leq i, j \leq d$. For $j \in \{i-1, i\}$ this is obvious, and for other values of j if $u_i = w_j$ then v, u_i, u_j is a 3-element cutset.

Assume now that $w_i = w_j$ for some $j \neq i$. We already have that $j \notin \{i-1, i+1\}$ and hence the vertices $u_i, u_{i+1}, u_j, u_{j+1}$ are all distinct. We consider two regions of the planar drawing of T: $\mathcal{R}_1$ is bounded by the 4-cycle $u_{i+1}vu_jw_i$ and does not contain the vertex u_i, and $\mathcal{R}_2$ is bounded by the 4-cycle $u_ivu_{j+1}w_i$ and does not contain the vertex u_{i+1}.

The neighbors of u_i that differ from v, u_{i+1} and w_i, lie in $\mathcal{R}_2$ and the neighbors of u_j that differ from v, u_{j+1} and w_i, must lie in $\mathcal{R}_1$. Hence $\{v, u_{i+1}, u_{j+1}, w_i\}$ separates u_i from u_j (see Figure 2.8), contradicting that T is 5-connected.

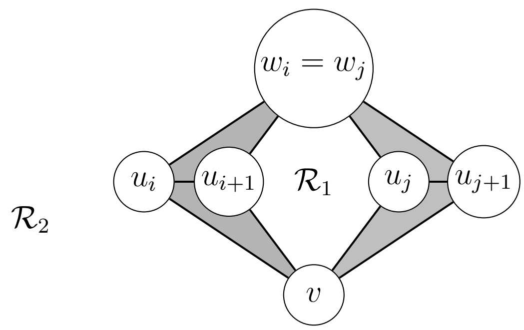


Figure 2.8: 4-element cutset $\{v, u_{i+1}, u_{j+1}, w_i\}$ in M_v when $w_i = w_j$. The shaded regions are unions of faces, so they have no additional vertices.

Now let v be a vertex of T with maximum degree, i.e., $d(v) = \Delta(T) = \Delta$. The mosaic graph M_v around v contains $2\Delta + 1$ vertices. If T contains a vertex that is not in M_v, then $2\Delta + 2 \leq n$, and the claimed inequality follows.

If every vertex of T is in M_v, then set of edges F not in M_v form a triangulation of the 2Δ-cycle $u_1w_1u_2w_2 \ldots u_\Delta w_\Delta u_1$ on the region different from $\mathcal{R}_v$; consequently $|F| = 2\Delta - 3$. Note that for any $1 \leq i < j \leq \Delta$, if $u_iu_j \in F$, then the 3-cycle u_iu_jv separates the vertices w_i and w_j, so u_i, u_j, v would be a cutset of size 3, a contradiction. If for any $j \notin \{i, i-1\}$, $u_iw_j \in F$, then the 4-cycle $u_iw_ju_jv$ has the vertices w_i and w_{i-1} on its different sides, giving a cutset of size 4, which is also a contradiction. Therefore every edge in F connects two vertices of $W = \{w_1, \ldots, w_\Delta\}$.

But then for every i, the edges $w_{i-1}u_i$ and u_iw_i lie on the boundary of the same face, giving $w_{i-1}w_i \in F$. Hence $w_1, \ldots, w_\Delta$ determines a Δ-gon (all of the sides are in F), and this Δ-gon is triangulated by the remaining edges of F. Lemma 18 applies. Say, w_{i-1}, w_i, w_{i+1} is a triangle with two edges on the boundary of the Δ-gon. Then $d(w_i) = 4$, contradicting the fact that T is 5-connected.

Finally, assume to the contrary that T has 13 vertices. Then $\Delta(T) \leq 5$, therefore T is 5-regular. The sum of degrees of T is odd, contradicting the Handshaking Lemma. $\square$

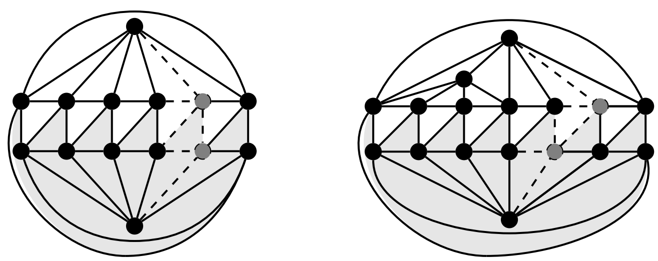


Figure 2.9: The triangulation T_n^5, which minimizes the Wiener index among all 5-connected triangulations of order $n = 2k \geq 12$ (left) and of order $n = 2k + 1 \geq 15$ (right). Gray vertices and dashed edges indicate the pattern to be repeated. The shaded region shows the mosaic graph around a degree $\lfloor \frac{n}{2} \rfloor - 1$ vertex.

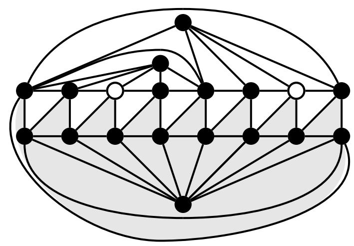


Figure 2.10: The 5-connected triangulation X. The two white vertices are at distance 4. The shaded region shows the mosaic graph around the degree 8 vertex.

For every $n \geq 12$, $n \neq 13$, the n-vertex triangulation T_n^5 is defined by Figure 2.9 (these will be our minimizers of the Wiener index). The following lemma is easy to verify and we leave the details to the reader.

Lemma 21. *Assume that $n \geq 12$, $n \neq 13$.*

(a) *T_n^5 is a 5-connected triangulation.*

(b) *T_n^5 has diameter 3.*

(c) *For n even, the degree sequence of T_n^5 is $\frac{n}{2}-1, \frac{n}{2}-1, 5, \ldots, 5$.*

(d) *For n odd, the degree sequence of T_n^5 is $\lfloor \frac{n}{2} \rfloor - 1, \lfloor \frac{n}{2} \rfloor - 2, 6, 6, 5, \ldots, 5$. (For $n = 15$, the terms in this sequence are* not *in decreasing order.)*

(e)

$$W(T_n^5) = \begin{cases} \frac{5n^2}{4} - 7n + 12, & \text{if } n \text{ is even,} \\ \frac{5n^2}{4} - 6n - \frac{9}{4}, & \text{if } n \text{ is odd.} \end{cases}$$

The 5-connected triangulation X of order 19, defined by Figure 2.10, has

$$W(X) = 335 = W(T_{19}^5). \tag{2.5}$$

We will also show that X is the *only* 5-connected triangulation that is not isomorphic to any T_n^5 and achieves the minimum Wiener index for its order. Note that as X is of diameter 4, the lower bound in Lemma 19 (f) cannot be used to compute $W(X)$. The different diameter, and also the different degree sequence, implies that $X \not\simeq T_{19}^5$.

We define the *extended mosaic graph* $M_v^\star$ by adding edges to M_v. Given a 5-connected triangulation G and a vertex v with mosaic graph M_v, we introduce the graph $M_v^\star$, on the vertex set of M_v, by setting

$$E(M_v^\star) = E(M_v) \cup \{w_i w_{i+1} : 1 \le i \le d(v), w_i w_{i+1} \in E(G)\}.$$

Note that $w_i w_{i+1} \in E(G)$ if an only if $d(u_{i+1}) = 5$. Let $\mathcal{R}_v^\star$ denote the extension of $\mathcal{R}_v$ by adding to it the faces bounded by the 3-cycles $u_{i+1} w_i w_{i+1}$ for all edges $w_i w_{i+1} \in E(G)$; let C_v denote the boundary cycle of $\mathcal{R}_v^\star$ and let $\mathcal{Q}_v^\star$ denote the other domain defined by the cycle C_v. Now all vertices of G that are not vertices of M_v and all edges of $E(G) \setminus E(M_v^\star)$ lie in the region $\mathcal{Q}_v^\star$ of the drawing of G.

We will use the following notation in the rest of the section. Given a 5-connected triangulation G and a vertex v, if a, b are vertices of C_v, then $P_v(a, b)$ denotes the path on the cycle C_v from a to b that follows the clockwise cyclic order. (So if $C_v = (w_1, w_2, \ldots, w_d)$ in clockwise cyclic order, then $P_v(w_1, w_2)$ is just the edge w_1w_2 with its endpoints, while $P_v(w_2, w_1)$ goes through all vertices of the cycle and misses only the edge w_1w_2.)

Lemma 22. *Let G be a 5-connected triangulation of order $n \geq 12$, and let v be a vertex of G with $d(v) = d$. Consider the extended mosaic graph $M_v^\star$. The following are true:*

(a) *Every vertex $z \in V(C_v)$ has an edge of $E(G) \setminus E(M_v^\star)$ incident upon it.*

(b) *If for some $z_1, z_2 \in V(C_v)$ we have $z_1z_2 \in E(G) \setminus E(M_v^\star)$, then z_1z_2 cuts $\mathcal{Q}_v^\star$ into two subregions, each containing a vertex of G in its interior, and $z_1, z_2 \in \{w_1, w_2 \ldots, w_d\}$.*

(c) *If n is even and $d(v) = \frac{n}{2} - 1$, then $G \simeq T_n^5$.*

Proof. Set $W = \{w_1, \ldots, w_d\}$ and $U = \{u_1, \ldots, u_d\}$.

(a): Observe that $W \subseteq V(C_v) \subseteq W \cup U$. Vertices in W have degree at most 4 in $M_v^\star$, and vertices of U have degree 5 in $M_v^\star$. If a vertex of U has degree 5 in G, then it is not a vertex of C_v. (a) follows.

(b): Let $z_1, z_2 \in V(C_v)$ where $z_1z_2 \in E(G) \setminus E(M_v^\star)$. Assume first that z_1z_2 cuts $Q_v^\star$ into two subregions, one of which (say $\mathcal{R}$) contains no vertices in its interior. We will show that $\mathcal{R}$ contains a (triangular) face f such that the boundary of f has two edges e_1, e_2 that are on the boundary of $\mathcal{R}$ and $z_1z_2 \notin \{e_1, e_2\}$. This is obviously true when the boundary of $\mathcal{R}$ is a 3-cycle. Otherwise the edges lying in the interior of $\mathcal{R}$ are edges of $E(G) \setminus E(M_v^\star)$ that form a triangulation of $\mathcal{R}$, and by Lemma 18 this triangulation contains two faces with two boundary edges on the boundary of

$\mathcal{R}$. One of these faces, f, does not have the edge z_1z_2 on its boundary. Let c be the common endvertex of the two edges e_1, e_2 on the boundary of $\mathcal{R}$. Then $c \notin \{z_1, z_2\}$ and c cannot have an edge from $E(G) \setminus E(M_v^\star)$ incident upon it, contradicting (a). So z_1z_2 cuts $Q_v^\star$ into two subregions, both of which contains a vertex in its interior. Now assume to the contrary that $\{z_1, z_2\} \cap U \neq \emptyset$. We may assume that $z_1 = u_1$. Then $z_2 = u_\ell$ for some $3 \leq \ell \leq d-1$ or $z_2 = w_j$ for some $3 \leq j \leq d-2$ ($j \neq 2$ and $j \neq d-1$, using Lemma 19 (b) for edges u_1u_2 and u_1u_d). If $z_2 = u_\ell$, then u_1vu_ℓ is a separating 3-cycle (as w_2 and w_d are in different regions of this cycle), and if $z_2 = w_j$ then $u_1w_ju_jv$ is a separating 4-cycle (as w_2 and w_d are in different regions); both of which contradict the 5-connectedness of G.

(c): Assume now that n is even and $d(v) = \frac{n}{2} - 1$. Lemma 20 gives $\Delta(G) = d(v)$. G has exactly one vertex, say x, not in M_v, and hence in the region $\mathcal{Q}_v^\star$. Then the already proven parts (a) and (b) imply that $E(G) \setminus E(M_v^\star) = \{z_1x : z_1 \in V(C_v)\}$. As $W \subseteq C_v$, $d(x) \geq |W| = d(v) = \Delta(G)$, we get $d(x) = \frac{n}{2} - 1$ and $W = C_v$, and each edge of the form w_iw_{i+1} is an edge of $M_v^\star$. (c) follows. $\square$

Theorem 23. *Assume that $n \geq 12$ and n is even. The triangulation T_n^5, which was defined in Figure 2.9, is the unique minimizer of the Wiener index among all 5-connected triangulations of order n.*

Proof. Let $n \geq 12$ be even and assume T is a 5-connected triangulation on $n = 2k$ vertices ($k \geq 6$). The degree sum of T is $2(3n-6) = 6n-12$, and Lemma 20 gives $\Delta(T) \leq \frac{n}{2} - 1$. By Lemma 13 the integer sequence $y_1, \ldots, y_n$ that sums to $6n - 12$, satisfies $5 \leq y_i \leq \frac{n}{2} - 1$ and has the largest sum of squares is the sequence $\frac{n}{2} - 1, \frac{n}{2} - 1, 5, 5, \ldots, 5$, which is exactly the degree sequence of T_n^5 by Lemma 21 (c). As T_n^5 has diameter 3, by Lemma 19 (f) T_n^5 indeed has the minimum Wiener index among all 5-connected n-vertex triangulations. We know that the degree sequence of T is the same as the degree sequence of T_n^5, so $T \simeq T_n^5$ follows from Lemma 22 (c). $\square$

To characterize extremal triangulations of odd order, we need a bit more information about their structure.

Lemma 24. *There are no* 5*-connected triangulations on* 21 *vertices with degree sequence* $8, 8, 8, 5, ..., 5$*.*

Proof. Assume that G is a 5-connected triangulation on 21 vertices with degree sequence $8, 8, 8, 5, \dots 5$. Let v be a degree 8 vertex, and let x_1, x_2, x_3, x_4 be the vertices in $V(G) \setminus V(M_v^\star)$. Set $X = \{x_1, x_2, x_3, x_4\}$, $U = \{u_i : 1 \leq i \leq 8\}$ and $W = \{w_i : 1 \leq i \leq 8\}$. Let b be the number of connected components of the subgraph of G induced by X and D_i be the component containig x_i, $c_i = |N(x_i) \cap X|$ and χ be the number of vertices of degree 8 in X.

Clearly $W \subseteq V(C_v)$, and by Lemma 22 (a) and (b) all vertices of $V(C_v) \cap U$ have degree 8 (consequently $|V(C_v) \cap U| \leq 2$). Assume that $z \in V(C_v) \cap U$; then $|N(z) \cap X| = 3$. Let $\{x_i, x_j, x_k\} = N(z) \cap X$, then without loss of generality $x_i x_j x_k$ is a path in G whose edges form faces with the edges $x_i z, x_j z, x_k z$. Moreover, if e, f are the two edges on C_v that are incident upon z, then the cyclic order of the edges that lie in or on the boundary of $\mathcal{Q}_v^\star$ around z is e, zx_i, zx_j, zx_k, f or f, zx_i, zx_j, zx_k, e; otherwise one of the triangles $zx_i x_j$ or $zx_j x_k$ is a separating triangle, which is a contradiction. Finally, $x_i x_k \notin E(G)$, as otherwise $zx_i x_k$ is a separating triangle.

Assume first that $U \cap V(C_v) \neq \emptyset$; without loss of generality $u_1 \in U \cap V(C_v)$, u_1 is adjacent to x_2, x_3, x_4 and $x_2 x_3 x_4$ is a path in G, consequently $x_2 x_4 \notin E(G)$. We have that either $b = 2$ and $D_1 = \{x_1\}$, or $b = 1$. Without loss of generality we may assume that the cyclic order of edges around u_1 in $\mathcal{Q}_v^\star$ is $u_1 w_1, u_1 x_2, u_1 x_3, u_1 x_4, u_1 w_8$. As $u_1 w_1, u_1 x_2$ (and also $u_1 x_4, u_1 w_8$) bound a common face, we have $w_1 x_2, w_8 x_4 \in E(G)$. Let $\mathcal{P}^\star$ be the region we get if we leave out from $\mathcal{Q}_v^\star$ the faces with u_1 on their boundary.

Consider the case when $|U \cap V(C_v)| = 2$, i.e. for some $j \neq 1$ the vertex u_j also has degree 8. If $N(u_j) \cap X = N(u_1) \cap X = \{x_2, x_3, x_4\}$ (which must happen when $b = 2$), we have that the path $x_2x_3x_4$ is induced in D_2. But then $u_1x_2u_jx_4$ is a separating 4-cycle, which is a contradiction. Therefore we must have $b = 1$, and without loss of generality for some $t \in \{2, 3\}$, $N(u_j) \cap X = \{x_1, x_t, x_{t+1}\}$ where $x_1x_t \in E(G)$ and $x_1x_{t+1} \notin E(G)$. This gives $|E(X)| \geq 3$, $|E(X,U)| = 6$, and (since the vertices of X have degree 5) $|E(X,W)| = 20 - 6 - 2|E(X)| \leq 8$. On the other hand, since $b = 1$, Lemma 22 (b) implies that the set of edges in $E(G) \setminus E(M_v^\star)$ that are incident upon W form the set $E(X,W)$, and (as w_1 and w_8 have degree 3 in $M_v^\star$) consequently $|E(X,W)| \geq 10$, a contradiction. Therefore we must have $|U \cap V(C_v)| < 2$, i.e. $U \cap V(C_v) = \{u_1\}$.

Now we have that $U \cap V(C_v) = \{u_1\}$. Assume that $b = 2$, so $D_1 = \{x_1\}$. Let s and t be the smallest and largest indices such that $w_s, w_t \in N(x_1)$; $1 \leq s < s + 4 \leq t \leq 8$. The path $w_sx_1w_t$ cuts the region $\mathcal{Q}_v^\star$ into two regions $\mathcal{Q}_1$ and $\mathcal{Q}_2$, where $\mathcal{Q}_1$ has u_1 on its boundary. Therefore x_2, x_3, x_4 lie inside $\mathcal{Q}_1$, $w_sw_t \in E(G) \setminus E(M_v^\star)$, and by Lemma 22 (b) for each $i : s \leq i \leq t$ $x_1w_i \in E(G)$. If $s \neq 1$ then $\{w_{s-1}, w_{s+1}, u_s, u_{s+1}, w_t, x_1\} \subseteq N(w_s)$, so w_s must have degree 8. If $s = 1$, then $\{u_1, u_2, w_2, x_2, w_t, x_1\} \subseteq N(w_s)$, so w_s has degree 8. Similar arguments imply that w_t also has degree 8. But then u_1, v, w_s, w_t all have degree 8, a contradiction. So we must have $b = 1$, i.e. x_1 is connected to at least one other vertex in X. If x_1 is connected to both x_2 and x_4, then (as the edges x_1x_2 and x_1x_4 lie in $\mathcal{P}^\star$) $u_1x_2x_1x_4$ is a separating 4-cycle, which is a contradiction. We may assume $x_1x_4 \notin E(G)$, so $c_1 \in \{1, 2\}$. Then $|E(X,W)| = 13 + 3\chi - 2c_1$. Since the number of degree 8 vertices in W is $1-\chi$, $|E(X,W)| = 10+3(1-\chi) = 13-3\chi$. This gives $2c_1 = 6\chi$, so 3 divides c_1, which is a contradiction. Thus we must have $U \cap V(C_v) = \emptyset$.

