Spring 2022

Tangled up in Tanglegrams

Drew Joseph Scalzo

Tangled up in Tanglegrams

by

Drew Joseph Scalzo

Bachelor of Science
The University of Akron 2020

Submitted in Partial Fulfillment of the Requirements

for the Degree of Masters of Science in

Mathematics

College of Arts and Sciences

University of South Carolina

2022

Accepted by:

Eva Czabarka, Director of Thesis

Laszlo Szekely, Reader

Tracey L. Weldon, Interim Vice Provost and Dean of the Graduate School

Dedication

I dedicate this Thesis to my parents.

Acknowledgments

First, I would like to thank my parents for all the emotional support you have provided me throughout this process. For being the second reader of this thesis, I would like to thank Dr. Laszlo Szekely. Also, a big thanks goes out to Stephen Smith for helping me with all the tex and formatting questions I had. Finally, I would like to thank my thesis advisor, Dr. Eva Czabarka. Without her suggestions, clarifications, and insight this thesis would not have been possible.

Abstract

Tanglegrams are graphs consisting of two rooted binary plane trees with the same number of leaves and a perfect matching between the two leaf sets. A Tanglegram drawing is a special way of drawing a Tanglegram; and a Tanglegram is called planar if it has a drawing such that the matching edges do not cross. In this thesis, we will discuss various results related to the construction and planarity of Tanglegrams, as well as demonstrate how to construct all the Tanglegrams of size 4 by looking at two types of rooted binary trees - Caterpillar and Complete Binary Trees. After augmenting a Tanglegram with an edge between its roots, we will prove that the Tanglegram crossing number of the original Tanglegram is greater than or equal to the crossing number of the augmented Tanglegram taken as a graph. We will show that the removal of a matching edge from a Tanglegram of size $n \geq 3$ decreases the Tanglegram crossing number by at most $n - 3$, and give a family of 1-edge panar Tanglegrams (one for every $n \geq 3$) of size n with Tangle crossing number $n - 3$, showing that the previous statement is sharp. We will also discuss various conditions on the nonplanarity of Tanglegrams.

Table of Contents

List of Figures

Chapter 1

Introduction

In this thesis, we begin in Chapter 2 by introducing the definitions needed to establish elementary results that are necessary to facilitate future discussions on Tanglegrams. We mention several results related to trees, due to their connection in the construction of Tanglegrams. At times, we make distinctions between multiple forms of notation as they relate to how it will be presented in this thesis.

In Chapter 3, we first focus on building our definition of Tanglegrams by talking about both complete binary and caterpillar trees, make a distinction between unlabelled and labelled Tanglegrams, and various operations that can be performed on Tanglegrams. Then, our discussion is broken several parts, all of which illustrate how to construct Tanglegrams of size 4. First, we consider the type of left and right rooted binary trees that will be the bases of our Tanglegrams. Then, we focus on determining all the matchings possible between the leaves of the two rooted binary trees. We also provide many illustrations of the process throughout the chapter.

In Chapter 4, we discuss the planarity of Tanglegrams and rely on the idea of the Tanglegram crossing number to generalize the planarity of Tanglegrams. Then, we focus on proving several results related to Tanglegrams and the augmentation of these graphs with an edge connecting the roots of the two rooted binary trees. Our discussion concludes with proving a Theorem relates the Tanglegram crossing number of a Tanglegram to a Tanglegram with one less matching edge.

Finally, we conclude this thesis in Chapter 5 by briefly touching on notable properties of induced Subtanglegrams, and in particular, extend the prior planarity arguments to these types of graphs.

Chapter 2

Elementary Graph Notions

To begin discussing anything of importance to Tanglegrams, we need to state quite a few central definitions.

Definition 1. For a set A and a non-negative integer k, $\binom{A}{k}$ denotes the collection of *k-element subsets of A.*

Definition 2. A *graph* $G = (V, E)$ consists of a set V of vertices and a set E of edges where $E \subseteq \binom{V}{2}$. We use the notation $e = xy$ for $e = \{x, y\}$, where $e \in E$ and $x, y \in V$; we assume that $V \neq \emptyset$.

Definition 3. We define the *order* of G to be the number of vertices of G; denoted by $|V|$ or $|V(G)|$. Additionally, we define the *size* of G to be the number of edges of G; denoted by $|E|$ or $|E(G)|$.

In this thesis, we make no distinction between $|V|$ and $|V(G)|$ unless otherwise stated. Similarly for $|E|$ or $|E(G)|$. Also, all graphs discussed in this thesis are simple finite graphs in that they have at most one edge between any two vertices in V. Note that our definition does not allow loops, which are edges whose two endpoints are the same.

Definition 4. We say a vertex v is *incident* with an edge e if v is an endpoint of e. Similarly, v is *adjacent* to another vertex u if v and u are connected by an edge; u and v are called *neighbors*.

Definition 5. The *complete graph* on n vertices, denoted K_n, is the graph where all vertices are pairwise adjacent; equivalently, it is the graph where every vertex is connected to every other vertex by an edge.

Complete graphs are of particular importance in this thesis, as they will allow us to generalize the planarity of graphs. Figure 2.1 depicts an illustration of K_3, K_4, and K_5, the complete graphs on $3, 4$, and 5 vertices, respectively.

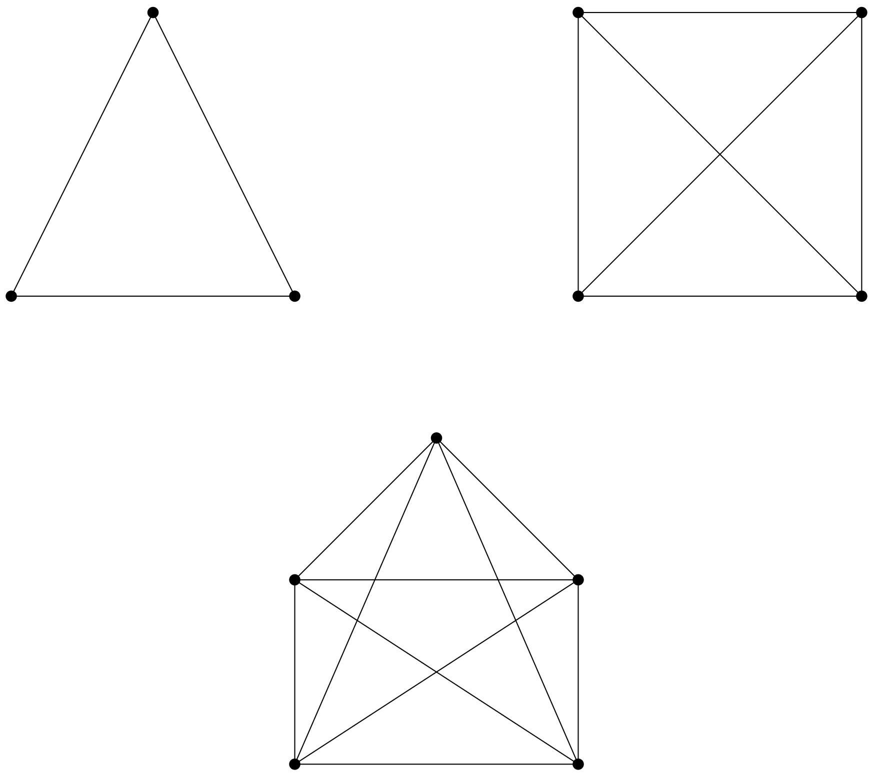


Figure 2.1 The complete graphs K_3 (top left), K_4 (top right), and K_5 (bottom).

Definition 6. Let $G = (V, E)$ and $G' = (V', E')$ both be graphs such that $V' \subseteq V$ and $E' \subseteq E$. Then, we say G' is a *subgraph* of G, and write it as $G' \subseteq G$. If $G' \subseteq G$ but $G' \neq G$, then G' is a *proper subgraph* of G.

Definition 7. Let $G = (V, E)$ and be a graph. Then, an *induced subgraph* $G' = (V', E')$ of G is a subgraph formed from a subset $V' \subseteq V$ and all of the edges of E (i.e. $E' = E \cap \binom{V'}{2}$) that connect pairs of vertices in that subset. Furthermore, a *spanning subgraph* of G is a subgraph that contains all the vertices of the original graph.

We now quickly mention some important notation revolving around the removal of one or more vertices or edges of a graph.

Notation 1. For a graph G and a vertex $v \in V(G)$, $G - v$ denotes the graph induced by $V(G) - \{v\}$. Similarly, let $A = \{v_1, v_2, \ldots, v_n\} \subseteq V(G)$ such that $v_i \in V$ for all $1 \leq i \leq n$. We denote the graph G with the removal of all the vertices contained in A by $G - A$, where the order of the removal of the vertices of A does not matter. Additionally, for $G = (V, E)$ and $v \notin V$, $G + v = (V \cup \{v\}, E)$.

Notation 2. For a graph G and an edge $e \in E(G)$, $G - e$ denotes the graph with vertex set $V(G)$ and edge set $E(G) - \{e\}$. Similarly, let $B = \{e_1, e_2, \ldots, e_n\} \subseteq E(G)$ such that $e_j \in E$ for all $1 \leq j \leq n$. We denote the graph G with the removal of all the edges contained in B by $G - B$, where the order of the removal of the vertices of B does not matter. Similar notation follows for the addition of a edge e: $G + e$.

Now, back to some key definitions needed to elaborate on future results.

Definition 8. The *degree* of a vertex v in a graph G, denoted by $\deg_G(v)$, is the number of edges incident to v; i.e. the number of edges that have v as an endpoint. We define the numbers $\delta(G)$ and $\Delta(G)$ as follows.

$$\delta(G) = \min\{\deg_G(v) \mid v \in V(G)\}$$
$$\Delta(G) = \max\{\deg_G(v) \mid v \in V(G)\}$$

Definition 9. Let $G = (E, V)$ be a graph. We define a *path* on G to be a sequence of $k+1$ <u>distinct</u> vertices $v_0 v_1 \ldots v_k$ (where $k \in \mathbb{N}$ such that for each $i \in [k]$, $v_{i-1}v_i \in E$). Note, v_0 and v_k are the *endvertices* of the path, and the *length* of the path is the number of edges on the path, which is k.

Definition 10. Let $G = (E, V)$ be a graph. We define a *cycle* on G to be a sequence of vertices $v_1 v_2 \ldots v_k v_1$ such that $v_1 \ldots v_k$ is a path P and $v_k v_1$ is an edge of the graph that is not on the path. Note that this implies that k is at least 3. The *length* of the cycle is the number of edges on the cycle, which is k.

Definition 11. A graph G is *connected* if there is a path between any two vertices in G.

Now that we have defined what is meant by a connected graph, we can discuss some important terms and concepts related to induced subgraphs and minimal spanning subgraphs.

Definition 12. Let $G = (V, E)$ be a graph. A *component* of G is the maximal connected subgraph of G. Note that this means components have disjoint vertex sets.

We claim that the components of a graph are induced subgraphs. So, consider a component H of G. By definition it is a subgraph, its vertex set is some $U \subseteq V$. Let $u_1, u_2 \in U$ such that $u_1u_2 \in E$. If $u_1u_2 \notin H$, then $H' = H + u_1u_2$ is a subgraph of G, still connected (as any two vertices of H' has a path between then even in H), and H' contains H as a proper subgraph, contradicting that H was a maximal connected subgraph. Thus all edges connecting vertices in U are edges of H, showing that H is induced.

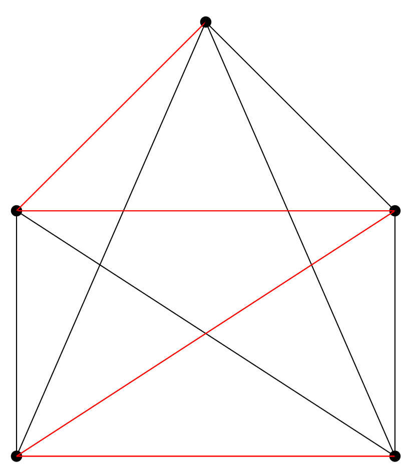


Figure 2.2 A minimal connected spanning subgraph of K_5.

