Combinatorial Problems in Graph Drawing and Knowledge Representation

BY

DESPINA STASI
B.S. Applied Mathematics (Illinois Institute of Technology) 2003
B.S. Computer Science (Illinois Institute of Technology) 2003
M.S. Mathematics (University of Illinois at Chicago) 2006

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois at Chicago, 2012

Chicago, Illinois

Defense Committee:

György Turán, Chair and Advisor
Robert H. Sloan, Secondary Advisor, Computer Science
Shmuel Friedland
Dhruv Mubayi
Marcus Schaefer, DePaul University

Copyright by

Despina Stasi

2012

To my husband,

Aleksandar Leposavić,

for his unwavering love and support

and his unshaken belief in me.

And to my parents,

and especially my mother,

for instilling in me a profound love

for education and knowledge.

ACKNOWLEDGMENTS

I am extremely grateful to my advisor, György Turán, without whose kind support, patience and insightful guidance this thesis would not be possible.

I owe my sincere gratitude to Robert H. Sloan, who served as co-advisor for this thesis, for his support, and especially his enthusiasm and guidance during a crucial semester of the thesis.

I want to express my warm and sincere thanks to Michael J. Pelsmajer and Marcus Schaefer, for affording me my first opportunity at research and for making it an instructive and pleasant experience.

Finally, I would like to thank Dhruv Mubayi and Shmuel Friedland, who graciously donated their time to serve as committee members for this thesis.

TABLE OF CONTENTS

CHAPTER **PAGE**

1 **INTRODUCTION** 1

2 **STRONG HANANI-TUTTE ON THE PROJECTIVE PLANE** 3
2.1 Introduction 3
2.2 From Embeddings to Drawings 5
2.2.1 Basic Observations and Redrawing Tools 5
2.2.2 K-graphs and iocr 8
2.3 Proof of the Main Theorem 11

3 **RANDOM HORN FORMULAS AND PROPAGATION CONNECTIVITY FOR DIRECTED HYPERGRAPHS** 16
3.1 Introduction 16
3.2 Preliminaries 18
3.3 A lower bound 19
3.4 An upper bound 21
3.5 Open Problems 26

4 **HORN MINIMIZATION: HYDRA FORMULAS AND (SOME) STEINER MINIMIZATION** 28
4.1 Introduction 28
4.2 Background 31
4.3 Steiner minimization 32
4.3.1 Steiner Basics 32
4.3.2 Algorithm 33
4.4 Hydra Formulas 38
4.5 The Hydra Number of Graphs 40
4.6 Line Graphs 43
4.7 The Hydra Number of Trees 48
4.7.1 Trees with Low Hydra Number 49
4.7.2 Lower Bounds for Trees 52
4.7.3 Spiders and Large In-degrees 53
4.7.4 Complete Binary Trees 56
4.8 Hydra Number and Total Interval Number 61
4.9 Minimal Directed Hypergraphs with All k-tuples Good 62
4.10 Open Problems 64

APPENDICES 65

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

Appendix A 66

CITED LITERATURE 73

VITA 77

LIST OF FIGURES

FIGURE		PAGE
1	Removing a self-intersection; illustration from [37].	6
2	Vertex x in E_{22} (*left*) and in A_2 (*right*); illustration from [36].	13
3	F_1, D_3 and B_1 (left to right) with labels; illustration from [36].	14
4	$E_{18} = K_{4,4} - u_4v_4$; illustration from [36].	14
5	Single-headed graph G_k for $k = 4$ from Theorem 4.6.4 (left) and its line graph (right).	47
6	The forbidden subgraph for caterpillars.	49
7	Part of the non-caterpillar tree T from the proof of Lemma 4.7.3.	51
8	Induction on the line graph $L(B_{4k+3})$ from the proof of Theorem 4.7.9.	58
9	Line graph $L(B_{3k})$ from the proof of Theorem 4.7.10.	60

LIST OF ABBREVIATIONS

CNF	Conjunctive Normal Form
a.a.s.	asymptotically almost surely
w.h.p.	with high probability

SUMMARY

In this thesis we consider some combinatorial problems in graph drawing and knowledge representation.

We begin with the problem of extending the Hanani-Tutte Theorem to the projective plane.

A graph is planar if and only if it can be drawn on the plane so that every pair of non-adjacent edges cross an even number of times. (Hanani-Tutte theorem [16; 43].)

Of course, necessity is trivial since any planar drawing of a graph G satisfies the conclusion. The surprising part of the theorem is that it is enough to find a drawing of G where no pair of independent edges crosses an odd number of times, to conclude that there is a planar drawing of the graph.

A weaker version of this theorem requires all adjacent pairs of edges to cross evenly as well as non-adjacent ones. This weaker version of the Hanani-Tutte theorem has been shown to hold on all orientable sufaces by Cairns and Nikolayevsky [15]. Later, an elementary proof which also extends to non-orientable surfaces was given by Pelsmajer, Schaefer, and Štefankovič [38].

In Chapter 2 we show that the strong version of the Hanani-Tutte theorem holds for the projective plane. Namely we show that:

If a graph G can be drawn in the projective plane so that every two non-adjacent edges cross evenly, then G can be embedded in the projective plane.

By Archdeacon, Glover, Huneke, and Wang [4; 22] any non-projective planar graph G must contain one of 35 minimal forbidden minors for the projective plane. The proof of the strong

Hanani-Tutte theorem on the projective plane relies on proving that for any drawing of any of the 35 minors on the projective plane, there must be a pair of non-adjacent edges that cross an odd number of times.

For our second problem we consider the property that in a random definite Horn formula of size-3 clauses over n variables, where every such clause is included with probability p, there is a pair of variables for which forward chaining produces all other variables. We show that *a.a.s.* the property does not hold for $p \leq 1/(11n \ln n)$, and *a.a.s.* it does hold for $p \geq (5 \ln \ln n)/(n \ln n)$.

We study random definite Horn formulas using the language of random uniform directed hypergraphs, a generalization of Erdős-Rényi random graphs. We construct a random directed 3-uniform hypergraph H over n vertices by including every hyperarc $u, v \to w$ with probability p, where p is a function of n. The (u, v)-propagation component is the closure of the pair (u, v) under forward chaining: starting from the pair u, v, and while there exists a hyperarc $x, y \to z$ in H such that x and y are both in the (u, v)-propagation component, and z is not, add z to the (u, v)-propagation component. A hypergraph H is propagation connected if for some pair (u, v), the (u, v)-propagation component includes all vertices.

The related property that every (u, v)-propagation component in a random 3-uniform directed hypergraph contains all vertices has a threshold function $p = 2 \ln n/n$ [31]. The property of propagation connectivity was introduced for undirected hypergraphs by Berke and Onsjö, who were motivated by a certain type of constraint satisfaction problems [8]. They showed that *a.a.s.* an undirected random 3-uniform hypergraph is not propagation connected for $p \leq 1/(n(\log n)^2)$, and *a.a.s.* it is propagation connected for $p \geq 1/(n(\log n)^{0.4})$. This undirected case has been improved by Coja-Oghlan, Onsjö and Watanabe to upper and lower bounds of the order $p = 1/n \log n$ [18].

We conclude Chapter 3 with some open problems on propagation connectivity of directed hypergraphs motivated by definite Horn formulas.

Finally we consider two problems in Horn minimization. Horn minimization is concerned with finding among all equivalent Horn formulas, a Horn formula with the minimum number of clauses. Horn minimization is NP-Complete [6], and cannot be approximated in polynomial time within a factor of $2^{\log^{O(1-o(1))} n}$ unless P=NP [11; 14]. However there exists an $(n-1)$-approximation algorithm for minimizing Horn formulas on n variables [25].

We begin Chapter 4 by briefly considering the problem of Steiner minimization introduced by Bhattacharya, DasGupta, Mubayi and Turán [11], where we are allowed to add new variables in a restricted way to find the smallest Horn formula ψ that is "equivalent" in a general sense to a given Horn formula φ, in that ψ implies a clause C over the original set of variables if and only if φ implies C. Steiner minimization is MAX-SNP-hard. We update slightly an efficient $O(n \log\log n/(\log n)^{1/4})$ approximation algorithm of [11] and its analysis to show that Steiner minimization can be efficiently approximated within a factor of $O(n/\log n)$.

A significant step in the Steiner minimization approximation algorithm involves minimizing the given Horn formula with respect to the number of distinct bodies it contains. Since this problem can be solved efficiently [3; 5; 23; 32], it provides a useful approach to Horn minimization. The following question can then be raised: given a Horn formula φ, and a set of distinct bodies what is the smallest Horn formula equivalent to φ?

This question brings us to the main focus of the Chapter 4: hydra formulas and hydra numbers. We define a hydra formula to be a definite Horn formula of size 3-clauses on n variables, possessing the property that the forward chaining procedure starting from any body of the hydra formula, produces all n variables. For the remainder of Chapter 4 we are interested in minimizing hydra formulas. In other words, we are interested in finding the minimum number of clauses in a Horn formula equivalent to a given hydra formula. This number depends entirely on the set of bodies that appear in the hydra formula, and as such it is really a parameter of the graph formed by the bodies of the hydra.

With this in mind, we define the hydra number of a graph, $h(G)$, to be the minimum number of clauses in any hydra formula with a specific set of bodies. Clearly $h(G)$ must be at least equal to the number of edges of the graph. In the case where equality holds we call the graph single-headed. It is not hard to see that hamiltonian graphs are single-headed. In fact, we show that a more general sufficient condition for single-headedness is the hamiltonicity of the line graph of one of its spanning subgraphs. On the other hand, this is not a necessary condition for single-headedness. On the contrary, we construct a family of single-headed graphs such that every spanning subgraph has a line graph with a large path cover number.

We proceed to give bounds for the hydra number. We show by construction that the hydra number of a graph is at most equal to the path cover number of the line graph of its spanning subgraphs, where neither the graph nor the subgraphs are allowed isolated vertices. On the other hand, the quantity $\ell(G)$, which counts the number of degree-2 vertices neighboring a degree-1 vertex in the graph G, gives a lower bound for $h(G)$.

SUMMARY (Continued)

The remainder of Chapter 4, investigates the hydra number of trees. We show that the only trees that are single-headed are stars, and the only trees with hydra number $|E(G)| + 1$ are caterpillars. We use the two results in the previous paragraph, to give close bounds for the hydra number of a complete binary tree, namely $\frac{9}{8}|E(G)| \leq h(G) \leq \frac{17}{15}|E(G)| + 1$. The two bounds coincide to give an exact answer for the hydra number of a spider tree T, specifically $h(T) = |E(T)| + \ell(T)$.

Spiders prove to be a useful family of trees for two more reasons. Firstly the construction of a family of spiders T_k is an example demonstrating the existence of graphs for which optimal hydra formulas require a high-in-degree variable, and where limiting the in-degree of the variables to a constant can double the number of clauses in excess of $|E(T_k)|$ required in such a formula. Secondly the hydra number of T_k is the maximum possible (as a function of $|V(G)|$) hydra number over all trees.

Finally we present a result on a problem that is related to hydra minimization, which asks for the minimum number of clauses in a Horn formula of size 3-clauses such that, for a fixed integer k, every k-tuple of the variables implies, through forward chaining, all other variables. We conclude the chapter with some open problems on hydra formulas.

CHAPTER 1

INTRODUCTION

Combinatorics is the study of discrete mathematical objects, including enumerating objects satisfying a certain property and finding minimal or maximal, or, in some sense optimal objects satisfying a property. Combinatorial problems arise in many areas of mathematics and the field is tightly intertwined with theoretical computer science. In this thesis, we consider combinatorial problems and applications in graph drawing and knowledge representation.

We begin with a graph drawing problem. By graph drawing we mean the field of graph theory concerned with the geometric representation of a graph. One of the basic questions in graph drawing involves determining the crossing number of a graph, the minimum number of edge crossings in any drawing of a graph in the plane. An assertion that the crossing number of a graph is zero is equivalent to an assertion that the graph is planar. Crossing numbers have been defined for drawings satisfying specific restrictions, for example, requiring the edges to be drawn as straight line segments. One may also only count the number of pairs of independent edges that cross, or the number of pairs of independent edges that cross an odd number of times. Such questions can also be posed for surfaces other than the plane. In Chapter 2 we show that graphs that can be embedded in the projective plane have independent odd crossing number equal to zero.

Knowledge representation is the field of artificial intelligence concerned with how an intelligent agent uses formal symbols to accurately and efficiently represent and reason about sets of facts and rules. For our purposes we envision a knowledge base consisting of rules and facts represented by Horn formulas, where reasoning is performed through resolution, and more specifically the forward chaining procedure.

In Chapter 3 we consider a problem motivated by a knowledge base consisting of Horn clauses of size 3 over n variables, selected uniformly at random. We consider the existence of a threshold for the property that some pair of variables implies through forward chaining all n variables.

As there are many equivalent Horn formulas to encode the knowledge of an agent, it is extremely useful to pick the shortest representation, a problem called Horn minimization. In Chapter 4 we study two problems in Horn minimization. Firstly we consider the problem of approximating the smallest Horn formula using new variables, that can be considered as internal to the knowledge base. Secondly we study the size of smallest hydra formulas, a subclass of definite Horn formulas.

CHAPTER 2

STRONG HANANI-TUTTE ON THE PROJECTIVE PLANE

This chapter is based on joint work with Michael J. Pelsmajer of Illinois Institute of Technology and Marcus Schaefer of DePaul University published in [36].

2.1 Introduction

In the plane there is a beautiful characterization of planar graphs known as the Hanani-Tutte theorem: a graph is planar if and only if it can be drawn in the plane so that every two non-adjacent edges cross an even number of times. Equivalently, any drawing of a non-planar graph in the plane must contain two non-adjacent edges that cross oddly.

There are several proofs of the Hanani-Tutte theorem, including the original 1934 proof by Hanani [16] and the 1970 proof by Tutte [43] (see [37] for more references). Our goal in the current chapter is to show that the result remains true in the projective plane.[1]

Theorem 2.1.1. *If a graph G can be drawn in the projective plane so that every two non-adjacent edges cross evenly, then G can be embedded in the projective plane.*

This is not the first result that indicates that the Hanani-Tutte theorem is not a special property of the plane. Using homology theory, Cairns and Nikolayevsky [15] showed that if a graph can be drawn on an orientable surface so that every pair of edges (not just non-adjacent

[1]A sphere with a crosscap. We assume that the reader is familiar with the basic terminology of drawings and embeddings in surfaces. For background see [34; 20].

ones) crosses an even number of times, then the graph can be embedded in that surface. Pelsmajer, Schaefer, and Štefankovič [38] gave a new, elementary proof of this weak Hanani-Tutte theorem that also establishes the result for non-orientable surfaces. Theorem 2.1.1 is the first time the strong version of the Hanani-Tutte theorem has been established for any higher-order surface.

There is an alternative view of the Hanani-Tutte theorem in terms of crossing numbers. The *crossing number* of a graph G, denoted by $\mathrm{cr}_S(G)$, is the minimum number of crossings in any drawing of G in surface S. Hence a graph G is embeddable in S if and only if $\mathrm{cr}_S(G) = 0$. The *odd crossing number* of G, denoted by $\mathrm{ocr}_S(G)$, is the minimum number of pairs of edges that cross oddly in any drawing of G in surface S. The *independent odd crossing number* of G, $\mathrm{iocr}_S(G)$, is the minimum number of pairs of non-adjacent edges that cross oddly in any drawing of G in surface S.

The strong Hanani-Tutte theorem can now simply be stated as "$\mathrm{iocr}(G) = 0$ implies $\mathrm{cr}(G) = 0$" and Theorem 2.1.1 becomes "$\mathrm{iocr}_{\mathrm{N}_1}(G) = 0$ implies $\mathrm{cr}_{\mathrm{N}_1}(G) = 0$" using N_1 as a symbol for the projective plane. The weak Hanani-Tutte theorem in this notation reads "$\mathrm{ocr}_S(G) = 0$ implies $\mathrm{cr}_S(G) = 0$" and is true for all surfaces S as we mentioned above. The crossing number point of view emphasizes the algebraic nature of the Hanani-Tutte theorem as argued by van der Holst in [44].