Therefore $W = V(C_v)$ and every vertex in W has degree 4 in $M_v^\star$, so every vertex in W has either one or 4 edges incident upon it from $E(G) \setminus E(M_v^\star)$. Set $F = E(W) \setminus E(M_v^\star)$ and $m_x = |E(X)|$. We have $2|F|+|E(W,X)| = \sum_{z \in W}(d(z)-4) = 14 - 3\chi$ and $|E(X,W)| = (\sum_{z \in X} d(z)) - 2m_x = 20 + 3\chi - 2m_x$.

If $\chi = 2$, then all vertices of W have at most one neighbor in X. Since the two vertices in X that have degree 8 have at least 5 neighbors in W, this implies that $5 + 5 \leq |W| = 8$, a contradiction. Therefore $\chi \in \{0, 1\}$; at least one vertex in W has degree 8.

Suppose $b = 1$. Then $3 \leq m_x \leq 5$, and $|E(X, W)| = 20 - 2m_x + 3\chi$. On the other hand by Lemma 22 (b) $F = \emptyset$, so $|E(X, W)| = 14 - 3\chi$. This gives $m_x = 3 + 3\chi$, consequently $\chi = 0$, $m_x = 3$, and exactly two vertices (say w_ℓ, w_q) in W have degree 8. But then at least one of the 4-cycles of the form $w_\ell x_i w_q x_j w_\ell$ is separating, which is a contradiction. Thus we must have $b > 1$.

Since $b > 1$, we must have either that the components spanned by X are a K_1 and a K_3, or the subgraph generated by X has exactly $4 - b$ edges (as all of its components are trees). In the former case $|E(X, W)| = 14 + 3\chi$, in the latter $|E(X, W)| = 12 + 2b + 3\chi \geq 16 + 3\chi > 14 - 3\chi \geq |E(X, W)|$, which is a contradiction. Therefore the components spanned by X are a K_1 and a K_3. We get that $2|F| + 14 + 3\chi = 14 - 3\chi$, which gives $\chi = 0$, and $F = \emptyset$. So exactly two vertices (say w_ℓ, w_q) in W have degree 8, and $E(W, V(G)) \setminus E(M_v^\star) = E(X, W)$. Without loss of generality the K_3 in X is formed by the vertices x_2, x_3, x_4. But then $X \subseteq N(w_\ell)$, so the subgraph generated by $\{x_2, x_3, x_4, w_\ell\}$ is a K_4. This is a contradiction, as one of the triangles $w_\ell x_2 x_3$, $w_\ell x_3 x_4$, $w_\ell x_2 x_4$ is separating, contradicting the 5-connectedness of G. $\square$

Lemma 25. *Let $n \geq 15$ be odd, and let G be a 5-connected triangulation of order n with degree sequence $d_1 \geq d_2 \geq d_3 \geq \ldots \geq d_n$. If $W(G) \leq W(T_n^5)$, then $d_1 = \lfloor \frac{n}{2} \rfloor - 1$, and one of the following holds:*

(a) *$n = 23$ and the degree sequence of G is $10, 8, 8, 5, \ldots, 5$.*

(b) *$d_2 \geq \lfloor \frac{n}{2} \rfloor - 2$, $d_3 + d_4 \leq 12$, and consequently $d_3 \leq 7$, $d_4 \leq 6$ and $d_5 = 5$.*

Proof. Set $n = 2k + 1$, then $k = \lfloor \frac{n}{2} \rfloor \geq 7$. As $d_1 \leq k - 1$, the only possible degree sequence for $k = 7$ is $(6, 6, 6, 5, \ldots, 5)$, which satisfies the conclusion. Hence we may assume $k \geq 8$. As T_n^5 has diameter 3, by Lemma 19 (f) and Lemma 21 (d) we must have

$$\sum_{x \in V(G)} d^2(x) \geq \sum_{x \in V(T_n^5)} d^2(x) = (k-1)^2 + (k-2)^2 + 2 \cdot 6^2 + 5^2(2k-3) = 2k^2 + 44k + 2.$$

Assume first that $d_1 \leq k-2$.

If $k = 8$, the only sequence possible is $6, 6, 6, 6, 6, 5, \ldots, 5$ whose sum of squares is $480 < 2 \cdot 8^2 + 44 \cdot 8 + 2$, which is a contradiction.

If $k = 9$, then

$$\sum_{x \in V(G)} d^2(x) \leq 3 \cdot 7^2 + 6^2 + 15 \cdot 5^2 = 558 < 560 = 2 \cdot 9^2 + 44 \cdot 9 + 2,$$

a contradiction.

If $k = 10$, then By Lemma 24

$$\sum_{x \in V(G)} d^2(x) \leq 2 \cdot 8^2 + 7^2 + 6^2 + 17 \cdot 5^2 = 638 < 642 = 2 \cdot 10^2 + 44 \cdot 10 + 2,$$

a contradiction.

If $k \geq 11$, then

$$\sum_{x \in V(G)} d^2(x) \leq 2(k-2)^2 + 8^2 + 5^2(2k-2) = 2k^2 + 42k + 22 < 2k^2 + 44k + 2,$$

a contradiction. So we proved that $d_1 = k-1$. If $k = 8$, the only sequences possible are $7, 7, 6, 5, \ldots, 5$ and $7, 6, 6, 6, 5 \ldots, 5$, which satisfy the conclusion. Hence we may assume $k \geq 9$.

Assume next that $d_2 \leq k-3$. If $k = 9$, the only sequence possible is $8, 6, 6, 6, 6, 5, \ldots, 5$ with degree square sum $558 < 2 \cdot 9^2 + 44 \cdot 9 + 2$. If $k = 10$,

$$\sum_{x \in V(G)} d^2(x) \leq 9^2 + 2 \cdot 7^2 + 6^2 + 17 \cdot 5^2 = 640 < 2 \cdot 10^2 + 44 \cdot 10 + 2.$$

If $k \geq 11$, then

$$\sum_{x \in V(G)} d^2(x) \leq (k-1)^2 + (k-3)^2 + 8^2 + 5^2(2k-2) = 2k^2 + 42k + 24 \leq 2k^2 + 44k + 2.$$

with equality only if $k = 11$ and the degree sequence is $10, 8, 8, 5, \ldots, 5$, i.e. G satisfies (a). Therefore we may assume that $d_2 \geq k - 2$.

As we already know that $d_1 = k - 1$ and $d_2 \geq k - 2$, using that $\sum d_i = 6(2k - 1)$, $d_3 + d_4 \leq 6(2k - 1) - (2k - 3) - 5(2k - 3) = 12$, so from $d_4 \leq d_3$ we get $d_4 \leq 6$ and from $d_4 \geq 5$ we get $d_3 \leq 7$. If $d_4 = 5$ then $d_5 = 5$, otherwise we have that $d_4 = d_3 = 6$ and $d_5 \leq 6(2k - 1) - (2k - 3) - 12 - 5(2k - 4) = 5$. $\square$

Lemma 26. *Let $n \geq 15$ be odd, $k = \lfloor \frac{n}{2} \rfloor$, and let G be a 5-connected triangulation of order n, with $W(G) \leq W(T_n^5)$. Let v be a vertex with $d(v) = \Delta(G) =: \Delta$. Consider the extended mosaic graph $M_v^\star$, and the sets $W = \{w_1, \ldots, w_\Delta\}$ and $U = \{u_1, \ldots, u_\Delta\}$. Let x_1, x_2 denote the two vertices not in $M_v^\star$. The following statements are true:*

(a) *$\Delta = k - 1$, at most 4 degrees of G are larger than 5, and for the largest 4 degrees $\Delta \geq d_2 \geq d_3 \geq d_4$ of G we have either $d_2 \geq k - 2$, $d_3 \leq 7$, $d_4 \leq 6$ and $d_3 + d_4 \leq 12$, or $n = 23$, $d_2 = d_3 = 8 = k - 3$, $d_4 = 5$.*

(b) *For $i \in \{1, 2\}$, there are vertices $a_i, b_i \in V(C_v)$ such that $N(x_i) \setminus \{x_{3-i}\} = V(P_v(a_i, b_i))$. (We refer to these a_i, b_i vertices in the forthcoming claims.) Furthermore, if $c \in V(P_v(a_1, b_1)) \cap V(P_v(a_2, b_2))$, then $c = a_1 = b_2$ or $c = a_2 = b_1$.*

(c) *$C_v = w_1 w_2 \ldots w_\Delta$, $d(x_i) \leq \Delta - 1$, and if $x_1 x_2 \notin E(G)$ then $d(x_i) \leq \Delta - 2$ and $n \geq 19$.*

(d) *If $x_1 x_2 \in E(G)$, then $G \simeq T_n^5$.*

(e) *If $x_1 x_2 \notin E(G)$ then $a_1 b_1, a_2 b_2 \in E(G) \setminus E(M_v^\star)$. Moreover, for $i \in \{1, 2\}$, if $c_i \in V(P_v(b_i, a_{3-i}))$ and $z \in V(G) \setminus \{x_1, x_2\}$, such that $c_i z \in E(G) \setminus E(M_v^\star)$, then $z \in V(P_v(b_{3-i}, a_i))$, and the neighbors of c_i in $P_v(b_{3-i}, a_i)$ form a consecutive sequence of vertices on this path.*

(f) *If $x_1 x_2 \notin E(G)$, then $a_1 = b_2$ or $a_2 = b_1$.*

(g) *If* $x_1x_2 \notin E(G)$*, then* $a_1 = b_2$ *and* $a_2 = b_1$*.*

(h) *If* $x_1x_2 \notin E(G)$*, then* $G \simeq X$ *and* $W(G) = W(T^5_{19})$*.*

Proof. Note that $n \geq 15$, so $k \geq 7$. Let G, v, x_1, x_2 be as in the conditions. Lemma 25 yields (a).

(b): Assume $i \in \{1, 2\}$. As $d(x_i) \geq 5$, x_i has at least 4 neighbors on C_v, so there are vertices $a_i, b_i \in N(x_i) \cap V(C_v)$ such that all vertices in $N(x_i) \setminus \{x_{3-i}\}$ lie on the path $P_v(a_i, b_i)$ and x_{3-i} does not lie in the interior of the subregion of $\mathcal{Q}^\star_v$ bounded by the cycle $x_iP_v(a_i, b_i)$. By Lemma 22 (b), no two vertices in $P_v(a_i, b_i)$ can be joined by an edge that is not in $M^\star_v$. As every vertex of C_v has at least one edge incident upon it from $E(G) \setminus E(M^\star_v)$, we must have $V(P_v(a_i, b_i)) \subseteq N(x_i)$, therefore $V(P_v(a_i, b_i)) = N(x_i) \setminus \{x_{3-i}\}$. As all edges incident upon x_1 or x_2 lie in the region $\mathcal{Q}^\star_v$ and do not cross, the two paths share at most their endvertices.

(c): Assume to the contrary that $c \in U \cap V(P_v(a_i, b_i))$. As $d(x_i) \geq 5$, $P_v(a_i, b_i)$ has at least 4 vertices. Therefore, there exists an internal vertex $c^\star$ of the path $P_v(a_i, b_i)$, such that the edge $cc^\star$ is an edge of this path. This means $c^\star \in W$ (see Lemma 22 (b)), $c^\star$ has at most 3 edges incident upon it in $E(M^\star_v)$, and the only edge in $E(G) \setminus E(M^\star_v)$ incident upon $c^\star$ is $c^\star x_i$, contradicting $d(c^\star) \geq 5$. Thus, $V(P_v(a_i, b_i)) \subseteq W$. By Lemma 22 (a) the internal vertices of the paths $P(b_i, a_{3-i})$ each have at least one edge of $E(G) \setminus E(M^\star_v)$ incident upon them. Thus, by the definition of a_i, b_i and $M^\star_v$, the internal vertices of the paths $P(b_i, a_{3-i})$ have to be adjacent to at least one other internal vertex of the paths $P(b_i, a_{3-i})$. Lemma 22 (b) implies that $V(P(b_i, a_{3-i})) \subseteq W$. So $V(C_v) = W$. By part (b), we have $|V(P_v(a_{3-i}, b_{3-i})) \setminus V(P_v(a_i, b_i))| \geq 2$. Hence $\Delta = |W| \geq |V(P_v(a_i, b_i))| + |V(P_v(a_{3-i}, b_{3-i})) \setminus V(P_v(a_i, b_i))|$. As $N(x_i) = V(P_v(a_i, b_i))$ or $V(P_v(a_i, b_i)) \cup \{x_{3-i}\}$, depending on whether x_1x_2 is non-edge or edge, the claimed upper bounds on $d(x_i)$ follow. Assume $x_1x_2 \notin E(G)$. As we have $5 \leq d(x_i) \leq \Delta - 2$, we have $\Delta \geq 7$, so $n \geq 17$. If $n = 17$, then we must have

$d(x_1) = d(x_2) = 5$. As $8 = d(x_1) + d(x_2) - 2 \leq |W| = 7$, this is a contradiction. (c) follows.

(d): Let $x_1x_2 \in E(G)$. Assume that $b_2 \neq a_1$. Then either $P_v(b_2, a_1)$ has an internal vertex c or $b_2a_1 \in E(G)$. In the first case, as $c \in C_v$, there is at least one edge in $E(G) \setminus E(M_v^\star)$ incident upon c by Lemma 22 (a). All edges of $E(G) \setminus E(M_v^\star)$ incident upon c must lie in the subregion of $\mathcal{Q}_v^\star$ bounded by the cycle $P_v(b_2, a_1)x_1x_2$. By Lemma 22 (b) these edges must be of the form cx_1 or cx_2. But none of these are edges of G, which is a contradiction. In the second case $b_2a_1 \in E(G)$, and the subregion of $\mathcal{Q}_v^\star$ bounded by the 4-cycle $b_2a_1x_1x_2$ has no vertices in its interior, so we must have either $x_1b_2 \in E(G)$ or $x_2a_1 \in E(G)$, a contradiction. So $a_1 = b_2$, and $a_2 = b_1$, and $V(C_v) = W$ by (c). All $w \in W$ are incident to 4 edges of $M_v^\star$, but $a_1, a_2 \in W$ are incident to 2 more edges, and vertices of $W \setminus \{a_1, a_2\}$ are incident to one more, so $d(a_1) = d(a_2) = 6$, and (a) gives $d_2 \geq k - 2$. Since a_1, a_2 and v are vertices with degree greater than 5, and G has at most 4 vertices with degree greater than 5, we get $\min(d(x_1), d(x_2)) = 5$, which gives that $G \simeq T_n^5$ as claimed.

For the remaining cases assume that x_1 and x_2 are not adjacent, so $n \geq 19$ and $\Delta \geq 8$. This also implies that $N(x_i) = V(P(a_i, b_i))$, so the paths $P(a_i, b_i)$ have at least 5 vertices.

(e): In this case the edges x_ia_i, x_ib_i lie on the boundary of the same face, so $a_ib_i \in E(G)$, and $a_ib_ix_i$ is a boundary of a face. Moreover, as $|V(P(a_i, b_i)| \geq 5$, $a_ib_i \notin E(M_v^\star)$. The rest of the statement is trivial if $a_1 = b_2$ and $a_2 = b_1$, so assume that is not the case. Consider the connected subregion $\mathcal{R}$ of $\mathcal{Q}_v^\star$ bounded by the cycle $P(b_1, a_2)P(b_2, a_1)$ (that has length at least 3 by the assumption); it has no vertices in its interior. Any edges between vertices of the cycle $P(b_1, a_2)P(b_2, a_1)$ are edges of this cycle or lie inside $\mathcal{R}$. This finishes the proof unless $a_1 \neq b_2$ and $a_2 \neq b_1$, so consider that to be the case. Let $c_i \in V(P_v(b_i, a_{3-i}))$ and $z \in V(G) \setminus \{x_1, x_2\}$ such that $c_iz \in E(G) \setminus E(M_v^\star)$. As c_i lies on the boundary of the connected subregion $\mathcal{R}$, Lemma 22 (b) gives that $z \in V(P(b_{3-i}, a_i))$, as claimed. Also, if $z_1, z_2 \in V(P(b_{3-i}, a_i))$ are different neighbors of c_i where $z_1 z_2$ is not an edge of the path $P_v(b_{3-i}, a_i)$, then by Lemma 22 (b) any internal vertex z_3 of the $z_1 - z_2$ subpath of $P(b_{3-i}, a_i)$ can only have the edge $z_3 c_i$ incident upon it from $E(G) \setminus E(M_v^\star)$. Since by Lemma 22 (a) z_3 must have an edge from $E(G) \setminus E(M_v^\star)$ incident upon it, (e) follows.

(f): Assume to the contrary that $a_1 \neq b_2$ and $a_2 \neq b_1$. By (e) we have $a_1 b_1, a_2 b_2 \in E(G) \setminus E(M_v^\star)$. Let $\mathcal{R}$ be the connected subregion of $\mathcal{Q}_v^\star$ bounded by the cycle $P_v(b_2, a_1) P_v(b_1, a_2)$. G has at most 4 vertices with degree greater than 5. As $d(v) = \Delta > 6$, $V(G) \setminus \{v\}$ has at most 3 vertices with degree greater than 5. In particular, C_v contains at most 3 vertices with degree greater than 5. As by (c) $V(C_v) = W$, each of a_1, a_2, b_1, b_2 has at least 4 edges incident upon them in $E(M_v^\star)$, and two edges incident upon them from $E(G) \setminus E(M_v^\star)$ (the edges $a_i b_i$, $a_i x_i$, $b_i x_i$). This gives that a_1, b_1, a_2, b_2 have degree at least 6, a contradiction. (f) follows.

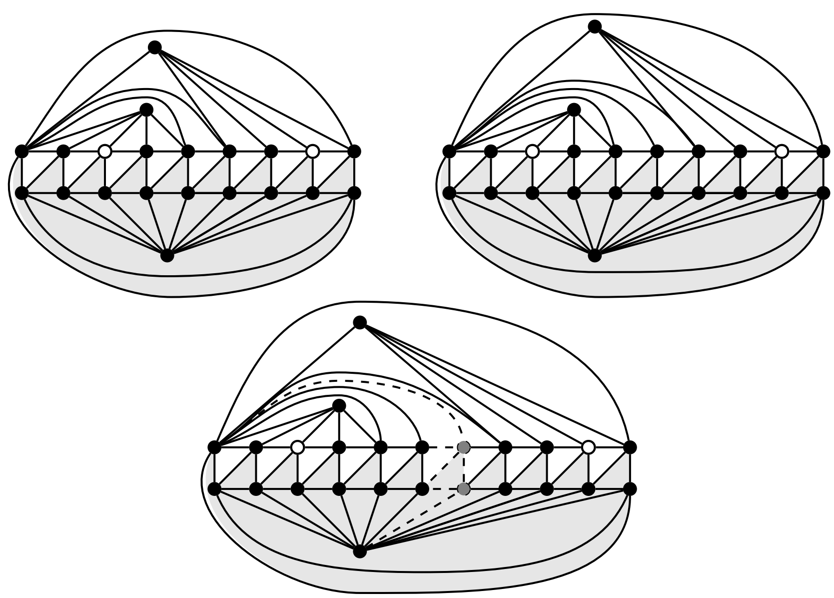


Figure 2.11: 5-connected triangulations of order $n = 21$, 23 and $n \geq 25$, which have the same degree sequence as T_n^5. The gray regions show the mosaic graphs around the vertex of degree $k-1$. The gray vertices and dashed edges on the triangulation of order 25 indicate the pattern to be repeated to get the construction for higher odd order. The two white vertices are at distance 4.