We will "cycle" back to graphs in a moment, but now it is time to throw away cycles to begin with and discuss cycle-free graphs.

Definition 13. A *tree* $T = (V, E)$ is a connected graph that is *acyclic*, or cycle-free. All the vertices with degree at most one are called *leaves*. Any other vertex in T is called an *internal vertex*. The *order* of T, denoted $|T|$, is the number of vertices of T. Note that the vertex in the single vertex tree a leaf.

Theorem 1. *For a graph T on n vertices the following are equivalent:*

1. *T is a tree.*

2. *There is a unique path between any two vertices of T.*

3. *T is minimally connected; i.e. T is connected but $T - e$ is not for all edges $e \in T(E)$.*

4. *T is maximally acyclic; i.e. T contains no cycles but $T + e$ does for $e = xy$, where x and y are any two non-adjacent vertices in T.*

5. *T is connected and has $n - 1$ edges.*

6. *T is acyclic and has $n - 1$ edges.*

Most of the proofs for the equivalences above can be found in [5]. We will show that latter result - if T is a tree with n vertices, then T has $n - 1$ edges, simply because the proof is a standard exercise.

Proposition 2. *If T is a tree with n vertices, then T has $n - 1$ edges.*

Proof. Let T be a tree such that $|T| = n$. We will work by induction on n.

Suppose $|T| = 1$. Then, T is a singleton point with no edges and the result holds.

Suppose $|T| = 2$. Then, T has two vertices that must be joined by an edge to adhere to T being a connected graph. As this path is unique, we can only have 1 edge in our graph. So, the result holds.

Now suppose the result holds for all $|T| = n$ and let $T' = (V', E')$ be a tree with $n + 1$ vertices. We claim that T' has a leaf. Take a longest path in a tree with $n \geq 2$ vertices. The endvertices of this path have degree 1. Otherwise you have either a cycle, or you can lengthen you path, contradicting the choice of longest path.

So, let v be a leaf of T', $u \in V(T')$ be a internal vertex, and $e = vu$ be the unique edge connecting v to u. By removing the leaf v and edge e, we have a resulting graph G' that is still connected, but with one less vertex: $|G'| = (n+1)-1 = n$. Then, by the induction hypothesis, G' has $n-1$ edges. Thus, we have that $|E(T')| = (n-1)+1 = n$, and so T' is a tree with n vertices and $n-1$ edges. By induction, we are finished. □

In this thesis, we need to specify a type of tree that will help us form future graphs. The particular type of tree we care to discuss is called a *binary tree*.

Definition 14. A *rooted tree* is a tree with a vertex specified as a root. If T is a rooted tree with root r, then for any non-root vertex y, the *parent* of y is the neighbor of y on the unique $r - y$ path in T; the root has no parent. If y is any vertex of T, then the children of y are those neighbors of y that are not the parent of y.

Note that this means that all neighbors of the root are the children of the root, and leaves have no children.

Definition 15. A *rooted binary tree* L is a rooted tree where any non-leaf vertex has precisely two children. Any non-leaf vertices are *internal vertices*. The *size* of L is the number of leaves L has. Note that “size” here differs from “size” in Definition 3, as we know exactly the number of edges in trees, we rarely speak about the number of edges in them in the coming discussion. A *cherry* of L are two leaves with the same parent.

Definition 16. A graph $G = (V, E)$ is *bipartite*, denoted $K_{m,n}$, if V admits a partition into 2 classes, or sets A and B with $|A| = m$ and $|B| = n$, such that every edge has its ends in each class and every 2 vertices of G in the same class must not be adjacent. A bipartite graph is *complete* if each pair of vertices from different classes are adjacent. Figure 2.3 gives three different drawings of $K_{3,3}$, the complete bipartite graph on two classes each containing 3 vertices.

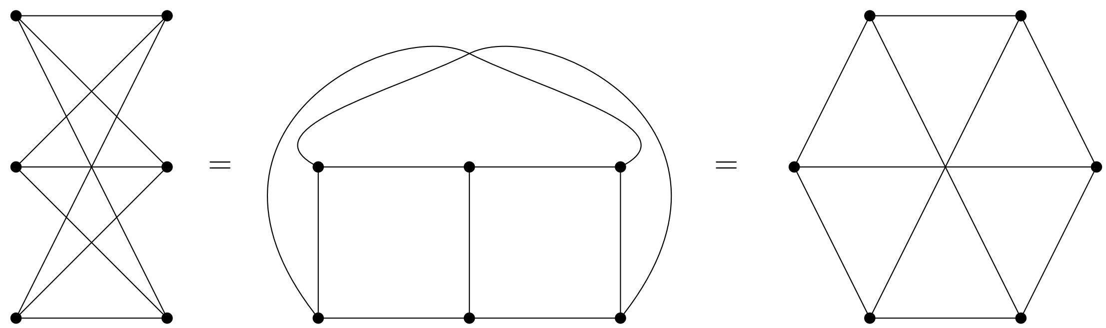


Figure 2.3 Three different drawings of the bipartite graph $K_{3,3}$. - Fig 1.6.2 from [5].

Definition 17. If $G = (V, E)$ is a graph, and $e = uv$ is an edge of G, then G/e (G *contracted on the edge* e) is the graph with vertex set $(V \setminus \{u, v\}) \cup \{w_{uv}\}$ and edge set $\{xy : xy \in E, \{x, y\} \cap \{u, v\} = \emptyset\} \cup \{xw_{uv} : x \in V - \{u, v\}, (xu \in E \text{ or } xv \in E)\}$. In other words, we remove the edge uv, identify the vertices u, v (the resulting vertex is called w_{uv}) and remove any duplicates from the resulting edges.

Definition 18. A *minor* of a graph G is a graph that can be obtained from a subgraph by edge contractions.

Definition 19. A *subdivision* of a graph G is a graph obtained by replacing edges of G with new paths of length at least one connecting the endpoints of the former edge. The new vertices introduced with the paths differ from the old vertices of G and new vertices introduced for different paths are distinct. A subdivision never decreases the number of vertices. Figure 2.5 illustrates this concept.

Definition 20. A *planar drawing* is a drawing where the vertices of the graph are represented by different points in the plane and edges are represented by simple

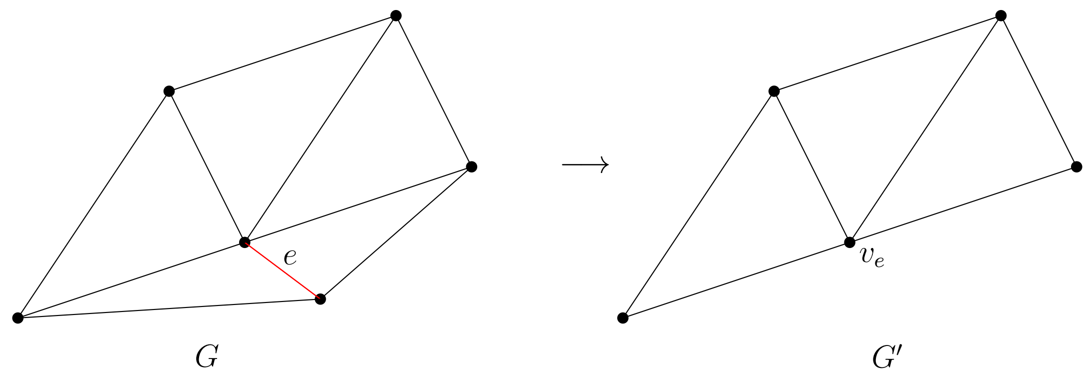


Figure 2.4 Contracting the edge e of G to obtain G'. This results in G' being a minor of G.

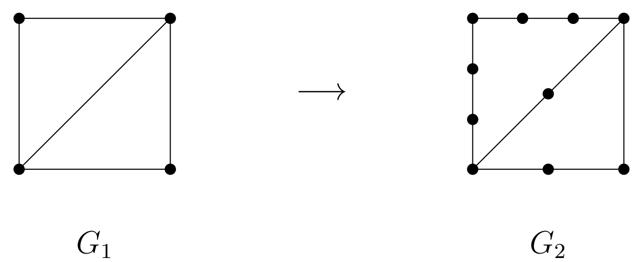


Figure 2.5 The graph G_2 is a subdivision of the graph G_1.

curves connecting their endpoints (i.e. the points representing their endvertices) such that the interior of these curves are disjoint from the set of points that are vertices of the graph and from each other. A *planar graph* is any graph that has a planar drawing.

Definition 21. A graph G is *planar* if it can be drawn in a way such that no edges intersect each other except at their endpoints.

We illustrate the concept of subdividing a graph in Figure 2.5. This idea is crucial to understanding the planarity generalization embedded in Kuratowski's Theorem.

Theorem 3 (Kuratowski's Theorem, [7]). *A graph is nonplanar if and only if it contains a subdivision of $K_{3,3}$ or K_5.*

Definition 22. A *matching* of a graph is a subset M of the edge set where each vertex has either one or zero edges incident to it in M. The matching M is perfect if every vertex is connected to exactly one edge in M.

In Chapter 3, we explore the properties and characteristics of a particular type of graph called a Tanglegram.

Chapter 3

Introduction to Tanglegrams

We now turn to a much deeper discussion on a peculiar type of graph known as a Tanglegram. Tanglegrams play an important role in phylogenetics, particularly in the theory of co-speciation. As we will see by its construction, the two rooted trees that form a Tanglegram represent the phylogenetic tree of hosts and the phylogenetic tree of their parasites. Although the results discussed in this thesis do not pertain to any significant biological discussion or applications, one could find themselves drawn to Tanglegrams simply for their inherit ability to illustrate the beauty of co-species relations; refer to [8] for more information.

Recall that in Definiton 15, we defined what a rooted binary tree is. The binary tree drawn in Figure 3.1 is called a *Caterpillar*, but different drawings of a Caterpillar easy to make. We define the two types of importance below.

Definition 23. A *Rooted Caterpillar*, denoted $\mathcal{C}_n$, of size n is the unique rooted binary tree with n leaves such that there are two leaves of distance $n-1$ from the root and for each $i \in \{1, 2, \ldots, n-2\}$, there is one leaf of distance i from the root. Figure 3.1 gives a progression of Caterpillar trees up to size 4.

Definition 24. A *Complete Binary Tree* is a rooted binary tree, with height k, where every leaf is at distance k from the root. All Complete Binary Trees have 2^k leaves. Figure 3.2 gives a progression of the Complete Binary Trees up to height 3.

By assigning a matching between the leaves of two arbitrary rooted binary trees of the same order, we can finally construct the Tanglegram.

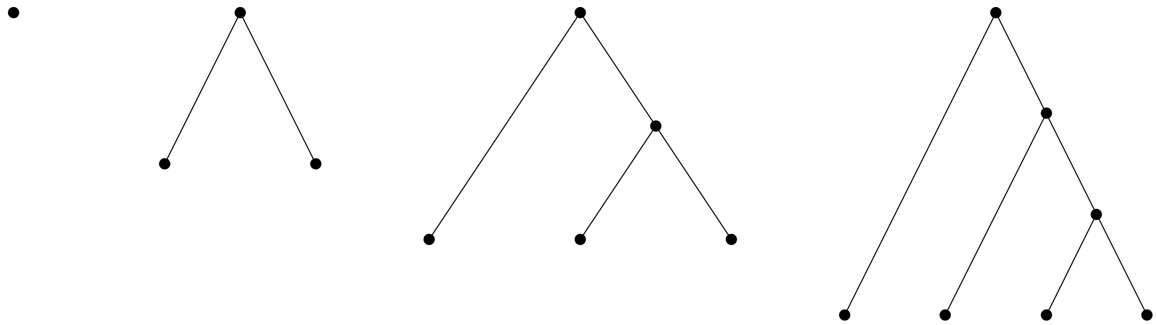


Figure 3.1 A progression of Caterpillar Trees of size 1 to 4.

