Hydra formulas and directed hypergraphs: A preliminary report*

Robert H. Sloan and **Despina Stasi**
University of Illinois at Chicago

György Turán
University of Illinois at Chicago,
University of Szeged

Abstract

We consider the problem of determining minimal Horn formula size for a subclass of Horn formulas. A *hydra formula* is a Horn formula consisting of size 3 definite Horn clauses, specified by a set of bodies of size 2, and containing clauses formed by these bodies and *all* possible heads. A hydra formula can be specified by the undirected graph formed by the bodies occurring in the formula. Thus minimal formula size for hydras can be considered as a graph parameter, the *hydra number*. We discuss how the hydra number relates to other quantities such as the path cover number of the line graph, characterize trees with low hydra number and give bounds for the hydra number of complete binary trees. We also discuss a related optimization problem and formulate several open problems.

1 Introduction

Horn minimization is the problem of finding a shortest possible Horn formula equivalent to a given formula. There are approximation algorithms, computational hardness and inapproximability results for this problem (Hammer and Kogan 1993; Bhattacharya et al. 2010; Boros and Gruber 2012). Estimating the size of a minimal formula is not well understood even in rather simple cases. The problem can also be viewed as a problem for directed hypergraphs. Special cases correspond to the well studied transitive reduction and minimum equivalent digraph problems for directed graphs.

Definition 1.1. *A definite 3-Horn formula φ is a* hydra[1] *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[2].

We consider the Horn minimization problem for hydras. Besides being a natural subproblem of Horn minimization, this 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 (Maier 1983; Guigues and Duquenne 1986; Angluin, Frazier, and Pitt 1992; Arias and Balcázar 2011). 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 (Bhattacharya et al. 2010). Hydras are a natural test case for this approach.

A hydra φ is determined by the undirected graph G formed by the bodies in φ, and thus the minimal formula size of a hydra can be viewed as a graph parameter $h(G)$, the *hydra number* of G. As the problem discussed in this paper has a simple and natural formulation as a combinatorial problem involving undirected graphs and directed hypergraphs, we discuss the connection between the logic and combinatorics formulations in Section 2, and for the rest of the paper we use combinatorial terminology.

We 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 3 we give some sufficient and necessary conditions for single-headedness. The hydra number is related to the path cover number of the line graph (Theorem 4.1, Example 4.2). It is shown in Theorem 5.1 that single-headed trees must be stars and that trees with hydra number $|E(G)| + 1$ must be caterpillars. In Theorem 6.1 we show that the hydra number of complete binary trees is between $\frac{13}{12}\,|E\,(G)|$ and $\frac{8}{7}|E\,(G)\,|$.

In Section 7 we consider the related problem of finding minimal definite 3-Horn formulas for which every k-tuple of variables implies all other variables, and we give almost matching lower and upper bounds. We conclude the paper by mentioning several open problems.

2 Preliminaries

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

*This material is based upon work supported by the National Science Foundation under Grant No. CCF-0916708.

[1]In Greek mythology the Lernaean Hydra is a beast possessing many heads.

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

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. The implicate C is a prime implicate 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. The *closure* $cl_\varphi(S)$ of a set of variables S with respect to φ is the set of variables marked by forward chaining started from S. A set of variables is *good* if its closure is the set of all variables.

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.

Proposition 2.1. *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 (Hammer and Kogan 1992). 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 $x, y \to head(C)$ occurs in φ and so C is not prime. □

A definite Horn formula may also be viewed as a *directed hypergraph*, where vertices are the variables occurring in the formula, and there is a hyperedge corresponding to each clause in the formula. The body (or tail) of the edge is the body of the clause and the head of the edge is the head of the clause. Definite 3-Horn formulas, in particular, are represented by *directed 3-hypergraphs* with hyperedges of the form $u, v \to w$.

Forward chaining, then, defines a notion of reachability in directed hypergraphs: a vertex v can be reached from a set of vertices S iff forward chaining started by marking vertices in S eventually marks vertex v. The set of vertices reachable from S in a hypergraph H is called the closure of S, and it is denoted by $cl_H(S)$. The set S is good if its closure is the whole vertex set of H.

Definition 2.2. *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 2.3. *The* hydra number $h(G)$ *of an undirected graph* $G = (V, E)$ *is*

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

Proposition 2.1 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. For the rest of the paper we are going to use the latter terminology.