(g): Assume to the contrary that $a_1 \neq b_2$ or $a_2 \neq b_1$. By (f), without loss of generality we have $a_1 = b_2$ and $a_2 \neq b_1$. By definition $a_1x_1, a_1x_2 \in E(G)$ and by (e) all vertices of $P_v(b_1, a_2)$ are neighbors of a_1. All other neighbors of a_1 are one of the 4 neighbors of a_1 in $M_v^\star$. Consequently $d(a_1) = 6 + |V(P_v(b_1, a_2))| \geq 8$, so $d(a_1) = d_2 \in \{k-1, k-2, k-3\}$, and if $d(a_1) = k-3$, then G contains no degree 6 vertices. As $V(C_v) = W$, we must have $d(a_2) = d(b_1) = 6$, as a_2, b_1 each have 4 neighbors in $M_v^\star$, and both are joined to $a_1 = b_2$, and to a single x_i, and not joined to anything else. By (a) $d(a_1) \geq k-2$, and, as v, a_1, a_2, b_1 are the 4 vertices of degree greater than 5, all other vertices (including x_1 and x_2) have degree 5. So every $w \in W \setminus \{a_1, a_2, b_1\}$ has 4 neighbors in $M_v^\star$, and is joined by an edge in $E(G) \setminus E(M_v^\star)$ to exactly one of the vertices x_1, a_1, x_2, and the paths $P(a_i, b_i)$ have 5 vertices each. As the sum of degrees is $6n - 12 = k - 1 + d(a_1) + 12 + 5(n-4)$, we get $d(a_1) = k-2$. As $d(a_1) \geq 8$, this gives $n \geq 21$. Figure 2.11 has the graph G for all $n \geq 21$. Since G has the same degree sequence as T_n^5 and $W(G) \leq W(T_n^5)$, by Lemma 19 (f) we must have $W(G) = W(T_n^5)$ and the diameter of G is at most 3. However, G has diameter at least 4, as demonstratred on Figure 2.11, a contradiction. (g) follows.

(h): By (g), $a_1 = b_2$ and $a_2 = b_1$. By (e) $a_1a_2 \in E(G)$. By (c) $V(C_v) = W$, and any edge from $E(G) \setminus E(M_v^\star)$ incident upon a vertex $w \in W \setminus \{a_1, a_2\}$ connects w to exactly one of x_1, x_2. Each a_i has 4 incident edges in $E(M_v^\star)$, and in addition, it is joined to exactly 3 more vertices: x_1, x_2, a_{3-i}. So $d(a_1) = d(a_2) = 7 = d_2 = d_3$. By (a), $d_3 + d_4 \leq 12$, consequently all vertices of $V(G) \setminus \{v, a_1, a_2\}$ (including x_1 and x_2) have degree 5. As $6n - 12 = k - 1 + 14 + 5(n-3)$, $n = 19$. We have that $G \simeq X$ and $W(G) = W(T_{19}^5)$ (see Figure 2.10). □

The following theorem now follows:

Theorem 27. *Let $n \geq 15$ be odd. If $n \neq 19$, then the unique minimizer of the Wiener index among 5-connected triangulations of order n is T_n^5. If $n = 19$, then there are precisely two minimizers, T_{19}^5 and X.*

Chapter 3

An Infinite Antichain of Planar Tanglegrams

3.1 History and Background

A tanglegram is an important data structure that represents coevolution and cospeciation in the biological sciences ([31]). In particular, bioinformaticians believe that a parameter called the tangle crossing number correlates with a number of parameters of import in various evolutionary models. Let me formally define these terms.

Definition 19. A *rooted tree* T is a tree with a distinguished vertex called the root. Given a vertex v in a rooted tree, and a neighbor y of v, we say that y is the parent of v if y is on the path from v to the root. Otherwise we say that y is a child of v. The rooted tree T is *binary* if every vertex has zero or two children.

Definition 20. A *tanglegram* of size n is an ordered triplet $\mathcal{T} = (T_1, T_2, M)$, where T_1 and T_2 are rooted binary trees with n leaves each, and M is a perfect matching between the two leaf sets. We will call T_1 the left tree and T_2 the right tree of the tanglegram. Two tanglegrams are considered the same if there is a graph isomorphism between them that fixes the roots of the left tree and the right tree.

In graph theory, we visualize and often identify a graph with a drawing or embedding of the graph in the plane. We will do the same for tanglegrams, but the embedding must follow some rules not normally present in the more general context of graphs. We call such an embedding a *layout*.

Definition 21. A *plane binary tree* is a rooted binary tree, in which the children of internal vertices are specified as left and right children. A plane binary tree is easy to draw on one side of a line, without edge crossings, such that only the leaves of the tree are on the line. We will say that the plane binary tree P is a *plane tree of the rooted binary tree* T, if P is isomorphic to T as a graph.

Definition 22. A *layout* (L, R, M) of the tanglegram $\mathcal{T} = (T_1, T_2, M)$ is given by a left plane binary tree L isomorphic to T_1, drawn in the halfplane $x \leq 0$, having its leaves on the line $x = 0$, a right plane binary tree R isomorphic to T_2 drawn in the halfplane $x \geq 1$, having its leaves on the line $x = 1$, and the perfect matching M between their leaves drawn in straight line segments. (See Figure 3.1.)

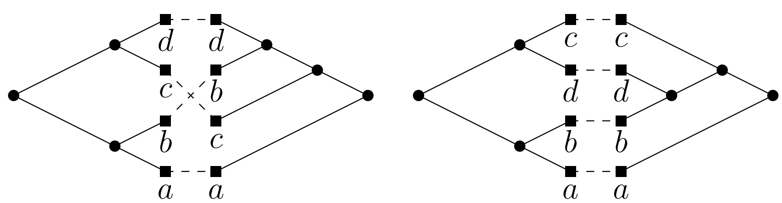


Figure 3.1: Two layouts of the same tanglegram. The leaf labels help to show that the matching was preserved under the isomorphisms of the left and right tree from one layout to the other. The layout on the right shows that the tanglegram that these layouts correspond to is planar.

Definition 23. The *tangle crossing number* of a tanglegram is the minimum crossing number (i.e., the minimum number of unordered crossing edge-pairs) among all of its layouts. The tanglegram is *planar* if it has a layout without any crossings.

In the biological sciences, one might consider the situation where the left tree is the phylogenetic tree of a collection of hosts, the right tree is the phylogenetic tree of their parasites, and the matching connects each host with its corresponding parasite, e.g. gophers and lice ([23]). In this case, the tangle crossing number has been related to the number of times that parasites switched hosts throughout the evolutionary process ([23]). Another interesting example was considered in [5] (pp. 204-206), where the left and right trees were considered to be gene trees rather than phylogenetic trees. In this case, the tangle crossing number correlates to the number of horizontal gene transfers. Tanglegrams have been well studied in the disciplines of phylogenetics and computer science (see e.g., [2], [3], [4], [12], [18], [25], [28], [33]).

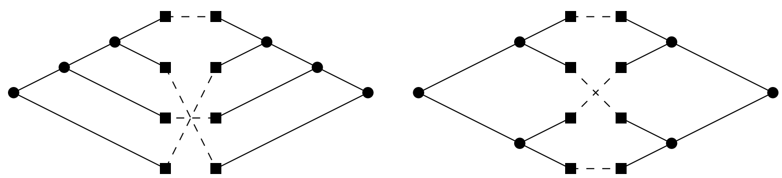
Let $\mathcal{T} = (T_1, T_2, M)$ be a tanglegram of size n, and suppose that the vertices of T_1 and T_2 are labeled. Let's count the number of plane trees of T_1 with respect to this labeling. There are $n-1$ internal vertices of T_1, and for each internal vertex, there are 2 choices for which child is its left child (forcing the remaining child to be its right child). Therefore, there are 2^{n-1} different plane trees of T_1 with respect to the specified labeling. The same result is true for T_2. For each layout of $\mathcal{T}$, we must choose a plane tree of T_1 and T_2 independently, so there are $2^{n-1} \cdot 2^{n-1} = 2^{2n-2}$ different layouts of $\mathcal{T}$. As the tangle crossing number is the minimum crossing number among *all* layouts of a given tanglegram, it is easy to imagine that this quantity is difficult to compute. In fact, computing the tangle crossing number was shown to be NP-hard ([18]), even when both trees are complete binary trees ([4]). This problem was, however, shown to be fixed-parameter tractable ([4]) , i.e., if we fix $k \geq 1$, there is a polynomial-time algorithm (polynomial in the size of the tanglegram) to determine if the tangle crossing number of a given tanglegram is at least k.

Czabarka, Székely, and Wagner ([13]) discovered a Kuratowski-like theorem that characterizes planar tanglegrams by two excluded induced subtanglegrams. To state this more precisely, we need a few more definitions.

Definition 24. Given a rooted binary tree T with root r and a non-empty subset B of its leaves, the *rooted binary subtree induced by* B, $T[B]$, is obtained as follows: Take the smallest subtree T' of T containing all vertices of B, and designate the vertex $\rho \in V(T')$ closest to r in T as the root of T'. This rooted tree is not necessarily binary, so we further suppress all vertices of degree 2 (except ρ) in T' to make it binary. The resulting rooted binary tree is $T[B]$.

Definition 25. Given a tanglegram $\mathcal{T} = (T_1, T_2, M)$ and a nonempty subset M' of M, the *subtanglegram induced by* M' is $\mathcal{T}[M'] = (T_1[B_1], T_2[B_2], M')$, where B_i is the set of leaves in T_i matched by M'. We say that $\mathcal{T}^*$ is an *induced subtanglegram of* $\mathcal{T}$ if there exists some $M^* \subseteq M$ such that $\mathcal{T}^* = \mathcal{T}[M^*]$. In this case, we write $\mathcal{T}^* \preceq \mathcal{T}$.

Theorem 28 (Czabarka, Székely, and Wagner [13]). *A tanglegram is planar unless is contains $\mathcal{T}_1$ or $\mathcal{T}_2$ as an induced subtanglegram (see Figure 3.2).*



$\mathcal{T}_1$ $\mathcal{T}_2$

Figure 3.2: The two nonplanar tanglegrams that form the characterization in Theorem 28

Theorem 28 provides a natural polynomial-time algorithm to determine if a tanglegram is planar or, equivalently, if a tanglegram has tangle crossing number at least 1. This result led to the following questions:

(i) For $k \geq 2$, is there a similar finite characterization for tanglegrams with tangle crossing number at least k?

(ii) For $k \geq 3$, is there a similar finite characterization for tanglegrams that have at least k pairwise crossing edges in every layout?

Note that $\preceq$ is a partial order on the set of tanglegrams, and that $\preceq$ is well-founded, i.e., it has no infinite strictly decreasing chains. This is easy to see as a proper subtanglegram of $\mathcal{T} = (T_1, T_2, M)$ must have a matching M' which satisfies the inequality $1 \leq |M'| < |M|$. In other words, if $\mathcal{T}'$ is a proper subtangle of $\mathcal{T}$, then $\mathcal{T}'$ is strictly smaller in size than $\mathcal{T}$. Furthermore, the tanglegram which consists of left and right trees with only one vertex each and an edge between them is the unique minimum element of the set of all tanglegrams with respect to $\preceq$.

Lemma 29. *Consider the general context where $(X, \trianglelefteq)$ is a well-founded partial order. Furthermore, let $\emptyset \neq U \subseteq X$ be an upward closed set and $M \subseteq U$ be the set of all minimal elements of U with respect to $\trianglelefteq$. Then:*

(a) *U is generated by M.*

(b) *If N generates U, then $M \subseteq N$.*

(c) *M forms an antichain.*

(d) *The poset $(X, \trianglelefteq)$ has no infinite antichain if and only if every upset is finitely generated.*

Proof. (a) Let $u \in U$ be arbitrary. We must show that there exists some $m \in M$ such that $m \trianglelefteq u$. If u is minimal, then $u \in M$, and we're done. Otherwise, there exists $u_1 \in U$ such that $u_1 \trianglelefteq u$ and $u_1 \neq u$. If u_1 is minimal, then $u_1 \in M$, and we are done. Otherwise, there exists $u_2 \in U$ such that $u_2 \trianglelefteq u_1$ and $u_2 \neq u_1$. Continue this process inductively. Since $u, u_1, u_2, u_3, \ldots$ forms a strictly decreasing chain and $(X, \trianglelefteq)$ is well-founded, this process must end after finitely many steps, i.e. there exists some $k \geq 1$ such that $u_k \in M$ and $u_k \trianglelefteq u$.

(b) Suppose that N generates U and that $x \in M$ is arbitrary. As $x \in U$ and N generates U, there exists some $y \in N$ such that $y \trianglelefteq x$. But x is minimal, so it must be the case that $y = x$, i.e. that $x \in N$.

(c) Suppose that $x, y \in M$ with $x \neq y$. Then since both x and y are minimal, $x \ntrianglelefteq y$ and $y \ntrianglelefteq x$, i.e. x and y are incomparable. Since x and y were arbitrary, M forms an antichain.

(d) First suppose that $(X, \trianglelefteq)$ has no infinite antichain. Then the result that every upset is finitely generated follows from (a), (b), and (c).

Suppose conversely that $(X, \trianglelefteq)$ has an infinite antichain, A, and let B be the upset generated by A. We will see that A is precisely equal to the set of minimal elements of B. By definition, if $b \in B$, then there exists some $a \in A$ such that $a \leq b$. Furthermore, suppose there exists some $b \in B$ such that $b \leq a_1$ for some $a_1 \in A$. Since $b \in B$ and A genereates B, there exists some $a_2 \in A$ such that $a_2 \leq b$. But then, by transitivity, $a_2 \leq a_1$. Since A is an antichain, we must have $b = a_1 = a_2$. Therefore, A is the set of minimal elements of B, as claimed. By part (b), A is a minimal generating set of B, so B is not finitely generated.

$\square$

The sets in consideration in questions (i) and (ii) both form upward closed sets with respect to $\preceq$ for each fixed value of k. As per the folklore result presented above as Lemma 29, showing that there are no infinite antichains in the partial order of tanglegrams with the induced subtanglegram relation would result in an affirmative answer to both questions for all values of k, and would result in a number of algorithmic consequences. Unfortunately, in Section 3.2 we will explicitly construct such an infinite antichain. The existence of this antichain is somewhat surprising because Kruskal's Tree Theorem ([26]) states that the partial order of rooted binary trees with respect to the induced subtree relation is a well partial order. Note that the existence of the infinite antichain shown in Section 3.2 does not imply a negative answer to questions (i) and (ii); it just cuts off the elegant proof technique outlined above.

The results presented in the remainder of this chapter are joint work with Éva Czabarka and László Székely.

3.2 The Infinite Antichain

Here we construct the infinite antichain mentioned in Section 3.1. We actually consider a subposet of the partial order of tanglegrams with the induced subtanglegram relation. In particular, we restrict ourselves to the set of tanglegrams in which the left and right trees are both rooted caterpillars of size n.

Definition 26. For $n \geq 2$, the *rooted caterpillar* C_n with n leaves is the rooted binary tree, whose $n-1$ internal vertices form a path, and the root is an endvertex of this path. Figure 3.3 shows C_4 as an example.

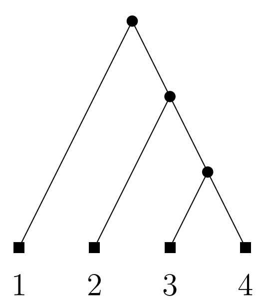


Figure 3.3: The unique rooted caterpillar C_4. The $n = 4$ leaves are labeled so that every label is equal to the distance from the leaf to the root except for the leaf labeled 4. The leaf labelled 4 is distance $3 = n - 1$ from the root.

Definition 27. A *catergram* of size n is a tanglegram (T_1, T_2, M) such that T_1 and T_2 are both isomorphic to C_n.

An important fact about rooted caterpillars is that any induced subtree of one is again a rooted caterpillar. This means in particular that the set of catergrams is closed under taking induced subtanglegrams, so the set of catergrams under $\preceq$ is a subposet of the set of all tanglegrams under $\preceq$. It will be helpful to label the leaves of the left and right rooted caterpillar of a catergram in a particular way. Note that, as is customary, $[n]$ denotes the set $\{1, 2, 3, \ldots, n\}$ and S_n denotes the symmetric group acting on $[n]$. For $\pi \in S_n$, we use the notation $\pi = (a_1, \ldots, a_n)$, if $\pi(i) = a_i$ for all $i \in [n]$.

Definition 28. For $n \geq 2$, the *distance labeling* of the leaves of C_n is the following: for each i, $1 \leq i \leq n-2$, the leaf labeled i is the one at distance i from the root, and the two leaves at distance $n-1$ are labeled arbitrarily by $n-1$ and n (see Figure 3.3).

In doing so, we can uniquely identify a catergram of size n with a permutation on $[n]$. This makes the large body of research that has been conducted on permutations applicable to the problem of finding an infinite antichain.

Definition 29. For $n \geq 2$ and $\pi \in S_n$, the *catergram* $\mathcal{T}_\pi$ is the tanglegram (C_n, C_n, M_π), where M_π is defined as follows: Using the distance labeling of the leaves of both caterpillars, match the leaf on the left tree labeled i with the leaf on the right tree labeled j if and only if $\pi(i) = j$ (see Figure 3.4).

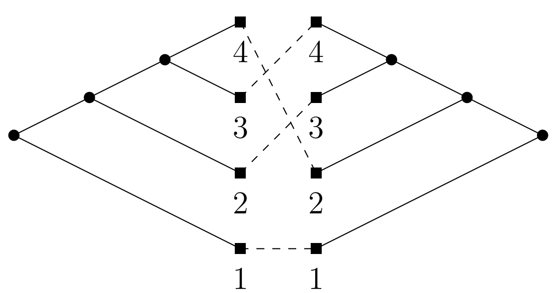


Figure 3.4: The catergram $\mathcal{T}_\pi$, where $\pi = (1, 3, 4, 2)$.

It is important to point out that, while every permutation defines a unique catergram, the same is not true in reverse; for each catergram, there is a set of permutations that define it as in Definition 29. The reason the set of catergrams of size n are not in one-to-one correspondence with the elements of S_n stems from the fact that in Definition 28, both leaves labeled n and $n-1$ are distance $n-1$ from the root. It will be helpful to classify which permutations define the same catergram.

Definition 30. Given a tanglegram $\mathcal{T} = (T_1, T_2, M)$, where the root of T_i is r_i, the *multiset of distance pairs*, $\mathbb{D}(\mathcal{T})$, contains exactly k copies of (d_1, d_2) if and only if there exists exactly k matching edges of the form $(x_1, x_2) \in M$ such that x_i is a leaf of T_i at distance d_i from r_i.

If two catergrams have the same distance pair multiset, then they are the same as tanglegrams.