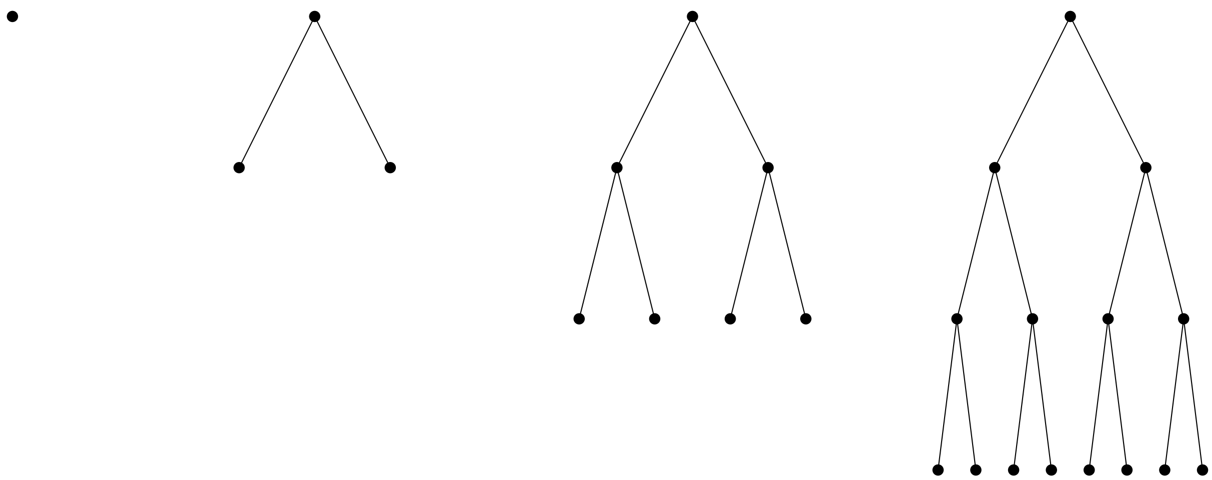


Figure 3.2 A progression of Complete Binary Trees of height 0 to 3.

Definition 25. A *Tanglegram* $(\mathcal{L}, \mathcal{R}, \sigma)$ *of size* n is a graph consisting of an n-leaf left binary tree $\mathcal{L}$ with a root r, an n-leaf right binary tree $\mathcal{R}$ with a root p, and a perfect matching σ between the leaves of $\mathcal{L}$ and $\mathcal{R}$. The *size* of a Tanglegram is the number of leaves in $\mathcal{L}$ or $\mathcal{R}$.

For much of graph theory, we do not care for labels of vertices on the graphs. The same follows for Tanglegrams. Some of the Tanglegrams depicted here have labels on their vertices and some do not.

Definition 26. Two Tanglegrams (L_1, R_1, σ_1) and (L_2, R_2, σ_2) are considered the same if there exists a bijection between the vertex sets of the two Tanglegram such that the following are preserved.

- The left root maps to the left root.
- The right root maps to the right root.
- The bijection makes a graph isomorphism between the two Tanglegrams.

If you consider size n Tanglegrams as a labeled graph, the bijection described in the previous definition defines an equivalence relation on these Tanglegrams, and the different equivalence classes of this relation can be considered as the different Tanglegrams. These Tanglegrams correspond to graphs where only the roots of the left tree and the right tree have labels (identifying which tree they belong to as roots).

Definition 27. A *Tanglegram Layout of $(\mathcal{L}, \mathcal{R}, \sigma)$* is a straight line drawing such that:

1. A left planar binary tree that is isomorphic to $\mathcal{L}$, with a root r and tree drawn in the plane where $x \leq 0$, and whose leaves are on the line $x = 0$.
2. A right planar binary tree that is isomorphic to $\mathcal{R}$, with a root p and tree drawn in the plane where $x \geq 1$, and whose leaves are on the line $x = 1$.
3. A perfect matching σ between their leaves drawn in straight line segments.

Definition 28. A *switch* on the Tanglegram Layout of a Tanglegram $(\mathcal{L}, \mathcal{R}, \sigma)$ is the following operation: select an internal vertex v of one of the two trees $\mathcal{L}$ and $\mathcal{R}$ and change the order of its two children. Then, draw the subtrees rooted at the children the same way as they were drawn before. Figure 3.3 illustrates a switch on a Tanglegram of size 4 at a root p.

To illustrate the difference between labeled and unlabeled Tanglegrams, consider the 2 Tanglegrams presented in Figure 3.4.

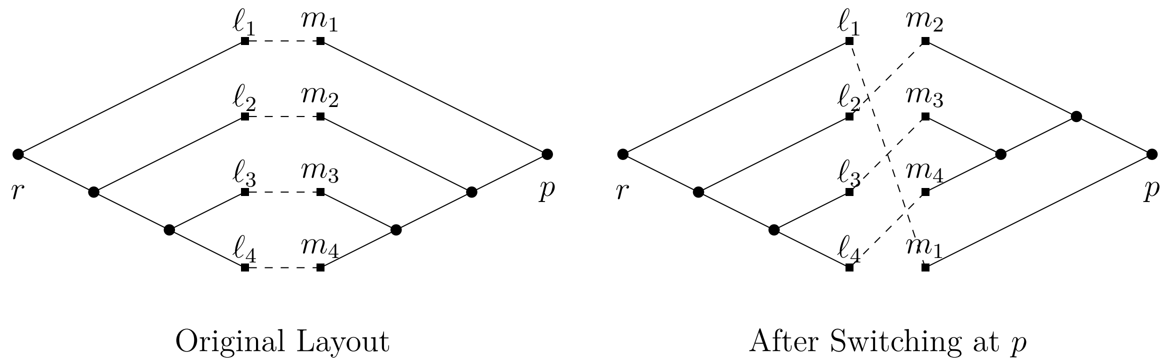


Figure 3.3 The results of a switch operation.

The left tree in both of them is a size 3 Tanglegram with root r, other internal vertex a, where r is adjacent to leaf ℓ_3, and a is adjacent to leaves ℓ_1 and ℓ_2. Also, the right tree in both of them has root p, other internal vertex b, where p is adjacent to leaf m_3, and b is adjacent to leaves m_1 and m_2.

The matching in the first one is $\{m_1\ell_1, m_2\ell_2, m_3\ell_3\}$ and in the second one is $\{m_1\ell_2, m_2\ell_1, m_3\ell_3\}$. They are different as labeled graphs, but the following isomorphism shows that they are the same Tanglegram:

$$f(v) = \begin{cases} v, & \text{if } v \notin \{\ell_1, \ell_2\} \\ \ell_1, & \text{if } v = \ell_2 \\ \ell_2, & \text{if } v = \ell_1 \end{cases}$$

The 16 pictures in the Figures 3.5 and 3.6 llustrate the different layouts of the same Tanglegram. You have 16 different layouts if you consider the Tanglegram as a labeled graph, but only 8 as a Tanglegram layout.

We then show that there are 16 different labelled Tanglegram layouts of this labelled Tanglegram. As unlabelled and labelled Tanglegrams they are all the same, but these are different labelled Tanglegram drawings of the same labelled Tanglegram. Figures 3.5 and 3.6 show all 16 drawings.

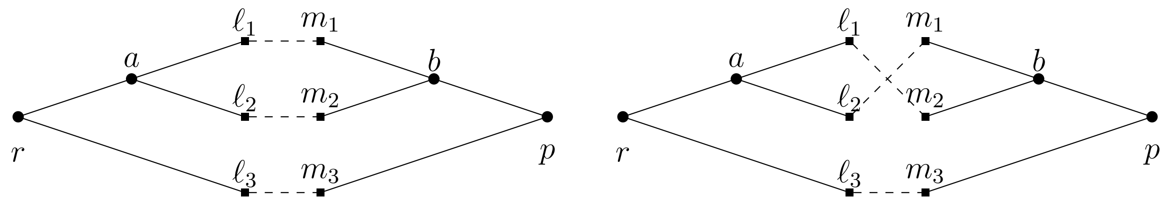


Figure 3.4 These Tanglegrams are different as labeled Tanglegrams, but are the same as unlabelled Tanglegrams.

Figure 3.7 gives a picture of a Tanglegram of size 4 using two Caterpillar trees as its two rooted binary trees. The perfect matching of its leaves is notated by dotted lines between its leaves. There happens to be an additional 12 distinct Tanglegrams of size 4, which is illustrated later in Figure 3.14.

We can determine the distinct number of Tanglegrams of a particular size by adhering to only those drawings that do not result in a symmetry of any other. There is a unique binary tree with 1 leaf: the singleton vertex. We can obtain all rooted binary trees with $n > 1$ leaves by using the rooted binary trees with $n - 1$ leaves and adding two children to one of their leaves. Now this means that there is only one Tanglegram of size 1: both the left- and right tree is a singleton vertex, and the matching connects these two vertices.

So for 2 leaves, we take the singleton vertex and add two leaves to it (this is both a rooted Caterpillar and a rooted Complete Binary tree of height 1) - there is only one such tree. For Tanglegrams of size 2, the left and right trees must be equal to this unique tree, and there is only one way to match them (there are two ways if we consider them as labeled graphs, but they are the same Tanglegram). See Figure 3.8.

For rooted binary trees on 3 leaves, there is one rooted (unlabeled) binary tree T on 2 leaves. As the two leaves are not distinguishable (a tree isomorphism takes one into the other) there is only one way to obtain a rooted binary tree on 3 -leaves by appending two children on a leaf of T. The resulting tree is a rooted Caterpillar; with

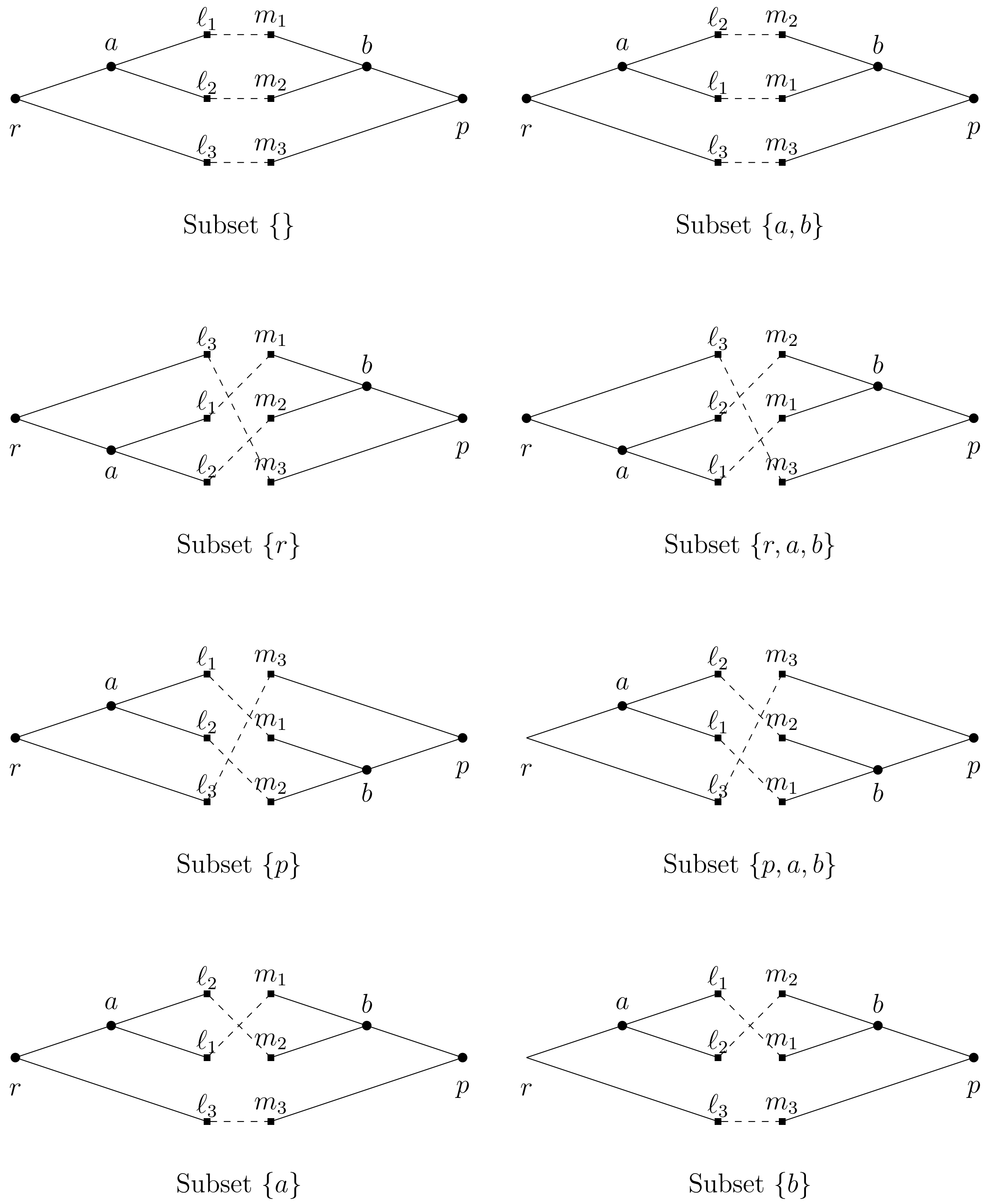


Figure 3.5 The first 4 pairs of layouts from corresponding set $X = \{r, a, b, p\}$.

