## Horn Formulas, Directed Hypergraphs, Lattices and Closure Systems: Related Formalisms and Applications

Edited by

Kira Adaricheva<sup>1</sup>, Giuseppe F. Italiano<sup>2</sup>, Hans Kleine Büning<sup>3</sup>, and György Turán<sup>4</sup>

- 1 Yeshiva University, US and Nazarbayev University, Kazakhstan adariche@yu.edu and kira.adaricheva@nu.edu.kz
- $\mathbf{2}$ Università di Roma II, IT, italiano@disp.uniroma2.it
- 3 Universität Paderborn, DE, kbcsl@upb.de
- 4 University of Illinois - Chicago, US, gytQuic.edu

#### - Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 14201 "Horn formulas, directed hypergraphs, lattices and closure systems: related formalisms and applications". The seminar brought together researchers working in various areas of mathematics and computer science, mostly in algebra, logic, date base theory, artificial intelligence and data mining. A key objective of the seminar has been to bring together a critical mass of researchers and to provide a platform for personal contacts and scientific interchange between the different disciplines in an atmosphere that will stimulate collaboration and lead to new partnerships. The goal was to crystallize the main research directions and to disseminate challenging open problems across the different research areas.

Seminar May 11-16, 2014 - http://www.dagstuhl.de/14201

1998 ACM Subject Classification F.2 Analysis of Algorithms and Problem Complexity, F.4 Mathematical Logic and Formal Languages, G.2.2 Graph Theory, H.2.8 Database Applications – Data mining, I.2.3 Deduction and Theorem Proving

Keywords and phrases Horn formulas, directed hypergraphs, lattices, closure system, databases, implicational systems and concept analysis

Digital Object Identifier 10.4230/DagRep.4.5.1

#### 1 **Executive Summary**

Kira Adaricheva Giuseppe F. Italiano Hans Kleine Büning György Turán

> License  $\bigcirc$  Creative Commons BY 3.0 Unported license © Kira Adaricheva, Giuseppe F. Italiano, Hans Kleine Büning, and György Turán

## Brief Introduction to the Topic

The Dagstuhl Seminar 14201 on "Horn formulas, directed hypergraphs, lattices and closure systems: related formalisms and applications" was motivated by the growing recognition in the respective research communities that theoretical research and applications of the areas would benefit from increasing the interaction between these fields of research.

Except where otherwise noted, content of this report is licensed

under a Creative Commons BY 3.0 Unported license Horn formulas, directed hypergraphs, lattices and closure systems: related formalisms and applications, Dagstuhl Reports, Vol. 4, Issue 5, pp. 1–26 Editors: Kira Adaricheva, Giuseppe F. Italiano, Hans Kleine Büning, and György Turán

DAGSTUHL Dagstuhl Reports REPORTS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

#### 14201 – Horn Formulas, Directed Hypergraphs, Lattices and Closure Systems

These areas deal with very closely related concepts, but have traditionally been studied within logic, algebra, combinatorics, database theory and artificial intelligence using different techniques and often exploring similar questions with somewhat different emphasis corresponding to the particular area. One of the basic results, the existence of GD-basis, was discovered independently and has different proofs in several areas, such as database theory, the theory of implicational systems and computational learning theory.

The principal objective of the seminar was, as formulated in the seminar announcement, to "bring together a critical mass of researchers and to provide a platform for personal contacts and scientific interchange between the different disciplines in an atmosphere that will stimulate collaboration and lead to new partnerships". In particular, it was hoped that the invitation of a large number of young researchers who would then become familiar with the related research in all the topics discussed, will contribute to the fruitful study of these areas as a more unified discipline in the next generation. Another, related, objective was to help crystallize the main research directions and to disseminate challenging open problems across the different research areas.