Definition 31. Assume $n \geq 2$. Given a permutation $\pi = (a_1, \ldots, a_n) \in S_n$, we define the (not necessarily different) permutations $\widehat{\pi}$, $\widetilde{\pi}$ as

$$\widehat{\pi}(i) = \begin{cases} a_i, & \text{if } i \leq n-2 \\ a_n, & \text{if } i = n-1 \\ a_{n-1}, & \text{if } i = n \end{cases} \quad \text{and} \quad \widetilde{\pi}(i) = \begin{cases} a_i, & \text{if } a_i \notin \{n-1, n\} \\ n-1, & \text{if } a_i = n \\ n, & \text{if } a_i = n-1; \end{cases},$$

and finally let $\pi^* = \widehat{(\widetilde{\pi})}$. We define the set $X_\pi = \{\pi, \widehat{\pi}, \widetilde{\pi}, \pi^*\}$.

Lemma 30. *Let $\pi = (a_1, \ldots, a_n) \in S_n$ be arbitrary. The following facts are obvious and are presented without proof:*

(a) *We have $\mathbb{D}(\mathcal{T}_\pi) = \{(1, a_1^*), (2, a_2^*), \ldots, (n-1, a_{n-1}^*), (n-1, a_n^*)\}$, where*

$$a_i^* = \begin{cases} a_i, & a_i < n \\ n-1, & a_i = n. \end{cases}$$

(b) $\widehat{(\widetilde{\pi})} = \widetilde{(\widehat{\pi})}$, $\pi = \widehat{(\widehat{\pi})} = \widetilde{(\widetilde{\pi})}$, *and* $\pi \notin \{\widehat{\pi}, \widetilde{\pi}\}$.

(c) $\rho \in X_\pi \Leftrightarrow X_\rho = X_\pi$.

(d) $\widehat{\pi} = \widetilde{\pi} \Leftrightarrow \{a_{n-1}, a_n\} = \{n-1, n\} \Leftrightarrow \pi = \pi^*$; *consequently* $|X_\pi| \in \{2, 4\}$.

(e) $\mathcal{T}_\rho = \mathcal{T}_\pi \Leftrightarrow \mathbb{D}(\mathcal{T}_\rho) = \mathbb{D}(\mathcal{T}_\pi) \Leftrightarrow \rho \in X_\pi$.

Now we are in a position to utilize some of the work done on permutations. In particular, we will consider the partial order of permutations under the *pattern* relation.

Definition 32. We say that two sequences of n numbers, $(a_1, \ldots, a_n), (b_1, \ldots, b_n) \in \mathbb{R}^n$, are *order isomorphic* if for all $i, j \in [n]$, we have $a_i < a_j$ iff $b_i < b_j$. Given a permutation $\pi \in S_n$ and non-empty subset $A \subseteq [n]$, where $a_1, \ldots, a_k$ lists the elements of A in increasing order, we denote by $\pi[A]$ the permutation in $S_{|A|}$ that is order isomorphic to $(\pi(a_1), \pi(a_2), \ldots, \pi(a_k))$. If $\rho \in S_m$ and $\pi \in S_n$, then we say that ρ is a *pattern in* π if $\pi[A] = \rho$ for some $A \subseteq [n]$. In this case, we write $\rho \leq \pi$.

Definition 33. Assume $\pi \in S_n$ and $\emptyset \neq A \subseteq [n]$. Then (with a slight abuse of notation) we denote by $\mathcal{T}_\pi[A]$ the induced subtanglegram $\mathcal{T}_\pi[M^*]$, where M^* is the matching containing edges of M incident upon leaves of the left tree that are labeled with elements of A.

Lemma 31. *The following statements hold:*

(a) *Let v be a leaf of C_n at distance i from the root r of C_n, and $y \neq v$ be another leaf that is at distance j from r. Let T be the binary tree induced by all leaves except v (so $T = C_{n-1}$) with root r^*. Then y is a leaf in T, and the distance of y from r^* is j if $j < i$, and $j - 1$ otherwise.*

(b) *For any $\pi \in S_n$ and non-empty $A \subseteq [n]$, we have $\mathcal{T}_\pi[A] = \mathcal{T}_{\pi[A]}$.*

(c) *Let $m \leq n$. For $\rho \in S_m$ and $\pi \in S_n$, we have $\mathcal{T}_\rho \preceq \mathcal{T}_\pi \Leftrightarrow \mathcal{T}_\rho = \mathcal{T}_{\pi[A]}$ for some $A \subseteq [n] \Leftrightarrow \sigma \leq \pi$ for some $\sigma \in X_\rho$.*

Proof. (a) is obvious.

(b) follows by strong induction on $k := |[n] \setminus A|$: If $k = 0$, then $A = [n]$ and the statement is trivial. If $k = 1$, then $A = [n] \setminus \{j\}$ for some $j \in [n]$, and the statement follows from (a). If $k > 1$, then set $j := \max([n] \setminus A)$, $B = A \cup \{j\}$,

$\ell := 1 + |\{a \in A : a < j\}|$ and $C = [|B|] \setminus \{\ell\}$. It is easy to see that $|B| = |A| + 1 < n$, $|C| = |B| - 1$, $\pi[A] = (\pi[B])[C]$, and $\mathcal{T}_\pi[A] = (\mathcal{T}_\pi[B])\,[C]$, so the statement follows from the induction hypothesis.

To see (c): Let $m \leq n$, $\rho \in S_m$ and $\pi \in S_n$. $\mathcal{T}_\rho \preceq \mathcal{T}_\pi$ iff $\mathcal{T}_\rho = \mathcal{T}_{\pi[A]}$ for some $A \subseteq [n]$ follows from the definitions and (b). By Lemma 30(e), the latter is equivalent with $\sigma \leq \pi$ for some $\sigma \in X_\rho$. $\square$

It is now clear how the poset of catergrams with the induced subtanglegram relation is related to the poset of permutations with the pattern relation. Laver [27], Pratt [32], Tarjan [38], and Spielman and Bóna [36] constructed infinite antichains of permutations for the partial order defined by permutation patterns, and we will build off of the results of Spielman and Bóna to construct the infinite antichain in our context. A first natural guess might be that the sequence of catergrams generated by the Spielman and Bóna permutations would themselves generate an antichain in the partial order defined by induced subtanglegrams. The exact opposite turns out to be true; these tanglegrams form a chain.

Definition 34 (Spielman and Bóna, [36])**.** For $i \in \mathbb{Z}^+$, set $\pi_i \in S_{[12+2i]}$ as

$$(\pi_i(1), \pi_i(2), \pi_i(3), \pi_i(4)) = (11 + 2i, 10 + 2i, 8 + 2i, 12 + 2i),$$
$$(\pi_i(9 + 2i), \pi_i(10 + 2i), \pi_i(11 + 2i), \pi_i(12 + 2i)) = (3, 2, 1, 5),$$

and for $j : 5 \leq j \leq 8 + 2i$,

$$\pi_i(j) = \begin{cases} 11 + 2i - j, & \text{if } j \text{ is odd} \\ 15 + 2i - j, & \text{if } j \text{ is even.} \end{cases}$$

So for example, the first two permutations in the Spielman and Bóna sequence are:

$$\begin{aligned} \pi_1 &= (13, 12, 10, 14, 8, 11, 6, 9, 4, 7, 3, 2, 1, 5) \\ \pi_2 &= (15, 14, 12, 16, 10, 13, 8, 11, 6, 9, 4, 7, 3, 2, 1, 5). \end{aligned}$$

Theorem 32. $\{\mathcal{T}_{\pi_i} : i \in \mathbb{Z}^+\}$ *is an infinite chain in the induced subtanglegram partial order.*

Proof. Let $i \geq 1$ be fixed, but arbitrary. According to Lemma 31(c), in order to show that $\mathcal{T}_{\pi_i} \preceq \mathcal{T}_{\pi_{i+1}}$, we need only show that there exists some $\sigma \in X_{\pi_i}$ such that $\sigma \leq \pi_{i+1}$. Set $A_i = [14+2i] \setminus \{2,4\}$. One can easily verify that $\widetilde{\pi_i} = \pi_{i+1}[A_i]$ (i.e., that $\widetilde{\pi_i} \leq \pi_{i+1}$), which proves the theorem. $\square$

We cannot use the permutations of Spielman and Bóna directly, but if we turn them "upside down", the resulting permutations generate a sequence of catergrams that do form an antichain. By turning π_i "upside down", we mean that for each j from 1 to $12+2i$, we replace $\pi_i(j)$ with $13+2i-\pi_i(j)$. We call the resulting permutation ρ_i. The two permutations π_i and ρ_i are related as follows: The smallest output of π_i is the largest output of ρ_i, the seond smallest output of π_i is the second largest output of ρ_i, etc.

Definition 35. For $i \in \mathbb{Z}^+$, we set $\rho_i \in S_{[12+2i]}$ as

$$(\rho_i(1), \rho_i(2), \rho_i(3), \rho_i(4)) = (2,3,5,1),$$
$$(\rho_i(9+2i), \rho_i(10+2i), \rho_i(11+2i), \rho_i(12+2i)) = (10+2i, 11+2i, 12+2i, 8+2i),$$

and for $j : 5 \leq j \leq 8+2i$,

$$\rho_i(j) = \begin{cases} j+2, & \text{if } j \text{ is odd} \\ j-2, & \text{if } j \text{ is even.} \end{cases}$$

So for example, the first two permutations in our sequence will be

$$\begin{aligned} \rho_1 &= (2,3,5,1,7,4,9,6,11,8,12,13,14,10) \\ \rho_2 &= (2,3,5,1,7,4,9,6,11,8,13,10,14,15,16,12). \end{aligned}$$

We are now ready to present the main result.

Theorem 33. $\{\mathcal{T}_{\rho_i} : i \in \mathbb{Z}^+\}$ *is an antichain with respect to the relation* $\preceq$.

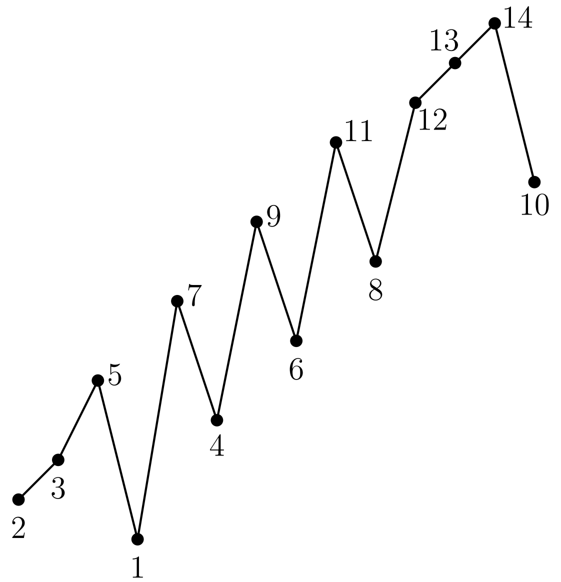
Permutations may be visualized as functions drawn in the standard cartesian plane with line segements between successive points to help parse the order of the points. While the proof of Theorem 33 below is completely general, it is helpful to consider and refer to the situation for specific permutations (say ρ_1 and ρ_3). See Figure 3.5 for the permutations of X_{ρ_1} and Figure 3.6 for ρ_3. These serve as useful archetypes.

Proof. In the proof we will use the fact that for any k and any $\gamma \in X_{\rho_k}$, the permutation γ has exactly two entries that are preceded by at least 3 larger elements: the entry 1 and the entry $8+2k$; moreover, if $\gamma \in \{\rho_k, \widetilde{\rho_k}\}$ then $8+2k$ is preceded by exactly 4 larger elements, but these 4 elements are not order isomorphic in ρ_k and $\widetilde{\rho_k}$.

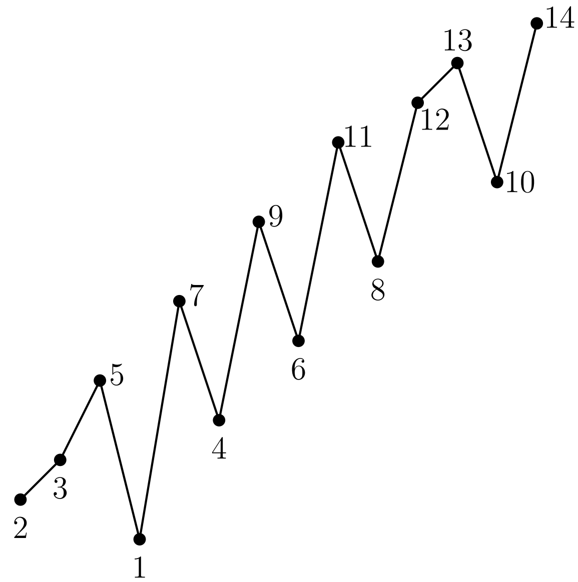
By Proposition 31(c), it is sufficient to show that for any $i < j$ and for any $\sigma \in X_{\rho_i}$, $\sigma \not< \rho_j$. By our starting remark, if $\sigma < \rho_j$, then the entries 1 and $8+2i$ in σ should map to the entries 1 and $8+2j$ in ρ_j, and the preceding larger elements must map to preceding larger entries; consequently $\widetilde{\rho_i} \not< \rho_j$. As $8+2j$ is the last entry of ρ_j, but not of $\widehat{\rho_i}$ or ρ_i^* (unless $\rho_i^* = \rho_i$), we get that $\widehat{\rho_i} \not< \rho_j$ and $\rho_i^* \not< \rho_j$. So what remains to be shown is $\rho_i \not< \rho_j$, which was essentially stated and proved in [36], but for completeness, we include a (somewhat different) proof here.

Suppose by way of contradiciton that $\rho_i < \rho_j$, i.e., entries of ρ_i map to entries of ρ_j in an order preserving fashion. By our earlier remarks, the first 4 elements of ρ_i must map to the first 4 elements of ρ_j and the last 6 elements of ρ_i must map to the last 6 elements of ρ_j, so we must map the sequence $(7, 4, 9, 6, \ldots, 7+2i, 4+2i)$ to $(7, 4, 9, 6, \ldots, 7+2j, 4+2j)$ by leaving out $2(j-i) \geq 2$ elements.

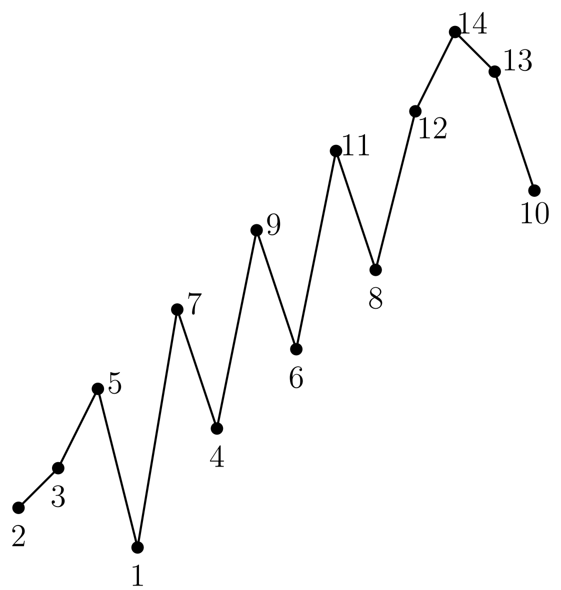
Some observations: Let x be an entry of the contiguous subsequence $(7, 4, 9, 6, \ldots, 7+2k, 4+2k)$ of ρ_k for arbitrary k. If x is even, then there are no entries that appear after x in ρ_k that are smaller than x, and x is preceeded by the entry $x+1$. If x is



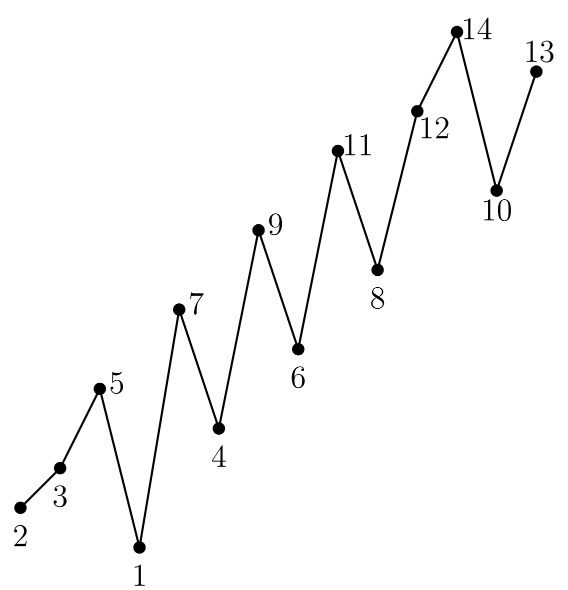
(a) $\rho_1 = (2, 3, 5, 1, 7, 4, 9, 6, 11, 8, 12, 13, 14, 10)$



(b) $\widehat{\rho_1} = (2, 3, 5, 1, 7, 4, 9, 6, 11, 8, 12, 13, 10, 14)$



(c) $\widetilde{\rho_1} = (2, 3, 5, 1, 7, 4, 9, 6, 11, 8, 12, 14, 13, 10)$



(d) $\rho_1^* = (2, 3, 5, 1, 7, 4, 9, 6, 11, 8, 12, 14, 10, 13)$

Figure 3.5: A visual representation of the elements of X_{ρ_1}.

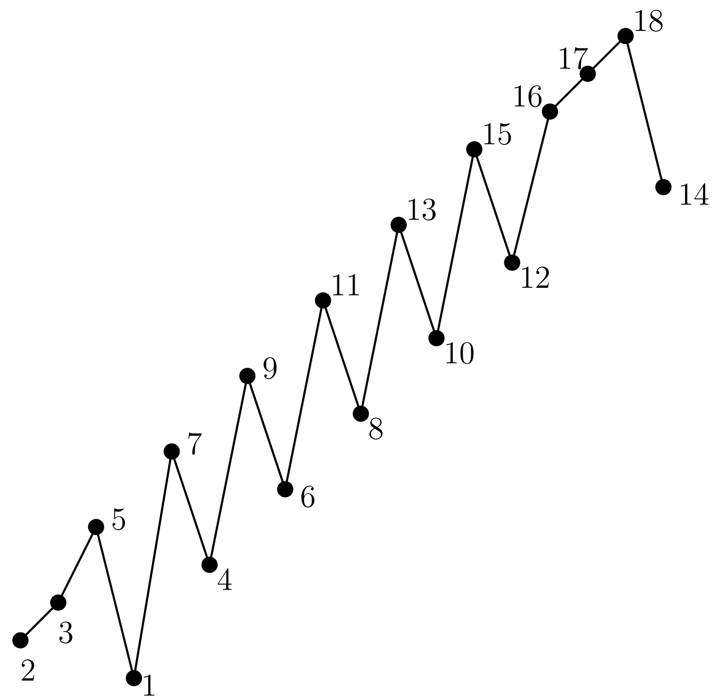


Figure 3.6: $\rho_3 = (2, 3, 5, 1, 7, 4, 9, 6, 11, 8, 13, 10, 15, 12, 16, 17, 18, 14)$

odd, then there are exactly two entries in ρ_k that follow x and are smaller than x, and they are both even.