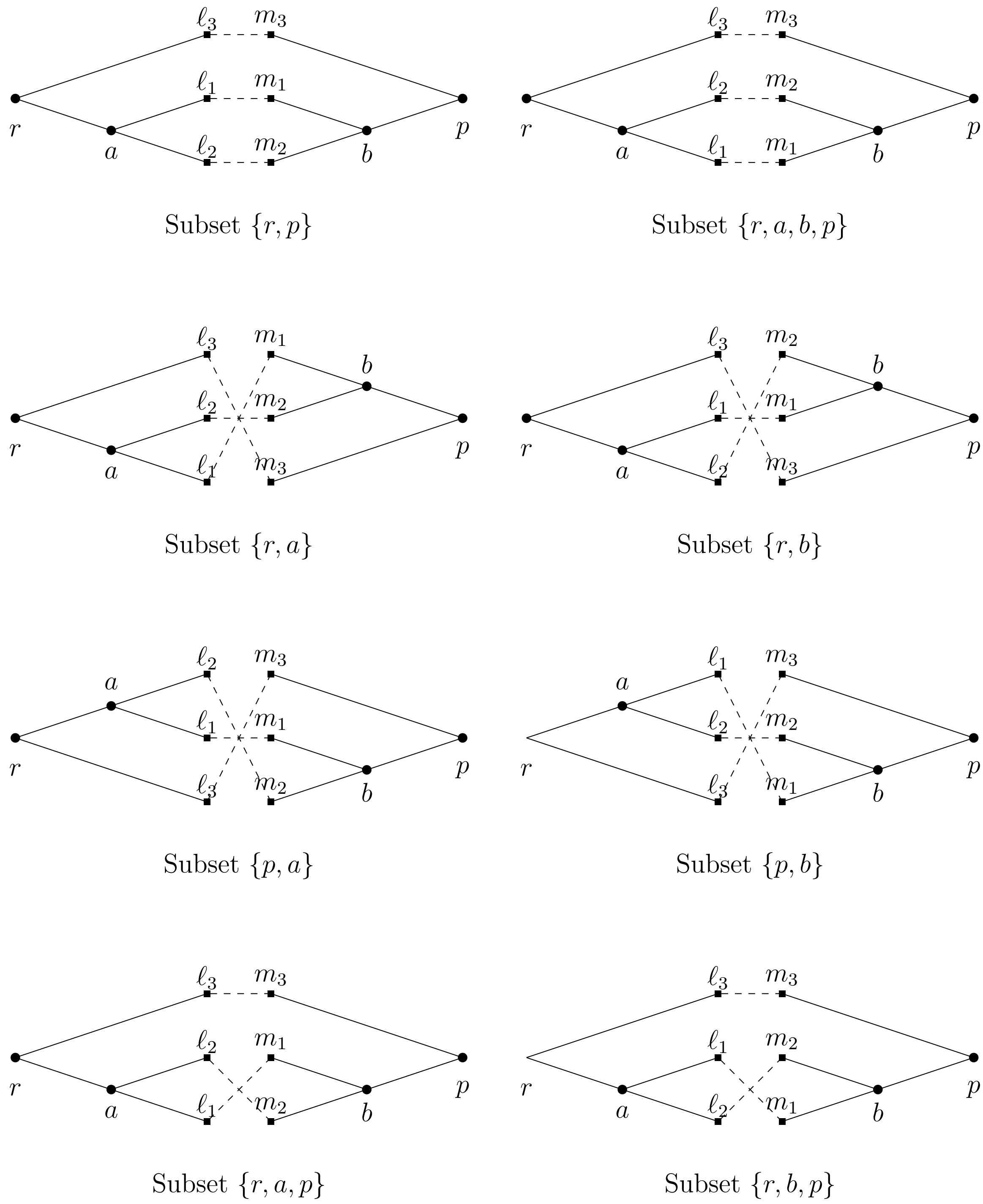


Figure 3.6 The second 4 pairs of layouts from corresponding set $X = \{r, a, b, p\}$.

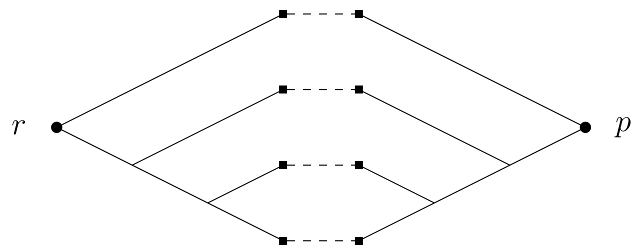


Figure 3.7 A Tanglegram of size 4 with two $\mathcal{C}_4$ graphs as the rooted binary trees.

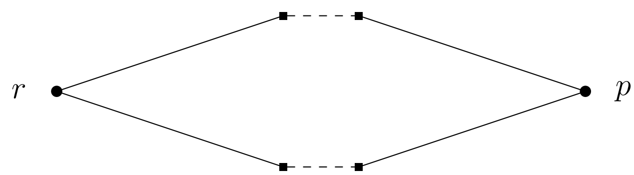


Figure 3.8 A Tanglegram of size 2 with two rooted Complete Binary trees.

two distinguishable leaves: the two leaves forming a cherry are not distinguishable from each other. Now the matching can do two things: the leafs connected to the root on the left- and right-tree can be matched to each other or not. The two possible size 3 Tanglegrams are discussed below.

As both the left and right tree must be a caterpillar tree, they have exactly one leaf that is the child of the root, and two other leaves at distance two from the root. So, we have two cases: the leaves that are children of the two roots are matched to each other, or they are not. Both of these cases result in a unique Tanglegram, as illustrated in Figure 3.9.

Now for size 4 tanglegrams. Note that the 4-leaf complete binary tree has two cherries; the 4-leaf caterpillar has a cherry (both of its leaves at distance 3 from the root), a leaf at distance one and another leaf at distance 2 from the root. If the two sides are both complete binary trees, we have two cases:

- If one of the cherries on the left is matched to a cherry on the right, then the

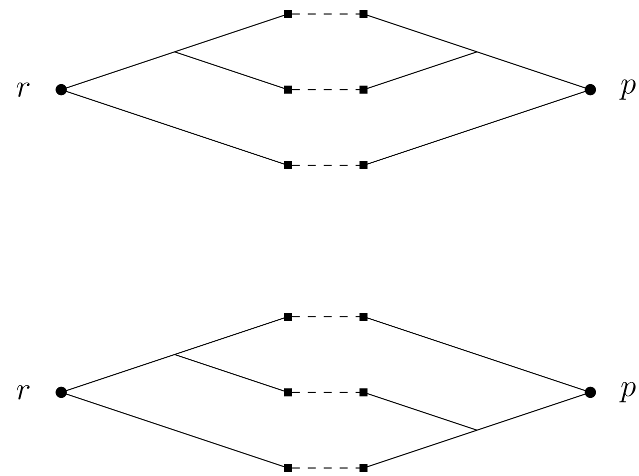


Figure 3.9 The two different Tanglegrams of size 3.

other cherry on the left must match to the other cherry on the right.This defines a unique (unlabeled) Tanglegram.

- If the previous case does not happen, then both leaves in a cherry need to be matched to two leaves that do not form a cherry. This defines a unique (unlabeled) Tanglegram.

If the left tree is a caterpillar and the right tree is a complete binary tree, then we have two cases:

- The cherry in the caterpillar is matched to a cherry in the binary tree: this defines a unique Tanglegram.
- The cherry in the caterpillar is not matched to a cherry in the binary tree: this defines a unique Tanglegram.

If the left tree is a complete binary tree and the right tree is a caterpillar, the two cases are very similar as the above ones.

Now, if both the left and right trees are caterpillars, we have the following cases:

- The cherry on the left is matched to the cherry on the right, in which case the following subcases occur.
  - The leaves at distance 1 from the root match to each other (and consequently so do leaves at distance 2 from the root).
  - The leaves at distance 1 from the root do not match to each other (and consequently a leaf at distance one from the root on one side matches to a leaf at distance two from the root on the other side).
- The cherry on the left is matched to leaves that are not part of the cherry on the other side (and consequently the cherry on the right is matched to leaves that are not part of the cherry on the left: this defines a unique Tanglegram.
- One leaf of the cherry on the left is matched to one leaf of the cherry on the right, but the other leaf of the cherry on the left is not matched to the cherry on the right. We have the following based on where the other leaf of the cherry matches to...
  - both the left and the right cherry matches to the leaf at distance one from the root on the other side (and consequently the leaves at distance two from the root match to each other).
  - both the left and the right cherry matches to the leaf at distance two from the root on the other side (and consequently the leaves at distance one from the root match to each other).
  - the left cherry leaf matches to the leaf at distance one from the root on the right and the right cherry matches to the leaf at distance two from the root on the left.
  - the left cherry leaf matches to the leaf at distance two from the root on the right and the right cherry matches to the leaf at distance one from the

root on the left.

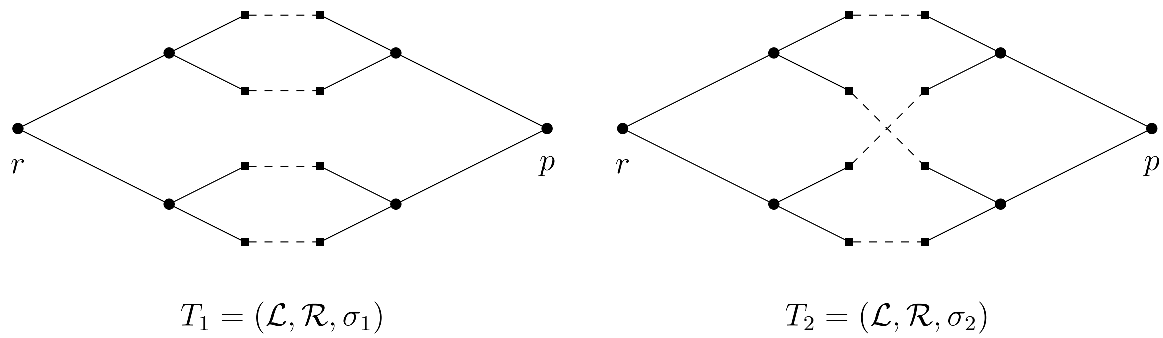


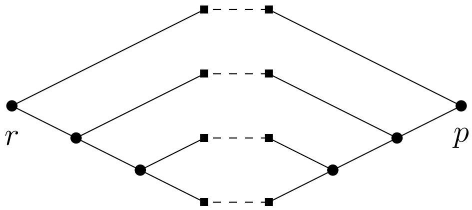
Figure 3.10 The 2 different Tanglegrams of size 4 with Complete Binary Trees as $\mathcal{L}$ and $\mathcal{R}$

Organizing all of our perfect matchings and figures, we conclude that there are13 Tanglegrams of size 4, Figure 3.14 illustrates all 13 layouts.