Our proof of the strong Hanani-Tutte theorem for the projective plane uses techniques developed for the Hanani-Tutte theorem and related results in the plane and higher-order surfaces by Pelsmajer, Schaefer and Štefankovič [37; 38] and combines them with ideas from

Mohar and Robertson on embeddings in the projective plane [33]; see Section 2.2. The proof will not naturally extend to any surface other than the projective plane, since it makes use of the list of minimal forbidden minors for the projective plane.

2.2 From Embeddings to Drawings

In this section we develop the necessary tools to deal with drawings in the projective plane. Some of these tools are extensions of well-known results for embeddings. All of them will play an important role in the proof of the strong Hanani-Tutte theorem for the projective plane.

2.2.1 Basic Observations and Redrawing Tools

Recall that a closed curve is *contractible* if it can be contracted to a point. In the projective plane a closed curve is contractible if and only if it passes through the crosscap an even number of times. [1]

Lemma 2.2.1. *Suppose that a graph drawn in the projective plane contains two vertices connected by three internally disjoint paths. Of the three cycles formed by pairs of paths, exactly one or exactly three are contractible.*

Proof. Let P_1, P_2, P_3 be the three paths. Call a path even (odd) if it passes through the crosscap an even (odd) number of times. Then a cycle is contractible if and only if it is formed by two paths of the same parity. By the pigeonhole principle, exactly two or three of the paths have the same parity. The lemma follows. □

---

[1] Any one-sided (or non-contractible) curve can serve as the crosscap. "Passing through the crosscap" means crossing that curve.

Lemma 2.2.1 is based on [34, Proposition 4.3.1].

For convenience, we say that a particular drawing of a graph is *iocr*-0 if no pair of non-adjacent edges crosses an odd number of times.

Lemma 2.2.2. *If a graph G drawn on the projective plane contains two vertex-disjoint non-contractible cycles, then the drawing is not iocr-*0*.*

Proof. In the projective plane any two non-contractible curves cross an odd number of times.[1] Therefore there must exist two edges in G, one in each of the two cycles, that cross oddly. These must be non-adjacent, as they belong to vertex-disjoint cycles, so the given drawing of G is not iocr-0. □

We will occasionally apply redrawing moves that lead to self-intersections of edges. These can be removed as shown in Figure 1.

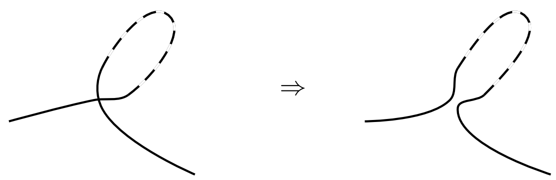

Figure 1. Removing a self-intersection; illustration from [37].

[1]Letting the first curve serve as the crosscap, we know that the other curve must pass through the crosscap an odd number of times since it is non-contractible.

A ΔY*-exchange* in G is a process that replaces a triangle in a drawing of G with a claw (a $K_{1,3}$). The three vertices of the triangle become the leaves of the claw.

Lemma 2.2.3. *Let G be a graph with $\mathrm{iocr}_{\mathrm{N}_1}(G) > 0$, and suppose G' can be obtained from G by a ΔY-exchange. Then $\mathrm{iocr}_{\mathrm{N}_1}(G') > 0$.*

Proof. Consider an iocr-0 drawing of G'. Let e_1, e_2 and e_3 be the three edges of the claw. Draw a new edge f_1 by closely following e_1 and e_2; similarly add f_2 following e_2, e_3 and f_3 following e_3, e_1. Remove any self-intersections as shown in Figure 1. If f_1 crosses an edge e of $G' - \{e_1, e_2, e_3\}$ oddly, then e must cross either e_1 or e_2 oddly; hence e is incident to f_1. Similarly f_2 and f_3 only cross adjacent edges oddly. Removing e_1, e_2, e_3 now yields an iocr-0 drawing of G, which implies that $\mathrm{iocr}_{\mathrm{N}_1}(G) = 0$. □

The following lemma shows, roughly speaking, that for an iocr-0 drawing it is not a single vertex that makes the difference between planarity and non-planarity.

Lemma 2.2.4. *Let $x \in V(G)$ and $H = G - x$. Suppose there is an iocr-0 drawing of G in the projective plane such that all cycles in H are contractible. Then G is planar.*

Proof. Consider the specified drawing of G in the projective plane.

Claim: We can redraw G so that each edge of H passes through the crosscap an even number of times, and the drawing is still iocr-0.

Let F be a spanning forest of H and select a root in every component of F. Process the edges of F in a breadth-first order as follows: suppose uv is an edge of F so that u is closer to the root of uv's component than v. Start contracting uv by moving v along uv towards u, pushing all crossings along with it; stop when uv passes through the crosscap an even number of times. Call *uv processed*. Note that this move does not change the parity of how often any processed edge of F other than uv passes through the crosscap, since the only edges whose parity is changed by the contraction are edges incident to v, and none of those can have been processed already. At the end, every edge of F passes through the crosscap an even number of times. Every edge in $E(H) - E(F)$ also uses the crosscap evenly, since it completes a contractible cycle with some edges in F.

We now remove the crosscap and replace it with a disk. We reconnect severed edges by simple curves within the newly added disk.

Any two such curves within the disk have to cross oddly, since their crossings with the disk boundary alternate. Since each edge of H passes through the disk an even number of times, its crossing parity with every other edge does not change. Hence, any two edges whose crossing parity changes must be incident to x, which means that the independent odd crossing number is not affected by replacing the crosscap with a disk. We have thus obtained an iocr-0 drawing of G in the sphere (and, thereby, the plane), which implies that the graph is planar by the Hanani-Tutte theorem for the plane. □

2.2.2 K-graphs and iocr

Let H be a subgraph of a graph G. An *H-component* or *H-bridge* is either an edge (and its endpoints) that does not belong to H but both of whose endpoints do, or a connected component of $G - V(H)$ together with all edges (and their endpoints) connecting this component to H.

H is a K_4-*graph* of G if it is a subdivision of K_4, and there is an H-component that is attached to all the vertices of degree 3 in H. H is a $K_{2,3}$-*graph* of G if it consists of three internally disjoint paths connecting two vertices, x and y, and there is an H-component that contains at least one internal vertex of each of the three paths. A K-*graph* is either a K_4-graph or a $K_{2,3}$-graph.

The following result is a well-known fact for embeddings (see [33, p326]). We relax the assumption that G is embedded to allow crossings, but control the crossings by requiring $\mathrm{iocr}_{\mathrm{N}_1}(G) = 0$.

Lemma 2.2.5. *Let G be a graph containing a K-graph. Then every iocr-0 drawing of G in the projective plane contains two non-contractible cycles in the K-graph.*

Proof. Fix an iocr-0 drawing of G in the projective plane; let H be the K-graph within G. Note that if a K-graph contains one non-contractible cycle, then it must contain two by Lemma 2.2.1. Hence, for a contradiction, we may assume that all the cycles in H are contractible.

Suppose first that H is a subdivision of a $K_{2,3}$. Then H is the union of three internally disjoint paths P_1, P_2, P_3 that have the same endpoints x and y, and there is an H-component B containing an internal vertex of each path. Consider three edges from these vertices of H to $V(B) - V(H)$, and let T be a minimal tree in B that contains those edges. Then T has three leaves, so it must contain a unique vertex z of degree 3. Let $G' = H \cup T$. By construction, G' is a subdivision of $K_{3,3}$. The drawing of G yields an iocr-0 drawing of G'. All the cycles in $G' - z$ are cycles in H, which are contractible by assumption. Applying Lemma 2.2.4 implies that G' is planar, which is a contradiction.

If H is a subdivision of a K_4, let S be the set of its 4 vertices of degree 3 and let B be an H-component that contains S. As above, there is a tree T in B such that the set of its leaves is S. Since T has four leaves it either has a unique vertex of degree 4 or two vertices of degree 3. Let $G' = H \cup T$. Since G' contains a K_5-minor it is not planar. If T contains a vertex z of degree 4 we proceed as in the case of $K_{2,3}$: the cycles in $G' - z$ are cycles in H and therefore contractible, but then Lemma 2.2.4 implies that G' is planar, which is a contradiction. If T contains two vertices u, v of degree 3, let Q be the path between u and v in T. Let uw be the first edge in the path (possibly $w = v$). Contract uw by moving u along uw to w and then identifying u and w. As we contract, we may create odd crossings and self-intersections. Only drawings of edges incident to u may contain self-intersections; remove them as shown in Figure 1. Any new odd crossing will be between an edge e_1 that was incident to u before the contraction, and an edge e_2 that had crossed uw oddly. But e_2 must have been incident to either u or w, since the drawing was iocr-0. After contraction both edges are incident to $u = w$, so the drawing is still iocr-0. In this fashion we can contract all the edges along the path Q until we have a single vertex of degree 4, with the drawing of H unchanged. Then we are back in the first case, which suffices. □

Apart from applying Lemma 2.2.5 directly, we will also use it in the following variant.

Corollary 2.2.6. *If a graph G contains two disjoint K-graphs, then* $\mathrm{iocr}_{\mathrm{N}_1}(G) > 0$.

Proof. Assume $\mathrm{iocr}_{\mathrm{N}_1}(G) = 0$. Applying Lemma 2.2.5 twice gives us two vertex-disjoint non-contractible cycles, which contradicts Lemma 2.2.2. □

2.3 Proof of the Main Theorem

If G cannot be embedded in the projective plane, then it must contain at least one of 35 minimal forbidden minors for the projective plane determined by Archdeacon, Glover, Huneke, and Wang [4; 22]. For a complete list of minimal forbidden minors and their names see [33][1] or [34, p198]. We show that all these graphs have independent odd crossing number larger than zero. This establishes Theorem 2.1.1 since from an iocr-0 drawing of a graph, one can easily obtain an iocr-0 drawing of any minor of the graph.[2]

The first twelve graphs are formed from two Kuratowski graphs by a disjoint union, a one-vertex identification, or a two-vertex identification and possibly deleting an edge between these vertices. It is easy to see that each contains two disjoint K-graphs. By Corollary 2.2.6 the independent odd crossing number of each of these twelve graphs is nonzero.

Of the remaining 23 minimal forbidden minors, C_7, E_{19}, D_{12}, E_{11}, E_{27}, D_9, G_1 [33, Fig. 3] and D_{17}, E_{20}, F_4 [33, Fig. 6] also contain two disjoint K-graphs, as observed in [33]. Again, by Corollary 2.2.6, the independent odd crossing number is nonzero for all of these graphs.

It is also known that each graph B_7, C_4, C_3, D_2 and E_2 [33, Fig. 6] can be obtained from graph A_2 through a sequence of ΔY-exchanges, and the graph E_5 can be obtained from the graph D_3 in the same way [33]. By Lemma 2.2.3 we need only show that $\mathrm{iocr}_{\mathrm{N}_1}(D_3) > 0$ and $\mathrm{iocr}_{\mathrm{N}_1}(A_2) > 0$ to prove a nonzero independent odd crossing number for all of these graphs.

---

[1] All references to [33] in this section are to the proof of Theorem 3.1 in that paper.

[2] We already used this in the proof of Lemma 2.2.5. This fact also underlies almost all proofs of the strong Hanani-Tutte theorem.

Thus we are left with seven graphs, E_{22}, A_2, D_3, F_1, B_1, E_{18}, and E_3. For each we will assume an iocr-0 drawing in the projective plane, then find a contradiction.

Consider E_{22}, letting x be its unique degree 4 vertex as seen in Figure 2. Every 4-cycle not containing x is disjoint from a $K_{2,3}$-graph, so it must be contractible, by Lemma 2.2.2. For any two 4-cycles that share exactly one edge, their symmetric difference forms a 6-cycle, and by Lemma 2.2.1 this 6-cycle is contractible. Any other cycle in $E_{22} - x$ is the symmetric difference of one of those 6-cycles C and a 4-cycle C', such that $C \cap C'$ is a path; it is also contractible by Lemma 2.2.1. Then Lemma 2.2.4 gives a planar drawing of E_{22}, a contradiction.

We deal with A_2 by a similar argument. Let x be the unique degree 6 vertex in A_2 (see Figure 2). Any triangle not containing x is disjoint from a K_4-graph and is therefore contractible (Lemma 2.2.2). Suppose that C is a minimal non-contractible cycle in $A_2 - x$. If C is not an induced cycle, then Lemma 2.2.1 yields a shorter non-contractible cycle, contradicting minimality. If C is an induced cycle, then C is a 4-cycle and there is a vertex $y \neq x$ adjacent to every vertex of C. Let u and v be a pair of non-adjacent vertices in C. The union of either u, v-path in C and the path uyv forms a 4-cycle that is not induced, so it is contractible. Then C is contractible as well by Lemma 2.2.1, a contradiction. Hence, every cycle in $A_2 - x$ is contractible. Then Lemma 2.2.4 gives a planar drawing of A_2, a contradiction.

For graphs F_1 and D_3 we borrow part of an argument from [33]. The cycles $v_1v_2v_3v_4$ and $v_1v_2v_3u_1$ in F_1 (see Figure 3) are each disjoint from a $K_{2,3}$-graph, so they must both be contractible by Lemma 2.2.2. But the vertices v_1, v_2, v_3, v_4, u_1 induce a $K_{2,3}$-graph in F_1, and at most one of its three cycles is contractible, which contradicts Lemma 2.2.5.

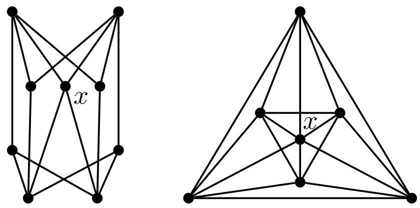

Figure 2. Vertex x in E_{22} (*left*) and in A_2 (*right*); illustration from [36].

For D_3 we apply the same argument to its cycles $v_1v_3v_2x$ and $v_1v_3v_2y$ (see Figure 3): Each is disjoint from a K_4-graph so both are contractible (Lemma 2.2.2). But there is a $K_{2,3}$-graph on vertices v_1, v_2, v_3, x, y, and at most one of its cycles is non-contractible, contradicting Lemma 2.2.5.

Next consider B_1. Any triangle containing exactly one of the vertices x, y, z is contractible since it is disjoint from a K_4-graph. Then by Lemma 2.2.1, the 4-cycles xu_1yu_2, xu_1zu_2, yu_1zu_2 are all contractible. But these are all the cycles in the $K_{2,3}$-graph on vertices $\{x, y, z\}, \{u_1, u_2\}$, which contradicts Lemma 2.2.5.

For the two remaining graphs we employ counting arguments. First consider E_{18}, which is $K_{4,4}$ with one edge removed (see Figure 4). Let $\{u_1, u_2, u_3, u_4\}$, $\{v_1, v_2, v_3, v_4\}$ be the two partite sets and let u_4v_4 be the missing edge. Each of u_4 and v_4 is contained in 9 induced 4-cycles. Furthermore, for each 4-cycle containing u_4 there is a disjoint 4-cycle containing v_4 [33]. Hence, at least one of the two cycles must be contractible by Corollary 2.2.6. So one of u_4, v_4, say u_4,

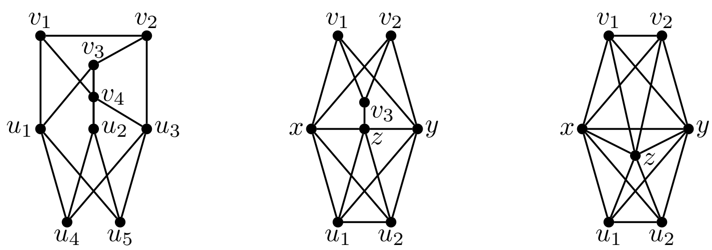

Figure 3. F_1, D_3 and B_1 (left to right) with labels; illustration from [36].

belongs to at most 4 non-contractible cycles. But $\{u_4, v_i, v_j, u_i, u_j\}$ induces a $K_{2,3}$-graph for each distinct pair i, j in $\{1, 2, 3\}$, and by Lemma 2.2.5 each one contains two non-contractible cycles. Since all these cycles are pairwise distinct, there are at least 6 non-contractible 4-cycles containing u_4, a contradiction.