Let x now be the first entry that is erased from ρ_j. The entries before x in ρ_i are mapped to the same entries, respectively, in ρ_j, and the entry x in ρ_i is mapped to a different entry that appears after x in ρ_j.

If x is even, then, as the entry $x+1$ is before x in ρ_i, x must map to an entry smaller than $x+1$ but is after x in ρ_j. As such an entry does not exist, x must be odd.

As x is odd, it is immediately followed by the even entry $x-3$ in both ρ_i and ρ_j, and preceeded by the entry $x-2$, which was not erased from ρ_j. As entry $x-2$ in ρ_i maps to entry $x-2$ in ρ_j, and entry x in ρ_i maps to an entry after x in ρ_j, it follows that entry $x-3$ in ρ_i must map to an entry that is after $x-3$ in ρ_j and is smaller than $x-3$. Since such an entry does not exist, $\rho_i \not< \rho_j$. $\square$

3.3 Planarity of the Tanglegrams in the Antichain

As a final point of this chapter, we show that the tanglegrams of Theorem 33 are in fact all planar. We do this by building off of the result of Czabarka, Szekely, and Wagner, presented here as Theorem 28.

Lemma 34. *A catergram $\mathcal{T}_\pi$ is planar if and only if none of* $(3,2,1,4)$, $(4,2,1,3)$, $(3,2,4,1)$, *or* $(4,2,3,1)$ *is a pattern of* π.

Proof. Since the tanglegram $\mathcal{T}_2$ referenced by Theorem 28 is not a catergram, it can never be an induced subtanglegram of $\mathcal{T}_\pi$. Therefore, Theorem 28 yields that $\mathcal{T}_\pi$ is planar if and only if $\mathcal{T}_1$ is not an induced tanglegram of $\mathcal{T}_\pi$. Inspecting Figure 3.2, it is clear that $\mathcal{T}_1$ is defined by the permutation $(3,2,1,4)$, i.e., $\mathcal{T}_1 = \mathcal{T}_{(3,2,1,4)}$. By Lemma 31(c), $\mathcal{T}_{(3,2,1,4)} \not\preceq \mathcal{T}_\pi$ if and only if $\sigma \not\leq \pi$ for any $\sigma \in X_{(3,2,1,4)} = \{(3,2,1,4),(4,2,1,3),(3,2,4,1),(4,2,3,1)\}$. □

Theorem 35. *For every $i \in \mathbb{Z}^+$ the catergram $\mathcal{T}_{\rho_i}$ is planar.*

Proof. By Lemma 34, it suffices to show that none of $(3,2,1,4)$, $(4,2,1,3)$, $(3,2,4,1)$, or $(4,2,3,1)$ is a pattern of ρ_i for any $i \geq 1$. As ρ_i does not contain a decreasing subsequence of length 3, $(3,2,1,4)$ and $(4,2,1,3)$ are not among its patterns. The last entry of the remaining $(3,2,4,1)$ and $(4,2,3,1)$ has three larger elements preceding it, and the first two elements are in decreasing order. If they are patterns of ρ_i, then 1 must map to either 1 or $8+2i$. If 1 maps to 1, then the other three elements must map to the sequence $(2,3,5)$, and if 1 maps to $8+2i$, then the remaining three elements must map to a subsequence of $(9+2i, 10+2i, 11+2i, 12+2i)$. As both of these are increasing, $(3,2,4,1)$ and $(4,2,3,1)$ are not patterns of ρ_i. □

Just having a proof that $\mathcal{T}_{\rho_i}$ is planar is somewhat unsatisfactory; one naturally wants to see a planar layout of this catergram.

First note that given a plane tree P of any rooted binary tree T with n uniquely labeled leaves, the drawing of P gives an ordering $(\ell_1, \ldots, \ell_n)$ of the labels by the order they appear on their line in the drawing. Moreover, if v is an internal vertex of T, then the set of leaves that are descendants of v, i.e., the leaves separated by v from the root, must appear in a contiguous block of $(\ell_1, \ldots, \ell_n)$. It is easy to see that if $(\ell_1, \ldots, \ell_n)$ is an ordering of the leaf labels such that for every internal vertex v of T the leaves that are descendants of v appear in a contiguous block of $(\ell_1, \ldots, \ell_n)$, then there is precisely one plane tree P of T that puts the leaves in the order $(\ell_1, \ldots, \ell_n)$ on its line of leaves.

If v is an internal vertex of the caterpillar C_n whose leaves are labeled according to our distance convention, then there is an $i \in [n]$ such that the set of leaves that are descendants of v are exactly the leaves labeled with entries that are at least i. Therefore a permutation $(\ell_1, \ldots, \ell_n) \in S_n$ arises from a plane tree of C_n precisely when for every $i \in [n]$, the entries bigger than i appear only on one side (left or right) of i in $(\ell_1, \ldots, \ell_n)$.

Definition 36. Given a rooted binary tree T on n leaves, which are labeled by the elements of $[n]$, we call a permutation $(\ell_1, \ldots, \ell_n) \in S_n$ *consistent* with T, if for every internal vertex v of T, the set of leaves that are descendants of v appear in a contiguous block of $(\ell_1, \ldots, \ell_n)$. A permutation $(\ell_1, \ldots, \ell_n)$ is *cater-good*, if it is consistent with the distance labeled caterpillar C_n (see Definition 28), i.e., for every $i \in [n]$, the entries bigger than i appear only one side (left or right) of i in $(\ell_1, \ldots, \ell_n)$.

Lemma 36. *The following facts are obvious and are presented without proof:*

(a) *The tanglegram (T_1, T_2, M), where the leaves of T_1 and T_2 are labeled, is planar if and only if there are permutations $\pi_1 = (a_1, \ldots, a_n)$ and $\pi_2 = (b_1, \ldots, b_n)$ of the leaf labels of T_i, such that π_i is consistent with T_i for $i = 1, 2$, and $M = \{a_i b_i : i \in [n]\}$.*

(b) *The catergram* T_σ *is planar if and only if there is a cater-good permutation* $(a_1, \ldots, a_n)$ *such that* $(\sigma(a_1), \ldots, \sigma(a_n))$ *is also cater-good. A planar layout is obtained by these permutation, putting leaves in their order on the lines* $x = 0$ *and* $x = 1$.

(c) *If a permutation* $(c_1, \ldots, c_n)$ *of* $[n]$ *is unimodal, then it is cater-good.*

(d) *For every* $i \in \mathbb{Z}^+$, *a planar drawing of* $\mathcal{T}_{\rho_i}$ *is given by the permutation* $(a_1, \ldots, a_{12+2i})$ *where:*

- $(a_1, a_2, a_3) = (1, 2, 3)$,
- *For* $j \in [3 + i]$, $a_{3+j} = 3 + 2j$,
- $(a_{7+i}, a_{8+i}, a_{9+i}) = (10 + 2i, 11 + 2i, 12 + 2i)$,
- *For* $j \in [3 + i]$, $a_{13+2i-j} = 2 + 2j$.

Note that the permutation $(a_1 = 1, \ldots, a_{12+2i})$ in (d) is unimodal, and consequently so is $(\rho_i(a_1), \ldots, \rho_i(a_{12+2i})) = (a_2, a_3 \ldots, a_{12+2i}, 1)$. Figure 3.7 gives the planar drawing of $\mathcal{T}_{\rho_4}$ determined by the permutation given in this lemma.

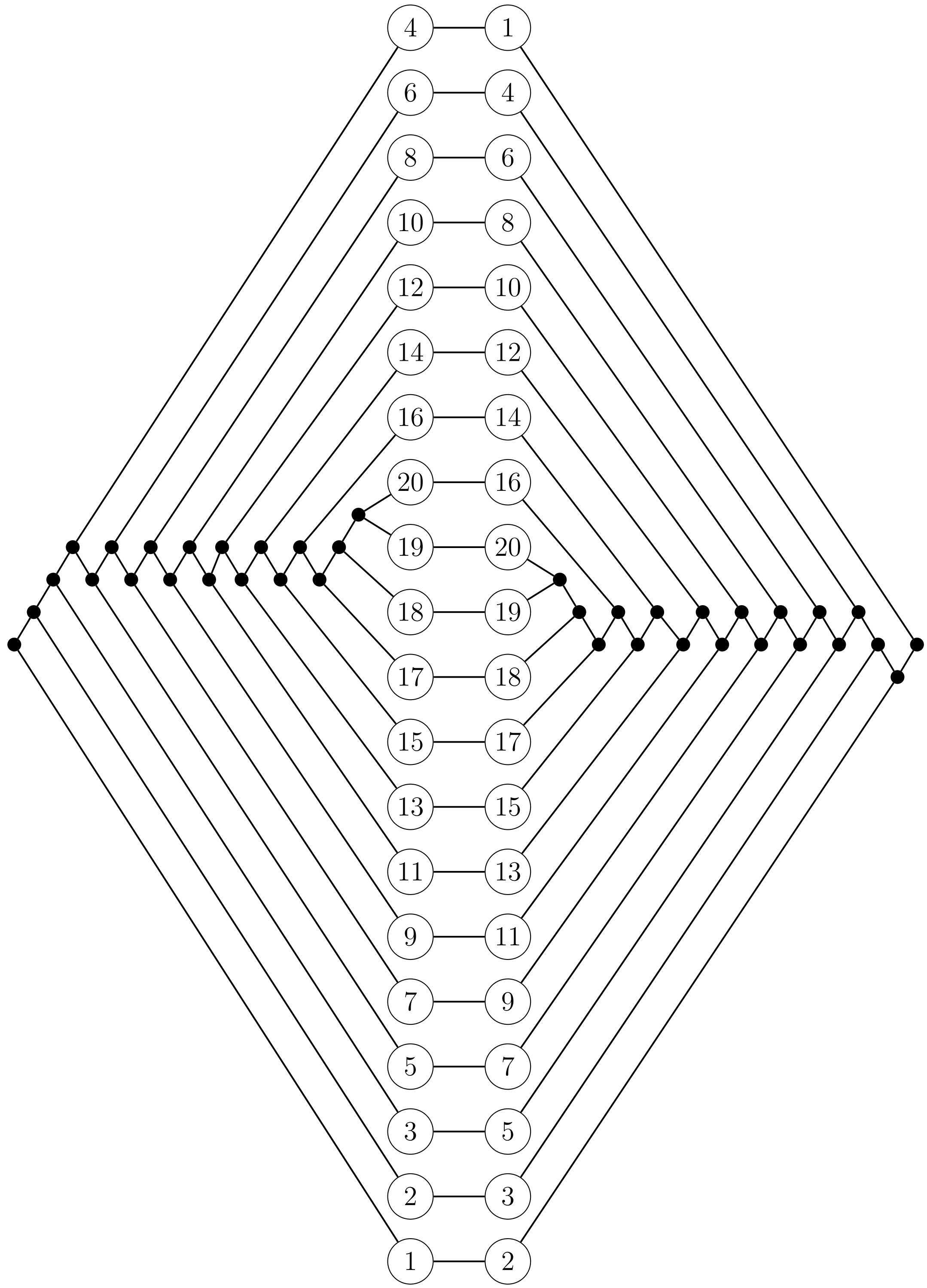


Figure 3.7: A planar drawing of $\mathcal{T}_{\rho_4}$ as described in Lemma 36 (d).

Chapter 4

Maximum Diameter of k-Colorable Graphs

4.1 History and Background

New ideas and results presented in this chapter are joint work with Éva Czabarka and László Székely.

The notion of distance is a natural and fundamental one, and questions about such notions can be of import to both theory and application. In what follows, we will focus on bounding from above the diameter of a graph in terms of some of its other parameters - its minimum degree and order to name just two. Being the maximum distance between any two vertices, the diameter provides a natural upper bound, or worst-case, for any problem which involves the distance metric in the graph theory sense.

There is a natural relationship between the minimum degree of a graph and its diameter, as increasing the minimum degree would force the graph to have more edges, which in turn may create new and shorter paths between pairs of vertices. Indeed, the authors of [1], [17], [21] and [29] independently proved the following:

Theorem 37. *For a fixed minimum degree $\delta \geq 2$, every connected graph G of order n satisfies* $\operatorname{diam}(G) \leq \dfrac{3n}{\delta+1} + O(1)$*, as $n \to \infty$.*

This upper bound is sharp (even for δ-regular graphs [6]), but the optimal constructions all have complete subgraphs whose order increases with n. Erdős, Pach, Pollack, and Tuza in [17] made note of this, and conjectured that if we consider the set of graphs without a clique of a given (fixed) size, then we can make a stronger statement about their diameter in terms of this additional parameter:

Conjecture 1 (Erdős, Pach, Pollack, and Tuza, [17])**.** *Let $r, \delta \geq 2$ be fixed integers and let G be a connected graph of order n and minimum degree δ.*

(i) *If G is K_{2r}-free and δ is a multiple of $(r-1)(3r+2)$ then, as $n \to \infty$,*

$$\begin{aligned} \operatorname{diam}(G) &\leq \frac{2(r-1)(3r+2)}{(2r^2-1)} \cdot \frac{n}{\delta} + O(1) \\ &= \left(3 - \frac{2}{2r-1} - \frac{1}{(2r-1)(2r^2-1)}\right)\frac{n}{\delta} + O(1). \end{aligned}$$

(ii) *If G is K_{2r+1}-free and δ is a multiple of $3r-1$, then, as $n \to \infty$,*

$$\operatorname{diam}(G) \leq \frac{3r-1}{r} \cdot \frac{n}{\delta} + O(1) = \left(3 - \frac{2}{2r}\right)\frac{n}{\delta} + O(1).$$

Furthermore, they created examples showing that the above conjecture, if true, is sharp, and showed part (ii) of the conjecture for $r = 1$. This conjecture proved to be quite challenging, and no additional progress was made for almost 20 years. In 2009, Czabarka, Dankelmann and Székely decided to attack the problem under a stronger assumption: Rather than considering graphs which are K_{k+1}-free, they instead considered graphs which are k-colorable. Under this stronger assumption, they were able to verify Conjecture 1 (ii) for $r = 2$.

Theorem 38 (Czabarka, Dankelmann, and Székely, [9])**.** *For every connected* 4*-colorable graph G of order n and minimum degree $\delta \geq 1$,* $\operatorname{diam}(G) \leq \frac{5n}{2\delta} - 1$.

Czabarka, Singgih and Székely ([10]) gave an infinite family of $(2r-1)$-colorable (hence K_{2r}-free) graphs with diameter $\dfrac{(6r-5)(n-2)}{(2r-1)\delta + 2r - 3} - 1$, providing a counterexample for Conjecture 1 (i) for every $r \geq 2$ and $\delta > 2(r-1)(3r+2)(2r-3)$. The question of whether Conjecture 1 (i) holds in the range $(r-1)(3r+2) \leq \delta \leq 2(r-1)(3r+2)(2r-3)$ remains open. The counterexample led Czabarka *et al.* ([10])

to the modified conjecture below, which no longer requires cases for the parity of the order of the excluded complete subgraphs:

Conjecture 2 (Czabarka, Singgih and Székely, [10]). *For every $k \geq 3$ and $\delta \geq \lceil \frac{3k}{2} \rceil - 1$, if G is a K_{k+1}-free (under a stronger hypothesis, k-colorable) connected graph of order n and minimum degree at least δ,* $\mathrm{diam}(G) \leq \left(3 - \frac{2}{k}\right) \frac{n}{\delta} + O(1)$.

Rather than considering the set of all graphs in Conjecture 2 for fixed k, δ, and n, Czabarka, Singgih and Székely proved in [11] that it is sufficient to consider only those graphs which generate *canonical clump graphs* (defined in Section 4.2). These graphs have a significant amount of structure that brings additional clarity to the problem. By considering a related packing problem (see Theorem 42 in Section 4.2), Czabarka, Singgih and Székely proved the following:

Theorem 39 (Czabarka, Singgih and Székely, [11]). *Assume $k \geq 3$. If G is a connected k-colorable graph of minimum degree at least δ, then*

$$\mathrm{diam}(G) \leq \frac{3k-4}{k-1} \cdot \frac{n}{\delta} - 1 = \left(3 - \frac{1}{k-1}\right) \frac{n}{\delta} - 1.$$

Theorem 39 displays an astounding amount of progress towards Conjecture 2, but there is still a gap between the proven and conjectured upper bounds, even with the stronger assumption. Czabarka, Székely, and I focused on closing this gap, and I show in Section 4.4 how to do this for $k = 3$ and $k = 4$.

4.2 Clump Graphs and Duality

As mentioned in Section 4.1, when attacking Conjecture 2 for given k, δ, and n, it is not necessary to consider the set of *all* graphs on n vertices which are k-colorable and have minimum degree at least δ. Instead, given an arbitrary such graph G', we will see that there exists a graph, G, such that G has the same relevant parameters as G', and G generates a *strongly canonical clump graph.* Therefore, we need only bound the diameter of graphs from this smaller family of highly structured graphs.

Given a k-colorable connected graph G of order n and minimum degree at least δ, choose a vertex x whose eccentricity is $\mathrm{diam}(G)$. Take a *fixed* good k-coloring of G. Let *layer* L_i denote the set of vertices at distance i from x, and a *clump* in L_i be a maximal set of vertices in L_i that have the same color. The number of layers is $\mathrm{diam}(G) + 1$. We call a graph *layered*, if such a vertex x and the distance layers $L_0 = \{x\}, L_1, \ldots, L_D$ are given. Let $c(i) \in \{1, 2, \ldots, k\}$ denote the number of colors used in layer L_i by our fixed coloration. We can assume without loss of generality that any two vertices in layer L_i in G which are differently colored are joined by an edge in G, and that two vertices in consecutive layers which are differently colored are also joined by an edge in G. We call this assumption *saturation* with respect to the fixed good k-coloring. Assuming saturation does not make loss of generality, as adding these edges does not decrease degrees, keeps the fixed good k-coloration, and does not reduce the diameter, while making the graph more structured for our convenience.

From the layered and saturated graph G above, we create an *unweighted clump graph* $H = H(G)$. Vertices of H correspond to the clumps of G. Two vertices of H are connected by an edge if there were edges between the corresponding clumps in G. H is naturally k-colored and layered, based on the coloration and layering of G. With a slight abuse of notation, we denote the layers of H by L_i as well. To create a *weighted clump graph*, we assign positive integer *weights* to each vertex of the unweighted clump graph. Blowing up vertices of H into as many copies as their weight is, we obtain a bigger k-colorable graph of the same diameter (we do not put edges between successors of the same vertex). In case the weights are the cardinalities of the clumps in G, after the blow-up of $H = H(G)$, we get back G. The degree of a vertex v in a blow-up of H, where v is a successor of a vertex w of H by blow-up, is the sum of the weights of neighbors of the vertex w in H. The number of vertices in a blow-up of H is the sum of the weights of all vertices in H.

Definition 37 (Czabarka, Singgih and Székely, [11])**.** A *canonical clump graph* is a clump graph $H = H(G)$ which is obtained in the manner outlined above from any graph G that satisfies all four conclusions of Theorem 40 below.