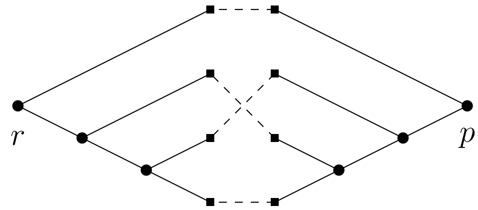
Now rooted binary trees with 5 leaves can be obtained from the following:

- The Complete Binary tree of height 2 by adding 2 children to one of its leaves (as the leaves are indistinguishable, there is only 1 such tree, neither Caterpillar nor Complete Binary).
- The rooted Caterpillar with 4 leaves by adding 2 children to a ...
  - Leaf at distance 1 from the root (neither Caterpillar nor Complete Binary)
  - Leaf at distance 2 (neither Caterpillar nor Complete Binary
  - One of the cherry leaves (Caterpillar)

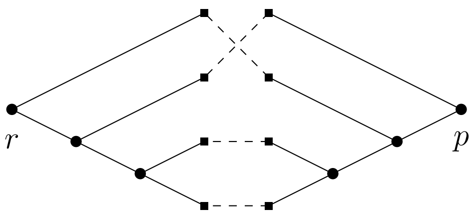
We can continue in this manner to find all the Tanglegrams of size n, but one will find that this number grows quite quickly. In consideration to the enumeration problem for Tanglegrams, [4] obtained an explicit formula for T_n of Tanglegrams with n leaves on each side. The following asymptotic formula holds for the counting sequence



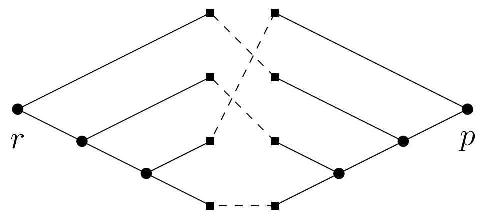
$T_3 = (\mathcal{L}, \mathcal{R}, \sigma_3)$



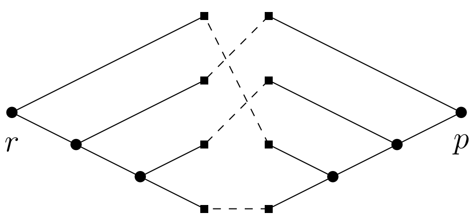
$T_4 = (\mathcal{L}, \mathcal{R}, \sigma_4)$



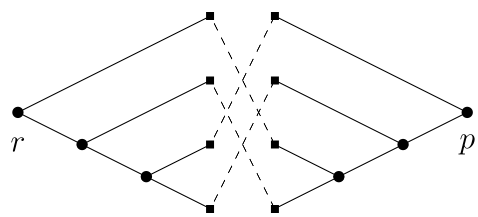
$T_5 = (\mathcal{L}, \mathcal{R}, \sigma_5)$



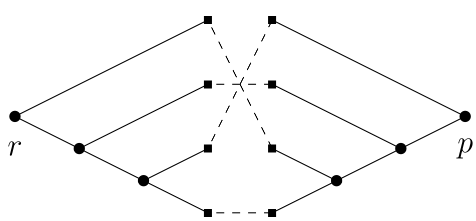
$T_6 = (\mathcal{L}, \mathcal{R}, \sigma_6)$



$T_7 = (\mathcal{L}, \mathcal{R}, \sigma_7)$

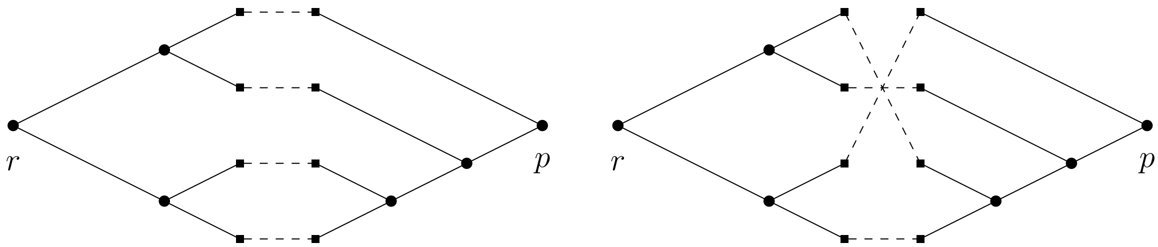


$T_8 = (\mathcal{L}, \mathcal{R}, \sigma_9)$



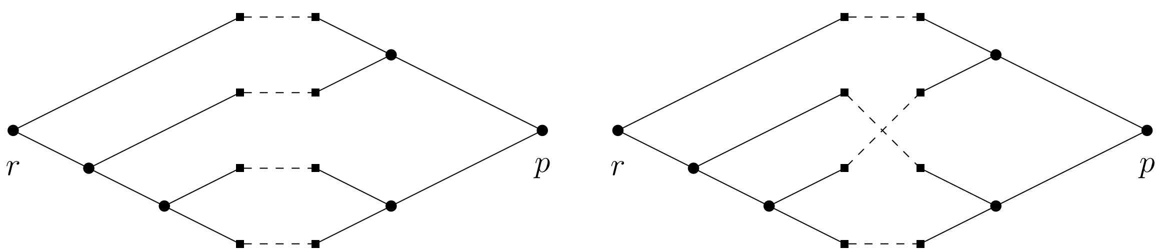
$T_9 = (\mathcal{L}, \mathcal{R}, \sigma_9)$

Figure 3.11 The 7 different Tanglegrams of size 4 with Rooted Caterpillars as $\mathcal{L}$ and $\mathcal{R}$.



$T_{10} = (\mathcal{L}, \mathcal{R}, \sigma_{10})$ $T_{11} = (\mathcal{L}, \mathcal{R}, \sigma_{11})$

Figure 3.12 The 2 different Tanglegrams of size 4 with Complete Binary Tree $\mathcal{L}$ and Rooted Caterpillar $\mathcal{R}$.



$T_{12} = (\mathcal{L}, \mathcal{R}, \sigma_{12})$ $T_{13} = (\mathcal{L}, \mathcal{R}, \sigma_{13})$

Figure 3.13 The 2 different Tanglegrams of size 4 with Rooted Caterpillar $\mathcal{L}$ and Complete Binary Tree $\mathcal{R}$.

$1, 1, 2, 13, 114, 1509, 25595, 535753, 13305590, 382728552, \ldots$

$$T_n \sim n! \cdot \frac{e^{\frac{1}{8}} 4^{n-1}}{\pi n^3},$$

thanks to work done by [2]. We will end with determing the number of size 4 Tanglegrams and encourage the reader to allocate what time they would have spent drawing out all the Tanglegrams of size 5 into something much more productive.

The next chapter generalizes the planarity argument of Kuratowski's Theorem, and applies it to Tanglegrams.

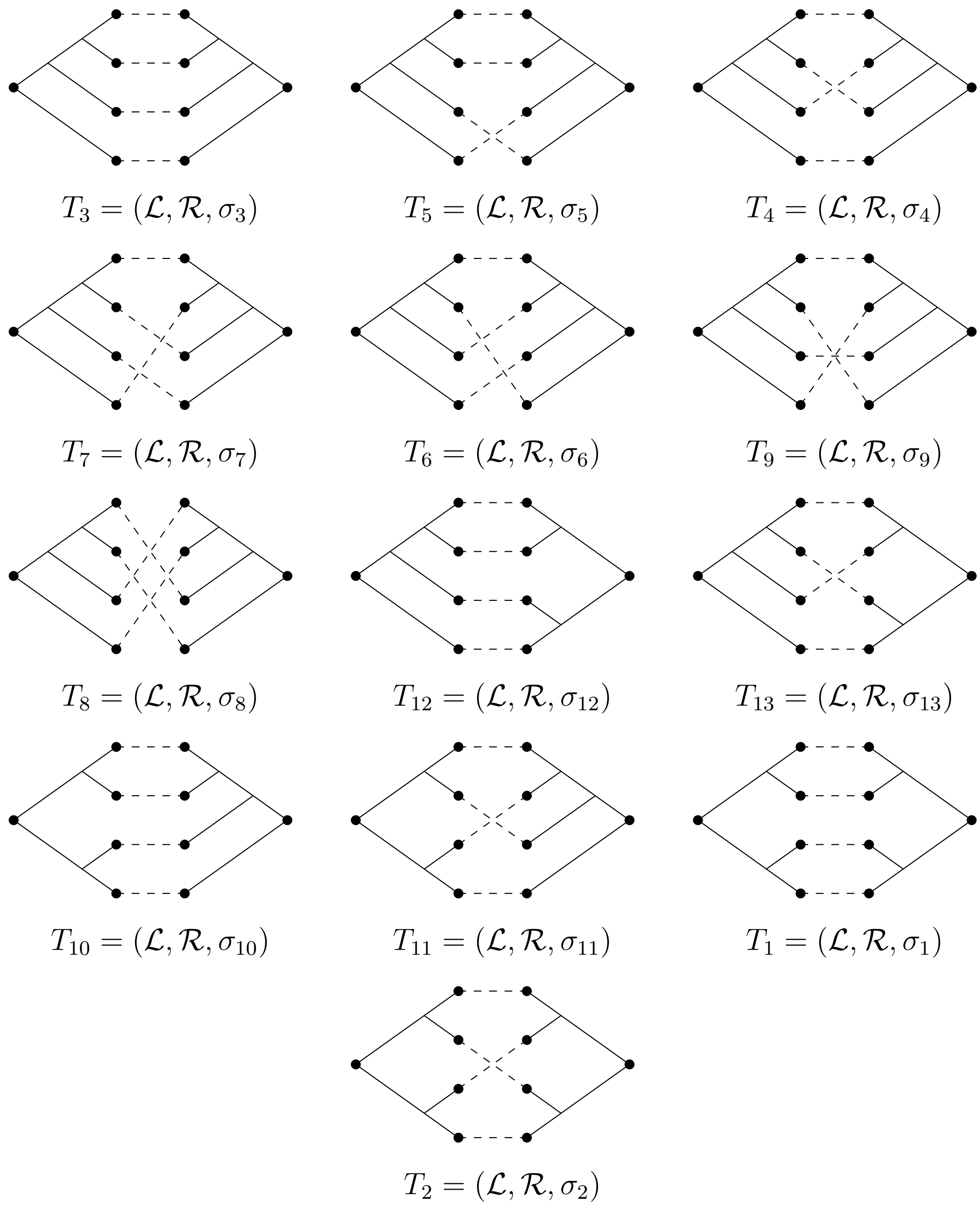


Figure 3.14 The 13 tanglegrams of size 4. T_2 and T_9 cannot be drawn without at least one crossing of the matching edges.

Chapter 4

Tanglegram Crossing Numbers and Planarity

We have discussed how to draw Tanglegrams in the plane (i.e. what a Tanglegram layout is). This definition ensures that in a layout two edges can cross only if they both are matching edges, but, as Figure 3.6 illustrates, in the different layouts of a Tanglegram the amount of crossing that occur can be different.

Definition 29. A *graph drawing*, D, is drawing of a graph where the vertices of the graph are represented by points and edges are represented by simple curves connecting their endpoints, and not going through any other vertices of the graph.

Definition 30. For a drawing D and edges e, f, let $cr_D(e, f)$ define the number of common interior points of e and f. Then, the *crossing number of the drawing D* is

$$cr(D) = \sum_{\{e,f\} \in \binom{E}{2}} cr_D(e, f).$$

The *crossing number of a graph*, denoted cr(G), is the minimal crossing number of all of its drawings.

Definition 31. The *Tangelgram crossing number of a Tanglegram T* is the minimal crossing number over all of its layouts.

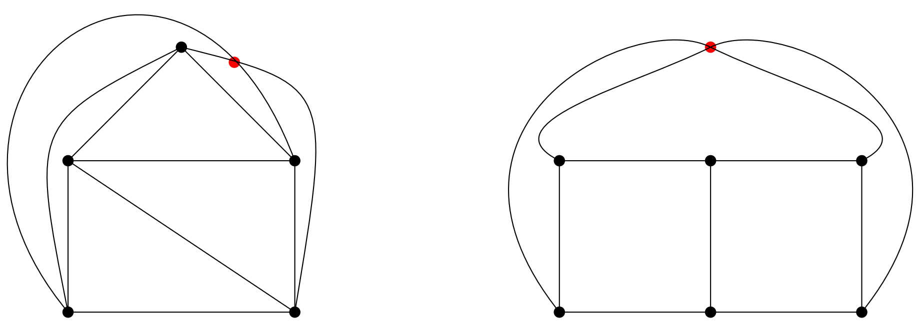
Definition 32. A *planar graph* is a graph that has a planar drawing, i.e. a drawing with crossing number 0. A plane graph is a planar graph together with a planar drawing (i.e. two different planar drawings of the same graph are different plane graphs)

Definition 33. A Tanglegram is *planar* if its Tanglegram crossing number is zero; in other words, if it has a layout without crossing matching edges. Otherwise, the tanglegram is called *nonplanar*.