Remark 2.4. For the rest of the paper 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.

3 The hydra number of graphs

In this section we note some simple properties of the hydra number.

Proposition 3.1. *For every graph* $G = (V, E)$ *with at least three vertices*

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

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 hyperedges 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 hyperedge. □

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

Definition 3.2. *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 hyperedge 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

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

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 (1) by adding the hyperedge $v_i, v_j \to v_{i+1}$, where $i + 1$ is meant $\bmod\ m$. Thus we obtain the following.

Proposition 3.3. *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 3.1. Indeed, every edge must occur as the body of at least two hyperedges as otherwise forward chaining cannot mark any further edges.

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) hyperedge of H.

Remark 3.4. *Assume that the directed hypergraph* $H = (V, F)$ *represents the graph* $G = (V, E)$ *and* $|V| \ge 4$. *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 3.3.

Proposition 3.5. *Let G be a connected graph and let G' be a connected spanning subgraph of G. 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 connected spanning subgraph of G, for every edge $(u, v) \in E(G) \setminus E(G')$ there is an edge $(v, w) \in E(G')$. The directed hypergraph H representing G obtained from H' by adding the hyperedge $u, v \to w$ to H' for each edge $(u, v) \in E(G) \setminus E(G')$ 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 3.6. *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 $|E(G)|$ hyperedges. Construct $\hat{H}$ from H by adding hyperedges $u, v \to w$ 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 $\hat{H}$), hyperedge $u, v \to w$ ensures all pairs in $E(G)$ can reach in $\hat{H}$ the new variable w as well. On the other hand, hyperedge $v, w \to z$ ensures that the new pairs (u, v) and (v, w) can reach all other variables. Finally, there are $|E(\hat{G})|$ hyperedges in H. □

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

Proposition 3.7. *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 path of length 2. If at least one of G_1, G_2 contains 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, v and w be the three vertices forming the path of length 2 between G_1 and G_2, where $u \in G_1$, $v \notin G_1 \cup G_2$, $w \in G_2$ and, without loss of generality, let G_1 contain at least two vertices. There is exactly one hyperedge of the form $u, v \to z$. Either z is a vertex in G_1, or it is w. In the first case, forward chaining started from z and one of its neighbors cannot mark w. In the second case consider the unique hyperedge of the form $v, w \to t$. Here t can either be a vertex in G_2 or it is u. In both cases, if forward chaining is started from v, w, no vertex in G_1 other than u can be marked. □

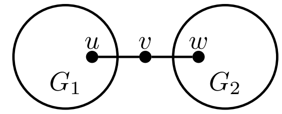

Figure 1: G_1, G_2 connected by a path of length 2.

The following is a simple bound relating the hydra number of a disconnected graph to the hydra numbers of its components.

Proposition 3.8. *Let G have k connected components $G_1, G_2, \ldots, G_k$ for $k \geq 2$. Then*

$$h(G) \leq \sum_{i=1}^{k} h(G_i) + 2k.$$

Proof. This follows directly by considering optimal directed hypergraph realizations of the connected components, and cyclically adding hyperedges

$$(x_1, y_1 \to x_2), (x_1, y_1 \to y_2), (x_2, y_2 \to x_3),$$

$$(x_2, y_2 \to y_3), \ldots, (x_k, y_k \to x_1), (x_k, y_k \to y_1),$$

where x_i, y_i are vertices in G_i for $i = 1, \ldots, k$. □

Equality holds when G is a matching. In the final version of the paper we will discuss sharper versions under certain assumptions on the components.

4 Line graphs

In this section we consider graph parameters 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 set of vertex-disjoint paths such that every vertex $v \in V$ is in exactly one path.

In Proposition 3.3 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 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 gives a general upper bound for the hydra number.

Theorem 4.1. *Let G be a connected graph and G' be a connected spanning subgraph of G. 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 3.5 it is sufficient to prove the bounds for G'.