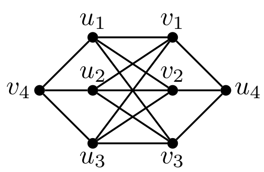

Figure 4. $E_{18} = K_{4,4} - u_4v_4$; illustration from [36].

Finally consider E_3, which is $K_{3,5}$. Let A, B be the two partite sets, with $|A| = 5$. By Lemma 2.2.5, for all $\{a_1, a_2, a_3\} \subset A$, $\{b_1, b_2\} \subset B$, exactly one of the three cycles in the $K_{2,3}$ induced by these 5 vertices is contractible. There are 30 possible choices of subsets of A and B as above, and every 4-cycle is present in 3 of the resulting subgraphs. Thus there are exactly 10 contractible 4-cycles in the given drawing of $K_{3,5}$. Some pair of vertices of B is shared by at least 4 of them, and two of these 4-cycles must also share a vertex of A. The union of these two 4-cycles is a $K_{2,3}$-graph, but by Lemma 2.2.5, the $K_{2,3}$-graph contains only one contractible cycle, a contradiction. $\square$

This concludes this chapter, as well as the part of the thesis concerned with graph drawing. The rest of the thesis concerns combinatorial questions in expressions in knowledge representation called Horn formulas.

CHAPTER 3

RANDOM HORN FORMULAS AND PROPAGATION CONNECTIVITY FOR DIRECTED HYPERGRAPHS

This chapter is based on joint work with Robert H. Sloan and György Turán at the University of Illinois at Chicago [41].

3.1 Introduction

Horn formulas are a subclass of CNF expressions, where every clause contains at most one unnegated variable. This class is tractable in the sense that many problems that are hard for CNF expressions in general are polynomially solvable for Horn formulas (such as satisfiability and equivalence). It is partly for this reason that Horn formulas are of basic importance in artificial intelligence and other areas. Random Horn formulas have been studied in [21; 19; 27; 31; 35].

A Horn formula is *definite* if it consists of clauses containing exactly one unnegated variable. We consider definite Horn formulas with clauses of size 3, i.e., with clauses of the form $(\bar{a} \vee \bar{b} \vee c)$, which can also be written as $a, b \rightarrow c$. Here a and b form the *body* of the clause and c is the *head* of the clause. Implication between a definite Horn formula φ and a definite Horn clause C can be decided by *forward chaining*: mark variables in the body of C and, while there is a clause in φ with all its body variables marked, mark its head variable as well. Then C is implied by φ iff its head gets marked.

We consider random definite Horn formulas with clauses of size 3 over n variables, where every clause is included with probability p. It follows directly from the results of [31] that $p = (2 \ln n)/n$ is a threshold probability for the following property: *every pair of variables implies every other variable* (see also [21] for a related result).

In this chapter we consider the property that *some pair of variables implies every other variable.* This property is closely related to the property of propagation connectivity for 3-uniform undirected hypergraphs, introduced recently by Berke and Onsjö [8]. They consider a marking process like forward chaining, except that now a vertex can be marked if it is contained in an edge whose other two vertices are already marked. A 3-uniform undirected hypergraph is *propagation connected* if there is a pair of vertices such that the marking process, starting from that pair, marks every vertex. Berke and Onsjö showed that for $p < 1/(n(\log n)^2)$ random hypergraphs are *a.a.s.* not propagation connected [8] and for $p > 1/(n(\log n)^{0.4})$ random hypergraphs are *a.a.s.* propagation connected [9]. The first result proves a *lower bound* for the transition from random hypergraphs being *a.a.s.* not propagation connected to random hypergraphs being *a.a.s.* propagation connected, and the second result proves an *upper bound* for the transition. We use the terms lower and upper bound in a similar sense throughout the chapter.

The Horn formula property mentioned above is equivalent to propagation connectivity for *directed* 3-uniform hypergraphs, where by a directed hypergraph we mean a hypergraph with each edge having a distinguished vertex called its head, and the other vertices called its body. (This is one of the possible definitions of a directed hypergraph. There are several other variants.) In the rest of the chapter we use the terminology of propagation connectivity for directed hypergraphs instead of Horn formulas. We show that random directed 3-uniform hypergraphs for $p \leq 1/(11n \ln n)$ are *a.a.s.* not propagation connected and for $p \geq (5 \ln \ln n)/(n \ln n)$ are *a.a.s.* propagation connected. The proofs are based on two versions of the "fanning-out process" (see, e.g., [29; 28]). For the upper bound we start the process by exploring a subset of the vertices and finding a maximal degree pair within that subset.

For the undirected hypergraph version of the problem, Coja-Oghlan, Onsjö and Watanabe concurrently and independently proved lower and upper bounds, both of the order $1/(n \ln n)$ [18]. It appears that their argument can be adapted to yield order $1/(n \ln n)$ lower and upper bounds in the directed case as well. The proofs we present here are simpler.

The lower and upper bounds are presented in Section 3 and 4. In the closing section we mention a few open problems.

3.2 Preliminaries

We consider 3-uniform directed hypergraphs H with directed edges of the form $u, v \to w$. The pair (u, v) is the *body* of the edge and w is the *head* of the edge. Note that the body is an unordered pair. The *degree* of a pair (u, v) is the number of vertices w that form an edge $u, v \to w$ with the pair. We refer to vertex w as a *successor* of (u, v). The (u, v)-propagation connected component (or simply (u, v)-component) of H is the set of vertices marked by the marking process starting with (u, v).

The probability model where a random directed hypergraph is formed over the vertex set $[n] = \{1, \ldots, n\}$ by including each edge $u, v \to w$ independently with probability p is denoted by $\mathcal{DH}(n,p)$. A *property* Q of directed hypergraphs of order n is a subset of $\mathcal{DH}(n,p)$ s.t. if $H \in Q$ and G is isomorphic to H then $G \in Q$. Q is *monotone increasing* if adding an edge to $H \in Q$ preserves the property. For any monotone increasing property of directed hypergraphs the probability that the property holds for a random directed hypergraph drawn from $\mathcal{DH}(n,p)$ is a monotone non-decreasing function of p (see [12, Th.2.1]).

Proposition 3.2.1. *[12, pp36] Suppose that Q is a monotone increasing property, let $0 \le p_1 < p_2 \le 1$ and $\mathcal{H}_i \in H(n,p_i)$ for $i = 1,2$. Then*

$$\Pr[\mathcal{H}_1 \in Q] \le \Pr[\mathcal{H}_2 \in Q].$$

We use the following versions of the Chernoff bounds [28].

Proposition 3.2.2. *If $X \in \mathrm{BIN}(n,p)$ and $t \ge 0$ then*

(a) $\Pr[X \ge \mathbb{E}[X] + t] \le \exp\left\{-\frac{t^2}{2\left(\mathbb{E}[X]+\frac{t}{3}\right)}\right\}$,

(b) $\Pr[X \le \mathbb{E}[X] - t] \le \exp\left\{-\frac{t^2}{2\mathbb{E}[X]}\right\}$.

3.3 A lower bound

We first give a lower bound for probabilities p such that a random directed hypergraph from $\mathcal{DH}(n,p)$ is *a.a.s.* propagation connected.

Theorem 3.3.1. *Let $p \le 1/(11n\ln n)$. In a random directed hypergraph drawn from $\mathcal{DH}(n,p)$* a.a.s. *every propagation connected component has size at most $11\ln n$.*

Proof. By monotonicity we may assume $p = 1/(11n\ln n)$.The following process is used to explore $H \in \mathcal{DH}(n,p)$. Start with two sets $A_0 = \{u,v\}$ and $B_0 = \emptyset$. The sets A_i and B_i

represent the sets of *discovered vertices* and *saturated pairs* at iteration i respectively, and put $m_i = |A_i|$. At iteration i of the process consider vertices u_i and v_i such that $u_i, v_i \in A_{i-1}$ and $(u_i, v_i) \notin B_{i-1}$. Find every edge $u_i, v_i \to w$ where $w \notin A_{i-1}$. Construct the set A_i so that it contains all vertices in set A_{i-1} plus all vertices w, where w is the head of an edge that was found in step i. Construct the set B_i by $B_i = B_{i-1} \cup \{(u_i, v_i)\}$. When every pair in A_i is saturated, we have discovered all the vertices in the component, and from then on we put $A_j = A_i$, $B_j = B_i$ for every $j > i$.

We need to show that this process stabilizes after a small number of steps with high probability. Define X_i to be the number of successors, in $V \setminus A_{i-1}$, of the pair (u_i, v_i) to be saturated. Each edge with body (u_i, v_i) and head in $V \setminus A_{i-1}$ is in the hypergraph with probability p, independently of the presence or absence of any other edge. Furthermore each such edge is considered at most once in the process. Thus $X_i \in \mathrm{BIN}(n - m_{i-1}, p)$.

Let $k = \lceil 11 \ln n \rceil$. If the process generates at least k vertices then this must happen in the first $\binom{k-1}{2}$ iterations. Thus the probability of generating at least k vertices is at most

$$\Pr\left[\sum_{i=1}^{\binom{k-1}{2}} X_i \geq k - 2\right]. \tag{3.1}$$

Let $X_i^+ \in \mathrm{BIN}(n,p)$ and replace the upper limit in the summation (3.1) by $\binom{k}{2}$ for convenience. Then, noting that $\sum_{i=1}^{\binom{k}{2}} X_i^+ \in \mathrm{BIN}\left(\binom{k}{2}n, p\right)$ and as such has mean $\binom{k}{2}np$, the probability (3.1) can be upper bounded by

$$\Pr\left[\sum_{i=1}^{\binom{k}{2}} X_i^+ \geq k-2\right] = \Pr\left[\sum_{i=1}^{\binom{k}{2}} X_i^+ \geq \binom{k}{2}np + k - 2 - \binom{k}{2}np\right].$$

Using the values of p and k we note that $np \sim 1/k$, giving $\binom{k}{2}np \sim k-2-\binom{k}{2}np \sim k/2$. Then the Chernoff bound (Proposition 3.2.2(a)) with $t = k-2-\binom{k}{2}np$ gives the upper bound

$$\exp\{-(3k)/16\} \sim \exp\{-(33\ln n)/16\} = o(n^{-2}),$$

which implies the theorem by the union bound. □

3.4 An upper bound

In this section we give a sufficient condition for probabilities p such that a random directed hypergraph from $\mathcal{DH}(n,p)$ is *a.a.s.* propagation connected.

Theorem 3.4.1. *For $p \geq (5\ln\ln n)/(n\ln n)$ a random directed hypergraph from $\mathcal{DH}(n,p)$ is* a.a.s. *propagation connected.*

Proof. The property of being propagation connected is monotone non-decreasing. By monotonicity we may assume $p = (5\ln\ln n)/(n\ln n)$. We use a modification of the process described in the previous section. First we consider *all* edges over the first $n/4$ vertices and find a highest-degree pair (u,v) in that subset. Starting from the successors of the pair we find a sufficiently large part of the rest of the component using a variant of the original process organized into phases as follows.

Let $m = \lceil (\ln n)/(\ln\ln n)\rceil$ and assume that we found a pair (u, v) with m successors $w_1, \ldots, w_m$ among the first $n/4$ vertices. Let $A_0 = \{w_1, \ldots, w_m\}$ be the initial set of discovered vertices and let C_0 be the $(3/4)n$ vertices not considered so far, forming the initial set of *available vertices*. In iteration i of the new process we pick an arbitrary set $D_{i-1} \subseteq C_{i-1}$ of $n/2$ available vertices, and we find all edges $u, v \to w$, where $u, v \in A_{i-1}$ and $w \in D_{i-1}$. If there are at least m distinct successors in D_{i-1} then let A_i be any m of these and put $C_i = C_{i-1} \setminus A_i$. Otherwise let $A_j = A_{i-1}$ for every $j \geq i$. We run this process for $\lceil 3\ln n \ln\ln n\rceil$ iterations.

The following lemma, analogous to bounds for graphs (see [12, Ch.3]), gives a bound for the maximal degree of a pair in $H \in \mathcal{DH}(n, p)$. Using monotonicity, this lemma is stated for the smaller and simpler probability $1/(n \ln n)$.

Lemma 3.4.2. *If* $p = 1/(n \ln n)$*, then the maximum degree of* $H \in \mathcal{DH}(n, p)$ *is* a.a.s. *at least* $(\ln 4n)/(\ln\ln 4n)$.

Proof. Let $d = \lceil (\ln 4n)/(\ln\ln 4n)\rceil$ and let the random variable Y_{ij} be the number of successors of pair (i, j) in H. Then $Y_{ij} \in \mathrm{BIN}\,(n-2, p)$ and since we are dealing with directed edges, the variables Y_{ij} are independent. Thus the probability that every degree is smaller than d is

$$(1 - \Pr[Y_{ij} \geq d])^{\binom{n}{2}} \leq \left(1 - \binom{n-2}{d} p^d (1-p)^{n-2-d}\right)^{\binom{n}{2}} < \left(1 - \frac{1}{2}\binom{n-2}{d} p^d\right)^{\binom{n}{2}},$$

if n is sufficiently large. For the last inequality we used the fact that $(1-p)^{n-2-d} = 1 - o(1)$. Using $1 - x < e^{-x}$, we need to show that

$$\left(\frac{pn}{d}\right)^d n^2 \to \infty.$$

This follows by taking logarithms and using the definitions of p and d. Specifically we use $(\ln 4n)/(\ln\ln 4n) \leq d \leq (2\ln 4n)/(\ln\ln 4n)$ to get

$$\begin{aligned}
d\ln\left(\frac{pn}{d}\right) + 2\ln n &> \left(\frac{\ln 4n}{\ln\ln 4n}\right)\ln\left(\frac{\ln\ln 4n}{2(\ln n)(\ln 4n)}\right) + 2\ln n && (3.2)\\
&= \left(\frac{\ln 4n}{\ln\ln 4n}\right)(\ln\ln\ln 4n - \ln 2 - \ln\ln n - \ln\ln 4n) + 2\ln n && (3.3)\\
&= \left(\frac{\ln 4n}{\ln\ln 4n}\right)\ln\ln\ln 4n - (\ln 4n)\left(1 + \frac{\ln\ln n + \ln 2}{\ln\ln 4n}\right) + 2\ln n && (3.4)\\
&= (\ln 4n)\frac{(\ln\ln\ln 4n - \ln 2)}{\ln\ln 4n} + (\ln n)\left(1 - \frac{\ln\ln n}{\ln\ln 4n}\right) - (\ln 4)\left(1 + \frac{\ln\ln n}{\ln\ln 4n}\right). && (3.5)
\end{aligned}$$

We may now conclude that the expression on the right tends to infinity, since the first term tends to infinity, the second term is positive for any value of n and the third term has a constant limit.

□

We also use a version of a lemma of [9] showing that *a.a.s.* every component is either small or contains every vertex. This lemma is similar to the gap theorem in [29] and its proof is included for completeness.

Lemma 3.4.3 ([9]). *If* $p = (5 \ln \ln n)/(n \ln n)$ *then* a.a.s. *every propagation connected component has either size n or size less than* $(\ln n)^2$.

Proof. If a set of vertices is a propagation connected component then there can be no edges with body in the component and head outside. Thus the probability that there is a component of size k is at most

$$\binom{n}{k} \cdot (1-p)^{\binom{k}{2}(n-k)}.$$

We show that for $(\ln n)^2 \leq k \leq n-1$ this quantity is $o(1/n)$.