Let us explore some interesting characteristics of planarity and crossing numbers of graphs.

Proposition 4. *For a connected plane graph with v vertices, e edges and f faces, we have $v - e + f = 2$.*

We can construct familiar drawings of K_5 and $K_{3,3}$ optimally with respect to their crossing number. Although we do not prove that the crossing numbers of K_5 and $K_{3,3}$ are not greater than 1, the graphs in 4.1 are drawn in such a way that K_5 and $K_{3,3}$ have only one crossing.



K_5 with $\mathrm{cr}(K_5) = 1$ $K_{3,3}$ with $\mathrm{cr}(K_{3,3}) = 1$

Figure 4.1 Drawings of K_5 and $K_{3,3}$ with their respective crossing numbers. The red dots mark the point where the edges cross.

As both graphs $K_{3,3}$ and K_5 in Figure 4.1 are drawn with one crossing, their crossing number is at most 1. As they are known to be non-planar, their crossing number equals 1. Kuratowski's theorem states that essentially these two graphs are the obstacles of planarity, as any nonplanar graph must contain a subdivision of one of then. We will explore similar obstacles for the non-planarity of Tanglegrams.

Using Theorem 3, we assert multiple important ideas. One, in particular, is if the graph crossing number of a Tanglegram augmented by an edge between the roots of its binary trees is nonzero, so is Tanglegram crossing number.

Proposition 5 ([4])**.** *Let $T = (\mathcal{L}, \mathcal{R}, \sigma)$ be a Tanglegram and let the roots of $\mathcal{L}$ and $\mathcal{R}$ be r and p, respectively. Let T^* be the underlying graph of T augmented with an edge between r and p; see Figure 4.2. Then, $crt(T) \geq cr(T^*)$.*

Proof. Consider an optimal layout of the Tanglegram T with $\mathrm{crt}(T)$ crossings. We can create a drawing D of T^* by drawing the edge between r and p in the optimal layout of T such that this edge creates no new crossings. Then we have

$$cr(T^*) \leq cr(D) = crt(T)$$

□

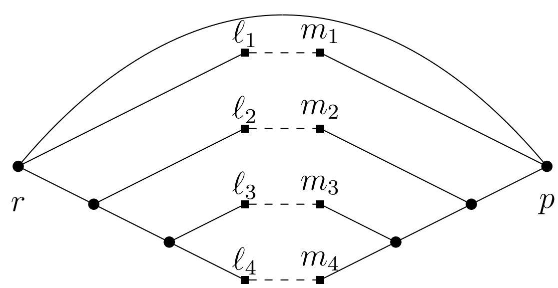


Figure 4.2 Drawing of an augmented Tanglegram T with an augmented edge connecting roots r and p.

Proposition 6 ([4]). *Let $T = (\mathcal{L}, \mathcal{R}, \sigma)$ be a Tanglegram layout and let the roots of $\mathcal{L}$ and $\mathcal{R}$ be r and p, respectively. Let T^* be the underlying graph of T augmented with an edge between r and p; see Figure 4.2. Then, the following are equivalent: (comes from [4])*

1. *crt*$(T) \geq 1$

2. *cr*$(T^*) \geq 1$

3. T^* *contains a subdivision of* $K_{3,3}$

Proof. Note that T^* has maximum degree 3, so it can not contain a subdivision of K_5. This means that Kuratovski's theorem implies that 2 and 3 are equivalent.

$1 \Rightarrow 2$ can be done with standard techniques using topology.

$2 \Rightarrow 1$ follows from Proposition 5. □

By the way we defined it in Definition 33, a Tanglegram is planar if its crossing number is zero. But from the view of ordinary crossing numbers, the size 4 Tanglegrams are simply a subdivision of the four vertex, 3-regular multi-graphs.

Definition 34. The Tanglegram is k-edge planar, if for any $M \subseteq \sigma$ with $|M| < k$, the Tanglegram induced by $\sigma - M$ is not planar, but there is an $M' \subseteq \sigma$ with $|M'| = k$ such that the Tanglegram induced by $\sigma - M'$ is planar.

Definition 35. A *multi-graph* is a graph where one or more of its vertices may be connected to any other vertices by more than one edge. This includes vertices being connected to itself via an edge, called a *loop*.

We first note the following remark, which can be seen since K_4 can be drawn without any crossings of the interior of its edges.

Remark 1. The complete graph on 4 vertices, K_4, is planar.

Any four vertex, 3-regular graph can be obtained by duplicating some edges of an appropriate subgraph of K_4, and the size 4 Tanglegrams are simply a subdivision

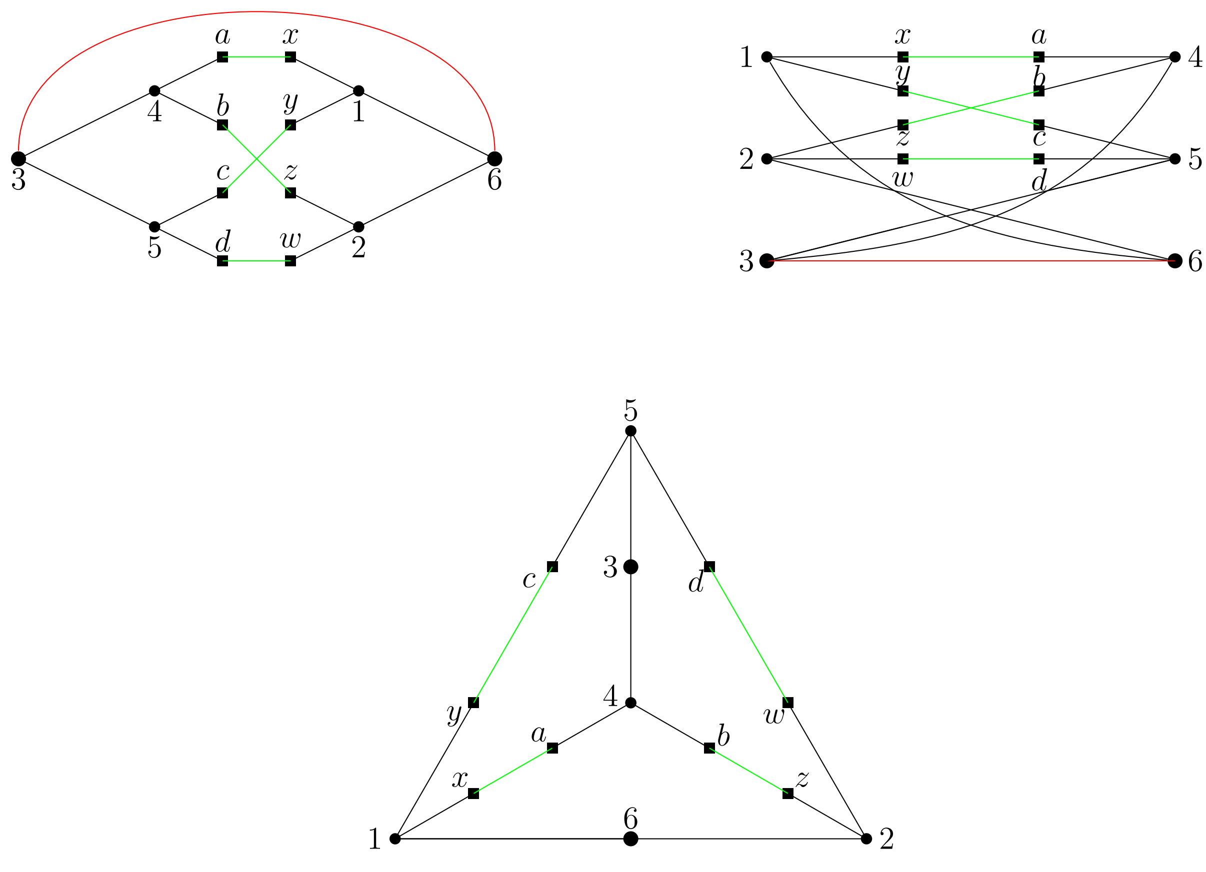


Figure 4.3 The extended Tanglegram T_2 is a subdivided $K_{3,3}$, but without the extra edge it is a subdivided K_4, which is a planar graph.

of the four vertex, 3-regular multi-graphs. If we view size 4 Tanglegrams as graphs, they have crossing number 0 (in other words, as graphs, they are planar). However, not all size 4 Tanglegrams are planar as Tanglegrams. In particular, the Tanglegrams T_2 and T_9 have crossing number 1, and we can use Proposition 5 to show that they are not planar. This is illustrated in Figures 4.3 and 4.4.

For now, we illustrate that planarity can be judged by the notion of how many edges can be removed until the graph becomes planar. In particular, we show that for a Tanglegram of size n, when we remove any one of the matching edges and suppress the two leafs it connected, the crossing number of the original Tanglegram and the

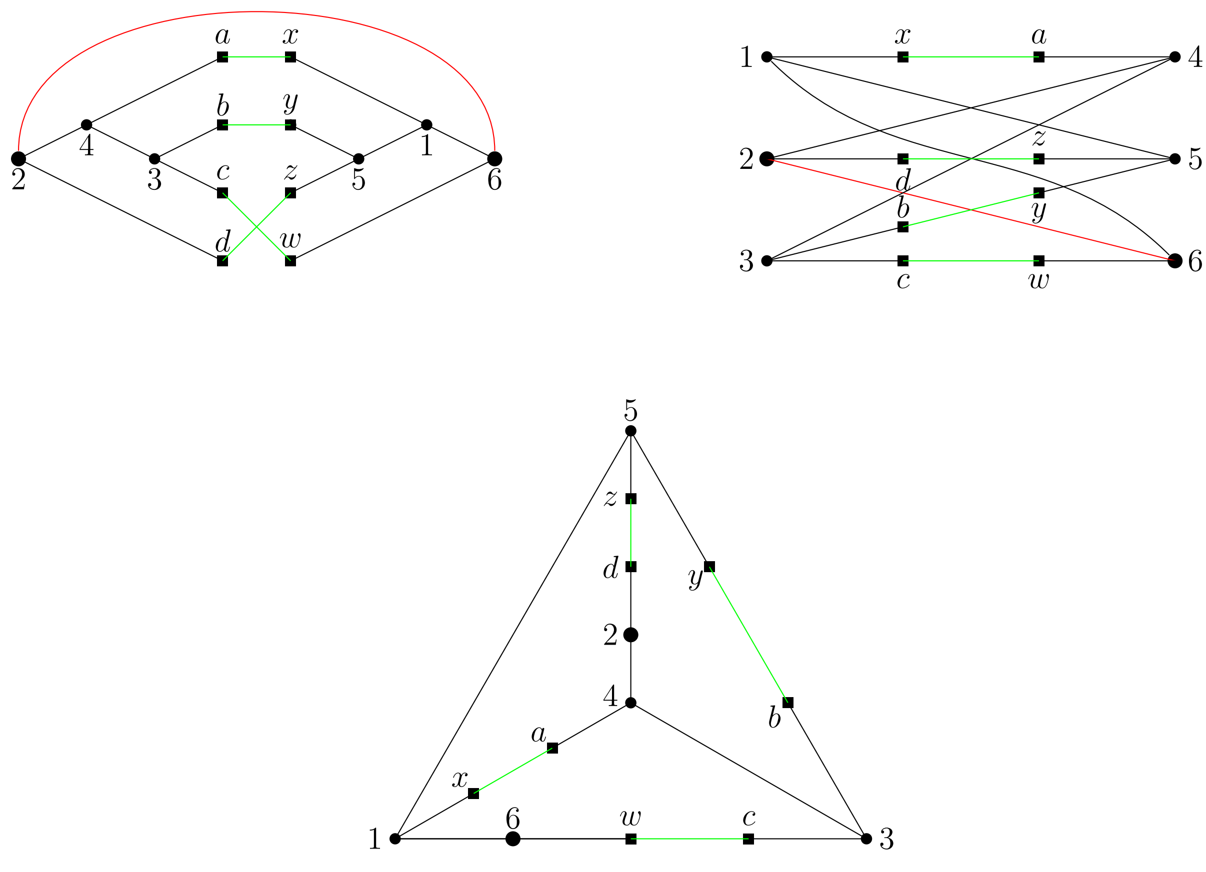


Figure 4.4 The extended Tanglegram T_9 is a subdivided $K_{3,3}$, but without the extra edge it is a subdivided K_4, which is a planar graph.

crossing number of the suppressed Tanglegram are related in a particularly nice way.