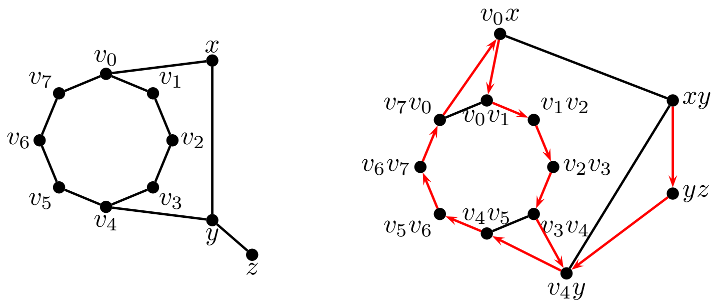

Figure 2: The graph (left) and its line graph (right) discussed in Example 4.2.

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 hyperedge $u, v \to w$ 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 $l_i - 1$ hyperedges: for each directed edge $(e, f) \in \vec{P_i}$, where $e = (u, v)$ and $f = (v, w)$, add a hyperedge $u, v \to w$ to H.

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

$$u_k, v_k \to x_1, u_k, v_k \to y_1, \text{ and}$$
$$u_i, v_i \to x_{i+1}, u_i, v_i \to y_{i+1}, \text{ for } 1 \leq i \leq k-1.$$

□

□

As the following example shows, the condition of Theorem 4.1(i) is sufficient but not necessary for a graph G to be single-headed.

Example 4.2. Graph G in Figure 2 with 11 vertices and 12 edges is single-headed and has no connected spanning subgraph with a hamiltonian line graph.

For a simple proof of single-headedness, consider the graph $\bar{G}$ be obtained from G by deleting edges (x, y) and (y, z). Then $L(\bar{G})$ is hamiltonian and so $\bar{G}$ is single-headed. But G is obtained from $L(\bar{G})$ using the operation of Proposition 3.6, and so G is single-headed as well. The directed 3-hypergraph H with 12 hyperedges representing G can be viewed in Figure 2 as directed edges in the line graph $L(G)$. For every directed edge $e \to f$ in the line graph $L(G)$, where $e = (u, v)$ and $f = (v, w)$ for some vertices u, v, w in G, the hyperedge $u, v \to w$ belongs to H, and H contains no other hyperedges.

By inspection, $L(G)$ contains no hamiltonian cycle. Now consider an arbitrary connected spanning subgraph G' of G. There are three independent paths connecting v_0 and v_4 in G and, by virtue of its connectivity, G' can exclude exactly one edge in at most two of these three paths, and must include all other edges of G. Then, with one exception, G' consists of two distinct components (one with exactly one vertex) connected by a path of length 2. Thus, by Proposition 3.7, G' is not single-headed and so $L(G')$ is non-hamiltonian. In the remaining case, when G' is obtained by deleting edge (v_0, x), its line graph is easily seen to be non-hamiltonian as well.

The final version of the paper will contain a more complicated example of a single-headed graph for which the line graphs of its connected spanning subgraphs do not even contain a hamiltonian path.

5 Trees with low hydra number

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

A *star* is a tree that contains no length-3 path. A *caterpillar* is tree for which deleting all vertices of degree one and their incident edges from the tree gives a path graph. We call this path the spine of T, and note that it is unique. Another useful characterization of caterpillars is that they do not contain the subgraph in Figure 3 (Harary and Schwenk 1971) (see also (West 2001, p.88)).

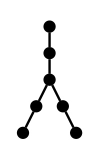

Figure 3: The forbidden subgraph for caterpillars.

Caterpillars have been instrumental in (Raychaudhuri 1995), 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 (Agnetis et al. 2001) for the same problem. For general graphs the problem is NP-hard. Furthermore, (Bertossi 1981) 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 5.1. *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.

The proof of Theorem 5.1 relies on lower bounds given for $h(T)$ in Lemmas 5.2 and 5.3. We first show that a tree that is not a star cannot be single-headed.

Lemma 5.2. *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, say $(s,t),(t,u),(u,v)$. Let T_s and T_t be the trees rooted at s and t respectively that we would get by removing (s,t) from T. Let hypergraph H represent T. If the body t,u has two heads, then we are done. Otherwise, assume without loss of generality that the head of the hyperedge with body t,u is a neighbor of u in T (possibly but not necessarily v).

Then for s to be reachable from (t,u) in H, some edge of T that is within T_t must be a body with a head in T_s. That will be hyperedge with a head that is not a neighbor of either of its body vertices in T. Thus its body cannot be single-headed so $h(T) > |E(T)|$. □

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

Lemma 5.3. *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 3. Let us call the central vertex of that forbidden subgraph u.