Theorem 40 (Czabarka, Singgih and Székely, [11])**.** *Assume $k \geq 3$. Let G' be a k-colorable connected graph of order n, diameter D and minimum degree at least δ. Then there is a saturated k-colored and layered connected graph G of the same parameters n and δ, with layers $L_0, \ldots, L_D$, for which the following hold for every i $(0 \leq i \leq D-1)$:*

(a) *If $c(i) = 1$, then $c(i+1) \leq k-1$.*

(b) *The number of colors used to color the set $L_i \cup L_{i+1}$ is $\min(k, c(i)+c(i+1))$. In particular, when $c(i)+c(i+1) \leq k$, then L_i and L_{i+1} do not share any color.*

(c) *If $c(i) = k$, then $i \geq 2$ and $c(i+1) \geq 2$.*

(d) *If $|L_i| > c(i)$, i.e., L_i contains two vertices of the same color, then $i > 0$ and $c(i) + \max\Big(c(i-1), c(i+1)\Big) \geq k$.*

It will be helpful for us to require one additional property on the canonical clump graphs of Definition 37.

Definition 38. We call a canonical clump graph with diameter $D \geq 2$ *strongly canonical* when it uses exactly one color in its first and last layers, i.e., when $c(0) = c(D) = 1$. We also consider any canonical clump graph with diameter 1 to be strongly canonical.

It is not difficult to see the following: If the graph G' in the assumption of Theorem 40 is layered with $|L_0'| = 1$ and $c'(D) = 1$, then the proof of Theorem 40 in [11] provides a layered graph G with $|L_0| = 1$ (and hence $c(0) = 1$), and $c(D) = 1$. Based on this observation, the following lemma implies that to resolve Conjecture 2, it is sufficient to consider only those graphs, G, that generate a strongly canonical clump graph.

Lemma 41. *Assume $k \geq 3$ and $D \geq 2$. Let G' be a k-colored layered connected graph of order n, diameter D, and minimum degree at least δ, with layers $L'_0, \ldots, L'_D$. Then there is a k-colored layered connected graph G of the same parameters, with layers $L_0, \ldots, L_D$, for which $c(0) = c(D) = 1$, and for each i $(0 \leq i \leq D-2)$, we have $c'(i) = c(i)$ and $L'_i = L_i$.*

Proof. As $|L'_0| = 1$ is necessary in a layered graph, we must have $c'(0) = 1$, and if $c'(D) = 1$, the choice $L'_i = L_i$ suffices. If $c'(D) > 1$, pick a color A in L'_D. If possible, choose A so that A also appears in L'_{D-2}. This ensures that for all colors B in L'_D such that $B \neq A$, there is a color C in L'_{D-2} such that $B \neq C$ (where $C = A$, if A appeared in L_{D-2}, otherwise any color in L_{D-2} works). Create a layered graph G from G' by moving all vertices in L'_D that are not colored A to the next-to-last layer, which will be L_{D-1}, and connect them to all vertices in $L_{D-2} = L'_{D-2}$ that have different color. Note that for all vertices of L_{D-1}, there is at least one such vertex. As we only changed the number of vertices in layers $D-1$ and D, and did not change the coloration of the vertices, the claim follows. □

Until now, we have been considering the problem of finding the maximum diameter of any graph defined by the fixed parameters k, δ, and n. Let's consider the following related problem: Fix k, δ, and the diameter D. For a k-colorable graph with minimum degree δ and diameter D, how small can the order be? Our weighted strongly canonical clump graphs give us the perfect means to formulate this question and articulate solutions. Imagine that H is a strongly canonical clump graph that arises from a graph G which is k-colorable, has minimum degree at least δ and has diameter D, and consider weighting the vertices. No matter what weights are chosen, the resulting blown-up graph will still be k-colorable and will still have diameter D. Therefore, one way to think about this new question is to find a weighting of the vertex set of H such that the resulting blown-up graph has the minimum possible order while still satisfying the minimum degree condition.

Let $\mathcal{H} = \mathcal{H}_{k,D,\delta}$ denote the family of strongly canonical clump graphs arising from k-colorable graphs with minimum degree at least δ and diameter D. Fix $H \in \mathcal{H}$. We formalize the ideas in the above argument by describing our new question as a linear program:

$$\text{Minimize} \sum_{x \in V(H)} w(x),$$

subject to the condition

$$\forall y \in V(H) \qquad \sum_{x \in V(H)\,:\, xy \in E(H)} w(x) \geq \delta. \tag{4.1}$$

The dual of 4.1 is the following packing problem:

$$\text{Maximize} \sum_{y \in V(H)} \delta \cdot u(y),$$

subject to the condition

$$\forall x \in V(H) \qquad \sum_{y \in V(H)\,:\, xy \in E(H)} u(y) \leq 1. \tag{4.2}$$

Using the inequalities of weak duality, Czabarka, Singgih and Székely were able to connect solutions of packing problem 4.2 to Conjecture 2.

Theorem 42 (Czabarka, Singgih and Székely, [11])**.** *Fix $k \geq 3$. Assume that there exists constants $\tilde{u} > 0$ and $C \geq 0$ such that for all D and δ, and for all $H \in \mathcal{H}_{k,D,\delta}$, the optimum of linear program 4.2 above is at least $\tilde{u}\delta D + C\delta$. Then for any k-colorable graphs G with minimum degree δ on n vertices, we have*

$$\operatorname{diam}(G) \leq \frac{1}{\tilde{u}} \frac{n}{\delta} - \frac{C}{\tilde{u}}.$$

Theorem 42 is a powerful tool, and applications of it have produced the impressive progress made towards Conjecture 2 thus far. It says that if we

1. fix $k \geq 3$,

2. choose $\tilde{u}$ and C (which will depend on k), and

3. for general D, δ, and $H \in \mathcal{H}_{k,D,\delta}$, find a feasible solution of 4.2 that is at least $\tilde{u}\delta D + C\delta$,

then we can bound from above the diameter of any k-colorable graph with minimum degree δ on n vertices. The difficulty of this technique lies in step 3 - the larger you make $\tilde{u}$ and C, the harder it becomes to create general feasible solutions. To illustrate this technique, I will include three concrete applications, each of increasing complexity. The first two applications (shown in Section 4.3) were given by Czabarka, Singgih, and Székely, and provide an upper bound on the diameter of the graphs in Conjecture 2 for general k. In Section 4.4, I will show how to use this tool to close the gap and verify Conjecture 2 for $k = 3$ and $k = 4$.

Remark 1. In Definition 38, there is a distinction made between strongly canonical clump graphs with diameter 1 and diameter at least 2. The applications of Theorem 42 in Section 4.3 were originally given with only the notion of canonical clump graphs (not strongly canonical) and thus there is no need to consider clump graphs with diameter 1 and diameter at least 2 separately.

4.3 Applications for General k

In this section, we will see two concrete applications of Theorem 42 which produce progressively better upper bounds on the diameter of the graphs of Conjecture 2 for general k. These two applications, in conjunction with that given in Section 4.4, illustrate a natural progression of ideas that one might try to fully prove Conjecture 2 under the stronger colorability assumption.

For the entirety this section, assume that $k \geq 3$ is fixed and that $H \in \mathcal{H}_{k,D,\delta}$ is also fixed. To use Theorem 42, we will construct weightings of the vertex set of H so that the neighborhood condition of packing problem 4.2 is satisfied. Recall that we use the notation L_i for the layers of the clump graph $H(G)$ as well as for the layers of G. Hence $c(i) = |L_i|$ if L_i denotes a layer of the clump graph. We denote a weighting of the vertex set of H by a function such as $u : V(H) \to [0,1]$. For notational convenience, when $X \subseteq V(H)$, and the weighting is given by u, we will define $u(X)$ to be the sum of the weights of the vertices in X, i.e., $u(X) := \sum_{x \in X} u(x)$.

The first idea that we will explore is to give every layer of H the same total weight, and to distribute each layer's weight equally amongst its vertices. It turns out that under this weighting scheme, the largest weight we can assign to each layer is $k/(3k-1)$. Thus, we formally define our first weighting scheme as follows: For $0 \leq i \leq D$, the weight of vertex v is given by $u_1(v) = \dfrac{k}{|L_i|(3k-1)}$ for every vertex $v \in L_i$.

Lemma 43. *For fixed $k \geq 3$ and fixed $H \in \mathcal{H}_{k,D,\delta}$, the weighting $u_1 : V(H) \to [0,1]$ described above is a feasible solution of packing problem 4.2. Furthermore, the ojective function evaluated at this feasible solution is at least $\tilde{u}\delta D + C\delta$, where $\tilde{u} = C = \dfrac{k}{3k-1}$.*

Proof. Let us first evaluate the objective function of packing problem 4.2 under the weighting u_1. For any layer L_i,

$$u_1(L_i) = \sum_{v \in L_i} u_1(v) = \sum_{v \in L_i} \frac{k}{|L_i|(3k-1)} = |L_i| \cdot \frac{k}{|L_i|(3k-1)} = \frac{k}{3k-1}.$$

Therefore,

$$\sum_{v \in V(H)} u_1(v) = \sum_{i=0}^{D} u_1(L_i) = (D+1)\frac{k}{3k-1}.$$

We plug this into the objective function of 4.2 to find

$$\sum_{v \in V(H)} \delta \cdot u_1(v) = \delta(D+1)\frac{k}{3k-1} = \frac{k}{3k-1} \cdot \delta \cdot D + \frac{k}{3k-1} \cdot \delta.$$

It remains to show that u_1 is a feasible solution of 4.2, i.e. that $u_1(N(v)) \leq 1$ for all $v \in V(H)$. To that end, let $v \in V(H)$ be arbitrary and suppose $v \in L_i$ for $0 < i < D$. It is always true that $L_i \setminus \{v\} \subseteq N(v)$ (by the saturation assumption), and in the worst case, $L_{i-1} \cup L_{i+1} \subseteq N(V)$. Therefore,

$$\begin{aligned} u_1(N(v)) &\leq u_1\Big(L_{i-1} \cup (L_i \setminus \{v\}) \cup L_{i+1}\Big) \\ &= u_1(L_{i-1}) + u_1(L_i) + u_1(L_{i+1}) - u_1(v) \qquad (4.3) \\ &= 3 \cdot \frac{k}{3k-1} - u_1(v). \end{aligned}$$

Since $|L_i| \leq k$, we have

$$\frac{1}{3k-1} \leq \frac{k}{|L_i|(3k-1)} = u_1(v). \qquad (4.4)$$

Combining Equations 4.3 and 4.4, we get

$$u_1(N(v)) \leq \frac{3k}{3k-1} - \frac{1}{3k-1} = 1.$$

If $v \in L_i$ for $i = 0$, then $N(v) \subseteq L_0 \cup L_1$. Therefore, $u_1(N(v)) \leq u_1(L_0) + u_1(L_1) = \frac{2k}{3k-1} < 1$. The same calculation can be done for $v \in L_D$. □

Theorem 44. *Assume $k \geq 3$. If G is a connected k-colorable graph of order n with minimum degree at least $\delta \geq 1$, then* $\operatorname{diam}(G) \leq \left(3 - \frac{1}{k}\right)\frac{n}{\delta} - 1$.

Proof. With Lemma 43 in place, we may apply Theorem 42 with $\tilde{u} = C = \frac{k}{3k-1}$. □

For our next idea, we design a new weighting for which every layer has the same total layer weight (as before), but we now utilize some structure to distribute differing amounts of a layer's weight amongst its vertices. This will allow us to improve the result of Theorem 44 to that of Theorem 39. Some conventions and notation will be helpful to describe the structural properties that we make use of.

The strongly canonical clump graph that we are considering already has layers $L_0, L_1, \ldots, L_D$ defined. We define for convenience the two additional layers L_{-1} and L_{D+1}, as $L_{-1} = L_{D+1} = \emptyset$.

Definition 39. For each i, $0 \le i \le D$, define the (possibly empty) set $S_i = \{x \in L_i : L_{i-1} \cup L_i \subseteq N(x)\}$. We set $S_{-1} = S_{D+1} = \emptyset$, in accordance with $L_{-1} = L_{D+1} = \emptyset$.

Suppose that $v \in L_i$ for some $0 \le i \le D$. We define our second weighting, u_2, as follows:

- If $|L_i| \le k-1$, then $u_2(v) = \dfrac{k-1}{|L_i|(3k-4)}$.
- If $|L_i| = k$, then $u(v) = \begin{cases} \dfrac{1}{3k-4}, & \text{if } v \in S_i, \\ \dfrac{1}{3k-4} - \dfrac{1}{(k-|S_i|)(3k-4)} & \text{otherwise.} \end{cases}$

Lemma 45 (Czabarka, Singgih and Székely, [11])**.** *For fixed $k \ge 3$ and fixed $H \in \mathcal{H}_{k,D,\delta}$, the weighting $u_2 : V(H) \to [0,1]$ described above is a feasible solution of packing problem 4.2. Furthermore, the ojective function evaluated at this feasible solution is at least $\tilde{u}\delta D + C\delta$, where $\tilde{u} = C = \dfrac{k-1}{3k-4}$.*

Proof. We again begin by first evaluating the objective function of packing problem 4.2 under the weighting u_2. For layers L_i with $|L_i| \le k-1$,

$$u_2(L_i) = \sum_{v \in L_i} u_2(v) = \sum_{v \in L_i} \frac{k-1}{|L_i|(3k-4)} = |L_i| \cdot \frac{k-1}{|L_i|(3k-4)} = \frac{k-1}{3k-4}.$$

For layers L_i with $|L_i| = k$,

$$\begin{aligned} u_2(L_i) &= \sum_{v \in S_i} u_2(v) + \sum_{v \in L_i \setminus S_i} u_2(v) \\ &= \sum_{v \in S_i} \frac{1}{3k-4} + \sum_{v \in L_i \setminus S_i} \left(\frac{1}{3k-4} - \frac{1}{(k-|S_i|)(3k-4)} \right) \\ &= |S_i| \cdot \frac{1}{3k-4} + |L_i \setminus S_i| \cdot \left(\frac{1}{3k-4} - \frac{1}{(k-|S_i|)(3k-4)} \right) \\ &= |S_i| \cdot \frac{1}{3k-4} + (k-|S_i|) \cdot \left(\frac{1}{3k-4} - \frac{1}{(k-|S_i|)(3k-4)} \right) \\ &= \frac{k-1}{3k-4}. \end{aligned}$$

Since every layer has total weight $\dfrac{k-1}{3k-4}$,

$$\sum_{v \in V(H)} u_2(v) = \sum_{i=0}^{D} u_2(L_i) = (D+1)\frac{k-1}{3k-4}.$$

We plug this into the objective function of 4.2 to find

$$\sum_{v \in V(H)} \delta \cdot u_2(v) = \delta(D+1)\frac{k-1}{3k-4} = \frac{k-1}{3k-4} \cdot \delta \cdot D + \frac{k-1}{3k-4} \cdot \delta.$$

It remains to show that u_2 is a feasible solution of 4.2, i.e. that $u_2(N(v)) \leq 1$ for all $v \in V(H)$. Let $v \in V(H)$ be arbitrary and suppose that $v \in L_i$ with $|L_i| \leq k-1$. In the worst case, $N(v) = L_{i-1} \cup (L_i \setminus \{v\}) \cup L_{i+1}$. Therefore,

$$\begin{aligned} u_2(N(v)) &\leq u_2\Big(L_{i-1} \cup (L_i \setminus \{v\}) \cup L_{i+1}\Big) \\ &= u_2(L_{i-1}) + u_2(L_i) + u_2(L_{i+1}) - u_2(v) \\ &= 3 \cdot \frac{k-1}{3k-4} - u_2(v). \end{aligned} \tag{4.5}$$

Since $|L_i| \leq k-1$, we have

$$\frac{1}{3k-4} \leq \frac{k-1}{|L_i|(3k-4)} = u_1(v). \tag{4.6}$$

Combining Equations 4.5 and 4.6, we get

$$u_2(N(v)) \leq \frac{3k-3}{3k-4} - \frac{1}{3k-4} = 1.$$

Now suppse that $v \in L_i$ with $|L_i| = k$. If $v \in S_i$, then the argument to show that $u_2(N(v)) \leq 1$ proceeds exactly as the agument just completed. Assume then that $v \in L_i \setminus S_i$. By definition, there exists some vertex $w \in L_{i-1} \cup L_{i+1}$ such that $vw \notin E(H)$. Since $|L_i| = k$, Theorem 40(c) and the contrapositive of Theorem 40(a) guarantee that whichever layer w is in has at least 2 vertices; we refer to the layer that contains w as L^*. Since $S_i \cup L^*$ forms a clique and H is k-colorable, $|S_i| + |L^*| \leq k$, and hence, $|S_i| \leq k - |L^*| \leq k-2$. Shifting this inequality around, we find that $k - |S_i| \geq 2$, which implies

$$u_2(v) = \frac{1}{3k-4} - \frac{1}{(k-|S_i|)(3k-4)} \geq \frac{1}{3k-4} - \frac{1}{2(3k-4)} = \frac{1}{2(3k-4)}.$$

Notice that the previous arguments have also shown that no matter how w is weighted, $u_2(w) \geq \min\left(\frac{1}{3k-4}, \frac{1}{2(3k-4)}\right) = \frac{1}{2(3k-4)}$. Therefore,

$$\begin{aligned}
u_2(N(v)) &\leq u_2\Big(\big(L_{i-1} \cup L_i \cup L_{i+1}\big) \setminus \{v, w\}\Big) \\
&= u_2(L_{i-1}) + u_2(L_i) + u_2(L_{i+1}) - u_2(v) - u_2(w) \\
&\leq 3 \cdot \frac{k-1}{3k-4} - 2 \cdot \frac{1}{2(3k-4)} \\
&= 1.
\end{aligned}$$

□

To prove Theorem 39, Czabarka, Singgih and Székely demonstrated Lemma 45 and applied Theorem 42 with $\tilde{u} = C = \frac{k-1}{3k-4}$.

In order to make further progress towards Conjecture 2, we need to not only have differing amounts of weight distributed amongst vertices in a given layer, but to also have differing total layer weights based on the structure of the given strongly canonical clump graph.

4.4 Closing the Gap for $k = 3, 4$

In this section, we include one more application of Theorem 42 to close the gap between the proven upper bound in Theorem 39, and the conjectured upper bound in Conjecture 2, for $k = 3$ and $k = 4$. In other words, for fixed $k \in \{3, 4\}$, our goal is to produce a feasible solution to packing problem 4.2 for arbitrary strongly canonical clump graphs $H \in \mathcal{H}_{k,D,\delta}$. We will require much more of the graph's structure to be elucidated than in Section 4.3 to define the weighting and to prove that it contains the necessary properties.

As the notation L_i refers to both the layers of the clump graph $H(G)$, and the layers of G, we can rephrase the arguments of Section 4.2 to better suit our situation. Keep in mind that $c(i) = |L_i|$ if L_i denotes a layer of the clump graph.

Lemma 46. *An unweighted k-colorable strongly canonical clump graph with layers $L_0, \ldots, L_D$ satisfies the following properties:*

(a) $|L_0| = 1$ *and whenever* $D > 1$, $|L_D| = 1$ *as well,*

(b) *If* $|L_i| = k$, *then* $2 \leq i \leq D - 1$ *and* $\min(|L_{i-1}|, |L_{i+1}|) \geq 2$, *and*

(c) *For* $i \in [D]$, *the edges that do not appear between* L_{i-1} *and* L_i *form a matching of size* $\max(k, |L_{i-1}| + |L_i|) - k$.

