2024

Problems in Graph Theory With Applications to Topology and Modeling RNA

Rayan K. Ibrahim
Virginia Commonwealth University

DISSERTATION

PROBLEMS IN GRAPH THEORY WITH APPLICATIONS TO TOPOLOGY AND MODELING RNA

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy at Virginia Commonwealth University.

by

RAYAN IBRAHIM

Bachelor of Science in Mathematics at the College of Staten Island

Master of Science in Mathematics at Virginia Commonwealth University

Directors:

Craig E. Larson,

Professor, Department of Mathematics and Applied Mathematics

Allison H. Moore,

Associate Professor, Department of Mathematics and Applied Mathematics

Virginia Commonwealth University
Richmond, Virginia
July, 2024

Abstract

In this thesis, we explore four projects. In the first project, we explore r-neighbor bootstrap percolation on a graph G. We establish upper bounds for the number of vertices required to percolate in the case that $r = 2$ for particular classes of graphs. In the second project, we study the structure of graphs with independence number two. We prove a lower bound on the number of edges of such graphs, related to an upper bound on the number of edges in a triangle-saturated graph, and give a sufficient forbidden induced subgraph condition for independence number two graphs. In the third project, we extend an existing method and provide a robust framework for studying and analyzing the structure of RNA molecules via chord diagrams and graph theory. In the fourth project, we prove that existing graph theoretic results on spanning tree enumeration in symmetric graphs may be applied to calculate determinants of simple theta curves.

Acknowledgements

I'd like to thank my mother Karima, father Khaled, and brothers Jad and Rami. Without their support, I would not be where and who I am today. I'd like to thank my close friends Binghui Jiang and Emily Horton for emotional support. I'd like to thank Cheryl Craddock and Cynthia Palumbo of the Verrazzano School at CUNY CSI for always lending an ear, guiding me, and giving me the confidence to pursue my dreams.

I would like to thank my advisors Craig Larson and Allison Moore, whose lessons and support have been invaluable to me. Craig's enthusiasm for interesting problems and knowledge of math culture are infectious and unmatched. Allison's proficiency, work ethic, and eagerness to work with and help students grow is awe-inspiring. My idea of success is becoming half as proficient as Allison and half as knowledgeable as Craig. I am grateful to have learned many things from Craig and Allison, both mathematical and non mathematical. Without their patience and support, this endeavor would not have been possible.

I would like to thank Neal Bushaw for agreeing to be on my committee and for exposing me to the world and culture of extremal combinatorics in one of my most memorable and favorite courses at VCU. I would also like to thank Danielle O'Donnol for graciously agreeing to be on my committee. I'd like to thank Dan Cranston for humoring me during my first year of graduate study whenever I had a graph theory problem I could not solve.

I'd like to acknowledge some individuals from my years in undergrad. I'd like to thank Kevin O'Bryant, who was one of the first to cultivate my interest in the beauty of mathematics and combinatorics. I'd like to thank Andrew C. Poje, whose Calculus I course in my first year of college inspired me to major in mathematics. Along with Andrew, I'd like to thank Alex Fabregat Tomás, who I met while working with Andrew on a research project and who showed me the beauty of TikZ. I'd like to thank Erika L.C. King, who served as my mentor during my REU experience at Hobart and William Smith Colleges. Finally I'd like to thank Mikael Vejdemo-Johansson.

I would like to thank my teachers from New Dorp High School in Staten Island,

New York, including Dr. Spindler, Mr. Dellatacoma, Mr. Vroom, Mr. LaMarca, Mr. Grusky, Mr. Hubbs, Ms. Brown, Ms. Lu, and surely many more.

I would like to thank my collaborators who have become close friends, Kevin McCall and Hudson LaFayette, for many stimulating discussions. Thank you for always lending an ear and for the myriad of mathematical email correspondence.

I would like to thank Andy Bristow and Aimee Ellington for their support in my teaching endeavors throughout graduate school. I'd like to thank all my fellow graduate students who I've met over the years, who are too many to list here. I'd like to thank all my students, who have made teaching a fun and exciting experience in my time at VCU. I'd like to thank Ellie Roberts and Iris Hammond for making the Department of Mathematics and Applied Mathematics a place of belonging. I'd like to thank Angela Reynolds for her work as the graduate program director.

I would like to acknowledge the support of the American Mathematical Society, allowing me to travel and attend MRC: Trees in Many Contexts and the Joint Math Meetings in 2022. I'd like to acknowledge the support of the National Science Foundation (DMS-2204148) and The Thomas F. and Kate Miller Jeffress Memorial Trust, Bank of America, Trustee. I would like to acknowledge individuals who wrote fantastic software which aided significantly in this work: McKay [80], McKay and Piperno [81], Miller [83], and Brinkmann, Goedgebeur, and Schlage-Puchta [20]. Section 2.6 contains joint work with Kevin McCall and Hudson LaFayette, and more broadly Chapter 2 would not have been possible without the innumerable correspondences with them. Chapter 5 contains joint work with Elpers and Moore that has been submitted for publication. Many thanks to Ken Baker, Dorothy Buck and Danielle O'Donnol for providing the LaTeX for Table 5.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This thesis investigates four independent projects involving combinatorics, graph theory, and applications of graph theory to the study of RNA and to topology.

In Chapter 2 we investigate r-neighbor bootstrap percolation. The r-neighbor bootstrap percolation process concerns the spread of infection iteratively throughout a graph. In particular, an initial infected set of vertices A_0 is chosen, and a vertex may become infected in an iteration if it has at least r neighbors which are already infected. Once infected, a vertex remains infected. It is worth noting that the 2-neighbor bootstrap problem has also appeared in popular math culture [15, Problems 34 & 35] [4].

Let $m(G, r)$, the r-percolation number, be the minimum number of vertices that can be chosen which infect the entire graph. If G is a graph, then we denote the complementary prism of G as $G\overline{G}$ (see Definition 2.18.) Duarte et al. [46] proved that when G and $\overline{G}$ are connected, $m(G\overline{G}, 2) \leq 5$. Our first result is an improvement on this bound.

Theorem 2.25. If G is a graph such that G and $\overline{G}$ are connected then $m(G\overline{G}, 2) \leq 3$.

We provide constructions showing for fixed diameter $d \geq 3$ and $r \geq 2$, the r-percolation number is unbounded. Let $\mathcal{G}$ be the class of graphs which are 2-connected and have diameter two. It is not yet known if there is a constant k such that for all $G \in \mathcal{G}$, we have $m(G, 2) \leq k$. We show graphs in $\mathcal{G}$ have at most two vertex-disjoint 2-forbidden subgraphs (see Definition 2.3), and explicitly describe the structure of those graphs with exactly two vertex-disjoint 2-forbidden subgraphs. Using this structure and Theorem 2.25, we prove Theorem 2.26.

Theorem 2.26. If G is a 2-connected graph with diameter two and contains two vertex-disjoint 2-forbidden subgraphs, then $m(G, 2) \leq 3$.

We say a graph is H-free if it does not contain H as an induced subgraph. We prove a bound on $m(G, 2)$ for graphs which are C_6-free or C_7-free which is an improvement over a bound for C_6-free graphs given by Cappelle et al. [29].

Theorem 2.36. Let G be 2-connected and diameter two. If G is C_6-free with $|G| \geq 6$ or C_7-free with $|G| \geq 7$ then $m(G, 2) \leq 3$.

In joint work with LaFayette and McCall, we proved an upper bound of two for the class of 2-connected, diameter two, and C_5-free graphs (Theorem 2.33) [67]. In this thesis, we use Theorem 2.33 and Theorem 2.36 to establish a base case for an inductive argument and prove Corollary 2.37.

Corollary 2.37. If G is 2-connected, has diameter two, and is C_k-free for any $5 \leq k \leq n$ then $m(G, 2) \leq \lceil (k-3)/2 \rceil + 1$.

In Chapter 3 we investigate properties of graphs with independence number two. An independent set is a set of vertices where no two vertices in the set are adjacent. The independence number of a graph G, denoted $\alpha(G)$, is the maximum cardinality of an independent set. Let $\boldsymbol{\alpha}_2$ be the class of graphs with independence number two. The diameter of a graph G is the maximum distance between any two vertices in G. In using `CONJECTURING` [72, 73], we obtained Proposition 3.13 as a conjecture and subsequently proved it.

Proposition 3.13. If G is claw-free, bull-free, and has diameter two, then $\alpha(G) = 2$.

This recovers a result implied by Maza, Jing, and Masjoody [78, Theorem 1.1]. Similarly, we obtain another conjecture resulting in Theorem 3.19, a lower bound on the number of edges in graphs in $\boldsymbol{\alpha}_2$. Interestingly, this bound relates to one of Barefoot et al. [10] for triangle-saturated graphs. Let $d(v)$ be the degree of a vertex v.

Theorem 3.19. If $G \in \boldsymbol{\alpha}_2$ with order n and G has diameter two, then the size of G is at least $\max_{v \in V(G)} d(v)(n - d(v)) - 1$.

In Chapter 4, we turn our attention to the use of graph theoretical methods in the modeling of secondary structures in RNA. RNA is a nucleic acid that is known to fold into complex secondary structures [49]; a type of structure of particular interest in this thesis is that of a pseudoknot. Chord diagrams are combinatorial objects that may be used to record hydrogen bonding involved in secondary structures. When RNA structures contain pseudoknots, the corresponding chord diagrams contain crossings (see Section 4.2 for definitions).

In Section 4.2.2 we investigate a method for quantifying the complexity of pseudoknotting in RNA that is currently utilized in the bpRNA-1m database due to Danaee et al. [44]. We first interpret this method with a robust graph theoretical framework, and make several observations about its application. We then generalize the method in Section 4.3 to quantify pseudoknotting in a manner that takes into account both nestings, crossings, and a distance-based threshold τ. A formal definition of pseudoknotting is given in Definition 4.26. This definition defines a pseudoknot in terms of a weighted vertex cover of a certain intersection graph constructed from a partition of the chord diagram representing the nucleotide sequence of the RNA molecule.

To implement our graph-theoretic definition of pseudoknots, we define an algorithm that constructs such a partition. Given a chord diagram D and a parameter τ, Algorithm 1 produces a partition of D. To quantify secondary structures other than pseudoknots, we also implement a routine called `classifyBases` as part of a methodology described in Section 4.4 and Figure 38. Applying Definition 4.26, Algorithm 1, and the methodology of Section 4.4, we perform an initial analysis of pseudoknots and other secondary structures present in the RNA sequences contained in the bpRNA-1m(90) database. We report on quantities of pseudoknots, observe the structures with the most pseudoknots, and count the number of pseudoknot types observed under new methods and compare with previous methods.

Finally, Chapter 5 contains joint work on spatial theta curves with Elpers and Moore [50]. The *θ-graph* is the unique multigraph consisting of two vertices and three parallel edges $\{a, b, c\}$ between them. A *theta curve* ϑ is a spatial embedding of the θ-graph in the three-sphere, taken up to ambient isotopy. Spatial graphs are generalizations of knots and links, with theta curves being a special case. Every theta curve contains three *constituent knots* K_{ij}, formed by taking pairs of edges $i, j \in \{a, b, c\}$. A theta curve is said to be *simple* if at least one of its constituent knots is unknotted [114].

The determinant of a knot K, denoted $\det(K)$, is the order of the homology of the branched double cover of K. We analogously define the determinant $\det(\vartheta)$ of a theta curve ϑ as the integer-valued invariant arising from the first homology of its Klein cover. When ϑ is simple, we prove that $\det(\vartheta)$ is the product of the determinants of the constituent knots.

Theorem 5.13. Let ϑ be a simple theta curve with constituent knots K_{ab}, K_{ac}, K_{bc}. Then

$$\det(\vartheta) = \det(K_{ab}) \cdot \det(K_{ac}) \cdot \det(K_{bc}).$$

Let ϑ be a simple theta curve. Then ϑ corresponds to a *strongly invertible knot*, which is a knot together with an orientation-preserving involution h on the three-sphere such that $h(K) = K$ and $\mathrm{Fix}(h)$ is a circle intersecting K in two points. In other words, a strongly invertible knot is one which exhibits a particular $\mathbb{Z}_2$ symmetry. To prove Theorem 5.13, we realize a graph that admits an involutive symmetry as the Tait graph G of a strongly invertible knot corresponding to the theta curve. The Goeritz matrix plays the role of the graph Laplacian, the matrix determinant of which calculates the determinant of the corresponding strongly invertible knot, or the tree weight of G. Using the Matrix Tree Theorem, spanning trees of graphs admitting involutive symmetry can be enumerated via a product formula involving two smaller, modified subgraphs [39, 123]. One can explicitly identify the two factors in the spanning tree enumeration formula with the determinants of the constituent knots of the theta curve, as realized by their Tait graphs.

1.1 Graph Theory Background and Notation

A *graph* $G = (V, E)$ consists of a vertex set $V = V(G)$ and an edge set $E = E(G)$. If $V(G) = \{v_1, \ldots, v_n\}$ then $E(G)$ is a set of unordered pairs of distinct elements in $V(G)$. In particular, all graphs are simple and undirected throughout this thesis unless otherwise stated. We say two vertices $x, y \in V(G)$ are adjacent if $(x, y) \in E(G)$, and write xy to represent the edge as shorthand. The *order* of G is $|V(G)|$ and the *size* of G is $|E(G)|$. The *complement* of G, denoted $\overline{G}$, is the graph with vertex set $V(\overline{G}) = V(G)$ and edges $E(\overline{G})$ such that $xy \in E(\overline{G})$ if and only if $xy \notin E(G)$. A graph H is a *subgraph* of G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. We say H is an *induced subgraph* if $E(H) = \{xy : x, y \in V(H) \text{ and } xy \in E(G)\}$. For a subset of vertices $A \subseteq V(G)$ we let $G[A]$ denote the induced subgraph with vertex set A. We write $G - H$ or $G - A$ to represent the induced subgraph $G[V(G) - V(H)]$ or $G[V(H) - A]$. A subgraph H of G is a *spanning subgraph* if $V(H) = V(G)$. The *open neighborhood* of a vertex $x \in V(G)$ in G, denoted $N_G(x)$, is the set of vertices in G which are adjacent to x. The vertices in $N_G(x)$ are the *neighbors* of x. The *degree* of x in G, denoted $d_G(x)$, is $|N_G(x)|$. The *closed neighborhood* $N[x] = N(x) \cup \{x\}$. The maximum and minimum degrees over $V(G)$ are denoted $\Delta(G)$ and $\delta(G)$ respectively. If every vertex in G has degree k for some fixed integer k, then we say G is *k-regular*, or *regular* when k is not specified. If H is a subgraph of G or A is a subset of vertices in G, then $N_H(x) = N_G(x) \cap V(H)$ and $N_A(x) = N_G(x) \cap A$. For two graphs G and H the *disjoint union* is the graph with vertex set $V(G) \cup V(H)$ and edge set $E(G) \cup E(H)$, and is written as $G \cup H$.

A graph is *connected* if for any pair of vertices there is a path between them. A graph is *disconnected* if there are at least two vertices with no path between them. A disconnected graph comprises a set of maximal (with respect to order) connected subgraphs called *components*. The *distance* between any two vertices $x, y \in V(G)$ is the length of a shortest path between x and y, denoted $d(x, y)$. The *diameter* of a graph is the maximum distance over all pairs of vertices. For example, in a graph with diameter two, every pair of vertices is either adjacent or share a common neighbor. By convention the diameter of a disconnected graph is ∞. For a vertex v and set of vertices A, we may informally say "v can travel to A in at most k steps" to mean that there is a vertex $u \in A$ such that $d(v, u) \leq k$. A graph G is *k-connected* if $|V(G)| \geq k$ and $G - A$ is connected for any $A \subseteq V(G)$ with $|A| \leq k - 1$. A *cut-vertex* is a vertex whose removal disconnects the graph. A *block* in a graph G is a maximal connected subgraph of G with no cut-vertex. Note that if a block is K_1 then it must be an isolated vertex, and if G has no cut-vertex then G itself is a block. If $A, B \subseteq V(G)$ we say A is *complete* to B if each vertex in A is adjacent to every vertex in B, and A is *anticomplete* to B if each vertex of A has no neighbor in B. We say A *dominates*

B if each vertex of B has at least one neighbor in A.

Throughout this dissertation, we say a graph G is H-free if G does not contain H as an induced subgraph. Note that in other contexts, the property H-free may refer to a graph which does not contain H as a (not necessarily induced) subgraph. In some cases, e.g. triangle-free, the two notions are equivalent. A set of edges $M \subseteq E(G)$ is a *matching* if no two edges $e, e' \in M$ share an endpoint. A *perfect matching* is a matching M such that every vertex in $V(G)$ is the endpoint of an edge in M.

The graph on n vertices with all possible edges is the *complete graph* K_n. A *clique* is a subgraph which is isomorphic to a complete graph. The *clique number* $\omega(G)$ is the maximum size of a clique in G. A *tree* is a connected graph T of order n with $n-1$ edges, or equivalently, a connected, acyclic graph. A *forest* is the disjoint union of a set of trees. The cycle on n vertices will be denoted C_n, and the path on n vertices will be denoted P_n.

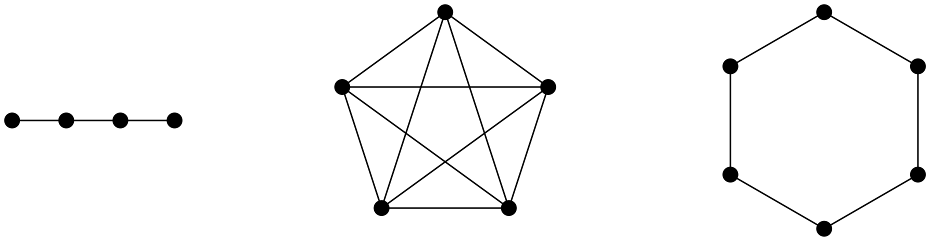


Fig. 1. From left to right: the graphs P_4, K_5, and C_6.

CHAPTER 2

2-NEIGHBOR BOOTSTRAP PERCOLATION

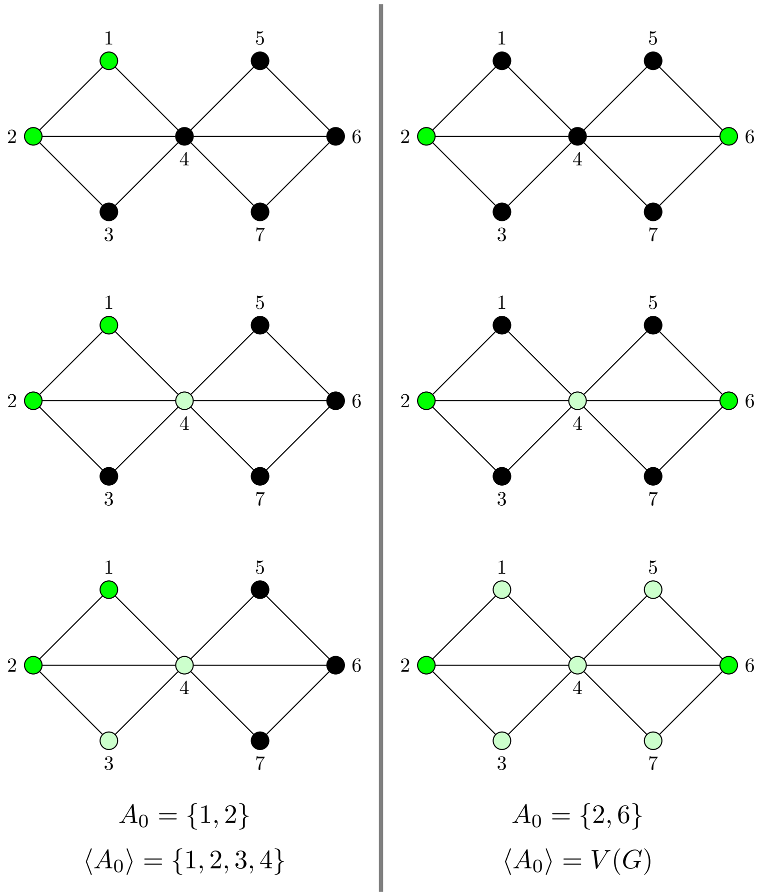
2.1 Overview

The process of r-neighbor bootstrap percolation is a deterministic cellular automaton defined on a graph, G. The process begins with an initial set of infected vertices $A_0 \subseteq V(G)$. In each subsequent round, an uninfected (sometimes called susceptible) vertex v becomes infected if v is adjacent to at least r previously infected vertices. Once infected, vertices remain infected forever. We use A_t to denote the set of all infected vertices as of round t. Symbolically,

$$A_t = A_{t-1} \cup \{v \in V(G) : |N(v) \cap A_{t-1}| \geq r\}.$$

The parameter r is called the *percolation threshold*. If G is a finite graph, then after a finite number of rounds, either all vertices of G become infected or the infection stops at some proper subset of $V(G)$. The set of infected vertices after the percolation process finishes is called the *closure* of A_0, denoted $\langle A_0 \rangle$. If $\langle A_0 \rangle = V(G)$, then we say that A_0 *r-percolates*, or *percolates* when r is clear from context.

Bootstrap percolation on graphs was introduced by Chalupa, Leath, and Reich [33] as a simple model to study magnetic materials. One model that has received much attention is when the vertices of A_0 are selected randomly; each vertex is selected independently and every vertex of G has probability p of being initially selected. After the initial step, the infection proceeds deterministically. This model has been studied extensively, for example in [5, 6, 7, 8, 9, 64]. The minimum size of a percolating set in a graph G with percolation threshold r is denoted $m(G, r)$. Observe that if $|V(G)|$ is at most r, then $m(G, r) = |V(G)|$. The quantity $m(G, r)$ has been studied extensively under particular degree conditions [52, 59, 121, 41, 95]. We may refer to $m(G, r)$ as the *r-percolation number*, or *percolation number* when r is clear from context.



$A_0 = \{1, 2\}$

$\langle A_0 \rangle = \{1, 2, 3, 4\}$

$A_0 = \{2, 6\}$

$\langle A_0 \rangle = V(G)$

Fig. 2. Two examples of the 2-neighbor bootstrap percolation process.

It is interesting to explore the structure of graphs for which $m(G, r) = r$, which are the extremal graphs in our context. We refer to graphs with $m(G, r) = r$ as *r-bootstrap-good*, or r-BG, and graphs for which $m(G, r) > r$ as *r-bootstrap-bad*, or r-BB. It may be enlightening to understand conditions which force $m(G, r) > r$; while one can often find an upper bound for $m(G, 2)$ via constructive methods or algorithmically, such as in [29], it is somewhat difficult to establish lower bounds for particular classes of graphs.

Here we mostly fix our attention to the problem of 2-neighbor bootstrap percolation and in particular the quantity $m(G, 2)$. The quantity $m(G, 2)$ is also known as the *P_3-hull number* of G [29]. A graph *convexity* on $V = V(G)$ is a collection $\mathcal{C}$ of subsets of V such that $\emptyset, V \in \mathcal{C}$ and $\mathcal{C}$ is closed under intersections. For example, for the cycle C_5 with $V(C_5) = \{1,2,3,4,5\}$ and $E(C_5) = \{(1,2),(2,3),(3,4),(4,5),(1,5)\}$ the collection $\mathcal{C} = \{\emptyset, \{1,2\}, \{3,4\}, \{1,2,3,4,5\}\}$ is a convexity on C_5. The sets in $\mathcal{C}$ are called *convex sets*. For a set $S \subseteq V$, the *convex hull* of S with respect to $\mathcal{C}$, denoted $H_{\mathcal{C}}(S)$, is the smallest set in $\mathcal{C}$ containing S. Note convex hulls are unique, as $\mathcal{C}$ is closed under intersections, i.e. $H_{\mathcal{C}}(S)$ is the intersection of all convex sets containing S. Convexities can be defined by a set of paths $\mathcal{P}$ in G such that S is convex if and only if for every pair of vertices $x, y \in S$, the vertices of every path in $\mathcal{P}$ containing x and y as endpoints are also in S. Here we focus on P_3 convexity, where S is P_3-convex if no vertex in $V(G) \setminus S$ has at least two neighbors in S. In other words, S is P_3-convex if and only if for every pair of vertices $x, y \in S$, we have $N(x) \cap N(y) \subset S$. We say S is a P_3-*hull set* if the convex hull with respect to P_3 convexity of S is $V(G)$. The P_3-*hull number* $h_{P_3}(G)$ is the size of a minimum P_3-hull set. The P_3-hull number has been studied extensively in recent years [28, 67, 25, 18, 58, 40, 79, 45]. Let $A_0 \subseteq V(G)$. The translation from 2-neighbor bootstrap percolation to P_3 convexity is as follows. The closure $\langle A_0 \rangle$ is the P_3-convex hull of A_0. The set of all P_3-convex sets in G is the set of all possible closures. If A_0 percolates, then A_0 is a P_3-hull set. The size of a minimum percolating set is the size of a minimum P_3-hull set, i.e. $m(G,2) = h_{P_3}(G)$.

For the earlier example, the P_3-convexity $\mathcal{C}$ on C_5 is

$$
\begin{aligned}
&\{\emptyset,\\
&\{1\}, \{2\}, \{3\}, \{4\}, \{5\},\\
&\{1,2\}, \{2,3\}, \{3,4\}, \{4,5\}, \{1,5\},\\
&\{2,3,4\}, \{3,4,5\}, \{1,4,5\}, \{1,2,5\}, \{1,2,3\},\\
&\{1,2,3,4,5\}\}.
\end{aligned}
$$

Notice that the closure of any subset of $V(C_5)$ appears in $\mathcal{C}$.

The decision problem “Given a graph G and integer k, is $m(G,2) \le k$” is NP-hard [32], however there do exist classes of graphs for which there are polynomial time algorithms (or constant upper bounds) to compute $m(G,2)$ exactly (see [8, 29, 40, 31, 46, 65, 45] and references therein). There are also classes of graphs for which the decision problem is (surprisingly) NP-complete, such as planar graphs with maximum degree four [45].

2.2 Is the Percolation Number Bounded?

We begin this section with a generalization of a result in Bushaw et al. [25]. Recall that the girth of a graph G is the length of the shortest cycle in G, or infinity if G is acyclic.

Proposition 2.1 ([25])**.** *If G is 2-BG and not P_3, then G has girth less than five.*

Thinking contrapositively, Proposition 2.1 gives a lower bound $3 \leq m(G, 2)$ in the case that G does not contain any triangles or any copy of C_4 as a subgraph. We can extend Proposition 2.1 to any threshold r.

Theorem 2.2. *Let $r \geq 2$ and let G be a graph with order at least $r+1$. If G is r-BG and is not the star with r leaves, then G has girth less than five.*

Proof. Fix $r \geq 2$ and let $I \subset V(G)$ where $|I| = r$ and I percolates. We assume G is not the star with r leaves. Since I percolates and $V(G) - I$ is nonempty, there is a vertex $x \in V(G) - I$ which is adjacent to every vertex in I. If I is not an independent set, then G contains a triangle. Assume now that I is an independent set. So $I' = I \cup \{x\}$ induces a star with r leaves. It cannot be that $V(G) = I'$, as by assumption G is not the star with r leaves. So $V(G) - I'$ is nonempty. Since I percolates, there must be at least one vertex $y \in V(G) - I'$ where y has at least r neighbors in I'. Thus either y is adjacent to x and adjacent to a neighbor of x, forming a triangle, or y shares at least two neighbors with x, forming a C_4. This proves that the girth of G must be less than five. $\square$

Intuitively, large enough girth restricts the ways in which an initial infected set can percolate; if one infects too few vertices initially, in this case the fewest possible, then a triangle or C_4 is forced past the initial iteration of infection. In contrast, having small girth does not give us much information, that is, all we would know is that there exists some small cycle somewhere in the graph.

Definition 2.3. Let G be a graph. A subgraph H of G is an *r-forbidden subgraph* if every vertex in H has at most $r-1$ neighbors in $G-H$.

The concept of an r-forbidden subgraphs is useful in understanding which vertices one must infect in order to percolate a graph.

Lemma 2.4 (Hedžet and Henning [62])**.** *Let $A \subseteq V(G)$ such that $\langle A \rangle = V(G)$. If H is an r-forbidden subgraph of G, then $V(H)$ must contain at least one vertex of A.*

Lemma 2.4 can be understood in the following way. Let H be an r-forbidden subgraph of G. Then, if every vertex of a percolating set is in $G-H$, the infection could not spread to H; every vertex in H has at most $r-1$ neighbors in $G-H$, too few to become infected. Note that $G-v$ for any vertex v is an r-forbidden subgraph trivially. In a triangle-free graph, for any edge uv the graph $G-\{u,v\}$ is an r-forbidden subgraph. The concept of r-forbidden subgraphs gives us one way to understand where the vertices of a percolating set must lie.

Cappelle et al. [29] prove the following upper bound.

Theorem 2.5 ([29, Theorem 10])**.** *If G is a 2-connected graph with diameter two, then $m(G,2) \leq \lceil \log_2(\Delta(G)+1) \rceil + 1$.*

One core idea used in establishing Theorem 2.5 and several others like it is what follows.

Observation 2.6. *Let G be a graph and $A \subseteq V(G)$. Any vertex in $V(G) - \langle A \rangle$ has at most $r-1$ neighbors in $\langle A \rangle$. That is, the graph induced by $V(G) - \langle A \rangle$ is an r-forbidden subgraph.*

The observation is immediate; if a vertex outside of $\langle A \rangle$ had at least r neighbors in $\langle A \rangle$, then that vertex would also be in $\langle A \rangle$.

Lemma 2.7. *Let G be a graph, let $A \subseteq V(G)$, and fix $r = 2$. Let $X \subseteq V(G) - \langle A \rangle$ be the set of vertices who have exactly one neighbor in $\langle A \rangle$. If X induces a connected subgraph, then for any $x \in X$, we have $X \subset \langle A \cup \{x\} \rangle$.*

Proof. Let G, A, and X be as described, and let $H = G[X]$. Let $x \in X$. Since H is connected, there is a path from x to every other vertex in H. Let x' be another vertex in H and P be a path between x and x'. Since every vertex in H has an infected neighbor in $\langle A \rangle$, if x is infected, each vertex along P beginning with the neighbor of x will become infected until x' is reached. □

Lemma 2.8. *Let G be a 2-connected graph with diameter two, $A \subseteq V(G)$ and fix $r = 2$. If $\langle A \rangle$ is a dominating set of G, then the graph induced by $V(G) - \langle A \rangle$ is connected.*

Proof. Let $H = G[V(G) - \langle A \rangle]$. If $\langle A \rangle$ is a dominating set, then every vertex in H has at least one neighbor in $\langle A \rangle$. By Observation 2.6 that for any $v \in V(H)$, the number of neighbors of v in $\langle A \rangle$ is at most one. Therefore, every vertex $v \in V(H)$ has a unique neighbor in $\langle A \rangle$.

Suppose toward contradiction that H is disconnected. Then there is a partition $V(H) = X \cup Y$ so that X is anticomplete to Y. Let $x \in X$ and $y \in Y$. Since G has diameter two and X is anticomplete to Y, it must be that x and y have a common neighbor $w \in \langle A \rangle$. We claim every vertex in H has w as a neighbor.

Indeed, for every vertex $y' \in Y$, it must be that y' and x share a common neighbor in $\langle A \rangle$. Similarly for every vertex $x' \in X$, it must be that x' and y share a common neighbor in $\langle A \rangle$. As w is the unique neighbor of both x and y in $\langle A \rangle$, it must be that w is also the unique neighbor of every x' and y'. Since every vertex in H has w as a unique neighbor in $\langle A \rangle$, this implies that w is a cut-vertex of G. This is a contradiction, as G is 2-connected. □

Lemma 2.9. *Let G be a 2-connected graph with diameter two. If H_1 and H_2 are vertex-disjoint 2-forbidden subgraphs of G, there must be at least one edge with an endpoint in H_1 and an endpoint in H_2.*

Proof. Let G, H_1, and H_2 be as described. Suppose there is no edge between H_1 and H_2. Let $G' = G - (H_1 \cup H_2)$. Since G has diameter two, every vertex in $V(H_1) \cup V(H_2)$ has a neighbor in G'. Thus $\langle V(G') \rangle = V(G')$ is a dominating set, and by Lemma 2.8, the graph $G[H_1 \cup H_2]$ is connected. □

The following proof is found in [29], except here it is in the language of r-neighbor bootstrap percolation.

Proposition 2.10 ([28, Lemma 2.2], [29, Lemma 4])**.** *Let G be a 2-connected graph with diameter two and $A \subseteq V(G)$. If $\langle A \rangle$ is a dominating set, then $m(G,2) \leq |A|+1$.*

Proof. By Lemma 2.8, the graph $H = G - \langle A \rangle$ is connected. By Lemma 2.7 for any $x \in V(H)$ we have that $A \cup \{x\}$ percolates G. □

Bushaw et al. [25] prove the following result about 2-BG graphs.

Theorem 2.11 ([25, Lemma 2.1])**.** *If G is 2-BG, then G has at most two blocks.*

This makes sense; blocks intersect in cut-vertices, and cut-vertices can act as bottlenecks for the infection. This can be observed in Figure 2, which shows a graph with two blocks. The cut-vertex stops the infection in one case, but there is a way in which an initial infected set $\{x, y\}$ has a chance to percolate if the cut-vertex is a common neighbor of x and y. If a graph has more than three blocks, then one of two scenarios may occur. Either all blocks share the same cut-vertex, or there are two blocks which do not share a cut-vertex (i.e. there are at least two cut-vertices). In the first case, each block is a 2-forbidden subgraph and so by Lemma 2.4, $m(G,2)$ is at least the number of blocks. In the second case, one can possibly infect two blocks intersecting in a cut-vertex as in Figure 2, in which case the infection cannot travel beyond the other cut-vertex, and otherwise choosing any two vertices which do not have a cut-vertex as a common neighbor will not result in percolation. Furthermore in the latter case, the leaves of the block graph of G are 2-forbidden subgraphs. Figure 3 shows two small examples illustrating both cases.

One can find several classes of graphs with unbounded percolation number for any threshold, the simplest being paths and cycles. For example, $m(C_k, 2) = \lceil n/2 \rceil$ and $m(C_k, r) = k$ for $r \geq 3$, as every vertex in the latter instance has degree strictly less than r. It is easy to construct families of graphs with unbounded r-percolation number in a trivial sense. That is, any vertex with degree less than the threshold must be in the initial infected set, and if $\Delta(G) < r$ then $m(G,r) = |G|$ (subgraphs induced on a single vertex are an r-forbidden subgraph). Furthermore from Theorem 2.11 we

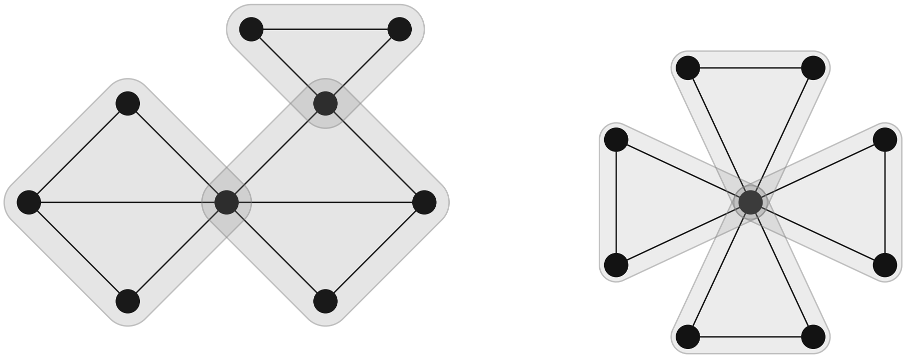


Fig. 3. Two examples of graphs with more than two blocks, with a friendship graph on the right. Blocks are highlighted in gray.

see a hint that connectivity is intimately tied with r-neighbor bootstrap percolation. The r-percolation number is unbounded for the class of graphs with a cut-vertex. To see this, one can take the friendship graphs F_n (see F_4 in Figure 3) or a class of generalized friendship graphs, where each block is a clique of size at least r. This shows it is possible to construct examples which are non-trivial in the sense of the degree sequence of G.

Another invariant, the diameter, also plays a key role in determining whether the r-percolation number is unbounded for a class of graphs. Intuitively, one would choose an initial infected set in which the vertices share many neighbors. Large diameter may affect which vertices are contained in a minimum percolating set. For example, when $r = 2$, a pair of vertices at distance three from one another would never percolate.

Recall that a graph G is k-connected if the removal of any set of $k-1$ vertices from G leaves a connected graph. Using Lemma 2.4, we can construct families of graphs for every threshold $r \geq 2$ and diameter $d \geq 2$, other than when $r = d = 2$, that have unbounded r-percolation number. The constructed graphs will be r-connected (and thus minimum degree at least r).

Example 2.12 (Unbounded Percolation Number, $d \geq 3, r \geq 2$)**.** The first construction is as follows. Fix $d \geq 3$ and $r \geq 2$, and let $t \geq 2$ and $s \geq r$. First, take the disjoint union sK_t, $X = K_{(r-1)t}$, and $d-1$ copies of K_r when $d \geq 4$ and zero other-

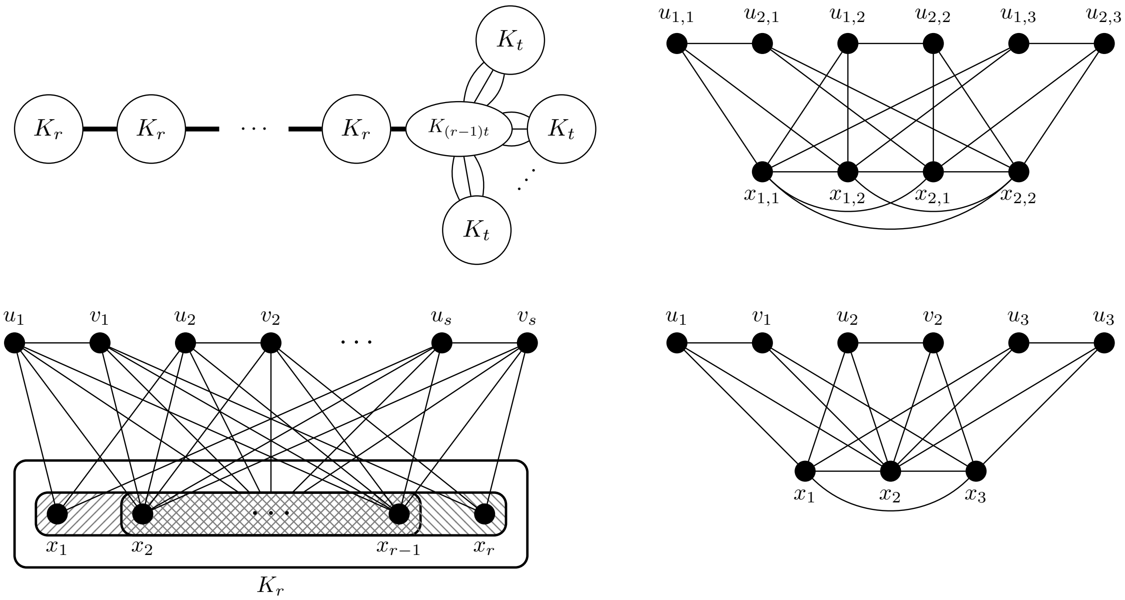


Fig. 4. The first construction on the top row, and the second construction on the bottom row. A schematic for the constructed graphs on the left, with examples on the right. In the top right is the constructed example for $d = 3$, $r = 3$, and $t = 2$. On the bottom right is the constructed example for $d = 2$ and $r = 3$.

wise. We refer to each copy of K_t as a leaf clique and X as the central clique. Let $u_{i,j}$ be the ith vertex in the jth copy of K_t, and let

$$V(X) = \bigcup_{k=1}^{t} \{x_{k,1}, x_{k,2}, \dots, x_{k,r-1}\}.$$

Set $N_X(u_{i,j}) = \{x_{i,1}, x_{i,2}, \dots, x_{i,r-1}\}$ for every $1 \leq i \leq t$ and $1 \leq j \leq s$. If $d = 3$ we are done. If $d \geq 4$, let $Y_1, Y_2, \dots, Y_{d-1}$ be the copies of K_r. then we add all possible edges between Y_i and Y_{i+1} for $1 \leq i \leq d-2$, and all possible edges between Y_{d-1} and X. Any vertex in Y_1 is distance d from any vertex in a leaf clique. If $d = 3$, then any vertex in a leaf clique has eccentricity three, as $d(u_{i,j}, u_{k,\ell}) = 3$ for $i \neq j$ and $k \neq \ell$. This achieves the desired diameter. When $r = t = 2$, this is the stacked book graph $B_{s,2}$.

Each leaf clique is an r-forbidden subgraph, so the percolation number is at least s, which is at least r. See Figure 4.