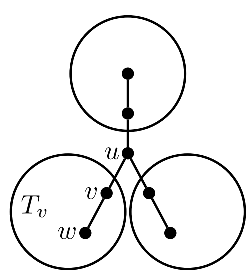

Figure 4: Part of the non-caterpillar tree T from the proof of Lemma 5.3.

Assume for contradiction that H is a hypergraph with $|E(T)|+1$ hyperedges 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 4.) Body u,v must have a head that is a neighbor 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 5.1. We need to prove the upper bounds. The single-headedness of stars is easily seen directly, or follows from Theorem 4.1(i). For T a caterpillar, the upper bound follows from Theorem 4.1(ii) as the line graph of a caterpillar contains a hamiltonian path. □

6 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 6.1. *For $d \geq 3$ it holds that*

$$h\left(B_d\right) \geq \frac{13}{12}\left|E\left(B_d\right)\right|, \textit{ and}$$

$$h\left(B_d\right) \leq \begin{cases} \frac{8}{7}\left|E\left(B_d\right)\right| & \textit{for } d \equiv 0 \bmod 3, \\ \frac{8}{7}\left|E\left(B_d\right)\right| + \frac{5}{7} & \textit{for } d \equiv 1 \bmod 3, \\ \frac{8}{7}\left|E\left(B_d\right)\right| + \frac{1}{7} & \textit{for } d \equiv 2 \bmod 3. \end{cases}$$

Proof. Let us begin with the upper bound. Note that all the expressions in the statement of the theorem's upper bounds evaluate to integers.

First let $d = 3k$ for some $k > 1$. Define $p(G)$ to be the minimum number of paths in a path cover of $L(G)$. By Theorem 4.1(ii) we need only show that $p(B_{3k})$ is $|E(B_{3k})|/7$. This we do inductively. By inspection, $p(B_3) = 2$. Considering the top three levels of the line graph $L(B_{3k})$ we note the following inductive inequality:

$$p\left(B_{3k}\right) \leq 2 + 8p\left(B_{3k-3}\right), \tag{2}$$

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 5.) Inequality (2) 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

$$p\left(B_{3k+1}\right) \leq 1 + 2p\left(B_{3k}\right),$$

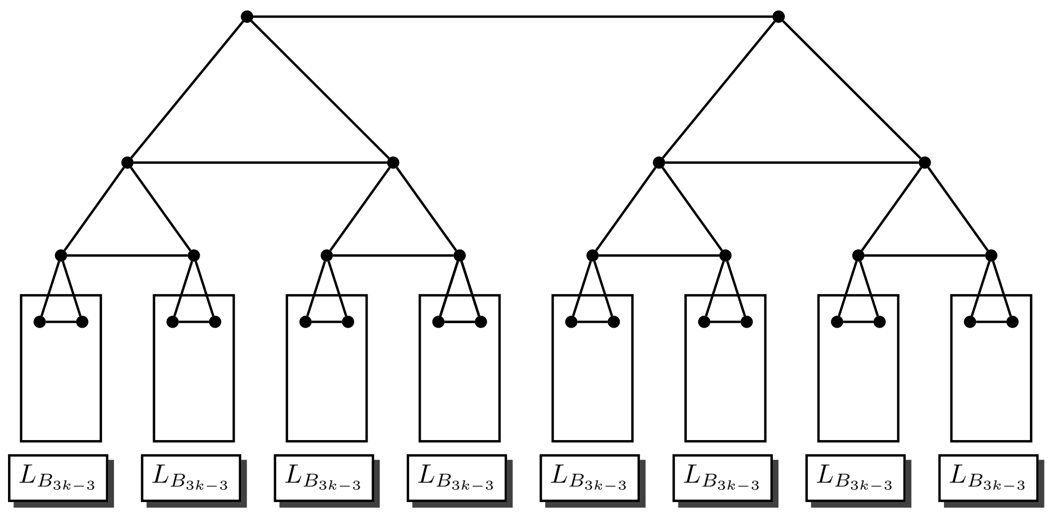

Figure 5: Line graph $L(B_{3k})$ from the proof of Theorem 6.1.