Theorem 7 ([1])**.** *Let $T = (\mathcal{L}, \mathcal{R}, \sigma)$ be a Tanglegram of size $n \geq 3$ and let $e \in \sigma$ be a matching edge of T. Then, we have that*

$$crt(T) - crt(T - e) \leq n - 3.$$

Consequently, in any optimal layout, any matching edge crosses at most $n - 3$ edges.

Proof. We will work by induction on n. We will first verify the base case $n = 3$.

We previously found all the Tanglegrams of size 3 in Chapter 3. They were all planar. Similarly, removing a matching edge, the Tanglegram still remains planar.

So,

$$\begin{aligned} \text{crt}(T) - \text{crt}(T-e) &\le n-3 \\ 0-0 &\le 3-3 \\ 0 &\le 0 \end{aligned}$$

Clearly, in any optimal layout, the matching edges cross $0 = 3-3$ other edges. Therefore, the theorem holds for $n = 3$.

Now let $n \ge 4$ and suppose that for every Tanglegram of size $n-1$ that

$$\text{crt}(T) - \text{crt}(T-e) \le n-4.$$

Fix some Tanglegram $T = (\mathcal{L}, \mathcal{R}, \sigma)$ with size n. Let $e \in \sigma$ be arbitrary. Let $e = uv$, where $u \in \mathcal{L}$ and $v \in \mathcal{R}$. We fix an optimal layout T' of $T - e$ to be such that

$$T' = (\mathcal{L}_u, \mathcal{R}_v, \sigma - e)$$

with the fewest the number of crossings.

Now let $w_{\mathcal{L}'}$ be the parent vertex of u and $\mathcal{L}'$ be the subtree rooted at the second child of $w_{\mathcal{L}'}$. Define $w_{\mathcal{R}'}$ in a similar manner. We have, as a result, two planar drawings of $\mathcal{L}$ whose sub-drawings of $\mathcal{L}_u$ agrees with the drawing of $\mathcal{L}_u$ in T':

1. One drawing with u immediately above the leaves of $\mathcal{L}'$.

2. One drawing with u immediately below the leaves of $\mathcal{L}'$.

Observe that the ordering of the leaves of $\mathcal{L}_u$ in each drawing of $\mathcal{L}$ is the same as in T'. Also, by performing a switch operation at $w_{\mathcal{L}'}$, we can obtain one of these drawings of $\mathcal{L}$. This is noted in $\mathcal{R}$ and $\mathcal{R}'$, and Figure 4.5 illustrates two potential positions of u and v in a drawing of T.

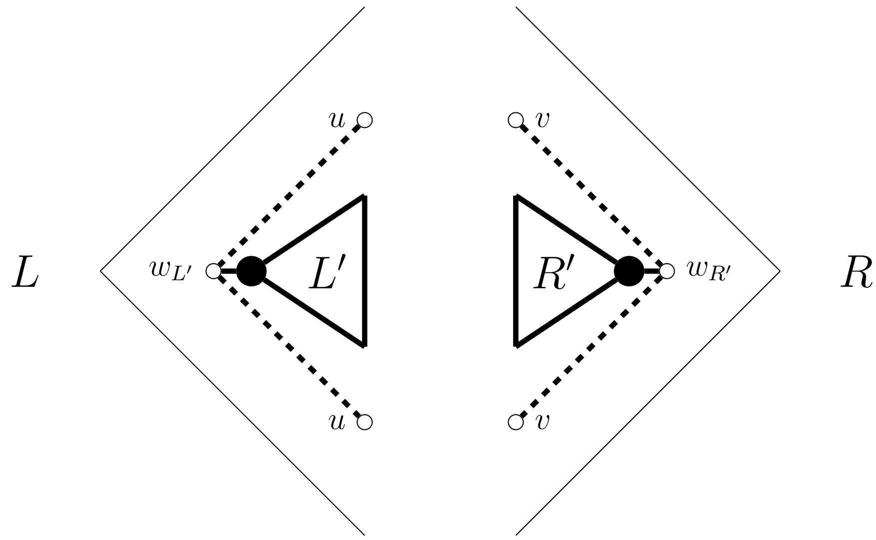


Figure 4.5 An illustration of the potential positions of u and v in the proof of Theorem 7.

All that is left to show is that there is some drawing D of T that uses one of these two drawings of $\mathcal{L}$ and $\mathcal{R}$ in which the matching edge $e \in \sigma$ crosses at most $n-3$ edges. Then, we will have that

$$\mathrm{crt}(T) \leq \mathrm{crt}(D) \leq \mathrm{crt}(T-e) + (n-3).$$

Or, equivalently

$$\mathrm{crt}(T) - \mathrm{crt}(T-e) \leq n-3.$$

We have two cases to deal with:

1. $\mathcal{L}'$ and $\mathcal{R}'$ each have exactly one leaf and they are matched in $\sigma - e$ or
2. There is a leaf in $\mathcal{L}'$ and a leaf in $\mathcal{R}'$ which are not matched with one another.

Case 1: Let e' be the edge matching the single leaves in $\mathcal{L}'$ and $\mathcal{R}'$. By the induction hypothesis e' crosses at most $n-4$ edges in the layout. Let the drawing of T be with u above $\mathcal{L}'$ and v above R' be such that e is parallel to e'. Then e crosses precisely those edges that e' crosses, so e crosses at most $n-4$ edges (See Figure 4.6 for an illustration of this case.)

Case 2: Let $u_{\mathcal{L}'} \in \mathcal{L}'$ be a leaf and $v_{\mathcal{R}'} \in \mathcal{R}'$ be a leaf such that $u_{\mathcal{L}'}$ and $v_{\mathcal{R}'}$ are not matched together. We define two more ideas below:

1. We say a leaf is *matched upward* if the leaf to which it is connected is at least as high as the lowest leaf in the respective tree.

2. We say a leaf is *matched downward* if the leaf to which it is connected is no higher than the highest leaf in the respective tree.

Let e_1 and e_2 be matching edges with endpoints $u_{\mathcal{L}'}$ and $v_{\mathcal{R}'}$, respectively. We have two subcases to consider:

1. Let $u_{\mathcal{L}'}$ and $v_{\mathcal{R}'}$ be both matched upward (respectively, downward). Draw the vertex u below (respectively, above) in $\mathcal{L}'$ and the vertex v below (respectively, below) in $\mathcal{R}'$. Then, e does not cross e_1 or e_2, and so, e crosses at most $n-3$ edges. Thus,

$$\text{crt}(T) - \text{crt}(T-e) \leq n-3.$$

2. Let $u_{\mathcal{L}'}$ be matched to a leaf higher (respectively, lower) than the leaves of $\mathcal{R}'$ and let $v_{\mathcal{R}'}$ be matched to a leaf lower (respectively, higher) than the leaves of $\mathcal{L}'$. Draw the vertex u directly below (respectively, above) the leaves of $\mathcal{L}'$ and v directly above (respectively, below) the leaves of $\mathcal{R}'$. Then, e does not cross e_1 or e_2, and so, e crosses at most $n-3$ edges. Thus,

$$\text{crt}(T) - \text{crt}(T-e) \leq n-3.$$

   Figure 4.6 illustrates both of these subcases.

Now consider an optimal drawing of T with $\text{crt}(T)$ many crossings. Take a matching edge e that crosses x other edges. The removal of e results in a subdrawing D of $T-e$ with $\text{crt}(T)-x$ crossings. Since $\text{crt}(T-e) \leq \text{crt}(D) = \text{crt}(T)-x$, we get that $x \leq \text{crt}(T) - \text{crt}(T-e) \leq n-3$.

Therefore, we have considered all possibilities and the Theorem is proven. □

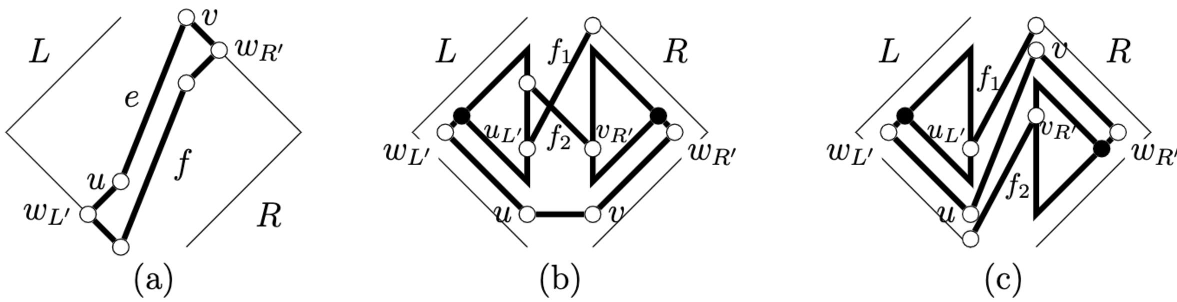


Figure 4.6 An illustration of the possible relations between $\mathcal{L}'$ and $\mathcal{R}'$ in the proof of Theorem 7: (a) $\mathcal{L}'$ and $\mathcal{R}'$ each have exactly one leaf and they are matched in $\sigma - e$. (b) $u_{\mathcal{L}'}$ and $v_{\mathcal{R}'}$ are not matched to each other and are both matched upward. (c) $u_{\mathcal{L}'}$ is matched to a leaf higher than the leaves of $\mathcal{R}'$, and $v_{\mathcal{R}'}$ is matched to a leaf lower than the leaves of $\mathcal{L}'$.

Definition 36. A Tanglegram T is *k-edge planar* if for any $M \subseteq \sigma$ with $|M| < k$, the Tanglegram induced by $\sigma - M$ is not planar, but there is an $M' \subseteq \sigma$ with $|M'| = k$ such that the Tanglegram induced by $\sigma - M'$ is planar.

Definition 37. For each $n \geq 4$, we define the *Caterpillar Tanglegram* $P_n = (L, R, \sigma)$ as follows: L and R are copies of the rooted caterpillar $\mathcal{C}_n$. We label the leaves of L as u_i, where i is the leafs distance from the root. Since there are precisely two leaves at distance $n-1$, we arbitrarily label one of these u_n instead. Similarly, the leaves of R are labeled using v_i. Finally, we construct $\sigma_n = \{u_i v_{n-i} \mid i \in [n-1]\} \cup u_n v_n$.

Theorem 8 ([1])**.** *For each $n \geq 4$, Caterpillar Tanglegram P_n is 1-edge planar and has a crossing number of $n-3$.*

Proof. Observe in Figure 8 that $\text{crt}(P_n) \leq n-3$, and that the removal of the $u_n v_n$ edge results in a planar Tanglegram. As for $n \geq 4$, or, $n-3 \geq 1$, the rest of the statement is proved if we show that $\text{crt}(P_n) \geq n-3$. We prove this by induction on n.

