Comenius University
Faculty of Mathematics, Physics and Informatics
Department of Applied Informatics
Bratislava, Slovakia

Covering edges of a hypergraph: complexity and applications

Master's Thesis

Vladimír Repiský
author

Mgr. Tibor Hegedüs
advisor

Bratislava May 2007

By this I declare that I wrote this thesis by myself, only with the help of the referenced literature, under the careful supervision of my thesis advisor.

Bratislava, May 2007 Vladimír Repiský

Abstract

Title: Covering edges of a hypergraph: complexity and applications
Author: Vladimír Repiský
Advisor: Mgr. Tibor Hegedüs

This thesis gives an overview of the edge covering problems in hypergraphs. We explore the complexity properties of covering the edges of a hypergraph by subhypergraphs of different types. We consider the problem of covering by clique hypergraphs, split hypergraph, and threshold hypergraphs. In total we prove NP-completness of 6 problems. We especially focus on the problem of covering by threshold hypergraphs, which has applications in the theory of machine learning. We give a reduction from the clique covering problem, so proving NP-hardness of this problem would imply NP-hardness of the problem of covering by threshold hypergraphs. Moreover, we propose a generalization of the concept of Dilworth number of a graph to hypergraphs. We give a polynomial algorithm for the computation of this number. We prove that the Dilworth number gives an upper bound for some important parameters like the diameter and domination number of a hypergraph.

Keywords: Threshold hypergraph, Hypergraph covering, NP-completeness

Acknowledgments

I am very grateful to Tibor Hegedüs, my supervisor, for his time, advices and support. His insights have been always very helpful.

I thank Ján Šefránek, my teacher, whose seminar in Artificial Intelligence inspired me to do research in this field.

This thesis is dedicated to my parents Vladimír Repiský and Katarína Repiská, I thank them for their love and support over the years.

I would also thank my friends, Michal Malý and Pavol Slamka, for our valuable discusions about this thesis.

Finally, I want to thank Andreas Brandstaedt, Midge Cozzens, Uri Peled, Jim Orlin, Van Bang Le and Mark D. Halsey for sending me their articles, to which I had not other access.

Contents

Chapter 1

Introduction

1.1 Motivation

The complexity properties of the covering problems in graphs have been extensively studied by many authors (see [GJ79] for a survey). In recent years, Aizenstein et al. [AHHP98] has proved that efficient learnability in one of the on-line learning models is related to the representation problem, e.g. to decide if a given Boolean formula in DNF has a representation of some type, for example as a union of k linearly separable functions. NP-hardness of representation problem for some concept class implies negative results for its efficient learnability in this model. Such results are given for many concept classes, however for an important concept class - the class of unions of two linearly separable functions, the status of the representation problem is unknown. The covering problem by threshold hypergraphs can be easily transformed to the representation problem of unions of two linearly separable functions. This thesis focuses on the complexity properties of covering problems in hypergraphs. These problems were not explored elsewhere, and their applications to the learning theory are a good motivation for a deeper examination of these problems.

1.2 Summary of thesis contributions

We prove NP-completness of the following covering problems for r-uniform hypergraphs:

- r-HYPERGRAPH CLIQUE PARTITION NUMBER
- r-HYPERGRAPH STABILITY NUMBER
- HYPERGRAPH SPLIT DIMENSION

These problems are shown to be NP-complete even in their fixed-parameter versions (for every fixed $k \geq 2$ and $r \geq 3$). Other NP-complete problems are mentioned as generalizations of their counterparts for graphs.

In the Chapter 6 we propose a concept of the Dilworth number of a hypergraph, which is a generalization of the Dilworth number for graphs. Similarly as the Dilworth number of graph, the Dilworth number of hypergraph has many nice properties. The hypergraphs with the Dilworth number 1 are exactly threshold hypergraphs of the type T_3. We give a polynomial algorithm for computation of this number. We prove that the Dilworth number gives an upper bound for some important parameters like diameter and domination number of a hypergraph.

In the Chapter 7 and 8 we present two ideas which could possibly lead to a proof of NP-hardness of recognizing hypergraphs with threshold dimension two.

We give a reduction of the COVERING BY TWO r-UNIFORM CLIQUES problem to the THRESHOLD DIMENSION TWO problem. At present, the status of the COVERING BY TWO r-UNIFORM CLIQUES problem is unknown. In case it turns out to be NP-complete, our reduction will imply NP-hardness of the THRESHOLD DIMENSION TWO problem.

We also propose a special 3-uniform hypergraph, which can be used in the proof. It exhibits similar properties as a special subgraph used by Yannakakis in his proof of NP-completness of THRESHOLD DIMENSION THREE problem for graphs [Yan82]. We give a computerized proof of the properties for this hypergraph.

Chapter 2

Terminology

We present here some basic terminology used in the thesis.

Graphs: A *graph* G is a pair (V, E) of *vertices* V and *edges* E. E is a set of unordered pairs of distinct vertices, called edges. Sometimes we denote the *vertex set* of G by $V(G)$ and the *edge set* by $E(G)$. A *clique* (or a *complete set*) V' is a graph with all possible edges, e.g. $\forall x, y \in V'$, $x \neq y$, $\{x, y\} \in E$. We also use the term clique for a subset V' of V if V' induces a clique in V. By *induced graph* by a subset V' of V we mean a graph $G(V', E')$ with the edge set $E' = \{e | e \in E \wedge \forall x \in e, x \in V'\}$. A *complementary graph* of G is a graph $G = (V, E')$ with edges $E = \{\{x, y\} | x, y \in V, \{x, y\} \notin E\}$. A *stable set* (or an *independent set*) is a graph with no edge. A stable set is also a subset V' of V, which induces a stable graph in G.

A *coloring* c of G by k colors is a function $c : V \rightarrow \{1, 2, \ldots, k\}$ which assigns a color to every vertex from V, in such a way, that none of the edges is *monochromatic*, e.g $\forall x, y \in V, x \neq y, \{x, y\} \in e, c(x) \neq c(y)$.

The *join* of graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ is the graph $G_1 \oplus G_2 = (V_1 \cup V_2, E_1 \cup E_2 \cup E_{12})$, where $E_{12} = \{\{x, y\} | x \in V_1, y \in V_2\}$.

Hypergraphs: A *hypergraph* is a generalization of a graph. Hypergraph $H(V, E)$ with the vertex set V and edge set E differs from a graph in the property that an edge can connect more than two vertices, that is, an edge is an arbitrary subset of the vertex set V. A *r-uniform hypergraph* $H(V, E)$ is a hypergraph whose all edges are of size r.

A *coloring* of a hypergraph H is defined similarly to coloring of a graph as a function $c : V \to \{1, 2, \ldots, k\}$ which assigns a color to every vertex from V, such that none of the edges is *monochromatic*, e.g $\forall e \in E$, $\exists x, y \in V, x \neq y, x, y \in e, c(x) \neq c(y)$.

The *join* of two r-uniform hypergraphs $H_1 = (V_1, E_1)$ and $H_2 = (V_2, E_2)$ is defined as the hypergraph $H_1 \oplus H_2 = (V_1 \cup V_2, E_1 \cup E_2 \cup E_{12})$, where $E_{12} = \{\{x_1, \ldots, x_n\} | ((\forall i, 1 \leq i \leq n)(x_i \in V_1 \cup V_2)) \wedge ((\exists i, 1 \leq i \leq n) x_i \in V_1) \wedge ((\exists j, 1 \leq j \leq n) x_j \in V_2)\}$.

Partially ordered sets: *Partially ordered set* (or a *poset*) is a structure **P**=(X,P), where X is set, and P is a *reflexive*, *antisymmetric* and *transitive* binary relation on X; i.e., $(x, x) \in P$ for each $x \in X$, if $(x, y) \in P$ and $(y, x) \in P$ than $x = y$ and if $(x, y) \in P$ and $(y, z) \in P$ than $(x, z) \in P$. We call X the *ground set* of the poset X, and we refer to P as a *partial order* on X. We use the notations $x \leq y$ in P, $y \geq x$ in P and $(x, y) \in P$ interchangeably, and we usually omit the reference to partial order P, when the reference to it is clear from the context. A similar structure to poset is a preorder. A *preorder* P is a structure P=(X,P), where X is a set, and P is a reflexive and transitive binary relation on X.