## Organization of the Seminar and Activities

The seminar brought together 40 participants working in various areas of mathematics and computer science, mostly in algebra, logic, date base theory, artificial intelligence and data mining. In order to establish the common ground for the discussion and use of terminology, the organizers planned five tutorial talks that were scheduled in the first two days of the seminar. There were the following:

- Endre Boros, in Horn Boolean functions
- Leonid Libkin, in Data Bases
- Marcel Wild, in Closure Systems
- Karell Bertet, in Implicational systems and Concept Analysis
- Georgio Ausiello, in Directed Hypergraphs

It is worth mentioning that two of the tutorial presenters, Endre Boros and Georgio Ausiello, are editors-in-chief of two leading journals that often publish papers associated with the topic of the seminar: Applied Discrete Mathematics and Theoretical Computer Science, respectively. Most other talks of the seminar were related to one or more of these big themes, and they were loosely grouped into sections of presentations in order to stimulate the discussion during regular sessions of the seminar, as well as the break time and follow-up informal meetings. The following grouping gives an approximation to the various topics reflected in the presentations:

- Closures: Duquenne, Wild, Rudolph, Khardon
- Implicational systems: Wild, Adaricheva, Bertet
- Databases: Petit, Libkin
- Directed hypergraphs: Berczi, Nanni, Turán, Ausiello
- Concept Analysis: Obiedkov, Napoli, Kuznetsov
- Horn formulas: Arias, Kučera, Stasi, Čepek
- Applications : Arias, Balcazar, Nation, Bertet
- Knowledge representation: Khardon, Marquis, Sloan
- Constraints: Tamaki
- Probability: Istrate, Balcazar
- Satisfiability: Kleine Büning, Kullmann, Wild

There was a relatively large group of participants attending the seminar without giving a presentation, many of them formed an active audience that initiated discussions and informal meetings. There were several young participants, who associated themselves with one or two presenters, some of them were co-authors in presented results. In some presentations, the participants touched the rare topics which remarkably found active response from the audience. For example, U. Nanni mentioned application of hypergraphs in the modeling of E-Learning. This seems to open a new connection to the theory Knowledge Spaces that were not previously considered by the organizers as the direction of common interests with the current seminar. S. Tamaki presented the overview of recent developments in constraint satisfiability, the topic of the separate Dagstuhl seminar in 2012, which is also familiar to a number of participants of the current seminar. This gave a nice connection to the earlier collective effort of researchers approaching a common research problem with different tools. On Wednesday most participants took part in the countryside walk, the venue of informal exchange of news and views. It was followed by a dinner and a session on open problems. 17 open problems were presented by the participants, some of them referring to the topics of presentations, several young participants introducing new topics. On Thursday night many informal groups were gathering in the after-dinner discussions. During the work of the whole seminar participants took advantage of open access to the comprehensive library, in particular, checking the collection of books authored by participants of the seminar.

## **Concluding Remarks**

We believe that the seminar was very successful in bringing together a critical mass of researchers from different communities and in providing a platform for personal contacts and scientific interchange between the participants. Due to its highly interdisciplinary nature, in order to stimulate collaborations and to foster possible interactions between different research communities, the organizers decided to schedule talks and discussions not only grouped according to topics but also so as to provide a vivid mix of different research questions and results. In particular, the first two days started with introductory talks (tutorials and surveys) delivered by leading experts, while the rest of the seminar included talks presenting current research and applications as well as problem sessions aimed at identifying core research problems coming from the different fields. Besides presentations, the program offered room for open discussions and informal working groups. As a major outcome, a special issue of the journal *Theoretical Computer Science*, co-edited by the organizers, will be devoted to the themes of the seminar. We hope this could serve as a reference material for future interdisciplinary research in the field. Schloss Dagstuhl and its staff provided a very convenient and stimulating environment. The seminar participants appreciated the cordial atmosphere which improved mutual understanding and inspiration. The organizers of this seminar wish to thank all those who helped to make the seminar a fruitful research experience.