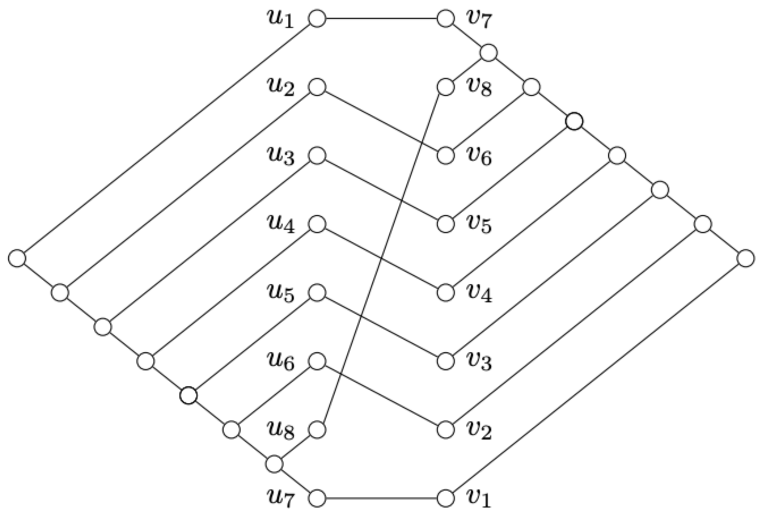


Figure 4.7 The Caterpillar Tangelgram P_8.

We have shown earlier that $\mathrm{crt}(P_4) = 1$ (refer back to Figure 4.4), so the statement is true for $n = 1$. Let $n \geq 4$ and assume that $\mathrm{crt}(P_n) = n - 3$. Consider P_{n+1}. Since P_4 is an induced subtanglegram of P_n, we have that $\mathrm{crt}(P_{n+1}) \geq 1$.

Consider an optimal layout of P_{n+1} with $\mathrm{crt}(P_{n+1}) \geq 1$ many crossings. D must contain a crossing pair of edges, so one of these edges is of the form $u_i v_{n-i}$ for some $i \in [n-1]$. The removal of $u_i v_{n-i}$ from P_{n+1} gives a copy of P_n (and an induced layout D of P_n from our optimal layout of P_{n+1}). Since $u_i v_{n-i}$ crossed at least one edge, we have that $n - 3 \leq \mathrm{crt}(P_n) \leq \mathrm{cr}(D) \leq \mathrm{crt}(P_{n+1}) - 1$, which gives that $n - 2 = (n+1) - 3 \leq \mathrm{crt}(P_{n+1})$. Thus, the result is acheived. $\square$

Chapter 5

Properties of Induced Subtanglegrams

Recall that an induced subgraph $G' = (V', E')$ of a graph $G = (V, E)$ is a graph consisting of vertex subset $V' \subseteq V$ and edge set $E' = E$. An *induced subtree* T' acts similarly on a tree T. We will focus primarily on induced binary subtrees, which is utilized quite often in the study of phylogenetics.

In a rooted plane binary tree B_T with a root r, we say a subset L of the leaves of B_T *induces* another rooted binary tree by taking the smallest subtree containing the leaves in L, designating the vertex r_L of this subtree closest to the old root as the new root, and finally suppressing all vertices of degree 2 other than r_L. See Figure 5.1 for an illustration of this process.

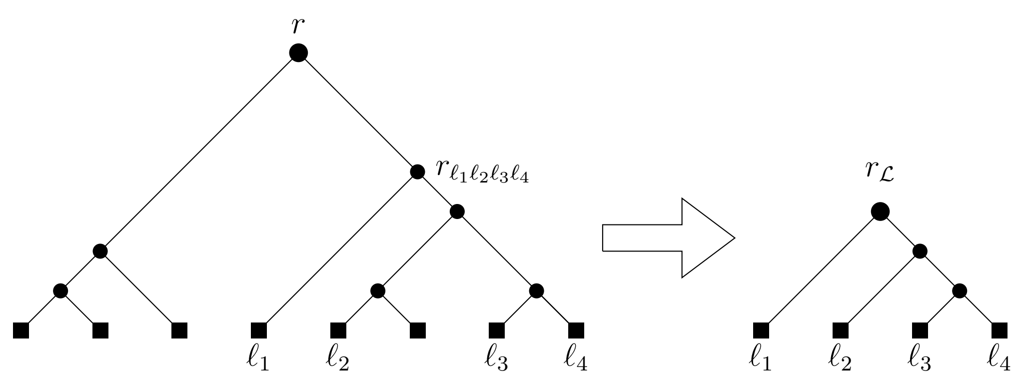


Figure 5.1 A rooted binary tree with root r, four leaves $\mathcal{L} = \{\ell_1, \ell_2, \ell_3, \ell_4\}$ selected, the vertex $r_{\ell_1\ell_2\ell_3\ell_4}$ and the tree induced by the selected leaves.

Consider a layout of a Tanglegram $T = (\mathcal{L}, \mathcal{R}, \sigma)$ with roots r and p of $\mathcal{L}$ and $\mathcal{R}$, respectively. Let $\sigma' \subseteq \sigma$ be a subset of the set of matching edges. The leaf sets of σ'

induce a left and right induced binary plane tree, which, after putting back the edges of σ between the corresponding leaves, define a layout of a Tanglegram T'.

Definition 38. We call T' the *induced subtanglegram* of a Tanglegram T; T' is induced by the matching edge set σ' (see Figure 5.2).

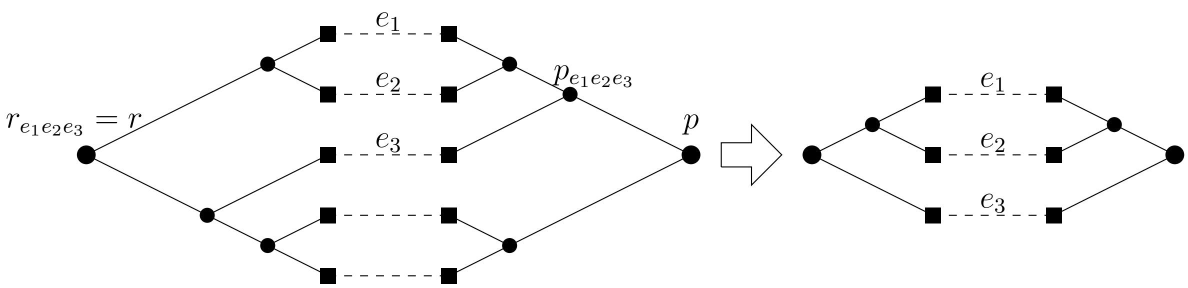


Figure 5.2 A Tanglegram T with matching edges $\sigma' = \{e_1, e_2, e_3\}$ selected, the vertices $r_{e_1e_2e_3}$ and p_{e_1,e_2,e_3}, and the subtanglegram induced by the selected edges.

We claim that given either a planar or non-planar drawing of a Tanglegram T of size $n \geq 4$, that we can always find an induced subtanglegram T' such that T' is planar. In particular, we assert that the largest possible T' that is planar will always be of size $n - \text{crt}(T)$.

Theorem 9. *Let $T = (\mathcal{L}, \mathcal{R}, \sigma)$ be a Tanglegram of size $n \geq 4$. Then, there exists an induced subtanglegram $T' = (\mathcal{L}', \mathcal{R}', \sigma')$ of T such that*

$$|T'| \geq n - crt(T).$$

Consequently, if T is k-edge-planar, then $k \leq crt(T)$.

Proof. Let T be a Tanglegram of size $n \geq 4$. We consider cases:

1. If T is planar, then any induced subtanglegram T' of T is planar, by the definition of how T' is constructed. So, the $\text{crt}(T) = 0$ and further,

$$|T'| = n \geq n - 0 = n.$$

2. If T is nonplanar, then keep deleting matching edges until T has a crossing number of zero and is planar. Since T is nonplanar, there was at least one crossing of its matching edges, so any resulting subtanglegram T' is such that $|T'| \geq n - \text{crt}(T)$.

We have shown that the removal of at most crt(T) many edges from T results in the planar Tanglegram. As by definition, we need to remove at most k edges from a k-edge-planar Tanglegram, $k \leq$crt(T). □

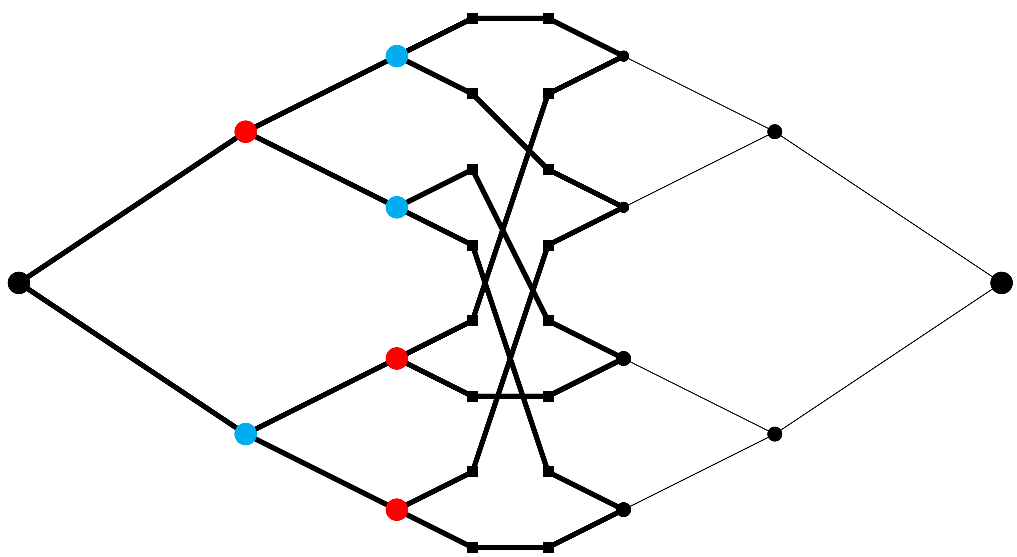


Figure 5.3 Red and blue nodes represent the two partitions of a subdivided $K_{3,3}$ with all vertices on the left side.

Recall that Proposition 7 states that every non-planar Tanglegram, the augmented graph T^* contains a subdivision of $K_{3,3}$. So, we state the next theorem which was proved in [4].

Theorem 10 ([4]). *Every non-planar Tanglegram contains T_2 or T_9 as an induced Subtanglegram.*

Now observe that Theorem 11 is stronger than the statement of Proposition 7, as it provides a subdivided $K_{3,3}$ such that three vertices of the $K_{3,3}$ lie an the left-subtree, and the other three are in the right subtree. As Figure 5.3 shows, if we find a subdivided $K_{3,3}$ in the augmented graph of a non-planar Tanglegram, this subdivided $K_{3,3}$ does not have to correspond to any induced subtanglegram. So it can be seen that Theorem 11 gives more structure than Proposition 7.

Bibliography

[1] R. Anderson, Shuliang Bai, F. Barrera-Cruz, É. Czabarka, G. Da Lozzo, N.L. F. Hobson, Jephian C.-H. Lin, A. Mohr, H.C. Smith, L.A. Székely, H. Whitlatch, Analogies between the crossing number and the tangle crossing number, Electronic Journal of Combinatorics, Volume 25, Issue 4 (2018).

[2] S. C. Billey, M. Konvalinka, and F. A. Matsen. On the enumeration of tanglegrams and tangled chains. JCT A, Volume 146, Pages 239-263 (2017).

[3] Czabarka, Éva, László A. Székely, and Stephan Wagner. “Inducibility in Binary Trees and Crossings in Random Tanglegrams.” arXiv.org, January 26, 2016. https://arxiv.org/abs/1601.07149,SIAM J. Discrete Math Vol. 31, No. 3, pp. 1732-1750 (2017).

[4] Czabarka, Éva, László A. Székely, and Stephan Wagner. “A Tanglegram Kuratowski Theorem.” arXiv.org, August 1, 2017. https://arxiv.org/abs/1708.00309, Journal of Graph Theory , Vol. 90, Issue 2 pp. 111-122 (2019).

[5] Diestel, R. (2010). Graph theory (5th edition). Springer.

[6] Gessel, Ira M. “Counting Tanglegrams with Species.” arXiv.org, July 2, 2021. https://arxiv.org/abs/1509.03867,Journal of Combinatorial Theory, Series A, Volume 184, November 2021, 105498.

[7] Kuratowski, Kazimierz (1930), "Sur le problème des courbes gauches en topologie" (PDF), Fund. Math. (in French), 15: 271–283.

[8] Scornavacca, Celine, Franziska Zickmann, and Daniel H. Huson. "Tanglegrams for Rooted Phylogenetic Trees and Networks." OUP Academic. Oxford University Press, June 14, 2011. https://doi.org/10.1093/bioinformatics/btr210.