We say that, two elements x and y are *comparable* when $(x, y) \in P$ or $(y, x) \in P$, otherwise they are *incomparable*. We write $x \bowtie y$, when x and y are incomparable. The *incomparability graph* $I(P) = (P, \bowtie)$ of a poset P=(X,P) is an undirected graph, where $x \bowtie y$, when x and y are incomparable in P. A *linear order* L=(X,L) is a partial order, where for every $x, y \in X$ holds $(x, y) \in L$ or $(y, x) \in L$. We define the *dimension* $d_O(P)$ of a partial order P as the minimum number of linear orders whose intersection is P.

Learning theory: A *domain* is a nonempty finite set X. A *concept* is any subset of X. A *concept class* C represents a nonempty set of concepts. The learning task is to identify an unknown concept over X from C using permitted types of queries. We mention here two types of queries: *membership query* and *equivalence query*. Let c be a target concept we want to learn. The membership query gets as input $x \in X$ and outputs true if $x \in c$ or false if $x \notin c$. The equivalence query gets as input a concept c'. If the concept c' is equal to c than it returns true, otherwise it returns a *counter-example*

e.g $x \in (c \setminus c') \cup (c' \setminus c)$. The *representation problem* $REP(F)$ for class F is defined as follows:
Instance: A set $X = \{x_1, x_2, \ldots, x_n\}$ of variables and a Boolean formula g over X.
Question: Is there some $f \in F$ such that f represents the same Boolean function as g?

Chapter 3

Cliques in Hypergraphs

3.1 Clique in a graph

In graph theory a *clique* is a well known concept. By *clique* we denote a subset C of vertices of a graph $G(V, E)$ if for all pairs of vertices $u, v \in C$ the edge $\{u, v\}$ belongs to E. Note that the complement of a clique is stable set and vice versa.

3.2 Clique in a hypergraph

For a r-uniform hypergraph $H(V, E)$ we call *r-uniform clique* a subset C of vertices V if for every subset R of C with size r E contains R.

We define a *clique number* of a hypergraph $H(V, E)$ as the minimum number l of subsets to whose the vertex set V can be partitioned into in such a way, that each of these l subsets induces a clique in H. We denote the clique number of a hypergraph H by $P_c(H)$. Sometimes this number is denoted in literature as the *clique cover number*.

Similarly we define *r-uniform stable set* (or *r-uniform independent set*) as a subset S of vertices V if for every subset R of C with size r, E does not contain R. We will drop from the name the r-uniform part on the places, where it is clear, that we mean r-uniform clique, or r-uniform stable set.

Moreover we define the *stability number* of a hypergraph $H(V, E)$ as the minimum number s of subsets to whose the vertex set V can be partitioned into in such a way, that each of these l subsets induces a stable set in H. We denote the stability number of a hypergraph H by $s(H)$.

3.3 Properties

We use the common symbol $\chi(H)$ for the *chromatic number* of a hypergraph H. At first we prove some relations between the chromatic number, the clique number and the stability number of an r-uniform hypergraph.

Theorem 1. *For every hypergraph H holds*

$$\chi(H) = s(H).$$

Proof. $\chi(H) \geq s(H)$: Suppose, that $V(H)$ can be colored by k colors. Let S_i be the set of vertices colored by color number i. The set S_i is stable, otherwise it would contain an edge e, which is a contradiction to the fact that vertices in the edge e cannot be all colored with the same color.

$\chi(H) \leq s(H)$: Let $V(H)$ can be partitioned to k stable sets. We can color vertices $V(H)$ such that to a vertex from stable set S_i we assign color i. This is a legal coloring, because no edge is monochromatic (vertices in it do not have the same color). □

Theorem 2. *For every hypergraph H holds*

$$s(H) = P_c(H'),$$

where H' is the complementary hypergraph of H.

Proof. Every stable set of H is a clique in the complementary hypergraph of H and vice versa. Therefore the stability number of H and clique number of H' are equal. □

Corollary 3. *For every hypergraph H holds*

$$\chi(H) = s(H) = P_c(H'),$$

where H' is the complementary hypergraph of H.

Denote by $cov(H)$ the minimal number of subsets $s_1, \ldots, s_{cov(H)}$ of $V(H)$, such that every s_i is a clique and for every $v \in V(H)$ there exists j, such that $v \in s_j$. Note that the sets s_i do not need to be disjoint.

Theorem 4. *For every hypergraph H holds*

$$P_c(H) = cov(H).$$

Proof. $P_c(H) \leq cov(H)$. Every subset of a set of vertices which induces a clique in H also induces a clique in H. If we have two sets s_j and s_k, inducing cliques in H, we can transform them into disjoint sets by setting $s'_j = s_j$ and $s'_k = s_k \setminus s_j$. These set induce cliques in H. So, from the covering of H by the sets inducing cliques in H we can construct a covering of H by disjoint sets inducing cliques in H.

$P_c(H) \geq cov(H)$. Every partitioning of $V(H)$ into disjoint sets is also a covering of $V(H)$, where the sets do not need to be disjoint. □

3.4 Complexity properties

In this section we will focus on the complexity properties of problems related to the concept of cliques in hypergraphs. Corollary 3 tells us, that the problem of determination of chromatic number of a hypergraph has the same complexity as determination of the clique and stability number of a hypergraph. We can formally describe the chromatic number problem for hypergraphs as follows:

r-HYPERGRAPH CHROMATIC NUMBER
Instance: A r-hypergraph H(V,E) and a positive integer $k \leq |V|$.
Problem: Is it true that $\chi(H) \leq k$?

Theorem 5 ([Lov73])**.** *The problem r-HYPERGRAPH CHROMATIC NUMBER is NP-complete, even for every fixed $r \geq 3$ and $k \geq 2$.*

As a corollary the problems dealing with stability and cliqueness of a hypergraph are NP-complete:

r-HYPERGRAPH CLIQUE PARTITION NUMBER
Instance: A r-hypergraph H(V,E) and a positive integer $k \leq |V|$.
Problem: Is it true that $P_c(H) \leq k$?

r-HYPERGRAPH STABILITY NUMBER
Instance: A r-hypergraph H(V,E) and a positive integer $k \leq |V|$.
Problem: Is it true that $s(H) \leq k$?

Theorem 6. *The problems r-HYPERGRAPH CLIQUE PARTITION NUMBER and r-HYPERGRAPH STABILITY NUMBER are NP-complete, even for every fixed $r \geq 3$ and $k \geq 2$.*

Proof. This is clear from the fact that r-HYPERGRAPH CHROMATIC NUMBER is NP-complete (see Theorem 5) and the Corollary 3.

□

We list some other problems related to clique partition number. They are all NP-complete, because their restrictions to graphs are NP-complete [GJ79][Kar72].

r-UNIFORM CLIQUE PROBLEM
Instance: r-uniform hypergraph $H(V, E)$, positive integer $k \leq |V|$.
Problem: Does H contain a clique of size k or more?

r-UNIFORM INDEPENDENT SET PROBLEM
Instance: r-uniform hypergraph $H(V, E)$, positive integer $k \leq |V|$.
Problem: Does H contain an independent set of size k or more?

r-UNIFORM VERTEX COVER
Instance: r-uniform hypergraph $H(V, E)$, positive integer $k \leq |V|$.
Problem: Is there a vertex cover of size k or less for H, i.e., a subset $V' \subseteq V$ and $|V'| \leq k$ such that for each edge $\{v_1, \ldots, v_r\} \in E$ at least one of $v_1, \ldots, v_r$ belongs to V'?

3.5 Covering of edges

We define the *clique edge covering number* (or the *clique dimension*) of a hypergraph $H(V, E)$ as a minimal number of subsets $V_1, V_2, \ldots, V_K$ of V such that each V_i induces a clique in H and such that for every edge $e \in E$ there is some V_i such that $e \subseteq V_i$. We denote the clique dimension of a hypergraph H by $D_c(H)$. Consider the following problem:

COVERING BY r-UNIFORM CLIQUES
Instance: r-uniform hypergraph $H(V, E)$, positive integer $k \leq |V|$.
Problem: Is it true that $D_c(H) \leq k$?