Example 2.13 (Unbounded Percolation Number, $d = 2, r \geq 3$)**.** The second construction is as follows. Fix $d = 2$ and $r \geq 3$ and let $s \geq r$. Let $G = K_r \cup sK_2$. We will call the copy of K_r the central clique X and the copies of K_2 leaf cliques $L_1, L_2, \ldots, L_s$. Let $V(L_i) = \{u_i, v_i\}$, and let $V(X) = \{x_1, x_2, \ldots, x_r\}$. Set $N_X(u_i) = \{x_1, x_2, \ldots, x_{r-1}\}$ and $N_X(v_i) = \{x_2, x_3, \ldots, x_r\}$ for all $1 \leq i \leq s$. The vertices $X' = N(u_i) \cap N(v_i) = \{x_2, x_3, \ldots, x_{r-1}\}$ are dominating vertices of G. Since G is connected with a dominating vertex, and G is not a clique, the diameter of G is two. Each leaf clique is an r-forbidden subgraph, so the percolation number is at least s, which is at least r. See Figure 4.

It is interesting to note that the second construction contains $r - 2$ dominating vertices, and yet through the power of r-forbidden subgraphs the r-percolation number is unbounded.

The question remains if one can construct a family of 2-connected graphs with diameter two and unbounded 2-percolation number. Certainly, one can look to the star, or consider the friendship graph and natural generalizations. These graphs have many mutually vertex-disjoint 2-forbidden subgraphs. However, they also have a cut-vertex, i.e. they are not 2-connected. In fact, if a graph G with diameter two has a cut-vertex v, then v is unique and dominating in G [122, Exercise 2.1.44]. For such a graph, we know $m(G, 2)$ exactly in terms of the number of connected components of $G - v$, or the number of blocks of G [29, Proposition 2].

In Examples 2.12 and 2.13, our main strategy for constructing families of graphs with unbounded percolation number was to ensure the existence of multiple mutually vertex-disjoint r-forbidden subgraphs. Theorem 2.14 will show that this strategy will not work in the case of 2-connected graphs with $d = r = 2$.

Theorem 2.14. *If G is 2-connected with diameter two, then G contains at most two vertex-disjoint 2-forbidden subgraphs.*

Proof. Let G be as described. Suppose toward contradiction that G contains three vertex-disjoint 2-forbidden subgraphs, say H_1, H_2, and H_3. Let $G' = G - \bigcup_{i=1}^{3} H_i$. By Lemma 2.9, there must be an edge xy with $x \in H_1$ and $y \in H_2$.

Since the unique neighbor of x in $G - H_1$ is y, and the unique neighbor of y in $G - H_2$ is x, for any $v \in G - (H_1 \cup H_2)$, we have v anticomplete to $\{x, y\}$. Let $z \in H_3$. Since G has diameter two, z must have a common neighbor with x in H_1 and a common neighbor with y in H_2. Since H_3 is 2-forbidden, this is impossible. $\square$

Example 2.15 (Two 2-forbidden subgraphs)**.** Examples of graphs with two vertex-disjoint 2-forbidden subgraphs are C_5 and the Petersen graph. In C_5, one can partition the graph into a P_3 and K_2 with a matching connecting them, or take two independent edges as the 2-forbidden subgraphs, with a remaining singleton. In the Petersen graph, the outer pentagon and the inner pentagram in the canonical drawing are each a 2-forbidden subgraph. See Figure 5.

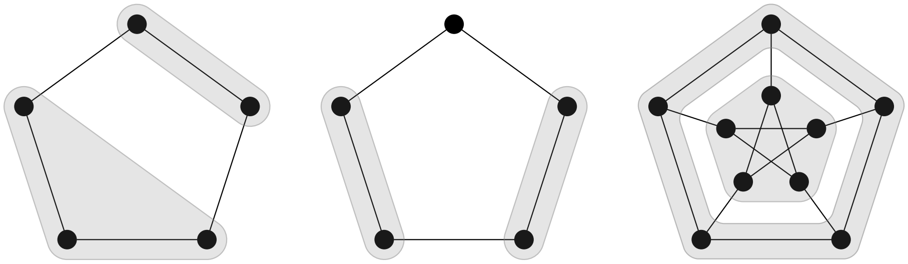


Fig. 5. Illustrations of vertex-disjoint 2-forbidden subgraphs in the five-cycle and the Petersen Graph.

Question 2.16. *Does there exist a constant k such that for any 2-connected graph G with diameter two, $m(G, 2) \leq k$?*

Question 2.16 was left open in [29]. Let $\mathcal{G}$ be the class of 2-connected graphs with diameter two. A step toward an answer to Question 2.16 may come in the form of an additional assumption on G, such as forbidding an induced subgraph, or bounding the degree. For example, consider cubic graphs in $\mathcal{G}$. We claim that such a graph G has at most ten vertices. Indeed, since G has diameter two, for a vertex $x \in V(G)$ there is a partition of $V(G)$ into sets D_0, D_1, and D_2 where D_i is the set of vertices at distance i from x. We have $|D_0| = 1$ and since G is cubic, $|D_1| = 3$. Since every vertex $y \in D_1$ has x as a neighbor, y can have at most two neighbors in D_2. This shows

$|D_2| \leq 6$, and $|V(G)| = |D_0| + |D_1| + |D_2| \leq 10$. Thus there are finitely many cubic graphs with diameter two, and those with 2-percolation number three are illustrated in Figure 6. The graph with ten vertices in this case is uniquely the Petersen graph, also known as a Moore graph. Moore graphs are discussed in more detail below.

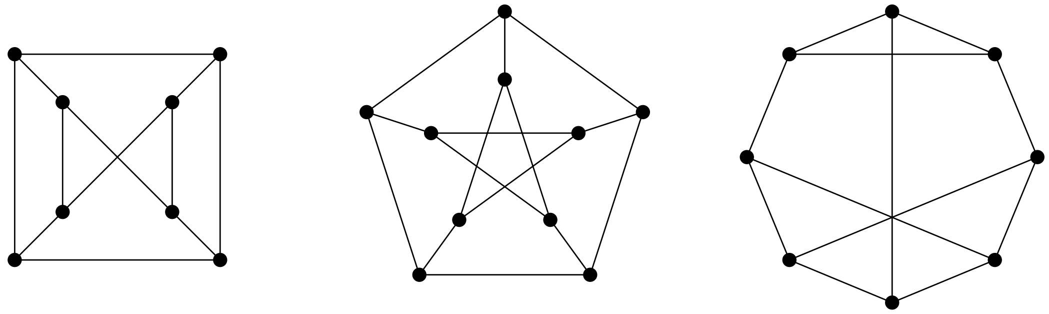


Fig. 6. The diameter two cubic graphs with $m(G, 2) = 3$.

Through efforts to construct graphs which are 2-BB, one realizes that girth equal to five is sufficient, but not necessary. That is, there are graphs which are 2-BB with girth less than five. Figure 6 shows two examples, one with girth four (left) and one with girth three (right). This is to say Theorem 2.1 is not true in the converse. Furthermore, there are infinitely many graphs with diameter $k \geq 3$ and girth at least five, with one construction involving a pair of vertices with an arbitrary number of paths of length k between them.

While there are infinitely many graphs with diameter at least three and girth at least five, the graphs with diameter two and girth five are more constrained. These graphs are known as Moore graphs. Moore graphs were originally defined in the following way. Let G be a graph with diameter k, maximum degree d, and $V(G) = \{v_1, \ldots, v_n\}$. Let D_i be the set of vertices at distance i from v_1. Then $|D_0| = 1$ and $|D_i| \leq d(d-1)^{i-1}$ for $i \geq 1$. This is indeed an upper bound on $|D_i|$, as in the worst case, each vertex in D_j has $d-1$ neighbors in D_{j+1} for $1 \leq j \leq i-1$. Note that $\cup_{i=0}^{k} D_i$ is a partition of $V(G)$. As a result,

$$\sum_{i=0}^{k} |D_i| = n \leq 1 + \sum_{i=1}^{k} d(d-1)^{i-1}. \tag{2.1}$$

Hoffman and Singleton [63] call the graphs attaining equality in Inequality 2.1 *Moore graphs of type* (d, k). Moore graphs are necessarily regular, and Singleton [107] showed an equivalent definition of Moore graphs are graphs with diameter k and girth $2k+1$, or in the case of diameter two graphs, girth five. For $k = 2$, Moore graphs exist only for $d = 2, 3, 7$ and possibly $d = 57$, with a unique graph for $d = 2, 3, 7$ [63]. It is currently unknown if a Moore graph of type $(57, 2)$ exists nor is it known if it must be unique [42]. The known Moore graphs of diameter two are C_5, the Petersen graph, and the Hoffman-Singleton graph, with 2-percolation numbers three, three, and four respectively. To the author's knowledge, there is no 2-connected graph with diameter two known to have $m(G, 2) = 4$ other than the Hoffman-Singleton graph.

This begs the question: are there infinitely many 2-connected graphs with diameter two with girth 3 or girth 4 (triangle-free) which are 2-BB? Furthermore, can the answers to these questions yield a construction of a 2-connected graph with diameter two different from the Hoffman-Singleton graph with 2-percolation number larger than three?

2.3 Matching Cuts and Complementary Prisms

Definition 2.17 (Matching cut, [55])**.** A *matching cut* is a matching in a connected graph whose removal disconnects the graph.

Matching cuts were first introduced by Graham in [56, Problem 20]. Results on graphs with a matching (or *simple*) cut, sometimes called *decomposable* graphs, appear in [55], where it was asked which graphs are decomposable. Later, Chvátal [38] showed that recognizing decomposable graphs is an NP-complete problem. One can alternatively define a matching cut in the following way. Define a partition $V(G) = X \cup Y$ to be a *cut*. Then (X, Y) is a matching cut if and only if each vertex in X has at most one neighbor in Y and each vertex in Y has at most one neighbor in X. In referring to matching cuts, we will always be clear about whether we are talking about the edge set or the separated parts.

Definition 2.18 (Complementary Prism)**.** The *complementary prism* of a graph G, denoted $G\overline{G}$, is the graph formed by the disjoint union of G and its complement $\overline{G}$

with a perfect matching joining pairs of corresponding vertices in G and $\overline{G}$.

Complementary prisms were introduced by Haynes et al. [61] as a special case of a more general complementary product, which is a generalization of the cartesian product. Complementary prisms are one class of graphs with a matching cut. One example of a well known complementary prism is the Petersen graph; see Figure 9 for several examples of complementary prisms.

Lemma 2.19 gives us a decomposition for graphs with diameter two and a matching cut.

Lemma 2.19. *Let G be a graph with a matching cut M separating G into subgraphs H_1 and H_2. Let $V(H_1) = X_1 \cup Y_1$ and $V(H_2) = X_2 \cup Y_2$ such that X_1 is anticomplete to X_2, and Y_1 and Y_2 comprise the endpoints of the edges in M. If for all $y \in Y_1$ and $y' \in Y_2$ we have $d_G(y, y') \leq 2$, then $G[Y_1 \cup Y_2]$ contains a complementary prism as a spanning subgraph.*

Proof. Let G be as described and let $G' = G[Y_1 \cup Y_2]$. Let $Y_1 = \{y_1, y_2, \ldots, y_k\}$ and $Y_2 = \{y'_1, y'_2, \ldots, y'_k\}$ so that $M = \{y_1y'_1, y_2y'_2, \ldots, y_ky'_k\}$.

If there is some i and j such that $y_iy_j \notin E(G[Y_1])$ and $y'_iy'_j \notin E(G[Y_2])$, then y_i and y'_j have no common neighbor in G', and are thus distance at least three apart in G, a contradiction. This implies that for any i and j, either $y_iy_j \in E(G[Y_1])$ or $y'_iy'_j \in E(G[Y_2])$. That is, non-edges y_iy_j in $G[Y_1]$ correspond to edges in $y'_iy'_j$ in $G[Y_2]$ and similarly non-edges $y'_iy'_j$ in $G[Y_2]$ correspond to edges y_iy_j in $G[Y_1]$. $\square$

Corollary 2.20. *If G is a graph with diameter two and a matching cut, then $V(G)$ can be partitioned into sets X_2, Y_1, and Y_2 such that*

- *$|Y_1| = |Y_2|$ with $G[Y_1 \cup Y_2]$ containing a complementary prism as a spanning subgraph with a perfect matching between Y_1 and Y_2, and*
- *X_2 dominates Y_2.*

Proof. Let G be a diameter two graph with a matching cut with matching M that separates G into subgraphs H_1 and H_2. There are partitions $V(H_1) = X_1 \cup Y_1$ and

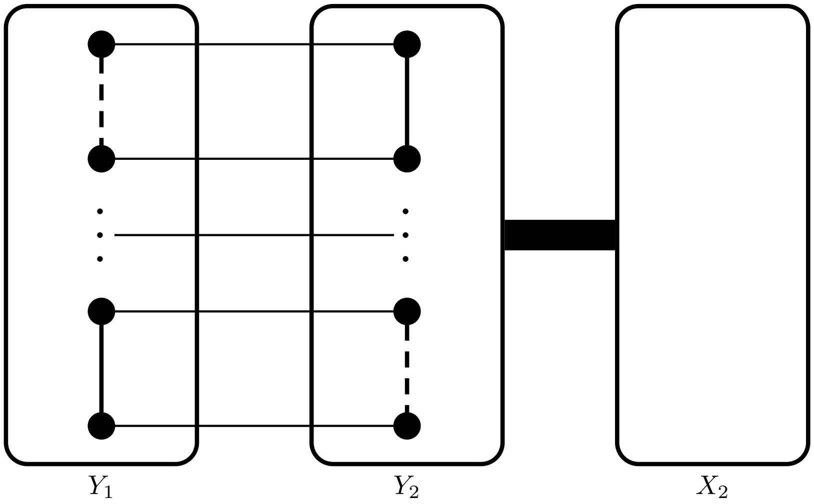


Fig. 7. A schematic of the decomposition admitted by a diameter two graph with a matching cut as described in Lemma 2.19.

$V(H_2) = X_2 \cup Y_2$ so that X_1 is anticomplete to X_2, and Y_1 and Y_2 are the endpoints of edges in M. If both X_1 and X_2 are nonempty, then the distance between any $x \in X_1$ and $x' \in X_2$ would be three, a contradiction. Without loss of generality assume X_1 empty and X_2 is possibly nonempty. So $|Y_1| = |Y_2|$. By Lemma 2.19, we have that $G[Y_1 \cup Y_2]$ contains a complementary prism as a spanning subgraph. No vertex of X_2 has a vertex in Y_1 as a neighbor, and since G has diameter two every vertex in Y_1 has a common neighbor with every vertex of X_2 in Y_2. Since each vertex of Y_1 has a unique neighbor in Y_2, every vertex in X_2 must dominate Y_2. This gives us the decomposition. □

Lemma 2.19 and Corollary 2.20 will be used later in the proofs of Lemma 2.22 and Theorem 2.26 respectively.

One potential strategy for constructing 2-BB graphs is to guarantee the resulting graph contains two vertex-disjoint 2-forbidden subgraphs, and utilize the arising structure with additional properties, e.g. the triangle-free property. We explain below the details of percolation on a triangle-free graph and matching cut.

Let G be a 2-connected graph with diameter two, and let G be triangle-free with order at least five and a matching cut separating G into subgraphs H and H'.

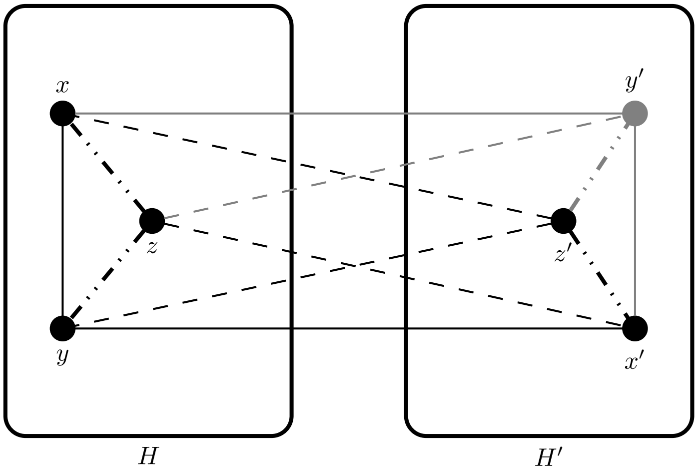


Fig. 8. The structure of a graph which is triangle-free, has diameter two, and has a matching cut, through the lens of percolation. Dashed lines represent non-edges. For a fixed vertex and all edges incident to that vertex which are dash-dotted, at most one dash-dotted edge exists in the graph. Here, z and z' represent any other vertex in H and H' respectively. At most one of $\{xz, yz\}$ and at most one of $\{x'z, y'z\}$ exist in the graph. As y' may possibly not exist, it is colored gray along with its incident edges.

Then H and H' are vertex-disjoint 2-forbidden subgraphs in G, and by Lemma 2.4, a percolating set must contain at least one vertex from H and H'. In order for a set of size two to percolate, one must choose vertices $x \in H$ and $x' \in H'$ where $xx' \notin E(G)$ and x and x' have a common neighbor y. Without loss of generality assume $y \in V(H)$. So $y \in \langle\{x, x'\}\rangle$, and there is possibly a $y' \in H'$ having x and x' as neighbors ($xyx'y'x$ is a four cycle) so that $y' \in \langle\{x, x'\}\rangle$. Now, two conditions hold showing that $\langle\{x, x'\}\rangle = \{x, x', y, y'\}$. Figure 8 highlights these conditions. Let $z \in H - \{x, y\}$ and $z' \in H' - \{x', y'\}$.

1. Since H and H' are 2-forbidden, z does not have x' or y' as a neighbor, and z' does not have x or y as a neighbor.

2. As G is triangle-free, z has at most one neighbor in $\{x, y\}$ and z' has at most one neighbor in $\{x', y'\}$.

Note if $|G| \leq 4$, then G is one of K_2, P_3, or C_4.

Using Lemma 2.19, 2-connected graphs with diameter two with a matching cut contain a spanning complementary prism. We show that there are finitely many complementary prisms which are triangle-free.

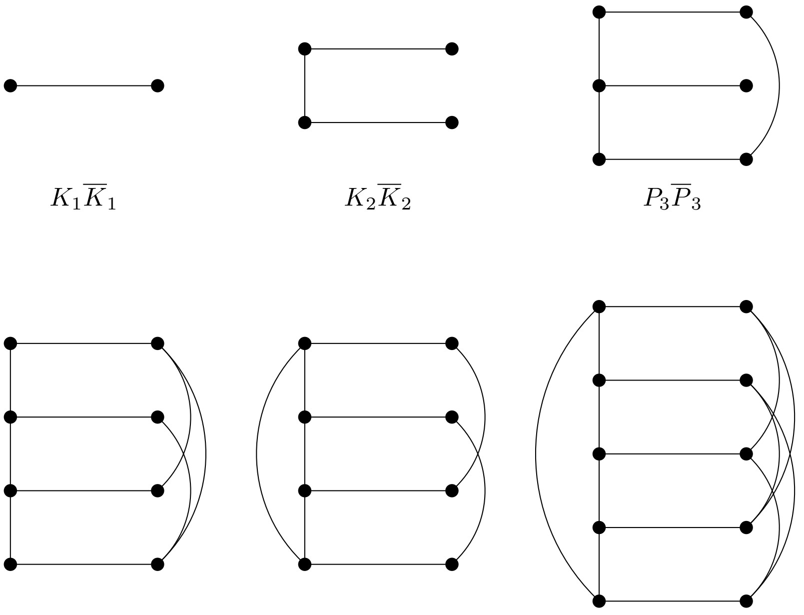
Proposition 2.21. *If* $G\overline{G}$ *is triangle-free, then* G *is one of* K_1*,* K_2*,* $\overline{K_2}$*,* $K_2 \cup K_1$*,* P_3*,* $2K_2$*,* P_4*,* C_4*, or* C_5*.*

Proof. The complement of a triangle-free graph has independence number at most two, and the complement of a graph with independence number at most two is triangle-free. Let G be a graph and assume $G\overline{G}$ is triangle-free. It follows then that G and $\overline{G}$ are triangle-free. However, since the complement of a triangle-free graph is a graph with independence number at most two, $\alpha(G)$ and $\alpha(\overline{G})$ are at most two. The neighborhood of any vertex in a triangle-free graph is an independent set. As G and $\overline{G}$ are both triangle-free with independence number two, this implies that $\Delta(G)$ and $\Delta(\overline{G})$ are at most two. If $|G| \geq 6$ then one of G or $\overline{G}$ must have a vertex with degree at least three. Thus $|G| \leq 5$, and $G\overline{G}$ is triangle-free for $G \in \{K_1, K_2, \overline{K_2}, K_2 \cup K_1, P_3, 2K_2, P_4, C_4, C_5\}$. □

Proposition 2.21 shows that there are nine graphs which are both triangle-free and have independence number two. Accounting for the graphs which are complements of each other, there are six unique triangle-free complementary prisms, namely $K_1\overline{K_1}, K_2\overline{K_2}, P_3\overline{P_3}, C_4\overline{C_4}, P_4\overline{P_4}$, and $C_5\overline{C_5}$. These are presented in Figure 9.

A graph G has a matching cut if and only if G can be partitioned into two 2-forbidden subgraphs. With this fact and the use of Lemma 2.19, we know the structure of 2-connected graphs with diameter two that contain two vertex-disjoint 2-forbidden subgraphs.

Lemma 2.22. *Let* G *be* 2*-connected with diameter two and let* G *contain two vertex-disjoint* 2*-forbidden subgraphs* H_1 *and* H_2 *such that* $V(G) \neq V(H_1) \cup V(H_2)$*. Letting* $V(H_1) = X_1 \cup Y_1$ *and* $V(H_2) = X_2 \cup Y_2$*, there is a partition of* $V(G)$ *into sets* $\{x\}$*,* X_1*,* X_2*,* Y_1*, and* Y_2 *such that* x *is complete to* X_1 *and* X_2*,* X_1 *is complete to* Y_1*,* X_2 *is complete to* Y_2*, and* $G[Y_1 \cup Y_2]$ *contains a complementary prism as a spanning subgraph.*



$K_1\overline{K}_1$ $K_2\overline{K}_2$ $P_3\overline{P}_3$

$P_4\overline{P}_4$ $C_4\overline{C}_4$ $C_5\overline{C}_5$

Fig. 9. The six triangle-free complementary prisms.

Proof. Let $G' = G - (H_1 \cup H_2)$. We claim that every vertex in G' has neighbors in both H_1 and H_2. Without loss of generality, suppose toward contradiction that there exists $x \in G'$ anticomplete to H_2. Since G has diameter two, x must have a common neighbor with every vertex $b \in H_2$. Since H_1 is a 2-forbidden subgraph, no vertex in H_1 can serve as a common neighbor of x and y. Thus, a common neighbor of x and y can only exist in G'. So $N_{G'}(x)$ dominates H_2. This implies there is no edge between H_1 and H_2 as H_2 is a 2-forbidden subgraph, with all outside edges going to G'. By Lemma 2.9, this is impossible. So x must have neighbors in both H_1 and H_2. Let $H_1' = H_1 - N_{H_1}(x)$ and $H_2' = H_2 - N_{H_2}(x)$. We claim that $|H_1'| = |H_2'|$. Indeed, let $y \in N_{H_1}(x)$. Since H_1 is a 2-forbidden subgraph of G, the only other neighbors of y lie in H_1. Thus in order for y to travel to H_2' in at most two steps, it must be that $N_{H_1'}(y)$ dominates $V(H_2')$. This is due to the fact that every vertex in H_2' has a unique common neighbor in H_1' with y. This shows that $|H_1'| \geq |H_2'|$. Mirroring the argument for $y' \in N_{H_2'}(x)$ yields $|H_2'| \geq |H_1'|$, and thus $|H_1'| = |H_2'|$. This then implies that every vertex in $N_{H_i}(x)$ dominates H_i', and there is a perfect matching between

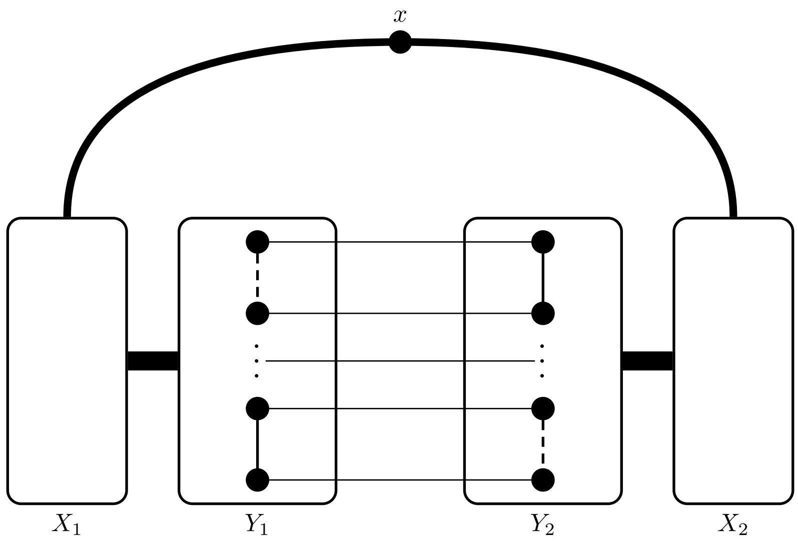


Fig. 10. The structure of a 2-connected graph with diamter two and two vertex-disjoint 2-forbidden subgraphs that do not partition the graph.

H_1' and H_2'. From this we also have $V(G') = \{x\}$; if G' contained any other vertex its only choice for neighbors is in G' as every vertex in H_1 and H_2 have an outside neighbor.

Since G has diameter two, any pair $a \in H_1'$ and $b \in H_2'$ are distance at most two apart. By Lemma 2.19 the graph induced on $V(H_1') \cup V(H_2')$ contains a complementary prism as a spanning subgraph. This completes the proof. □

If $|H_1'| = |H_2'| = 1$ then G must be C_5 (see Figure 5 center). Figures 7 and 10 taken together illustrate the structure of 2-connected graphs with diameter two and with two vertex-disjoint 2-forbidden subgraphs.

Theorem 2.23 (Duarte et al. [46]). *Let G be a graph. If G and $\overline{G}$ are connected, then $m(G\overline{G}, 2) \leq 5$. If G has k components with $k \geq 2$, then $m(G\overline{G}, 2) = k + 1$.*

The bound for graphs G such that G and $\overline{G}$ are connected can be improved. By combining two theorems in [61] we have the following theorem.

Theorem 2.24 (Haynes et al. [61, Theorem 2, Theorem 3]). *Let G be a graph with order at least two. Then $\operatorname{diam}(G\overline{G}) = 2$ if and only if $\operatorname{diam}(G) = \operatorname{diam}(\overline{G}) = 2$.*

In Theorem 2.24, diameter two complementary prisms are a special case of Theorem 2.23. Here we are able to improve the bound from five to three.

Theorem 2.25. *If G is a graph such that G and $\overline{G}$ are connected then $m(G\overline{G}, 2) \leq 3$.*

Proof. First, we claim that $G\overline{G}$ contains a matched five cycle with a pendant as in Figure 11. Start with vertices $v_1, v_2 \in V(G)$ which are independent. Since G is connected, v_2 has a neighbor $v_3 \in V(G)$. Then there are corresponding vertices in $\overline{G}$, namely $\overline{v_1}$, $\overline{v_2}$, and $\overline{v_3}$ with $\overline{v_1}$ and $\overline{v_2}$ adjacent and $\overline{v_2}$ and $\overline{v_3}$ nonadjacent. Since $G\overline{G}$ is a complementary prism, either v_1 and v_3 are adjacent or $\overline{v_1}$ and $\overline{v_3}$ are adjacent. In either case, the subgraph H induced by $\{v_1, v_2, v_3, \overline{v_1}, \overline{v_2}, \overline{v_3}\}$ is a five cycle with a pendant attached, which requires three vertices to infect. Without loss of generality let $\{v_1, \overline{v_2}, \overline{v_3}\}$ be a percolating set of H.

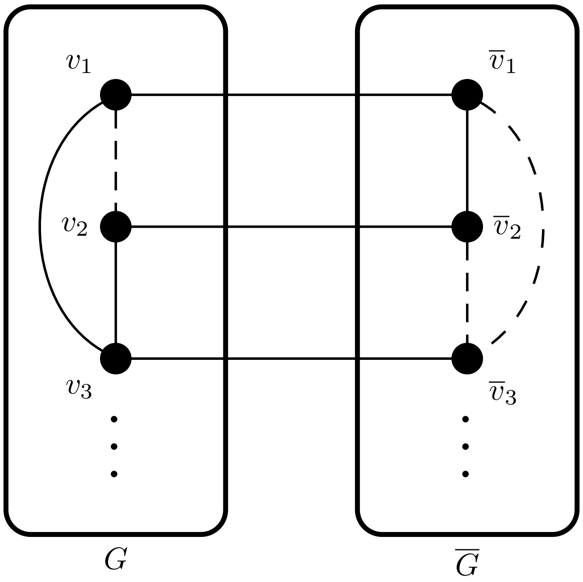


Fig. 11. Any complementary prism $G\overline{G}$ with G and $\overline{G}$ connected contains a matched five cycle with a pendant, with three matched vertices in each part, up to permutation of v_1, v_2, v_3 and interchange of G and $\overline{G}$.

Next, we claim that $\{v_1, \overline{v_2}, \overline{v_3}\}$ is a percolating set of G. Let $A_G = \{v_1, v_2, v_3\}$ and $A_{\overline{G}} = \{\overline{v_1}, \overline{v_2}, \overline{v_3}\}$. Since G is connected, there is some $v \in V(G)$ not yet infected with at least one neighbor in A_G. Note that as $G\overline{G}$ is a complementary prism, the non neighbors of v in A_G are the neighbors of $\overline{v}$ in $A_{\overline{G}}$. That is, either v has two neighbors in A_G or $\overline{v}$ has two neighbors in $A_{\overline{G}}$ (note this does not work if $|A_G| \leq 2$ and $|A_{\overline{G}}| \leq 2$). Using the definition of a complementary prism, all pairs v and $\overline{v}$ come in two forms:

1. The vertex v has at least one neighbor in A_G and $\overline{v}$ has at least one neighbor in $A_{\overline{G}}$.
2. Either v is anticomplete to A_G or $\overline{v}$ is anticomplete to $A_{\overline{G}}$.

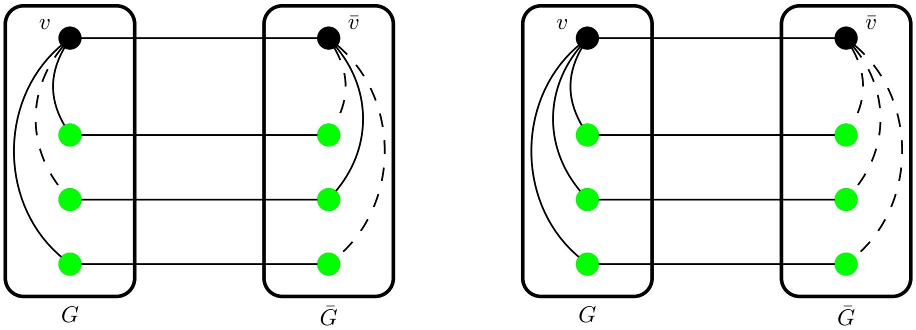


Fig. 12. The two ways in which pairs of vertices in a complementary prism can have neighbors in a particular infected set.

For the first type, without loss of generality, v must have at least two neighbors in A_G, and thus $\overline{v}$ has one vertex in $A_{\overline{G}}$ and v as infected neighbors. That is, both vertices in a type (1) pair become infected. For type (2) pairs, either v or $\overline{v}$ becomes infected, but not both. So, every pair $(v, \overline{v})$ come in two forms; either v and $\overline{v}$ are both infected, or exactly one of v and $\overline{v}$ is infected. We focus on the latter subset of pairs.

Without loss of generality, if all of G is infected in $G\overline{G}$, then by Lemma 2.7, $\overline{G}$ must also become infected; $V(G)$ dominates $V(\overline{G})$ and $\overline{G}$ is connected. Suppose G and $\overline{G}$ are not eventually infected. Then there are vertices $u, \overline{u}$ and $v, \overline{v}$ such that u is infected, $\overline{u}$ is uninfected, v is uninfected, and $\overline{v}$ is infected. We will call $(u, \overline{u})$ and $(v, \overline{v})$ opposite pairs. Since $G\overline{G}$ is a complementary prism, if u is not adjacent to v, then $\overline{u}$ is adjacent to $\overline{v}$ and if $\overline{u}$ is not adjacent to $\overline{v}$, then u is adjacent to v. That is to say, $G\overline{G}$ must have one of uv or $\overline{uv}$ as an edge. Without loss of generality assume u is not adjacent to v. Then $\overline{u}$ has u and $\overline{v}$ as infected neighbors. That is $\overline{u}$ must

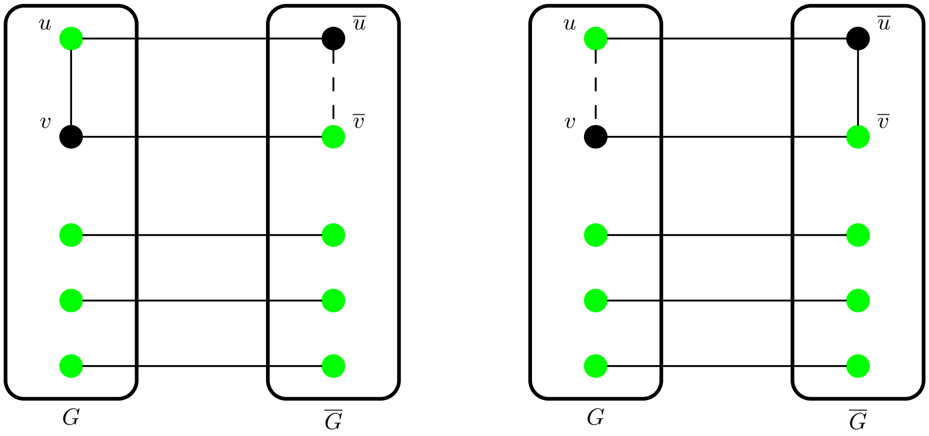


Fig. 13. Opposite pairs in a complementary prism at an arbitrary point of infection.

eventually become infected. Iterating this comparison, either there exists an opposite pair, or all vertices in either G or $\overline{G}$ are infected.

□

Complementary prisms have been studied in the related concepts of monophonic convexity and geodetic convexity [89, 90, 30]. The house graph and its complement, the path on five vertices, highlight the difference between the P_3-hull number and the monophonic interval number in Theorem 4.6 of [90] and the geodetic hull number in Theorem 3.8 of [30]. See Figure 14.

Theorem 2.26. *If G is a 2-connected graph with diameter two and contains two vertex-disjoint 2-forbidden subgraphs, then $m(G,2) \leq 3$.*

Proof. Let G be a graph which is 2-connected with diameter two. Let H_1 and H_2 be two vertex disjoint 2-forbidden subgraphs of G , and let $G' = G - (H_1 \cup H_2)$.

Case 1. G' nonempty. By Lemma 2.22, we have $V(G') = \{x\}$, and a partition of $V(G)$ into $\{x\}$, $Y_i = V(H_i) - N_{H_i}(x)$, and $X_i = N_{H_i}(x)$ for $i = 1, 2$, where $G[Y_1 \cup Y_2]$ contains a complementary prism as a spanning subgraph, and X_i is complete to Y_i. It is a fact that for any graph G, either G or $\overline{G}$ is connected. Without loss of generality assume $G[Y_1]$ is connected. A set comprising one vertex in X_1, one in X_2, and one in Y_1 is a percolating set. So $m(G,2) \leq 3$.

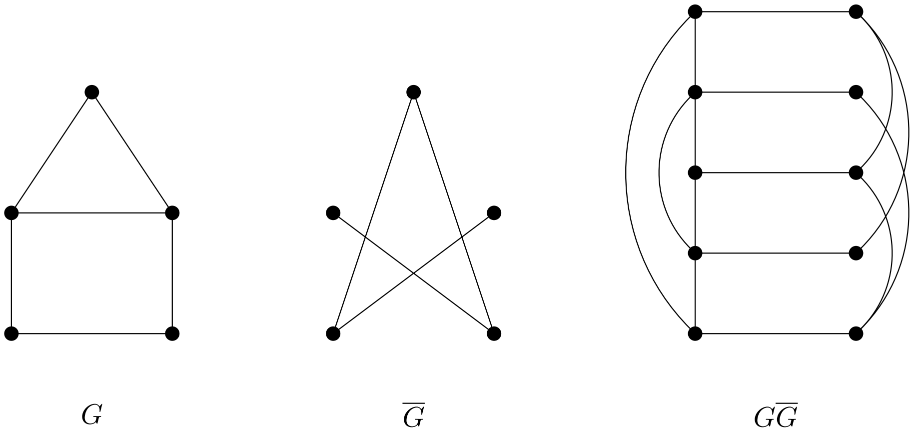


Fig. 14. The house graph G, its complement $\overline{G} \cong P_5$, and its complementary prism. The monophonic number and geodetic hull number of $G\overline{G}$ are two, but the P_3-hull number is three.

Case 2. G' empty. There is a matching cut M separating H_1 and H_2. By Corollary 2.20, we have $V(H_1) = X_1 \cup Y_1$ and $V(H_2) = X_2 \cup Y_2$ such that $G[Y_1 \cup Y_2]$ contains a spanning complementary prism and X_i is complete to Y_i for $i = 1, 2$. By Theorem 2.25 it follows that $m(G[Y_1 \cup Y_2], 2) \leq 3$. Note that since G is 2-connected, we have $|Y_1| = |Y_2| \geq 2$. So, infecting $G[Y_1 \cup Y_2]$ leads to X_1 and X_2 becoming infected.

□

There are exactly six triangle-free complementary prisms. Let G be a 2-connected graph with diameter two and let G be triangle-free with two vertex-disjoint 2-forbidden subgraphs H_1 and H_2. If $V(G) = V(H_1) \cup V(H_2)$, then using Lemma 2.19 and Proposition 2.21 G must either be one of the six triangle-free complementary prisms, P_3, or the graph constructed by taking $K_2\overline{K_2}$ and adding an independent set which is complete to the $\overline{K_2}$ part (this includes C_5). See Figure 15 (left) for an illustration of the last of the mentioned graphs.

Otherwise if $V(G) - (V(H_1) \cup V(H_2))$ is nonempty, G exhibits the decomposition in Lemma 2.22. Referencing Figure 10, if G is triangle-free then $G[Y_1 \cup Y_2]$ can only be isomorphic to K_2, as otherwise $G[Y_1]$ or $G[Y_2]$ contains an edge and forms a triangle with a vertex in X_1 or X_2 respectively. Thus in this case G can only take the form where $G[Y_1 \cup Y_2] \cong K_2$, and X_1 and X_2 contain an arbitrary number of vertices each. In other words, G is the disjoint union of a singleton x with an edge uv with an arbitrary number of paths of length two between x and u and between x and v (See Figure 15 (center)). Notice C_5 can be expressed as the left and center graphs in Figure 15, and the 2-forbidden subgraphs are illustrated in Figure 5.

In the opposite direction, one can also find a 2-connected graph with diameter two which is 2-BB with arbitrarily many triangles. Take the disjoint union $K_n\overline{K_n}$ with an independent set of any size which is complete to the $\overline{K_n}$ part.

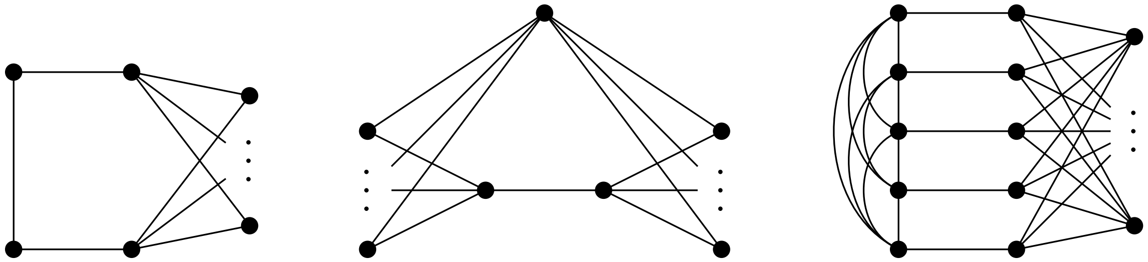


Fig. 15. Each graph is 2-connected with diameter two. Left: A triangle-free graph with two vertex-disjoint 2-forbidden subgraphs which partition the graph. Center: A triangle-free graph with two vertex-disjoint 2-forbidden subgraphs which do not partition the graph. Right: The graph $K_5\overline{K_5}$ with an independent set complete to the $\overline{K_5}$ part.

2.4 Local Connectivity and Strongly r-BG Graphs

We say a graph is *strongly r-BG* if every r-set of vertices percolate the graph. The star with at least two leaves is an example of a graph which is not strongly 2-BG, and also a counterexample to Theorem 2.27 with the assumption that G is connected rather than 2-connected.

Theorem 2.27. *Let G be a 2-connected graph of order $n \geq 4$. Then G is strongly 2-BG if and only if G does not contain a 2-forbidden subgraph H with $2 \leq |H| \leq n-2$.*

Proof. Assume G is strongly 2-BG. Suppose toward contradiction that G contains a 2-forbidden subgraph H with $2 \leq |H| \leq n-2$. Let $G' = G - H$, and note $|G'| \geq 2$.