For the following definition, and also for the rest of this section, assume that we are given a k-colorable strongly canonical clump graph H with layers $L_0, \ldots, L_D$. As in Section 4.3, we define for convenience the two additional layers L_{-1} and L_{D+1}, as $L_{-1} = L_{D+1} = \emptyset$. We will frequently be using the sets S_i defined in Definition 39. In addition:

Definition 40. For each i, $0 \leq i \leq D$, we call layer L_i *big* if $|S_i| > \frac{k}{2}$, and we call L_i *small* if it is not big.

Remark 2. Note that if L_i is big, then $i \notin \{0, D\}$.

Lemma 47. *Assume* $D \geq 2$. *Let* H *be an unweighted* k*-colorable strongly caonical clump graph with layers* $L_0, \ldots, L_D$. *The following is true for each* i, $0 \leq i \leq D$:

(a) $|L_i| \leq k - |S_{i-1}|$ *and* $|L_i| \leq k - |S_{i+1}|$.

(b) $|S_i| \leq k - 1$.

(c) *If* L_i *is big, then* $1 \leq i \leq D - 1$ *and* L_{i-1}, L_{i+1} *are small.*

(d) *If* $|L_i| = 1$, *then* $L_i = S_i$.

(e) $|L_i \setminus S_i| \leq k - |S_i| - |S_{i+1}|$ *and* $|L_{i+1} \setminus S_{i+1}| \leq k - |S_i| - |S_{i+1}|$.

(f) *If* $|S_i| = k - 1$, *then* $L_i = S_i$ *and for* $j = i \pm 1$, $|L_j| = |S_j| = 1$.

(g) *If* $k \in \{3, 4\}$ *and* L_i *is big, then* $|S_i| = k - 1$.

Proof. (a) This follows from the fact that $S_{i-1} \cup L_i$ and $S_{i+1} \cup L_i$ both form complete subgraphs in H, which is k-colorable.

(b) This follows from (a) and the fact that $L_{i-1} \cup L_{i+1}$ contains at least one vertex.

(c) This follows from (a) and the definition of *big*.

(d) Since $|L_i| = 1$, Lemma 46 (b) tells us (by contrapositive) that $\max(|L_{i-1}|, |L_{i+1}|) \leq k-1$. By Lemma 46 (c), the vertex in L_i is adjacent to every vertex in $L_{i-1} \cup L_{i+1}$.

(e) Since $S_i \cup S_{i+1} \cup (L_i \setminus S_i)$ forms a complete graph in the k-colorable graph H, we have $|L_i \setminus S_i| \leq k - |S_i| - |S_{i+1}|$. Similarly, $S_i \cup S_{i+1} \cup (L_{i+1} \setminus S_{i+1})$ forms a complete graph, and hence $|L_{i+1} \setminus S_{i+1}| \leq k - |S_i| - |S_{i+1}|$.

(f) Suppose $|S_i| = k - 1$. Then $1 \leq i \leq D - 1$ and by (a), $|L_{i-1}| = |L_{i+1}| = 1$. Part (d) already implies that for $j = i \pm 1$, $|L_j| = |S_j| = 1$. Using Lemma 46 (b), we further get $|L_i| \leq k - 1$, so $k - 1 = |S_i| \leq |L_i| \leq k - 1$. The claim follows.

(g) Finally, suppose L_i is big. Then by definition, $k/2 < |S_i|$, and by (b), $|S_i| \leq k - 1$. For $k \in \{3, 4\}$, these inequalities force $|S_i| = k - 1$.

□

Definition 41. Let H be an unweighted k-colorable strongly canonical clump graph with layers $L_0, \ldots, L_D$. If for some $s \geq 1$, the contiguous segment of $2s + 1$ layers, $L_i, L_{i+1}, \ldots, L_{i+2s}$, satisfies the three conditions below, then we say that the contiguous segment is Type 1 if $s = 1$, and Type 2 if $s > 1$.

(i) For each $j : 1 \leq j \leq s$ the layer L_{i+2j-1} is big (thus, L_{i+2j-2}, L_{i+2j} are small).

(ii) $i = 0$ or L_{i-1} is small.

(iii) $i + 2s = D$ or L_{i+2s+1} is small.

Definition 42. Let H be an unweighted k-colorable strongly canonical clump graph with layers $L_0, \dots, L_D$. Assume that $t \geq 0$. We say that the contiguous segment of $t+1$ layers, $L_i, L_{i+1}, \dots, L_{i+t}$, is Type 3, if the following hold:

(i) For each $j : i \leq j \leq i+t$ the layer L_j is small.

(ii) If $i \neq 0$ then $i > 2$ and L_{i-2} is big (thus, L_{i-1}, L_{i-3} are small).

(iii) If $i+t \neq D$ then $i+t < D-2$ and L_{i+t+2} is big (thus, L_{i+t+1}, L_{i+t+3} are small).

Note that in a Type 3 segment every layer is small.

The following Lemma easily follows from the definition of strongly canonical clump graphs and Lemma 47.

Lemma 48. *Let H be an unweighted k-colorable strongly canonical clump graph with diameter $D \geq 2$. Then the layers $L_0, \dots, L_D$ can be partitioned into segments of Type 1, Type 2 and Type 3. Moreover, if $k \in \{3, 4\}$ and L_j is a layer in a Type 1 or Type 2 segment, then $L_j = S_j$ and $|L_j| \in \{1, k-1\}$.*

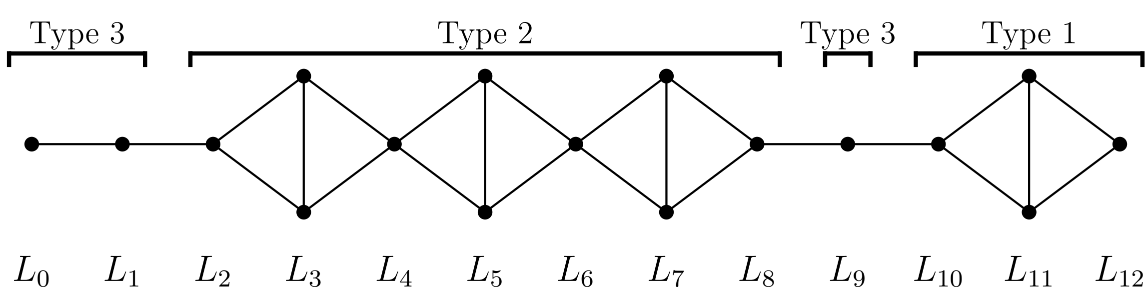


Figure 4.1: An example strongly canonical clump graph with $k = 3$. The clump graph has diameter 12, and each layer is directly (vertically) above their respective labels. The layers are partitioned into segments of Type 1, 2, and 3.

Now that we have partitioned the layers of H (an arbitrary strongly canonical clump graph from $\mathcal{H}_{k,D,\delta}$), we can assign weights to its vertices, i.e. we will specify the output of the function $u : V(H) \to [0, 1]$. Because Lemmas 41 and 48 require $D \geq 2$, we differentiate between the cases $D = 1$ and $D \geq 2$.

For $H \in \mathcal{H}_{k,1,\delta}$ (i.e. for H with $\mathrm{diam}(H) = 1$), we simply set $u(v) = 1/(|H| - 1)$ for every vertex $v \in V(H)$. For $H \in \mathcal{H}_{k,D,\delta}$ with $D \geq 2$ and $v \in V(H)$, we set $u(v)$ as follows:

- If $v \in L_i$ and L_i is of Type 1: $u(v) = \begin{cases} \dfrac{2}{3k-2}, & \text{if } |L_i| = k-1, \\ \dfrac{k+2}{2(3k-2)} & \text{otherwise.} \end{cases}$
- If $v \in L_i$ and L_i is of Type 2:

$$u(v) = \begin{cases} \dfrac{1}{2(k-1)}, & \text{if } |L_i| = k-1, \\ \dfrac{k+2}{2(3k-2)}, & \text{if } |L_i| = 1 \text{ and } L_i \text{ is not the first or last layer of the segment,} \\ \dfrac{3k+2}{4(3k-2)}, & \text{otherwise.} \end{cases}$$

Remark 3. As $k \geq 3$, $\dfrac{k+2}{2(3k-2)} < \dfrac{3k+2}{4(3k-2)} < \dfrac{k}{3k-2}$. In other words, the end-layers of Type 1 and Type 2 segments have weight smaller than $\dfrac{k}{3k-2}$.

- If $v \in L_i$ and L_i is of Type 3:

$$u(v) = \begin{cases} \dfrac{k}{(3k-2)|L_i|}, & \text{if } |L_i| \leq k/2 \\ \dfrac{2}{3k-2}, & \text{if } |L_i| > k/2 \text{ and } v \in S_i \\ \dfrac{k-2|S_i|}{(3k-2)(|L_i|-|S_i|)}, & \text{otherwise.} \end{cases}$$

Remark 4. Note that as $|S_i| \leq \dfrac{k}{2}$, we get $u(v) \geq 0$. Also, $u(v) \geq \dfrac{2}{3k-2}$ in the first and second cases of the Type 3 assignment.

Remark 5. We will see in the proof of Lemma 49 that the total weight on any Type 3 layer is $\dfrac{k}{3k-2}$.

Lemma 49. *For fixed $k \in \{3,4\}$ and fixed $H \in \mathcal{H}_{k,D,\delta}$, the weighting $u : V(H) \to [0,1]$ described above is a feasible solution of the packing problem 4.2. Furthermore, the ojective function evaluated at this feasible solution is at least $\tilde{u}\delta D + C\delta$, where $\tilde{u} = C = \dfrac{k}{3k-2}$.*

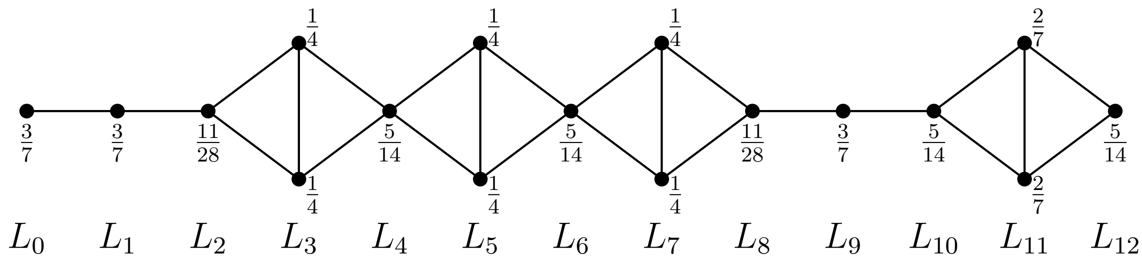


Figure 4.2: An example strongly canonical clump graph with $k = 3$. The weighting that is assigned to each vertex by the description above is shown using the partition of Figure 4.1.

Proof. Suppose that $H \in \mathcal{H}_{k,1,\delta}$. We begin by showing that the easy weighting given to $V(H)$ satisfies the requirements of Lemma 49. Since $\delta \geq 1$ and $D = 1$, H is the complete graph on $|H|$ vertices. It follows that the sum of the weights of the neighbors of every vertex is $(|H| - 1) \cdot \dfrac{1}{|H| - 1} = 1$, so the proposed weighting is a feasible solution of packing problem 4.2. Furthermore, the objective function of 4.2 under this weighting evaluates to $\delta \cdot \dfrac{|H|}{|H| - 1} = \delta \cdot \left(1 + \dfrac{1}{|H| - 1}\right) > \delta$. Since $\tilde{u} + C = \dfrac{k}{3k-2} + \dfrac{k}{3k-2} = \dfrac{2k}{3k-2} < 1$ when $k \geq 3$, $\tilde{u}\delta + C\delta = (\tilde{u} + C) \cdot \delta < \delta$. This completes the argument for the case when $D = 1$.

Now suppose that $H \in \mathcal{H}_{k,D,\delta}$ with $D \geq 2$, and suppose that $V(H)$ has been given the weighting above based on the partitioning of Lemma 48. To see that the objective function evaluated at the solution u is large enough, let us first add up all of the weight assigned to vertices of H. In particular, we will see that

$$u\Big(V(H)\Big) = \frac{k}{3k-2} \cdot (D + 1).$$

If $\{L_i, L_{i+1}, L_{i+2}\}$ is a Type 1 segment, then by Lemmas 47(g) and 47(f), $|L_{i+1}| = k - 1$ and $|L_i| = |L_{i+1}| = 1$. Therefore,

$$u(L_{i-1}) + u(L_i) + u(L_{i+1}) = \frac{k+2}{2(3k-2)} + (k-1) \cdot \frac{2}{3k-2} + \frac{k+2}{2(3k-2)}$$

$$= \frac{k+2}{3k-2} + \frac{2k-2}{3k-2} \tag{4.7}$$
$$= 3 \cdot \frac{k}{3k-2}.$$

If $\{L_i, L_{i+1}, \ldots, L_{i+2s}\}$ is a Type 2 segment, then Lemmas 47(g) and 47(f) again imply that the $s+1$ small layers of the segment all have size 1 and the s big layers of the segment all have size $k-1$. Therefore,

$$\begin{aligned}
&u(L_i) + u(L_{i+1}) + \cdots + u(L_{i+2s}) \\
&\quad = \underbrace{2 \cdot \frac{3k+2}{4(3k-2)}}_{\text{end layers}} + \underbrace{s(k-1) \cdot \frac{1}{2(k-1)}}_{\text{big layers}} + \underbrace{(s-1) \cdot \frac{k+2}{2(3k-2)}}_{\text{internal small layers}} \\
&\quad = \frac{3k+2}{2(3k-2)} + \frac{s}{2} + \frac{(s-1)(k+2)}{2(3k-2)} \\
&\quad = \frac{3k+2}{2(3k-2)} + \frac{s(3k-2)}{2(3k-2)} + \frac{sk+2s-k-2}{2(3k-2)} \\
&\quad = \frac{3k+2+3sk-2s+sk+2s-k-2}{2(3k-2)} \\
&\quad = \frac{2k+4sk}{2(3k-2)} \\
&\quad = \frac{2k(1+2s)}{2(3k-2)} \\
&\quad = (1+2s) \cdot \frac{k}{3k-2}.
\end{aligned}$$

If L_i is a Type 3 layer and $|L_i| \leq k/2$, then

$$u(L_i) = |L_i| \cdot \frac{k}{(3k-2)|L_i|} = \frac{k}{3k-2}.$$

Finally, suppose that L_i is a Type 3 layer with $|L_i| > k/2$. Then all vertices in S_i get weight $\frac{2}{3k-2}$, and all vertices in $L_i \setminus S_i$ get weight $\frac{k-2|S_i|}{(3k-2)(|L_i|-|S_i|)}$. Therefore,

$$u(L_i) = |S_i| \cdot \frac{2}{(3k-2)} + (|L_i| - |S_i|) \cdot \frac{k-2|S_i|}{(3k-2)(|L_i|-|S_i|)} = \frac{k}{3k-2}.$$

The above calculations show that $u(V(H)) = \frac{k}{3k-2} \cdot (D+1)$, as desired. We can now evaluate the objective function of 4.2:

$$\sum_{y \in V(H)} \delta \cdot u(y) = \delta \cdot \frac{k}{3k-2} \cdot (D+1) = \tilde{u}\delta D + C\delta,$$

where $\tilde{u} = C = \frac{k}{3k-2}$.

Now we will check that u is a feasible solution of 4.2, i.e. we check that $u(N(y)) \leq 1$ for all $y \in V(H)$. First suppse that $y \in L_i$, where L_i is in a segment of Type 1:

- Suppose L_i is the middle layer, which is big by definition. By utilizing Equation 4.7, we find

$$\begin{aligned} u(N(y)) &\leq u(L_{i-1}) + u(L_i \setminus \{y\}) + u(L_{i+1}) \\ &= u(L_{i-1} \cup L_i \cup L_{i+1}) - u(y) \\ &= 3 \cdot \frac{k}{3k-2} - \frac{2}{3k-2} \\ &= 1. \end{aligned}$$

- Suppose that L_i is one of the end-layers. Then Lemmas 47(g) and 47(f) imply that $|L_i| = 1$, i.e. $L_i \setminus \{y\} = \emptyset$. By Remarks 3 and 5, we find

$$\begin{aligned} u(N(y)) &\leq u(L_{i-1}) + u(L_i \setminus \{y\}) + u(L_{i+1}) \\ &= u(L_{i-1}) + u(L_{i+1}) \\ &\leq \frac{k}{3k-2} + (k-1) \cdot \frac{2}{3k-2} \\ &= 1. \end{aligned}$$

Next, suppose that $y \in L_i$, where L_i is in a segment of Type 2:

- Suppose that L_i is a big layer. By Remark 3, $u(N(y))$ is the largest when L_i is the second or second-to-last layer in the segment. Therefore,

$$u(N(y)) \leq u(L_{i-1}) + u(L_i \setminus \{y\}) + u(L_{i+1})$$

$$
\begin{aligned}
&= u(L_{i-1}) + u(L_i) + u(L_{i+1}) - u(y) \\
&\le \frac{3k+2}{4(3k-2)} + (k-1)\cdot\frac{1}{2(k-1)} + \frac{k+2}{2(3k-2)} - \frac{1}{2(k-1)} \\
&= \frac{3k+2}{4(3k-2)} + \frac{1}{2} + \frac{k+2}{2(3k-2)} - \frac{1}{2(k-1)} \\
&= \frac{11k^2-15k+2}{12k^2-20k+8} \\
&\le 1, \qquad \text{when } k \ge 3.
\end{aligned}
$$

- Now suppose that L_i is a small layer. In this case, $L_i = \{y\}$.

  - If L_i is not an end layer, then

$$
\begin{aligned}
u(N(y)) &\le u(L_{i-1}) + u(L_i \setminus \{y\}) + u(L_{i+1}) \\
&= u(L_{i-1}) + u(L_{i+1}) \\
&= 2(k-1)\cdot\frac{1}{2(k-1)} \\
&= 1.
\end{aligned}
$$

  - If L_i is an end layer, then by Remarks 3 and 5,

$$
\begin{aligned}
u(N(y)) &\le u(L_{i-1}) + u(L_i \setminus \{y\}) + u(L_{i+1}) \\
&= u(L_{i-1}) + u(L_{i+1}) \\
&\le \frac{k}{3k-2} + (k-1)\cdot\frac{1}{2(k-1)} \\
&= \frac{k}{3k-2} + \frac{1}{2} \\
&= \frac{5k-2}{6k-4} \\
&\le 1, \qquad \text{when } k \ge 3.
\end{aligned}
$$

Finally, suppose that L_i is a Type 3 layer.