Theorem 7. *The problem COVERING BY r-UNIFORM CLIQUES is NP-complete.*

Proof. According to the work of Kou, Stockmeyer and Wong [KSW78] and Orlin [Orl77], the COVERING BY CLIQUES problem for graphs is NP-complete. Therefore this problem is also NP-complete for hypergraphs. □

We are not aware of the complexity of the COVERING BY r-UNIFORM CLIQUES problem for a fixed k. This version of the problem is tractable (solvable in polynomial time) for ordinary graphs as was shown by Gramm et al. [GGHN06].

Chapter 4

Split Hypergraphs

4.1 Introduction

A *split graph* is a graph whose vertex set can be partitioned into a clique and a stable set. Let (K, S) be such a partition, where K is a clique and S is a stable set; we call (K, S) a *split partition* and connote the split graph with this partition as (K, S). Split graphs were first defined by Földes and Hammer [FH77]. Later on, these graphs were independently introduced by Chernyak and Chernyak [CC86] as *polar graphs* . A survey of the split graphs can be found in [Mer03] or [MP95]. We define a *split hypergraph* similarly: as a hypergraph whose vertex set can be partitioned into a clique and a stable set. Split hypergraphs were also defined by Sloan and Turán in [ST97]. They also presented a polynomial algorithm for recognition of split hypergraphs.

4.2 Covering of edges

The *split dimension* (or the *split edge covering number*) of a hypergraph is the least number of split subhypergraphs covering its edges. We denote the split dimension of a hypergraph H by $d_s(H)$. For graphs it was shown by Chernyak and Chernyak [CC91] and later by Peled and Mahadev [MP95], that for every fixed $k \geq 3$, the problem of determining if a graph has split dimension at most k is NP-complete. We modify the proof of Peled and Mahadev to show that the problem of determining if a given r-uniform hypergraph has split dimension at most k is NP-complete, even for every fixed $k \geq 2$ and $r \geq 3$.

Lemma 8. *Let H be a hypergraph and k a positive integer; then $d_s(H) \leq k$ if and only if $V(H)$ can be partitioned into k or fewer cliques and a maximal stable set.*

The word "maximal" in the Lemma is not necessary. We can transfer the vertices to stable set until it becomes maximal. Using this transformation, the rest of H will be still partitioned to k or fewer cliques.

Proof. Let $V(H)$ can be partitioned into k or fewer cliques K_i and a maximal stable set S. Let S_i be a graph induced in H by $K_i \cup S$. Clearly each S_i is a split graph, therefore $d_S(H) \leq k$. Now assume that $E(H)$ can be partitioned to k or fewer split subhypergraphs G_i. Each G_i has some split partition (K_i, S_i). The stable set S is $S = V(H) \setminus \bigcup_i K_i$. We can drop common vertices from the cliques K_i to make them disjoint.

□

We consider the following problem:

HYPERGRAPHS SPLIT DIMENSION
Instance: A r-uniform hypergraph $G(V, E)$ and a positive integer $k \leq |V|$.
Problem: Is it true that $d_S(H) \leq k$?

We define a graph modification:

Definition 1. *Let $H(V, E)$ be a r-uniform hypergraph, we define H_k^r as $H_k^r = H \oplus I_{(r-1)*k+1}$.*[1] *$I_{(r-1)*k+1}$ is a stable set with $(r-1)*k+1$ vertices.*

Theorem 9. *For every r-uniform hypergraph $H(V, E)$ holds:*

$$P_c(H) \leq k \text{ if and only if } d_S(H_k^r) \leq k.$$

Proof. Assume, that $V(H)$ can be partitioned to k or fewer cliques. Then H_k^r can be partitioned into these k or fewer cliques and a stable set $I_{(r-1)*k+1}$. According to the Lemma 8 $d_S(H_k^r) \leq k$.

If $d_S(H_k^r) \leq k$ then according to the Lemma 8 $V(H_k^r)$ can be partitioned to k or fewer cliques K_i and a maximal stable set S. If $V \cap S \neq \emptyset$, then the stable set S is not a subset of $I_{(r-1)*k+1}$ and $I_{(r-1)*k+1} \subseteq \bigcup_i K_i$ holds, because the H_k^r was constructed with all edges between $V(H)$ and $I_{(r-1)*k+1}$

[1] Operation $\oplus$ is join of hypergraphs, for explanation see Chapter 2.

and in the stable set there cannot be vertices from $V(H)$ and $I_{(r-1)*k+1}$. $I_{(r-1)*k+1} \subseteq \bigcup_i K_i$ can not hold, because each K_i can have at most $r-1$ vertices in $I_{(r-1)*k+1}$ and $I_{(r-1)*k+1}$ has $(r-1)*k+1$ vertices. So $V \cap S = \emptyset$ and $V \subseteq \bigcup_i K_i$. □

Theorem 10. *The HYPERGRAPHS SPLIT DIMENSION problem is NP-complete, even for every fixed $r \geq 3$ and $k \geq 2$.*

Proof. This problem is in NP, because we can decide in a polynomial time if a given hypergraph is split.

We reduce the problem r-HYPERGRAPH CLIQUE PARTITION NUMBER to this problem. According to the Theorem 9 $P_c(H) \leq k$ if and only if $d_S(H_{S_k}) \leq k$, so for every given hypergraph H for the CLIQUE PARTITION NUMBER problem we create H_{S_k} and check if $d_S(H_{S_k}) \leq k$. □

Chapter 5

Threshold Hypergraphs

In this section we present some known results about the threshold graphs and hypergraphs. It is based mostly on the book from Peled and Mahadev [MP95] and the article from Reiterman et al. [RRST85].

5.1 Motivation

The notion *threshold graph* was first defined by Chvátal and Hammer [CH75]. These are simply graphs for which there exists a linear-threshold function separating independent sets and non-independent sets. The motivation for their study comes from many directions like computer science, psychology, sociology, scheduling theory, and many others. To display the variety of usage of threshold graphs, they can be applied in set-packing problems [CH73, CH75], parallel processing [HZ77] or resource allocation [Ord85]. For us the most important work is that of Yannakakis [Yan82], where he proved NP-completness of recognition of graphs with threshold dimension of at most three. In his proof he used a class of graphs named *difference graphs* (or *chain graphs*), which are very close to the threshold graphs. Many properties of difference graphs hold also for threshold graphs and vice versa.

5.2 Threshold graphs

Threshold graphs are defined as follows:

Definition 2. *A graph $G = (V, E)$ is a threshold graph if there exist non-*

negative real numbers w_v *(*$v \in V$*) and* t *such that*

$$\sum_{v \in U} w_v \leq t \text{ if and only if } U \subseteq V \text{ is a stable set}$$

There exist many equivalent characterizations of threshold graphs. The following theorem from [MP95] gives us six different characterizations of threshold graphs.

Theorem 11 ([MP95]). *For a graph* $G(V, E)$*, the following statements are equivalent:*

- G *is a threshold graph.*
- G *does not have induced subgraphs isomorphic to* P_4*,* C_4 *or* $2K_2$*.*
- G *is a split graph* $G(K, S)$ *and the neighborhoods of the vertices of* S *are nested.*
- G *can be constructed from the one-vertex graph by repeatedly adding an isolated vertex or a dominating vertex.*
- *The vicinal preorder of* G *is total.*
- *There exist non-negative real numbers* w_v *(*$v \in V$*) and* t *such that for distinct vertices* u *and* v*,*

$$w_u + w_v > t \text{ if and only if } \{u, v\} \in E$$

5.3 Definition

M. CH. Golumbic [Gol80] first considered the problem of generalization of threshold graphs to r-uniform hypergraphs. He proposed the following three generalizations.

Definition 3. *A hypergraph* $H(V, E)$ *is a threshold hypergraph of type* T_1 *if and only if there exist (positive integer) numbers* w_v *(*$v \in V$*) and a threshold* t*, such that*

$$\sum_{v \in U} w_v \leq t \text{ if and only if } U \subseteq V \text{ is a stable set}$$

Definition 4. *A hypergraph $H(V, E)$ is a threshold hypergraph of type T_2 if and only if there exist (positive integer) numbers w_v $(v \in V)$ and a threshold t, such that for every subset $A \subseteq V$ of size r*

$$A \in E \text{ if and only if } \sum_{v \in A} w_v > t$$

Sometimes we define the numbers w_v $(v \in V)$ as a labeling function c, such that $c(v) = w_v$ for every $v \in V$.