Executive Summary	
Kira Adaricheva, Giuseppe F. Italiano, Hans Kleine Büning, and György Turán	1
Overview of Talks	
Effective bases of closure systems <i>Kira V. Adaricheva</i>	6
Query learning of Horn implications: complexity and connections to the GD-basis <i>Marta Arias</i>	6
Logical entailment among partial implications Jose Luis Balcazar	7
From digraphs to dypergraphs: paths and arborescences Kristof Berczi	7
Closure lattices and implication bases Karell Bertet	8
Horn functions: a combinatorial view <i>Endre Boros</i>	11
Shortest CNF representations of Horn functions: complexity issues, decomposition schemes, and lower bounds	19
Some variations on lexicographic enumeration of closed subsets and canonical bases of implications Vincent Duquenne	12
Random Horn satisfiability. Some (old) results and some new(er) surprises. Gabriel Istrate	14
$\operatorname{GFODDs}$ – a representation and calculus for structured functions over possible worlds	
Roni Khardon	14
Closed Sets in Relational Data Roni Khardon	15
Expressive Power of Quantified Horn Formulas Hans Kleine Büning	15
Hydras: Complexity on general graphs and a subclass of trees Petr Kučera	16
Good SAT representations of boolean functions, with applications to XORs Oliver Kullmann	16
On Complexity of Computing Concept Lattices, Implication Bases and Dualization Sergei O. Kuznetsov	17
Databases and lattice theory Leonid Libkin	18
Knowledge compilation and Horn formulae <i>Pierre Marquis</i>	18

## Kira Adaricheva, Giuseppe F. Italiano, Hans Kleine Büning, and György Turán

Directed hypergraphs: a leverage to investigate properties of Petri nets, workflow nets, and more Umberto Nanni	. 18
Formal Concept Analysis, Variations and Applications Amedeo Napoli	. 19
Mathematical Methods for Analyzing Genomic DataJames B. Nation	. 21
Enumeration of minimal dominating sets of a graph and related notions <i>Lhouari Nourine</i>	. 21
Parameterized Ceteris Paribus Preferences over Atomic Conjunctions Sergei Obiedkov	. 22
RQL: a Query Language for Implications      Jean-Marc Petit	. 23
On the Succinctness of Closure Operator Representations Sebastian Rudolph	. 23
The hydra number of a graph Despina Stasi	. 24
The Complexity of Robust Satisfiability of the Constraint Satisfaction Problem (Survey)	94
Some combinatorial problems for directed hypergraphs	. 24
Participants	. 26



## **3.1** Effective bases of closure systems

Kira V. Adaricheva (Yeshiva University – New York, US)

License 
 © Creative Commons BY 3.0 Unported license
 © Kira V. Adaricheva

 Joint work of Adaricheva, Kira V.; Nation, James B.

 Main reference K. V. Adaricheva, J. B. Nation, "On implicational bases of closure systems with unique critical sets," Discrete Applied Mathematics, Vol. 162, pp. 51–69, 2014.
 URL http://dx.doi.org/10.1016/j.dam.2013.08.033

In this talk the overview is presented of recent results inspired by the study of closure systems with unique critical sets. Many of these results, however, are of a more general nature. Among those is the statement that every optimum basis of a finite closure system, in D. Maier's sense, is also right-side optimum. This connects the results given in D. Maier [5] and G. Ausiello et al. in [3], in different frameworks of closure systems.

We introduce the K-basis of a closure system, which is a refinement of the canonical basis of V. Duquenne and J.L. Guigues [4], and discuss a polynomial algorithm to obtain it.

Then we focus on closure systems with unique critical sets, and some subclasses of these where the K-basis is unique. A further refinement in the form of the E-basis is possible for closure systems without D-cycles. The latter basis was first introduced in K. Adaricheva et al. [2], in the family of the ordered direct bases.

One of the main achievements is to demonstrate a polynomial algorithm to recognize the D-relation from a K-basis. Consequently, closure systems without D-cycles can be effectively recognized. While the E-basis achieves an optimum in one of its parts, the optimization of the others is an NP- complete problem. New results are obtained in finding the optimum bases in some classes of convex geometries, K. Adaricheva [1].