If $n/2 \leq k \leq n-1$ then replacing $\binom{n}{k}$ with $\binom{n}{n-k}$ gives the upper bound

$$\left(\frac{ne}{n-k}\right)^{n-k} \exp\left\{-p\binom{k}{2}(n-k)\right\} = \exp\left\{-(n-k)\left(p\binom{k}{2} - \ln\left(\frac{ne}{n-k}\right)\right)\right\}.$$

As $p\binom{k}{2} = \Omega\left(n \ln \ln n / \ln n\right)$, the probability is upper bounded by $\exp\{-\Omega\left(n \ln \ln n / \ln n\right)\}$.

Else $(\ln n)^2 \leq k < n/2$ and the analogous calculation gives the upper bound

$$\exp\left\{-k\left(\ln k + \frac{p(k-1)(n-k)}{2} - (\ln n + 1)\right)\right\}.$$

Here $n-k$ can be replaced by $n/2$ and then substituting the values of p and k we can lower bound $\ln k + p(k-1)(n-k)/2 - (\ln n + 1)$ by

$$\ln n\left(\frac{5 \ln \ln n}{4} - 1\right) + \ln \ln n\left(2 - \frac{5}{4 \ln n}\right) - 1 = \Omega\left(\ln n \ln \ln n\right).$$

Since $k \geq (\ln n)^2$ we get an upper bound of the form $\exp\{-\Omega\left((\ln n)^3\right)\}$. $\square$

Returning to the proof of Theorem 3.4.1, let us say that we are successful if we find a pair of degree m among the first $n/4$ vertices, we can run the iterative process for $\lceil \ln n \ln\ln n \rceil$ iterations, always finding m new vertices, and the event described in Lemma 3.4.3 occurs. In this case, after the last iteration we found a component of size $(\ln n)^2$, and by Lemma 3.4.3 the hypergraph is propagation connected.

The number Z_i of edges added in the ith iteration has distribution $\mathrm{BIN}\left(\binom{m}{2}\frac{n}{2}, p\right)$. Using the Chernoff bound (Proposition 3.2.2(b)) for the probability that there are fewer than m such edges we get

$$\Pr\left[Z_i < m\right] \leq \exp\left\{-\frac{(\mathbb{E}[Z_i]-m)^2}{2\mathbb{E}[Z_i]}\right\} < \exp\left\{-\frac{\ln n}{41 \ln\ln n}\right\}. \tag{3.6}$$

The last bound is due to $\mathbb{E}[Z_i] \sim m^2np/4 \sim (5\ln n)/(4\ln\ln n)$ and $\mathbb{E}[Z_i] - m \sim (\ln n)/(4\ln\ln n)$ which gives

$$\frac{(\mathbb{E}[Z_i]-m)^2}{2\mathbb{E}[Z_i]} \sim \frac{\ln n}{40 \ln\ln n}.$$

Since we saturate more than one pair in an iteration it is possible that the same vertex is discovered by more than one edge. The probability of such a conflict is at most

$$\binom{\binom{m}{2}}{2} \cdot \frac{n}{2} \cdot p^2 = O\left(\frac{(\ln n)^2}{n(\ln\ln n)^2}\right). \tag{3.7}$$

If $Z_i \geq m$ and there are no conflicts in iteration i then we found at least m new vertices in that iteration. Hence, using Lemmas 3.4.2, 3.4.3 and the bounds (3.6) and (3.7), the probability of failure is

$$o(1) + O\left(\frac{(\ln n)^3}{n \ln \ln n} + \exp\left\{-\frac{\ln n}{41 \ln \ln n}\right\} \ln n \ln \ln n\right) = o(1).$$

In the previous step we use the fact that $((\ln n)^3)/(n \ln \ln n) \to 0$ and that the logarithm of the term $\exp\{-(\ln n)/(40 \ln \ln n)\} \ln n \ln \ln n$ is

$$-\frac{\ln n}{40 \ln \ln n} + \ln \ln n + \ln \ln \ln n \to -\infty.$$

□

The approach used in this chapter seems to require edge probabilities of the order of magnitude given in Theorem 3.4.1, as the probability should be large enough to produce a sufficient number of newly discovered vertices, starting from an initial set of size determined by the maximum degree calculation of Lemma 3.4.2.

3.5 Open Problems

It follows from the result of [31] mentioned in the introduction and Theorem 3.3.1 that the fraction of pairs (u, v) such that the (u, v)-component has size n grows from 0 to 1 over the interval between $p = \Omega(1/(n \ln n))$ and $O(\ln n/n)$. It would be interesting to have more detailed information about growth over this interval.

Forward chaining can be studied for different ranges of the parameters. For example, [35] gives phase transition results on forward chaining where the marking process is started from a positive fraction of the vertices and p is of the form c/n^2. Forward chaining could be studied beyond just determining the size of the component produced. The process producing a propagation connected component can also be viewed, using the terminology of propositional logic, as a resolution derivation of consequences, or using the terminology of directed graphs, as a hyperpath [7]. Considering combinatorial parameters of the structure of a propagation connected component, such as its depth, could also be of interest from the point of view of knowledge base applications.

The Internet and other large networks motivated the study of random graphs in models different from the standard models of fixed edge probabilities or fixed number of edges. For evolving knowledge bases, modeled by random Horn formulas, we are not aware of any such work. A possible choice would be to consider random subformulas of a given formula (corresponding to 'true' knowledge). For random graphs such a model has been studied, e.g., in [17].

CHAPTER 4

HORN MINIMIZATION: HYDRA FORMULAS AND (SOME) STEINER MINIMIZATION

This chapter is based on joint work with Robert H. Sloan and György Turán at the University of Illinois at Chicago [42; 40].

4.1 Introduction

We consider a problem concerning the minimal number of hyperarcs in directed hypergraphs with prescribed reachability properties. In this chapter, a directed hypergraph $H = (V, F)$ has size-3 hyperarcs of the form $u, v \to w$ where u, v is called the *body* (or tail) and w is called the *head* of the hyperarc. Reachability is defined by a marking procedure known as *forward chaining*. A vertex $w \in V$ is *reachable* from a set $S \subset V$ if the following process marks w: start by marking vertices in S, and as long as there is a hyperarc $a, b \to c$ such that both a and b are marked, mark c as well.

Given an undirected graph $G = (V, E)$, we would like to find the minimal number of hyperarcs in a directed hypergraph $H = (V, F)$, such that for every pair (u, v), the set of vertices reachable from $\{u, v\}$ in H is the whole vertex set V if $(u, v) \in E$, and it is $\{u, v\}$ otherwise. In other words, given a set of bodies, we look for the minimal total number of heads assigned to these bodies such that every body can reach every vertex. The minimum is called the *hydra number* of G, denoted by $h(G)$.

The problem is a combinatorial reformulation of a special case of the *minimization problem for propositional Horn formulas*. A *hydra formula* φ is a definite Horn formula with clauses of size 3 such that every body occurring in the formula occurs with all possible heads. The minimal number of clauses needed to represent φ is the hydra number of the undirected graph G corresponding to the bodies in φ.

Horn formulas are a basic knowledge representation formalism. Horn minimization is the problem of finding a shortest possible Horn formula equivalent to a given formula. Special cases correspond to the well studied transitive reduction and minimum equivalent digraph problems for directed graphs [6]. Estimating the size of a minimal formula is not well understood even in rather simple cases. Horn minimization was shown to be NP-Complete by Ausiello, D'Atri and Saccà [6]. On the other hand, Hammer and Kogan showed that reducing a Horn formula over n variables, to a prime and irredundant form is an $(n-1)$-approximation algorithm for general Horn minimization [25]. A $2^{\log^{1-\epsilon}(n)}$-inapproximability result for Horn minimization of definite Horn formulas assuming NP$\not\subseteq$DTIME$(n^{\text{polylog}(n)})$ was given by Bhattacharya, DasGupta, Mubayi and Turán [11]. That result we strengthened recently by Boros and Gruber [14].

Besides being a natural subproblem of Horn minimization, the hydra minimization problem may also be of interest for the following reason. The Horn *body minimization* problem is the problem of finding, given a definite Horn formula, an equivalent Horn formula with the minimal number of distinct bodies. There are efficient algorithms for this problem [3; 5; 23; 32]. Thus one possible approach to Horn minimization is to find an equivalent formula with the minimal number of bodies and then to select as few heads as possible from the set of heads assigned to the bodies. This approach is indeed used in an approximate Horn minimization algorithm [11], and is still used in Section 4.3 of this chapter. Hydras are a natural test case for the body minimization approach.

Specifically, [11] proposed the problem of Steiner minimization of definite Horn formulas: given a Horn formula φ over a variable set X, find a *Steiner X-equivalent* Horn formula ψ with the smallest number of clauses. Horn formula ψ is allowed to include variables not in X (in a restricted manner), and so it is not usually equivalent to φ. However φ and ψ are equivalent in the following sense: for any clause C over X, ψ implies C *if and only if* φ implies C. Bhattacharya et al showed that Steiner minimization is MAX-SNP-hard but also gave an efficient $O(n \log\log n/(\log n)^{1/4})$ approximation algorithm, where n is the number of variables in the original definite Horn formula [11]. Coincidentally, we improve the result slightly with a simple update of the algorithm and its analysis (Theorem 4.3.4) to gain a factor of $(\log\log n)/(\log n)^{3/4}$, and show that this is sharp (Theorem 4.3.6).

The chapter is organized as follows. Section 4.2 contains some background on Horn formulas. After a small detour into Steiner minimization (Section 4.3), and a discussion of the motivating Horn minimization problem for hydra formulas (Section 4.4), we use the rest of the chapter to present various results on hydra numbers.

It is easy to see that $|E(G)| \leq h(G) \leq 2|E(G)|$ for every graph G on at least three vertices. Graphs satisfying the lower bound are called *single-headed*. In Section 4.5 we give some sufficient and some necessary conditions for single-headedness. The hydra number of a graph is related to the path cover number of the line graph of its spanning subgraphs (Theorem 4.6.1, Theorem 4.6.4). Corollary 4.6.2 gives the upper bound $|E(G)| + \lfloor |V(G)|/2 \rfloor$ for graphs with no isolated vertices based on Theorem 4.6.1(ii).

Section 4.7 considers hydra number of trees. It is shown in Theorem 4.7.1 that single-headed trees are precisely the stars and that trees with hydra number $|E(G)| + 1$ are precisely the caterpillars. We give a lower bound for the hydra number of a tree (Theorem 4.7.4) and we show that this is sharp for spider trees (Corollary 4.7.6). We conclude the discussion on spiders by considering the in-degree of vertices in optimal hydras representing a spider family (Theorem 4.7.7). We then consider binary trees: we show that the hydra number of a complete binary tree is between $\frac{9}{8}|E(G)|$ and $\frac{17}{15}|E(G)|+1$ (Theorem 4.7.9), and note that a connectivity restriction on the spanning subgraph in Theorem 4.6.1 can significantly worsen the upper bound on the hydra number (Theorem 4.7.10).

In Section 4.8 we give a weaker upper bound for the hydra number, but use it to determine the maximal hydra number of a tree. Finally in Section 4.9 we consider the related problem of finding the minimal number of hyperarcs for which every k-tuple of vertices is good, and we give almost matching lower and upper bounds. We conclude this chapter by mentioning several open problems.

4.2 Background

A *definite Horn clause* is a disjunction of literals where exactly one literal is unnegated. Such a disjunction can also be viewed as an implication, for example the clause $\bar{x} \vee \bar{y} \vee z$ is equivalent to the implication $x, y \to z$. The tuple x, y is the *body* and the variable z is the *head* of the clause. The size of a clause is the number of its literals. A definite d-Horn formula is a conjunction of definite Horn clauses of size d. A clause C is an implicate of a formula φ if every truth assignment satisfying φ satisfies C as well, denoted as $\varphi \models C$. The implicate C is *prime* if none of its proper subclauses is an implicate.

Implication between a definite Horn formula φ and a definite Horn clause C can be decided by *forward chaining*: mark every variable in the body of C, and while there is a clause in φ with all its body variables marked, mark the head of that clause as well. Then φ implies C iff the head of C gets marked.

4.3 Steiner minimization

In this section we make a small detour into Steiner minimization, and improve the approximation algorithm from [11], which uses body minimization. We begin by defining a *Steiner extension over X* and *Steiner X-equivalence.*

4.3.1 Steiner Basics

Definition 4.3.1 (Steiner extension over X [11])**.** *Let φ be a Horn formula over the variable set V, and $X \subset V$. We say that φ is a* Steiner extension over X *if every variable in φ not in X appears as the head of a clause whose body is composed entirely of variables in X, or as the only variable in the body of a clause whose head is a variable in X.*

Definition 4.3.2 (Steiner X-equivalence [11])**.** *Horn formulas φ and ψ are* Steiner X-equivalent *if:*

i. For every clause C over X, $\varphi \models C$ iff *$\psi \models C$, and*

ii. Both φ and ψ are Steiner extensions over X.

The Steiner equivalence of definite Horn formulas can be decided efficiently, unlike more general notions of extended equivalence of definite Horn formulas [11].

In constructing a Horn formula that is Steiner X-equivalent to a given definite Horn φ in Theorem 4.3.4, we will require the following decomposition theorem from [11]. It guarantees that a bipartite graph can be efficiently decomposed into complete bipartite graphs with not too many copies of each vertex.

Theorem 4.3.3 ([11]). *Let G be a bipartite graph, with parts A and B, and let $a = |A|$, $b = |B|$ and $a \geq b$. Then a partition of G into t different complete bipartite graphs $\{G_i\}_{1 \leq i \leq t}$ with $V(G_i) \cap A = A_i$ and $V(G_i) \cap B = B_i$ such that*

$$\sum_{1 \leq i \leq t} |A_i| + |B_i| = O\left(\tfrac{ab}{\ln a} + a\ln(b+1)\right)$$

can be found in polynomial time.

4.3.2 Algorithm

We construct a polynomial $O(n/\log n)$ approximation algorithm for Steiner minimization of definite Horn formulas. Our algorithm is a slight variation of the $O(n \log\log n/(\log n)^{1/4})$ approximation algorithm given in [11] with an improved analysis.

Theorem 4.3.4. *There is an efficient $O(n/\log n)$ approximation algorithm for Steiner minimization of definite Horn formulas, where n is the number of variables in the original formula.*

Proof. We use the following algorithm.

Input: A definite Horn formula φ

Algorithm STEINER-APPROX:

Construct ψ, a prime and irredundant Horn formula equivalent to φ.

if ψ has at most $n/\log n$ clauses **then** return ψ

else return ψ'=STEINER(MIN-BODY(ψ)).

By Hammer and Kogan [24] we can produce a prime and irredundant Horn CNF ψ equivalent to φ in quadratic time. If ψ is already small the result follows, so we may assume that ψ contains more than $n/\log n$ clauses.

Apply to ψ the procedure MIN-BODY, which can be any efficient body minimization algorithm, to output a Horn formula χ = MIN-BODY(ψ) with the minimum number of distinct bodies. Finally we apply the STEINER(χ) procedure [11] to produce our approximation to the smallest Horn formula Steiner X-equivalent to φ.

Specifically, during the STEINER procedure we construct a bipartite graph G_χ, such that:

i. Bodies(χ) and Heads(χ), the set of bodies and heads respectively, form the two parts of vertex partition, and

ii. Edge $(b, h) \in E(G_\chi)$ for every $b \in$ Bodies(χ) and $h \in$ Heads(χ) with $(b \to h) \in \chi$.

This bipartite graph is used to produce a Horn formula ψ' that is Steiner X-equivalent to ψ in the following way:

(1) Decompose the edge set of G_χ into t complete bipartite graphs $G_i = \{B_i, H_i\}$ for $1 \leq i \leq t$ such that $B_i \subset$ Bodies(χ) and $H_i \subset$ Heads(χ).

(2) For each G_i introduce the new variable $y_i \notin X$ and add the clauses $b \to y_i$ and $y_i \to h$ for $b \in B_i$ and $h \in H_i$ to the new formula ψ'.

We guarantee that the size of ψ' will not be too large and that the decomposition of G_χ can be performed efficiently using Theorem 4.3.3.