Pick $A \subseteq V(G')$ as an initial infected set with $|A| = 2$. Notice that every vertex in H has at most one neighbor in G'. However, as G is strongly 2-BG, it must be that $\langle A \rangle = V(G)$. That is, the infection must spread from G' to H, and may only do so if at least one vertex of H has at least two neighbors in G'. This is a contradiction.

Now assume G does not contain a 2-forbidden subgraph H with $2 \leq |H| \leq n-2$. Let $A \subset V(G)$ such that $|A| = 2$ and let $H = G[\langle A \rangle]$. Suppose toward contradiction that $V(H) \neq V(G)$, i.e. A does not percolate. Let $G' = G - H$. Every vertex $v \in V(G')$ has at most one neighbor in H, as otherwise if v has two or more neighbors in H, then $v \in H$. This implies G' is a 2-forbidden subgraph of G. By assumption $|G'| \leq n - 2$, and since G is 2-connected, $|G|' \neq 1$. As $2 \leq |G'| \leq n - 2$, this is a contradiction. □

For $n = 3$, the triangle is a counterexample to Theorem 2.27. That is, K_3 is strongly 2-BG, but it contains a 2-forbidden subgraph of order two in the form of an edge. It remains an open problem of whether Theorem 2.27 may be generalized for graphs G with order at least $2r$ containing no r-forbidden subgraph H with $r \leq |H| \leq n - r$.

A graph is *locally connected* if the open neighborhood of each vertex in the graph induces a connected graph [34].

Proposition 2.28 (Bushaw et al. [25])**.** *Let G be a connected graph of order at least three. If G is locally connected, then G is 2-BG.*

A sketch of Proposition 2.28 is shown in Figure 16.

In [25] it is noted that any pair of adjacent non-pendant vertices percolate the graph. We further observe that the only connected graph which is locally connected with at least one pendant vertex is K_2, and that any two non-adjacent vertices with a common neighbor percolate. The converse of Proposition 2.28 is not true. One example of a graph which is 2-BG but not locally connected is given by C_n with edges added between every pair of vertices at distance two. This is known as the square of a cycle (see Figure 17). In the case that G has diameter two, we can show that the property of local connectedness is equivalent to being strongly 2-BG.

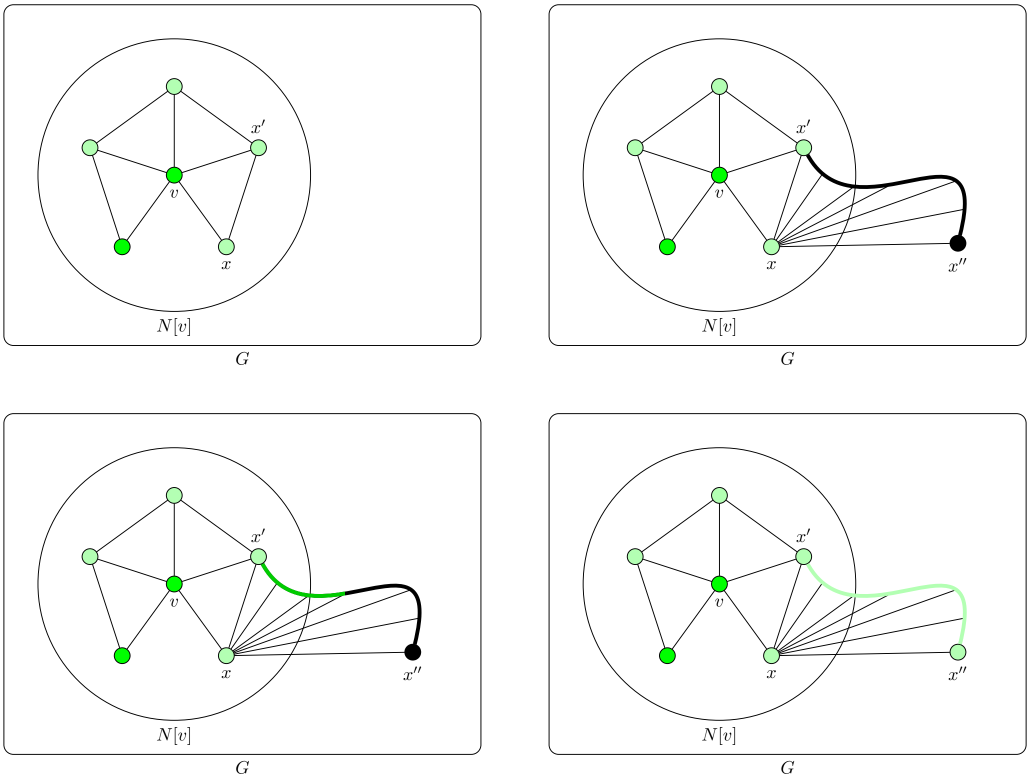


Fig. 16. The infection process in a locally connected graph is shown.

Corollary 2.29. *Let G be a locally connected graph. Then G is strongly 2-BG if and only if G has diameter at most two.*

Proof. Assume G is strongly 2-BG. Then by definition, every pair of vertices percolates G. If two vertices x and y are distance at least three from each other, then they have no common neighbors. That is, $\langle\{x, y\}\rangle = \{x, y\}$. Thus, every pair of vertices in G must be at distance two or less from each other.

Assume G has diameter at most two. Let $x, y \in V(G)$ such that x and y are not adjacent. Then since G has diameter two, x and y must share a common neighbor z. This implies $z \in \langle\{x, y\}\rangle$, and in particular infecting x and y leads to two infected pairs of adjacent vertices, namely $\{x, z\}$ and $\{y, z\}$. By Proposition 2.28 any pair of adjacent vertices percolates G. So $\langle\{x, y\}\rangle = V(G)$. We've shown that any pair of adjacent or non-adjacent vertices percolates G. □

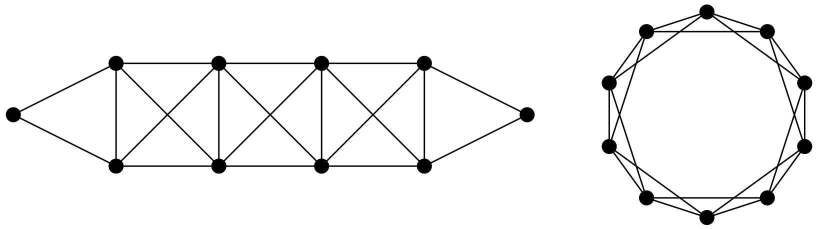


Fig. 17. A locally connected graph with diameter five which is not strongly 2-BG (left), and a strongly 2-BG graph which is not locally connected (right).

2.5 Maximal Percolating Sets and Their Structure

In this subsection we will assume that the threshold $r = 2$.

Definition 2.30 (Maximal Percolating)**.** A k-set $A \subseteq G$ is *maximal percolating* if there is no other k-set $A' \subseteq V(G)$ such that $\langle A \rangle \subset \langle A' \rangle$.

If there is a k-set of vertices which percolate, then $m(G, 2) \leq k$. However, if no k-set percolates the graph, then it may be useful to define a notion of maximal in relation to closures. We may gain some insight into the structure of G through the closure of k-sets of vertices which are not properly contained in the closure of another k-set. Maximal percolating sets are a natural object to define and study. Lemma 2.32 allows us to narrow down the structure of the subgraphs induced by maximal percolating k-sets, along with their interaction with the rest of the graph.

We begin with an observation.

Observation 2.31. *If G has diameter at most two, then $G[\langle A \rangle]$ has diameter at most two for any k-set $A \subseteq V(G)$ with $k \geq 2$.*

Observation 2.31 comes from the fact that G has diameter two. If any pair of vertices $x, y \in \langle A \rangle$ are nonadjacent, then they must have a common neighbor z in G. Since z is adjacent to both x and y, it must be that $z \in \langle A \rangle$. Thus every pair of vertices in $\langle A \rangle$ are adjacent or share a common neighbor.

This line of thinking may be a useful tool to prove things about the structure of a closure and it's interaction with the remainder of the graph. That is, any hereditary property of G is by definition a property of $G[\langle A \rangle]$. So global properties of the graph may give us insight into its local structure with regards to closures of vertex sets.

Lemma 2.32. *Let G be a 2-connected graph with diameter two, let $A \subseteq V(G)$ be a maximal percolating k-set with $\langle A\rangle \neq V(G)$ and $k \geq 3$, and let $H = G[\langle A\rangle]$. Then*

(i) H is two-connected.

(ii) Every vertex in $V(H)$ has at least one neighbor in $G - H$.

(iii) A is an independent set.

(iv) Every vertex in $V(H)$ has at most two neighbors in A.

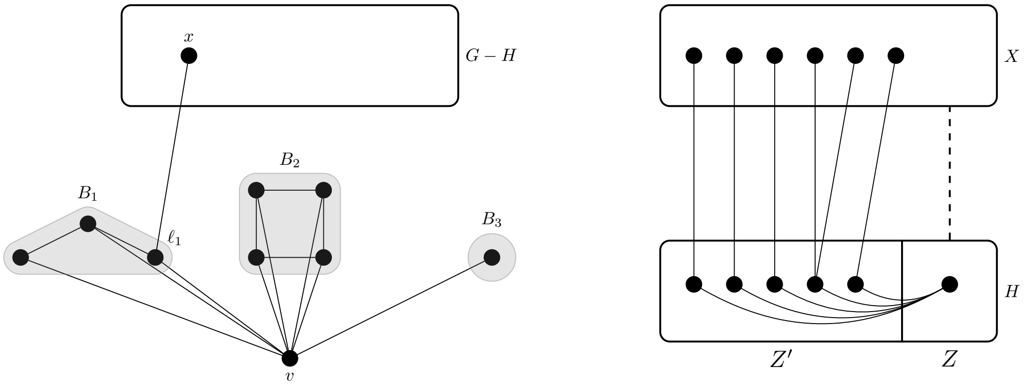


Fig. 18. Cases in the proof of Lemma 2.32. Left: Case (i), blocks are highlighted without their common cut-vertex v. Right: Case (ii).

Proof. (i) Suppose toward contradiction that H is not two-connected. By Observation 2.31, the diameter of H is two. Thus H must have a unique cut-vertex v with blocks mutually intersecting at v, and v must be dominating [122, Exercise 2.1.44]. Since each block minus v induces a 2-forbidden subgraph of H, by Lemma 2.4 each block contains exactly one vertex from A and thus there are k blocks. Let $V(H) = \{v\} \cup \bigcup_{i=1}^{k} B_i$ where B_i is the set of vertices in the ith block not including v (see Figure 18.) Let $A = \{\ell_1, \ell_2, \ldots, \ell_k\}$ with $\ell_i \in B_i$. by Corollary 2.29, any choice of ℓ_i works; since v is dominating, each block induces a graph which is locally connected with diameter two and thus strongly 2-BG. We claim that A is not maximal percolating. Indeed, if every vertex of $H - v$ has no neighbor in $G - H$, then v is a cut-vertex of G, a contradiction. So without loss of generality, we let $\ell_1 \in B_1$ be a vertex with a neighbor x in

$G-H$. But then $\langle\{x,\ell_2,\ldots,\ell_k\}\rangle$ contains A (and thus $\langle A\rangle$), as the vertices x and $\ell_2,\ldots,\ell_k$ infect v, and then ℓ_1 becomes infected by x and v. This shows A is not maximal percolating, a contradiction. So, H must be two-connected.

(ii) Let $X \subseteq V(G-H)$ be the vertices with a neighbor in H. Note since $G-H$ is 2-forbidden, vertices in X have exactly one neighbor in H. Let $V(H) = Z \cup Z'$ such that vertices in Z have no neighbor $G-H$, and vertices in Z' have at least one neighbor in $G-H$.

Since G has diameter two, each $z \in Z$ has a common neighbor with every vertex in X, and since each $z \in Z$ has no neighbors in $G-H$, the common neighbor must be in Z'. Notice that for $z', z'' \in Z'$, we have $N_X(z') \cap N_X(z'') = \emptyset$ (otherwise there is $x \in X$ where x has two neighbors in Z'.) So, for any $z \in Z$, it must be that z dominates Z'. If $|Z| \geq 2$, then any two vertices percolate H and thus A is not maximal percolating. If $Z = \{z\}$ then by (i) notice $H-z$ is connected. Since z dominates $H-z$ and H is 2-connected, H must be locally connected. By Observation 2.31, the diameter of H is two. Thus by Corollary 2.29 $\langle\{x,z\}\rangle$ contains H for any $x \in H-z$, a contradiction to the assumption that A is maximal percolating.

(iii) Suppose two vertices $x, y \in A$ are adjacent. By (ii), every vertex in A has a neighbor in $G-H$. Let z be the neighbor of x in $G-H$ and let $A' = \{z\} \cup A \setminus \{x\}$. Notice that infecting A' results in x becoming infected from z and y. Then $\langle A'\rangle$ contains A and thus A is not maximal percolating, a contradiction.

(iv) Suppose there is a vertex $x \in V(H) - A$ with at least three neighbors in A. Let $a \in A$ be a neighbor of x. By (ii) a has a neighbor $z \in G-H$, and similar to the proof of (iii), we let $A' = \{z\} \cup A \setminus \{a\}$. Then x becomes infected by $A \setminus \{a\}$, and a becomes infected by x and z. That is, $\langle A'\rangle$ contains $\langle A\rangle$, a contradiction. $\square$

2.6 Forbidden Induced Subgraphs

In joint work with Hudson LaFayette and Kevin McCall, we build upon the results in [25]. In [67], we prove Theorem 2.33, which is an improved version of a conjecture in [25].

Theorem 2.33 (Ibrahim, LaFayette, and McCall [67])**.** *Let G be a 2-connected graph with diameter two. If G is C_5-free, then $m(G, 2) = 2$.*

Proof. Let G be as described. Let $A \subset V(G)$ be maximal with $|A| = 2$ and let $H = G[\langle A \rangle]$. We claim H contains at least two dominating vertices. Let v be a vertex in H with a neighbor w in $G - H$, and suppose toward contradiction that v is not dominating in H. Then v has a non-neighbor x in H. Since G has diameter two, v and x have a common neighbor y. Since $v, x \in \langle A \rangle$ and y is adjacent to both v and x, it must be that y is in H. Similarly, x and w must have a common neighbor z, and z must be in $G - H$ as otherwise w would not be in $G - H$. That is, every vertex in $G - H$ has at most one vertex in H as a neighbor. By this observation and the independence of v and x, it must be that $\{v, w, x, y, z\}$ induces a C_5 in G. Since G is C_5-free, this is a contradiction. Thus, v must be dominating in H.

We have shown that for any edge uv with $u \in V(G - H)$ and $v \in V(H)$, it must be that v is dominating in H. Since G is 2-connected, there must be more than one vertex in H with a neighbor in $G - H$. In other words, H contains at least two dominating vertices. For two dominating vertices $x, y \in V(H)$, if x' is the neighbor of x in $G - H$, then $\langle \{x', y\} \rangle$ contains H. This implies A is not maximal percolating. □

We now prove a structural lemma.

Lemma 2.34. *Let G be a 2-connected graph with diameter two and let $A \subseteq V(G)$. If $H = G[\langle A \rangle]$, then $V(G - H) = X \cup Y$ where every vertex in X has exactly one neighbor in $V(H)$ and every vertex in Y has no neighbor in $V(H)$. Furthermore, if $G[X]$ is connected, then $m(G, 2) \leq |A| + 1$. Otherwise if $G[X]$ is disconnected, Y must be nonempty.*

Proof. Any vertex in $V(G - H)$ has at most one neighbor in $\langle A \rangle$, as otherwise such a vertex would have become infected. Thus, $G - H$ is a 2-forbidden subgraph of G and we can partition $V(G - H)$ into sets X and Y with every vertex in X having exactly one neighbor in $V(H)$ and Y anticomplete to $V(H)$.

If $G[X]$ is connected, then by Lemma 2.7 we have for any $x \in X$ that $X \subset \langle A \cup \{x\} \rangle$. In order for any vertex in $y \in Y$ to travel to a vertex in H in at most two steps, it must first travel to X. Since every vertex in X has a unique neighbor in $V(H)$, it must be that y has at least $|V(H)|$ neighbors in X. Thus $Y \subset \langle A \cup \{x\} \rangle$.

If $G[X]$ is disconnected, then by Proposition 2.10, it cannot be that Y is empty. □

Lemma 2.7 and Proposition 2.10 provide some insight in how an infection may grow by adding an additional vertex to some initial infected set. Let G be a 2-connected graph with diameter two. Recall that when the percolation threshold is 2, any vertex not contained in the closure of a set has at most one neighbor in that closure. For $A \subseteq V(G)$ the partition $V(G) = \langle A \rangle \cup X \cup Y$ is a helpful way to view G, where X is the set of vertices with exactly one neighbor in $\langle A \rangle$ and Y is the set of vertices with no neighbor in $\langle A \rangle$.

Theorem 2.35 (Cappelle et al. [29])**.** *Let G be 2-connected and diameter two. If G is C_6-free then $m(G, 2) \leq 4$*

We can improve the upper bound here by one.

Theorem 2.36. *Let G be 2-connected and diameter two. If G is C_6-free with $|G| \geq 6$ or C_7-free with $|G| \geq 7$ then $m(G, 2) \leq 3$.*

Here we present two proofs, one for C_6-free graphs, and another for both C_6-free and C_7-free.

Proof (C_6-free). Assume $\alpha(G) \leq 2$. Any maximal independent set is a dominating set, so by Lemma 2.7 we have $m(G, 2) \leq \alpha(G) + 1 \leq 3$. So we may assume $\alpha(G) \geq 3$. Let A be a maximal percolating 3-set of vertices in G. If $\langle A \rangle = V(G)$ then we are done. Otherwise, $\langle A \rangle \neq V(G)$ and by Lemma 2.32 we may assume A is an independent set of size three. Since G has diameter two, each pair of vertices in A must have a common neighbor, and by Lemma 2.32 since A is maximal percolating, there is no vertex in $\langle A \rangle$ adjacent to all of A. That is, the common neighbors of pairs in A are unique. Let the set of common neighbors be A'. Notice since A is maximal percolating and $A \subseteq \langle A' \rangle$, we have $\langle A \rangle = \langle A' \rangle$. This implies that A' is maximal percolating, and thus an independent set. But then $A \cup A'$ induces a copy of C_6 in G, a contradiction. $\square$

Proof 2. Let G be a 2-connected graph with diameter two such that G is C_6-free. Let $a, b \in V(G)$ such that $ab \notin E(G)$ and let H be the graph induced on $\langle \{a, b\} \rangle$. Consider the partition $V(G) = V(H) \cup X \cup Y$, with every vertex of X having exactly one neighbor in $V(H)$ and every vertex of Y having no neighbor in H. Note, every vertex $y \in Y$ must have at least $|H| \geq 2$ distinct neighbors in X. This is because y has no neighbor in H, the diameter of G is 2, and every vertex in X has exactly one neighbor in H. That is, 2 distinct vertices in H cannot have the same common neighbor in X with y, as that common neighbor would then belong to H. We will use this fact later.

If $G[X]$ is connected, by Lemma 2.34 we have $A \cup \{x\}$ percolates for any $x \in X$.

If $G[X]$ is disconnected, by Lemma 2.34 Y is nonempty. Pick any vertex $y \in Y$ and consider $\langle \{a, b, y\} \rangle$. If $X \subset \langle \{a, b, y\} \rangle$ then since every vertex in Y has at least two distinct neighbors in X, we have $Y \subset \langle \{a, b, y\} \rangle$ and thus $\langle \{a, b, y\} \rangle = V(G)$. Since G has diameter two, there is at least one vertex in each component which has a neighbor in Y. For each i, since every vertex in X_i has a neighbor in H, and X_i is connected, if one vertex becomes infected in X_i, every vertex will become infected in X_i. In other words, if $V(X_i) \cap \langle \{a, b, y\} \rangle$ is nonempty, then $V(X_i) \subset \langle \{a, b, y\} \rangle$. From these facts, if $Y \subset \langle \{a, b, y\} \rangle$, then since every component X_i has a neighbor in Y, we have $X \subset \langle \{a, b, y\} \rangle$ and thus $\langle \{a, b, y\} \rangle = V(G)$. Note that this includes the special case $Y = \{y\}$.

We assume then that there are vertices in Y and components in X which are not contained in $\langle \{a, b, y\} \rangle$, in other words $Y' = Y - \langle \{a, b, y\} \rangle$ and $X' = X - \langle \{a, b, y\} \rangle$ are nonempty. Since G has diameter two, for every vertex $x \in X'$ there is a common neighbor $w \in Y'$ of x and y. That is to say, y has at least one neighbor in Y'.

What follows is illustrated in Figure 19. Since a and b are independent, they have a common neighbor $z \in V(H)$. Since the diameter of G is 2, there is a common neighbor $y' \in X \cap \langle \{a, b, y\} \rangle$ of z and y (since z and y are infected, so too is y').

Similarly, there is common neighbor $w' \in X'$ of w and b. If $w' \in \langle \{a, b, y\} \rangle$ then w would have become infected. So y' and w' are not in the same component. Now $C = zbw'wyy'z$ is a cycle of length 6. However since G is C_6-free, C must contain a chord. No chord can have w or w' as an endpoint, as otherwise they would be in $\langle \{a, b, y\} \rangle$. Of the remaining possibilities, no chord can have y as an endpoint as y is distance 2 from z and b, and lastly $y'b$ is not an edge, as then $y' \in \langle \{a, b\} \rangle$.

The proof is identical in the case that G is C_7-free but not C_6-free, except y' is adjacent to a instead of z. This is the reason why we pick independent vertices a and b at the start (otherwise, choosing a and b such that $ab \in E(G)$ is sufficient in proving C_6 case). □

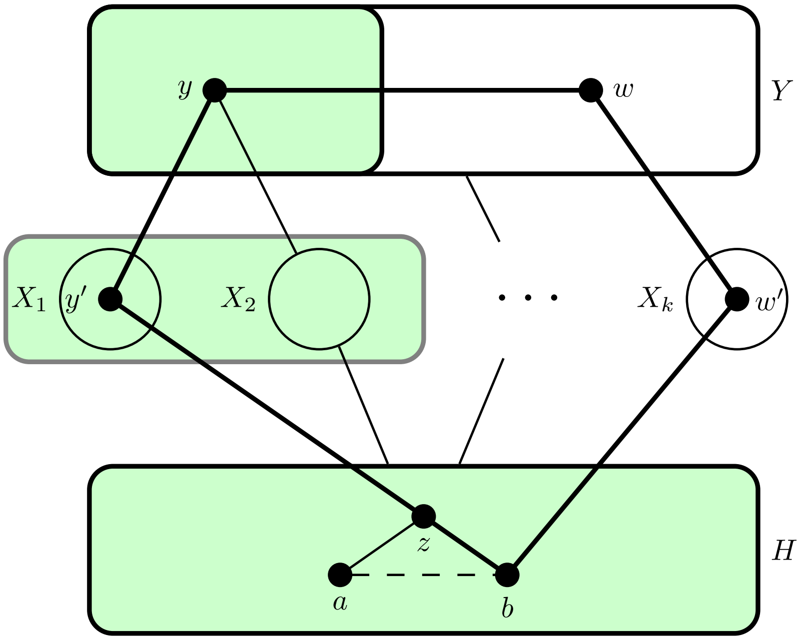


Fig. 19. A 2-connected, diameter two graph partitioned into three sets with a C_6 displayed.

The bound is sharp for the Möbius ladder on eight vertices, seen in Figure 6 (left). This proof method generalizes. We use induction on k to prove an upper bound on $m(G, 2)$ for C_k-free graphs in terms of k.

Corollary 2.37. *If G is 2-connected, has diameter two, and is C_k-free for any $5 \leq k \leq n$ then $m(G, 2) \leq \lceil (k-3)/2 \rceil + 1$.*

Proof. Let G be a 2-connected, diameter two graph with order n, and let k be an integer such that $5 \leq k \leq n$. For $k = 5$ we have $m(G, 2) \leq 2$ by Theorem 2.33, and for $k = 6$ and $k = 7$ Theorem 2.36 gives $m(G, 2) \leq 3$. We assume true up to $k = 7$ and consider $k \geq 8$. Assume G is C_k-free. If G is C_t-free for some $t \leq k - 1$ then by the induction hypothesis $m(G, 2) \leq \lceil (t-3)/2 \rceil + 1 \leq \lceil (k-3)/2 \rceil + 1$. So we may assume G contains every cycle C_t for $5 \leq t \leq k-1$ as an induced subgraph. In particular, G contains a copy of C_{k-3}, say C, as an induced subgraph. Let A be a minimum percolating set of C. In other words, A is a set of $\lceil (k-3)/2 \rceil$ vertices and $\langle A \rangle$ contains C. Let $H = G[\langle A \rangle]$ and let $a, b \in V(C)$ such that $d_C(a, b) = 2$. Consider the partition $V(G) = V(H) \cup X \cup Y$, with every vertex of X having exactly one neighbor in $V(H)$ and every vertex of Y having no neighbor in H. Note that every vertex in Y has at least $|H|$ neighbors in X.

If $G[X]$ is connected, by Lemma 2.34 we have $A \cup \{x\}$ percolates for any $x \in X$. So we may assume $G[X]$ is disconnected and Y is nonempty. Let $y \in Y$ and consider $A' = A \cup \{y\}$. If $X \subset \langle A' \rangle$ or $Y \subset \langle A' \rangle$ then $V(G) = \langle A' \rangle$ and we are done.

What follows is illustrated in Figure 20. Assume now that $Y' = Y - \langle A' \rangle$ and $X' = X - \langle A' \rangle$ are nonempty. Let $y' \in X \cap \langle A' \rangle$ be the common neighbor of y and a. Note that y has some neighbor $w \in Y'$. Let $w' \in X$ be the common neighbor of w and b.

Now if x is the common neighbor of a and b in C, then $C - x$ is an induced path of length $k - 4$ in G. Let $A'' = \{w, w', y, y'\} \cup V(C - x)$. Notice $|A''| = k$. We claim A'' induces a cycle of length k in G. Indeed, $C - x$ is an induced path, and there is no additional edge from any of $\{w, w', y, y'\}$ to $C - x$. For the remaining 3 possible edges, if $w' \sim y'$ or $w' \sim y$ then w' would have been infected. If $w \sim y'$ then w would have been infected. Thus A'' induces a cycle of length k, a contradiction. Thus $\langle A' \rangle = V(G)$.

□

Theorem 2.38. *If G is 2-connected, diameter two, and $2K_2$-free then $m(G, 2) \leq 3$.*

Proof. Let $a, b \in V(G)$ such that $ab \notin E(G)$ and consider $\langle \{a, b\} \rangle$. Let $V(G) = \langle \{a, b\} \rangle \cup X \cup Y$ be the partition with every vertex in X having exactly one neighbor

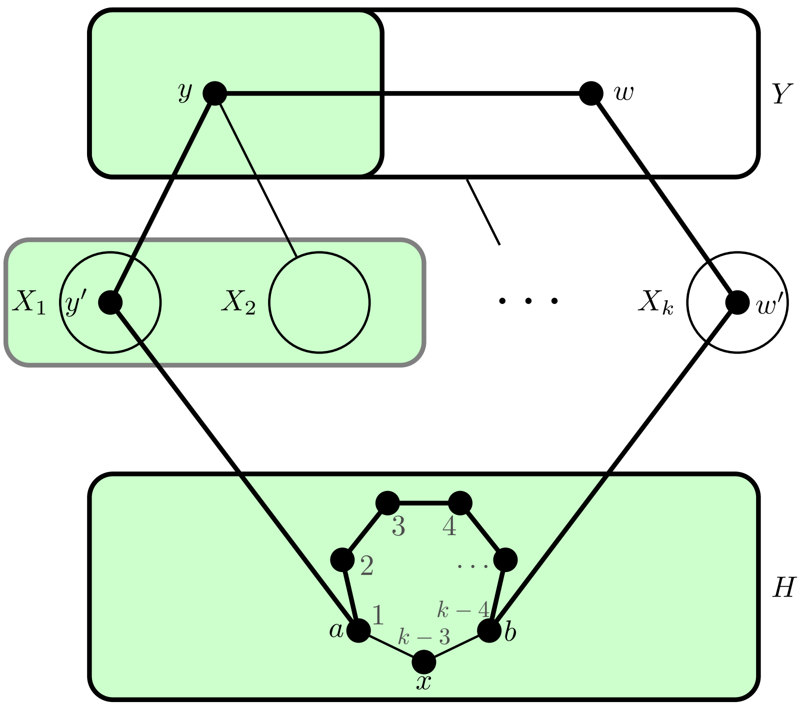


Fig. 20. An illustration of the setup in Corollary 2.37. There is an induced cycle of length $k-3$ and consequently an induced path between a and b of length $k-4$.

in $\langle\{a,b\}\rangle$ and every vertex in Y anticomplete to $\langle\{a,b\}\rangle$.

If $G[X]$ is connected, by Lemma 2.34 we have $A \cup \{x\}$ percolates for any $x \in X$. So we may assume $G[X]$ is disconnected and Y is nonempty. Let $y \in Y$ and consider $\langle\{a,b,y\}\rangle$. If either $X' = X - \langle\{a,b,y\}\rangle$ or $Y' = Y - \langle\{a,b,y\}\rangle$ are empty, then $V(G) \subseteq \langle\{a,b,y\}\rangle$. Thus both X' and Y' are nonempty.

The following is illustrated in Figure 21. Since G has diameter two, y must have at least one neighbor $w \in Y'$ which y uses to travel to X' in at most 2 steps. Now y has a common neighbor $y' \in X \cap \langle\{a,b,y\}\rangle$ with a and w has a common neighbor $w' \in X'$ with b. We claim the edges ay' and bw' induce a copy of $2K_2$. Indeed

1. $ab \notin E(G)$ by assumption.
2. by' and aw' are not in $E(G)$, as otherwise y' or w' would be in $\langle\{a,b\}\rangle$.
3. $y'w' \notin E(G)$, as then w would be in $\langle\{a,b,y\}\rangle$.

Thus, it cannot be the case that X' and Y' are both nonempty. That is, $\langle\{a, b, y\}\rangle = V(G)$. $\square$

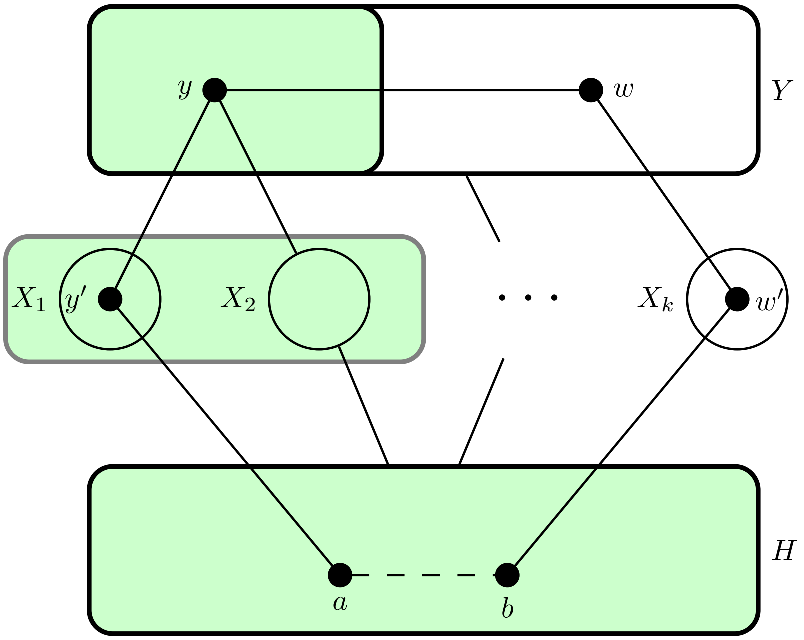


Fig. 21. An illustration of the setup in Theorem 2.38. Here the edges ay' and bw' induce a $2K_2$.

The bound in Theorem 2.38 is sharp for two infinite family of graphs described at the end of Section 2.3 and illustrated in Figure 15. These are the triangle-free, 2-connected graphs with diameter two and two vertex-disjoint 2-forbidden subgraphs.

CHAPTER 3

GRAPHS WITH INDEPENDENCE NUMBER TWO

Let G be a finite, simple graph. We recall a few definitions. A set $I \subseteq V(G)$ is *independent* if no two vertices in I are adjacent. The *independence number* of G, denoted $\alpha(G)$, is the maximum size of an independent set. Figure 22 shows three examples of graphs with independent sets. The graphs with independence number one are the complete graphs. What can one say then about the graphs with independence number two? Let $\boldsymbol{\alpha}_2$ be the class of graphs with independence number at most two. The complement of any graph $G \in \boldsymbol{\alpha}_2$ is triangle-free and vice versa. This suggests that, while graphs in $\boldsymbol{\alpha}_2$ are "one step up" from cliques, their structure is diverse as with the well-studied triangle-free graphs.

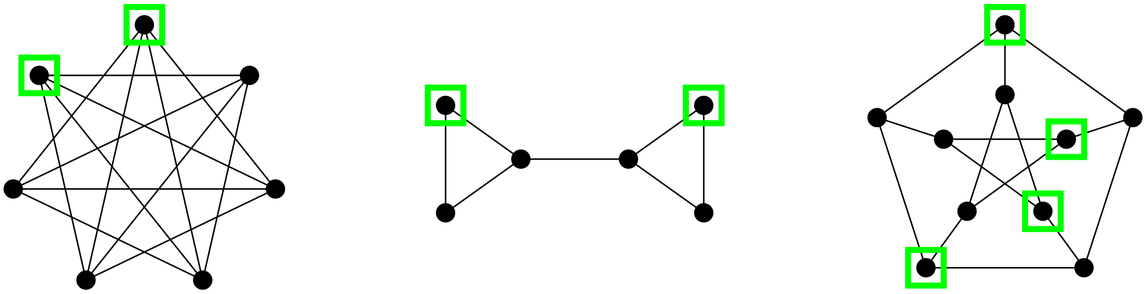


Fig. 22. Three graphs with maximum independent sets highlighted with green squares. On the left is $\overline{C_7}$, and on the right is the Petersen graph.

Definition 3.1 (α-critical)**.** An edge $e \in E(G)$ is *critical* if $\alpha(G - e) = \alpha(G) + 1$. A graph G is *α-critical* if G has independence number α and every edge $e \in E(G)$ is critical.

An example of a graph in $\boldsymbol{\alpha}_2$ which is α-critical is C_5. The removal of any edge in C_5 yields P_5, which has independence number three. We say a graph G is *triangle-saturated* (or maximally triangle-free) if G is triangle-free and for any $e \in E(\overline{G})$ the graph $G + e$ contains a triangle. Graphs in $\boldsymbol{\alpha}_2$ which are α-critical are exactly the complements of triangle-saturated graphs.

3.1 Motivation – Hadwiger's Conjecture

Recall that the chromatic number of a graph G, denoted $\chi(G)$, is the minimum number of colors needed to color the vertices of G so that no two adjacent vertices share the same color. One observation to be made is that $\chi(G) \geq \omega(G)$, where $\omega(G)$ is the maximum size of a clique in G. However, this lower bound is not always sharp; the gap between $\chi(G)$ and $\omega(G)$ can be arbitrarily large. One method of showing this is due to Mycielski, who showed it is possible to construct graphs which are triangle-free (clique number two) with arbitrarily large chromatic number [88].

Definition 3.2. Let G and H be graphs, $|H| \leq |G|$, and let $V(H) = \{v_1, v_2, \ldots, v_k\}$. We say G contains H as a minor if there are disjoint subsets $V_1, V_2, \ldots, V_k$ of $V(G)$ satisfying the following:

1. $G[V_i]$ is connected.
2. For every $v_iv_j \in E(H)$, there is an edge between V_i and V_j.

Alternatively, one can say H can be obtained from G by a sequence of edge contractions and vertex deletions.

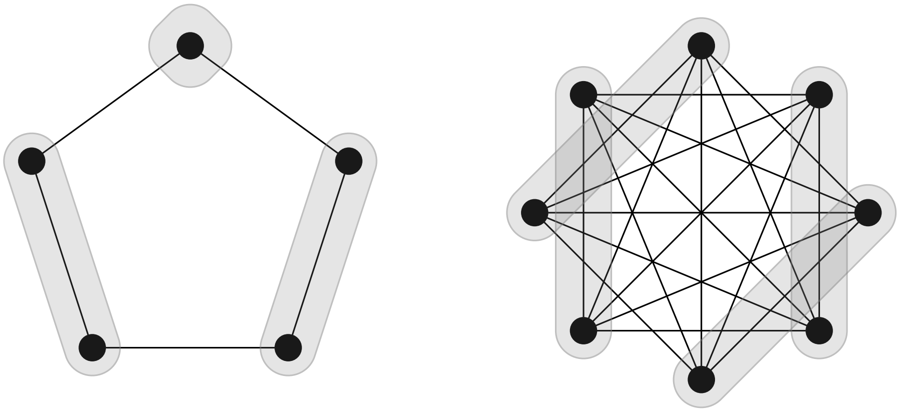


Fig. 23. Left: A K_3 minor in C_5. Right: A K_4 minor in $\overline{C_8}$.

In particular, we are interested in K_t minors. We let $H \preceq G$ denote "G contains H as a minor", or "G is *contractible* to H".

Conjecture 3.3 (Hadwiger [60])**.** *If $\chi(G) \geq t$ then $K_t \preceq G$.*

We say a graph is HW if the graph satisfies Conjecture 3.3. Hadwiger's Conjecture gives a structural justification for when $\chi(G) > \omega(G)$. Graphs in $\boldsymbol{\alpha}_2$ are of particular interest for Hadwiger's Conjecture, and this special case was first considered by Duchet and Meyniel [47]. There are many results for particular classes of graphs, however, Hadwiger's conjecture is not known to be true for all graphs in α_2.

Since no two adjacent vertices are assigned the same color in a proper coloring of a graph, the sets of vertices with the same color, or color classes, are also independent sets. Thus if $G \in \boldsymbol{\alpha}_2$ and $n = |V(G)|$, then $\chi(G) \geq n/2$. Conjecture 3.3 then implies that for a graph $G \in \boldsymbol{\alpha}_2$, we have $K_t \preceq G$ for $t \geq \lceil n/2 \rceil$. There is, however, an equivalent statement of Hadwiger's Conjecture for graphs in $\boldsymbol{\alpha}_2$.

Conjecture 3.4. *If $G \in \boldsymbol{\alpha}_2$ then G contains a $K_{\lceil |V(G)|/2 \rceil}$ minor.*

Plummer, Stiebitz, and Toft [93] show that Conjecture 3.4 implies Conjecture 3.3 for graphs in $\boldsymbol{\alpha}_2$. That is, to show that G is HW, one need only exhibit a complete minor of size exactly $\lceil n/2 \rceil$, rather than a complete minor of size $\chi(G) \geq \lceil n/2 \rceil$.

Observation 3.5. *Fix $t \geq 1$. If for every α-critical graph G with $\alpha(G) = t$ we have that G is HW, then every graph with independence number t is HW.*

Observation 3.5 follows from the fact that every graph with independence number α has a spanning α-critical subgraph, which can be obtained by removing non-critical edges iteratively.

Question 3.6 (Seymour)**.** *Let $G \in \alpha_2$. Does there exist $c > 1/3$ so that $K_{\lceil cn \rceil} \preceq G$?*

For a surveys on Hadwiger's Conjecture, see Seymour [105] and Cameron and Vušković [26].

3.2 Some General Structure

We collect several known and basic structural facts.

Lemma 3.7. *Let* $G \in \boldsymbol{\alpha}_2$ *and assume* G *has diameter two. Then* G *contains at least* $\lfloor n/2 \rfloor$ *vertices of degree at least* $\lfloor n/2 \rfloor$ *when* n *is odd and* $\delta = \lfloor n/2 \rfloor$*, or otherwise contains at least* $\lceil n/2 \rceil$ *vertices of degree at least* $\lceil n/2 \rceil$*.*

Proof. Let $G \in \boldsymbol{\alpha}_2$. Assume that G has diameter two. Let v be a vertex of minimum degree δ. If $\delta \geq \lceil n/2 \rceil$, we are done. Assume now $\delta < \lceil n/2 \rceil$. Let N be the δ vertices adjacent to v and C be the vertices at distance 2 from v. Because $\alpha(G) = 2$, the vertices in C form a clique, and $|C| = n - \delta - 1$. Since G has diameter two, every vertex in C has at least one neighbor in N. Let $x \in C$. If n is even, or if n is odd with $\delta < \lfloor n/2 \rfloor$, then

$$\begin{aligned} d(x) &\geq (|C| - 1) + 1 \\ &= n - \delta - 1 \\ &\geq n - \left(\left\lfloor \frac{n}{2} \right\rfloor - 1 \right) - 1 \\ &= n - \left\lfloor \frac{n}{2} \right\rfloor \\ &= \left\lceil \frac{n}{2} \right\rceil \end{aligned}$$

Otherwise, if n is odd and $\delta = \lfloor n/2 \rfloor$, then $|C| = \lfloor n/2 \rfloor$ and $d(x) \geq \lfloor n/2 \rfloor$ for each $x \in C$. □

We recall a well-known theorem of Brooks.