Now we define a partial order $\ll$ on the vertices of a hypergraph in this way:

Definition 5. *For $x, y \in V$ $x \ll y$ if and only if for any $\{x_1, x_2, \ldots, x_{r-1}\} \in [V \backslash \{x, y\}]^{r-1}$ (that is, for any $r-1$-element subset of $V \backslash \{x, y\}$), $\{x, x_1, \ldots, x_{r-1}\} \in E$ implies $\{y, x_1, \ldots, x_{r-1}\}$*

Definition 6. *A hypergraph $H(V, E)$ is a threshold hypergraph of type T_3 if and only if the partial order $\ll$ on V is total, e.g for all $x, y \in V$ either $x \ll y$ or $y \ll x$ or both hold.*

5.4 Properties

Golumbic has also asked the question whether these definitions are equivalent. This was answered negatively by Reiterman et al. [RRST85]. They also gave us some interesting theorems about threshold hypergraphs. We will mention some of them here.

Theorem 12 ([RRST85])**.** *For every r-uniform hypergraph $H(V, E)$ property T_1 implies property T_2 and T_2 implies property T_3. Conversely the property T_3 does not imply property T_2 and the property T_2 does not imply property T_1. If $r = 2$ properties T_1,T_2 and T_3 are equivalent.*

Reiterman et al. [RRST85] also characterized T_3 hypergraphs in terms of forbidden configurations.

Definition 7. *Let H(V,E) be a r-uniform hypergraph. A forbidden configuration in H is a finite sequence of not necessarily distinct vertices $x_1, x_2, ..., x_r, y_1, y_2, ..., y_r$; $x_1 \notin \{y_2, ..., y_r\}$,$y_1 \notin \{x_2, ..., x_r\}$ in V such that $\{x_1, x_2, ..., x_r\} \in E$,$\{y_1, y_2, ..., y_r\} \in E$,$\{y_1, x_2, ..., x_r\} \notin E$,$\{x_1, y_2, ..., y_r\} \notin E$.*

Theorem 13 ([RRST85])**.** *Hypergraph $H(V,E)$ is a T_3 hypergraph if and only if it does not contain a forbidden configuration.*

Another characterization of T_3 and T_2 hypergraphs is given by Reiterman et al. [RRST85] using the concept of a system of generators.

Definition 8. *Let H(V,E) be a r-uniform hypergraph. We say that $E' \subset E$, $E' \neq \emptyset$ is a system of generators of H if there exists a linear ordering $\prec$ of V such that $e = \{x_1, x_2, \ldots, x_r\} \in E$ if and only if $(\exists e' = \{x'_1, x'_2, \ldots, x'_r\} \in E')(\forall i \in \{1, 2, \ldots, r\})(x'_i \prec x_i)$*

Theorem 14 ([RRST85])**.** *Hypergraph $H(V,E)$ is a T_3 hypergraph if and only if it has a system of generators.*

Theorem 15. *[[RRST85]] Hypergraph $H(V,E)$ is a T_2 hypergraph if and only if it has a system of generators with only one generator.*

Chapter 6

Dilworth Number of a hypergraph

6.1 Introduction

In this section we use definitions from [FH78] and [MP95]. For a graph (V, E) we define a vicinal preorder as a binary relation $\precsim$ on V as

$$x \precsim y \text{ if and only if } N(x) \subseteq N(y) \cup \{y\}$$

(For $v \in V$, $N(v)$ denotes the set of neighbors of v.)

Since $\precsim$ is a reflexive and transitive relation, $\precsim$ is a preorder. If $x \precsim y$ but $y \precsim x$ does not hold, then we denote this by $x \prec y$. If $x \precsim y$ and $y \precsim x$ holds, then we denote this by $x \sim y$.

A *chain* is defined as a set of mutually comparable vertices. All the elements in a chain can be sorted into a sequence $x_1, x_2, \ldots, x_k$ such that $x_i \precsim x_j$ holds for all i, j satisfying $1 \leq i < j \leq k$. Vertices of every graph G can be partitioned into chains. The least number of such chains, into which the vertices of the graph can be partitioned, is called *Dilworth number* and is denoted by $D(G)$. As an opposite of a chain we define an *antichain* as a set of mutually incomparable elements.

According to the well-known Dilworth theorem [Dil50], the Dilworth number is equal to the size of the largest antichain.

6.2 Generalization to hypergraphs

We use the relation defined in the Definition 5 for the definition of a vicinal preorder of a hypergraph:

> For $x, y \in V$, $x \precsim y$ if and only if for any $\{x_1, x_2, \ldots, x_{r-1}\} \in [V \setminus \{x, y\}]^{r-1}$, $\{x, x_1, \ldots, x_{r-1}\} \in E$ implies $\{y, x_1, \ldots, x_{r-1}\}$

We define the *Dilworth number of a hypergraph* as the least number of chains, into which the vertices of the hypergraph can be partitioned.

As we know from the previous chapter, threshold graphs are exactly graphs with Dilworth number 1. For hypergraph it holds that threshold hypergraphs of the type T_3 are precisely the hypergraphs with Dilworth number 1.

6.3 Computation of the Dilworth number

There exist several algorithms for the computation of the Dilworth number for graphs. An $O(n^3)$ algorithm was proposed by Mahadev [Mah84]. An even faster algorithm can be given if we use matrix multiplication. In that case the computation time of such an algorithm will be $O(f(n))$, where $f(n)$ is the time required to multiply two $n \times n$ matrices. Currently the best know algorithm for the matrix multiplication is from Coppersmith and Winograd and requires time $O(n^{2.376})$ [CW87]. For graphs with small Dilworth number k the algorithm from Felsner et al. is efficient, which recognizes whether a graph has the Dilworth number k in the time $O(k^2 n^2)$ [FRS03]. We will use the Mahadev's algorithm as a model for our algorithm.

Theorem 16. *The Dilworth number of a r-uniform hypergraph can be computed in the time $O(n^{r+1})$.*

Proof. Let $H(V, E)$ is a r-uniform hypergraph with $V = \{x_1, \ldots, x_n\}$. We will construct a transitive directed graph G' on V with edges:

$$E(G') = \{(x_i, x_j) | (x_j \prec x_i) \text{ or } (x_i \sim x_j \text{ and } i < j)\}$$

Each clique in G' forms a chain of G and every chain of G forms a clique in G'. So the Dilworth number of G is the size of the minimal clique partition of G'. There is an $O(n^2)$ algorithm for transitive orientable graphs [Gol80], which computes such a partition. So this part of this algorithm is not "slow".

The construction of G' can be done in the time $O(n^{r+1})$. For every pair of vertices x,y we scan the sets of $r-1$ vertices to determine, whether $x \prec y$ or $x \sim y$ holds. So the total time will be $O(n^{r-1} * n^2)$=$O(n^{r+1})$. □

The Dilworth number for hypergraphs gives us an upper bound for some parameters of a hypergraph, such as the diameter and domination number. These parameters are NP-hard to compute even for graphs, therefore a polynomial algorithm for the computation of a Dilworth number for hypergraphs can be very useful.

We use the definition of a *diameter* of a hypergraph from Ye [Ye03]. Let $H = (V, E)$ be a r-uniform hypergraph, $E = \{e_1, e_2, \ldots, e_m\}$. A path P in H from x_1 to x_{s+1} is a vertex-edge alternative set $x_1e_1x_2e_2, \ldots, x_s, e_s, x_{s+1}$ such that $\{x_i, x_{i+1}\} \subseteq e_i$ $(i = 1, 2, \ldots, s)$ and $x_i \neq x_j$, $e_i \neq e_j$ $(i \neq j)$, where s is called the length of path P. The distance of vertices x and y, $dist(x, y)$ is the minimum length of a path which connects x and y. We denote the diameter of H by $d(H)$ and is defined as $d(H) = max\{dist(x, y)|x, y \in V\}$.

The following two theorems were proved for graphs by Földes and Hammer [FH78]. We prove them for hypergraphs.

