Hydras: Directed Hypergraphs and Horn Formulas*

Robert H. Sloan[1], Despina Stasi[1], and György Turán[1,2]

[1] University of Illinois at Chicago
[2] Hungarian Academy of Sciences and University of Szeged, Research Group on Artificial Intelligence

Abstract. We consider a graph parameter, the *hydra number*, arising from an optimization problem for Horn formulas in propositional logic. The hydra number of a graph $G = (V, E)$ is the minimal number of hyperarcs of the form $u, v \to w$ required in a directed hypergraph $H = (V, F)$, such that for every pair (u, v), the set of vertices reachable in H from $\{u, v\}$ is the entire vertex set V if $(u, v) \in E$, and it is $\{u, v\}$ otherwise. Here reachability is defined by the standard forward chaining or marking algorithm.

Various bounds are given for the hydra number. We show that the hydra number of a graph can be upper bounded by the number of edges plus the path cover number of its line graph, and this is a sharp bound for some graphs. On the other hand, we construct graphs with hydra number equal to the number of edges, but having arbitrarily large path cover number. Furthermore we characterize trees with low hydra number, give bounds for the hydra number of complete binary trees, discuss a related optimization problem and formulate several open problems.

1 Introduction

We consider a problem concerning the minimal number of hyperarcs in directed hypergraphs with prescribed reachability properties. In this paper, 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 $(u, v) \in E$, 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

* An earlier version of this paper entitled "Hydra formulas and directed hypergraphs: A preliminary report" appears in the online proceedings of the International Symposium on Artificial Intelligence and Mathematics 2012.

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*[3] of G, denoted by $h(G)$.

The problem is a combinatorial reformulation of a special case of the *minimization problem for propositional Horn formulas*. 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. There are approximation algorithms, computational hardness and inapproximability results for this problem [1–4]. Special cases correspond to the well studied transitive reduction and minimum equivalent digraph problems for directed graphs. Estimating the size of a minimal formula is not well understood even in rather simple cases. 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.

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 [5–8]. 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 [2]. Hydras are a natural test case for this approach.

The paper is organized as follows. Section 2 contains some background, including a discussion of the motivating Horn minimization problem. The rest of the paper presents 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 some necessary conditions for single-headedness. In Section 4 we show that the hydra number is related to the path cover number of the line graph (Theorem 13, Theorem 14). In Section 5 it is shown that single-headed trees are precisely the stars and that trees with hydra number $|E(G)| + 1$ are precisely the non-star caterpillars (Theorem 16). In Section 6 we show that the hydra number of a complete binary tree is between $\frac{13}{12}\,|E\,(G)|$ and $\lceil \frac{8}{7}|E\,(G)\,| \rceil$ (Theorem 19).

In Section 7 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 the paper by mentioning several open problems.

Due to space limitations, the proof of Theorem 19 is omitted. Further results on hydra numbers will appear in [9].

---

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

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. 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.

Definition 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[4].

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. *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 [10]. 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 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 paper.

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

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

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

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

Proposition 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 5. 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.

For the remainder of the paper we use hypergraph terminology.

3 The Hydra Number of Graphs

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

Proposition 6. *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. Graphs satisfying the lower bound are of particular interest as they represent 'most compressible' hydras.

Definition 7. *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

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

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 hyperarc $v_i, v_j \to v_{i+1}$, where $i + 1$ is meant modulo m. Thus we obtain the following.

Proposition 8. *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 6. 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 9. 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 8.

Proposition 10. *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 hyperarc $u, v \to w$ 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 11. *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 12. *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, 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 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 8 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 connected subgraph gives a general upper bound for the hydra number.

Theorem 13. *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 10 it is sufficient to prove the bounds for G'.

For 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 $f = (v, w)$.

For 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 $1 \le i \le k-1$. □

The condition of Theorem 13(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 14. *There is a family of single-headed graphs G_k with $\Theta(k)$ edges such that for every spanning, connected subgraph $G' \subseteq G_k$, $L(G')$ has path cover number $\Theta(k)$.*

Proof. Consider the sequence of graphs $\{G_k : k \ge 1\}$ constructed from an $8k$-cycle, with vertices $v_0, \ldots, v_{8k-1}$, and pendant edges $x_i v_{4i}$ and $y_i v_{4k+4i}$ for $0 \le i \le k-1$. Add a vertex z_i and the edges (x_i, y_i), (y_i, z_i), for each i, $0 \le i \le k-1$, corresponding to the construction in Proposition 11.

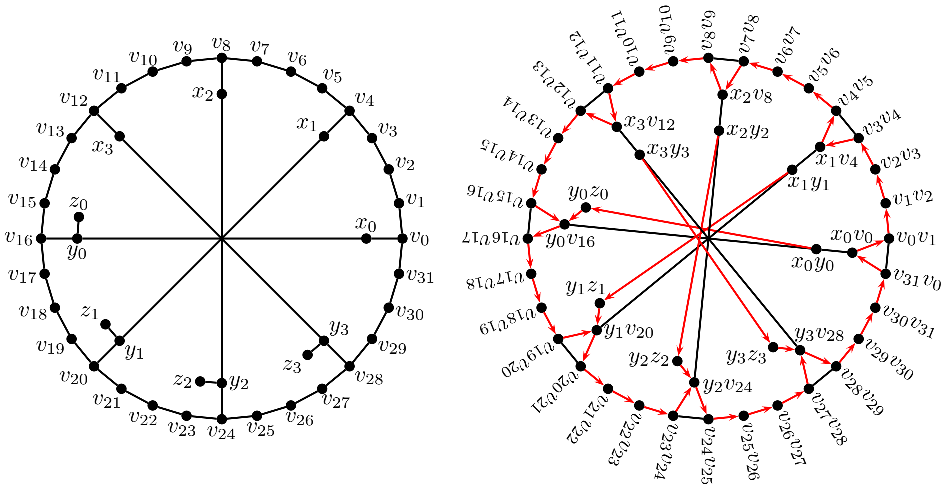

Fig. 1. Single-headed graph G_k for $k = 4$ from Theorem 14 (left) and its line graph (right).