Theorem 3.8 (Brooks [22])**.** *Let* G *be a connected graph with maximum degree* Δ*. If* G *is a complete graph or odd cycle, then* $\chi(G) = \Delta + 1$*. Otherwise,* $\chi(G) \leq \Delta$*.*

Proposition 3.9. *If* G *is a connected graph with* $\alpha(G) = 2$ *and* $\Delta(G) < n/2$*, then* $G \cong C_5$*.*

Proposition 3.9 is a consequence of Brooks' Theorem.

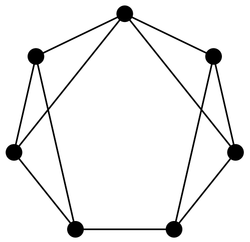


Fig. 24. A graph of order seven in $\boldsymbol{\alpha}_2$. Six vertices have degree $3 = \lfloor 7/2 \rfloor$ and exactly one with degree $4 = \lceil 7/2 \rceil$. The graph has exactly $\lceil n/2 \rceil$ vertices of degree exactly $\lceil n/2 \rceil$.

Proof 1. (Using Brooks' Theorem). Note that for any graph G, the inequality

$$\chi(G) \geq n/\alpha(G)$$

holds. Let G be an order n graph with $\alpha(G) = 2$ and $\Delta(G) < n/2$. Then $\chi(G) \geq n/2$. If G is not a complete graph or odd cycle, then by Theorem 3.8, we have $\Delta(G) \geq \chi(G) \geq n/2$. This contradicts our assumption that $\Delta(G) < n/2$.

Otherwise, if G is a complete graph or odd cycle, G must be isomorphic to C_5. $\square$

We provide another proof, without the use of Brooks' Theorem.

Proof 2. If G has diameter 3, then G is two cliques A, B joined by at least one edge xy. Let $|A| \geq |B|$ and $x \in A$. Then $d(x) \geq n/2$. So G must have diameter two. For any vertex v, we can partition G into $\{v\}$, the neighbors N of v, and a clique C of size $n - d(v) - 1$. Since the diameter of G is 2, every vertex in C has at least one neighbor in N, or degree at least $|C|$. As $d(v) \leq \Delta < n/2$, we have $|C| \geq \lceil n/2 \rceil - 1$. Furthermore, as every vertex $w \in C$ must also have degree at most Δ, we have $\lfloor n/2 \rfloor \geq d(w) \geq |C|$. Then

$$\lfloor n/2 \rfloor \leq |C| \leq \lceil n/2 \rceil - 1,$$

i.e. $|C| = \lfloor n/2 \rfloor$. Then $|N| = \lfloor n/2 \rfloor$ for any v, and thus G is regular of degree $\lfloor n/2 \rfloor$. As $\lfloor n/2 \rfloor = n/2$ for even n, it must be that n is odd. Note since the sum of degrees must be even and G is regular, n must be of the form $4k + 1$ for $k = 1, 2, \ldots$ so that that n is odd and $\lfloor n/2 \rfloor$ is even. So $|C| = |N| = 2k$. We will show that

$|N| \leq 2$. If $w \in C$, then $d(w) = \lfloor n/2 \rfloor$ and C being a clique implies w has exactly one neighbor in N. As G is regular of degree $\lfloor n/2 \rfloor$, it cannot be the case that some vertex in N dominates C. Thus there exist $x, x' \in N$ and $w, w' \in C$ so that xw and $x'w'$ are independent edges in G. Since $\alpha(G) = 2$, $N - x$ and $N - x'$ must induce cliques. Otherwise, a non-adjacent pair in $N - x$ forms an independent set with w, and similarly with $N - x'$ and w'. It follows that x and x' are independent with exactly one neighbor in C, as they each have $\lfloor n/2 \rfloor - 2$ neighbors in N, v as a neighbor, and at least one neighbor in C. If $|N| = 2$ then we have C_5. Assume $|N| \geq 4$. Then $N' = N - \{x, x'\}$ is nonempty. Let $y \in N'$. Since $N - x$ and $N - x'$ induce cliques, y must be adjacent to every vertex in N. It follows that y has no neighbors in C, otherwise $d(y) > \lfloor n/2 \rfloor$. Since $|C| = |N| \geq 4$ and both x and x' have unique neighbors in C, there are vertices in C with no neighbors in N, a contradiction. So $|N| = 2$ and C_5 is the unique graph with independence number 2 and maximum degree less than $n/2$. □

Definition 3.10. Let G be a graph. A *clique cover* of G is a partition of $V(G)$ so that each part induces a clique in G. The *clique cover number* of G, denoted $\overline{\chi}(G)$, is the minimum cardinality of a clique cover.

Proposition 3.11. *Let G be a graph with independence number α at most 2. If the diameter of G is 3, then $\overline{\chi}(G) = 2$.*

Proof. Let $u, v \in V(G)$ so that $d(u, v) = 3$. Suppose there exists $x \in V(G) - N[u] \cup N[v]$. Then $\{x, u, v\}$ is an independent set of size 3, a contradiction. Thus for all $x \in V(G)$, we have $x \in N[u]$ or $x \in N[v]$, however $x \notin N[u] \cap N[v]$, as then $d(u, v) = 2$. So $V(G) = N[u] \cup N[v]$ and $N[u] \cap N[v] = \emptyset$, that is the neighborhoods form a partition.

Suppose $N[u]$ does not induce a clique. Then there are independent vertices $x, x' \in N[u]$. Then $\{x, x', v\}$ forms an independent set of size 3, a contradiction. So $N[u]$ forms a clique. The argument is identical for $N[v]$. □

Observation 3.12. *If $G \in \boldsymbol{\alpha}_2$ and $\overline{\chi}(G) = 2$, then G has a dominating edge.*

Recall, any graph $G \in \boldsymbol{\alpha}_2$ is the complement of a triangle-free graph. That is, graphs in $\boldsymbol{\alpha}_2$ are already characterized by forbidding the empty graph $\overline{K_3}$ as an induced subgraph. The claw and the bull are illustrated in Figure 25.

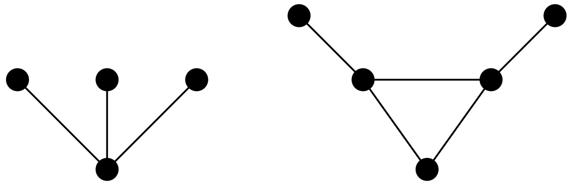


Fig. 25. The claw graph (left) and bull graph (right)

The following proposition is implied by Theorem 1.1 in [78]. We provide an independent proof here.

Proposition 3.13. *If G is claw-free, bull-free, and has diameter two, then $\alpha(G) = 2$.*

Proof. Let G be claw-free, bull-free, and $\operatorname{diam}(G) = 2$. Suppose toward contradiction that $\alpha(G) > 2$, and let $X = \{x_1, x_2, x_3\}$ be an independent set of size 3 in G. Since $\operatorname{diam}(G) = 2$ and X is an independent set, each pair of vertices (x_i, x_j) has a common neighbor y_{ij} in $G - X$. Since G is claw-free, the y_{ij} are distinct. Let $Y = \{y_{12}, y_{13}, y_{23}\}$. Notice, since G is bull-free, Y is also an independent set. Since Y is an independent set, $x_1y_{23}, x_2y_{13}, x_3y_{12}$ are not edges in G, as otherwise there would exist a claw induced by $\{x_i\} \cup Y$. Thus $H = G[X \cup Y] \cong C_6$.

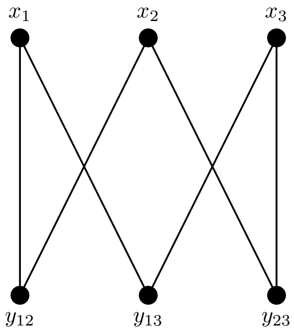


Fig. 26. Proof of Proposition 3.13: The independent set X and the common neighbors of pairs form an induced C_6.

Since $\operatorname{diam}(G) = 2$, the pairs $(x_1, y_{13}), (x_2, y_{12})$, and (x_3, y_{23}) each have common neighbors in $G - H$. Let Z be the set of these common neighbors, and let $H' =$

$G[X \cup Y \cup Z]$. Since $H \cong C_6$, the only other possible edges in H' have at least one endpoint in Z. Consider the following three cases.

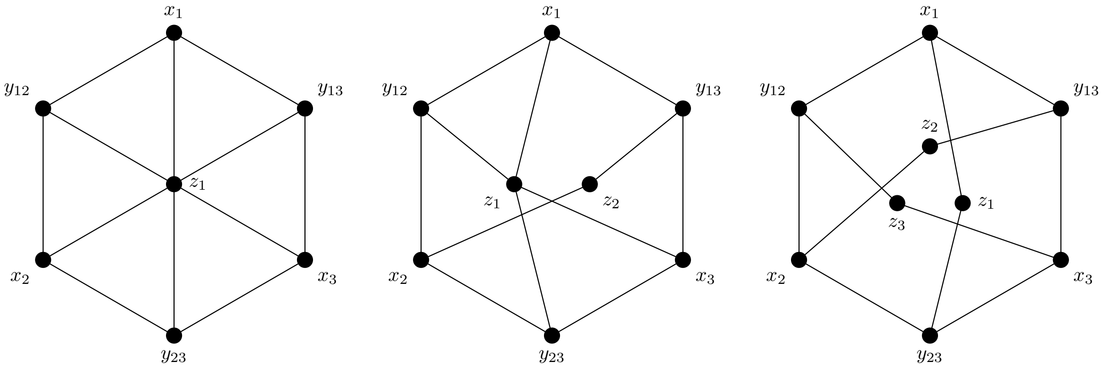


Fig. 27. Proof of Proposition 3.13: From left to right, Case 1, Case 2, and Case 3.

Case 1. $Z = \{z_1\}$. For $z_1 \in Z$, the sets $z_1 \cup X$ and $z_1 \cup Y$ each induce a claw. This is a contradiction, since G is claw-free.

Case 2. $Z = \{z_1, z_2\}$ where without loss of generality, z_1 is the common neighbor of (x_1, y_{23}) and (x_3, y_{12}), and z_2 is the common neighbor of (x_2, y_{13}). Consider $H - z_2$. If z_1 is adjacent to y_{13} or x_2, then H would contain an induced claw as z_1 would dominate either X or Y, a contradiction. So z_1 is independent from y_{13} and x_2. However, then there are bulls induced by $\{x_1, x_2, y_{12}, y_{13}, z_1\}$ and $\{x_2, x_3, y_{13}, y_{23}, z_1\}$, a contradiction.

Case 3. $Z = \{z_1, z_2, z_3\}$ where z_i is the common neighbor of (x_i, y_{jk}) where $j, k \in \{1, 2, 3\} - \{i\}$ with $j \neq k$. Let M_i be the set of edges with one endpoint being z_i and one endpoint in H. Note, $2 \leq |M_i| \leq 6$ for each i.

Case 3a. There exists an M_i such that $|M_i| = 2$. Without loss of generality, we assume $i = 1$. Then there are claws induced by $\{x_3, y_{23}, y_{13}, z_1\}$ and $\{x_1, x_2, y_{12}, z_1\}$.

Case 3b. There exists an M_i such that $|M_i| = 3$. Without loss of generality, we assume $i = 1$. So $M_1 = \{z_1y_{12}, z_1x_3, z_1w\}$ for some $w \in H$. Thus, there is an induced bull whose triangle is formed by either $\{z_1, y_{12}, w\}$ or $\{z_1, x_3, w\}$, and whose leaves are diametrically opposed on H.

Case 3c. If $|M_i| \geq 5$ for any i, then by the pigeonhole principle three of the five neighbors of z_i in H form an independent set, and thus along with z_i an induced claw.

Case 3d. The case of $|M_i| = 4$ for all i remains. If there is a z_i whose neighbors in H induce $2K_2$, then we are done by Case 2. If there is a z_i whose neighbors in H induce a disjoint union of P_3 and K_1, then the endpoints of the P_3, the vertex of K_1, and z_i induce a claw. We assume then that the neighbors of each z_i in H induce a P_4. Let v_1, v_2, v_3, v_4 be the neighbors of z_1 in H in order as they appear in the cycle (i.e. $\{v_1, v_3\} \subset X$ and $\{v_2, v_4\} \subset Y$ or vice versa). Let $v_1' \neq v_2$ be the other neighbor of v_1. Then $\{v_1, v_1', z_i, v_2, v_4\}$ induces a bull.

We've now shown that if G is claw-free, bull-free, and diameter two, it cannot have an independent set of size 3. This concludes the proof. $\square$

In Proposition 3.13, notice that the diameter assumption is required; any sufficiently large path or cycle (or graph containing these as an induced sub graph) is a counterexample to the claim that a claw-free and bull-free graph of unrestricted diameter has independence number equal to 2.

3.3 The Size of Graphs With Independence Number Two

Because independence number two and triangle-free graphs are complements, the α-critical graphs in $\boldsymbol{\alpha}_2$ are the complements of triangle-saturated graphs. If the removal of any edge in G results in an independent set of size 3, then the addition of any edge in $\overline{G}$ results in a triangle. The size of triangle-free graphs, and more generally K_t-free graphs, have been studied extensively, most notably with Mantel's Theorem and Turán's Theorem [77, 115].

Theorem 3.14 (Mantel [77])**.** *If a graph G is triangle-free, then the size of G is at most $\lfloor n^2/4 \rfloor$ with equality if and only if $G = K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.*

While this upper bound is sharp, the extremal graph is unique. What about the size of the other triangle-free graphs? Barefoot et al. [10] answer this for triangle-saturated graphs, generalizing Mantel's Theorem.

Theorem 3.15 (Barefoot et al. [10]). *If G is a triangle-saturated graph, then the size of G is either $(n-k)k$ for some integer k, or at most $\lfloor (n-1)^2/4 \rfloor + 1$.*

Theorem 3.15 then implies the following for α-critical graphs in $\boldsymbol{\alpha}_2$.

Corollary 3.16. *If G is an α-critical graph in $\boldsymbol{\alpha}_2$, then the size of G is at least $\lfloor n^2/4 \rfloor - 1$.*

This comes from the fact that an upper bound on the size of a graph gives a lower bound on the size of its complement. In particular, here the lower bound is

$$\binom{n}{2} - \left\lfloor \frac{(n-1)^2}{4} \right\rfloor - 1.$$

If n is odd, then $(n-1)^2$ is a multiple of 4 and the floor function can be dropped. If n is even, then the quantity $(n-1)^2/4 = n^2/4 - n/2 + 1/4$ has integer part $n^2/4 - n/2$ and fractional part $1/4$. In other words, in this case $\lfloor (n-1)^2/4 \rfloor = n(n-2)/4$. Thus,

$$\binom{n}{2} - \left\lfloor \frac{(n-1)^2}{4} \right\rfloor - 1 = \begin{cases} (n^2-1)/4 - 1 & \text{if } n \text{ odd} \\ n^2/4 - 1 & \text{if } n \text{ even} \end{cases},$$

or in other words, $\binom{n}{2} - (\lfloor (n-1)^2/4 \rfloor + 1) = \lfloor n^2/4 \rfloor - 1$.

Using the `AUTOMATED CONJECTURING PROGRAM` (see [72] and [73]) we acquire a similar bound in the form of Conjecture 3.17. While this similar bound is only as good as that of Theorem 3.15 for every graph in $\boldsymbol{\alpha}_2$, the independent proof may be insightful. Let $\text{Even}_G(v)$ (resp. $\text{Odd}_G(v)$) be the number of vertices of even (resp. odd) distance from a vertex v in G. We omit the subscript if G is clear from context.

Conjecture 3.17. *If G is a connected graph with independence number two, then $|E(G)| \geq \max_{v \in V(G)} |\text{Even}(v)| \cdot |\text{Odd}(v)| - 1$.*

If G has diameter two and order n, then for every vertex $v \in V(G)$, the vertices of odd distance from v are precisely the neighbors of v. More generally, if v is a vertex of eccentricity two in G then $|\mathrm{Even}(v)| \cdot |\mathrm{Odd}(v, G)| = (n - d(v))d(v)$. This quantity is maximized for vertex degrees which are as close to $n/2$ as possible, with a maximum possible value of $\lfloor n^2/4 \rfloor$.

Theorem 3.18. *If $G \in \boldsymbol{\alpha}_2$ with diameter two and minimum degree δ, then the size of G is at least $\delta(n - \delta) - 1$.*

Proof. Let G be a graph with $\alpha(G) = 2$ and $\mathrm{diam}(G) = 2$.

Case 1. Assume every vertex has degree at least $\lfloor n/2 \rfloor$. If n is even, then by the well-known Handshaking Lemma,

$$|E(G)| \geq \frac{n\delta}{2} \geq \frac{n\left(\frac{n}{2}\right)}{2} = \frac{n^2}{4}.$$

If n is odd, then $\lfloor n/2 \rfloor = (n-1)/2$.

Case 2. We assume $\delta(G) \leq \lfloor n/2 \rfloor - 1$. Since G has diameter two, for any vertex v one can partition $V(G)$ into $\{v\}$, the neighbors N of v, and the remaining vertices C. Note that C is a clique, as otherwise two independent vertices with v would make an independent set of size 3. Additionally, every vertex in C has at least one neighbor in N, and so the degree of any vertex in C is at least $|C| \geq \lceil n/2 \rceil \geq \delta + 1$. If $G \in \boldsymbol{\alpha}_2$ and $|V(G)| = 3$, then $G \cong P_3$ and

$$|E(G)| = 2 = \delta(G)(n - \delta(G)) - 1.$$

We now assume $|E(G)| \geq \delta(G)(n - \delta(G)) - 1$ is true for graphs of order smaller than n and consider a graph $G \in \boldsymbol{\alpha}_2$ with $|V(G)| > n$. Let $\delta(G) = \delta$.
Let v be a vertex of minimum degree δ. There is a partition of N into sets A, A' and B, where:

- A is the set of δ-degree vertices with no neighbor in C.
- A' is the set of δ-degree vertices with at least one neighbor in C.
- B is the set of vertices with degree at least $\delta + 1$

Since $|N| = \delta$, every vertex in A dominates N. Furthermore, each vertex in B either has exactly one neighbor in C (and thus dominates N with its remaining

$\delta-1$ possible neighbors), or has at least two neighbors in C. Each vertex in A' must have a non-neighbor in C, as otherwise it would have degree larger than δ. Let $w \in C$. Then w has $|C|-1 \geq \delta$ vertices in C as neighbors. In addition w has at least one neighbor in N. Let $k = |N(w) \cap N|$. Then $d(w) \geq \delta + k$. Let $k' = |N(w) \cap A'|$. We add edges between each vertex in $N(w) \cap A'$ and its non-neighbor in C, and remove w to obtain a new graph G' with $|E(G)| = |E(G')| + \delta + k - k'$. Notice that $\delta(G') = \delta$ as every δ-degree neighbor of w has had an edge added to it. Every vertex in N in G' either dominates N or has at least one neighbor in $C - w$, and every vertex in $C - w$ still has at least one neighbor in N, so the diameter of G' is 2. We've removed a vertex and added edges, so $\alpha(G') \leq 2$. Then

$$\begin{aligned}
|E(G)| &= |E(G')| + \delta + k - k' \\
&\geq (n-1-\delta)\delta - 1 + \delta + k - k' \\
&= (n-\delta)\delta - 1 - \delta + \delta + k - k' \\
&\geq (n-\delta)\delta - 1,
\end{aligned}$$

where the first inequality follows from the induction hypothesis, and the second inequality is because $k \geq k'$.

□

Theorem 3.19. *If $G \in \boldsymbol{\alpha}_2$ with order n and G has diameter two, then the size of G is at least $\max_{v \in V(G)} d(v)(n - d(v)) - 1$.*

Proof. The proof will follow a case analysis based on the clique cover number $\overline{\chi}(G)$ (see Definition 3.10.)

Case 1. Assume $\overline{\chi}(G) = 2$. Since $\overline{\chi}(G) = 2$ there is a partition $V(G) = A \cup B$ where A and B induce cliques. Let $|A| = a$ and $|B| = b$ and assume without loss of generality that $a \geq b$. First assume n is even. For some $k \geq 0$ we let $a = n/2 + k$ and $b = n/2 - k$. There are at least $n/2 - k$ edges with exactly one endpoint in A and an endpoint in B, as otherwise there is some vertex $y \in B$ with no neighbor in A and some vertex $x \in A$ with no neighbor in B and $d(x, y) = 3$, a contradiction.

Thus

$$\begin{aligned}|E(G)| &\geq \binom{n/2+k}{2} + \binom{n/2-k}{2} + \frac{n}{2} - k \\ &= \frac{n^2}{4} + k^2 - k \\ &\geq \frac{n^2}{4}.\end{aligned}$$

In the case that n is odd, $a = (n+1)/2 + k$, $b = (n-1)/2 - k$, and

$$\begin{aligned}|E(G)| &\geq \binom{(n+1)/2+k}{2} + \binom{(n-1)/2-k}{2} + \frac{n+1}{2} - k \\ &= \frac{n^2}{4} + k^2 + \frac{3}{4} \\ &\geq \frac{n^2}{4} + \frac{3}{4}.\end{aligned}$$

Hence, $|E(G)| \geq \lceil n^2/4 \rceil$. In both cases we have $|E(G)| \geq \lfloor n^2/4 \rfloor \geq \max_{v \in V(G)} d(v)(n - d(v))$.

Case 2. Assume $\overline{\chi}(G) = 3$. By Lemma 3.7 there are at least $\lfloor n/2 \rfloor$ vertices of degree at least $\lfloor n/2 \rfloor$ in G. Let $x \in V(G)$ such that $d(x) \geq \lfloor n/2 \rfloor$ and let $G' = G - x$. If $\overline{\chi}(G') = 3$ for every such x, then we move on to Case 3. Otherwise, we pick x so that $\overline{\chi}(G') = 2$.

Let $A \cup B$ be a partition of $V(G')$, where A and B each induce a clique in G'. Assume without loss of generality $|A| \geq |B|$. Since $\alpha(G) = 2$ the set $C = V(G) - N[x]$ is a clique in G. There is a partition of $E(G)$ into four sets, namely $E(G[A])$, $E(G[B])$, the set $E(A, B)$ of edges with one endpoint in A and one in B, and the set X of edges with x as an endpoint. Let $d(x) = \lfloor n/2 \rfloor + k$ for some $k \geq 0$. Note then that

$$|C| = n - (\lfloor n/2 \rfloor + k) - 1 = \lceil n/2 \rceil - k - 1.$$

Since C is partitioned by A and B, and C is a clique, we have $|E(A, B)| \geq |C \cap A| \cdot |C \cap B|$. As $|C \cap A| + |C \cap B| = |C|$, the minimum of this product occurs when $|C \cap A| = 1$ or $|C \cap B| = 1$. So

$$|E(A, B)| \geq \lceil n/2 \rceil - k - 2.$$

Since $|A| + |B| = n - 1$, we let $|A| = \lceil (n-1)/2 \rceil + \ell$ so that $|B| = \lfloor (n-1)/2 \rfloor - \ell$ for some $\ell \geq 0$. We have

$$\begin{aligned} |E(G)| &= |E(G[A])| + |E(G[B])| + |E(A,B)| + |X| \\ &\geq \binom{\lceil (n-1)/2 \rceil + \ell}{2} + \binom{\lfloor (n-1)/2 \rfloor - \ell}{2} + \left\lceil \frac{n}{2} \right\rceil - k - 2 + \left\lfloor \frac{n}{2} \right\rfloor + k \\ &= \binom{\lceil (n-1)/2 \rceil + \ell}{2} + \binom{\lfloor (n-1)/2 \rfloor - \ell}{2} + n - 2 \end{aligned}$$

Putting all this together, if n is odd then

$$\begin{aligned} |E(G)| &\geq \binom{(n-1)/2 + \ell}{2} + \binom{(n-1)/2 - \ell}{2} + n - 2 \\ &= \frac{n^2}{4} - \frac{5}{4} + \ell^2 \\ &\geq \left\lfloor \frac{n^2}{4} \right\rfloor - 1 + \ell^2. \qquad (n \not\equiv 0 \pmod 4) \end{aligned}$$

Note that since n is odd, $n \equiv 1, 3 \pmod 4$. Thus $n^2/4$ has remainder $1/4$ or $3/4$, giving the last inequality. If n is even then

$$\begin{aligned} |E(G)| &\geq \binom{n/2 + \ell}{2} + \binom{(n-2)/2 - \ell}{2} + n - 2 \\ &= \frac{n^2}{4} - 1 + \ell^2 + \ell \\ &\geq \left\lfloor \frac{n^2}{4} \right\rfloor - 1. \end{aligned}$$

Similar to Case 1, this proves the bound.

Case 3. Assume $\overline{\chi}(G) \geq 3$ and for any vertex x with $d(x) \geq \lfloor n/2 \rfloor$, we have $\overline{\chi}(G - x) \geq 3$. Note that if $\overline{\chi}(G) \geq 4$, then $\overline{\chi}(G - x) \geq 3$ for any $x \in V(G)$. Since for any $G \in \boldsymbol{\alpha}_2$ we have $\mathrm{diam}(G) = 3$ if and only if $\overline{\chi}(G) = 2$, it must be that $\overline{\chi}(G) \geq 3$ implies $\mathrm{diam}(G) = 2$. This allows us to use induction on n because the diameter is preserved via vertex removal.

The claim is true for $n \leq 6$, and verifiable via brute force. Let $S \subseteq V(G)$ such that for any $s \in S$, we have $(n - d(s))d(s) = \max_{v \in V(G)} (n - d(v))d(v)$. Let $s \in S$. If s is dominating in G then $(n - d(s))d(s) - 1 = n - 2$ and we are done. So we may assume $\Delta = \Delta(G) < n-1$.

Case 3a. Assume $d_G(s) \neq \Delta$. Then there is a vertex $x \in V(G) - \{s\}$ such that $d_G(x) = \Delta$. Since $d_G(s) < n-1$, there exists $w \in V(G)$ such that w is not adjacent to s. If $xs \in E(G)$ then $x \neq w$ and we let $G' = G - x + ws$. Otherwise let $G' = G - x$. If $xs \notin E(G)$ it may be the case that $x = w$, however in both $G - x$ and $G - x + ws$, we have $d_{G'}(s) = d_G(s)$. Define $\mathbb{1}_{xs}$ to be 1 if $xs \in E(G)$ and 0 otherwise. Then

$$\begin{aligned}
|E(G)| &= |E(G')| + d_G(x) - \mathbb{1}_{xs} \\
&\geq \max_{v \in V(G')} (n(G') - d_{G'}(v))d_{G'}(v) - 1 + \Delta - \mathbb{1}_{xs} \\
&\geq (n(G') - d_{G'}(s))d_{G'}(s) - 1 + \Delta - \mathbb{1}_{xs} \\
&= (n - 1 - d_{G'}(s))d_{G'}(s) - 1 + \Delta - \mathbb{1}_{xs} \\
&= (n - d_{G'}(s))d_{G'}(s) - d_{G'}(s) - 1 + \Delta - \mathbb{1}_{xs} \\
&= (n - d_G(s))d_G(s) - d_G(s) - 1 + \Delta - \mathbb{1}_{xs} && (d_G(s) = d_{G'}(s)) \\
&= \max_{v \in V(G)} (n - d_G(v))d_G(v) - d_G(s) - 1 + \Delta - \mathbb{1}_{xs} \\
&\geq \max_{v \in V(G)} (n - d_G(v))d_G(v) - 1 && (d_G(s) \leq \Delta - 1)
\end{aligned}$$

Case 3b. Assume $d_G(s) = \Delta$. By Lemma 3.7, we have $\Delta \geq \lfloor n/2 \rfloor$ and G has at least $\lfloor n/2 \rfloor$ vertices of degree at least $\lfloor n/2 \rfloor$. Let $u \in V(G)$ so that

$$\lfloor n/2 \rfloor \leq d_G(u) \leq \Delta.$$

We claim any such vertex u must be in S. If $\Delta = \lceil n/2 \rceil$, then u must have degree Δ or $\Delta - 1$, which implies $u \in S$. Otherwise, assume $\Delta > \lceil n/2 \rceil$. If $\Delta > \lceil n/2 \rceil$ and $d(u) \neq \Delta$, then $(n - d(u))d(u) > (n - d(s))d(s)$ as $d(u)$ would be closer to $n/2$ than $d(s) = \Delta$. This implies again, $d(u) = \Delta$ and $u \in S$. Consequently, $|S| \geq \lfloor n/2 \rfloor$, which is at least 2 for $n \geq 4$.

Let $s' \in S - \{s\}$. Without loss of generality assume $d(s) \geq d(s')$. If s and s'

are independent, let $G' = G - s$. Then $d_{G'}(s') = d_G(s')$ and

$$\begin{aligned}
|E(G)| &= |E(G')| + d_G(s) \\
&\geq \max_{v \in V(G')} (n(G') - d_{G'}(v))d_{G'}(v) - 1 + d_G(s) \\
&\geq (n(G') - d_{G'}(s'))d_{G'}(s') - 1 + d_G(s) \\
&= (n - 1 - d_{G'}(s'))d_{G'}(s') - 1 + d_G(s) \\
&= (n - d_{G'}(s'))d_{G'}(s') - d_{G'}(s') - 1 + d_G(s) \\
&= (n - d_G(s'))d_G(s') - d_G(s') - 1 + d_G(s) && (d_G(s') = d_{G'}(s')) \\
&= \max_{v \in V(G)} (n - d_G(v))d_G(v) - d_G(s') - 1 + d_G(s) \\
&\geq \max_{v \in V(G)} (n - d_G(v))d_G(v) - 1 && (d_G(s) \geq d_G(s'))
\end{aligned}$$

Thus we may assume S induces a clique.

Now assume $d_G(s) > d_G(s')$. Let $w \in V(G)$ such that w is not adjacent to s'. Let $G' = G - s + ws'$. Then $|E(G)| = |E(G')| + d_G(s) - 1$, $d_G(s) = d_{G'}(s)$, and since $d_G(s) - 1 \geq d_G(s')$, we can mimic the previous argument.

We now assume the remaining case: S induces a clique and all vertices in S share the same degree. We will show that this cannot occur. As previously shown, $|S| \geq \lfloor n/2 \rfloor$. As $\overline{\chi}(G) \geq 3$, it must be that $V(G) - S$ is nonempty. Let $v \in V(G) - S$ such that v has a neighbor $s \in S$. Then $d(v) \leq \lfloor n/2 \rfloor - 1$, as otherwise $d(v) \geq \lfloor n/2 \rfloor$ implies $v \in S$. Since $|S| \geq \lfloor n/2 \rfloor$, we let $|S| = \lfloor n/2 \rfloor + k$ for $k \geq 0$. Then $d(v) = \lfloor n/2 \rfloor - k - 1$, as otherwise $(n - d(v))d(v) \geq (n - d(s))d(s)$ for $s \in S$. Since G has diameter two, there is a partition of $V(G)$ into sets $\{v\}$, $N(v)$, and a clique C. Thus

$$\begin{aligned}
|C| &= n - d(v) - 1 \\
&= n - (\lfloor n/2 \rfloor - k - 1) - 1 \\
&= \lceil n/2 \rceil + k.
\end{aligned}$$

Then every vertex in $|C|$ has degree at least $\lceil n/2 \rceil + k$ (each vertex in C has at least one neighbor in $N(v)$), which implies $C \subseteq S$. But

$$|C| = \lceil n/2 \rceil + k \geq |S|.$$

In the case n is odd, we reach a contradiction. Otherwise, if n is even, it is possible $C = S$. We assumed v has a neighbor $s \in S$, however this implies $s \notin C$, which is a contradiction.

□

CHAPTER 4

MODELING RNA PSEUDOKNOT STRUCTURE VIA CHORD DIAGRAMS AND INTERSECTION GRAPHS

4.1 Biological Overview

Ribonucleic acid (RNA) is a molecule essential to many functions of life, notably gene expression, cellular communication, and the storage and transfer of genetic information. Typically a single-stranded structure, RNA is composed of a sequence of nucleotides containing a ribose sugar to which one of four nitrogenous bases (nucleobases) is attached: adenine (A), guanine (G), cytosine (C), and uracil (U). A sugar-phosphate backbone is formed by the binding of nucleotides in which the ribose sugars are linked together via phosphate groups. The primary structure of RNA refers to this nucleotide sequence, typically oriented from the $5'$ end to the $3'$ end [49].

It is well known that RNA molecules fold into a variety of secondary and tertiary structures related to their natural functions [120, 49]. Secondary and tertiary structures derive from both complementary Watson-Crick base pairing (A-U and G-C) and non-Watson-Crick pairs formed less commonly between bases (e.g. G-U or A-A pairing). There are several types of RNA, including ribosomal (rRNA), messenger (mRNA), and transfer (tRNA) RNA. Secondary structures of RNA may be schematically represented with diagrams. An example of a secondary structure schematic of a tRNA structure from *Escherichia coli* is given in Figure 28 [84, 85], and a hypothetical structure is shown in Figure 29. Common secondary structure motifs include, single-stranded regions, hairpin stems, hairpin loops, bulges and internal loops, junctions, multiloops, and pseudoknots [49, Figure 2.8], [35, Figure 3], [16, Table 1], [96, Figure 1.1], [44, Figure 1].

A pseudoknot is a motif of RNA secondary structure that is not topologically knotted, but does represent a three-dimensional folding pattern. Pseudoknots were first recognized in the study of the turnip yellow mosaic virus [98, 48], but the term

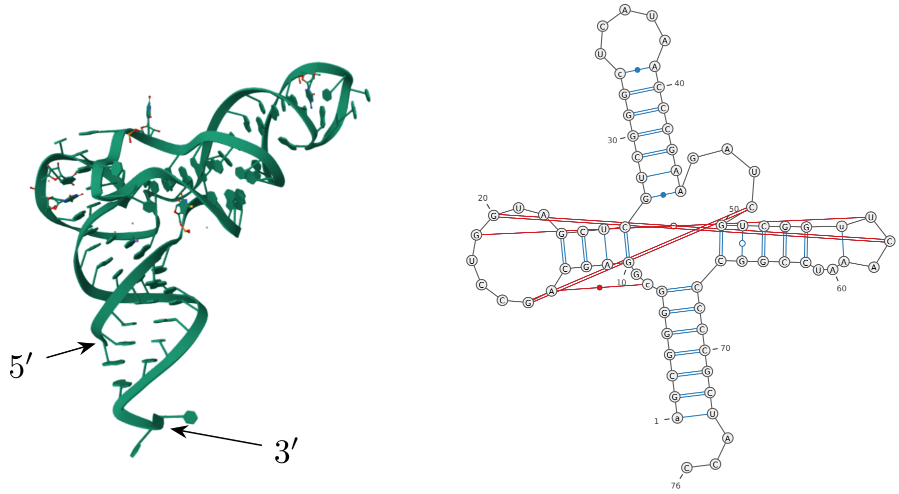


Fig. 28. Left: A 3D rendering of RCSB PDB 5L4O (or PDB_652 in the bpRNA-1m database [44]), a transfer RNA molecule present in *Escherichia coli* and obtained via x-ray diffraction [84, 85]. Image made using Mol* [102]. Right: Secondary structure schematic from [44].

was coined in Studnicka et al. [110]. The simplest type of pseudoknotted structure is an H-type pseudoknot, formed when nucleobases along the loop of a hairpin form bonds with nucleobases elsewhere along the sequence [19]. See Figure 29. A wide variety of pseudoknot motifs have been described and characterized, for example the basic H-type, K-type, L-type, and M-type motifs [70, 3]. In Bon et al. [16] it is shown that these basic motifs constitute the irreducible pseudoknots of topological genus equal to one (see section 4.5.5 for discussion on genus).

RNA secondary structures may be represented graphically with linear or circular *chord diagrams*, in which bonded pairs are represented by arcs ('chords') along a line segment or circle. An early version of a chord diagram representing a secondary structure appears in [119] as a connection to earlier models used in secondary structure prediction via a base-pairing matrix [111]. In [119], an RNA secondary structure is defined as a simple planar graph on a set of n labeled points such that there is a path along the n points representing the primary structure, with other edges that pair points representing bonds between bases [119, Definition 2.1].

Secondary structures represented by such planar graphs correspond to crossing-

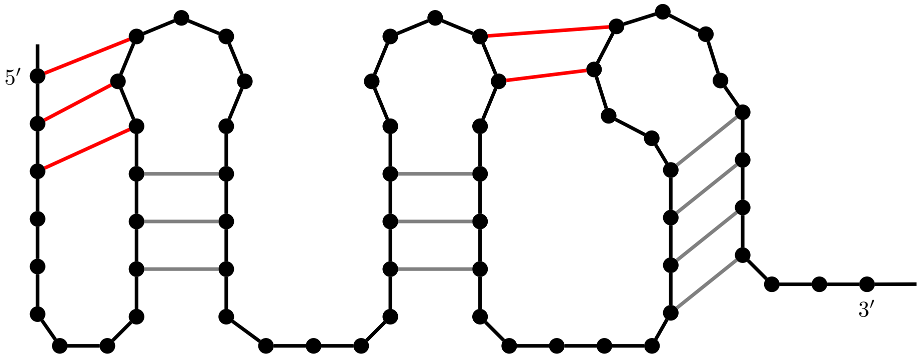


Fig. 29. A schematic example of a (hypothetical) segment of RNA consisting of a sequence of 63 nucleotides together with base-pairing information, i.e., an RNA secondary structure. The base sequence is oriented from the 5′-end to the 3′-end with black edges representing the sugar-phosphate backbone, and with gray and red edges representing hydrogen bonds in base pairing. The structure contains several hairpins forming two distinct pseudoknot motifs: an H-type pseudoknot (left) and an K-type pseudoknot (right).

less chord diagrams, which have since been studied extensively as models for understanding RNA secondary structure [16, 97, 117, 118, 91, 101]. Chord diagrams appear as objects of interest in knot theory and topology [94], [37, §3.4, §4] [71, §6],[69], and in enumerative combinatorics [1, 91, 119, 118, 66]. It is well known that the number of crossingless chord diagrams with $2n$ points and n chords is the nth Catalan number [108]. There is also interest in enumerating and studying chord diagrams with crossings under particular constraints [36, 112, 92, 2].

Roughly speaking, a pseudoknot in RNA occurs when its chord diagram contains chords that cross. In section 4.2, we will review several definitions related to pseudoknots from the literature, and give a more precise reformulation of pseudoknotting in terms of chord diagrams and intersection graphs. In section 4.3 we define an algorithm which reduces the complexity of a chord diagram for the purposes of quantifying pseudoknot complexity, and in section 4.4 we apply the algorithm and do a preliminary analysis on RNA secondary structures found in the bpRNA-1m database. The bpRNA-1m database stores 102,318 RNA secondary structures pulled from seven different sources [27, 124, 99, 68, 23, 57, 12] with their base pairing information, as well as annotation of various structural motifs such as loops (hairpin, multi, internal, exterior) and pseudoknots [44].

4.2 Combinatorial Theory

4.2.1 Chord Diagrams

Definition 4.1 (Linear Chord Diagram)**.** A *linear chord diagram* is a set of n points on an oriented line segment together with a (partial) matching of the points.

Note here we use *linear* chord diagrams as opposed to circular chord diagrams. A linear chord diagram can be obtained from a circular chord diagram by cutting the circle at a point, however the resulting diagram depends on the choice of point. In particular the choice of cut changes the ordering of the points by cyclic permutation.

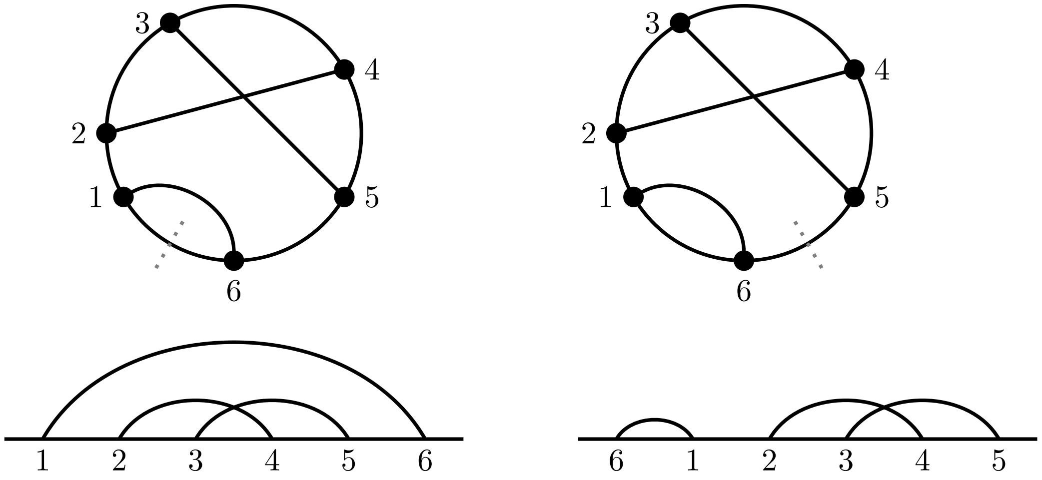


Fig. 30. Two different linear chord diagrams obtained from the same circular chord diagram, differing by a cyclic permutation of the points.

We denote chords as pairs $c_i = (\ell_i, r_i)$. Note as chord diagrams are matchings, $\ell_i < r_i$, and no two chords share an end point. By convention, chords are indexed by left endpoint.