Theorem 17. *For every r-uniform hypergraph H holds:*

$$d(H) \leq D(H) + 1.$$

Proof. Let the diameter of H be k, then H contains an induced path with $k+1$ vertices. Now assume, that two intermediate vertices x_i, x_j in this path are comparable, so let $x_i \precsim x_j$. Let the path look like this:

$$x_1e_1x_2e_2, \ldots, e_{i-1}, x_i, e_i, \ldots, e_{j-1}, x_j, e_j, \ldots, x_s, e_s, x_{s+1}$$

(the proof will be analogous if x_j will be before x_i in the path). Then we can omit the part of the path $x_i, e_i, \ldots, e_{j-1}$, and if the edge e_{i-1} does not contain x_j we replace this edge by $(e_{i-1} \setminus \{x_i\}) \cup \{x_j\}$. Such an edge must exist in E because $x_i \precsim x_j$. So no two intermediate vertices on the path are comparable, thus the Dilworth number is at least $k - 1$. □

Definition 9. *A dominating set is a subset S of vertices such that every vertex not in S is adjacent to some vertex in S. The domination number of H, denoted by $\gamma(H)$, is the minimum size of a dominating set of H.*

Theorem 18. *If H(V,E) is a r-uniform hypergraph with no isolated vertices, than*

$$\gamma(H) \leq D(H).$$

Proof. Let S be a minimum dominating set such that the subhypergraph of H induced by S has the largest possible number of edges. We claim that S is an antichain. Assume otherwise, that $x \precsim y$ for $x, y \in S$. If there is an edge between x and some vertex from $S \setminus \{x\}$ (that is, an edge containing both vertices), then the set $S \setminus \{x\}$ will be dominating. This is a contradiction with the minimality of S. So there is no edge in E with x and some vertex from $S \setminus \{x\}$. But H has no isolated vertices, so there must be an edge e between x and some vertex z from $V \setminus S$. Because $x \precsim y$ holds, there is also an edge $(e \setminus \{x\}) \cup \{y\}$ in E. Then the set $(S \setminus \{x\}) \cup \{z\}$ is a minimum dominating set, which has more edges than S, a contradiction. Therefore S is an antichain and $\gamma(G) = |S| \leq D(G)$. □

Chapter 7

Hypergraph Threshold Dimension

7.1 Introduction

In this chapter we will deal with the *threshold dimension problem*. Formally the *threshold dimension* $d_t(G)$ of a graph G is defined as the minimal number of threshold subgraphs of G needed to cover the edges $E(G)$. For a r-uniform hypergraph $H(V, E)$ we define the threshold dimension analogously, as the minimal number the threshold subhypergraphs needed to cover the edges $E(H)$. According to the types of threshold subhypergraphs we distinguish between T_1 threshold dimension $d_{T_1}(G)$, T_2 threshold dimension $d_{T_2}(G)$ and T_3 threshold dimension $d_{T_3}(G)$.

7.2 Relationship to the Learning theory

The threshold dimension problem has also applications in machine learning, in the learning theory. We mention here a theorem from Aizenstein et al. [AHHP98].

Theorem 19 ([AHHP98]). *Let F be a polynomially reasonable and polynomially size-bounded class of Boolean functions. If REP(F) is NP-hard under* $\leq_m^p$ *reductions, then F is not learnable with membership and equivalence queries unless NP=co-NP.*

Aizenstein et al. [AHHP98] also asked whether the representation problem for the class of unions of two linearly separable (threshold) functions is NP-hard. This problem can be translated to the language of graph theory.

Let f be a monotone Boolean 3-DNF formula, $f(x_1, \ldots, x_n) = x_{i1}x_{j1}x_{k1} \vee x_{i2}x_{j2}x_{k2} \vee \cdots \vee x_{im}x_{jm}x_{km}$. We construct a 3-uniform hypergraph $H_f = (V, E)$ with the edge set $E = \{\{x_{il}, x_{jl}, x_{kl}\} | 1 \leq l \leq m\}$ and the vertex set $V = \{x_1, \ldots, x_n\}$.

Theorem 20. *A monotone Boolean 3-DNF formula f can be represented as a union of two linearly separable functions if and only if H_f has threshold dimension 2 of the type T_1.*

Proof. Let f is representable as an union of two linearly separable functions, e.g $f(x_1, \ldots, x_n) = 1$ if and only if $l_1(x_1, \ldots, x_n) = 1$ or $l_2(x_1, \ldots, x_n) = 1$, where l_1 and l_2 are linearly separable functions. Let these two be: $l_1(x_1, \ldots, x_n) = [\sum_{i=1}^{n} w1_i * x_i > \theta_1]$ and $l_2(x_1, \ldots, x_n) = [\sum_{i=1}^{n} w2_i * x_i > \theta_2]$. The two 3-uniform T_1 threshold hypergraphs will be $H_1 = (V_1, E_1)$ and $H_2 = (V_2, E_2)$ with the sets: $E_1 = \{\{v_i, v_j, v_k\} | v_i, v_j, v_k \in V, v_i \neq v_j \neq v_k, w1_i + w1_j + w1_k > \theta_1\}$, $V_1 = \bigcup_{e \in E_1} e$, $E_2 = \{\{v_i, v_j, v_k\} | v_i, v_j, v_k \in V, v_i \neq v_j \neq v_k, w2_i + w2_j + w2_k > \theta_2\}$ and $V_2 = \bigcup_{e \in E_2} e$. Clearly the set $T \subseteq V$ is not stable if and only if there exists a subset of $T' \subseteq T$ with three elements such that $T' \in E_1$ or $T' \in E_2$.

On the other side from two 3-uniform T_1 threshold hypergraphs H_1,H_2 and their values of threshold functions we can create two linearly separable functions l_1, l_2. If for example some threshold hypergraph H_i does not contain all vertices, the value of this vertex in the function l_i will be zero. □

So if we will be able to prove NP-hardness of recognizing hypergraphs with threshold dimension 2 of the type T_1, it will imply that the class of unions of two linearly separable functions is not learnable with membership and equivalence queries unless NP=co-NP.

7.3 The Problem

The threshold dimension problem is defined as follows:

THRESHOLD DIMENSION
Instance: A r-uniform hypergraph H(V,E), a type of threshold dimension $t \in \{1, 2, 3\}$ and a positive integer $k \leq |V|$
Problem: Is it true that $d_{T_t(H)} \leq k$?

From the work of Chvátal and Hammer [CH75] we know that the determination of the threshold dimension of graphs is NP-complete problem. Furthermore, as was proved by Yannakakis [Yan82], it is NP-complete for every fixed $k \geq 3$. Therefore THRESHOLD DIMENSION is an NP-complete problem, even for every fixed $k \geq 3$. We are interested in the question whether this problem is NP-complete for fixed $k = 2$.

We are unable to answer this question, but we give a reduction of the COVERING BY r-UNIFORM CLIQUES ($k = 2$) problem to the THRESHOLD DIMENSION ($k = 2$) problem. If it will be proven that the COVERING BY r-UNIFORM CLIQUES problem is NP-complete for fixed $k = 2$, this reduction will imply NP-hardness of the THRESHOLD DIMENSION problem for fixed $k = 2$.

7.4 The Reduction

Here we give the reductions of the COVERING BY r-UNIFORM CLIQUES ($k = 2$) problem to THRESHOLD DIMENSION ($k = 2$) for every type of threshold hypergraphs T_1, T_2 and T_3.

Definition 10. *For 3-uniform hypergraph $H(V, E)$ we define 3-uniform hypergraph H_T as follows: $H_T(V_T, E_T)$, where vertices are $V_T = \{V \cup W\}$ and $W = \{w_0, w_1, w_2\}$ and edges are $E_T = \{E \cup F\}$, where $F = \{\{w, w', v\} | v \in V \wedge w, w' \in W \wedge w \neq w'\} \cup \{\{w, v, v'\} | v, v' \in V \wedge w \in W \wedge v \neq v'\}$.*

Theorem 21. *Let $H(V, E)$ be a 3-uniform hypergraph. For the hypergraph H_T constructed according to the definition 10 the following holds:*

$$c(H) \leq 2 \text{ if and only if } c_{T_3}(H_T) \leq 2$$

Proof. If $c(H) = 1$, then the graph H_T is T_3 threshold hypergraph, because the sorting of vertices $w_0, w_1, w_2, v1_1, ..., v1_n$, where $(v1_1, ..., v1_n \in V)$, satisfies Definition 6 of T_3 property.