By Proposition 11, G_k is single-headed, since a cycle with attached pendant edges has a hamiltonian line graph. We will show that for an arbitrary connected spanning subgraph $G' \subseteq G_k$ the path cover number of $L(G')$ is at least $k/4$.

Define D_i to be the set of vertices in the ith 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 $S = \{P_j : 1 \le j \le s\}$ of the vertices of $L(G')$.

Lemma 15. *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 $e \in D_i \backslash {D_i}'$. Since G' is both spanning and connected, 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 15 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 connected. 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.

5 Trees with Low Hydra Number

In this section 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 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 Fig.2 [11] (see also [12, p.88]).

Caterpillars have been instrumental in [13], 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 [14] for the same problem. For general graphs the problem is NP-hard. Furthermore, [15] proves that finding a hamiltonian path is NP-complete even for line graphs.

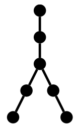

Fig. 2. The forbidden subgraph for caterpillars.

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 16. *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 17. *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 12. □

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

Lemma 18. *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 Fig.2. Let us call the central vertex of that forbidden subgraph u.

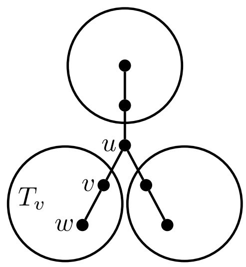

Fig. 3. Part of the non-caterpillar tree T from the proof of Lemma 18.

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 Fig.3.) 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 16). We need to prove the upper bounds. The single-headedness of stars is easily seen directly, or follows from Theorem 13(i). For T a caterpillar, the upper bound follows from Theorem 13(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 19. *For $d \geq 3$ it holds that*

$$\frac{13}{12}\left|E\left(B_d\right)\right| \leq h\left(B_d\right) \leq \left\lceil \frac{8}{7}\left|E\left(B_d\right)\right| \right\rceil .$$

Proof omitted due to space constraints.

7 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 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. [12]). 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 20. *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 (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. □

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 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?

Acknowledgement This material is based upon work supported by the National Science Foundation under Grant No. CCF-0916708. We would like to thank Daniel Uber for useful discussions.

References

1. Ausiello, G., D'Atri, A., Saccà, D.: Minimal representation of directed hypergraphs. SIAM J. Comput. **15**(2) (1986) 418–431
2. Bhattacharya, A., DasGupta, B., Mubayi, D., Turán, G.: On approximate Horn formula minimization. In Abramsky, S., Gavoille, C., Kirchner, C., Meyer auf der Heide, F., Spirakis, P.G., eds.: ICALP (1). Volume 6198 of Lecture Notes in Computer Science., Springer (2010) 438–450
3. Boros, E., Gruber, A.: Hardness results for approximate pure Horn CNF formulae minimization. In: International Symposium on AI and Mathematics (ISAIM). (2012)
4. Hammer, P.L., Kogan, A.: Optimal compression of propositional Horn knowledge bases: complexity and approximation. Artificial Intelligence **46** (1993) 131–145
5. Maier, D.: The Theory of Relational Databases. Computer Science Press (1983)
6. Guigues, J., Duquenne, V.: Familles minimales d'implications informatives résultant d'un tableau de données binaires. Mathématiques et Sciences Humaines **95** (1986) 5–18
7. Angluin, D., Frazier, M., Pitt, L.: Learning conjunctions of Horn clauses. Machine Learning **9** (1992) 147–164
8. Arias, M., Balcázar, J.L.: Construction and learnability of canonical Horn formulas. Machine Learning **85** (2011) 273–297
9. Stasi, D.: Combinatorial Problems in Graph Drawing and Knowledge Representation. PhD thesis, University of Illinois at Chicago (Forthcoming Summer 2012)
10. Hammer, P.L., Kogan, A.: Horn functions and their DNFs. Inf. Process. Lett. **44** (1992) 23–29
11. Harary, F., Schwenk, A.: Trees with hamiltonian squares. Mathematika **18** (1971) 138–140
12. West, D.B.: Introduction to Graph Theory. 2 edn. Prentice Hall (2001)
13. Raychaudhuri, A.: The total interval number of a tree and the Hamiltonian completion number of its line graph. Inf. Process. Lett. **56** (1995) 299–306
14. Agnetis, A., Detti, P., Meloni, C., Pacciarelli, D.: A linear algorithm for the Hamiltonian completion number of the line graph of a tree. Inf. Process. Lett. **79** (2001) 17–24
15. Bertossi, A.A.: The Edge Hamiltonian Path Problem is NP-complete. Inf. Process. Lett. **13**(4/5) (1981) 157–159