Definition 4.2. For any two chords c_1 and c_2, there are three possibilities:

(a) c_1 and c_2 form a *crossing*: $\ell_1 < \ell_2 < r_1 < r_2$.

(b) c_1 and c_2 form a *nesting*: $\ell_1 < \ell_2 < r_2 < r_1$.

(c) c_1 and c_2 are *independent*: $\ell_1 < r_1 < \ell_2 < r_2$.

Further, a k-crossing is a set of chords $(\ell_1, r_1), (\ell_2, r_2), \ldots, (\ell_k, r_k)$ such that $\ell_1 < \ell_2 < \cdots < \ell_k < r_1 < r_2 < \cdots < r_k$. A k-nesting is a set of chords $(\ell_1, r_1), (\ell_2, r_2), \ldots, (\ell_k, r_k)$ such that $\ell_1 < \ell_2 < \cdots < \ell_k < r_k < r_{k-1} < \cdots < r_1$.

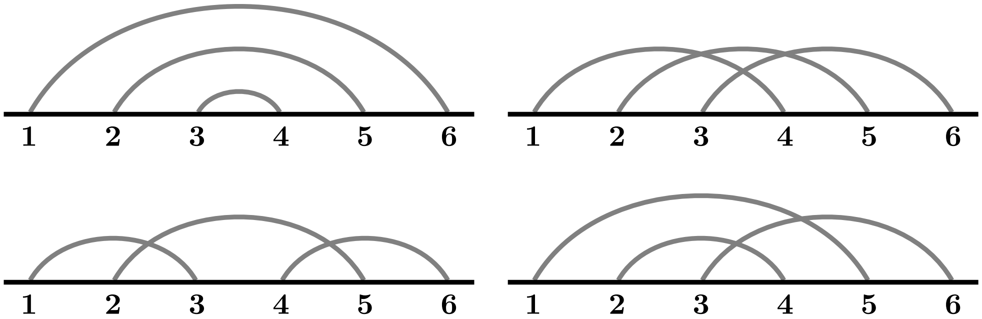


Fig. 31. Top left: A 3-nesting. Top right: A 3-crossing. Bottom left: Two 2-crossings. Bottom right: A 2-nesting and two 2-crossings.

The following definitions will become useful in formalizing pseudoknots in chord diagrams. We use the notation for the open interval between a and b as $(a, b) = (\min\{a, b\}, \max\{a, b\})$. In this way we need not specify $a \leq b$ or $b \leq a$, which may be ambiguous when comparing the end points of chords.

Definition 4.3 (Chord Obstructed)**.** Let c and c' be chords and let $U = (\ell, \ell') \cup (r, r')$, i.e. U can be thought of as the set of bases between the left endpoints and right endpoints of c and c'. We say c and c' are *chord obstructed* if there is some chord c'' such that either $\ell'' \in U$ or $r'' \in U$. A set of three or more chords $S = \{c_1, c_2, \ldots, c_k\}$ is chord obstructed if c_i and c_{i+1} are chord obstructed for some $i = 1, 2, \ldots, k-1$.

The chord obstructed relation throughout this chapter is only applied to consecutive chords with respect to the left endpoint ordering from the $5'$ to $3'$ end.

Definition 4.4 (Segment)**.** A *segment* of a chord diagram is a maximal nonempty set of chords $S = \{c_1, c_2, \ldots, c_k\}$ forming a k-nesting such that S is not chord obstructed.

Note that the set of segments $\mathcal{S}$ partitions the set of chords C. In crossingless chord diagrams, there is a natural poset structure on the set of segments defined by $S \prec S'$ if S' is nested in S.

Definition 4.5 (Intersection graph)**.** The *intersection graph* G of a chord diagram D is the graph whose vertices are the chords of D, and such that two vertices in G are adjacent if their corresponding chords in D form a crossing.

Example 4.6. Figure 30 left shows a 2-crossing nested within a chord, and right shows a chord which is independent from a 2-crossing. The intersection graphs corresponding with these chord diagrams are isomorphic. Specifically, the intersection graphs are an edge and an isolated vertex.

Remark 4.7. Definition 4.5 is well known, described more generally in [82], where chords in a chord diagram with n points are thought of as subintervals of $[1, n]$ and two vertices are adjacent if their corresponding intervals have nonempty intersection. See also [51]. Variations on the concept of an intersection graph arise numerous times in the biology literature under different names. For example, in [70], there is the notion of a conflict graph, whose vertex set comprises helices in the RNA structure and edges signify the crossing of chords corresponding to the helices. In Shu et al. [106] the concept of an element-contact graph is introduced, in particular the stem-loop-contact graph (SLCG) [106, Figure 7]. The segment graph of the bpRNA database [44] is another such example.

We will later compare the following graph theoretic invariants in Section 4.5.

Definition 4.8 (Cliques)**.** The maximum *clique number* $\omega(G)$ is the maximum size of an complete subgraph of G.

Definition 4.9. A *vertex cover* of a graph G is a set $A \subseteq V(G)$ such that for every edge $xy \in E(G)$ either $x \in A$ or $y \in A$. The *vertex cover number* of a graph G, denoted $\beta(G)$, is the number of vertices in a minimum vertex cover.

Definition 4.10. The *weight* of a vertex cover A in a vertex-weighted graph G is $\sum_{v \in A} w(v)$.

Note that a *minimum weight vertex cover* of a weighted graph G is not necessarily a minimum cardinality vertex cover. (Consider for example the path graph P_3 with weights 1,5,3).

If C is a vertex cover of G, then the graph $G-C$ contains no edges, as by definition every edge of G must have an endpoint in C. Similarly, if I is an independent set of G, then every edge of G has at least one end point in $G - I$. Thus we have the following observation.

Observation 4.11. *Let G be a graph and let I and C be an independent set and vertex cover respectively. Then $V(G) \setminus I$ is a vertex cover, and $V(G) \setminus C$ is an independent set. Moreover, the complement of a minimum weight vertex cover is a maximum weight independent set.*

For any $A \subset V(G)$ let $\omega(A)$ be the sum of the weights of vertices in A. Let $\Omega = \omega(V(G))$. Let C be a minimum weight vertex cover, and let $I = V(G) \setminus C$. If I was not a maximum weight independent set, then there is some independent set I' with $\omega(I') > \omega(I)$, and corresponding vertex cover $C' = V(G) - I'$ where

$$\omega(C') = \Omega - \omega(I') < \Omega - \omega(I) = \omega(C).$$

This is a contradiction, as C is a minimum weight vertex cover. Thus I must be an independent set of maximum weight. Similarly, all maximum weight independent sets correspond to minimum weight vertex covers.

Definition 4.12 (Genus)**.** The *genus* of a chord diagram D, denoted $\gamma(D)$, is half of the rank of the adjacency matrix of the intersection graph with $\mathbf{Z}_2$ coefficients [86].

Equivalently, the genus of a chord diagram is equal to the topological genus of the surface obtained by regarding the chord diagram as a band surgery diagram [16]. One way to calculate the genus g of a chord diagram D is via the formula

$$g = \frac{P - L}{2}$$

where P is the number of chords (or base pairs) and L is the number of closed loops in the corresponding double-line diagram [16] (see Figure 32.)

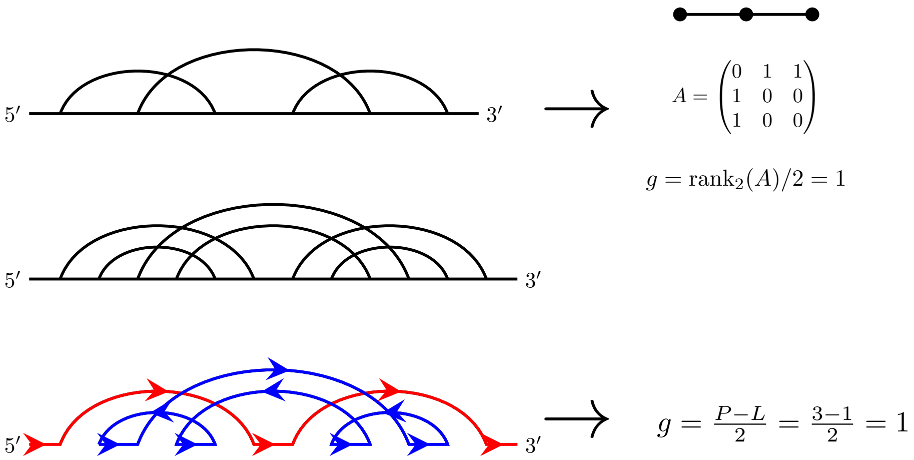


Fig. 32. Two methods of calculating the genus of a given chord diagram.

For a graph G we define $\mathrm{rank}(G)$ (resp. $\mathrm{rank}_p(G)$) to be the rank of the adjacency matrix of G with coefficients in $\mathbb{R}$ (resp. coefficients in $\mathbb{F}_p$).

Theorem 4.13. *Let D be a chord diagram and let G be the intersection graph of D. If G is acyclic, then $\gamma(D) = \beta(G)$.*

Proof. Let n be the order of G. The statement is true for $n \leq 4$ (see Figure 33.) We will use induction on n.

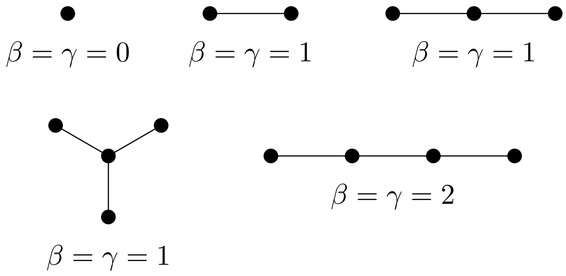


Fig. 33. The trees of order four or less and their genus and vertex cover numbers.

Let T be a tree on $n \geq 5$ vertices with adjacency matrix A. If T has a vertex v with at least two leaf neighbors, then we remove a leaf ℓ adjacent to v to form T'. By the induction hypothesis, $\mathrm{rank}_2(T')/2 = \beta(T')$. Adding back ℓ to form T, we have $\mathrm{rank}_2(T') = \mathrm{rank}_2(T)$ as the row corresponding to ℓ in A is identical to the rows corresponding to the other leaf neighbors of v. Note in a graph, for any vertex with a leaf neighbor, there is a minimum vertex cover containing that vertex. Thus, we have $\beta(T) = \beta(T')$.

Now we assume every vertex of T has at most one leaf as a neighbor. Then there is some leaf ℓ with a neighbor v of degree two; let $\{\ell, w\} = N(v)$. Let $T' = T - \{v, \ell\}$. By the induction hypothesis, $\operatorname{rank}_2(T')/2 = \beta(T')$.

First we claim $\beta(T) = \beta(T') + 1$. Indeed, no minimum vertex cover of T' will cover the edge $v\ell$, so T requires one more vertex in addition to a minimum vertex cover of T' to cover all edges of T.

Next we claim that $\operatorname{rank}_2(T) = \operatorname{rank}_2(T') + 2$. The adjacency matrix of T is given by

$$A = \begin{array}{c|cccc|cc} & & & & w & v & \ell \\ & & & & * & 0 & 0 \\ & & & & \vdots & \vdots & \vdots \\ & & & & * & 0 & 0 \\ w & * & \cdots & * & 0 & 1 & 0 \\ \hline v & 0 & \cdots & 0 & 1 & 0 & 1 \\ \ell & 0 & \cdots & 0 & 0 & 1 & 0 \end{array}$$

where the upper right block corresponds to the adjacency matrix A' of T'. Applying the row operation $\operatorname{row} w - \operatorname{row} \ell = \operatorname{row} v$ and column operation $\operatorname{col} w - \operatorname{col} \ell = \operatorname{col} w$ we obtain the matrix

$$B = \begin{array}{c|cccc|cc} & & & & w & v & \ell \\ & & & & * & 0 & 0 \\ & & & & \vdots & \vdots & \vdots \\ & & & & * & 0 & 0 \\ w & * & \cdots & * & 0 & 0 & 0 \\ \hline v & 0 & \cdots & 0 & 0 & 0 & 1 \\ \ell & 0 & \cdots & 0 & 0 & 1 & 0 \end{array}.$$

Since row and column operations preserve the rank, we have $\operatorname{rank}_2(B) = \operatorname{rank}_2(A)$, and since rows v and ℓ are linearly independent from every other row in B, we have $\operatorname{rank}_2(B) = \operatorname{rank}_2(A') + 2$. This shows that $\operatorname{rank}_2(T)/2 = \beta(T)$. □

Finally, we remark that the genus of a chord diagram D containing an r-nesting $C = \{c_1, \cdots, c_r\}$ that is not chord obstructed is equal to the genus of the diagram with $r-1$ chords of C removed, i.e. $D - \{c_2, \cdots, c_r\}$. The rank of a matrix can be thought of as the maximum number of linearly independent rows or columns in the matrix. It can be seen from the adjacency matrix A of the intersection graph that $c_2, \ldots, c_r$ correspond to identical rows in A, and are thus linearly dependent.

4.2.2 Pseudoknotted Structures

As discussed above, analyzing the role of pseudoknots in various RNA processes motivates the study of pseudoknot complexity for comparison, characterization of motifs, and prediction of RNA secondary structure [44, 16, 97, 109]. Practical definitions of a pseudoknot vary in the biological literature [16, 97]. Because we will ultimately be interested in quantifying pseudoknots aggregated by the bpRNA-1 database [44], we start from definitions presented there. In this database, a pseudoknotted structure is characterized as having base pair positions that cross in the sense of Definition 4.2(a); the working definition of a 'pseudoknot base pair' is one belonging to a minimal set of base pairs that results in a pseudoknot-free structure once removed. There is of course some ambiguity in these concepts. The simple example of two adjacent H-type pseudoknots demonstrates that such a minimal set is not unique. However, Danaee et al. [44] give one answer to the question of *how many* pseudoknots exist in a given secondary structure by providing an algorithm that identifies a minimal set of base pairs. We review their algorithm, and translate corresponding notions of pseudoknotting into the language of chord diagrams, as follows.

To make the notion of a pseudoknot, and more specifically the annotation of multiple pseudoknots, more precise, [44] introduces the notion of a segment of RNA secondary structure. An RNA segment is described as a region of duplexed RNA, possibly containing bulges or internal loops. In combinatorial terms, RNA segments correspond with the segments $s \in \mathcal{S}$ of the chord diagram D representing the secondary structure, as in Definition 4.4. The segments partition the set of chords C, and ordering the base sequence from the 5'-end to 3'-end indexes each segment by its leftmost endpoint. In [44], the following is observed; we provide a restatement in terms of chord diagrams.

Theorem 4.14 (Danaee et al. [44]). *Let $\mathcal{S}$ be the segment partition of a linear chord diagram D and let $S, S' \in \mathcal{S}$. If there are chords $c \in S$ and $c' \in S'$ such that c and c' are crossed, then any pair of chords from S and S' cross.*

Proof. Recall that a segment of size k is a maximal k-nesting in which pairs of consecutive chords are not chord obstructed. Let $\mathcal{S}$ be a segment partition of D and let $S, S' \in \mathcal{S}$ where $S < S'$ in the indexing of segments by left endpoints. Let $c \in S$ and $c' \in S'$ such that c and c' cross. Let $\ell_{\max}$ be the maximum left endpoint of S and $r_{\min}$ and $r_{\max}$ be the minimum and maximum right endpoints of S respectively.

By the left endpoint indexing, all left endpoints of chords in S' must be greater than $\ell_{\max}$. If some left endpoints of S' are less than $r_{\min}$ and some greater, then S' is chord obstructed by the right endpoints of S. If all left endpoints of S are greater than $r_{\min}$, either S is chord obstructed or no chords of S and S' cross. So all left endpoints of S' must lie between $\ell_{\min}$ and $r_{\min}$. Since c and c' cross, and S is not chord obstructed, it must be that $r' > r_{\max}$, i.e. c' crosses every chord in S. No other right endpoint of S' is less than $r_{\max}$, as otherwise S or S' are chord obstructed. Thus any pair of chords from S and S' cross. ☐

In other words, Theorem 4.14 says that two segments $s, s' \in S$ cross whenever any chords $c \in s, c' \in s'$ cross. The *segment graph* $G_{\mathcal{S}}$ of a chord diagram is a variation on an intersection graph (see Definition 4.5) whose vertex set is the set of segments, where two vertices are adjacent if their segments cross [44]. Vertices in $G_{\mathcal{S}}$ are weighted by the number of chords contained in their corresponding segments. The terminology *PK-segment* refers to any segment $s \in \mathcal{S}$ that crosses any other segment. Let $H_{\mathcal{S}} \subset G_{\mathcal{S}}$ be the subgraph with isolated vertices removed, referred to as the *PK-segment graph* in [44].

In [44], pseudoknotted structures are identified by finding a maximum weight independent set I in $H_{\mathcal{S}}$ via a heuristic approach, with an exact algorithm used in the specific case of components which are paths. As discussed above, the set $P = V(G_{\mathcal{S}}) - I$ is a minimum weight vertex cover of $G_{\mathcal{S}}$. That is, P is a set of segments, of minimum cardinality in C, that when removed from D leave a pseudoknot-free structure.

We may now formally quantify the size of a pseudoknotted structure according to the conventions of the bpRNA-1m database, as implied by the algorithms of Danaee et al. [44].

Definition 4.15. (Pseudoknotted Structures - *Segment Graph Method*) A secondary structure is *pseudoknotted* if its segment graph contains at least one edge and is called *pseudoknot-free* otherwise. The *number of pseudoknots* in a pseudoknotted structure is the minimum cardinality over all vertex covers of minimum weight of the segment graph.

Any segment contained in a minimum cardinality minimum weight vertex cover may be called simply 'a pseudoknot.' It is important to note that in [44], the number of pseudoknots is not the number of chords in the corresponding cover, in general.

Example 4.16. Figure 34 shows the chord diagram representing the secondary structure of tRNA (76-MER) found in *Escherichia coli* with 8 segments. The PK-segments are $\{2, 3, 4, 5, 6, 8\}$. The maximum weight independent set in $H_{\mathcal{S}}$ is $\{3, 8\}$. The minimum weight vertex cover in $G_{\mathcal{S}}$ is $\{2, 4, 5, 6\}$, indicating four pseudoknots in this structure according to the conventions of [44].

Definition 4.17 (Crossingless Secondary Structures)**.** Let D be a chord diagram with no crossings, and let $\mathcal{S}$ be the segment partition of D. We say an unpaired base is nested in a segment S if it is between the left and right endpoints of the innermost chord of S. A *stem* in D is a k-nesting with no other bases between the endpoints of any two consecutive chords. Segments are composed of stems. Let S be a segment and let c and c' be two consecutive chords in a segment S, with $\ell < \ell'$. If there are sequences of unpaired bases in the intervals (ℓ, ℓ') and (r, r'), i.e. $\ell' - \ell \geq 2$ and $r - r' \geq 2$, then those two sequences together comprise an *interior loop*. If exactly one of (ℓ, ℓ') or (r, r') contain a sequence of unpaired bases, that sequence is a *bulge*. If two or more segments $\mathcal{T}$ are nested in S, then there is a *multiloop* composed of all unpaired bases b nested in S and not nested in any segment in $\mathcal{T}$. Note multiloops may have length zero (that is, a multiloop may be the empty set). The *exterior loop* of D is the set of all unpaired bases which are not nested in any chord. If b_0 is

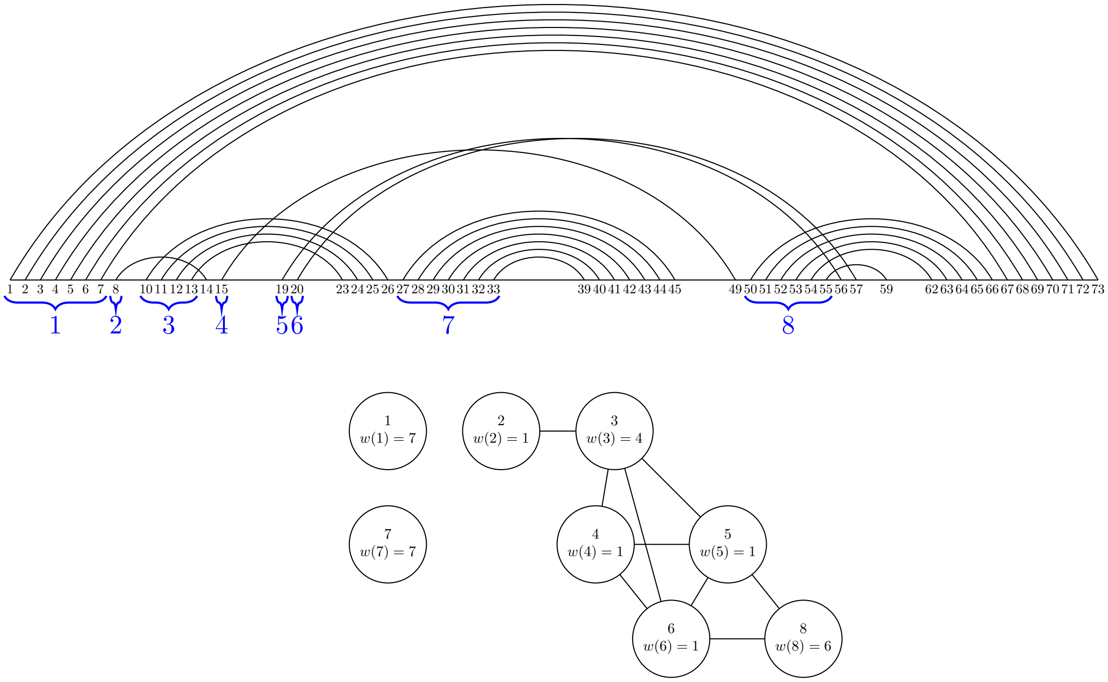


Fig. 34. (Top) Linear chord diagram of transfer RNA molecule of type 76-MER from *Escherichia coli* (PDB_652) with segments labeled. (Bottom) Segment graph $G_{\mathcal{S}}$ with weights.

the first paired base and b_f is the last paired base in the base sequence, the set of unpaired bases less than b_0 and greater than b_f are a part of the exterior loop called the *dangling ends*. The exterior loop can be thought of as the multiloop gained from an imaginary base pair bonding of the $5'$-end and $3'$ end, under which all chords are nested.

The secondary structures defined in Definition 4.17 are illustrated in Figure 35.

4.3 The τ-Segment Graph Method

4.3.1 Qualitatively Similar Pseudoknotted Structures

Identifying nested chords in a chord diagram is a common strategy for reducing the complexity of the combinatorial analysis of secondary structures because it allows for the reduction of helices to a single chord, or alternatively to a single vertex in an

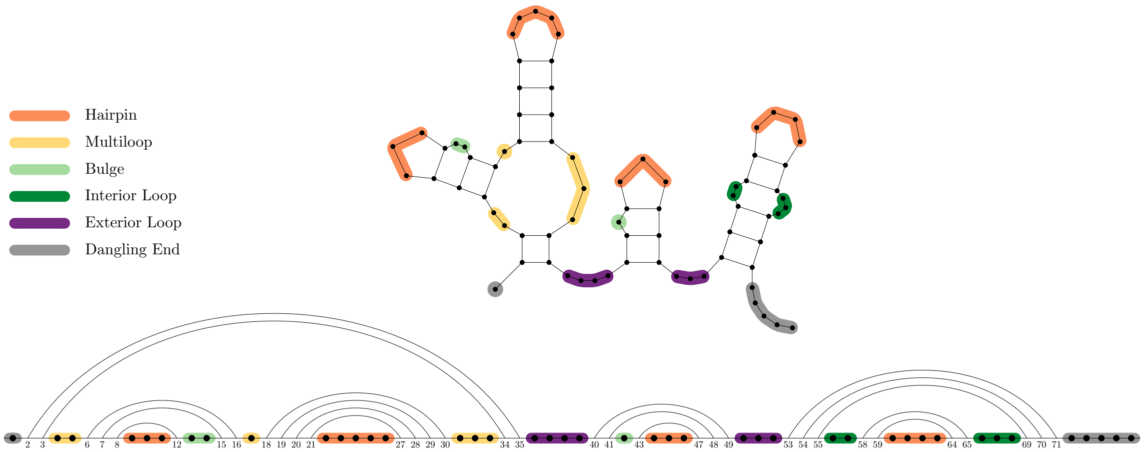


Fig. 35. An illustrated example of secondary structures in a crossingless chord diagram. There are eight stems in this structure

intersection graph. For example, nested chords are identified as segments in [44] (Definition 4.4 above), as ‘stacks’ in [16], as ‘shadows’ in [97], or as edges in tree diagrams of pseudoknot-free structures in [11] and elsewhere. Such reductions also preserve some invariants of interest, for example the topological genus (see Definition 4.12 and [16]). In contrast to r-nestings, r-crossings in chord diagrams are typically left untouched by simplification algorithms. Consequently, the existence of an r-crossing implies the existence of at least r pseudoknots according to the conventions of [44].

To further reduce complexity and to more effectively relate similar secondary structures, we propose in this section an alternative method for quantifying the size of a pseudoknotted structure and a new simplification algorithm that identifies both r-crossings and r-nestings in chord diagrams. In addition to handling complexity issues arising from r-crossings, this method will also eliminate some discrepancies resulting from weighted graphs and include a parameter accounting for distance between nucleotides. Before defining the algorithm, let us consider two motivating examples.

Example 4.18 (Discrepancies due to weights)**.** Consider two weighted segment graphs, each isomorphic to P_3, with weights $(1, 5, 3)$ and $(1, 5, 4)$, respectively. Such graphs represent nearly identical K-type secondary structures which differ only by a single bonded pair in the third stem. By [44] and Definition 4.15, the minimum cardinality over vertex covers of minimum weight determines the number of pseudoknots: 2 and 1, respectively. Note that the addition of one pair results in a *decrease* in the number of pseudoknots.

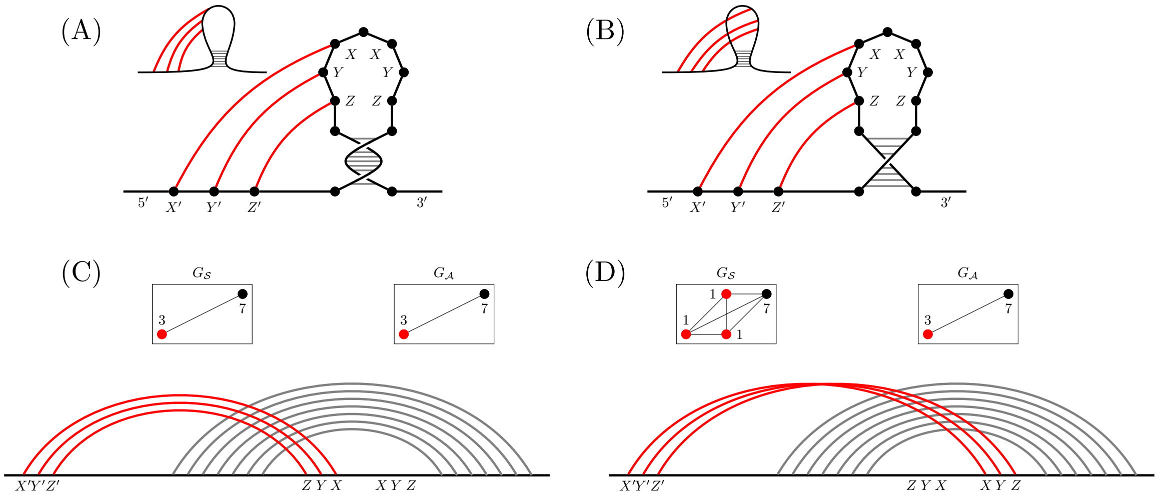


Fig. 36. Two closely related instances of hypothetical bonding between two hairpins.

Example 4.19 (Discrepancies due to r-crossings)**.** Consider a nucleotide sequence that contains a repetitious subsequence appearing in reverse. For an example, we follow Figure 36. This structure contains subsequence $\sigma = X'Y'Z'$, complementary sequence $\sigma' = ZYX$, and reverse complementary sequence $\overline{\sigma}' = XYZ$. Bonds formed between σ and σ' result in an r-nesting (Figure 36(A)), whereas bonds formed between σ and $\overline{\sigma}'$ form an r-crossing (Figure 36(B)). We assume here that either set of bonds are possible. In particular, the parity of number of half-turns in the helical stem may determine whether σ' or $\overline{\sigma}'$ is nearer to σ in a 3D conformation. Despite the heuristic similarity of the resulting pseudoknotted structures, their segment graphs differ significantly. By the conventions of [44] (Definition 4.15) the two structures are of pseudoknotting size 1 and 3, respectively. Moreover, the difference increases with additional base pairing in the hairpin stem.

The general observation is that variances in bonding from spatial conformations (e.g. helical twisting) may result in a quantification of pseudoknotting that is artificially high. Of 30 structures exhibiting the most pseudoknotting in Table 1, there are

28 which contain both a complementary and reverse complementary sequence (possibly not contiguous) near the site of the 3-crossing. One example RNA structure involving a 3-crossing and 5-nesting is bpRNA_CRW_55315. This structure contains a 3-crossing with left bases GCA at indices 2107, 2108, and 2112 and right bases AGU at indices 2164, 2165, and 2167, which is the reverse of the triple UGA at indices 2162, 2163, and 2164.

4.3.2 Augmented Segments and Distance in Chord Diagrams

In this subsection, we give a partitioning procedure that identifies r-crossings in a manner similar to that of r-nestings. We implement the procedure in an algorithm that incorporates an additional distance parameter τ in terms of the nucleotide sequence.

Definition 4.20 (Augmented segment)**.** An *augmented segment* of a chord diagram is a maximal nonempty set of chords $S = \{c_1, c_2, \ldots, c_k\}$ forming a k-nesting or a k-crossing which is not chord obstructed.

Revisiting Examples 4.18 and 4.19 above, notice that an intersection graph $G_{\mathcal{A}}$ produced with augmented segments would yield two pseudoknotted structures of size 1 in Example 4.18 and two structures of size 2 in Example 4.19. Let us now define notions of distance and nearness in chord diagrams.

Definition 4.21 (Chord distance)**.** The *chord distance* between two chords $c_1 = (\ell_1, r_1)$ and $c_2 = (\ell_2, r_2)$ is

$$d(c_1, c_2) = \max\{|\ell_1 - \ell_2|, |r_1 - r_2|\}.$$

Definition 4.22 (τ-near)**.** Two chords c_1, c_2 are *τ-near* if $d(c_1, c_2) \leq \tau$ and c_1 and c_2 are not chord obstructed. A k-crossing or k-nesting $\{c_1, \ldots, c_k\}$ is τ-near if for each $i = 1, \ldots, k-1$ we have c_i and c_{i+1} are τ-near.

Definition 4.23 (τ-Segment)**.** A *τ-segment* of a chord diagram is a maximal nonempty set of chords $S = \{c_1, c_2, \ldots, c_k\}$ forming a τ-near k-nesting or a τ-near k-crossing.

As with segments or augmented segments, the set of τ-segments $\mathcal{S}_\tau$ also partition the set of chords C. Notice that if two chords are τ-near, then they cannot be independent. Indeed, if (ℓ_1, r_1) and (ℓ_2, r_2) were independent, then the right endpoint for one chord would be between the left endpoints ℓ_1 and ℓ_2. Hence, if two chords are τ-near, they must either be crossed or nested. We have the following statement.

Theorem 4.24. *Let $\mathcal{S}_\tau$ be the τ-segment partition of a linear chord diagram D and let $S_1, S_2 \in \mathcal{S}_\tau$. If there are chords $c \in S_1$ and $c' \in S_2$ such that c and c' are crossed, then for every pair $c \in S_1$ and $c' \in S_2$, the chords c and c' are crossed.*

Proof. The proof is analogous to that of Theorem 4.14. $\square$

Generalizing the segment graph, we may now define the τ-segment intersection graph.

Definition 4.25 (G_τ)**.** The τ-segment intersection graph G_τ of a chord diagram is the weighted graph whose vertex set is the set of τ-segments, where two vertices are adjacent if the τ-segments cross. When $\tau = 0$, define $G_{\tau=0} := G_{\mathcal{S}}$. When $\tau = \infty$, the graph $G_{\mathcal{A}} := G_\infty$ is the intersection graph of the augmented segment partition.

The notation D_τ will indicate the chord diagram corresponding with G_τ, in which each τ-segment corresponds with a chord. As with $G_{\mathcal{S}}$ and $G_{\mathcal{A}}$ we use the notation $D_{\mathcal{S}}$ and $D_{\mathcal{A}}$ analogously. We may now formally revise the method for quantifying the size of pseudoknotted structures in RNA.

Definition 4.26. (Pseudoknotted Structures- *τ-Segment Graph Method*) A secondary structure is *τ-pseudoknotted* if G_τ contains at least one edge and is called *pseudoknot-free* otherwise. For $\tau \geq 1$, the *number of pseudoknots* is the minimum cardinality of a vertex cover of G_τ. For $\tau = 0$, the number of pseudoknots is the minimum cardinality over all vertex covers of minimum weight of the segment graph.

The definition in the case of $\tau = 0$ is explicitly made to agree with the conventions of [44].

Example 4.27. Figure 37. The maximum distance between any two non-chord obstructed chords is three. That is, the τ-segment partition is the same for all $\tau \geq 3$.

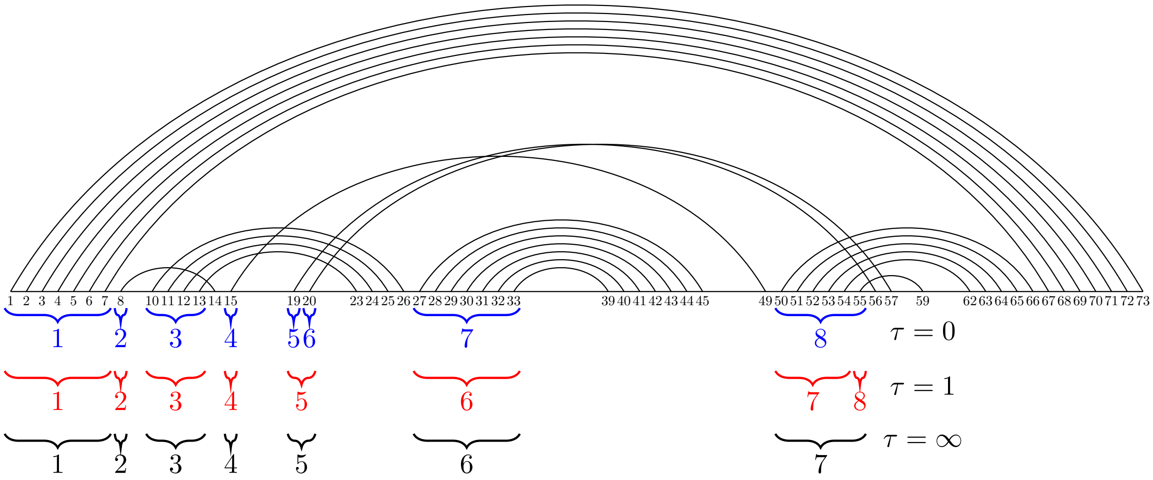


Fig. 37. Different tau-partitions of bpRNA_PDB_652 for $\tau = 0, 1$, and ∞.

This RNA structure has four pseudoknots according to the conventions of [44], and two pseudoknots according to Definition 4.26. The main difference comes from chords $(19, 56)$ and $(20, 57)$ becoming part of the same segment.

Algorithm 1 implements the τ-segment partition procedure. The input to the algorithm is a chord diagram with set of chords C (a list of base pairs indexed by left endpoint) and a non-negative integer parameter τ. Selecting a pair of chords $c, c' \in C$, the algorithm loops to build the τ-segment containing c. If $\tau = 0$, it checks whether the c, c' are nested and not chord obstructed. If $\tau > 0$, the algorithm checks whether c and c' are τ-near and not chord obstructed. If the criterion is met, c' is added to the segment containing c and the next pair is selected. If not, the segment is closed. The next unvisited pair of chords are then selected and the process repeats to build the next segment until all chords have been exhausted.

See Appendix 5.4.1 for a version of Algorithm 1 implemented in Matlab in the function called `findSegments`.

4.4 Methodology

In this section, we apply the definitions and algorithm above to the bpRNA-1m(90) database [44]. The database bpRNA-1m(90) is a subset of bpRNA-1m re-

Algorithm 1 Tau-Segment Partition

```
Input: $C = \{c_1, c_2, \ldots, c_k\} = \{(\ell_1, r_1), (\ell_2, r_2), \ldots, (\ell_k, r_k)\}$, $\tau$

Output: $\mathcal{C}_\tau$

  $S \leftarrow \{(\ell_1, r_1)\}$

  if $\tau > 0$ then

      for $1 \le i \le k-1$ do
          $d \leftarrow \max\{|\ell_i - \ell_{i+1}|, |r_i - r_{i+1}|\}$            ▷ Distance.
          if $d \le \tau$ and $\neg$ isChordObstructed$(c_i, c_{i+1})$ then
              $S \leftarrow$ append$(S, (\ell_{i+1}, r_{i+1}))$
          else
              $\mathcal{C}_\tau \leftarrow$ append$(\mathcal{C}_\tau, S)$            ▷ Store Segment
              $S \leftarrow \{c_{i+1}\}$            ▷ Initialize new segment
          end if
      end for
  else
      for $1 \le i \le k-1$ do
          if isNested$(c_i, c_{i+1})$ and $\neg$ ChordObstructed$(c_i, c_{i+1})$ then
              $S \leftarrow$ append$(S, c_{i+1})$
          else
              $\mathcal{C}_\tau \leftarrow$ append$(\mathcal{C}_\tau, S)$            ▷ Store Segment
              $S \leftarrow \{c_{i+1}\}$            ▷ Initialize new segment
          end if
      end for
  end if
  $\mathcal{C}_\tau \leftarrow$ append$(\mathcal{C}_\tau, S)$            ▷ Account for final segment.
```

stricted to the 28,370 RNA secondary structures with less than 90% sequence similarity. This database contains 3,320 RNA structures reported to contain at least one pseudoknot, i.e. structures whose segment graphs contain at least one edge, with a total of 7164 pseudoknots reported. To analyze the data, we perform the segment and τ-segment graph methods and analyze secondary structures by implementing the workflow algorithm of Figure 38 in MATLAB, and described as follows.

Chord diagrams associated with RNA structures are stored as Matlab arrays. The input to the algorithm is the set of chords C from a chord diagram D and an integer parameter τ. To carry out any of the above methods, we first call the `findSegments` function to create a segment partition of the chord diagram. The

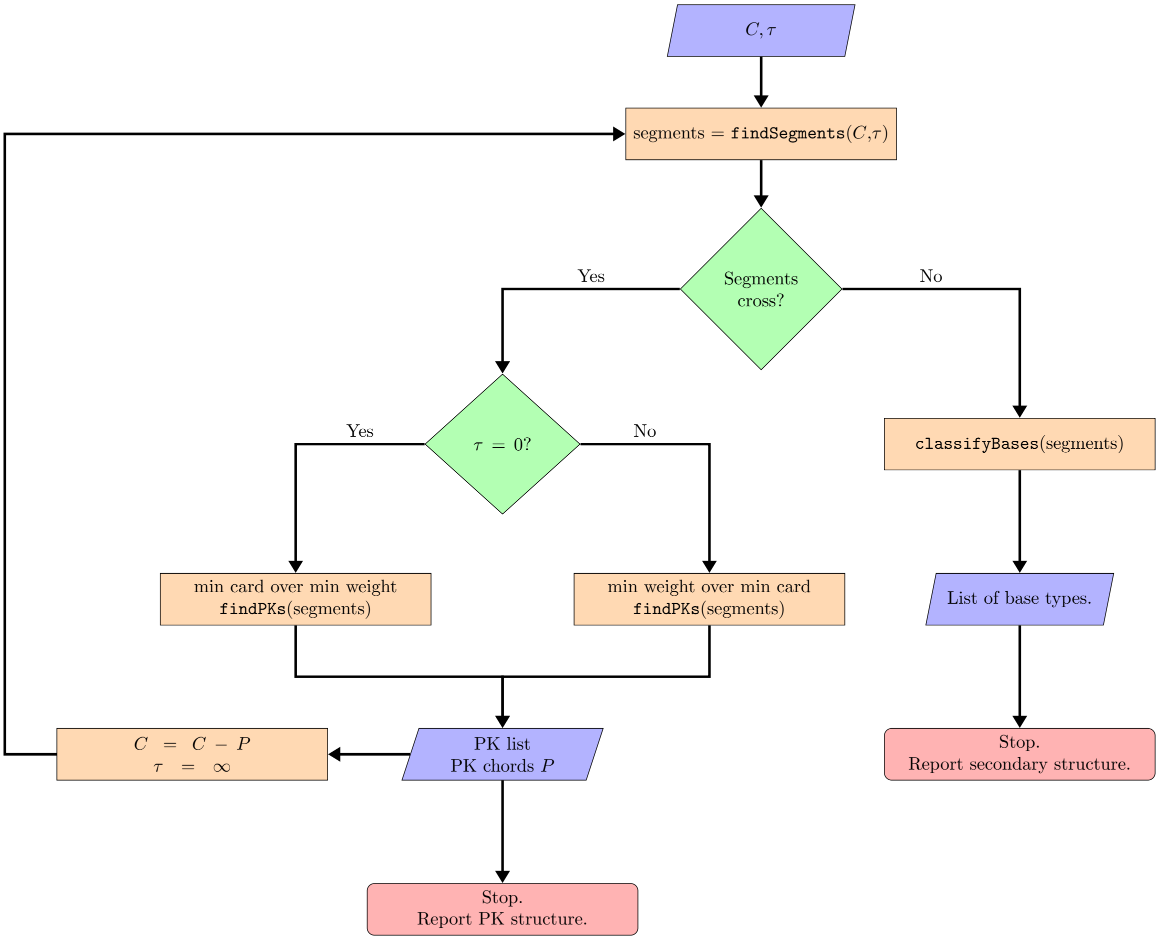


Fig. 38. A workflow diagram illustrating the process outlined in Section 4.4.

parameter determines which segment partition is created, with $\tau = 0, \infty, 1 < \tau < \infty$ corresponding to the segment partition, augmented segment partition, and τ-partition, respectively. See Figure 37

Depending on whether the segment partition contains any segments which cross each other, one of two subroutines is implemented. If the segment partition contains no segments which cross, then the secondary structures of the chord diagram are analyzed by the function `classifyBases`. This function outputs the primary base sequence with each base classified as belonging to one of the secondary structure motifs given in Definition 4.17.