Let $c(H) = 2$. So the edges of H can be covered by two cliques: C_1 and C_2 ($C_1, C_2 \subseteq V$, C_1, C_2 induces a r-clique in H). Then there exist two sets V_1 and V_2, which are subsets of V_T, such that V_1 and V_2 induces a T_3 hypergraph in H_T. Let $V_1 = C_1 \cup W$ and $V_2 = C_2 \cup W$. When we sort the vertices in V_1 in the following way: $w_0, w_1, w_2, v1_1, ..., v1_n$, where $(v1_1, ..., v1_n \in C_1)$, then this sorting satisfies Definition 6 of T_3 property. Similarly the sorting $w_0, w_1, w_2, v2_1, ..., v2_m$, where $(v2_1, ..., v2_m \in C_2)$, satisfies Definition 6 of T_3 and V_2 induces a T_3 hypergraph in H_T too. All edges of H_T are covered by these two T_3 hypergraphs.

On the other side, now assume that H_T can be covered by two or less T_3 hypergraphs. Let the two vertex sets be V_1 and V_2. We distinguish between two cases:

In case, when one of V_1, V_2 is empty (H_T can be covered by one T_3 hypergraph), for example V_1, than $V_2 = V_T$ must hold. Assume that the set $V'' = V_T \setminus W$ is not a clique; then there is a set of distinct vertices $s = \{v_1, v_2, v_3\} \notin E, (v_0, v_1, v_2 \in V)$. Then V_T contains a forbidden configuration: $w_0, v_1, v_2, v_3, w_1, w_2$ - a contradition with the fact that V_T is a T_3 hypergraph. Therefore V'' is a clique and together with an empty set creates a partition of V into two induced cliques in H.

In the second case, when V_1 and V_2 is not empty. We claim that both V_1 and V_2 must contain w_1, w_2 and w_0 or one of them is equal to V_T.

Assume that one of them (for example V_1) does not contain some of the vertices w_1, w_2 or w_0. Let it be w_j. Since we need to cover the edges from the set $\{\{w_j, w', v\} | v \in V \wedge w' \in W \wedge w_j \neq w'\}$, V_2 must contain w_1, w_2 and w_0. But to cover the edges from the set $\{\{w_j, v, v'\} | v, v' \in V \wedge v \neq v'\}$ it must contain also all vertices from V. This is a contradiction with the assumption that one T_3 hypergraph is not enough to cover H_T.

So we assume that both V_1 and V_2 contain w_1, w_2 and w_0. Assume, that $V_1 \setminus W$ (or $V_2 \setminus W$) is not a clique, then there is a set of vertices $s = \{v_1, v_2, v_3\} \notin E$, $v_1, v_2, v_3 \in V_1 \setminus W$. Then similarly V_1 contains a forbidden configuration: $w_0, v_1, v_2, v_3, w_1, w_2$ - a contradiction with the fact that V_T is a T_3 hypergraph. □

Theorem 22. *Let $H(V, E)$ be a 3-uniform hypergraph. For the hypergraph H_T constructed according to the definition 10 the following holds:*

$$c(H) \leq 2 \text{ if and only if } c_{T_2}(H_T) \leq 2$$

Proof. If $c(H) = 1$, then the graph H_T is T_2 threshold hypergraph, because the sorting of vertices $w_0, w_1, w_2, v1_1, ..., v1_n$, where $(v1_1, ..., v1_n \in V)$, satisfies Characterization 15 of T_2 property, e.g the set $\{w_0, w_1, v1_1\}$ is the generator.

Assume that H can be covered by two cliques (C_1 and $C_2 \subseteq V$, C_1, C_2 induce a r-clique in H). Then H_T can be covered by two T_2 hypergraphs $V_1 = C_1 \cup W, V_2 = C_2 \cup W$. V_1 (and V_2) induces a T_2 hypergraph in H_T. Indeed, when we sort the vertices in V_1 in the following way: $w_0, w_1, w_2, v1_1, ..., v1_n(v1_1, ..., v1_n \in C_1)$, then this sorting satisfies Characterization 15 of T_2 property and the set $\{w_0, w_1, v1_1\}$ is the generator.

The second implication follows from the theorem 12, which says that a T_2 hypergraph is also a T_3 hypergraph and from the previous theorem 21. □

Theorem 23. *Let $H(V, E)$ be a 3-uniform hypergraph. For the hypergraph H_T constructed according to the definition 10 the following holds:*

$$c(H) \leq 2 \text{ if and only if } c_{T_1}(H_T) \leq 2$$

Proof. If $c(H) = 1$, then the graph H_T is a T_1 threshold hypergraph, because we can set $c(w) = 1$, $w \in W$, $c(v) = 5$ for $v \in V$ and $t = 4$, and this labeling and threshold satisfies the definition of T_1.

Assume that H can be covered by two cliques (C_1 and $C_2 \subseteq V$, C_1, C_2 induce a r-clique in H). Then H_T can be covered by two T_1 hypergraphs $V_1 = C_1 \cup W$, $V_2 = C_2 \cup W$. V_1 (and V_2) induces a T_1 hypergraph in H_T. Indeed, if we set labeling $c(w) = 1$, $w \in W$, $c(v) = 5$ for $v \in V_1$ and $t = 4$, then this labeling and threshold satisfies the definition of T_1.

The second implication follows from the theorem 12, which says that a T_1 hypergraph is also a T_3 hypergraph and from the previous theorem 21. □

Theorem 24. *If the COVERING BY r-UNIFORM CLIQUES problem is NP-complete for fixed $k = 2$, then the problem THRESHOLD DIMENSION for fixed $k = 2$ is NP-complete for threshold dimension of the type T_2 and T_3 and NP-hard for threshold dimension of the type T_1.*

Proof. The THRESHOLD DIMENSION for $k = 2$ and threshold dimension of the type T_2 and T_3 is in NP, because we know polynomial algorithms for deciding whether a given hypergraph is a threshold hypergraph of the type T_2 or T_3 [RRST85]. We can use the Theorems 21, 22 or 23 to reduce the COVERING BY r-UNIFORM CLIQUES ($k = 2$) problem to the THRESHOLD DIMENSION for k=2 with the threshold dimension of the type T_1, T_2 or T_3. □

Chapter 8

The Partial Order Dimension

8.1 Introduction

Yannakakis [Yan82] has proved that the problem of determining if a given partial order has dimension at most 3 is NP-complete. A result contained in his proof is that the problem of determining if the edges of a given graph can be covered by 3 difference subgraphs has the same complexity. And similarly the problem of determing if the edges of a given graph can be covered by 3 threshold graphs is NP-complete. We give here the proof of Yannakakis together with some background definitions.

From Chapter 2 we know what a partial order is and how the dimension of a partial order is defined. We can imagine a poset P=(X,P) as a directed acyclic graph (DAG) on X. Assume that a partial order P is partitioned into two sets S,S' so that there is no edge directed from S' to S in P. Define B(P) as the bipartite graph with nodes X and the following set of edges: $\{(x, y) \mid x \in S, y \in S', x \bowtie y\}$. The bipartite graph B=(X,Y,P) is called a *chain graph* if there is no induced $2K_2$ in G. In other words for any pair of vertices $u, v \in X$ (or Y), either $N(u) \subseteq N(v)$ or $N(v) \subseteq N(u)$ (The symbol N(v) stands for the set of neighbors of vertex v). Chain graphs are also called *difference graphs* [HPS90]. For a bipartite graph G, let ch(G) be the minimum number of difference (chain) graphs of G that cover all the edges of G.

According to results of Dushnik and Miller [DM41] we know the necessary and sufficient conditions for a partial order to have dimension 2. These conditions gave us polynomial algorithm for testing if an arbitrary partial order has dimension at most 2. However for $k \geq 3$ this problem is NP-complete.