Let $ALG(\varphi)$ be the number of clauses in the Horn formula produced by STEINER-APPROX with input φ, and $OPT_S(\varphi)$ be the minimal number of clauses in any Horn formula that is Steiner X-equivalent to φ. Also let k and ℓ be respectively the number of distinct bodies and heads in MIN-BODY(ψ),

It is not hard to see that the minimal number of clauses in a Horn formula that is Steiner X-equivalent to φ is larger or equal to the number of distinct bodies in the MIN-BODY(φ), and by the equivalence of φ and ψ:

$$OPT_S(\varphi) \geq k.$$

Let us first deal with the case $k \geq \ell$. Then, as in [11]:

$$\frac{ALG(\varphi)}{OPT_S(\varphi)} \leq c \cdot \frac{\frac{k\ell}{\ln k} + k \ln \ell}{k} = c \cdot \left(\frac{\ell}{\ln k} + \ln \ell \right) = O\left(\frac{\ell}{\ln \ell} \right) = O\left(\frac{n}{\ln n} \right),$$

where the first inequality is due to $OPT_S(\varphi) \geq k$ and Theorem 4.3.3, the second step is due to the inequality $k \geq \ell$ and the last step is due to the fact that $x/\ln x$ is a monotone increasing function, and that there can only be as many heads as there are variables.

We now move to the case $k < \ell$. We will use the following lemma.

Lemma 4.3.5 ([13]). *Let φ be an arbitrary irredundant definite Horn CNF formula over n variables, and let φ' be a Horn CNF formula equivalent to φ with the minimum number of clauses. Then $|\varphi'| \geq \sqrt{|\varphi|}$.*

Proof. We follow the $(n-1)$-approximation algorithm for Horn minimization of Hammer and Kogan [25]. Let m be the number of clauses in φ and m' be the number of clauses in φ'. Define h to be the number of heads in φ'. Every definite clause C in φ' can be derived by resolving at most $h-1$ clauses of φ', as those are the maximum number of steps the forward chaining procedure would require to mark the head of C. Since φ is irredundant there can be at most $(h-1)\cdot m'$ number of clauses in φ, i.e. $m \leq (h-1)\cdot m'$. But $h \leq m'$, so $m \leq (m')^2$. □

Recall that $|\psi| > n/\log n$. Then, by Lemma 4.3.5 and the equivalence of φ and ψ, $OPT(\varphi) = OPT(\psi) \geq \sqrt{n/\log n}$. On the other hand [11] showed that $OPT_S(\varphi) \leq \frac{1}{4}(OPT(\varphi))^2$. Consequently

$$OPT_S(\varphi) \geq 2 \cdot \sqrt{OPT(\varphi)} \geq 2 \cdot \sqrt[4]{n/\log n}.$$

Now to finish the proof we note that:

$$\begin{aligned}
\frac{ALG(\varphi)}{OPT_S(\varphi)} &\leq c \cdot \left(\frac{\frac{k\ell}{\ln \ell} + \ell \ln k}{OPT_S(\varphi)} \right) \leq c \cdot \left(\frac{\ell}{\ln \ell} + \frac{\ell \ln OPT_S(\varphi)}{OPT_S(\varphi)} \right) \\
&\leq c' \cdot \left(\frac{n}{\ln n} + n \cdot \frac{\ln\left(2 \cdot \sqrt[4]{n/\log n}\right)}{2 \cdot \sqrt[4]{n/\log n}} \right) \\
&= O\left(\frac{n}{\ln n} + n^{3/4} \cdot (\log n)^{5/4} \right) = O\left(\frac{n}{\ln n} \right),
\end{aligned}$$

where the first inequality is due to Theorem 4.3.3, the second step is due to the inequality $OPT_S(\varphi) \geq k$ and the last step is due to the inequality $l \leq n$, the fact that $x/\ln x$ is a monotone increasing function, and the inequality $OPT_S(\varphi) \geq 2 \cdot \sqrt[4]{n/\log n}$. □

Next we show that the bound given in Theorem 4.3.4 for the STEINER-APPROX algorithm is sharp.

Theorem 4.3.6. *For every integer n, there exists a definite Horn formula φ_n such that*

$$\frac{ALG(\varphi_n)}{OPT_S(\varphi_n)} = \Omega\left(\frac{n}{\log n}\right).$$

Proof. We construct a definite Horn formula φ on variables $x_1, x_2, \ldots, x_n$ by taking all vertices x_i with $1 \leq i \leq n/2$ as single-variable bodies and all vertices x_j for $n/2 < j \leq n$ as heads. (For simplicity we may assume that n is even.) Then G_φ is a complete bipartite graph with equal parts, and the minimum formula that is Steiner X-equivalent to φ can be produced by adding a single new variable $y \notin \{x_i : 1 \leq n\}$, and the two clauses $(b \to y) \wedge (y \to h)$ for each clause $b \to h$ in φ. Then $OPT_S = n$. But the procedure in the proof of Theorem 4.3.3 instead decomposes $G(\varphi)$ into $(n/(2q))^2 = \Theta\left(n^2/\log^2 n\right)$ complete bipartite subgraphs $K_{q,q}$, with $q = \lfloor \log(n/2)/\ln(2e) \rfloor = \Theta(\log n)$. Each vertex participates in $\Theta(n/\log n)$ distinct $K_{q,q}$'s, and each $K_{q,q}$ contributes $\Theta(\log n)$ distinct clauses for $ALG(\varphi)$, the formula that is Steiner X-equivalent to φ, produced by STEINER-APPROX. Then $ALG(\varphi)$ has

$$\Theta(\log n) \cdot \Theta\left(\frac{n^2}{\log^2 n}\right) = \Theta\left(\frac{n^2}{\log n}\right)$$

clauses and the ratio $\frac{ALG(\varphi)}{OPT_S(\varphi)}$ is at least

$$\frac{\Theta\left(n^2/(\log n)\right)}{n} = \Omega\left(\frac{n}{\log n}\right).$$

□

As we saw minimizing the number of distinct bodies in a Horn formula can provide useful first steps in Horn minimization. Then a natural question is given a Horn formula φ and the minimum number of distinct bodies, how many heads can we expect in MIN-BODY(φ)? We investigate this question on hydras, a subclass of Horn formulas.

4.4 Hydra Formulas

We now turn our attention to the main focus of this chapter, hydra formulas and hydra number.

Definition 4.4.1. *A definite 3-Horn formula φ is a* hydra *formula, or a hydra, if for every clause $x, y \to z$ in φ and every variable u, the clause $x, y \to u$ also belongs to φ.*

For example, $(x, y \to z) \wedge (x, y \to u) \wedge (x, z \to y) \wedge (x, z \to u)$ is a hydra[1].

In the following proposition we note that every prime implicate of a hydra is a clause occurring in the hydra itself (this is not true for definite 3-Horn formulas in general). Thus minimization for hydras amounts to selecting a minimal number of clauses from the hydra that are equivalent to the original formula.

---

[1]Redundant clauses like $x, y \to x$ are omitted for simplicity.

Proposition 4.4.2. *Every prime implicate of a hydra belongs to the hydra.*

Proof. First note that all prime implicates of a definite Horn formula are definite Horn clauses [24]. Let us consider a hydra φ and a definite Horn clause C. If the body of C is of size 1, or it is of size 2 but it does not occur as a body in φ then forward chaining cannot mark any further variables, thus C cannot be an implicate. If the body of C has size at least 3 then it must contain a body x, y occurring in φ, otherwise, again, forward chaining cannot mark any further variables. But then the clause $x, y \to \mathit{head}(C)$ occurs in φ and so C is not prime. □

A definite Horn formula may also be viewed as a directed hypergraph of the type described in the introduction, and the two descriptions of forward chaining are equivalent. The *closure* $cl_H(S)$ of a set of vertices S with respect to H is the set of vertices marked by forward chaining started from S. A set of vertices is *good* if its closure is the set of all vertices.

For completeness, we restate the main notions used in this chapter.

Definition 4.4.3. *A directed 3-hypergraph* $H = (V, F)$ represents *an undirected graph* $G = (V, E)$ *if*

i. $(u, v) \in E$ *implies* $cl_H(u, v) = V$,

ii. $(u, v) \notin E$ *implies* $cl_H(u, v) = \{u, v\}$.

Definition 4.4.4. *The* hydra number $h(G)$ *of an undirected graph* $G = (V, E)$ *is*

$$\min\{|F| \,:\, H = (V, F)\, \mathit{represents}\, G\}.$$

Proposition 4.4.2 implies that the minimal formula size of a hydra φ and the hydra number of the undirected graph G formed by the bodies in φ are the same.

Remark 4.4.5. *For the rest of the chapter we assume that every variable in a hydra occurs in some body, or, equivalently, that graphs contain no isolated vertices. The removal of a variable occurring only as a head decreases minimal formula size by one, and, similarly, the removal of an isolated vertex decreases the hydra number by one.*

For the remainder of the chapter we use hypergraph terminology.

4.5 The Hydra Number of Graphs

In this section we note some simple properties of the hydra number. We start with an easy bound.

Proposition 4.5.1. *For every graph $G = (V, E)$ with at least three vertices*

$$|E(G)| \leq h(G) \leq 2|E(G)|.$$

Proof. For the upper bound construct a hypergraph of size $2|E(G)|$ by first ordering the edges of G, and then using each edge as the body of two hyperarcs whose heads are the two endpoints of the next edge in G. For the lower bound, note that each edge of G must be a body of at least one hyperarc. □

Equality holds in the upper bound when G is a matching, and only in this case.

Graphs satisfying the lower bound are of particular interest as they represent 'most compressible' hydras.

Definition 4.5.2. *A graph G is* single-headed *if $h(G) = |E(G)|$.*

A graph is single-headed iff there is a hypergraph $H = (V, F)$ such that every edge of G has *exactly* one head assigned to it, every hyperarc body in H is an edge of G and every edge of G is good in H. Cycles, for example, are single-headed, as shown by the directed hypergraph

$$(v_1, v_2 \to v_3), (v_2, v_3 \to v_4), \ldots, (v_{k-1}, v_k \to v_1). \tag{4.1}$$

Adding edges to the cycle preserves single-headedness. For example, the graph obtained by adding edge (v_i, v_j) is represented by the directed hypergraph obtained from (4.1) by adding the hyperarc $v_i, v_j \to v_{i+1}$, where $i + 1$ is meant modulo m. Thus we obtain the following.

Proposition 4.5.3. *Hamiltonian graphs are single-headed.*

We will discuss stronger forms of this statement in the next section. Matchings, on the other hand, satisfy the upper bound in Proposition 4.5.1. Indeed, every edge must occur as the body of at least two hyperarcs as otherwise forward chaining cannot mark any further vertices.

We call a body u, v single-headed (resp., multi-headed) with respect to a directed hypergraph H representing a graph G, if it is the body of exactly one (resp., more than one) hyperarc of H.

Remark 4.5.4. *Assume that the directed hypergraph $H = (V, F)$ represents the graph $G = (V, E)$ and $|V| \geq 4$. If $u, v \to w \in F$ and u, v is single-headed in H then w must be a neighbor of u or v. Indeed, otherwise $cl_H(u, v) = \{u, v, w\} \subset V$. This is a fact which we use numerous times in our proofs without referring to it explicitly.*

The following proposition generalizes the argument proving Proposition 4.5.3.

Proposition 4.5.5. *Let G be a graph with no isolated vertices and let G' be a spanning subgraph of G with no isolated vertices. Then*

$$h(G) \leq h(G') + |E(G)| - |E(G')|.$$

If G' is single-headed then G is also single-headed.

Proof. Let H' be a directed hypergraph of size $h(G')$ representing G'. Since G' is a spanning subgraph of G and contains no isolated vertices, for every edge $(u, v) \in E(G) \smallsetminus E(G')$ there is an edge $(v, w) \in E(G')$. The directed hypergraph H representing G obtained from H' by adding the hyperarc $u, v \to w$ to H' for each edge $(u, v) \in E(G) \smallsetminus E(G')$ satisfies the requirements. The second statement follows trivially. □

A second proposition gives a sufficient condition for single-headedness based on single-headedness of a non-spanning subgraph.

Proposition 4.5.6. *Let G be a connected graph and $(u, v) \notin E(G)$. Construct the graph $\hat{G}$ with vertex set $V(\hat{G}) = V(G) \cup \{w\}$ and edge set $E(\hat{G}) = E(G) \cup \{(u, v), (v, w)\}$, for some $w \notin V(G)$. If G is single-headed then $\hat{G}$ is single-headed.*

Proof. Let H be a directed hypergraph representing G and containing exactly $|E(G)|$ hyperarcs. Construct $\hat{H}$ from H by adding hyperarcs $u, v \to w$ and $v, w \to z$, where z is a neighbor of v in G guaranteed to exist by the connectivity of G. Since all pairs in $E(G)$ reach both u and v

in H (and in $\hat{H}$), hyperarc $u, v \to w$ ensures all pairs in $E(G)$ can reach in $\hat{H}$ the new variable w as well. On the other hand, hyperarc $v, w \to z$ ensures that the new pairs (u, v) and (v, w) can reach all other variables. Finally, there are $|E(\hat{G})|$ hyperarcs in H. □

Next we see a general sufficient condition for a graph *not* to be single-headed.

Proposition 4.5.7. *Let G be the union of two disjoint subgraphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, connected by a cut-edge. If both G_1 and G_2 contain at least two vertices then G is not single-headed.*

Proof. Assume that G is single-headed and let H be a directed hypergraph demonstrating this. Let $u \in G_1$, $v \in G_2$, and (u, v) be the cut-edge. There is exactly one hyperarc of the form $u, v \to z$ in H. If z is in G_1 (resp., in G_2) then forward chaining started from z, u (resp., z, v) cannot mark any vertices in G_2 (resp., G_1) other than v (resp., u). □

4.6 Line Graphs

In this section we consider a graph parameter that can be used to prove bounds on the hydra number. The *line graph* $L(G)$ of G has vertex set $V(L(G)) = E(G)$ and edge set $E(L(G)) = \{(e, f) | e \neq f \in E(G) \text{ and } e \cap f \neq \emptyset\}$. A (vertex-disjoint) *path cover* of G is a set of vertex-disjoint paths such that every vertex $v \in V$ is in exactly one path. The *path cover number* of G is the smallest integer k such that G has a path cover containing k paths.

In Proposition 4.5.3 we noted that hamiltonian graphs are single-headed. This can be extended to show that hamiltonicity of the line graph is also sufficient for single-headedness. Note that hamiltonicity of the line graph is a strictly weaker condition than hamiltonicity.

Hamiltonicity of the graph is easily seen to imply hamiltonicity of the line graph, and a triangle with a pendant edge shows that the converse fails. Furthermore, the path cover number of the line graph of any spanning subgraph with no isolated vertices gives a general upper bound for the hydra number.

Theorem 4.6.1. *Let G be a graph with no isolated vertices and let G' be a spanning subgraph of G with no isolated vertices. Then the following statements are true:*

i. If $L(G')$ is hamiltonian then G is single-headed.

ii. If $L(G')$ has a path cover of size k then $h(G) \leq |E(G)| + k$.

Proof. By Proposition 4.5.5 it is sufficient to prove the bounds for G'.

Proof of i. Let C be a hamiltonian cycle in $L(G')$. Direct the edges of C so that $\vec{C}$ is a directed hamiltonian cycle. The directed hypergraph H satisfying the requirements is constructed by adding a hyperarc $u, v \to w$ for each directed edge $(e, f) \in \vec{C}$, where $e = (u, v)$ and $f = (v, w)$. □

Proof of ii. Let $\{P_i\}_1^k$ be the minimum path cover of $L(G')$ and let l_i be the number of vertices of the path P_i. Direct the edges of each path P_i so that $\vec{P_i}$ is a directed path. Let $e_i = (x_i, y_i)$ and $f_i = (u_i, v_i)$ be the first and last edges in $\vec{P_i}$, respectively (if $\vec{P_i}$ is a single vertex then $e_i = f_i$).

We construct a directed hypergraph H representing G' and satisfying the requirements as follows. First, for each path $\vec{P_i}$ of at least 2 vertices we add $l_i - 1$ hyperarcs: for each directed edge $(e, f) \in \vec{P_i}$, where $e = (u, v)$ and $f = (v, w)$, add a hyperarc $u, v \to w$ to H.