If the segment partition contains segments which cross, then the pseudoknots of the chord diagram are analyzed by the `findPKs` process. The corresponding weighted segment graph is produced by the function `makeSegmentGraph`. To identify pseudoknots in the segment graph, the process carries out the following:

1. Find list of all maximal independent sets I.
2. If $\tau = 0$:
    (a) Calculate weight of each set $I \in \mathcal{I}$.
    (b) Subset all maximum weight sets $\mathcal{I}' \subseteq \mathcal{I}$.
    (c) Subset maximum cardinality maximum weight sets $\mathcal{I}'' \subseteq \mathcal{I}' \subseteq \mathcal{I}$.
3. If $\tau > 0$:
    (a) Subset all maximum cardinality sets $\mathcal{I}' \subseteq \mathcal{I}$.
    (b) Calculate weight of each set $I \in \mathcal{I}'$.
    (c) Subset maximum weight maximum cardinality sets $\mathcal{I}'' \subseteq \mathcal{I}' \subseteq \mathcal{I}$.
4. Dualize the independent sets $\mathcal{I}''$ to vertex covers $\mathcal{P}$.
5. Select first vertex cover $P \in \mathcal{P}$ with respect to the lexicographical ordering from the indexing of segments by left endpoints.

Step (1) applies the Bron-Kerbosch algorithm [21, 14] to find all maximal cardinality independent sets. The importance of the lexicographical ordering in step (5) will become apparent after Example 4.28 below.

The output of the process in Figure 38 is a vertex cover P. In the case that $\tau = 0$, this vertex cover represents a pseudoknotted structure by Definition 4.15, where the number pseudoknots is quantified by the minimum cardinalities of vertex covers of minimum weight. In the cases where $\tau > 0$, the minimum weight minimum cardinality vertex cover represents a τ-pseudoknotted structure by Definition 4.26.

For structures containing pseudoknots, we may still want to analyze the secondary structures of the RNA molecule. Therefore the last process removes all chords which compose a pseudoknot, leaving a set of chords $C' = C - P$. In the case of $\tau = 0$, the set of chords C' is crossingless. In the case $\tau > 0$, it may be that chords in C' cross, however C' comprises independent crossings and nestings, i.e. no two segments cross in the augmented segment partition of C'. Setting $\tau = \infty$ (which corresponds to an augmented segment partition), we enter the secondary structure classification subroutine assessing secondary structure using Definition 4.17, with independent crossings handled as nestings (and thus a type of stem).

Example 4.28. Here we apply the $\tau = 0$ reduction method to the RNA structure 3DIG from the Protein Data Bank [104, 103] (see Figure 39). There are two choices for a minimum cardinality minimum weight vertex cover, here of cardinality three. Namely, both covers contain the two segments highlighted in red, and differ by whether the cover contains segment $\{(40, 52)\}$ or segment $\{(41, 54)\}$. Removing either cover yields a chord diagram with no crossings, and finding the secondary structures, we may obtain two different pseudoknot types depending on the cover removed. If we remove the cover containing segment $\{(40, 52)\}$, then the pseudoknot corresponding to $\{(40, 52)\}$ is a pseudoknot connecting a bulge to a bulge. However if instead we remove the cover containing segment $\{(41, 54)\}$ then the pseudoknot corresponding to $\{(41, 54)\}$ is a pseudoknot connecting an interior loop to another interior loop. As a result, the two choices for a vertex cover have a different effect on the secondary structure classification and consequently pseudoknot typing.

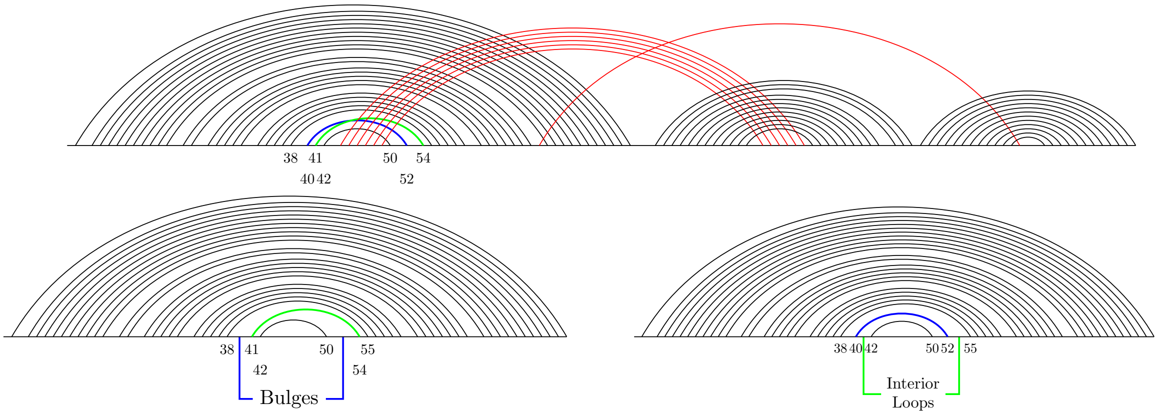


Fig. 39. RNA structure with bpRNA reference name PDB_455 and PDB reference name 3DIG. Structures such as nestings not relevant for the discussion have been omitted for clarity.

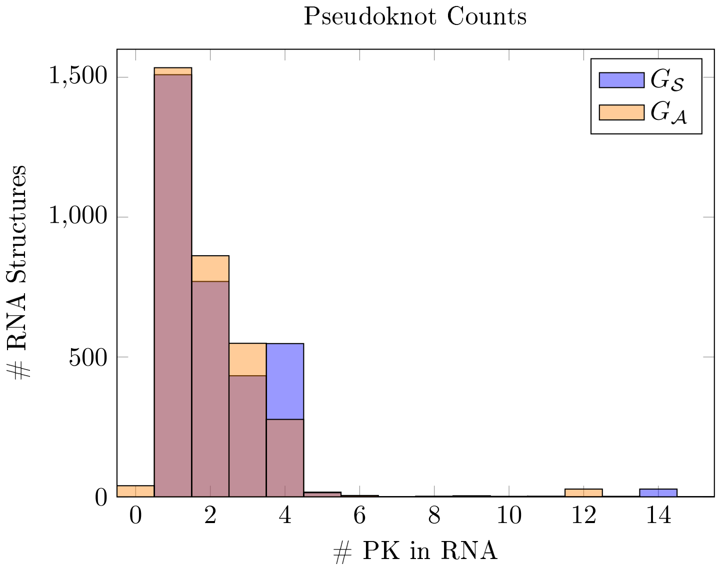
4.5 Analysis

We first applied the τ-Segment Graph Method with $\tau = 0$ to independently verify the quantities reported in [44]. By Definition 4.15, structures are pseudoknotted when their segment graphs contain at least one edge, and the number of pseudoknots is quantified by the minimum cardinalities of vertex covers of minimum weight. In agreement with [44], we obtained 3,320 graphs containing at least one edge in $G_{\mathcal{S}}$ from RNA structures in bpRNA-1m(90) and a total quantity of 7164 pseudoknots, as determined by the sum over the cardinalities of the vertex covers.

We next applied the τ-Segment Graph Method at $\tau = \infty$ (the augmented segment graph method) for every structure in bpRNA-1m(90), and found a minimum vertex cover for each one. The number of pseudoknots (the sum of vertex cover numbers over all graphs) was 6548 with this method.

4.5.1 Structures With the Most Pseudoknots

Applying the method with $\tau = 0$, we found 31 unique RNA structures containing 13 or more pseudoknots. These structures are listed in Table 1. When the method was applied with $\tau = \infty$, we found that the same 31 structures contained the most pseudoknots amongst all structures in the database. With the exception of the last



# PK	$G_{\mathcal{S}}$	$G_{\mathcal{A}}$
0	0	40
1	1509	1534
2	770	862
3	433	549
4	548	277
5	17	15
6	3	5
7	0	1
8	2	2
9	2	4
10	2	0
11	2	2
12	1	28
13	2	1
14	28	0
15	1	0

Fig. 40. A comparison of total number of pseudoknots between the segment graph method ($\tau = 0$) and augmented segment graph method ($\tau = \infty$). For each method, the number of structures containing r pseudoknots is displayed for each row r of the table.

structure, *Oceanobacillus iheyensis*, from $\tau = 0$ to $\tau = \infty$ there was a uniform decrease in pseudoknotting by two that resulted from a single 3-crossing being consolidated into one segment by τ-reduction. This uniformity in behavior is explained by the fact that all but the last structure are of type 23s prokaryotic ribosomal RNA, originating in various bacterial organisms [49].

4.5.2 Notable Decreases in Pseudoknot Quantity

Over the entire bpRNA-1m(90) database, a total of 573 structures had a decrease in numbers of pseudoknots when analyzed with the $\tau = \infty$ versus $\tau = 0$ methods. Of these, 531 structures decreased in quantity of pseudoknots by 1, 41 structures by 2, and 1 structure decreased by 3 (bpRNA_CRW_55316, *Plasmodium falciparum*). Of the 41 structures which decreased by 2, there were 6 unique RNA types with 36 of them being of type 23S.

Structures that changed from having a nonzero quantity of pseudoknots to zero

ID	Domain	Organism	$G_{\mathcal{S}}$	$G_{\mathcal{A}}$
CRW_55315	Eukaryota	Euglena gracilis	15	13
CRW_55268	Bacteria	Acinetobacter calcoaceticus	14	12
CRW_55269	Bacteria	Aeromonas hydrophila	14	12
CRW_55271	Bacteria	Bartonella bacilliformis	14	12
CRW_55275	Bacteria	Burkholderia mallei	14	12
CRW_55276	Bacteria	Bordetella pertussis	14	12
CRW_55279	Bacteria	Clostridium botulinum B	14	12
CRW_55283	Bacteria	Citrobacter freundii	14	12
CRW_55284	Bacteria	Campylobacter jejuni	14	12
CRW_55285	Bacteria	Chlamydophila psittaci 6BC	14	12
CRW_55287	Bacteria	Deinococcus radiodurans	14	12
CRW_55290	Bacteria	Enterococcus faecalis	14	12
CRW_55291	Bacteria	Erysipelothrix rhusiopathiae (str. 715)	14	12
CRW_55292	Bacteria	Haemophilus influenzae (operons A-F)	14	12
CRW_55295	Bacteria	Leptospira interrogans	14	12
CRW_55296	Bacteria	Lactococcus lactis	14	12
CRW_55297	Bacteria	Listeria monocytogenes	14	12
CRW_55298	Bacteria	Listeria monocytogenes	14	12
CRW_55299	Bacteria	Mycoplasma genitalium	14	12
CRW_55303	Bacteria	Neisseria gonorrhoeae	14	12
CRW_55305	Bacteria	Pseudomonas aeruginosa	14	12
CRW_55306	Bacteria	Plesiomonas shigelloides	14	12
CRW_55307	Bacteria	Ruminobacter amylophilus	14	12
CRW_55308	Bacteria	Rickettsia prowazekii (str. Madrid E)	14	12
CRW_55312	Bacteria	Staphylococcus carnosus	14	12
CRW_55313	Bacteria	Thermotoga maritima	14	12
CRW_55314	Eukaryota	Chlamydomonas reinhardtii	14	12
CRW_55317	Eukaryota	Spinacia oleracea	14	12
CRW_55338	Eukaryota	Cyanophora paradoxa	14	12
CRW_55270	Bacteria	Bacillus anthracis	13	11
PDB_647	Bacteria	Oceanobacillus iheyensis	13	11

Table 1. The 31 RNA structures with at least 13 pseudoknots when analyzed with the segment graph method. The rightmost column compares the number of pseudoknots in each structure via the τ-segment graph method with $\tau = \infty$. All structures with the exception of PDB_647 are RNA type 23S ribosomal RNA.

pseudoknots are shown in Table 2. Of these, one structure (*Homo sapiens*) decreased from 2 to 0 pseudoknots. All other structures decreased from 1 to 0 pseudoknots.

4.5.3 Maximum Values of τ and Persistence of Partitions

Let $\tau \geq 1$. As distances between chords are finite, there is a minimum value of τ, say τ_m, such that for any $\tau_* \geq \tau_m$ the τ_*-segment partition and the τ_m-segment partition are identical. The quantity τ_m is precisely the minimum value of τ such that the τ-segment partition is equivalent to the augmented segment partition. For all structures in bpRNA-1m(90), we calculate τ_m by first finding the augmented segment partition, and then finding the τ-segment partition for each $\tau > 0$ until the τ-segment partition is equal to the augmented segment partition. We find that the average τ_m is 13.035 and the median is 8. The mean absolute deviation is 10.96 and the median absolute deviation is 2. There are 323 structures with τ_m at least 17, and 33 structures with τ_m at least 100.

ID	Domain	Organism	Length	Method
CRW_1213	Bacteria	Actinomyces israelii	734	CSA
CRW_1219	Bacteria	Actinomyces israelii	1145	CSA
CRW_1563	Bacteria	Clavibacter sp. R1_2_cr	476	CSA
CRW_1725	Bacteria	Arthrobacter sp.	300	CSA
CRW_17723	Bacteria	Lachnospira multipara	977	CSA
CRW_17729	Bacteria	Moorella thermoautotrophica	869	CSA
CRW_17730	Bacteria	Moorella thermoautotrophica	821	CSA
CRW_17811	Bacteria	Thermoanaerobacter acetoethylicus	770	CSA
CRW_17823	Bacteria	Thermoanaerobacter ethanolicus	650	CSA
CRW_17834	Bacteria	Thermoanaerobacterium thermosulfurigenes	930	CSA
CRW_20267	Bacteria	Marigold phyllody phytoplasma	1015	CSA
CRW_20554	Bacteria	Mycoplasma collis	372	CSA
CRW_20606	Bacteria	Beet leafhopper transmitted virescence phytoplasma	700	CSA
CRW_20626	Bacteria	Potato witches'-broom phytoplasma	658	CSA
CRW_20629	Bacteria	Paulownia witches'-broom phytoplasma	698	CSA
CRW_3719	Bacteria	Actinomycetales sp.	472	CSA
CRW_3726	Bacteria	Actinomycetales sp.	503	CSA
CRW_3729	Bacteria	Actinomycetales sp.	502	CSA
CRW_3732	Bacteria	Actinomycetales sp.	478	CSA
CRW_4109	Bacteria	Mycobacterium xenopi	942	CSA
CRW_4363	Bacteria	Streptomyces abikoensis	1177	CSA
		⋮		
CRW_4401	Bacteria	Streptomyces mobaraensis	1197	CSA
CRW_4409	Bacteria	Streptomyces olivoreticuli	1216	CSA
CRW_4416	Bacteria	Streptomyces salmonis	1136	CSA
CRW_4449	Bacteria	coryneform actinomycete B755	679	CSA
CRW_4908	Bacteria	Acidocella facilis	922	CSA
CRW_4910	Bacteria	Acidiphilium angustum	977	CSA
CRW_4918	Bacteria	Acidiphilium sp.	944	CSA
CRW_7455	Bacteria	unidentified eubacterium 37SW-1	277	CSA
CRW_7488	Bacteria	Proteobacteria sp	484	CSA
CRW_7494	Bacteria	uncultured alpha proteobacterium	410	CSA
CRW_7502	Bacteria	uncultured alpha proteobacterium	222	CSA
CRW_7614	Bacteria	Nitrobacter sp.	452	CSA
CRW_7802	Bacteria	Rhodovulum euryhalinum	1138	CSA
CRW_7938	Bacteria	Sphingomonas asaccharolytica	629	CSA
CRW_8046	Bacteria	uncultured alpha proteobacterium	751	CSA
CRW_8048	Bacteria	uncultured alpha proteobacterium	730	CSA
CRW_8050	Bacteria	uncultured alpha proteobacterium	690	CSA
PDB_567	artificial sequences	synthetic construct	35	X-RAY
PDB_512	Eukaryota	Homo sapiens	12	X-RAY

Table 2. The 40 RNA structures with nonzero quantity of pseudoknots when analyzed with the segment graph method but zero pseudoknots using $\tau = \infty$ segment graph method. One structure (*Homo sapiens*) decreased from 2 to 0 pseudoknots. All other structures decreased from 1 to 0 pseudoknots.

Structures with large τ_m contain correspondingly large bulges and internal loops; large τ_m results from large gaps between chords which are nested but not τ-near for many values of τ. We verify this by keeping track of τ-segment partitions during the process of calculating τ_m. In each iteration, we log whether the RNA structure's τ-segment partition is equal to its $(\tau+1)$-segment partition by appending a 1 or 0 to a τ-diff string. Structures with large τ_m have long substrings of consecutive 1s.

Example 4.29 (τ_m Analysis)**.** Taking the structure in Figure 41, we obtain the following τ-diff list:

$$0100111111011111111111111111111111111111111111111111111111$$

$$1111111111111111111111111111111111111111111111111111111110.$$

From this we produce the array

$$[0, \ell_1, 0, 0, \ell_2, 0, \ell_3, 0] = [0, 1, 0, 0, 6, 0, 104, 0],$$

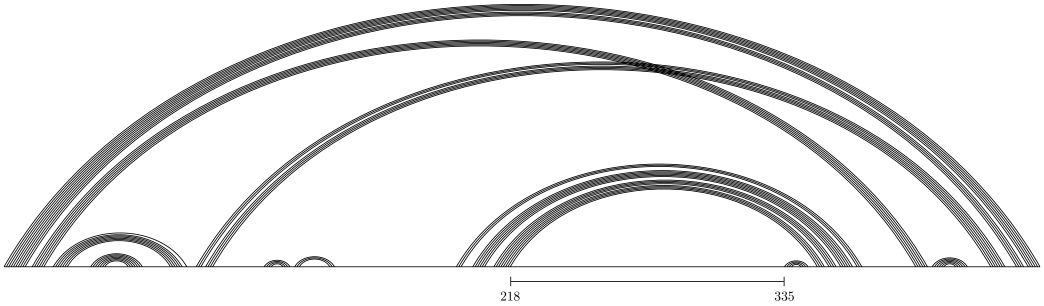


Fig. 41. RNA structure bpRNA_RFAM_4761, displaying an interior loop with one part of length $335 - 218 = 117$, which causes a large value of τ_m.

where each ℓ_i indicates the persistence of the τ-segment partition where τ is equal to the position of ℓ_i in the array. The number of 0s plus 1 is the number of unique τ-segment partitions. In this case, these are the τ-segments for $\tau = 1, 2, 4, 5, 12, 117$, with the $\tau = 2$ partition persisting for one value of τ, the $\tau = 5$ partition persisting for 6 values of τ, and the $\tau = 12$ partition persisting for 104 values of τ. The value of τ_m here is 117.

In sum, persistent τ-segment partitions are indicative of large bulges and internal loops.

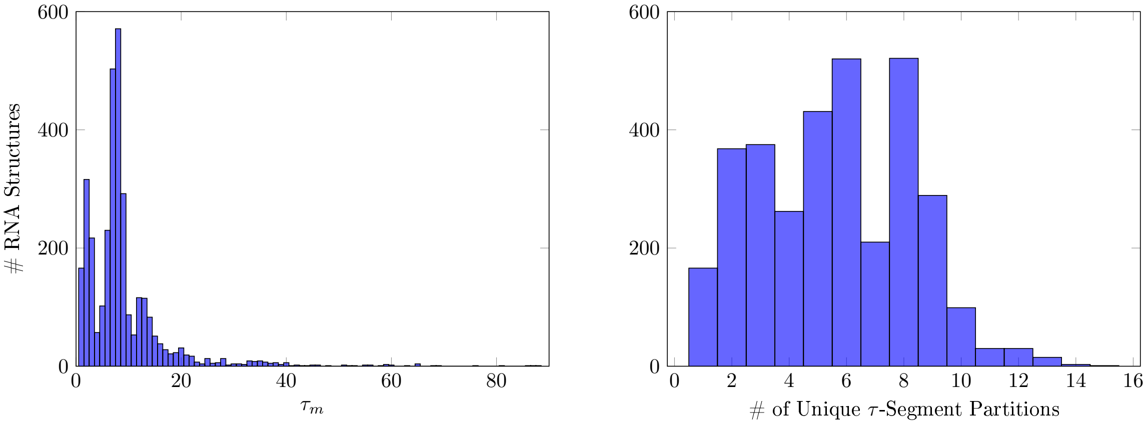
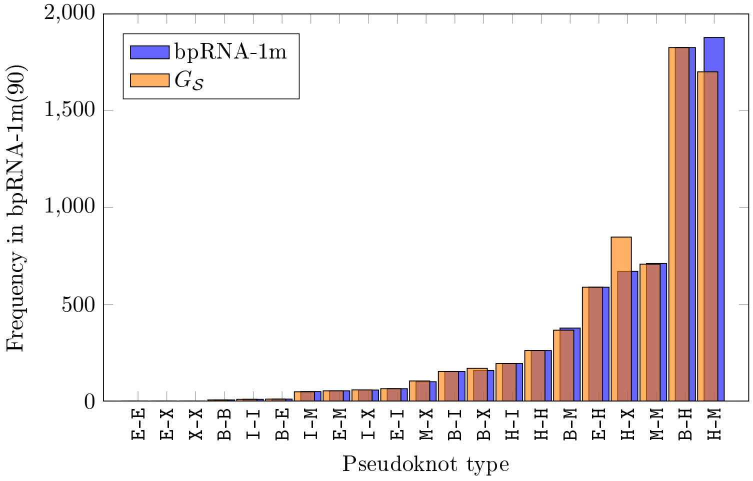


Fig. 42. Left: Distribution of τ_m over all RNA structures in bpRNA-1m(90) restricted to values of τ_m within one and a half standard deviations from the mean. Right: Distribution of number of unique τ-segment partitions.

4.5.4 Classifying Bases

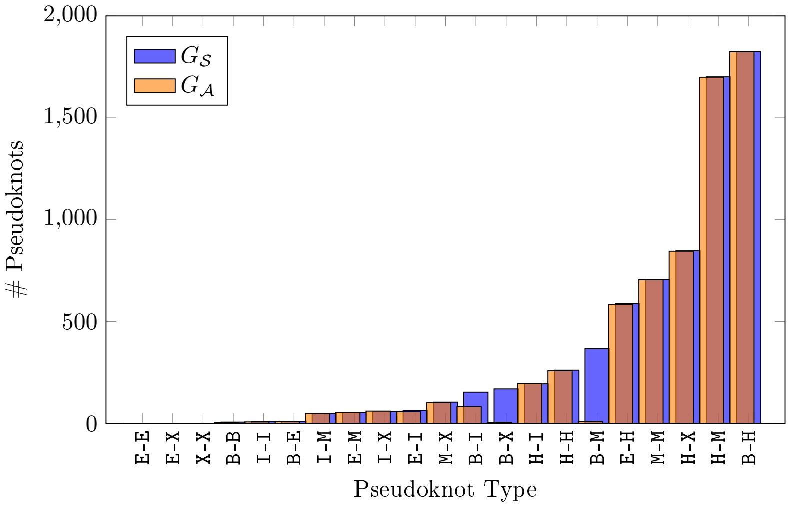
We implement the `classifyBases` routine with the $\tau = 0$ segment method to analyze secondary structures and compare with Figure 8B from [44]. By labelling and removing pseudoknots as done in [44], we obtain crossingless chord diagrams. Unpaired bases are labelled as one of several secondary structures in Definition 4.17, and remaining paired bases as stems. We observe slight discrepancies in pseudoknot type counts, though the general shape of the distribution is the same. Original pseudoknot type counts reported in [44] were obtained directly from the bpRNA-1m database.



PK type	bpRNA-1m	$G_\mathcal{S}$
E-E	0	0
E-X	0	0
X-X	0	0
B-B	6	6
I-I	9	9
B-E	10	10
I-M	49	48
E-M	53	53
I-X	57	58
E-I	64	64
M-X	100	104
B-I	153	153
B-X	158	169
H-I	194	194
H-H	261	261
B-M	377	366
E-H	588	588
H-X	670	847
M-M	711	707
B-H	1826	1826
H-M	1878	1701

Fig. 43. A comparison of counts of pseudoknot types. The bpRNA-1m (blue) data is reported in [44], while the $G_\mathcal{S}$ (orange) are counts found using the same method with our secondary structure definitions.

A discrepancy in the $\tau = 0$ method arises from the labelling of multiloops and external loops, as some structures in [44] have bases in the external loop being labelled as a multiloop base. This is a bpRNA software bug, and has since been fixed in a fork of [43]. The structure bpRNA_CRW_10025 is one example in which the secondary structure labelling is incorrect; bases 357, 385-393, 433-440, 604-633, 690-695, and 728-742 are labelled as part of a multiloop in bpRNA-1m, but by our definition they are part of the exterior loop. This is supported by the fact that counts of



PK type	$G_{\mathcal{S}}$	$G_{\mathcal{A}}$
E-E	0	0
E-X	0	0
X-X	0	0
B-B	6	5
I-I	9	7
B-E	10	8
I-M	48	48
E-M	53	54
I-X	58	60
E-I	64	57
M-X	104	102
B-I	153	82
B-X	169	5
H-I	194	196
H-H	261	258
B-M	366	9
E-H	588	584
M-M	707	705
H-X	847	845
H-M	1701	1699
B-H	1826	1824

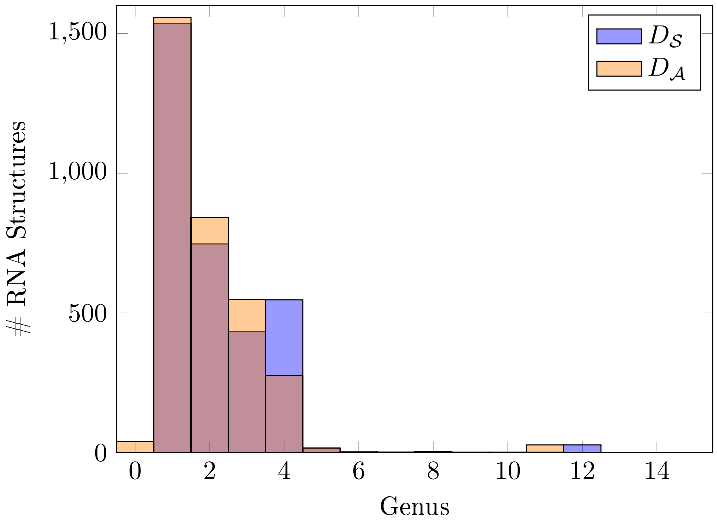
Fig. 44. A comparison of pseudoknot types arising from the $\tau = 0$ method and $\tau = \infty$ method.

pseudoknot types differ in Figure 43 only when an exterior loop or multiloop is part of the pseudoknot type.

Figure 44 (left) shows a comparison of pseudoknot type counts between the $\tau = 0$ method and the $\tau = \infty$ method. Figure 44 (right) shows the same comparison, except with a slightly modified $\tau = \infty$ method. In the modified method, the $\tau = \infty$ segment graph is weighted by the cardinality of the augmented segments, as in the $\tau = 0$ method. This gives us a weighted augmented segment graph $G^*_{\mathcal{A}}$. Instead of taking a minimum cardinality vertex cover with respect to the lexicographic order, we look over all minimum cardinality vertex covers and choose one with minimum weight (again, with respect to lexicographic order). In doing this, we obtain a more comparable distribution to that of the $\tau = 0$ method.

4.5.5 Calculation of Genus and Clique Numbers

After implementing the τ-segment graph method with $\tau = 0$ and $\tau = \infty$ we calculate the genus and maximum clique numbers of $D_{\mathcal{S}}$ and $D_{\mathcal{A}}$ and the segment graphs $G_{\mathcal{S}}$ and $G_{\mathcal{A}}$ respectively in MATLAB. The results are reported in Figure 45, Table 3, and Figure 46.



Genus	$D_\mathcal{S}$	$D_\mathcal{A}$
0	0	40
1	1536	1558
2	747	841
3	434	548
4	547	277
5	15	17
6	3	2
7	2	2
8	2	4
9	2	0
10	1	2
11	2	28
12	28	1
13	1	0

Fig. 45. A comparison of the genus of the segment graph and the augmented segment graph.

$\omega(G)$	$G_\mathcal{S}$	$G_\mathcal{A}$
1	0	40
2	3214	3177
3	60	96
4	46	7

Table 3. The clique number for each segment and augmented segment graph.

Out of 3320 segment graphs, 3208 are forests, and out of 3320 augmented segment graphs, 3210 are forests. Table 4 shows the frequency of forests with a given maximum tree size. This is important to note in the context of using genus and vertex covers for pseudoknot quantification. By Theorem 4.13, if the intersection graph of a chord diagram D is a forest F, then the genus $\gamma(D)$ is equal to $\beta(F)$. From this, we see that the genus of a corresponding chord diagram of an RNA structure is a robust quantifier of pseudoknot complexity. This is further supported by the bubble charts in Figure 46.

Max Tree Order	$G_{\mathcal{S}}$	$G_{\mathcal{A}}$
1	0	40
2	1449	1410
3	998	999
4	637	637
5	27	27
6	71	71
7	6	6
8	7	7
9	7	7
10	1	1
11	3	3
12	2	2

Table 4. For each segment and augmented segment graph that is a forest, we calculate the maximum order of a tree component.

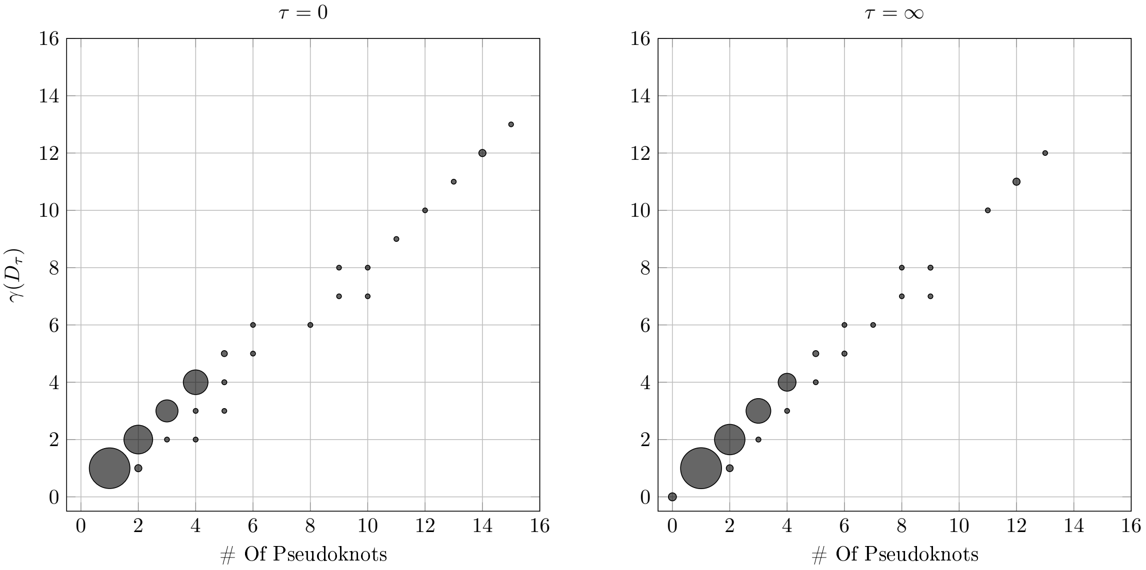


Fig. 46. Two bubble chart comparisons. Left: Comparison between genera of the chord diagrams $D_{\mathcal{S}}$ and pseudoknot count via the $\tau = 0$ method. Right: Comparison between genera of the chord diagrams $D_{\mathcal{A}}$ and pseudoknot count via the $\tau = \infty$ method.

CHAPTER 5

DETERMINANTS OF SIMPLE THETA CURVES

This chapter is joint work with Elpers and Moore [50].

5.1 Introduction

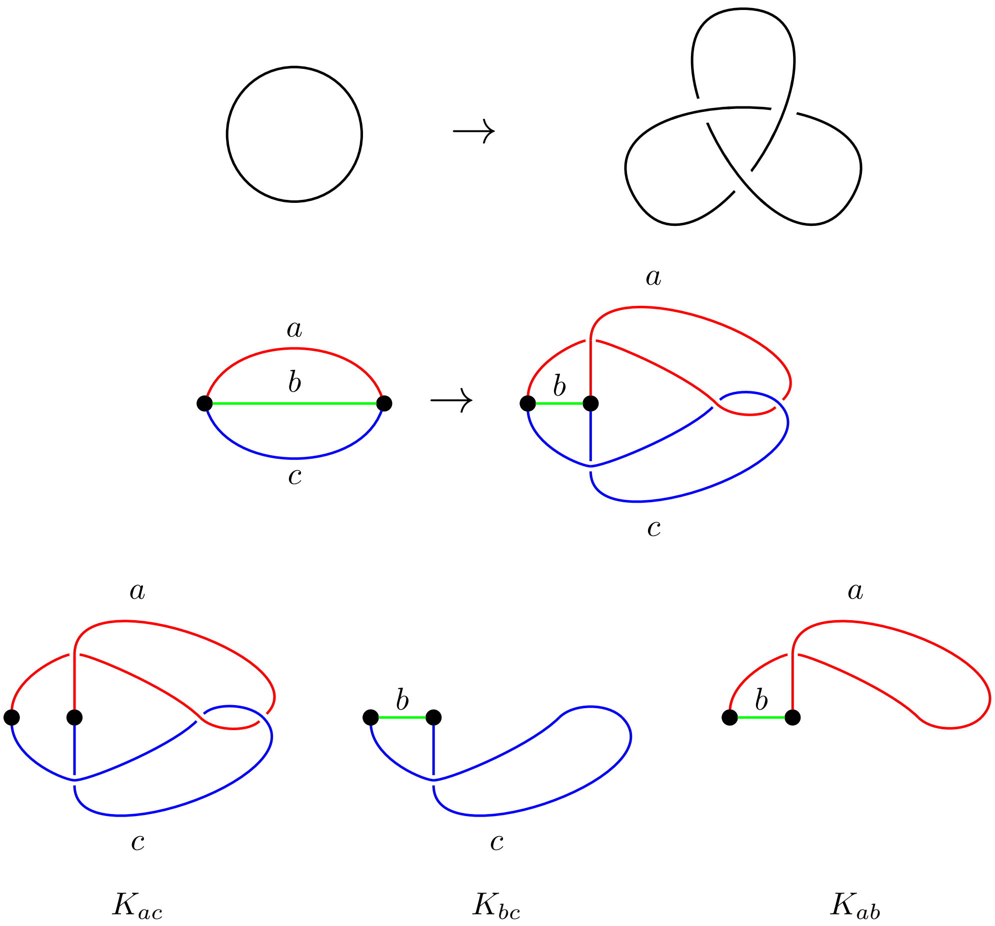
A *theta curve* ϑ is an embedding of the θ-graph in the three-sphere, up to equivalence by ambient isotopy. The θ-graph is the unique abstract graph consisting of two vertices connected by three parallel edges. Theta curves and other spatial graphs are generalizations of knots and links. In this article we study an integer-valued invariant of theta curves that we call the determinant $\det(\vartheta)$. Like the well-known determinant of links, this invariant can be defined as the order of the torsion subgroup of the first homology of a certain branched covering space (see Definition 5.1). Every theta curve contains three *constituent knots* K_{ij}, formed by taking pairs of edges $i, j \in \{a, b, c\}$. A *simple* theta curve is one which contains at least one constituent knot that is unknotted [114]. For example, among prime theta curves of up to seven crossings, all 90 in the Litherland-Moriuchi table [87, 75] are simple (see Table 5).

The relationship between $\det(\vartheta)$ for simple theta curves and the determinants of its constituent knots is described by the following statement.

Theorem 5.13. Let ϑ be a simple theta curve with constituent knots K_{ab}, K_{ac}, K_{bc}. Then

$$\det(\vartheta) = \det(K_{ab}) \cdot \det(K_{ac}) \cdot \det(K_{bc}).$$

Our methods for proving Theorem 5.13 come from graph theory. By assumption one constituent, say K_{ac}, is an unknot with $\det(K_{ac}) = 1$. We relate the determinants of the other two constituent knots with counts of weighted spanning trees of Tait graphs that are derived from a diagram of ϑ. As it will turn out, the determinants of constituent knots of a theta curve provide a geometric interpretation of a purely graph-theoretic spanning tree enumeration formula. More specifically, Ciucu, Yan, and



K_{ac} K_{bc} K_{ab}

Fig. 47. A knot is an embedding of the circle and a θ-curve is an embedding of the θ-graph. The three constituent knots of this particular θ-curve are the figure eight knot and two unknots.

Zhang [39] and Zhang and Yan [123] applied the Matrix Tree Theorem to enumerate the spanning trees of a graph admitting an involutive symmetry via a product formula involving two smaller graphs. In our context, we realize a graph that admits an involutive symmetry as the Tait graph of a strongly invertible knot (see Section 5.2) corresponding to the theta curve. The Goeritz matrix plays the role of the graph Laplacian, the determinant of which calculates the tree weight. We explicitly identify the two factors in the spanning tree enumeration formula with the determinants of the constituent knots of the theta curve, as realized by their Tait graphs.

5.2 Knots and Spatial Theta Curves

We consider knots and theta curves to be smoothly embedded in the three-sphere, up to equivalence by ambient isotopy. Label edges of a theta curve by the letters $\{a, b, c\}$, which may be thought of as non-identity elements of the Klein group $V \cong \mathbb{Z}_2 \times \mathbb{Z}_2$. A theta curve is a special type of *Klein graph* (see [53]), meaning a trivalent spatial graph endowed with a 3-edge coloring. Theta curves are also 3-Hamiltonian, meaning all of its $\{i, j\}$-colored subgraphs are connected. This means its constituents are knots, rather than links.

Recall that the cyclic double cover $\widehat{X}_2$ of the complement of a knot, $X = S^3 - N(K)$, is the regular covering space corresponding with the kernel of a homomorphism $\pi_1(X, x) \to H_1(X; \mathbb{Z}) \to \mathbb{Z} \to \mathbb{Z}_2$. The branched double cover $\Sigma_2(S^3, K)$ may be obtained by gluing a solid torus to the boundary of $\widehat{X}$ via the map $(z_1, z_2) \mapsto (z_1, z_2^2)$ to extend the covering to a branched covering map $\Sigma_2(S^3, K) \to S^3$. It is a standard fact of knot theory that the branched double cover of a knot is a rational homology sphere and the determinant of a knot may be defined by $\det(K) := |H_1(\Sigma_2(S^3, K); \mathbb{Z})|$. See for reference [74, Chapter 7].

Given a theta curve $\vartheta \in S^3$ and complement $Y = S^3 - N(\vartheta)$, we may similarly construct a covering space corresponding to the map $\pi_1(Y, x) \to H_1(Y; \mathbb{Z}) \to \mathbb{Z}_2 \times \mathbb{Z}_2$. This can be completed to a closed, oriented 3-manifold acted on by V by gluing solid cylinders and cubes to the boundary in a procedure explicitly described by Gille and Robert in [53, Proposition 2.6]. This manifold is the *Klein cover* $\Sigma_\vartheta := \Sigma(S^3, \vartheta)$ and has ϑ as the branching locus.

Definition 5.1. Let Σ_ϑ denote the Klein cover of a theta curve in S^3. The *determinant* $\det(\vartheta)$ of ϑ is the order of $H_1(\Sigma_\vartheta; \mathbb{Z})$.

One may visualize the Klein cover by iterating the branched double cover construction. One first constructs $\Sigma_2(S^3, K_{ac})$, branched over one of the constituent knots $K_{ac} = e_a \cup e_c$, then constructs a second branched covering of the manifold $\Sigma_2(S^3, K_{ac})$ branched over the knot $\tilde{e}_b$ that is the lift of the edge e_b. This will also yield $\Sigma_\vartheta \cong \Sigma_2(\Sigma_2(S^3, K_{ac}), \tilde{e}_b)$. The Klein cover of ϑ is unique, and so the order of a, b, c in this procedure does not matter. A proof that the Klein cover of a spatial

Klein graph in S^3 is unique up to diffeomorphism may be found in [53, Proposition 2.8]. Consequently, the determinant $\det(\vartheta)$ is a well-defined integer invariant of theta curves in the three-sphere.