- Suppose that $|L_i| \le \frac{k}{2}$ or $y \in S_i$ with $|L_i| > \frac{k}{2}$. Then by Remark 4 $u(y) \ge \frac{2}{3k-2}$. Furthermore, L_{i-1} and L_{i+1} are either other Type 3 layers or end-layers of Type

1 or Type 2 segments, so $u(L_{i-1}), u(L_i), u(L_{i+1}) \leq \frac{k}{3k-2}$. Therefore,

$$\begin{aligned} u(N(y)) &\leq u(L_{i-1}) + u(L_i \setminus \{y\}) + u(L_{i+1}) \\ &= u(L_{i-1}) + u(L_i) + u(L_{i+1}) - u(y) \\ &\leq 3 \cdot \frac{k}{3k-2} - \frac{2}{3k-2} \\ &= 1. \end{aligned}$$

- The last case that remains is when $y \in L_i \setminus S_i$ and $|L_i| > \frac{k}{2}$. By definition, there exists some $j \in \{i-1, i+1\}$ and some $w \in L_j$ such that $w \notin N(y)$. Similary to the previous case, we are done if we can show that $u(y) + u(w) \geq \frac{2}{3k-2}$. Since $y \notin N(w)$, $w \notin S_j$, i.e. $L_j \neq S_j$. Lemma 47 (d) implies that $|L_j| \neq 1$. Therefore, L_j is not an end-layer of a Type 1 or Type 2 segment, i.e. L_j is of Type 3. If $|L_j| \leq \frac{k}{2}$ then $u(w) \geq \frac{2}{3k-2}$ by Remark 4, and we're done. Otherwise, using the inequalities of Lemma 47(e), we get

$$\begin{aligned} u(v) + u(w) &= \frac{k - 2|S_i|}{(3k-2)(|L_i| - |S_i|)} + \frac{k - 2|S_j|}{(3k-2)(|L_j| - |S_j|)} \\ &\geq \frac{k - 2|S_i|}{(3k-2)(k - |S_i| - |S_j|)} + \frac{k - 2|S_j|}{(3k-2)(k - |S_i| - |S_j|)} \\ &= \frac{2}{3k-2}. \end{aligned}$$

□

Theorem 50. *Assume $k = 3$ or $k = 4$. If G is a connected k-colorable graph of order n with minimum degree at least $\delta \geq 1$, then* $\operatorname{diam}(G) \leq \left(3 - \frac{2}{k}\right) \frac{n}{\delta} - 1$.

Proof. With Lemma 49 in place, we may apply Theorem 42 with $\tilde{u} = C = \dfrac{k}{3k-2}$. □

It is natural to want to extend the ideas above to larger values of k. However, I will describe some calculations which are a stong indication that new ideas are needed to make progress for $k = 5$ and higher. In what follows, fix $k = 5$. Suppose that we wanted to partition the layers of a given strongly canonical clump graph into segments of Type 1, 2, and 3, as we did above. Then a segment of Type 1 would be three consecutive layers, L_i, L_{i+1}, and L_{i+2} such that L_i and L_{i+2} are small layers, while L_{i+1} is a big layer. Furthermore, layers L_{i-1} and L_{i+3} should both be small layers, as otherwise the layers L_i, L_{i+1}, and L_{i+2}, would belong to a longer Type 2 segement. As before, the small layers L_{i-1} and L_{i+3} will either be Type 3 layers, or end-layers of Type 1 or Type 2 segements. We would like Type 3 layers to have total weight exactly $\frac{5}{13} = \frac{k}{3k-2}$, and it is reasonable to assume that end-layers of Type 1 and Type 2 segements have total weight at most $\frac{5}{13} = \frac{k}{3k-2}$. In other words, we may assume that layers L_{i-1} and L_{i+3} both have total weight at most $5/13$.

With the ideas above in mind, consider a Type 1 segment with the additional specifications that $|L_i| = 2$, $|L_{i+1}| = 3$, $|L_{i+2}| = 2$, $L_i = S_i$, $L_{i+1} = S_{i+1}$, and $L_{i+2} = S_{i+2}$. This situation is drawn in Figure 4.3. It turns out that this configuration demonstrates why the technique of Theorem 50 cannot be extended. Assume that the layers L_{i-1} and L_{i+3} have total weight $5/13$ each, and note that $L_{i-1} \subseteq N(u)$ for all $u \in L_i$ and $L_{i+3} \subseteq N(v)$ for all $v \in L_{i+2}$. In this worst-case scenario, we then ask the question: How large can the total dual weight of $L_i \cup L_{i+1} \cup L_{i+2}$ be, while still maintaining the neighborhood condition for each vertex of $L_i \cup L_{i+1} \cup L_{i+2}$?

Let A, B, C, D, E, F, and G be variables representing the weights of the seven vertices of $L_i \cup L_{i+1} \cup L_{i+2}$ as in Figure 4.3. We can formulate the question above as a linear program:

$$\text{Maximize } A + B + C + D + E + F + G$$

subject to the conditions

- $B + C + D + E \leq 8/13$
- $A + C + D + E \leq 8/13$
- $A + B + D + E + F + G \leq 1$
- $A + B + C + E + F + G \leq 1$

- $A + B + C + D + F + G \leq 1$
- $C + D + E + G \leq 8/13$
- $C + D + E + F \leq 8/13$.

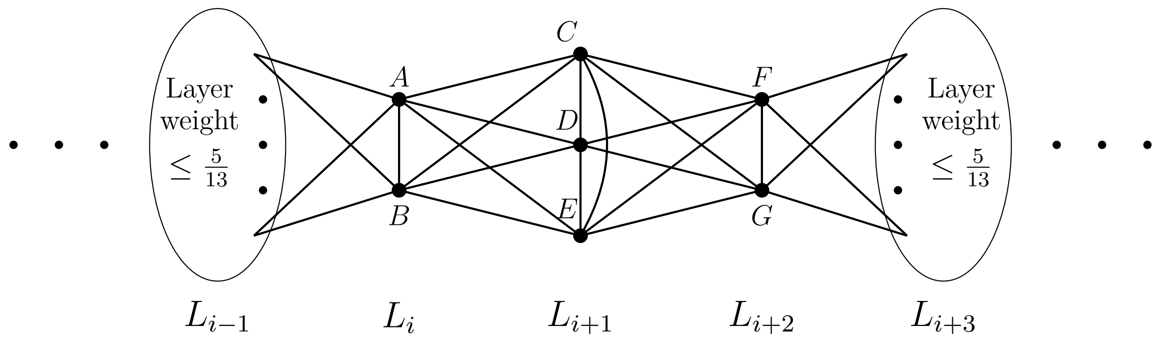


Figure 4.3: Layers L_i, L_{i+1}, and L_{i+2} form a Type 1 segment of a strongly canonical clump graph with $k = 5$. This segment demonsrates that the ideas of Theorem 50 cannot readily extend to $k \geq 5$.

Solving this linear program (we used Sage, see Appendix A), we find the optimal solution to be about 1.146. In order for the technique of Theorem 50 to work, the average dual weight of the three layers of any Type 1 segment must be at least $\frac{k}{3k-2}$. In other words, the total dual weight on $L_i \cup L_{i+1} \cup L_{i+2}$ needs to be at least $3 \cdot (5/13) \approx 1.154$. Since the optimal dual weighting for the Type 1 segment in consideration is strictly less than $3 \cdot (5/13)$, it is easy to imagine a strongly canonical clump graph which consists of many copies of this problematic Type 1 segment with only one Type 3 layer in between each copy (see Figure 4.4). For such a graph, the weighting used for Theorem 50 would not have sufficient weight to close the gap to Conjecture 2. Furthermore, the more copies of the problematic Type 1 segment that we use, the further away from the desired bound we will be.

It should be noted that, even though the weighting scheme used to prove Theorem 50 cannot be extended to general $k \geq 5$, I have not found any reason to suspect that Conjecture 2 is false. I have used Sage to find the optimal dual weighting of many graphs, including ones like that shown in Figure 4.4. The optimal solutions of my tests may not have followed the scheme used for Theorem 50, but they all have had total dual weight at least $\frac{k}{3k-2}(D+1)$. To make additional progress, we will likely have to modify the weighting scheme of Theorem 50 (maybe by altering the partitioning process) or we will have to create a new weighting scheme entirely.

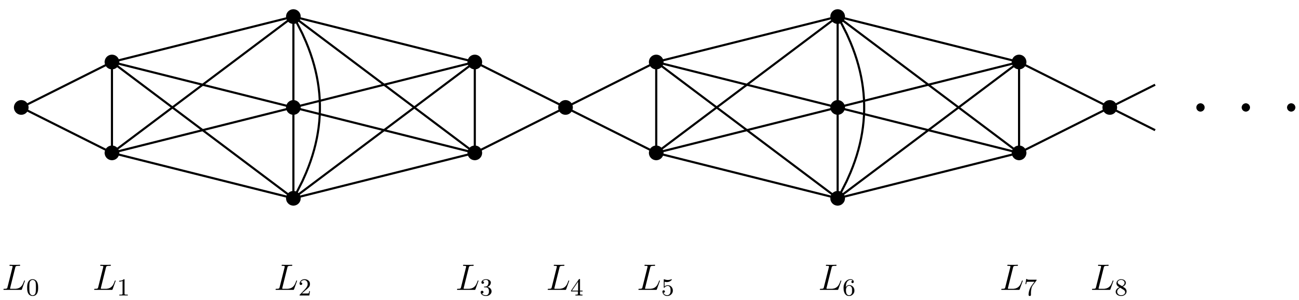


Figure 4.4: If we consider repeatedly alternating a Type 3 layer with a single vertex, and the problematic Type 1 segment, we'll get a strongly canonical clump graph for which the weighting scheme used for Theorem 50 is insufficient.

Bibliography

[1] D. Amar, I. Fournier, and A. Germa. "Ordre minimum d'un graphe simple de diamètre, degré minimum et connexité donnés". In: *Ann. Discrete Math.* **17** (1983), pp. 7–10.

[2] M.S. Bansal, Wen-Chieh Chang, O. Eulenstein, and D. Fernandez-Baca. "Generalized binary tanglegrams: algorithms and applications", pp. 114–125. In: *Bioinformatics and Computational Biology*, First International Conference, BICoB 2009, New Orleans, LA, USA, April 8–10, 2009.

[3] S. C. Billey, M. Konvalinka, and F.A. Matsen. "On the enumeration of tanglegrams and tangled chains". In: *Journal of Combinatorial Theory* **A146** (2017), pp. 239–263.

[4] K. Buchin, M. Buchin, J. Byrka, M. Nöllenburg, Y. Okamoto, R.I. Silveira, and A. Wolff. "Drawing (complete) binary tanglegrams. Hardness, approximation, fixed-parameter tractability". In: *Algorithmica* **62** (2012), pp. 309–332.

[5] A. Burt and R. Trivers. *Genes in Conflict*. Belknap Harvard Press, Cambridge, Mass., 2008. ISBN: 978-0674027220.

[6] L. Cacetta and W.F. Smyth. "Graphs of maximum diameter". In: *Discrete Math.* **102** (1992), pp. 121–141.

[7] Z. Che and K.L. Collins. "An upper bound on Wiener indices of maximal planar graphs". In: *Discrete Appl. Math.* **258** (2019), pp. 76–86. DOI: `https://doi.org/10.1016/j.dam.2018.11.026`.

[8] É. Czabarka, P. Dankelmann, T. Olsen, and L. A. Székely. "Wiener index and remoteness in triangulations and quadrangulations". In: *Discrete Mathematics & Theoretical Computer Science* **23**(1) (2021).

[9] É. Czabarka, P. Dankelmann, and L. A. Székely. "Diameter of 4-colorable graphs". In: *Europ. J. Comb.* **30** (2009), pp. 1082–1089.

[10] É. Czabarka, I. Singgih, and L. A. Székely. "Counterexamples to a conjecture of Erdős, Pach, Pollack and Tuza". In: *J. Combin. Theory* **B151** (2021), pp. 38–45.

[11] É. Czabarka, I. Singgih, and L. A. Székely. "On the diameter of k-colorable graphs". In: *Electronic J. of Comb* **28**(3) (2021), P3.52.

[12] É. Czabarka, L.A. Székely, and S. Wagner. "Inducibility in binary trees and crossings in tanglegrams". In: *SIAM J. Discrete Math.* **31**(3) (2017), pp. 1732–1750.

[13] É. Czabarka, L.A. Székely, and S. Wagner. "A tanglegram Kuratowski theorem". In: *Journal of Graph Theory* **90**(2) (2019), pp. 111–122.

[14] R. Diestel. *Graph Theory.* Fifth Edition. Graduate Texts in Mathematics 173, Springer, Berlin, Heidelberg, 2017. ISBN: 978-3642142789.

[15] A. A. Dobrynin, R. Entringer, and I. Gutman. "Wiener index of trees. Theory and applications". In: *Acta Appl. Math.* **66** (2001), pp. 211–249.

[16] R. C. Entringer, D. E. Jackson, and D. A. Snyder. "Distance in graphs". In: *Czech. Math J.* **26** (1976), pp. 283–296.

[17] P. Erdős, J. Pach, R. Pollack, and Z. Tuza. "Radius, diameter, and minimum degree". In: *J. Combin. Theory* **B47** (1989), pp. 279–285.

[18] H. Fernau, M. Kaufmann, and M. Poths. "Comparing trees via crossing minimization", pp. 457–469. In: *Proc. 25th Intern. Conf. Found. Softw. Techn. Theoret. Comput. Sci.* (FSTTCS'05), Lecture Notes in Computer Science **3821**, Springer-Verlag, New York City, NY, 2005; journal version *J. Computer System Sciences* **76** (2010), pp. 593–608.

[19] M. Fischermann, A. Hoffmann, D. Rautenbach, L. Székely, and L. Volkmann. "Wiener index versus maximum degree in trees". In: *Discr. Appl. Math.* **122** (2002), pp. 127–137.

[20] D. Ghosh, E. Győri, A. Paulos, N. Salia, and O. Zamora. "The maximum Wiener index of maximal planar graphs". In: *Journal of Combinatorial Optimization* **40**(4) (2020), pp. 1121–1135. DOI: `https://doi.org/10.1007/s10878-020-00655-4`.

[21] D. Goldsmith, B. Manvel, and V. Faber. "A lower bound for the order of the graph in terms of the diameter and minimum degree". In: *J. Combin. Inform. Syst. Sci.* **6** (1981), pp. 315–319.

[22] E. Győri, A. Paulos, and Chuanqi Xiao. "Wiener index of quadrangulation graphs". In: *Discrete Applied Mathematics* **289** (2021), pp. 262–269. DOI: `https://doi.org/10.1016/j.dam.2020.11.016`.

[23] M.S. Hafner and S.A. Nadler. "Phylogenetic trees support the coevolution of parasites and their hosts". In: *Nature* **332** (1988), pp. 258–259.

[24] F. Jelen and E. Triesch. "Superdominance order and distance of trees with bounded maximum degree". In: *Discrete Appl. Math.* **125**(2-3) (2003), pp. 225–233.

[25] M. Konvalinka and S. Wagner. "The shape of random tanglegrams". In: *Advances in Applied Mathematics* **78** (2016), pp. 76–93.

[26] J.B. Kruskal. "Well-quasi-ordering, the tree theorem, and Vázsonyi's conjecture". In: *Transactions of the American Mathematical Society*, American Mathematical Society **95**(2) (1960), pp. 210–225. DOI:10.2307/1993287.

[27] R. Laver. "Well-quasi-orderings and sets of finite sequences". In: *Math. Proc. Camb. Philos. Soc.* **79** (1976), pp. 1–10.

[28] F.A. Matsen, S.C. Billey, A. Kas, and M. Konvalinka. "Tanglegrams: a reduction tool for mathematical phylogenetics". In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **15**(1) (2018), pp. 343–349.

[29] J.W. Moon. "On the diameter of a graph". In: *Mich. Math. J.* **12** (1965), pp. 349–351.

[30] T. Olsen. *Distance Related Graph Invariants in Triangulations and Quadrangulations of the Sphere.* Ph.D. Dissertation, University of South Carolina, Columbia, SC, 2020.

[31] R.D.M. Page (ed). *Tangled Trees. Phylogeny, Cospeciation and Coevolution.* University of Chicago Press, Chicago, IL, 2002. ISBN: 978-0226644660.

[32] V.R. Pratt. "Computing permutations with double-ended queues, parallel stacks and parallel queues". In: *Proc. ACM Symp. Theory of Computing* **5** (1973), pp. 268–277.

[33] C. Scornavacca, F. Zickmann, and D.H. Huson. "Tanglegrams for rooted phylogenetic trees and networks". In: *Bioinformatics* **27**(13) (2011), pp. 248–256.

[34] R. Shi. "The average distance of trees". In: *Systems Sci. Math. Sci.* **6**(1) (1993), pp. 18–24.

[35] A. V. Sills and H. Wang. "On the maximal Wiener index and related questions". In: *Discr. Appl. Math.* **160** (2012), pp. 1615–1623.

[36] D.A. Spielman and M. Bóna. "An infinite antichain of permutations". In: *Electronic Journal of Combinatorics* **7**N2 (2000). DOI: `https://doi.org/10.37236/1540`.

[37] L.A. Székely, S. Wagner, and Hua Wang. "Problems related to graph indices in trees", pp. 3–31. In: *Recent Trends in Combinatorics*, eds. A. Beveridge, J.R. Griggs, L. Hogben, G. Musiker, and P. Tetali. The IMA Volumes in Mathematics and Applications 159, Springer-Verlag, New York City, NY, 2016. ISBN: 978-3319242965.

[38] R.E. Tarjan. "Sorting using networks of queues and stacks". In: *J. of the ACM.* **19** (1972), pp. 341–346.

[39] H. Wang. “The extremal values of the Wiener index of a tree with given degree sequence”. In: *Discr. Appl. Math.* **156** (2009), pp. 2647–2654.

[40] H. Wiener. “Structural determination of paraffin boiling points”. In: *J. Amer. Chem. Soc.* **69** (1947), pp. 17–20.

[41] H. Wiener. “Correlation of heats of isomerization, and differences in heats of vaporization of isomers, among the paraffin hydrocarbons”. In: *J. Amer. Chem. Soc.* **69** (1947), pp. 2636–2638.

[42] X. D. Zhang, Y. Liu, and M. Han. “The maximum Wiener index of trees with given degree sequences”. In: *MATCH Commun. Math. Comput. Chem.* **64** (2010), pp. 661–682.

[43] X. D. Zhang, Q, Xiang, L. Xu, and R. Pan. “The Wiener index of trees with given degree sequences”. In: *MATCH Commun. Math. Comput. Chem.* **60** (2008), pp. 623–644.

Appendix A

Code

```
#Problem case broken down into individual vertices
p = MixedIntegerLinearProgram()
v = p.new_variable(real=True, nonnegative=True)
A, B, C, D, E, F, G = v['A'], v['B'], v['C'], v['D'], v['E'],
    v['F'], v['G']
p.set_objective(A + B + C + D + E + F + G)
p.add_constraint(B + C + D + E <= 8/13)
p.add_constraint(A + C + D + E <= 8/13)
p.add_constraint(A + B + D + E + F + G <= 1)
p.add_constraint(A + B + C + E + F + G <= 1)
p.add_constraint(A + B + C + D + F + G <= 1)
p.add_constraint(G + C + D + E <= 8/13)
p.add_constraint(F + C + D + E <= 8/13)
round(p.solve(), 3)
```

Figure A.1: Solving the linear program at the end of Section 4.4 to find the maximum dual weighting on the problematic Type 1 segment.