If $k = 1$ then we complete the construction of H by adding two hyperarcs, $u_1, v_1 \to x_1$ and $u_1, v_1 \to y_1$. If $k > 1$ then we complete the construction by adding the $2k$ hyperarcs

$$(u_k, v_k \to x_1), (u_k, v_k \to y_1) \text{ and } (u_i, v_i \to x_{i+1}), (u_i, v_i \to y_{i+1}),$$

for $1 \leq i \leq k - 1$. □

□

For a dense graph G, the trivial upper bound for the hydra number, $2|E(G)|$, given by Proposition 4.5.1 is unnecessarily large. The following corollary of Theorem 4.6.1(ii) gives an upper bound as a function of the number of vertices.

Corollary 4.6.2. *Let G be a graph with no isolated vertices. Then*

$$h(G) \leq |E(G)| + \left\lfloor \frac{|V(G)|}{2} \right\rfloor.$$

Proof. Let $M \subseteq E(G)$ be a maximal set of independent edges of G, and let m be the number of edges it contains. Clearly $m \leq \lfloor |V(G)|/2 \rfloor$. Since G contains no isolated vertices, and M is maximal, each vertex u that is not an endpoint of an edge in M, must have a neighbor v that is an endpoint of an edge in M. Form the set $N \subset E(G)$ to consist of one such edge $(u, v) \in E(G)$ for each u not covered by M. Then $G' = (V, M \cup N)$ is a spanning subgraph of G containing no isolated vertices. Furthermore G' is the union of m components G_i, each of which consists of one edge with (possibly) pendant edges attached to its endpoints. Each component G_i has a line graph $L(G_i)$ consisting of one or two cliques sharing a vertex, and as such can be covered by one path. Then the path cover number of $L(G')$ is at most $m \leq \lfloor |V(G)|/2 \rfloor$. □

Remark 4.6.3. *We expect this bound to be very crude in most cases and certainly weaker than that path cover bound of Theorem 4.6.1, since it does not use the structure of G at all. That is also what makes it very accessible, as it is expressed solely in terms of the size of the edge and vertex sets of G.*

The condition of Theorem 4.6.1(i) is sufficient but not necessary for a graph G to be single-headed. In fact, there exist single-headed graphs such that the line graph of any of the connected spanning subgraphs has a large path cover number.

Theorem 4.6.4. *There is a family of single-headed graphs G_k with $\Theta(k)$ edges such that for every spanning subgraph $G' \subseteq G_k$ with no isolated vertices, $L(G')$ has path cover number $\Theta(k)$.*

Proof. Consider the sequence of graphs $\{G_k : k \geq 1\}$ constructed from an $8k$-cycle, with vertices $v_0, \dots, v_{8k-1}$, and pendant edges $x_i v_{4i}$ and $y_i v_{4k+4i}$ for $0 \leq i \leq k-1$. Add a vertex z_i and the edges (x_i, y_i), (y_i, z_i), for each i, $0 \leq i \leq k-1$, corresponding to the construction in Proposition 4.5.6.

By Proposition 4.5.6, G_k is single-headed, since a cycle with attached pendant edges has a hamiltonian line graph. We will show that for an arbitrary spanning subgraph $G' \subseteq G_k$ with no isolated vertices the path cover number of $L(G')$ is at least $k/4$.

Define D_i to be the set of vertices in the i_{th} diagonal of $L(G')$, namely $x_i v_{4i}$, $x_i y_i$, $y_i z_i$, and $y_i v_{4k+4i}$. Consider an arbitrary path cover $S = \{P_j : 1 \leq j \leq s\}$ of the vertices of $L(G')$.

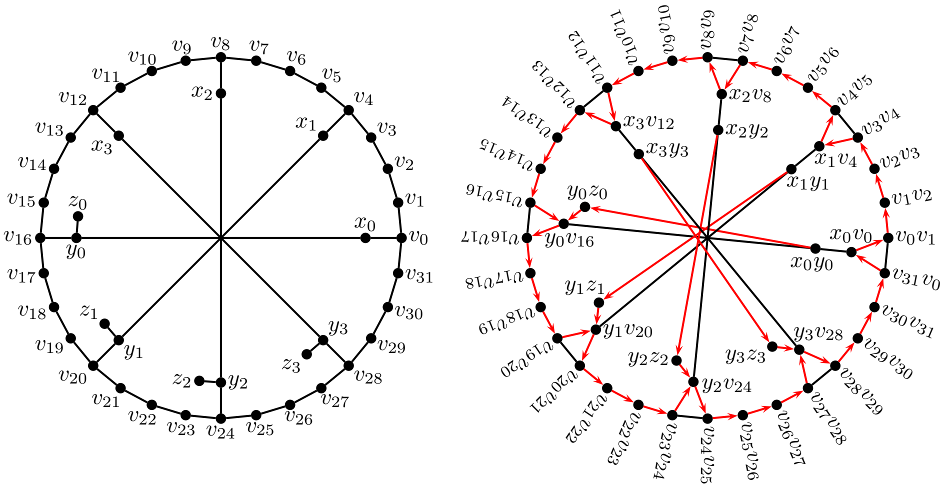

Figure 5. Single-headed graph G_k for $k = 4$ from Theorem 4.6.4 (left) and its line graph (right).

Lemma 4.6.5. *Let $D_i' = D_i \cap V(L(G'))$, and let $G[D_i']$ be the subgraph of $L(G')$ induced by D_i'. If $G[D_i']$ does not contain an endpoint of a path in S, then $D_i' = D_i$ and one path in S covers all vertices in D_i.*

Proof. Suppose that $G[D_i']$ does not contain an endpoint of a path in S, and assume for contradiction that $e \in D_i \setminus D_i'$. Since G' is spanning and contains no isolated vertices, it must contain the edge (y_i, z_i), and so $e \neq y_i z_i$. Also $e \notin \{x_i y_i, y_i v_{4k+4i}\}$, else $y_i z_i$ would be a degree-1 vertex in $L(G')$, and thus it would be an endpoint of a path in S. Furthermore $e \neq x_i v_{4i}$, otherwise S must have a path endpoint in the triangle $\{x_i y_i, y_i z_i, y_i v_{4k+4i}\}$. Thus D_i is contained in the vertex set of $L(G')$, and due to the structure of the diagonal and the assumption that no path endpoints of S fall in D_i, all vertices in the diagonal are covered by exactly one path P of S. □

Define X_i to include all vertices in D_i along with the cycle vertices $v_{4i-3}v_{4i-2}$, $v_{4i-2}v_{4i-1}$, $v_{4i-1}v_{4i}$, $v_{4i}v_{4i+1}$, $v_{4i+1}v_{4i+2}$, $v_{4i+2}v_{4i+3}$, and their antipodes on the circle $v_{4k+4i-3}v_{4k+4i-2}$, $v_{4k+4i-2}v_{4k+4i-1}$, $v_{4k+4i-1}v_{4k+4i}$, $v_{4k+4i}v_{4k+4i+1}$, $v_{4k+4i+1}v_{4k+4i+2}$, $v_{4k+4i+2}v_{4k+4i+3}$.

Let $X_i{}' = X_i \cap V(L(G'))$. We claim that the subgraph $G[X_i{}']$ induced by the vertex set $X_i{}'$ contains at least one endpoint of a path in S. Suppose not. By Lemma 4.6.5 all vertices in D_i are in $L(G')$. A case analysis shows that all other vertices in X_i must be present, otherwise a degree-1 vertex is introduced in $G[X_i{}']$ or G' is not both spanning and without isolated vertices. Thus there must be a path P in S going through all the vertices of X_i. A further case analysis shows that this is not possible.

There are $k/2$ disjoint sets $X_i{}'$ and so there are at least $k/4$ paths in S. □

A more involved case analysis gives at least two endpoints of paths of S in $X_i{}'$, and so at least $k/2$ paths in S.

4.7 The Hydra Number of Trees

Proposition 4.5.5, and by extension Theorem 4.6.1 suggest that many edges in a graph G do not necessarily improve $h(G)$, and so considering the hydra number of sparse graphs would be interesting. It is natural then to investigate the hydra number of the most sparse of connected graphs, the tree.

4.7.1 Trees with Low Hydra Number

We begin the discussion of the hydra number of trees, with trees having low hydra numbers, that is, hydra number $|E(T)|$ or $|E(T)|+1$.

A *star* is a tree that contains no length-3 path. A *caterpillar* is a tree for which deleting all vertices of degree one and their incident edges from the tree gives a path. We call this path the spine of T, and note that it is unique. A useful characterization of caterpillars is that they do not contain the subgraph in Figure 6 [26] (see also [45, p.88]).

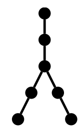

Figure 6. The forbidden subgraph for caterpillars.

Caterpillars have been instrumental in [39], where finding maximal caterpillars starting from the leaves of the tree was the basis for a polynomial algorithm used to find a minimum hamiltonian completion of the line graph of a tree (which is the same as finding a minimum path cover). A linear algorithm was later put forth by [1] for the same problem. For general graphs the problem is NP-hard. Furthermore, [10] proves that finding a hamiltonian path is NP-complete even for line graphs.

Stars are the only trees that are single-headed, and caterpillars are the only non-star trees that can attain $h(T) = |E(T)| + 1$.

Theorem 4.7.1. *Let T be a tree. Then*

i. $h(T) = |E(T)|$ *if and only if T is a star.*

ii. $h(T) = |E(T)| + 1$ *if and only if T is a non-star caterpillar.*

We first show that a tree that is not a star cannot be single-headed.

Lemma 4.7.2. *If T is a tree that is not a star, then $h(T) \geq |E(T)| + 1$.*

Proof. Since T is not a star, it contains a path of length three. The middle edge is a cut-edge between two components of at least two vertices, hence we can apply Proposition 4.5.7. □

In fact a hypergraph that represents a non-caterpillar tree requires even more hyperarcs.

Lemma 4.7.3. *If T is a tree that is not a caterpillar then $h(T) > |E(T)| + 1$.*

Proof. A non-caterpillar tree T contains the subgraph in Figure 6. Let us call the central vertex of that forbidden subgraph u.

Assume for contradiction that H is a hypergraph with $|E(T)| + 1$ hyperarcs that represents T. Let the two-headed body of H be α.

We claim α must have a head in every non-singleton subtree attached to u that does not contain both vertices of α. Suppose not. Let v be a neighbor of u, and let T_v be a non-singleton subtree of T not containing any heads of α, and also not containing both vertices of α. Finally let $w \in V(T_v)$ be a neighbor of v. (See Figure 7.) Body u, v must have a head that is a neighbor

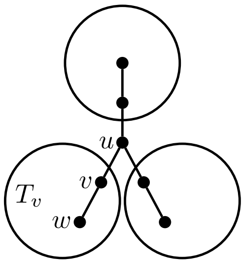

Figure 7. Part of the non-caterpillar tree T from the proof of Lemma 4.7.3.

of v in T_v: among the vertices in T_v, only v itself can be a head to a body completely outside T_v; so if u, v has only heads outside of T_v, then u, v cannot reach w. Body u, v must also have a head outside T_v, because otherwise only vertices in T_v and u would be reachable from u, v in H. So u, v must be α, which contradicts α having no heads in T_v.

Since there are at least three non-singleton subtrees attached to u, it must be that two of those subtrees each contain one head of α, and the third subtree contains both vertices of α. The two heads of α must not be adjacent to α, because they are in different subtrees. Those two heads also cannot be adjacent to each other. Therefore, the only vertices reachable from α in H are α's two heads and α itself. □

Proof of Theorem 4.7.1. We need to prove the upper bounds. The single-headedness of stars is easily seen directly, or follows from Theorem 4.6.1(i). For T a caterpillar, the upper bound follows from Theorem 4.6.1(ii) as the line graph of a caterpillar contains a hamiltonian path. □

4.7.2 Lower Bounds for Trees

Let T^- be the tree obtained from T by removing all leaves, and define $\ell(T)$ to be the number of leaves in T^-. The following theorem generalizes Lemma 4.7.3.

Theorem 4.7.4. *For a tree T that is not a star,*

$$h(T) \geq |E(T)| + \left\lceil \frac{\ell(T)}{2} \right\rceil .$$

Proof. Assume for contradiction that H is a hypergraph representing T with $|E(T)| + \lceil \ell(T)/2 \rceil - 1$ hyperarcs. Let v be a leaf in T^- and u be its unique neighbor in T^- and define T_v to be the subtree of T rooted at v and containing all leaf neighbors of v in T.

Claim 1. *For every subtree T_v there exists a multi-headed body α in H such that either*

(i) two heads of α are contained in T_v, or

(ii) both vertices of α are contained in T_v, or

(iii) one head of α is contained in T_v and α is the body u, v.

Proof of Claim. Otherwise body u, v cannot reach every vertex outside T_v and every vertex inside T_v. □

Then each subtree T_v "uses" the body or two heads of a multi-headed hyperarc. We will show that even in the case where the sum of bodies and heads of multi-headed bodies is maximum, we cannot "cover" all non-singleton subtrees.

Let b be the number of multi-headed bodies in H, and d be the average outdegree for the multi-headed bodies. The number of multi-headed hyperarcs in H can be counted by bd, but also by $b + \lceil \ell(T)/2 \rceil - 1$. On the other hand the sum of bodies and heads, if all heads are distinct, is counted by $b + bd = 2b + \lceil \ell(T)/2 \rceil - 1$, and so can be maximized by maximizing the number of multi-headed bodies, i.e. setting $b = \lceil \ell(T)/2 \rceil - 1$. Then $d = 2$, and all multi-headed hyperarcs are in fact two-headed.

So in this best case scenario there are $\lceil \ell(T)/2 \rceil - 1$ two-headed bodies and $\ell(T) - 2$ (resp. $\ell(T) - 1$) heads of a multi-headed body if $\ell(T)$ is even (resp. odd). But there are $\ell(T)$ subtrees and only $\ell(T) - 2$ (alternatively $\ell(T) - 1$ if $\ell(T)$ is odd) of them can be "covered" by multi-headed bodies and their heads. □

Theorem 4.7.4 can also be formulated for general graphs containing induced subgraphs that are trees (i.e. graphs that contain $\ell(G)$ degree≥ 2 neighbors of degree-1 vertices). The proof is also the same.

Theorem 4.7.5. *Let G^- be the graph obtained from G by removing all degree-1 vertices, and define $\ell(G)$ to be the number of degree-1 vertices in G^-. If $\ell(G) > 1$ then,*

$$h(G) \geq |E(G)| + \left\lceil \frac{\ell(G)}{2} \right\rceil .$$

4.7.3 Spiders and Large In-degrees

A *spider* is a tree with at most one vertex of degree at least 3, and all other vertices having degree at most 2. Let u be the vertex with degree at least 3. For convenience, we call the unique path from a leaf of a spider to u, including the endpoints, a *leg*. The upper bound from Theorem 4.6.1(ii) is sharp for spider trees (even if we require the subgraph G' to be connected) and so is the lower bound from Theorem 4.7.4.

Corollary 4.7.6. *If T is a spider tree with ℓ legs of length at least two, then*

$$h(T) = |E(T)| + \left\lceil \frac{\ell}{2} \right\rceil .$$

Proof. The upper bound is a corollary of Theorem 4.6.1(ii) as the path cover number of $L(T)$ is $\lceil \ell/2 \rceil$, since that is the minimum number of caterpillars in which the edges of T can be decomposed, with each caterpillar, except possibly the last one, containing two legs of T.

The lower bound is a corollary of Theorem 4.7.4, since $\ell = \ell(T)$. □

Another interesting parameter of a graph G is the largest in-degree of a vertex in an optimal hypergraph H representing G. The following two theorems demonstrate a family of trees T_k with large in-degree in the optimal hypergraph representing T_k.

Let T_k be a spider tree with exactly k legs all of which have length 2.