#### References

- K. Adaricheva, Optimum basis of finite convex geometry, ArXiv e-prints 2012, available at http://arxiv.org/abs/1205.3236
- 2 K. Adaricheva, J. B. Nation and R. Rand, Ordered direct implicational basis of a finite closure system, Disc. Appl. Math. 161(2013), 707–723.
- **3** G. Ausiello, A. D'Atri and D. Saccá, *Minimal representation of directed hypergraphs*, Journal on computing **15** (1986), 418–431.
- 4 J. L. Guigues and V. Duquenne, Familles minimales d'implications informatives résultant d'une tables de données binares, Math. Sci. Hum. 95 (1986), 5–18.
- 5 D. Maier, Minimum covers in the relational database model, JACM 27 (1980), 664–674.

## 3.2 Query learning of Horn implications: complexity and connections to the GD-basis

Marta Arias (UPC – Barcelona, ES)

License 

 Greative Commons BY 3.0 Unported license

 Marta Arias

 Joint work of Arias, Marta; Balcázar, Jose Luis; Tirnauca, Cristina

 Main reference M. Arias, J. L. Balcázar, "Construction and learnability of canonical Horn formulas," Machine
 Learning Journal, Vol. 85(3):273–297, 2011.

 URL http://dx.doi.org/10.1007/s10994-011-5248-5

In this talk we will review our results on the connections between learning Horn formulas and a minimal representation of Horn implications (the Guigue-Duquenne basis or GD-basis). The talk will cover a review of known upper and lower bounds on the complexity of learning these formulas by presenting a well-known algorithm and the notion of certificates. Then, we will introduce a new learning model based on closure queries which we think captures the essence of the difficulty of learning Horn implications. This is joint work with José L. Balcázar and Cristina Tirnauca.

## 3.3 Logical entailment among partial implications

Jose Luis Balcazar (UPC – Barcelona, ES)

 License © Creative Commons BY 3.0 Unported license
 © Jose Luis Balcazar
 Joint work of Balcázar, Jose Luis; Atserias, Albert
 Main reference J. L. Balcázar, "Redundancy, Deduction Schemes, and Minimum-Size Bases for Association Rules," Logical Methods in Computer Science, 6(2:4), 2010.
 URL http://dx.doi.org/10.2168/LMCS-6(2:4)2010

Relaxed implication connectives are a relatively natural concept. Probabilistic versions defined by resorting to thresholding conditional probability have been proposed in different research communities. Luxemburger introduces them as 'partial implications'; the association rules of the data mining world are partial implications that impose the additional condition that the consequent is a single propositional variable.

One gets easily a logic language of partial implications. For that language, we review known syntactic characterizations of one formula entailing another, and of two formulas entailing another, as well as the handling of additional classical implications, as described in the paper linked to by the following DOI: http://dx.doi.org/10.2168/LMCS-6(2:4)2010.

We report as well on our latest developments towards a full characterization of entailment among partial implications, from arbitrary sets of premises. This progress is obtained, essentially, by recasting the entailment property in the form of a linear program and reasoning on its dual. This part is joint work (unpublished for the time being) with Albert Atserias.

## 3.4 From digraphs to dypergraphs: paths and arborescences

Kristof Berczi (Eötvös Lorand University – Budapest, HU)

License  $\textcircled{\mbox{\scriptsize \ensuremath{\varpi}}}$  Creative Commons BY 3.0 Unported license  $\textcircled{\mbox{\scriptsize \ensuremath{\varpi}}}$  Kristof Berczi

Directed hypergraphs can be defined in many ways. We consider the following definition: a directed hypergraph (or dypergraph for short) is a pair H = (V, E) where V is a finite ground set, E is a family of hyperedges (subsets of V), and each hyperedge e has a designated node called the head of e, while other nodes in e are called tails. A path is an alternating sequence  $v_1, e_1, v_2, e_2, ..., e_k, v_{k+1}$ , where  $v_i$  is a tail node of  $e_i$  and  $v_{i+1}$  is the head of  $e_i$ . A node v is said to be weakly reachable from another node u if there is a path from u to v. The reason for calling this reachability condition 'weak' is that in some applications a different notion of reachability is used. However, an advantage of the present definition is that analogues of Menger's theorems extend naturally.