5.2.1 Simple Thetas and Strongly Invertible Knots

Consider the case that ϑ is simple. Up to relabeling, we may assume K_{ab} is an unknot. Then ϑ, together with this unknotted constituent, corresponds with a strongly invertible knot in the three-sphere as follows. The branched cover $\Sigma_2(S^3, K_{ac})$ is diffeomorphic to S^3, and the lift $\tilde{e}_b$ consists of two pre-images of e_b joined at the two vertices of ϑ on the branching set. Recall that a knot K in S^3 is *strongly invertible* if there is an orientation-preserving involution h on S^3 such that $h(K) = K$ and $\mathrm{Fix}(h)$ is a circle intersecting K in two points [100]. In our context, $\tilde{e}_b$ is strongly invertible. We write $\tilde{e}_b = (K, h)$ to emphasize the involution. For the reverse correspondence, let (K, h) be any strongly invertible knot in the three-sphere. As a consequence of the Smith conjecture, $\mathrm{Fix}(h)$ is unknotted and by definition, (K, h) intersects $\mathrm{Fix}(h)$ in two points. The quotient K/h is an embedded closed arc. Thus $\mathrm{Fix}(h) \cup K/h = e_a \cup e_b \cup e_c$ is a simple theta curve.

Recall that the branched double cover of any knot in S^3 is a rational homology sphere with first homology of odd order. Thus in the case ϑ is simple, $\det(\vartheta) = \det(K, h)$ is an odd integer. Note also that in the quotient under the action of the involution, a right-handed (respectively, left-handed) crossing in (K, h) descends to a right-handed clasp in ϑ, as in Figure 48. We will make use of this observation later.

5.2.2 Goeritz Matrices

The determinant of a knot or link can be calculated combinatorially as the determinant of an integral matrix associated to a knot diagram, due to a well-known construction of Goeritz and Trotter [54, 113] We review this following [74, Chapter 9] and apply it to simple theta curves below. Let K be a knot with diagram D_K. Then D_K admits two checkerboard colorings φ of the regions $X = \{X_0, \ldots, X_m\}$ of D_K, that is, there are two assignments $\varphi : X \to \{0, 1\}$ where $\varphi(X_i) \neq \varphi(X_j)$ when

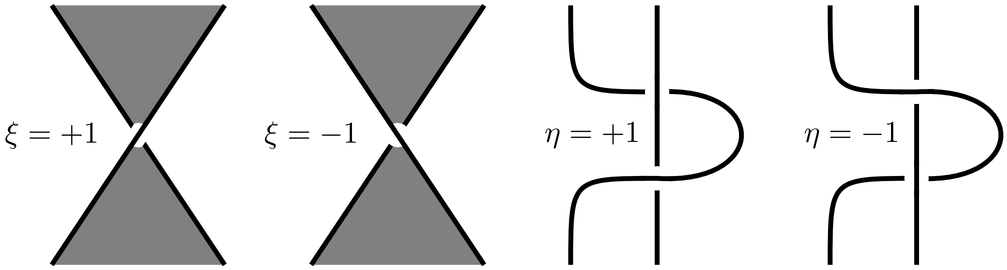


Fig. 48. Sign conventions for incidence numbers of crossings in a checkerboard shading (left) and for clasps in the quotient theta curve (right). The shading indicates regions X_i colored by an assignment $\varphi(X_i) = 0$. Both ξ and η are independent of strand orientation.

X_i and X_j share a boundary curve.

To each crossing c of D_K we associate a sign $\xi(c)$ with the convention in Figure 48. Let $\{X_i \mid \varphi(X_i) = 0\} = \{B_0, \dots, B_n\}$ and let C_{ij} be the set of crossings where B_i and B_j meet. We may then associate an $(n+1) \times (n+1)$ matrix $\widetilde{Q}_{D_K}$ to the diagram D_K of K with respect to the choice in shading φ. The matrix $\widetilde{Q}_{D_K} = [q_{ij}]$ is defined by

$$q_{ij} = \begin{cases} -\sum\limits_{c \in C_{ij}} \xi(c), & \text{if } i \neq j \\ \\ -\sum\limits_{k \neq i} q_{ik}, & \text{if } i = j. \end{cases}$$

From this, $\det(K) = |\det(Q_{D_K})|$ where the *Goeritz matrix* Q_{D_K} is the $n \times n$ matrix obtained from $\widetilde{Q}_{D_K}$ by deleting any row and column. The result is independent of the choices in the knot diagram, the checkerboard coloring, labelling of regions, and the row and column selected for deletion.

5.3 Symmetric Weighted Graphs

Let $G = (V(G), E(G))$ denote a graph and its vertex and edge sets. We will assume that graphs are undirected, but permit multi-edges, self-loops and edge weights. Recall that a *spanning tree* $T \subseteq G$ is a connected acyclic subgraph with $V(T) = V(G)$. For a graph G endowed with edge weights $\omega(e)$, define $\omega(G) \coloneqq \prod_{e \in E(G)} \omega(e)$. Define the *tree weight* of G by

$$\tau(G) = \sum_{T \subseteq G} \omega(T) = \sum_{T \subseteq G} \prod_{e \in E(T)} \omega(e), \tag{5.1}$$

where the sum is over spanning trees of G. For graphs with edge weights all equal one, $\tau(G)$ is simply the number of spanning trees of G. We restrict our attention to edge weights $\omega(e)$ in the multiplicative group $\{-1, +1\}$, and so tree weights will take on integer values. There is a well known method to count spanning trees, or more generally the tree weight, using the graph Laplacian.

Definition 5.2. Let G be a weighted graph with $V(G) = \{v_0, \ldots, v_n\}$. An $(n+1) \times (n+1)$ matrix called the *Laplacian* of G, $\widetilde{L}_G = [\ell_{ij}]$, is defined by

$$\ell_{ij} = \begin{cases} -\omega_{ij} & \text{if } i \neq j \\ -\sum_{k \neq i} \omega(e_{ik}) & \text{if } i = j. \end{cases} \tag{5.2}$$

Here, ω_{ij} is the sum of edge weights over all edges connecting v_i and v_j.

The following theorem is often attributed to Kirchhoff, and different versions are due to Bott-Mayberry [17] and Tutte [116]. See also the exposition in [13].

Theorem 5.3 (Matrix Tree Theorem)**.** *Let G be a graph and let L be the reduced Laplacian of G, obtained by deleting any row and column from $\tilde{L}$. Then $|\det(L)| = \tau(G)$.*

5.3.1 Tait Graphs

The procedure for calculating the determinant from a Goeritz matrix yields an equivalent graph theoretic method using the *Tait graph* of a knot diagram. Let φ be a checkerboard coloring of a diagram D_K of K. The fact that there exists a checkerboard coloring φ for any D_K can be proven in the following way: By forgetting crossing information, a knot diagram yields a planar four-valent graph G. Observe that the dual $G^{\perp}$ cannot contain any odd cycles, otherwise G would contain a vertex of odd degree. Therefore $G^{\perp}$ is bipartite, so the faces of G are two-colorable.

Thus, the coloring determines a pair of planar dual graphs G and $G^{\perp}$. The vertices $V(G)$ correspond with the shaded regions $B = \{X_i \mid \varphi(X_i) = 0\}$ and the vertices $V(G^{\perp})$ with unshaded regions $W = \{X_i \mid \varphi(X_i) = 1\}$. Edges in both graphs correspond to incidences between regions at crossings, with edge weight $\omega(e) = \omega(e^{\perp}) = \xi(c)$. Examples of Tait graphs are shown in Figure 51.

Combining the Matrix Tree Theorem and the Goeritz matrix formulation of the determinant of a knot, we have

$$\det(K) = |\det(Q_{D_K})| = |\det(L_G)| = \tau(G)$$

where Q_{D_K} is the reduced Goeritz matrix of K corresponding with any diagram D_K of K, L_G is the reduced Laplacian of the corresponding Tait graph G, and where $\tau(G)$ is tree weight of G.

We now collect several lemmas pertaining to signed graphs that we will need in later sections.

It is a standard result in graph theory that the number of spanning trees of a planar graph G is equal to the number of spanning trees of $G^{\perp}$. For graphs with arbitrary edge weights, this statement is false; for a counterexample, take a triangle with edge weights 1, 2, 3. For edge weights in the multiplicative group $\{-1, +1\}$, though, the statement generalizes as follows.

Lemma 5.4. *For planar graphs with edge weights in* $\{-1, +1\}$, $|\tau(G)| = |\tau(G^{\perp})|$.

Proof. Let G be a planar graph and $G^{\perp}$ its dual. There is a bijection $E(G) \to E(G^{\perp})$ which sends $e \in E(G)$ to $e^{\perp} \in E(G^{\perp})$. In particular, every edge e borders faces F_1 and F_2 and $e^{\perp} = (F_1, F_2) \in E(G^{\perp})$ is the edge uniquely corresponding to e. As a result, there is a bijection f that associates each spanning tree $T \subseteq G$ with a spanning tree $f(T) \subseteq G^{\perp}$, where

$$f(T) = G^{\perp} - \{e^{\perp} : e \in T\} = (G - T)^{\perp}.$$

See for example [76]. We extend f to weighted trees by assigning dual edges the same weight, i.e., $\omega(e) = \omega(e^{\perp})$.

By assumption $\omega(e) \in \{-1, 1\}$ for all $e \in E(G)$. Recall that $\omega(G) := \prod_{e \in E(G)} \omega(e)$.

From f we can deduce that for any spanning tree $T \subseteq G$,

$$\omega(G) = \tau(T)\tau(f(T))$$

and there are two cases.

Case 1. $\omega(G) = 1$. Then it must be the case that for all spanning trees T, $\tau(T) = \tau(f(T)) = 1$ or $\tau(T) = \tau(f(T)) = -1$. That is, for all spanning trees T, $\tau(T) = \tau(f(T))$.

Case 2. $\omega(G) = -1$. Then it must be the case that for all spanning trees T, $\tau(T) = 1$ and $\tau(f(T)) = -1$ or $\tau(T) = -1$ and $\tau(f(T)) = 1$. That is, for all spanning trees T, $\tau(T) = -\tau(f(T))$.

So either $\tau(G) = \tau(G^{\perp})$ or $\tau(G) = -\tau(G^{\perp})$. □

Lemma 5.5. *Let G be a graph containing an edge $e = (v, w)$ of weight $\omega(e)$. Let G' be $G - \{e\} \cup \{e_1, e_2\}$ where $e_1 = (v, w) = e_2$ are edges of weight $\frac{1}{2}\omega(e)$. Then $\tau(G) = \tau(G')$.*

Proof. For every spanning tree T in G that contains e there exist exactly two spanning trees in G', each of tree weight $\frac{1}{2}\omega(T)$. □

A version of the following lemma is proved in [39, Lemma 6] for unweighted graphs. Here, we are interested in counting spanning trees where the graphs inherit edge weights from the crossings of knot diagrams, and edge subdivisions will occur in the Tait graphs of our constituent knots. Hence, we extend their lemma to the specific case of graphs with edge weights $\omega(e) \in \{-1, +1\}$.

Lemma 5.6. *Let G_0 be a graph with vertices V_0 and edges E_0. Let a, b and x be three vertices distinct from V_0. Construct a graph $G = (V, E)$ by taking $V = V_0 \cup \{a, b\}$, and letting $E = E_0 \cup (a, b) \cup S$, where S is any set of edges of the form (v, a) or (v, b), where $v \in V_0$; specify the edge weight of $\omega(a, b) = \pm 1/2$. Construct a graph $G' = (V', E')$ by taking $V' = V_0 \cup \{a, x, b\}$, and letting $E' = E_0 \cup (a, x) \cup (b, x) \cup S$, where S is as in G; specify the edge weights $\omega(a, x) = \omega(b, x) = \pm 1$, in agreement* *with the sign of* $\omega(a,b)$*. Then*

$$|2\tau(G)| = |\tau(G')|.$$

Proof. As in the proof of [39, Lemma 6], we may partition the spanning trees of G into two sets $C_1 \cup C_2$, where spanning trees in C_1 contain edge (a,b) and spanning trees in C_2 do not contain edge (a,b). Likewise, partition the spanning trees of G' into three sets $C1' \cup C2' \cup C3'$, where spanning trees in C_1' contain both $(a,x),(b,x)$, where trees in C_2' contain (a,x) but not (b,x), and trees in C_3' contain (b,x) but not (a,x).

There exists a bijection $f : C_1' \to C_1$ obtained by deleting the vertex x and adding the edge (a,b). If the weights of the edges $(a,b) \in G$ and $(a,x),(b,x)$ in G' are all positive, then the bijection satisfies $\omega(f(T')) = \frac{1}{2}\omega(T')$ for all spanning trees T' in C_1', whereas if the edge weights are all negative, then $\omega(f(T')) = -\frac{1}{2}\omega(T')$. There are also bijections $g : C_2' \to C_2$ and $h : C_3' \to C_2$, obtained by contracting the edge (a,x) or (b,x), respectively. In this case, when the edge weights are all positive, then $\omega(g(T')) = \omega(T')$ and $\omega(h(T')) = \omega(T')$, whereas if the edge weights are all negative, then $\omega(g(T')) = \omega(T')$ and $\omega(h(T')) = \omega(T')$.

Finally, observe that

$$\begin{aligned}\tau(G') &= \sum_{T' \in C_1'} \omega(T') + \sum_{T' \in C_2'} \omega(T') + \sum_{T' \in C_3'} \omega(T') \\ &= \pm 2 \sum_{T \in C_1} \omega(T) \pm 2 \sum_{T \in C_2} \omega(T) \\ &= \pm 2\tau(G),\end{aligned}$$

where the sign in front of the summation is positive/negative when the edge weights $(a,b) \in G$ and $(a,x),(b,x) \in G'$ are all positive/negative, respectively. □

Remark 5.7. Non-simple graphs containing self-loops or multiedges may result from Tait graphs of knot diagrams. For edges that are self-loops, $\tau(G) = \tau(G - e)$.

5.3.2 Spanning Trees of Graphs With Involutive Symmetry

We will now show how the relationship between the determinants of the constituent knots and theta curve is described by counting spanning trees of graphs with involutive symmetry. Here, $G = (V(G), E(G))$ is the weighted Tait graph of (K, h), with symmetry from the involution h. The following algorithms constructing the graphs G_L and G_R are due to Zhang-Yan [123, Theorem 2.1], generalizing unweighted versions due to Ciucu-Yan-Zhang in [39, Theorem 4]. The involution h partitions $V(G)$ into three sets: $V_L \cup V_C \cup V_R$, where $V_L = \{v_1, \cdots, v_n\}$ consists of vertices on the left side of the axis of involution, $V_R = \{v_1', \cdots, v_n'\}$ are vertices on the right, and $V_C = \{w_1, \cdots, w_m\}$ are vertices lying on the axis.

Definition 5.8. (G_R and G_L [123]). Two weighted graphs G_L and G_R are obtained from G as follows. To form G_R:

(i) Take the subgraph of G induced by $V_R \cup V_C$.

(ii) For every edge $e = (w_i, w_j)$ along the axis of involution, reduce the weight by half.

To form G_L:

(i) Take the subgraph of G induced by V_L together with a new vertex u.

(ii) For each edge $e = (v_i, v_i')$ with weight $\omega(e)$, add an edge (u, v_i) with weight $2\omega(e)$.

(iii) For each edge $e = (v_i, w_j)$ add an edge (u, v_i) with weight $\omega(e)$.

With these defined,

Theorem 5.9. *(Zhang-Yan [123, Theorem 2.1]) Suppose that $G = (V(G), E(G))$ is a weighted graph with an involution h and that G_L, G_R, and V_C are defined as above. Then the tree weight of G is given by*

$$\tau(G) = 2^{m-1}\tau(G_L)\tau(G_R)$$

where m is the number of vertices of V_C.

Remark 5.10. Because (K, h) is strongly invertible, the edge set $E(G)$ of the Tait graph G contains edges of the form $e = (v_i, v'_j)$ only if $i = j$. Thus we have omitted items (3) and (2′) from the definition appearing in [123].

5.4 Proof of Theorem 5.13

Assume that $K_{ac} = e_a \cup e_c$ is unknotted, and call $K_{ab} = e_a \cup e_b$ and $K_{bc} = e_b \cup e_c$.

As described in Section 5.2.1, K_{ac} can be viewed as the fixed set $\mathrm{Fix}(h)$ of the involution for some strongly invertible knot $(K, h) = \tilde{e}_b$. We may assume that any diagram of K is symmetric with respect to h and view K_{ac} as a vertical axis α with the point at infinity. By definition, (K, h) intersects α in exactly two points. This partitions $\alpha \cup \{\infty\} = e_a \cup e_c$ into 'two rooms' along which the diagram admits a uniform checkerboard coloring pattern from wall to wall, as in Figure 49. More precisely, symmetry implies that given any edge $e \in E(G)$ or $e \in E(G^\perp)$, either α intersects e in exactly one point, α and e are disjoint, or α intersects e in e. Recall that there are two choices of a checkerboard shading φ of diagram of (K, h). For exactly one choice of shading φ, the following holds for all edges:

$$\begin{aligned} &\text{In } e_a: \quad e \cap e_a = e \text{ and } e^\perp \cap e_a = 1 \text{ point}, \\ &\text{In } e_c: \quad e \cap e_c = 1 \text{ point and } e^\perp \cap e_c = e^\perp. \end{aligned} \tag{5.3}$$

The other choice in shading will yield an equivalent statement interchanging e and $e^\perp$. As a consequence of (5.3), we have:

Lemma 5.11. *A checkerboard shading of (K, h) in S^3 induces a checkerboard shading on the constituent knots K_{ab} and K_{bc} in the quotient diagram of $\vartheta = K/h \cup \alpha$ in S^3.*

Proof. Choose a checkerboard shading in a symmetric diagram of (K, h). Quotient via the involution h to obtain a theta curve $\vartheta = K/h \cup e_a \cup e_c$. In $K_{ab} = K/h \cup e_a$, the checkerboard shading at the crossings of (K, h) along e_a descend in the quotient to checkerboard shaded clasps, and the shaded crossings along e_c descend to shaded 'fingers.' See Figure 49. Similarly in $K_{bc} = K/h \cup e_c$, the checkerboard shading at crossings along e_c descends to shaded clasps, and the shaded crossings along e_a

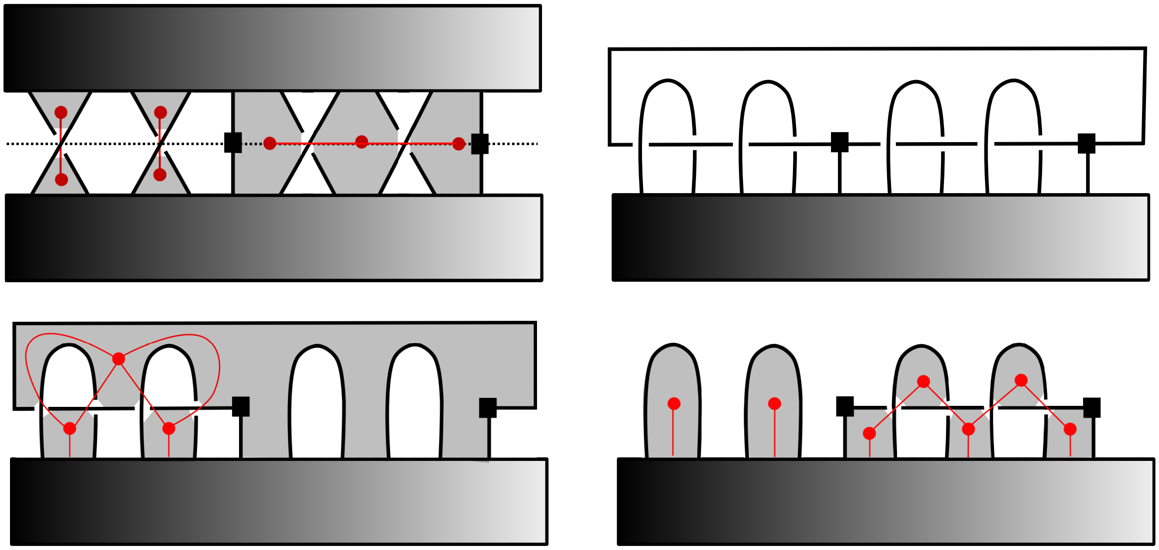


Fig. 49. (Top left) Checkerboard shading on (K, h) showing 'two rooms' of shading patterns along the axis. (Top right) $\vartheta = K/h \cup e_a \cup e_c$. (Bottom) The shading of (K, h) induces a shading on the constituent knots K_{bc} (left) and K_{ab} (right) of ϑ. Parts of the Tait graphs are indicated in red.

descend to shaded fingers. Away from the axis, the checkerboard shadings in the diagram of K_{ab} and K_{bc} agree with that of (K, h). $\square$

Let G, G_{ab}, G_{bc} denote the Tait graphs for $(K, h), K_{ab}, K_{bc}$, respectively.

Lemma 5.12. *For one choice of checkerboard shading of* (K, h)*, we have* $2^{m-1}\tau(G_R) = \tau(G_{ab})$ *and* $\tau(G_L) = \tau(G_{bc})$*. With other choice,* $2^{m-1}\tau(G_R^{\perp}) = \tau(G_{bc}^{\perp})$ *and* $\tau(G_L^{\perp}) = \tau(G_{ab}^{\perp})$*.*

In the second case, Lemma 5.12 implies $|\tau(G)| = |\tau(G^{\perp})|$.

Proof of Lemma 5.12. Fix a symmetric diagram of (K, h) and by convention, let e_c be the unbounded arc of the axis. The Tait graph G of the diagram is symmetric, planar and connected. The edge weights $\omega(e) \in \{-1, +1\}$ for $e \in G$ are induced from the incidence numbers $\xi(c)$ at the crossings, where the signs of the $\xi(c)$ depend on the choice in the checkerboard shading of the diagram. Specify the shading φ of (K, h) so that (5.3) holds. With this choice, the unbounded region of the diagram is unshaded. By Lemma 5.11, φ induces a shading on the constituent knot $K_{ab} = e_b \cup e_a$

corresponding with Tait graph G_{ab}. Under the action of the involution, edges in G that are disjoint from α map bijectively to edges in G_{ab}. Edges intersecting α in a point do not map to edges in G_{ab}, and edges that lie along α map to a subdivided edge in G_{ab}. In particular, to form G_{ab}:

(i) Take the subgraph of G induced by $V_R \cup V_C$.

(ii) For every edge $e = (w_i, w_j)$ along the axis of involution, subdivide e into $(w_i, x) \cup (x, w_j)$ and set $\omega(w_i, x) = \omega(x, w_j) = \frac{1}{2}\omega(e)$.

This nearly agrees with the definition of G_R; Lemma 5.6 then implies $2^{m-1}\tau(G_R) = \tau(G_{ab})$.

Consider now K_{bc} with Tait graph G_{bc}. Under the action of the involution, edges in G that are disjoint from α map bijectively to edges in G_{bc}. Edges that lie along α do not map to edges in G_{bc}. Edges $e = (v_i, v_i')$ in G that intersect α in a point map to a pair of edges in G_{bc}. (This pair of edges is dual to a subdivided edge in $G_{bc}^{\perp}$; see Figure 49). In particular, to form G_{bc}:

(i) Take the subgraph of G induced by V_L together with a new vertex u.

(ii) For each edge $e = (v_i, v_i')$ with weight $\omega(e)$, add a pair of edges edges $e_1 = (u, v_i) = e_2$ each with weight $\omega(e)$.

(iii) For each edge $e = (v_i, w_j)$ add an edge (u, v_i) with weight $\omega(e)$.

This nearly agrees with the definition of G_L; the difference is the factor of 2 in the edge weight in item (ii), which here manifests as a pair of edges. Thus $\tau(G_L) = \tau(G_{bc})$.

Finally, let us consider the other choice in shading. Equation 5.3 becomes an equivalent statement with e and $e^{\perp}$ interchanged. Duality preserves connectedness, planarity, symmetry and edge weights. The above argument applies, *mutatis mutandis*: interchange $G^{\perp}$ and G, and interchange K_{ab} and K_{bc}. $\square$

We can now prove the main result:

Theorem 5.13. *Let ϑ be a simple theta curve with constituent knots K_{ab}, K_{ac}, K_{bc}. Then*

$$\det(\vartheta) = \det(K_{ab}) \cdot \det(K_{ac}) \cdot \det(K_{bc}).$$

Proof. By Lemma 5.4, the determinants of (K, h), K_{ab}, and K_{bc} may be calculated by the tree weights of G, G_{ab}, and G_{bc}, respectively (or equivalently by the tree weights of $G^{\perp}, G^{\perp}_{ab}$, and $G^{\perp}_{bc}$). Hence, by Theorem 5.9, we have

$$\begin{aligned} \det(\vartheta) &= \tau(G) \\ &= 2^{m-1}\tau(G_L)\tau(G_R) \\ &= \tau(G_{ab})\tau(G_{bc}) \\ &= \det(K_{ab})\det(K_{ac})\det(K_{bc}). \end{aligned}$$

□

5.4.1 Examples

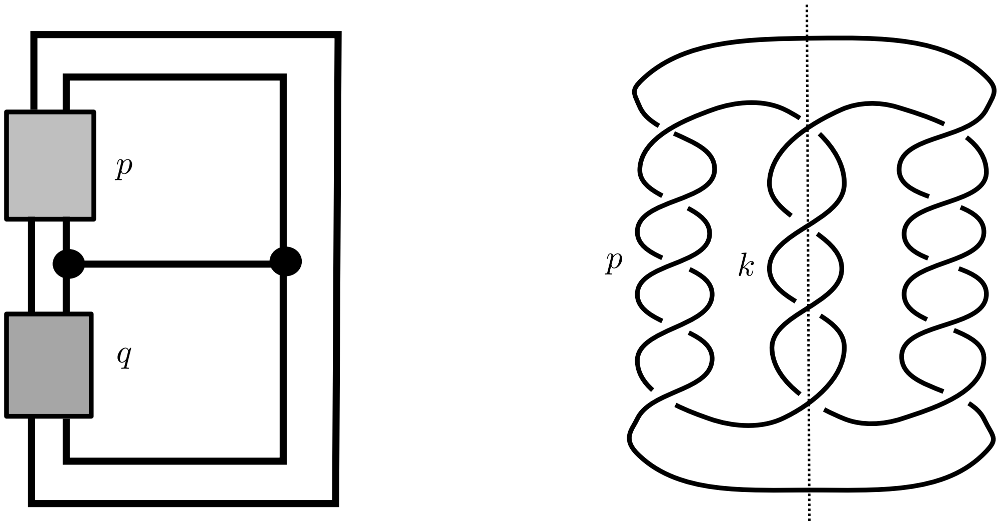


Fig. 50. The ϑ-curve $\vartheta(p, q)$ with constituent knots $T(p + q, 2)$, $T(p, 2)$, and U. The $P(p, q, p)$ pretzel knot is the corresponding strongly invertible knot.

Example 5.14. Let $\vartheta(p, q)$, with p odd and $q = 2k$ even, be the ϑ-curve pictured in Figure 50. The three constituent knots are the unknot, and the torus knots $T(p+q, 2)$ and $T(p, 2)$. By Theorem 5.13, $\det(\vartheta) = \det(T(p + q, 2)) \cdot \det(T(p, 2)) = (p + q) \cdot p = p^2 + pq$. The pretzel knot $P(p, k, p)$ is the strongly invertible knot that corresponds with $\vartheta(p,q)$, and it also has determinant p^2+pq.

Example 5.15. Consider the strongly invertible knot $(K,h) = 9_{48}$, pictured with an axis of involution in Figure 51. Its quotient under the involution, together with the axis, forms a spatial theta curve whose diagram contains 9 crossings. The two constituent knots $K_{ab} = 3_1$ and $K_{bc} = 6_1$ are shown in the figure. The determinants of $3_1, 6_1, 9_{48}$ are $3, 9, 27$. The Tait graphs $G, G^{\perp}$ for (K,h) and $G_{ij}, G_{ij}^{\perp}$ for K_{ij} are also illustrated.

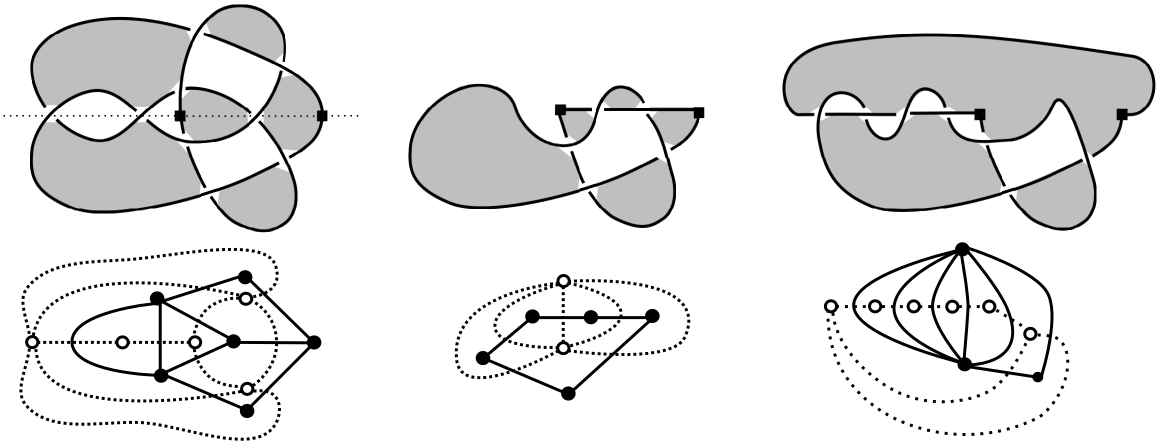


Fig. 51. (Left) Strongly invertible knot 9_{48} with axis of involution. (Center) Constituent knot K_{ab} is the trefoil 3_1. (Right) Constituent knot K_{bc} is the knot 6_1. Corresponding Tait graphs and their duals are shown below.

Example 5.16. Constituent knots for all theta curves in the Litherland-Moriuchi table [87, 75] were previously determined by Baker, Buck and O'Donnol in [24, Table 2]. We can now apply Theorem 5.13 to compute the values of $\det(\vartheta)$ for all of the theta curves tabulated. An augmented table including the values of $\det(\vartheta)$ and $\det(K)$ for constituent knots is displayed as Table 5 below.

ϑ	C. Knots			det(K_{ij})	det(ϑ)
$\mathbf{3_{1}}$	2x0_1	3_1		1,3	3
$\mathbf{4_{1}}$	2x0_1	4_1		1,5	5
$\mathbf{5_{1}}$	3x0_1			1	1
$\mathbf{5_{2}}$	2x0_1	3_1		1,3	3
$\mathbf{5_{3}}$	2x0_1	5_1		1,5	5
$\mathbf{5_{4}}$	0_1	3_1	5_1	3,5	15
$\mathbf{5_{5}}$	2x0_1	5_2		1,7	7
$\mathbf{5_{6}}$	2x0_1	5_2		1,7	7
$\mathbf{5_{7}}$	0_1	3_1	5_2	3,7	21
$\mathbf{6_{1}}$	3x0_1			1	1
$\mathbf{6_{2}}$	2x0_1	3_1		1,3	3
$\mathbf{6_{3}}$	0_1	3_1	4_1	3,5	15
$\mathbf{6_{4}}$	0_1	3_1	4_1	3,5	15
$\mathbf{6_{5}}$	2x0_1	6_1		1,9	9
$\mathbf{6_{6}}$	2x0_1	6_1		1,9	9
$\mathbf{6_{7}}$	2x0_1	6_1		1,9	9
$\mathbf{6_{8}}$	0_1	4_1	6_1	5,9	45
$\mathbf{6_{9}}$	2x0_1	6_2		1,11	11
$\mathbf{6_{10}}$	2x0_1	6_2		1,11	11
$\mathbf{6_{11}}$	2x0_1	6_2		1,11	11
$\mathbf{6_{12}}$	0_1	3_1	6_2	3,11	33
$\mathbf{6_{13}}$	0_1	4_1	6_2	5,11	55
$\mathbf{6_{14}}$	2x0_1	6_3		1,13	13
$\mathbf{6_{15}}$	2x0_1	6_3		1,13	13
$\mathbf{6_{16}}$	0_1	3_1	6_3	3,13	39
$\mathbf{7_{1}}$	3x0_1			1	1
$\mathbf{7_{2}}$	3x0_1			1	1
$\mathbf{7_{3}}$	3x0_1			1	1
$\mathbf{7_{4}}$	3x0_1			1	1
$\mathbf{7_{5}}$	2x0_1	3_1		1,3	3
$\mathbf{7_{6}}$	2x0_1	3_1		1,3	3
$\mathbf{7_{7}}$	2x0_1	3_1		1,3	3
$\mathbf{7_{8}}$	0_1	3_1	3_1	3,3	9
$\mathbf{7_{9}}$	0_1	3_1	3_1	3,3	9
$\mathbf{7_{10}}$	0_1	3_1	3_1	3,3	9
$\mathbf{7_{11}}$	2x0_1	5_2		1,7	7
$\mathbf{7_{12}}$	2x0_1	4_1		1,5	5
$\mathbf{7_{13}}$	2x0_1	4_1		1,5	5
$\mathbf{7_{14}}$	0_1	4_1	4_1	5,5	25
$\mathbf{7_{15}}$	2x0_1	5_1		1,5	5
$\mathbf{7_{16}}$	2x0_1	5_1		1,5	5
$\mathbf{7_{17}}$	2x0_1	5_1		1,5	5
$\mathbf{7_{18}}$	0_1	5_1	5_2	5,7	35
$\mathbf{7_{19}}$	2x0_1	5_2		1,7	7
$\mathbf{7_{20}}$	2x0_1	5_2		1,7	7
$\mathbf{7_{21}}$	2x0_1	5_2		1,7	7
$\mathbf{7_{22}}$	0_1	3_1	5_2	3,7	21
$\mathbf{7_{23}}$	0_1	4_1	5_2	5,7	35
$\mathbf{7_{24}}$	0_1	4_1	5_2	5,7	35
$\mathbf{7_{25}}$	2x0_1	7_1		1,7	7
$\mathbf{7_{26}}$	0_1	3_1	7_1	3,7	21
$\mathbf{7_{27}}$	0_1	5_1	7_1	5,7	35
$\mathbf{7_{28}}$	2x0_1	7_2		1,11	11
$\mathbf{7_{29}}$	2x0_1	7_2		1,11	11
$\mathbf{7_{30}}$	2x0_1	7_2		1,11	11
$\mathbf{7_{31}}$	0_1	3_1	7_2	3,11	33
$\mathbf{7_{32}}$	0_1	5_2	7_2	7,11	77
$\mathbf{7_{33}}$	2x0_1	7_3		1,13	13
$\mathbf{7_{34}}$	2x0_1	7_3		1,13	13
$\mathbf{7_{35}}$	0_1	3_1	7_3	3,13	39
$\mathbf{7_{36}}$	0_1	5_1	7_3	5,13	65
$\mathbf{7_{37}}$	0_1	5_2	7_3	7,13	91
$\mathbf{7_{38}}$	2x0_1	7_4		1,15	15
$\mathbf{7_{39}}$	2x0_1	7_4		1,15	15
$\mathbf{7_{40}}$	0_1	3_1	7_4	3,15	45
$\mathbf{7_{41}}$	0_1	3_1	7_4	3,15	45
$\mathbf{7_{42}}$	0_1	5_2	7_4	7,15	105
$\mathbf{7_{43}}$	2x0_1	7_5		1,17	17
$\mathbf{7_{44}}$	2x0_1	7_5		1,17	17
$\mathbf{7_{45}}$	0_1	3_1	7_5	3,17	51
$\mathbf{7_{46}}$	0_1	3_1	7_5	3,17	51
$\mathbf{7_{47}}$	0_1	3_1	7_5	3,17	51
$\mathbf{7_{48}}$	0_1	5_1	7_5	5,17	85
$\mathbf{7_{49}}$	0_1	5_2	7_5	7,17	119
$\mathbf{7_{50}}$	2x0_1	7_6		1,19	19
$\mathbf{7_{51}}$	2x0_1	7_6		1,19	19
$\mathbf{7_{52}}$	2x0_1	7_6		1,19	19
$\mathbf{7_{53}}$	2x0_1	7_6		1,19	19
$\mathbf{7_{54}}$	2x0_1	7_6		1,19	19
$\mathbf{7_{55}}$	0_1	3_1	7_6	3,19	57
$\mathbf{7_{56}}$	0_1	3_1	7_6	3,19	57
$\mathbf{7_{57}}$	0_1	4_1	7_6	5,19	95
$\mathbf{7_{58}}$	0_1	5_2	7_6	7,19	133
$\mathbf{7_{59}}$	2x0_1	7_7		1,21	21
$\mathbf{7_{60}}$	2x0_1	7_7		1,21	21
$\mathbf{7_{61}}$	2x0_1	7_7		1,21	21
$\mathbf{7_{62}}$	2x0_1	7_7		1,21	21
$\mathbf{7_{63}}$	2x0_1	7_7		1,21	21
$\mathbf{7_{64}}$	0_1	3_1	7_7	3,21	63
$\mathbf{7_{65}}$	0_1	4_1	7_7	5,21	105

Table 5. Theta curves through seven crossings, their constituents, and their determinants. All of the constituent knots in the Litherland-Moriuchi table [87, 75] are simple.