Theorem 4.7.7. *The central vertex u has $\lfloor k/2 \rfloor$ heads in an optimal hydra H representing T_k.*

Proof. First note that T_k contains $2k$ edges and $2k + 1$ vertices. From Corollary 4.7.6 we know that an optimal hydra H representing T_k contains $|E(T_k)| + \lceil k/2 \rceil$ hyperarcs, and, from the proof of Theorem 4.7.4, we know that there are $\lfloor k/2 \rfloor$ multi-headed bodies in H all of which have two heads.

Let u be the central vertex of T_k, Furthermore, as in the proof of Theorem 4.7.4, for every non-singleton subtree T_{v_i} rooted at a neighbor v_i of the central vertex u there exists a multi-headed body α such that either

i. two heads of α are contained in T_{v_i}, or

ii. both vertices of α are contained in T_{v_i}, or

iii. one head of α is contained in T_{v_i} and α is the body u, v_i.

In fact, with a small exception for k odd, all heads and bodies in each subtree T_{v_i} described above must be single-headed, otherwise there are more than $|E(T_k)| + \lceil k/2 \rceil$ hyperarcs in H. We will examine the subtrees in pairs and show that, for each pair, u is a head for at least one hyperarc. (If k is odd one subtree is allowed to contain both a two-headed body and the heads of a multi-headed body.)

Let w_i be the neighbor of v_i in T_{v_i}. If T_{v_i} satisfies case i., then hyperarc $v_i, w_i \to u$ is in H. If T_{v_i} satisfies case ii., then there exists subtree T_{v_j} such that T_{v_j} falls in case i., i.e. v_j and w_j are the two heads of body v_i, w_i. Finally subtree T_{v_i} may satisfy case iii., in which case there exists a subtree T_{v_j} such that v_j is the second head of body u, v_i. All pairs in $\{u, v_i, w_i, v_j, w_j\}$ other than u, v_i are single-headed, and vertices u, v_i and w_j are heads of single-headed bodies only. Then bodies v_i, w_i and v_j, w_j must both have u as their head (also u, v_j has a head $v_{j'}$ in a different subtree, and v_i is a head of body outside T_{v_i} and T_{v_j}).

Then we can partition our subtrees into $\lfloor k/2 \rfloor$ pairs (and possibly a single subtree) each containing at least one hyperarc with head u. Finally u has in-degree at least $\lceil k/2 \rceil$ in an optimal hydra H representing T_k. □

If we restrict the in-degree of the vertices of $V(T_k)$ in a hypergraph H representing T_k, then we increase the size of an optimal hypergraph H representing T_k by a term of order $|V(T_k)|$.

Theorem 4.7.8. *Let $d \geq 1$ be a constant. If H is an optimal hypergraph representing T_k, such that no vertex in T_k is a head of more than d hyperarcs, then the number of hypearcs in H is at least $|E(T_k)| + k - d$.*

Proof. Let T_{v_i} be the subtree of T_k rooted at vertex v_i neighboring u, and w_i be the other neighbor of v_i. There are k such subtrees, and at most d of them can contain an edge (v_i, w_i), whose endpoints form a single-headed body in H, as the only edge incident to (v_i, w_i) is (u, v_i). For any other subtree T_{v_j}, v_j, w_j is a multi-headed body in H. Then H contains at least $|E(T_k)| + k - d$ hyperarcs. Since $k = (|V(G)| - 1)/2$, this is an increase in the size of the optimal H by about $|V(G)| - d$. □

4.7.4 Complete Binary Trees

In this section we obtain upper and lower bounds for $h(G)$ when G is a complete binary tree. A *complete binary tree* of depth d, denoted B_d, is a tree with $d + 1$ levels, where every node on levels 1 through d has exactly 2 children. B_d has $2^{d+1} - 1$ vertices and $2^{d+1} - 2$ edges.

Theorem 4.7.9. *For $d \geq 3$ it holds that*

$$\frac{9}{8} |E(B_d)| \leq h(B_d) \leq \frac{17}{15} |E(B_d)| + 1.$$

Proof. The lower bound follows from Theorem 4.7.4, since there are 2^{d-1} leaves in $(B_d)^-$ (all vertices in the $(d - 1)$ level of the tree).

For the upper bound, define $p(T)$ to be the minimum number of edge-disjoint caterpillar subtrees required to cover the *vertices* of a tree T. Note that this is equal to the minimum path cover number of $L(T')$ over all subgraphs $T' \subset T$ with no isolated vertices.

First let $d = 4k+3$ for some $k > 1$. We show that there is a spanning subgraph B' of B_{4k+3} with no isolated vertices with a path cover containing at most $(2/15)|E(B_{4k+3})| + 2/15$ paths. Specifically we delete every fourth level of edges in the binary tree, leaving a forest of B_3's. This corresponds to deleting every fourth level of vertices in $L(B_{4k+3})$, to obtain $L(B')$. By inspection, $p(B_3) = 2$. Considering the top three levels of the line graph $L(B_{4k+3})$, we note the following inductive inequality:

$$p\left(B_{4k+3}\right) \leq 2 + 16p\left(B_{4k-1}\right), \tag{4.2}$$

which holds because we can construct a path cover of $L(B')$ by covering the top copy of $L(B_3)$ in $L(B')$ with 2 paths, and noting the 16 copies of $L(B_{4k-3})$ remaining in $L(B')$ (See Figure 8.)

Inequality (4.2) and the inductive hypothesis together complete the $d = 4k+3$ case as follows

$$p\left(B_{4k+3}\right) \leq 2\sum_{i=0}^{k} 16^i = \frac{2}{15}\left|E\left(B_{4k+3}\right)\right| + \frac{2}{15}. \tag{4.3}$$

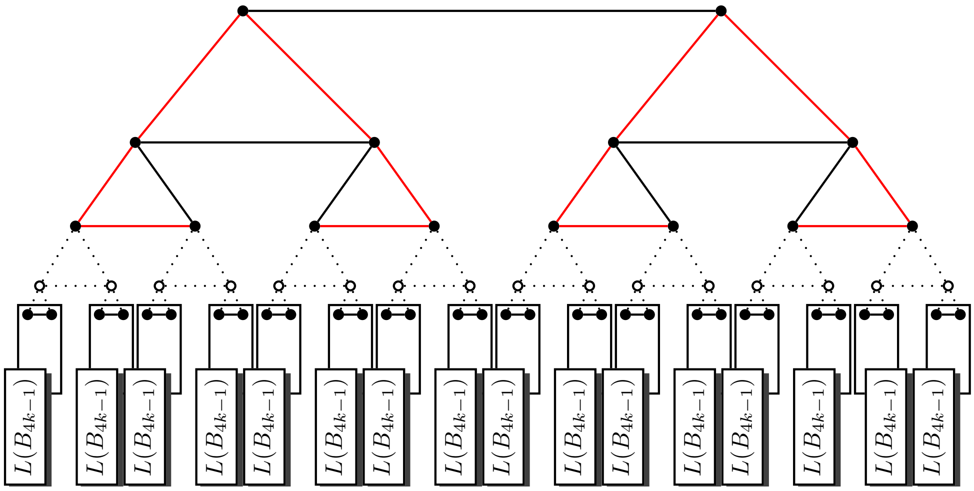

Figure 8. Induction on the line graph $L(B_{4k+3})$ from the proof of Theorem 4.7.9.

For the case $d = 4k+2$, construct $L(B')$ by deleting the third level of vertices of $L(B_{4k+2})$. Cover the vertices of the top two levels of $L(B_{4k+2})$ with one path, and note that there are 8 copies of $L(B_{4k-1})$ covering the remaining vertices of $L(B')$. Thus by Equation 4.3:

$$p\left(B_{4k+2}\right) \leq 1 + 8p\left(B_{4k-1}\right) \leq \frac{2}{15}\left|E\left(B_{4k+2}\right)\right| + \frac{1}{5}.$$

The case $d = 4k+1$ is handled similarly. Construct $L(B')$ by deleting the second level of vertices of $L(B_{4k+1})$. Cover the vertices of the top level of $L(B')$ with one path, and note that there are 4 copies of $L(B_{4k-1})$ covering the remaining vertices of $L(B')$. Once again, by Equation 4.3:

$$p\left(B_{4k+1}\right) \leq 1 + 4p\left(B_{4k-1}\right) \leq \frac{2}{15}\left|E\left(B_{4k+1}\right)\right| + \frac{11}{15}.$$

Finally let $d = 4k$. This time we cover the vertices of the top level of $L(B_{4k})$ with one path, and note that there are 2 copies of $L(B_{4k-1})$ covering the remaining vertices of $L(B_{4k})$. Thus

$$p(B_{4k}) \leq 1 + 2p(B_{4k-1}) \leq \frac{2}{15}|E(B_{4k})| + 1.$$

□

The following theorem demonstrates the significance of considering non-connected subgraphs of G when upper-bounding $h(G)$ in Theorem 4.6.1(ii). Even for trees, there are examples where the path cover number of the line graph of the tree is significantly larger than the hydra number of the tree.

Theorem 4.7.10. *Let $g(T)$ to be the minimum number of paths in a path cover of $L(T)$. Then*

$$h(B_d) \leq g(B_d) - \Theta(|V(B_d)|).$$

Proof. We claim that the path cover number of $L(B_d)$ is

$$g(B_d) = \begin{cases} \frac{1}{7}|E(B_d)| & \text{for } d \equiv 0 \bmod 3, \\ \frac{1}{7}|E(B_d)| + \frac{5}{7} & \text{for } d \equiv 1 \bmod 3, \\ \frac{1}{7}|E(B_d)| + \frac{1}{7} & \text{for } d \equiv 2 \bmod 3. \end{cases} \tag{4.4}$$

By [39], $g(B_d)$ is equal to the minimum number of caterpillars in a caterpillar-decomposition of the edges of B_d. [39] also provides an algorithm to calculate this number. Since the bottom

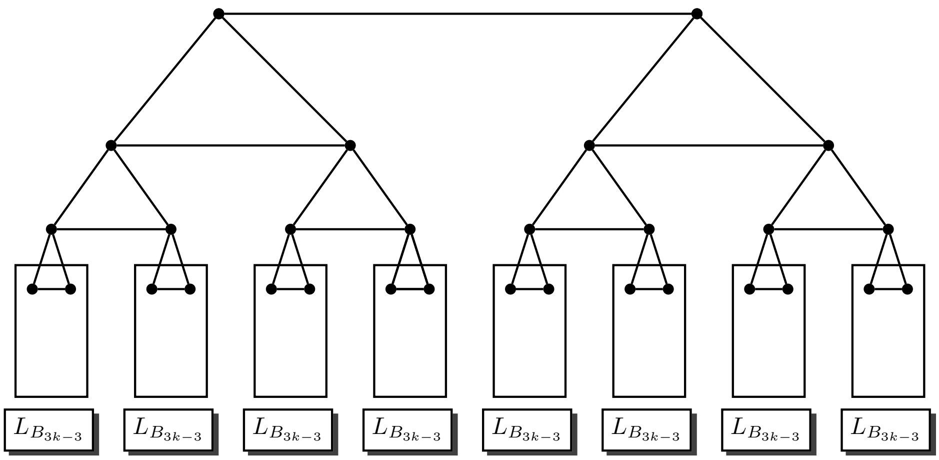

Figure 9. Line graph $L(B_{3k})$ from the proof of Theorem 4.7.10.

levels of B_d are furthest from the (bi)center of the tree, it will be optimal to cover the bottom-most parts of the tree with as few caterpillars as possible. Thus we distinguish the three cases in 4.4, and form inductive equations for each case.

We first let $d = 3k$ for some $k > 1$. We show inductively that $g(B_{3k})$ is $|E(B_{3k})|/7$. By inspection, $g(B_3) = 2$. Considering the top three levels of the line graph $L(B_{3k})$ we note the following inductive inequality:

$$g\left(B_{3k}\right) = 2 + 8g\left(B_{3k-3}\right), \tag{4.5}$$

which holds because we can construct a path cover by covering the top copy of $L(B_3)$ in $L(B_{3k})$ with 2 paths, and noting the 8 copies of $L(B_{3k-3})$ remaining in $L(B_{3k})$ (See Figure 9.) Inequality (4.5) and the inductive hypothesis together complete the $d = 3k$ case.

For the case $d = 3k + 1$, cover the vertices of the top level of $L(B_{3k+1})$ with one path and note that there are 2 copies of $L(B_{3k})$ covering the remaining vertices of $L(B_{3k+1})$. Thus

$$g\left(B_{3k+1}\right) = 1 + 2g\left(B_{3k}\right),$$

which together with the formula for the $d = 3k$ case gives the desired result.

Finally the $d = 3k + 2$ case is handled in a similar way. This time we cover the vertices of the top two levels of $L(B_{3k+2})$ with one path and note that there are 4 copies of $L(B_{3k})$ remaining. Thus

$$g\left(B_{3k+2}\right) = 1 + 4g\left(B_{3k}\right).$$

□

4.8 Hydra Number and Total Interval Number

It is worth noting that, for a triangle-free graph, we can upper bound the hydra number by a different graph metric, the total interval number of the graph. This follows from the theorem:

Theorem 4.8.1 ([39; 30]). *For a triangle-free graph G, the total interval number of G, $\tau(G)$ is equal to $|E(G)| + g(G)$.*

By Theorem 4.6.1(ii) for a triangle-free graph G,

$$h(G) \leq |E(G)| + g(G) = \tau(G).$$

Corollary 4.8.2. *For a connected triangle-free graph G, $h(G) \leq \tau(G)$.*

This corollary also suggests an answer to the following question: what is the maximal hydra number among n-vertex trees? It is, in fact, the maximal total interval number for a tree, *i.e.* $\lfloor (5|V(G)| - 3)/4 \rfloor$ [2; 30] or equivalently $\lfloor (5|E(G)| + 2)/4 \rfloor$. This bound on the hydra number is sharp and we have already seen a family of trees that attains it: the family T_k of Section 4.7.3.

Theorem 4.8.3. *The maximum hydra number of a tree G is $\lfloor (5|V(G)| - 3)/4 \rfloor$, and this is sharp.*

Proof. By [2] and Corollary 4.8.2, for any graph G, $h(G) \leq \tau(G) \leq \lfloor (5|V(G)| - 3)/4 \rfloor$. On the other hand recall that T_k is a spider tree with k legs of length 2, $|V(T_k)| = 2k+1$, $|E(T_k)| = 2k$ and hydra number $h(T_k) = |E(T_k)| + \lceil k/2 \rceil = \lfloor (5|V(T_k)| - 3)/4 \rfloor$. Then $\lfloor (5|V(T_k)| - 3)/4 \rfloor$ is the maximum hydra number of any tree, and it is attained by T_k. □

4.9 Minimal Directed Hypergraphs with All k-tuples Good

In this section we consider a problem related to hydra numbers. Given n and a number k $(2 \leq k \leq n-1)$, let $f(n,k)$ be the minimal number of hyperarcs in an n-vertex 3-uniform directed hypergraph H such that every k-element subset of the vertices is good for H. The case $k = 2$ is just the hydra number of complete graphs and so $f(n,2) = \binom{n}{2}$.

We use Turán's theorem from extremal graph theory (see, e.g. [45]). The *Turán graph* $T(n, k-1)$ is formed by dividing n vertices into $k-1$ parts as evenly as possible (i.e., into parts of size $\lfloor n/(k-1) \rfloor$ and $\lceil n/(k-1) \rceil$) and connecting two vertices iff they are in different parts. The number of edges of $T(n, k-1)$ is denoted by $t(n, k-1)$. If $k-1$ divides n then

$$t(n, k-1) = \left(1 - \frac{1}{k-1}\right) \frac{n^2}{2}.$$

Turán's theorem states that if an n-vertex graph contains no k-clique then it has at most $t(n, k-1)$ edges and the only extremal graph is $T(n, k-1)$. Switching to complements it follows that if an n-vertex graph has no empty subgraph on k vertices then it has at least $\binom{n}{2} - t(n, k-1)$ edges.