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

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

For the lower bound consider a directed hypergraph H representing B_d. Color the vertices of $L(B_d)$ red (resp., black) if the corresponding edge is single-headed (resp., multi-headed) in H. If $e = (u, v)$ is single-headed then the head w of the unique hyperedge with body u, v is a neighbor of u or v. If w is a neighbor of u (resp., v) then let $f = (u, w)$ (resp., $f = (v, w)$). Direct the edge (e, f) in $L(B_d)$ from e to f. Let $\vec{R}$ be the directed graph formed by the directed edges and their endpoints (some of which may be colored black). Note that red (resp., black) vertices have outdegree one (resp. zero) in $\vec{R}$.

We claim that $\vec{R}$ is acyclic. The only cycles in $L(B_d)$ are triangles, thus a simple directed cycle in $\vec{R}$ must consist of red vertices, and have length two or three. This corresponds in B_d to two incident single-headed edges, resp., three single-headed edges forming a star. None of these edges participate in a hyperedge with a head that is not involved in the cycle. Thus these edges could only reach vertices of B_d involved in the cycle.

Thus $\vec{R}$ has the following structure: it is a directed forest, where every tree in the forest has a black root, all its other vertices are red, and its edges are directed towards the root.

Consider a triangle at the bottom of $L(B_d)$ (please consult Figure 5). By the structure of $\vec{R}$ at least one of the two bottom vertices of this triangle is either black or a red leaf in $\vec{R}$.

Let b denote the number of black vertices in $L(B_d)$ and r denote the number of red leaves on the last level of $L(B_d)$. We have $2^{d-1} = \lceil |E(B_d)|/4 \rceil$ triangles in the bottom level of $L(B_d)$. Therefore, $\max\{b, r/2\} \geq |E(B_d)|/12$.

To conclude the lower bound, we show that both b and $r/2$ are lower bounds on the the difference between the number of edges in H and $|E(B_d)|$. For b, this is obvious.

Consider a red leaf $e = (v, w)$ at the bottom of $L(B_d)$ and assume without loss of generality that v is a leaf of B_d. Then, as e is the only edge incident to v, a hyperedge with a single-headed body pointing to v would make e into a non-singleton in $\vec{R}$. Thus the hyperedge with head v must have a multi-headed body. Hence there are at least r hyperedges with multi-headed bodies. If there are s multi-headed bodies then $r - s \geq s$, so $r - s \geq r/2$. ☐

7 Minimal definite 3-Horn formulas with all k-tuples implying all variables

In this section we consider a problem related to hydra minimization. Given variables $x_1, \ldots, x_n$ and a number $k\,(2 \leq k \leq n-1)$, find a shortest definite 3-Horn formula φ such that for every k-element subset S of variables $cl_\varphi(S) = \{x_1, \ldots, x_n\}$, i.e., every k-element subset of variables is good for φ. We denote the size of a shortest such formula by $f(n, k)$.

The case $k = 2$ is just hydra minimization for complete graphs and it follows from Proposition 3.3 that the shortest formula has size $\binom{n}{2}$. This was already noted in (Langlois et al. 2009) along with the stronger result that for some minimal formula every prime implicate has a resolution derivation where every intermediate clause has size 3 as well.

We use Turán's theorem from extremal graph theory (see, e.g. (West 2001)). 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 7.1. *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 φ is a definite 3-Horn formula with all k-tuples good. Then every k-element set S of variables must contain at least one pair of vertices forming a body in φ, otherwise forward chaining started from S cannot mark any variables. Thus the undirected graph formed by the bodies in φ contains no empty subgraph on k vertices, and the lower bound follows by Turán's theorem.

For the upper bound we construct a formula 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 clauses as in (1) (with the exception of the last edge closing the cycle). For every other edge (u, v), introduce a clause $u, v \rightarrow w$ 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* clauses 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. □

8 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 n-vertex graphs, and, in particular, among n-vertex trees? Can the line graphs of single-headed graphs have arbitrarily high path cover numbers? Can the path cover number of the line graph be used to get a lower bound for the hydra number?

Acknowledgement We would like to thank Daniel Uber for useful discussions.

References

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

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

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

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

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

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

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

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

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

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

Langlois, M.; Mubayi, D.; Sloan, R. H.; and Turán, G. 2009. Combinatorial problems for Horn clauses. In Lipshteyn, M.; Levit, V. E.; and Mcconnell, R. M., eds., *Graph Theory, Computational Intelligence and Thought*. Berlin, Heidelberg: Springer-Verlag. 54–65.

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

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

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