Appendix A

MATLAB CODE FOR RNA METHODOLOGY

```matlab
function [pairType,pairDist,isValid] = pairChecker(chord1,chord2,chordlist)
%function [pairType,pairdist,isValid] = pairChecker(chord1,chord2,chordlist)
%
% Checks whether two chords are nested or crossed, computes their distance,
% and optionally checks if their is a paired base between their endpoints
%
% Input:  chord1/2....... 1x2 arrays of the form [l_i r_i]
%         chordlist...... the list of chords to check if there is a paired
%              base between the left or right end points of chords 1 and 2
%
% Output: pairType.......... -1 for nesting, 1 for crossing, 0 for neither
%         pairDist.......... max(abs(l_2-l_1), abs(r_2-r_1))
%         isValid........... Is this a valid k-nest or k-cross. True if
%                            there is no paired base between l1,l2 or r1,r2
%
%
% - chord1 and chord2 should be input in order by left endpoints, but the
% code will check for this.
% - If only chord1 and chord2 are passed, will simply check if they are a
% crossing or nesting and how far they are.
%
%
% Example: chordlist = [1 3; 2 5; 4 6; 7 12; 8 11; 9 10; 13 18; 17 21]
%
%       [pairType,pairdist,isValid] = pairChecker([7 12],[8 11],chordlist)
%       pairType is -1, pairdist is 1, isValid is True
%
%       [pairType,pairdist,isValid] = pairChecker([13 18],[17 21],chordlist)
%       pairType is 1, pairdist is 4, isValid is True
%
%       [pairType,pairdist,isValid] = pairChecker([4 6],[9 10],chordlist)
%       pairType is 0, pairdist is 5, isValid is False


%Some checks to make sure input is formatted correctly.
%Check to make sure the chords are in the format [l_i r_i]
if ~isequal(size(chord1),[1 2]) || ~isequal(size(chord2),[1 2])
    error('Input chords must be size 1x2 array')
end

l1 = chord1(1);
r1 = chord1(2);

l2 = chord2(1);
r2 = chord2(2);

if l1 >= r1 || l2 >= r2
    error('Chords are not formatted correctly as [l_i r_i].')
end

%Check if chords are ordered by left endpoint

if l1 > l2
    tmp = chord1;
```

```
    chord1 = chord2;
    chord2 = tmp;
end

%End checks

pairDist = max([abs(l2-l1) abs(r2-r1)]); %The chord distance

if l2 > l1 && r2 > r1 && l2 < r1
    pairType = 1; %Crossed e.g. [1 3] and [2 4]
elseif l2 > l1 && r2 < r1
    pairType = -1; %Nested e.g. [1 4] and [2 3]
else
    pairType = 0; %Independent e.g. [1 3] and [4 5]
end

%If the global chord list is provided, then we can check if there is a
%paired base between the two chords.
if nargin == 3
    isValid = true;
    leftRange = sort([l1 l2]);
    rightRange = sort([r1 r2]);
    %Left endpoint check
    jt1 = chordlist > leftRange(1) & chordlist < leftRange(2);
    %Right endpoint check
    jt2 = chordlist > rightRange(1) & chordlist < rightRange(2);
    %The chord list contains all bases which are paired. So if there's any
    %endpoint of a chord within the range of l1,l2 or r1,r2, then there is
    %a paired base between chord1 and chord2, so isValid=false.
    if any(jt1,'all') || any(jt2,'all')
        isValid = false;
    end
end


end
```

```
function [segments] = findSegments(chords,tau)
%function [segments] = findSegments(chords,tau)
%   Input: chords................. list of base pairs
%          tau.................... 0=bpRNA, 1<tau<=inf = tau-Segments
%   Output: segments.............. cell array, each cell is a segment
%
%   A segment is a maximal non-chord obstructed k-nesting.
%   An augmented segment is a
%   If tau=0,  segments are maximal nestings
%   Otherwise tau is used in the distance condition, and segments can be
%   crossings.
numChords = size(chords,1);
segments={};
currentSegment = chords(1,:);
if tau < 0
    error('Enter a value of tau greater or equal to 0.')
end

for ii = 1:numChords-1
    chord1 = chords(ii,:);
    chord2 = chords(ii+1,:);
    [pairType,pairDist,isValid] = pairChecker(chord1,chord2,chords);

    %Comparison Dependent on input for tau.
    %tau=0 --> bpRNA segment definition
    % 1 <= tau < infty --> G_tau
    % tau = inf --> G_Aug = G_{tau_max}
```

```
    if tau == 0
        %bpRNA way.. every segment must be a nesting
        if pairType == -1 && isValid
            currentSegment = [currentSegment; chord2];
        else
            segments{end+1} = currentSegment;
            currentSegment = chord2;
        end
    else
        %Just need to check for maximally non chord obstructed sets of
        %chords within distance tau
        %If tau=inf, then pairDist < tau always true, looks for maximally
        %non chord obstructed groups of chords
        if pairDist <= tau && isValid
            currentSegment = [currentSegment; chord2];
        else
            segments{end+1} = currentSegment;
            currentSegment = chord2;
        end
    end

end
segments{end+1} = currentSegment;

end
```

```
function [segmentgraph] = makeSegmentGraph(segments)
%function [segmentgraph] = makeSegmentGraph(segments)
%   Input:  Takes a cell array of segments from findSegments
%   Output: Intersection graph of segments
%


numSegments = numel(segments);
segmentgraph=graph(logical(zeros(numSegments))); %#ok<*LOGL>
segmentgraph.Nodes.Weights = cellfun(@numel,segments)'/2;
if numSegments==1
    return
end
possibleEdges = nchoosek(1:numSegments,2);


for ii = 1:size(possibleEdges,1)
    s = possibleEdges(ii,1);
    t = possibleEdges(ii,2);
    chord1 = segments{s}(1,:);
    chord2 = segments{t}(1,:);
    pairType = pairChecker(chord1,chord2);
    %Theorem: if c1 in S1 and c2 in S2 cross, then all chords in S1 cross
    %all chords in S2, so just need to check one pair to see if segments
    %cross.
    if pairType == 1
        segmentgraph = addedge(segmentgraph,s,t);
    end
end


end
```

```
function [thePKs,numPKs] = findPKs(segmentgraph,weight_first)
%function [thePKs,numPKs] = findPKs(segmentgraph,use_weights)
% Input: segmentgraph........ from makeSegmentGraph
%        weight_first......... boolean.
```

```
%
% If weight_first is true, then minimize weight, then minimize cardinality
% If weight_first is false, then minimize cardinality, then minimize weight
%
%Warning: Use BK max independent set algorithm only on the connected
%components. Number of sets explodes with multiple components
%(multiplicative).
g = segmentgraph;

%Get list of components, and get rid of isolated vertices.
[comps,compsizes] = conncomp(g,'OutputForm','cell');
jt = compsizes==1;
comps(jt) = []; %Done this way for some debugging purposes.
weights = g.Nodes.Weights;
thePKs = cell(1,numel(comps)); %at least one PK per connected component
for ii = 1:numel(comps)
    V = comps{ii}; %Vertex set of component ii
    if weight_first %Min cardinality over all min weight covers
        if numel(V) == 2
            %if component is size 2, pick the vertex with minimum weight
            [~, ind] = min(weights(V));
            thePKs{ii} = V(ind);
        else
            H = subgraph(g,V); %Component jj
            A_H = full(adjacency(H)); %Adj matrix of H
            M = BK_MaxIS(A_H)'; %Each row is max'l ind set in H
            W = H.Nodes.Weights; %Weights of vertices of H
            W_I = M*W; %Sum of weights for each maximal independent set
            W_max = max(W_I); %Find maximum weight sum
            jt = W_I == W_max; %Find all ind sets of max weight
            Mp = M(jt,:);
            thecards = sum(Mp,2);
            jt = thecards == max(thecards); %Find max cardinality max weight ind set
            Mp = Mp(jt,:);
            Mvc = ones(size(Mp))-Mp;
            Mvc = sortrows(Mvc,'descend'); %Lex order
            thePKs{ii} = V(logical(Mvc(1,:)));
        end
    else %Min weight over all min cardinality covers
        H = subgraph(g,comps{ii});
        M = BK_MaxIS(full(adjacency(H)))';
        thecards = sum(M,2);
        jt = thecards == max(thecards); %Find maximum cardinality ind sets
        M = M(jt,:);
        W = H.Nodes.Weights; %Weights of vertices of H
        Mvc = ones(size(M))-M;
        Mvc = sortrows(Mvc,'descend'); %Lex order
        W_I = Mvc*W; %Sum of weights for each minimum vertex cover
        [~,jt] = min(W_I);
        thePKs{ii} = V(logical(Mvc(jt,:)));
    end

end

numPKs = sum(cellfun(@numel,thePKs));
end
```

```
function [secondary] = classifyBases(segments,seqLength)
%function [secondary] = classifyBases(segments,seqLength)
% Takes a segment partition where no two segments cross each other.
% If each segment is a nesting, then the chord diagram is crossingless
%
% Note after PK removal, the only crossings in the chord diagram come from
% segments which are crossings (independent crossings).
```

```
% If there is at least one segment which is a crossing, that is a 'twisted
% stem', and we will reverse the order of the right endpoints to make it
% look like it was a nesting originally, collect the chords all again, and
% remake the segment partition in this new chord diagram with no crossings.
%
% Outputs structure 'secondary' with fields
%   Internal loops
%   Bulges
%   Hairpins
%   Multiloops
%   External Loop
%   Dangling Ends
%   Stems
%   Twisted Stems

%Check to see which segments are twisted stems, if there are any.
segmentstmp = segments;
segmentTypes = zeros(1,numel(segments));
S = [];
ST = [];
for ii = 1:numel(segmentstmp)
    chords = segmentstmp{ii};
    nchords = size(chords,1);
    if nchords == 1
        segmentTypes(ii) = -1;
        S = [S; chords'];
        continue
    else
        pairType = pairChecker(chords(1,:),chords(2,:));
        segmentTypes(ii) = pairType;
        if pairType == 1 %if segment is a crossing
            segmentstmp{ii} = [chords(:,1) chords(end:-1:1,2)];
            ST = [ST; chords(:)];
        else
            S = [S; chords(:)];
        end
    end
end
S = sort(S);
ST = sort(ST);
%New chord list where every segment is now a nesting
chords = vertcat(segmentstmp{:});
%New segment list, to ensure maximal nestings. For example, in the
%reduction method, if you have an r-crossing properly containing a
%k-crossing, then identifying the r-crossing to a single chord creates an
%effective k+1-nesting. Except we don't identify, so we really get a k+r
%nesting.
segments = findSegments(chords,inf);

unpaired = setdiff(1:seqLength,chords(:));

minpaired = min(chords(:,1));
maxpaired = max(chords(:,2));

I = {}; % Internal loop regions
B = {}; % Bulge regions
H = {}; % Hairpin regions
M = {}; % Multiloop regions
X = {}; % Xternal loop regions
E = [1:(minpaired-1) (maxpaired+1):seqLength]; % Dangling
unpaired=setdiff(unpaired,E);
%bulges and internal loops are found within a segment.
for jj = 1:numel(segments)
    s = segments{jj};
```

```
    %outer = [outer; s(1,:)];
    for kk = 1:size(s,1)-1 %
        %Remember, right endpoints are in descending order in nestings
        L = (s(kk,1) + 1):(s(kk+1,1)-1); %Left side
        R = (s(kk+1,2) + 1):(s(kk,2)-1); %Right side
        LandR = [L R];
        if ~isempty(L) && ~isempty(R) %Check if both nonempty first
            %Internal Loop if theres a gap between lefts and rights
            I = [I {LandR}];
        elseif ~isempty(L) || ~isempty(R) %o/w check >= 1 nonempty
            %One of L or R is empty, so the nonempty one is a bulge
            B = [B {LandR}];
        end
        unpaired = setdiff(unpaired, LandR);
    end
    %Check if this segment flanks a hairpin by checking if the sequence
    %of bases contained by the innermost chord has no paired base
    %Also, if it does have a paired base, then there are multiloops to
    %find
    inner = s(end,:);
    testseq = (inner(1)+1):(inner(2)-1);
    basesContained = intersect(testseq,unpaired); %Bases under segment
    if ~isempty(basesContained)
        if all(diff(basesContained)==1)
            H = [H {basesContained}];
            unpaired = setdiff(unpaired,basesContained);
        else
            %Want all the unpaired bases under this segment which is
            %not underneath any other segment
            tmp = [];
            for kk = 1:numel(basesContained)
                base = basesContained(kk);
                jt = find(base > chords(:,1) & base < chords(:,2),1,'last');
                if isequal(chords(jt,:),inner)
                    %Check to see if the 'minimal containing chord' is
                    %inner
                    tmp = [tmp base];
                end
            end
            M = [M {tmp}];
            unpaired = setdiff(unpaired,tmp);
        end
    end
end


%So far we've gotten all the internal loops, bulges, and hairpins
%
%Unpaired bases not contained under any chord are part of the external
%loop
tmp = [];
for jj = 1:numel(unpaired)
    %External loop check
    jt=unpaired(jj) >= chords(:,1) & unpaired(jj) <= chords(:,2);
    if all(~jt) %if unpaired(jj) does not live underneath a chord
        tmp = [tmp unpaired(jj)];
    end
end
unpaired=setdiff(unpaired,tmp);
X = {};
%Find each region
dt = [0 find(diff(tmp) ~= 1) numel(tmp)];
for jj = 1:numel(dt)-1
    X = [X {tmp(dt(jj)+1:dt(jj+1))}];
```

```
end
if ~isempty(unpaired)
    error("Unpaired base not labeled")
end
%Need sequences of consecutive chords to order and 'index' the
%exterior regions from left to right.
secondary.I = I;
secondary.B = B;
secondary.H = H;
secondary.M = M;
secondary.X = X;
secondary.E = E;
secondary.S = S;
secondary.ST = ST;

end
```

REFERENCES

[1] Hüseyin Acan. "On a uniformly random chord diagram and its intersection graph". In: *Discrete Math.* 340.8 (2017), pp. 1967–1985. DOI: `10.1016/j.disc.2016.11.004`.

[2] Oswin Aichholzer, Ruy Fabila-Monroy, Philipp Kindermann, Irene Parada, Rosna Paul, Daniel Perz, Patrick Schnider, and Birgit Vogtenhuber. "Perfect matchings with crossings". In: *Algorithmica* 86.3 (2024), pp. 697–716. DOI: `10.1007/s00453-023-01147-7`.

[3] Christos Andrikos, Evangelos Makris, Angelos Kolaitis, Georgios Rassias, Christos Pavlatos, and Panayiotis Tsanakas. "Knotify: An Efficient Parallel Platform for RNA Pseudoknot Prediction Using Syntactic Pattern Recognition". In: *Methods and Protocols* 5.1 (2022).

[4] G. Antonick. "Béla Bollobás: The Spread of Infection on a Square Grid". In: *The New York Times* 81 (Aug. 2013).

[5] József Balogh and Béla Bollobás. "Bootstrap percolation on the hypercube". In: *Probab. Theory Related Fields* 134.4 (2006), pp. 624–648. DOI: `10.1007/s00440-005-0451-6`.

[6] József Balogh, Béla Bollobás, Hugo Duminil-Copin, and Robert Morris. "The sharp threshold for bootstrap percolation in all dimensions". In: *Trans. Amer. Math. Soc.* 364.5 (2012), pp. 2667–2701. DOI: `10.1090/S0002-9947-2011-05552-2`.

[7] József Balogh, Yuval Peres, and Gábor Pete. "Bootstrap percolation on infinite trees and non-amenable groups". In: *Combin. Probab. Comput.* 15.5 (2006), pp. 715–730. DOI: 10.1017/S0963548306007619.

[8] József Balogh and Gábor Pete. "Random disease on the square grid". In: *Proceedings of the Eighth International Conference "Random Structures and Algorithms" (Poznan, 1997)*. Vol. 13. 3-4. 1998, pp. 409–422. DOI: 10.1002/(SICI)1098-2418(199810/12)13:3/4<409::AID-RSA11>3.3.CO;2-5.

[9] József Balogh and Boris G. Pittel. "Bootstrap percolation on the random regular graph". In: *Random Structures Algorithms* 30.1-2 (2007), pp. 257–286. DOI: 10.1002/rsa.20158.

[10] Curtiss Barefoot, Karen Casey, David Fisher, Kathryn Fraughnaugh, and Frank Harary. "Size in maximal triangle-free graphs and minimal graphs of diameter 2". In: vol. 138. 1-3. 14th British Combinatorial Conference (Keele, 1993). 1995, pp. 93–99. DOI: 10.1016/0012-365X(94)00190-T.

[11] Giorgio Benedetti and Stefano Morosetti. "A graph-topological approach to recognition of pattern and similarity in RNA secondary structures". In: *Biophysical Chemistry* 59.1 (1996), pp. 179–184. DOI: https://doi.org/10.1016/0301-4622(95)00119-0.

[12] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. "The Protein Data Bank". In: *Nucleic Acids Research* 28.1 (Jan. 2000), pp. 235–242. DOI: 10.1093/nar/28.1.235. eprint: https://academic.oup.com/nar/article-pdf/28/1/235/9895144/280235.pdf.

[13] Norman Biggs. *Algebraic graph theory*. Second. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1993, pp. viii+205. ISBN: 0-521-45897-8.

[14] Berk Birand. *Bron-Kerbosch maximal independent set and maximal clique algorithms*. 2014.

[15] Béla Bollobás. *The art of mathematics*. Coffee time in Memphis. Cambridge University Press, New York, 2006, pp. xvi+359. ISBN: 978-0-521-69395-0. DOI: 10.1017/CBO9780511816574.

[16] Michael Bon, Graziano Vernizzi, Henri Orland, and A. Zee. "Topological Classification of RNA Structures". In: *Journal of Molecular Biology* 379.4 (2008), pp. 900–911. DOI: https://doi.org/10.1016/j.jmb.2008.04.033.

[17] R. Bott and J. P. Mayberry. "Matrices and trees". In: *Economic activity analysis*. Edited by Oskar Morgenstern. John Wiley, Sons, Inc., New York; Chapman, and Hall, Ltd., London, 1954, pp. 391–400.

[18] Boštjan Brešar and Mario Valencia-Pabon. "On the P_3-hull number of Hamming graphs". In: *Discrete Appl. Math.* 282 (2020), pp. 48–52. DOI: 10.1016/j.dam.2019.11.011.

[19] Ian Brierley, Simon Pennell, and Robert J. C. Gilbert. "Viral RNA pseudoknots: versatile motifs in gene expression and replication". In: *Nature Reviews Microbiology* 5.8 (Aug. 2007), pp. 598–610. DOI: 10.1038/nrmicro1704.

[20] Gunnar Brinkmann, Jan Goedgebeur, and Jan-Christoph Schlage-Puchta. "Ramsey numbers $R(K_3, G)$ for graphs of order 10". In: *Electron. J. Combin.* 19.4 (2012), Paper 36, 23. DOI: 10.37236/2548.

[21] Coen Bron and Joep Kerbosch. "Algorithm 457: finding all cliques of an undirected graph". In: *Commun. ACM* 16.9 (Sept. 1973), pp. 575–577. DOI: `10.1145/362342.362367`.

[22] R. L. Brooks. "On colouring the nodes of a network". In: *Proc. Cambridge Philos. Soc.* 37 (1941), pp. 194–197.

[23] J. Brown. "The ribonuclease P database". In: *Nucleic Acids Research* 26.1 (Jan. 1998), pp. 351–352. DOI: `10.1093/nar/26.1.351`.

[24] Dorothy Buck, Danielle O'Donnol, and Kenneth L. Baker. *Unknotting numbers for prime θ-curves up to seven crossings.* 2018. arXiv: `arXiv:1710.05237[math.GT] [math.GT]`.

[25] Neal Bushaw, Blake Conka, Vinay Gupta, Aidan Kierans, Hudson Lafayette, Craig Larson, Sarah Loeb, Kevin McCall, Andriy Mulyar, Scott Taylor, Evan Wainwright, Evan Wilson, and Guanyu Wu. "Bootstrap percolation via automated conjecturing". In: *Ars Math. Contemp.* 23.3 (2023). DOI: `10.26493/1855-3974.2340.a61`.

[26] Kathie Cameron and Kristina Vušković. "Hadwiger's conjecture for some hereditary classes of graphs: a survey". In: *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS* 131 (2020), pp. 20–37.

[27] Jamie J Cannone, Sankar Subramanian, Murray N Schnare, James R Collett, Lisa M D'Souza, Yushi Du, Brian Feng, Nan Lin, Lakshmi V Madabusi, Kirsten M Müller, Nupur Pande, Zhidi Shang, Nan Yu, and Robin R Gutell. "The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs". In: *BMC Bioinformatics* 3.1 (Jan. 2002). DOI: `10.1186/1471-2105-3-2`.

[28] Márcia R. Cappelle, Erika M. M. Coelho, Hebert Coelho, Braully R. Silva, Fábio Protti, and Uéverton S. Souza. "P_3-hull number of graphs with diameter two". In: *The proceedings of Lagos 2019, the tenth Latin and American Algorithms, Graphs and Optimization Symposium (LAGOS 2019)*. Vol. 346. Electron. Notes Theor. Comput. Sci. Elsevier Sci. B. V., Amsterdam, 2019, pp. 309–320.

[29] Márcia R. Cappelle, Erika M. M. Coelho, Hebert Coelho, Braully R. Silva, Uéverton S. Souza, and Fábio Protti. "P_3-convexity on graphs with diameter two: computing hull and interval numbers". In: *Discrete Appl. Math.* 321 (2022), pp. 368–378. DOI: 10.1016/j.dam.2022.07.013.

[30] Diane Castonguay, Erika M. M. Coelho, Hebert Coelho, and Julliano R. Nascimento. "On the geodetic hull number for complementary prisms II". In: *RAIRO Oper. Res.* 55 (2021), S2403–S2415. DOI: 10.1051/ro/2020089.

[31] C. C. Centeno, L. D. Penso, D. Rautenbach, and V. G. Pereira de Sá. "Geodetic number versus hull number in P_3-convexity". In: *SIAM J. Discrete Math.* 27.2 (2013), pp. 717–731. DOI: 10.1137/110859014.

[32] Carmen C. Centeno, Mitre C. Dourado, Lucia Draque Penso, Dieter Rautenbach, and Jayme L. Szwarcfiter. "Irreversible conversion of graphs". In: *Theoret. Comput. Sci.* 412.29 (2011), pp. 3693–3700. DOI: 10.1016/j.tcs.2011.03.029.

[33] Jan Chalupa, Paul L. Leath, and Guillaume R. Reich. "Bootstrap percolation on a Bethe lattice". In: *Journal of Physics C: Solid State Physics* 12.1 (1979), pp. L31–L35. DOI: 10.1088/0022-3719/12/1/008.

[34] Gary Chartrand and Raymond E. Pippert. "Locally connected graphs". In: *Časopis Pěst. Mat.* 99 (1974), pp. 158–163.

[35] Ricky X. F. Chen, Christian M. Reidys, and Michael S. Waterman. "RNA secondary structures with given motif specification: combinatorics and algorithms". In: *Bull. Math. Biol.* 85.3 (2023), Paper No. 21, 24. DOI: `10.1007/s11538-023-01128-5`.

[36] William Y. C. Chen, Eva Y. P. Deng, Rosena R. X. Du, Richard P. Stanley, and Catherine H. Yan. "Crossings and nestings of matchings and partitions". In: *Trans. Amer. Math. Soc.* 359.4 (2007), pp. 1555–1575. DOI: `10.1090/S0002-9947-06-04210-3`.

[37] S. Chmutov, S. Duzhin, and J. Mostovoy. *Introduction to Vassiliev knot invariants*. Cambridge University Press, Cambridge, 2012, pp. xvi+504. ISBN: 978-1-107-02083-2. DOI: `10.1017/CBO9781139107846`.

[38] V. Chvátal. "Recognizing decomposable graphs". In: *J. Graph Theory* 8.1 (1984), pp. 51–53. DOI: `10.1002/jgt.3190080106`.

[39] Mihai Ciucu, Weigen Yan, and Fuji Zhang. "The number of spanning trees of plane graphs with reflective symmetry". In: *J. Combin. Theory Ser. A* 112.1 (2005), pp. 105–116. DOI: `10.1016/j.jcta.2005.01.007`.

[40] Erika M. M. Coelho, Hebert Coelho, Julliano R. Nascimento, and Jayme L. Szwarcfiter. "On the P_3-hull number of some products of graphs". In: *Discrete Appl. Math.* 253 (2019), pp. 2–13. DOI: `10.1016/j.dam.2018.04.024`.

[41] Michael Dairyko, Michael Ferrara, Bernard Lidický, Ryan R. Martin, Florian Pfender, and Andrew J. Uzzell. "Ore and Chvátal-type degree conditions for bootstrap percolation from small sets". In: *J. Graph Theory* 94.2 (2020), pp. 252–266. DOI: `10.1002/jgt.22517`.

[42] C. Dalfó. "A survey on the missing Moore graph". In: *Linear Algebra Appl.* 569 (2019), pp. 1–14. DOI: 10.1016/j.laa.2018.12.035.

[43] Padideh Danaee. *bpRNA*. https://github.com/padidehdanaee/bpRNA. 2018.

[44] Padideh Danaee, Mason Rouches, Michelle Wiley, Dezhong Deng, Liang Huang, and David Hendrix. "bpRNA: large-scale automated annotation and analysis of RNA secondary structure". In: *Nucleic Acids Research* 46.11 (May 2018), pp. 5381–5394. DOI: 10.1093/nar/gky285.

[45] Lucia Draque Penso, Fábio Protti, Dieter Rautenbach, and Uéverton S. Souza. "On P3-Convexity of Graphs with Bounded Degree". In: *Algorithmic Aspects in Information and Management*. Ed. by Qianping Gu, Pavol Hell, and Boting Yang. Cham: Springer International Publishing, 2014, pp. 263–274.

[46] Marcio Antônio Duarte, Lucia Penso, Dieter Rautenbach, and Uéverton dos Santos Souza. "Complexity properties of complementary prisms". In: *J. Comb. Optim.* 33.2 (2017), pp. 365–372. DOI: 10.1007/s10878-015-9968-5.

[47] P. Duchet and H. Meyniel. "On Hadwiger's number and the stability number". In: *Graph theory (Cambridge, 1981)*. Vol. 62. North-Holland Math. Stud. North-Holland, Amsterdam-New York, 1982, pp. 71–73. ISBN: 0-444-86449-0.

[48] P. Dumas, D. Moras, C. Florentz, R. Giegé, P. Verlaan, A. Van Belkum, and C. W.A. Pleij. "3-D Graphics Modelling of the tRNA-Like 3'-End of Turnip Yellow Mosaic Virus RNA: Structural and Functional Implications". In: *Journal of Biomolecular Structure and Dynamics* 4.5 (Apr. 1987), pp. 707–728. DOI: 10.1080/07391102.1987.10507674.

[49] David Elliott and Michael Ladomery. *Molecular Biology of RNA*. Oxford University Press, Aug. 2023. ISBN: 9780191975820. DOI: `10.1093/hesc/9780199671397.001.0001`.

[50] Matthew Elpers, Rayan Ibrahim, and Allison H. Moore. *Determinants of Simple Theta Curves and Symmetric Graphs*. 2022. arXiv: `2211.00626 [math.GT]`.

[51] Hubert de Fraysseix. "A characterization of circle graphs". In: *European J. Combin.* 5.3 (1984), pp. 223–238. DOI: `10.1016/S0195-6698(84)80005-0`.

[52] Daniel Freund, Matthias Poloczek, and Daniel Reichman. "Contagious sets in dense graphs". In: *European J. Combin.* 68 (2018), pp. 66–78. DOI: `10.1016/j.ejc.2017.07.011`.

[53] Catherine Gille and Louis-Hadrien Robert. "A signature invariant for knotted Klein graphs". In: *Algebr. Geom. Topol.* 18.6 (2018), pp. 3719–3747. DOI: `10.2140/agt.2018.18.3719`.

[54] Lebrecht Goeritz. "Knoten und quadratische Formen". In: *Math. Z.* 36.1 (1933), pp. 647–654. DOI: `10.1007/BF01188642`.

[55] R. L. Graham. "On primitive graphs and optimal vertex assignments". In: *Ann. New York Acad. Sci.* 175 (1970), pp. 170–186.

[56] R. L. Graham and S. Silverman. "Problem 20". In: *Proceedings of the Calgary International Conference on Combinatorial Structures and their Applications held at the University of Calgary, Calgary, Alberta, Canada, June, 1969*. Ed. by Richard Guy, Haim Hanani, Norbert Sauer, and Johanan Schönheim. Gordon and Breach Science Publishers, New York-London-Paris, 1970, pp. 499–500.

[57] S. Griffiths-Jones. “Rfam: an RNA family database”. In: *Nucleic Acids Research* 31.1 (Jan. 2003), pp. 439–441. DOI: `10.1093/nar/gkg006`.

[58] Luciano N. Grippo, Adrián Pastine, Pablo Torres, Mario Valencia-Pabon, and Juan C. Vera. “On the P_3-hull number of Kneser graphs”. In: *Electron. J. Combin.* 28.3 (2021), Paper No. 3.32, 9. DOI: `10.37236/9903`.

[59] Karen Gunderson. “Minimum degree conditions for small percolating sets in bootstrap percolation”. In: *Electron. J. Combin.* 27.2 (2020), Paper No. 2.37, 22. DOI: `10.37236/6937`.

[60] H. Hadwiger. “Über eine Klassifikation der Streckenkomplexe”. In: *Vierteljschr. Naturforsch. Ges. Zürich* 88 (1943), pp. 133–142.

[61] Teresa W. Haynes, Michael A. Henning, Peter J. Slater, and Lucas C. van der Merwe. “The complementary product of two graphs”. In: *Bull. Inst. Combin. Appl.* 51 (2007), pp. 21–30.

[62] Jaka Hedžet and Michael A. Henning. 3-*Neighbor bootstrap percolation on grids*. 2023. arXiv: `2307.14033 [math.CO]`.

[63] A. J. Hoffman and R. R. Singleton. “On Moore graphs with diameters 2 and 3”. In: *IBM J. Res. Develop.* 4 (1960), pp. 497–504. DOI: `10.1147/rd.45.0497`.

[64] Alexander E. Holroyd. “Sharp metastability threshold for two-dimensional bootstrap percolation”. In: *Probab. Theory Related Fields* 125.2 (2003), pp. 195–224. DOI: `10.1007/s00440-002-0239-x`.

[65] Wing-Kai Hon, Ton Kloks, and Hsiang-Hsuan Liu. “On the P3-convexity of some classes of graphs with few P4s and permutation graphs”. In: *Proceedings of the Australasian Computer Science Week Multiconference.* ACSW

’16. Canberra, Australia: Association for Computing Machinery, 2016. ISBN: 9781450340427. DOI: 10.1145/2843043.2843055.

[66] J. A. Howell, T. F. Smith, and M. S. Waterman. “Computation of generating functions for biological molecules”. In: *SIAM J. Appl. Math.* 39.1 (1980), pp. 119–133. DOI: 10.1137/0139010.

[67] Rayan Ibrahim, Hudson LaFayette, and Kevin McCall. “Bootstrap percolation, connectivity, and graph distance”. In: *The Art of Discrete and Applied Mathematics* (Mar. 2024). DOI: 10.26493/2590-9770.1694.1cf.

[68] F. Juhling, M. Morl, R. K. Hartmann, M. Sprinzl, P. F. Stadler, and J. Putz. “tRNAdb 2009: compilation of tRNA sequences and tRNA genes”. In: *Nucleic Acids Research* 37.Database (Jan. 2009), pp. D159–D162. DOI: 10.1093/nar/gkn772.

[69] Christian Kassel and Vladimir Turaev. “Chord diagram invariants of tangles and graphs”. In: *Duke Math. J.* 92.3 (1998), pp. 497–552. DOI: 10.1215/S0012-7094-98-09215-8.

[70] Marcel Kucharík, Ivo L. Hofacker, Peter F. Stadler, and Jing Qin. “Pseudoknots in RNA folding landscapes”. In: *Bioinformatics* 32.2 (Oct. 2015), pp. 187–194. DOI: 10.1093/bioinformatics/btv572.

[71] Sergei K. Lando and Alexander K. Zvonkin. *Graphs on surfaces and their applications.* Vol. 141. Encyclopaedia of Mathematical Sciences. With an appendix by Don B. Zagier, Low-Dimensional Topology, II. Springer-Verlag, Berlin, 2004, pp. xvi+455. ISBN: 3-540-00203-0. DOI: 10.1007/978-3-540-38361-1.

[72] C. E. Larson and N. Van Cleemput. "Automated conjecturing I: Fajtlowicz's Dalmatian heuristic revisited". In: *Artificial Intelligence* 231 (2016), pp. 17–38. DOI: `10.1016/j.artint.2015.10.002`.

[73] C. E. Larson and N. Van Cleemput. "Automated conjecturing III: Property-relations conjectures". In: *Ann. Math. Artif. Intell.* 81.3-4 (2017), pp. 315–327. DOI: `10.1007/s10472-017-9559-5`.

[74] W. B. Raymond Lickorish. *An introduction to knot theory*. Vol. 175. Graduate Texts in Mathematics. Springer-Verlag, New York, 1997, pp. x+201. ISBN: 0-387-98254-X. DOI: `10.1007/978-1-4612-0691-0`.

[75] Rick Litherland. "The Alexander module of a knotted theta-curve". In: *Math. Proc. Cambridge Philos. Soc.* 106.1 (1989), pp. 95–106. DOI: `10.1017/S0305004100068018`.

[76] László Lovász. *Combinatorial problems and exercises*. second. AMS Chelsea Publishing, Providence, RI, 2007, p. 642. ISBN: 978-0-8218-4262-1. DOI: `10.1090/chel/361`.

[77] W Mantel. "Opgaven28". In: *Wiskd. Opgaven Met Oplossingen* 10 (1907), pp. 60–61.

[78] Sebastián González Hermosillo de la Maza, Yifan Jing, and Masood Masjoody. *On the structure of (claw, bull)-free graphs*. 2019. DOI: `10.48550/ARXIV.1901.00043`.

[79] Kevin McCall. "Investigations in the semi-strong product of graphs and bootstrap percolation". PhD thesis. 2023. DOI: `10.25772/0FN2-VX93`.

[80] Brendan McKay. *Description of graph6, sparse6, and digraph6 encodings*. `https://users.cecs.anu.edu.au/~bdm/data/formats`.

[81] Brendan D. McKay and Adolfo Piperno. "Practical graph isomorphism, II". In: *J. Symbolic Comput.* 60 (2014), pp. 94–112. DOI: 10.1016/j.jsc.2013.09.003.

[82] Terry A. McKee and F. R. McMorris. *Topics in intersection graph theory.* SIAM Monographs on Discrete Mathematics and Applications. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999, pp. viii+205. ISBN: 0-89871-430-3. DOI: 10.1137/1.9780898719802.

[83] Kyle Miller. *KnotFolio.* https://github.com/kmill/knotfolio. 2019.

[84] Auriane Monestier, Alexey Aleksandrov, Pierre-Damien Coureux, Michel Panvert, Yves Mechulam, and Emmanuelle Schmitt. "The structure of an E. coli tRNAfMet A1–U72 variant shows an unusual conformation of the A1–U72 base pair". In: *RNA* 23.5 (Jan. 2017), pp. 673–682. DOI: 10.1261/rna.057877.116.

[85] Auriane Monestier, Yves Mechulam, and Emmanuelle Schmitt. *Structure of an E.coli initiator tRNAfMet A1-U72 variant.* 2017. DOI: 10.2210/pdb5L4O/pdb.

[86] Gadi Moran. "Chords in a circle and linear algebra over GF(2)". In: *J. Combin. Theory Ser. A* 37.3 (1984), pp. 239–247. DOI: 10.1016/0097-3165(84)90048-7.

[87] Hiromasa Moriuchi. "An enumeration of theta-curves with up to seven crossings". In: *J. Knot Theory Ramifications* 18.2 (2009), pp. 167–197. DOI: 10.1142/S0218216509006884.

[88] J. Mycielski. "Sur le coloriage des graphs". In: *Colloq. Math.* 3 (1955), pp. 161–162. DOI: 10.4064/cm-3-2-161-162.

[89] Julliano R. Nascimento, Erika M.M. Coelho, Hebert Coelho, and Jayme L. Szwarcfiter. "On the Complexity of the P3-Hull Number of the Cartesian Product of Graphs". In: *Electronic Notes in Discrete Mathematics* 55 (Nov. 2016), pp. 169–172. DOI: `10.1016/j.endm.2016.10.042`.

[90] P. K. Neethu, Ullas Chandran S. V., and Julliano R. Nascimento. "On the monophonic convexity in complementary prisms". In: *Discrete Appl. Math.* 343 (2024), pp. 224–230. DOI: `10.1016/j.dam.2023.10.024`.

[91] R. C. Penner and Michael S. Waterman. "Spaces of RNA secondary structures". In: *Adv. Math.* 101.1 (1993), pp. 31–49. DOI: `10.1006/aima.1993.1039`.

[92] Vincent Pilaud and Juanjo Rué. "Analytic combinatorics of chord and hyperchord diagrams with k crossings". In: *Adv. in Appl. Math.* 57 (2014), pp. 60–100. DOI: `10.1016/j.aam.2014.04.001`.

[93] Michael D. Plummer, Michael Stiebitz, and Bjarne Toft. "On a special case of Hadwiger's conjecture". In: *Discuss. Math. Graph Theory* 23.2 (2003), pp. 333–363. DOI: `10.7151/dmgt.1206`.

[94] Michael Polyak and Oleg Viro. "Gauss diagram formulas for Vassiliev invariants". In: *Internat. Math. Res. Notices* 11 (1994), pp. 445–453. DOI: `10.1155/S1073792894000486`.

[95] Daniel Reichman. "New bounds for contagious sets". In: *Discrete Math.* 312.10 (2012), pp. 1812–1814. DOI: `10.1016/j.disc.2012.01.016`.

[96] Christian Reidys. *Combinatorial computational biology of RNA*. Pseudoknots and neutral networks. Springer, New York, 2011, pp. x+257. ISBN: 978-0-387-76730-7. DOI: `10.1007/978-0-387-76731-4`.

[97] Christian M. Reidys, Fenix W. D. Huang, Jørgen E. Andersen, Robert C. Penner, Peter F. Stadler, and Markus E. Nebel. “Topology and prediction of RNA pseudoknots”. In: *Bioinformatics* 27.8 (Feb. 2011), pp. 1076–1085. DOI: `10.1093/bioinformatics/btr090`. eprint: `https://academic.oup.com/bioinformatics/article-pdf/27/8/1076/17102827/btr090.pdf`.

[98] K. Rietveld, R. Van Poelgeest, C.W.A. Pleij, J.H. Van Boom, and L. Bosch. “The tRNA-Uke structure at the 3’ terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA”. In: *Nucleic Acids Research* 10.6 (1982), pp. 1929–1946. DOI: `10.1093/nar/10.6.1929`.

[99] M. A. Rosenblad. “SRPDB: Signal Recognition Particle Database”. In: *Nucleic Acids Research* 31.1 (Jan. 2003), pp. 363–364. DOI: `10.1093/nar/gkg107`.

[100] Makoto Sakuma. “On strongly invertible knots”. In: *Algebraic and topological theories (Kinosaki, 1984)*. Kinokuniya, Tokyo, 1986, pp. 176–196.

[101] William R. Schmitt and Michael S. Waterman. “Linear trees and RNA secondary structure”. In: *Discrete Appl. Math.* 51.3 (1994), pp. 317–323. DOI: `10.1016/0166-218X(92)00038-N`.

[102] David Sehnal, Sebastian Bittrich, Mandar Deshpande, Radka Svobodová, Karel Berka, Václav Bazgier, Sameer Velankar, Stephen K Burley, Jaroslav Koča, and Alexander S Rose. “Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures”. In: *Nucleic Acids Research* 49.W1 (May 2021), W431–W437. DOI: `10.1093/nar/gkab314`. eprint: `https://academic.oup.com/nar/article-pdf/49/W1/W431/38842088/gkab314.pdf`.

[103] Alexander Serganov. *Crystal Structure of the Thermotoga Maritima Lysine Riboswitch Bound To S-(2-aminoethyl)-L-cysteine.* 2008. DOI: 10.2210/pdb3DIG/pdb.

[104] Alexander Serganov, Lili Huang, and Dinshaw J. Patel. "Structural insights into amino acid binding and gene control by a lysine riboswitch". In: *Nature* 455.7217 (Sept. 2008), pp. 1263–1267. DOI: 10.1038/nature07326.

[105] Paul Seymour. "Hadwiger's conjecture". In: *Open problems in mathematics.* Springer, [Cham], 2016, pp. 417–437.

[106] Wenjie Shu, Xiaochen Bo, Zhiqiang Zheng, and Shengqi Wang. "A novel representation of RNA secondary structure based on element-contact graphs". In: *BMC Bioinformatics* 9.1 (Apr. 2008). DOI: 10.1186/1471-2105-9-188.

[107] Robert Singleton. "There is no irregular Moore graph". In: *Amer. Math. Monthly* 75 (1968), pp. 42–43. DOI: 10.2307/2315106.

[108] Richard P. Stanley. *Enumerative combinatorics. Vol. 1.* Vol. 49. Cambridge Studies in Advanced Mathematics. With a foreword by Gian-Carlo Rota, Corrected reprint of the 1986 original. Cambridge University Press, Cambridge, 1997, pp. xii+325. ISBN: 0-521-55309-1. DOI: 10.1017/CBO9780511805967.

[109] David W Staple and Samuel E Butcher. "Pseudoknots: RNA Structures with Diverse Functions". In: *PLoS Biology* 3.6 (June 2005), e213. DOI: 10.1371/journal.pbio.0030213.

[110] Gary M. Studnicka, Georgia M. Rahn, Ian W. Cummings, and Winston A. Salser. "Computer method for predicting the secondary structure of single-stranded RNA". In: *Nucleic Acids Research* 5.9 (1978), pp. 3365–3388. DOI: 10.1093/nar/5.9.3365.

[111] Ignacio Tinoco, Olke C. Uhlenbeck, and Mark D. Levine. “Estimation of Secondary Structure in Ribonucleic Acids”. In: *Nature* 230.5293 (Apr. 1971), pp. 362–367. DOI: `10.1038/230362a0`.

[112] Jacques Touchard. “Sur un problème de configurations et sur les fractions continues”. In: *Canad. J. Math.* 4 (1952), pp. 2–25. DOI: `10.4153/cjm-1952-001-8`.

[113] H. F. Trotter. “Homology of group systems with applications to knot theory”. In: *Ann. of Math. (2)* 76 (1962), pp. 464–498. DOI: `10.2307/1970369`.

[114] Vladimir Turaev. “Knotoids”. In: *Osaka J. Math.* 49.1 (2012), pp. 195–223.

[115] Paul Turán. “Eine Extremalaufgabe aus der Graphentheorie”. In: *Mat. Fiz. Lapok* 48 (1941), pp. 436–452.

[116] W. T. Tutte. “The dissection of equilateral triangles into equilateral triangles”. In: *Proc. Cambridge Philos. Soc.* 44 (1948), pp. 463–482. DOI: `10.1017/s030500410002449x`.

[117] M.S. Waterman and T.F. Smith. “RNA secondary structure: a complete mathematical analysis”. In: *Mathematical Biosciences* 42.3 (1978), pp. 257–266. DOI: `https://doi.org/10.1016/0025-5564(78)90099-8`.

[118] Michael S. Waterman. “Combinatorics of RNA hairpins and cloverleaves”. In: *Stud. Appl. Math.* 60.2 (1979), pp. 91–96. DOI: `10.1002/sapm197960291`.

[119] Michael S. Waterman. “Secondary structure of single-stranded nucleic acids”. In: *Studies in foundations and combinatorics*. Vol. 1. Adv. Math. Suppl. Stud. Academic Press, New York-London, 1978, pp. 167–212. ISBN: 0-12-599101-0.

[120] J.D. Watson, T.A. Baker, S.P. Bell, A. Gann, M. Levine, and R. Losick. *Molecular Biology of the Gene*. Pearson Education, 2013. ISBN: 9780321896704.

[121] Alexandra Wesolek. “Bootstrap percolation in Ore-type graphs”. In: (2019). DOI: 10.48550/ARXIV.1909.04649.

[122] Douglas B. West. *Introduction to graph theory*. Prentice Hall, Inc., Upper Saddle River, NJ, 1996, pp. xvi+512. ISBN: 0-13-227828-6.

[123] Fuji Zhang and Weigen Yan. “Enumerating spanning trees of graphs with an involution”. In: *J. Combin. Theory Ser. A* 116.3 (2009), pp. 650–662. DOI: 10.1016/j.jcta.2008.10.004.

[124] C. Zwieb. “tmRDB (tmRNA database)”. In: *Nucleic Acids Research* 31.1 (Jan. 2003), pp. 446–447. DOI: 10.1093/nar/gkg019.