Theorem 4.9.1. *If* $k \leq (n/2) + 1$ *then*

$$\binom{n}{2} - t(n, k-1) \leq f(n, k) \leq \binom{n}{2} - t(n, k-1) + (k-1).$$

Proof. Suppose H is a 3-uniform directed hypergraph with all k-tuples good. Then every k-element set S of vertices must contain at least one body of a hyperarc in H, otherwise forward chaining started from S cannot mark any vertices. Thus the undirected graph formed by the bodies in H contains no empty subgraph on k vertices, and the lower bound follows by Turán's theorem.

For the upper bound we construct a directed hypergraph based on the complement of $T(n, k-1)$ over the vertex set $\{x_1, \ldots, x_n\}$, consisting of $k-1$ cliques of size differing by at most 1. Assume that each clique has size at least 3. In each clique do the following. Pick a hamiltonian path, direct it, and introduce hyperarcs as in (4.1) (with the exception of the last edge closing the cycle). For every other edge (u, v), introduce a hyperarc $u, v \to w$ where w is a vertex on the hamiltonian path that is adjacent to u or v. For each edge e closing a hamiltonian cycle, add *two* hyperarcs with body e, and heads the endpoints of the first edge on the hamiltonian path of the next clique (where 'next' assumes an arbitrary cyclic ordering of the cliques). For cliques of size 2 the single edge in the clique plays the role of the unassigned edge and the construction is similar. □

4.10 Open Problems

We list only a few of the related open problems. As computing hydra numbers is a special case of Horn minimization, it would be interesting to determine the computational complexity of computing hydra numbers and recognizing single-headed graphs. What is the maximal hydra number among connected n-vertex graphs? Can the path cover number of the line graph be used to get a lower bound for the hydra number of trees?

APPENDICES

Appendix A

PERMISSIONS

Figure 1 was reprinted from J. Combin. Theory Ser. B, 97(4), Pelsmajer, M. J., Schaefer, M., and Štefankovič, D., Removing even crossings, 489–500, Copyright (2007), with permission from Elsevier.

Material in chapter 2 was reprinted from SIAM J. Discrete Math., 23(3), Pelsmajer, M. J., Schaefer, M., and Stasi, D., Strong Hanani-Tutte on the projective plane, 1317–1323, Copyright (2007), with permission from SIAM.

Subject: RE: Permission to use material in doctoral thesis
From: "Kelly Thomas" <Thomas@siam.org>
Date: 5/15/2012 12:20 PM
To: "Despina Stasi" <stasi@math.uic.edu>

Dear Ms. Stasi:

SIAM is happy to give permission to reuse material from the article mentioned below, but please be aware that Figure 1 requires permission from the original source.

You may use the usual style of your school for acknowledgment in the credit line.

Sincerely,

Kelly Thomas
Managing Editor, SIAM

From: despina@gmail.com [mailto:despina@gmail.com] **On Behalf Of** Despina Stasi
Sent: Monday, May 14, 2012 5:18 PM
To: Kelly Thomas
Subject: Permission to use material in doctoral thesis

Dear Ms. Thomas,

I am writing to request permission to use the following material from your publication *Pelsmajer, M. J., Schaefer, M., and Stasi, D.: Strong Hanani-Tutte on the projective plane. SIAM J. Discrete Math., 23(3):1317--1323, 2009* in my doctoral thesis. This material will appear as originally published. Unless you request otherwise, I will use the conventional style of the Graduate College of the University of Illinois at Chicago as acknowledgment.

Sincerely,

Despina Stasi
Dept. Mathematics, Statistics, and Computer Science
University of Illinois at Chicago

Office: 1221 SEO
Phone: 312-413-2167
Web: www.math.uic.edu/~stasi

ELSEVIER LICENSE
TERMS AND CONDITIONS

Jul 31, 2012

This is a License Agreement between Despina Stasi ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Despina Stasi
Customer address	MSCS, University of Illinois at Chicago Chicago, IL 60607-7045
License number	2910380871342
License date	May 15, 2012
Licensed content publisher	Elsevier
Licensed content publication	Journal of Combinatorial Theory, Series B
Licensed content title	Removing even crossings
Licensed content author	Michael J. Pelsmajer,Marcus Schaefer,Daniel Štefankovič
Licensed content date	July 2007
Licensed content volume number	97
Licensed content issue number	4
Number of pages	12
Start Page	489
End Page	500
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
Format	electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	Combinatorial Problems in Graph Drawing and Knowledge Representation
Expected completion date	May 2012
Estimated size (number of pages)	70
Elsevier VAT number	GB 494 6272 12

Permissions price	0.00 USD
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
Total	0.00 USD
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's

Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Website**: The following terms and conditions apply to electronic reserve and author websites:
Electronic reserve: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:
This license was made in connection with a course,

This permission is granted for 1 year only. You may obtain a license for future website posting,
All content posted to the web site must maintain the copyright information line on the bottom of each image,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com , and
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

17. **Author website** for journals with the following additional clauses:

All content posted to the web site must maintain the copyright information line on the bottom of each image, and the permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version. A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx . As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

18. **Author website** for books with the following additional clauses:
Authors are permitted to place a brief summary of their work online only.
A hyper-text must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx. or for books to the Elsevier homepage at http://www.elsevier.com

20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

21. **Other Conditions**:

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK500780116.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

CITED LITERATURE

1. Agnetis, A., Detti, P., Meloni, C., and Pacciarelli, D.: A linear algorithm for the Hamiltonian completion number of the line graph of a tree. Inf. Process. Lett., 79:17–24, 2001.

2. Andreae, T. and Aigner, M.: The total interval number of a graph. Journal of Combinatorial Theory, Series B, 46(1):7 – 21, 1989.

3. Angluin, D., Frazier, M., and Pitt, L.: Learning conjunctions of Horn clauses. Machine Learning, 9:147–164, 1992.

4. Archdeacon, D.: A Kuratowski theorem for the projective plane. Journal of Graph Theory, 5(3):243–246, 1981.

5. Arias, M. and Balcázar, J. L.: Construction and learnability of canonical Horn formulas. Machine Learning, 85:273–297, 2011.

6. Ausiello, G., D'Atri, A., and Saccà, D.: Minimal representation of directed hypergraphs. SIAM J. Comput., 15(2):418–431, 1986.

7. Ausiello, G., Italiano, G. F., Laura, L., Nanni, U., and Sarracco, F.: Classification and traversal algorithmic techniques for optimization problems on directed hyperpaths. Technical Report n.18, Dipartimento di Informatica e Sistemistica "Antonio Ruberti", Università di Roma "La Sapienza", 2010.

8. Berke, R. and Onsjö, M.: Propagation connectivity of random hypergraphs. In Stochastic Algorithms: Foundations and Applications, SAGA, pages 117–126. Springer LNCS 5792, 2009. An update appears in [9].

9. Berke, R. and Onsjö, M.: Propagation connectivity of random hypergraphs. Technical Report 1342-2812, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, October 2009.

10. Bertossi, A. A.: The Edge Hamiltonian Path Problem is NP-complete. Inf. Process. Lett., 13(4/5):157–159, 1981.

11. Bhattacharya, A., DasGupta, B., Mubayi, D., and Turán, G.: On approximate Horn formula minimization. In ICALP (1), eds, S. Abramsky, C. Gavoille, C. Kirchner, F. Meyer auf der Heide, and P. G. Spirakis, volume 6198 of Lecture Notes in Computer Science, pages 438–450. Springer, 2010.

12. Bollobás, B.: Random Graphs. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2001.

13. Boros, E.: personal communication, 2012.

14. Boros, E. and Gruber, A.: Hardness results for approximate pure Horn CNF formulae minimization. In International Symposium on AI and Mathematics (ISAIM), 2012.

15. Cairns, G. and Nikolayevsky, Y.: Bounds for generalized thrackles. Discrete Comput. Geom., 23(2):191–206, 2000.

16. Chojnacki, C. (Hanani, H): Über wesentlich unplättbare Kurven im drei-dimensionalen Raume. Fundamenta Mathematicae, 23:135–142, 1934.

17. Chung, F. R. K. and Horn, P.: The spectral gap of a random subgraph of a graph. Internet Mathematics, 4(2):225–244, 2007.

18. Coja-Oghlan, A., Onsjö, M., and Watanabe, O.: Propagation connectivity of random hypergraphs. In Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, volume 6302 of Lecture Notes in Computer Science, pages 490–503. Springer Berlin / Heidelberg, 2010.

19. Demopoulos, D. D. and Vardi, M. Y.: The phase transition in the random HornSAT problem. In Computational Complexity and Statistical Physics (Santa Fe Institute Studies in the Sciences of Complexity Proceedings), pages 195–220. Oxford University Press, Inc., 2006.

20. Diestel, R.: Graph theory, volume 173 of Graduate Texts in Mathematics. Berlin, Springer-Verlag, third edition, 2005.

21. Dunne, P. E. and Bench-Capon, T. J. M.: A sharp threshold for the phase transition of a restricted satisfiability problem for Horn clauses. J. Log. Algebr. Program., 47(1):1–14, 2001.

22. Glover, H. H., Huneke, J. P., and Wang, C. S.: 103 graphs that are irreducible for the projective plane. J. Comb. Theory, Ser. B, 27(3):332–370, 1979.

23. Guigues, J. and Duquenne, V.: Familles minimales d'implications informatives résultant d'un tableau de données binaires. Mathématiques et Sciences Humaines, 95:5–18, 1986.

24. Hammer, P. L. and Kogan, A.: Horn functions and their DNFs. Inf. Process. Lett., 44:23–29, 1992.

25. Hammer, P. L. and Kogan, A.: Optimal compression of propositional Horn knowledge bases: complexity and approximation. Artificial Intelligence, 46:131–145, 1993.

26. Harary, F. and Schwenk, A.: Trees with hamiltonian squares. Mathematika, 18:138–140, 1971.

27. Istrate, G.: The phase transition in random Horn satisfiability and its algorithmic implications. Random Struct. Algorithms, 20(4):483–506, 2002.

28. Janson, S., Łuczak, T., and Ruciński, A.: Random Graphs. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, 2000.

29. Karp, R. M.: The transitive closure of a random digraph. Random Struct. Algorithms, 1(1):73–94, 1990.

30. Kratzke, T. M. and West, D. B.: The total interval number of a graph II: Trees and complexity. SIAM J. Discrete Math., 9(2):339–348, 1996.

31. Langlois, M., Mubayi, D., Sloan, R. H., and Turán, G.: Combinatorial problems for Horn clauses. In Graph Theory, Computational Intelligence and Thought, pages 54–65. Springer LNCS 5420, 2009.

32. Maier, D.: The Theory of Relational Databases. Computer Science Press, 1983.

33. Mohar, B. and Robertson, N.: Disjoint essential cycles. J. Combin. Theory Ser. B, 68(2):324–349, 1996.

34. Mohar, B. and Thomassen, C.: Graphs on surfaces. Johns Hopkins Studies in the Mathematical Sciences. Baltimore, MD, Johns Hopkins University Press, 2001.

35. Moore, C., Istrate, G., Demopoulos, D. D., and Vardi, M. Y.: A continuous-discontinuous second-order transition in the satisfiability of random Horn-SAT formulas. Random Struct. Algorithms, 31(2):173–185, 2007.

36. Pelsmajer, M. J., Schaefer, M., and Stasi, D.: Strong Hanani–Tutte on the projective plane. SIAM J. Discrete Math., 23(3):1317–1323, 2009.

37. Pelsmajer, M. J., Schaefer, M., and Štefankovič, D.: Removing even crossings. J. Combin. Theory Ser. B, 97(4):489–500, 2007.

38. Pelsmajer, M. J., Schaefer, M., and Štefankovič, D.: Removing even crossings on surfaces. In European Conference on Combinatorics, Graph Theory and Applications (EuroComb '07), eds, J. N. Alberto Márquez and O. Serra, Electronic Notes in Discrete Mathematics, pages 85–90. Elsevier, August 2007.

39. Raychaudhuri, A.: The total interval number of a tree and the Hamiltonian completion number of its line graph. Inf. Process. Lett., 56:299–306, 1995.

40. Sloan, R. H., Stasi, D., and Turán, G.: Hydras: Directed hypergraphs and Horn formulas. To appear in 38th International Workshop on Graph Theoretic Concepts in Computer Science (WG 2012).

41. Sloan, R. H., Stasi, D., and Turán, G.: Random Horn formulas and propagation connectivity in directed hypergraphs. Submitted.

42. Sloan, R. H., Stasi, D., and Turán, G.: Hydra formulas and directed hypergraphs: A preliminary report. In ISAIM, 2012.

43. Tutte, W. T.: Toward a theory of crossing numbers. J. Combinatorial Theory, 8:45–53, 1970.

44. van der Holst, H.: Algebraic characterizations of outerplanar and planar graphs. European J. Combin., 28(8):2156–2166, 2007.

45. West, D. B.: Introduction to Graph Theory. Prentice Hall, 2 edition, 2001.

VITA

Name

Despina Stasi

Education

B.S., *High Honors*, Applied Mathematics, Illinois Institute of Technology, Chicago, Illinois, 2003
B.S., *High Honors*, Computer Science, Illinois Institute of Technology, Chicago, Illinois, 2003
M.S., Mathematics, University of Illinois at Chicago, Chicago, Illinois, 2006
Ph.D., Mathematics, University of Illinois at Chicago, Chicago, Illinois, 2012

Research and Teaching Experience

Visiting Lecturer, Dept. Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, 2011 – 2012

- Multivariable Calculus, Intermediate Algebra.

Research Assistant, Dept. Computer Science, University of Illinois at Chicago, 2010–2011

- Theoretical Foundations of Evolving Knowledge Bases.

Teaching Assistant, Dept. Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, 2004 – 2010

- Introduction to Computer Science, Introduction to Data Structures, Intermediate Algebra, Finite Mathematics, Finite Mathematics for Business Students, Calculus for Business Students, Single Variable Calculus, Multivariable Calculus.

Service, Scholarships, Honors

University of Illinois at Chicago

- Peter L. Hammer student award at the *International Symposium on Artificial Intelligence and Mathematics*, Fort Lauderdale, Florida, January 9-12, 2012.
- Participated in the organization of the *Workshop in Graph Theory and Combinatorics in Memory of Uri Peled,* University of Illinois at Chicago, February 21-22, 2010.
- Graduate Student Teaching Award (Honorable Mention Spring 2009).

Illinois Institute of Technology

- CASP Scholar (Cyprus-America Scholarship Program).
- Illinois Institute of Technology International Student Scholarship.
- Dean's List, Fall 2000 – Fall 2003.

Publications

Sloan, R. H., Stasi, D., and Turán, G.: Hydras: Directed hypergraphs and Horn formulas. To appear in 38th International Workshop on Graph Theoretic Concepts in Computer Science (WG 2012).

Sloan, R. H., Stasi, D., and Turán, G.: Hydra formulas and directed hypergraphs: A preliminary report. In ISAIM, 2012.

Sloan, R. H., Stasi, D., and Turán, G.: Random Horn formulas and propagation connectivity in directed hypergraphs. Submitted.

Pelsmajer, M. J., Schaefer, M., and Stasi, D.: Strong Hanani-Tutte on the projective plane. SIAM J. Discrete Math., 23(3):1317–1323, 2009.

Selected Talks

Hydras: directed hypergraphs and Horn formulas, 38th International Workshop on Graph Theoretic Concepts in Computer Science (WG 2012), Jerusalem, Israel, June 2012.

Hydra formulas and directed hypergraphs: A preliminary report, Special Session on Boolean and pseudo-Boolean Functions, International Symposium on Artificial Intelligence and Mathematics (ISAIM 2012), Fort Lauderdale, Florida, January 2012.

Random Horn formulas and propagation connectivity in directed hypergraphs, Midwest Theory Day, University of Chicago, December 2010.

Propagation connectivity in random directed hypergraphs and its connection to Horn formulas, Discrete Applied Math Seminar, Illinois Institute of Technology, November 2010.

A criterion for embeddability in the projective plane, 47th Midwest Graph Theory Conference (MIGHTY XLVII). Illinois Institute of Technology, November 2008.