8.2 The partial order dimension problem

Lemma 25. *For every partial order P, $ch(B(P)) \leq d_O(P)$.*

Proof. Let P has dimension d. Then there are d linear extensions $L_1, \ldots, L_d$ of P, whose intersection is P. Let L be a linear extention of P, we define L' as the bipartite graph with the vertex set $V(B(P))$ and the edge set $\{\{x,y\} | x \in X, y \in Y, (y,x) \in L\}$. The graphs $L'_1, \ldots, L'_d$ are difference graphs and cover all edges of B(P). □

Corollary 26. *For every partial order P, $d_O(P) \leq 3 \Rightarrow ch(B(P)) \leq 3$.*

The following problem is known [GJ79] to be NP-complete.

GRAPH CHROMATIC NUMBER 3
Instance: Given a graph $G(V,E)$.
Problem: Is the chromatic number $\chi(G) \leq 3$?

We reduce the GRAPH CHROMATIC NUMBER 3 problem to the partial order dimension 3 problem:

PARTIAL ORDER DIMENSION 3
Instance: Given a partial order.
Problem: Is the dimension $d_O(P) \leq 3$?

In the reduction we will construct for any graph G a partial order P in such a way that $d_O(P) \leq 3$ if and only if $\chi(G) \leq 3$. We use the construction from [MP95]. Let $G = (V,E)$, with $V = \{u_1, \ldots, u_n\}, E = \{e_1, \ldots, e_m\}$ is a graph. We will construct a partial order P on the union of two disjoint sets S and S'. The set S is partitioned into the sets Q and R, and the set S' is partitioned into Q' and R'. The sets are

$$Q = \{u_{ik} | \text{ vertex } u_i \text{ is incident with edge } e_k\},$$

$$R = \{u_{ia}, u_{ib} | u_i \text{ is a vertex}\},$$

$$Q' = \{u'_{ik} | \text{ vertex } u_i \text{ is incident with edge } e_k\},$$

$$R' = \{u'_{ia}, u'_{ib} | u_i \text{ is a vertex}\}.$$

We call the vertices in Q' and R' primed versions of its counterparts in the sets Q and R. The partial order P is defined by these rules:

- Each vertex in R is smaller than each vertex in $R' \cup Q'$ with the same first index, except for its primed version.
- Each vertex in Q is smaller than each vertex in R'.
- Each vertex u_{ik} in Q is smaller than a vertex u'_{jk} in Q' if $k < l$.
- A vertex $u_{ik} \in Q$ is smaller than a vertex $u'_{jk} \in Q'$ if $e_k = \{u_i, u_j\}$.
- A vertex $u_{ik} \in Q$ is smaller than a vertex $u'_{jl} \in Q$ if $k < l$.
- There are no other relations in P.

From this definition we can see that the edges of $B(P)$ are defined by these conditions:

- There is an edge from each vertex in R to its primed version in R' and to all vertices of $R' \cup Q'$ with a different first index.
- There is an edge from $u_{ik} \in Q$ to $u'_{jl} \in Q'$ if $k > l$.
- There is an edge from each vertex in Q to its primed version in Q'.

Lemma 27 ([Yan82]). *$ch(B(P)) \leq 3 \Rightarrow \chi(G) \leq 3$*

Proof. Let $ch(B(P)) \leq 3$ and D_1, D_2 and D_3 are three difference subgraphs of $B(P)$ that cover its edges. The subgraph H_i of $B(P)$ induced by all vertices with the first index i has three connected components: the edge $\{u_{ia}, u'_{ia}\}$, the edge $\{u_{ib}, u'_{ib}\}$ and the subgraph induced by u_{ik}, u'_{ik}, where $i \leq k \leq m$. Since every D_l is a difference graph, no D_l can contain two edges from different components of H_i. H_i has three components, so all the edges of the third component are in the same D_l, we color vertex u_i with color l. To show that this is a valid coloring, assume that there are two adjacent vertices u_i,u_j colored with the same color l. Let they be connected by the edge e_k=$\{u_i, u_k\}$. Then according to the definition of the coloring $\{u_{ik}, u'_{ik}\}, \{u_{jk}, u'_{jk}\} \in B_l$. This is a contradition with the assumption that B_l is a difference graph, because these two edges induce a $2K_2$ in $B(P)$.

□

Lemma 28 ([Yan82])**.** $\chi(G) \leq 3 \Rightarrow d_O(P) \leq 3$

Proof. The proof can be found in the article [Yan82] or in the book [MP95]. □

Theorem 29 ([Yan82])**.** *The problem PARTIAL ORDER DIMENSION 3 is NP-complete.*

Proof. It follows from Corollary 26 and Lemmas 27 and 28. □

The proof follows from this lemma:

Lemma 30 ([Yan82])**.** *Let B is a bipartite graph whose vertices are partitioned into the stable sets X and Y. Let B' be obtained from B by adding all the edges between vertices in X (i.e making X into a clique). Then $ch(B) = D_t(B')$.*

And finally the theorem about NP-completness of the threshold dimension problem.

Theorem 31 ([Yan82])**.** *For every fixed $k \geq 3$, it is NP-complete to determine if a given graph G has threshold dimension at most k, even if G is a split graph.*

The reduction was given from the GRAPH CHROMATIC NUMBER 3 problem. One could think that in the case of hypergraph threshold dimension 2 a similar proof can be done using the reduction from HYPERGRAPH CHROMATIC NUMBER 2. However such a proof would not be trivial.

Let's have a look at the coloring of $B(P)$ in the proof. The subgraph S induced by the vertices u_{ik},u_{jk},u'_{ik} and u'_{jk} represents an edge $e_k = \{u_i, u_j\}$ in the graph. This graph is a $2K_2$ with edges $\{u_{ik}, u'_{ik}\}$ and $\{u_{jk}, u'_{jk}\}$. Clearly edges of S cannot be covered by one difference graph. The difference graph covering the edge $\{u_{ik}, u'_{ik}\}$ represents the color of the vertex u_i in the original graph.

Here we present a hypergraph, which could be used in a similar proof for hypergraphs. Three from its edges can be covered by two T_3 threshold subhypergraphs only in such a way that there is a bijection between this covering of its (1st, 2nd and 5th) edges and the possible coloring of three vertices connected by an edge in a 3-uniform hypergraph.

Theorem 32. *The following 3-uniform hypergraph:* $H = (V, E)$*, where* $V = \{0, 1, 2, 3, 4\}$*,* $E = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7\}$ *and*
$e_1 = \{0, 1, 4\}$
$e_2 = \{0, 2, 3\}$
$e_3 = \{0, 2, 4\}$
$e_4 = \{0, 3, 4\}$
$e_5 = \{1, 2, 3\}$
$e_6 = \{1, 2, 4\}$
$e_7 = \{1, 3, 4\}$
can be covered by two T_3 *threshold subhypergraphs only by one of these* 6 *possible coverings:*

$cover_1(e_1) = 1$ $cover_2(e_1) = 1$ $cover_3(e_1) = 1$
$cover_1(e_2) = 1$ $cover_2(e_2) = 2$ $cover_3(e_2) = 2$
$cover_1(e_3) = 1$ $cover_2(e_3) = 1$ $cover_3(e_3) = 2$
$cover_1(e_4) = 1$ $cover_2(e_4) = 1$ $cover_3(e_4) = 2$
$cover_1(e_5) = 2$ $cover_2(e_5) = 2$ $cover_3(e_5) = 1$
$cover_1(e_6) = 2$ $cover_2(e_6) = 1$ $cover_3(e_6) = 1$
$cover_1(e_7) = 2$ $cover_2(e_7) = 1$ $cover_3(e_7) = 1$

$cover_4(e_1) = 2$ $cover_5(e_1) = 2$ $cover_6(e_1) = 2$
$cover_4(e_2) = 2$ $cover_5(e_2) = 1$ $cover_6(e_2) = 1$
$cover_4(e_3) = 2$ $cover_5(e_3) = 2$ $cover_6(e_3) = 1$
$cover_4(e_4) = 2$ $cover_5(e_4) = 2$ $cover_6(e_4) = 1$
$cover_4(e_5) = 1$ $cover_5(e_5) = 1$ $cover_6(e_5) = 2$
$cover_4(e_6) = 1$ $cover_5(e_6) = 2$ $cover_6(e_6) = 2$
$cover_4(e_7) = 1$ $cover_5(e_7) = 2$ $cover_6(e_7) = 2$

Proof. The proof explores all different coverings by two T_3 threshold subhypergraphs and for every coloring it either displays that it is a valid covering, or displays a forbidden configuration, which violates the characterization of T_3 hypergraphs. The proof was done by a computer program and it is omitted here because of space limitations. The complete proof can be found on the website [Rep]. □

Chapter 9

Future directions

This thesis provides answers to some questions, but several important open problems in this area still remain unsolved.

Hypergraph threshold dimension two: Is the problem of recognizing hypergraphs with threshold dimension two NP-hard? The problem for graphs and fixed dimension 3 is NP-complete, and the reduction is done from the graph coloring problem by three colors, which is NP-complete. Since the coloring problem for hypergraphs is NP-complete for two colors, it is not unreasonable to expect that the problem of recognizing hypergraphs with threshold dimesion 2 is NP-hard. We propose a particular 3-uniform hypergraph with specific properties which could be used in such a proof. Another possibility is to prove the NP-completeness of the COVERING BY r-UNIFORM CLIQUES problem for $k = 2$ and use our reduction.

Covering graphs coverable by three threshold graphs by four threshold graphs: The problem of coloring 3-colorable graphs with 4 colors is NP-hard [GJS74]. Since the proof of NP-completeness of the THRESHOLD DIMENSION problem was done using a reduction from the graph coloring problem, we ask whether the problem of covering graphs coverable by three threshold graphs by four threshold graphs is NP-complete.

Bibliography

[AHHP98] Howard Aizenstein, Tibor Hegedus, Lisa Hellerstein, and Leonard Pitt. Complexity theoretic hardness results for query learning. *Computational Complexity*, 7(1):19–53, 1998. 9, 33, 34

[CC86] Zh A Chernyak and A A Chernyak. About recognizing (α, β) classes of polar graphs. *Discrete Mathematics*, 62(2):133–138, 1986. 21

[CC91] Arkady A. Chernyak and Zhanna A. Chernyak. Split dimension of graphs. *Discrete Mathematics*, 89(1):1–6, 1991. 21

[CH73] V. Chvátal and P.L. Hammer. Set-packing and threshold graphs. *Univ. Waterloo Res. Report, CORR*, pages 73–21, 1973. 25

[CH75] V. Chvátal and P.L. Hammer. Aggregation of inequalities in integer programming. (STAN-CS-75-518), 1975. 25, 35

[CW87] D. Coppersmith and S. Winograd. Matrix multiplication via arithmetic progressions. *STOC '87: Proceedings of the nineteenth annual ACM conference on Theory of computing*, pages 1–6, 1987. 30

[Dil50] R. P. Dilworth. A decomposition theorem for partially ordered sets. *Annals of Mathematics*, (51):161–166, 1950. 29

[DM41] B. Dushnik and E.W. Miller. Partially ordered sets. *American Journal of Mathematics*, (63):600–610, 1941. 39

[FH77] S. Földes and P.L. Hammer. Split graphs. *Congressus Numerantium*, 19:311–315, 1977. 21

[FH78] S. Földes and P.L. Hammer. The dilworth number of a graph. *Annals of Discrete Mathematics*, 2:211–219, 1978. 29, 31

[FRS03] Stefan Felsner, Vijay Raghavan, and Jeremy Spinrad. Recognition algorithms for orders of small width and graphs of small dilworth number. *Order*, 20(4):351–364, 2003. 30

[GGHN06] Jens Gramm, Jiong Guo, Falk Hüffner, and Rolf Niedermeier. Data reduction, exact, and heuristic algorithms for clique cover. *Proceedings of the 8th Workshop on Algorithm Engineering and Experiments (ALENEX '06)*, pages 86–94, 2006. Extended version to appear under the title "Data reduction and exact algorithms for clique cover" in *ACM Journal of Experimental Algorithmics*. 19

[GJ79] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY, USA, 1979. 9, 18, 40

[GJS74] M. R. Garey, D. S. Johnson, and L. Stockmeyer. Some simplified np-complete problems. *STOC '74: Proceedings of the sixth annual ACM symposium on Theory of computing*, pages 47–63, 1974. 45

[Gol80] Martin Charles Golumbic. *Algorithmic Graph Theory and Perfect Graphs*. Academic Press, 1980. 26, 30

[HPS90] P. L. Hammer, U. N. Peled, and X. Sun. Difference graphs. *Discrete Appl. Math.*, 28(1):35–44, 1990. 39

[HZ77] Peter B. Henderson and Yechezkel Zalcstein. A graph-theoretic characterization of the pv_{chunk} class of synchronizing primitives. *SIAM J. Comput.*, 6(1):88–108, 1977. 25

[Kar72] Richard M. Karp. Reducibility among combinatorial problems. *Complexity of Computer Computations*, pages 85–103, 1972. 18

[KSW78] L. T. Kou, L. J. Stockmeyer, and C. K. Wong. Covering edges by cliques with regard to keyword conflicts and intersection graphs. *Communications of the ACM*, 21(2):135–139, 1978. 19

[Lov73] L. Lovasz. Colorings and coverings of hypergraphs. *Proceedings of the 4th Southeastern Conference on Combinatorics, Graph Theory, and Computing*, pages 3–12, 1973. 17

[Mah84] N. V. R. Mahadev. Stability numbers in structured graphs. *PhD Thesis*, 1984. 30

[Mer03] Russell Merris. Split graphs. *Eur. J. Comb.*, 24(4):413–430, 2003. 21

[MP95] N. V. R. Mahadev and U. N. Peled. Threshold graphs and related topics. *Annals of Discrete Mathematics*, 56, 1995. 21, 25, 26, 29, 40, 42

[Ord85] E. T. Ordman. Threshold coverings and resource allocation. *Proc. 16th Southeastern International Conf. on Graph Theory, Combinatorics and Computing*, (49):99–113, 1985. 25

[Orl77] J. Orlin. Contentment in graph theory: covering graphs with cliques. *lndag. Math.*, 39:406–424, 1977. 19

[Rep] Vladimír Repiský. Proof of the theorem 35. *http://vrepisky.googlepages.com/theorem*. 43

[RRST85] J. Reiterman, V. Rodl, E. Sinajova, and M. Tuma. Threshold hypergraphs. *Discrete Mathematics*, (54):193–200, 1985. 25, 27, 28, 38

[ST97] Robert H. Sloan and Gyorgy Turan. Learning from incomplete boundary queries using split graphs and hypergraphs. *European Conference on Computational Learning Theory*, pages 38–50, 1997. 21

[Yan82] M. Yannakakis. On the complexity of partial order dimension problem. *SIAM Journal on Algebraic and Discrete Methods*, 3:351–358, 1982. 10, 25, 35, 39, 41, 42

[Ye03] Miaolin Ye. Size of k-uniform hypergraph with diameter d. *Discrete Math.*, 260(1-3):285–293, 2003. 31

Index

Abstrakt

Názov: Pokrývanie hrán hypergrafu: zložitosť a aplikácie
Autor: Vladimír Repiský
Školitel: Mgr. Tibor Hegedüs

Táto práca prináša prehľad o problémoch pokrývania hrán v hypergrafoch. Skúmame v nej zložitostné vlastnosti pre problémy pokrývania hrán v hypergrafoch podhypergrafmi rôznych typov. Uvažujeme pokrývanie klikovými hypergrafmi, rozdelenými hypergrafmi, a prahovými hypergrafmi. Celkovo dokazujeme NP-úplnosť pre 6 problémov. Špeciálne sa zameriavame na problém pokrývania prahovými hypergrafmi, ktorý má aplikácie v teórii strojového učenia. Poskytujeme redukciu z problému pokrývania klikami, takže potencionálny dôkaz NP-tažkosti tohto problému by implikoval NP-ťažkosť problému pokrývania prahovými hypergrafmi. Okrem toho navrhujeme zovšeobecnenie pojmu Dilworthového čísla grafu na hypergrafy. Prinášame polynomiálny algoritmus na výpočet tohto čísla. Dokazujeme, že Dilworthovo číslo stanovuje hornú hranicu pre niektoré dôležité vlastnosti hypergrafov, ako sú polomer a dominačné číslo hypergrafu.

Keywords: Prahový hypergraph, Pokrývanie hypergrafu, NP-úplnosť