In this talk, we show how to extend Menger's theorem to dypergraphs. A method for 'cropping' dyperedges is also presented, giveing a powerful tool for handling rooted edge-connectivity problems in dypergraphs.

### 3.5 Closure lattices and implication bases

Karell Bertet (University of La Rochelle, FR)

Formal Concept Analysis (FCA) [14] allows to analyze and extract information from a context – i. e. a binary table – where implicit knowledge between data can be represented either by a concept lattice or implications. In data mining, various symbolic methods stemming from FCA have been studied for (un)supervised classification, association rules extraction and frequent pattern extraction. These methods focus on relevant concepts, closures, or implicational rules and their minimal generators. Implicational system corresponds to the set of representation of Horn clauses of a logic formula, and the closed set lattice is composed of all the true subsets of variables. A closure lattice can also be used as the hierarchical scheme of an ontology. More recently, extensions to more complex data were also proposed as pattern structures in [13], i.e. when there exists a Galois connexion between objects and their space of description. Let us cite extensions to histograms description of objects [25], or logical space description [9]. Thus, implicit knowledge can be extracted from any closure system as contexts, but also implications systems, pattern structures.

The first reference book on lattice theory is due to Birkhoff in 1940 [4], where the lattice structure is introduced as an algebraic structure provided with two operators: the lower and upper bounds. However, in recent works [20, 6], Monjardet establishes that the algebraic lattice structure appeared for the first time in the works of Dedekind in 1900 as the dualgruppe term, and then with several forms and terminologies between 1928 and 1936 in the works of Merge, Klein, Blind, Birkhoff, Öre or Von Neuman. Monjardet also mentions that lattices appeared for the first time in 1847 in its Boolean form in the works of Boole on the algebra of logic, then in the works of Pierce in 1880. The lattice term was proposed by Birkhoff during the first sympsosium on lattices in 1938, while Merge proposed the System von Dingenr term, and Öre the structure term.

In [8, 1], the lattice is defined as an ordered structure (i. e. transitive, antysymetric and reflexive) defined by the existence of particular elements called upper and lower bounds. The notion of irreducible elements [1, 19] of a lattice allow the conception of compact representations of lattices from which it is possible to reconstruct them. The fundamental result of lattice theory, stemming from [1], establishes the bijection between any lattice and the Galois lattice defined from its binary table of irreducible elements. A direct consequence is the existence of bijective links between lattices, reduced contexts (via the table) and a set of implicational rules (via the canonical (direct) basis):

- 1. Any complete lattice is the concept lattice of its binary table; Any reduced context is isomorphic to the irreducible table of its concept lattice.
- 2. Any complete lattice is isomorphic to the closed set lattice of its canonical (direct) basis of implicational rules; Any implicational system on its canonical (direct) form is isomorphic to the canonical (direct) basis of its closed set lattice.

In [1], the authors introduce the Galois lattice term, while the Galois connexion term was defined by Ore in 1944 in [23]. This fundamental bijection is a consequence of the property of a closure operator of any Galois connexion, and can therefore be extended to a closed set lattice [7] and to the very rich closure systems, i. e. sets systems provided with a closure operator. Indeed, we find closure operators in numerous domains, whether in logic, in databases, in combinatoric, or even in data analysis ....

The concept lattice was introduced in the 1990s by Wille within the framework of the Formal Concepts Analysis (FCA), with a reference book dated 1999 [14]. A concept lattice is a graph defined from a binary data (or context) object × attributes. This lattice composed of concepts connected by the generalization/specialisation relation supplies a very intuitive representation of the data. FCA is a knowledge representation approach whose use is increasing this last decade in various domains of computer science, such as data mining, knowledge representation, databases or information retrieval. Indeed, the technological improvements enable to use these structures for large data in these domains though they are exponential in space/time (worst case), and make the development of a large number of applications possible. The need for efficient algorithms to manipulate these structures is a major challenge.

Generation problems of these objects stemming from lattice theory belong to the more general class of problems having an input of size n, and an output of size N bounded by  $2^n$ :

- The closed set lattice and the canonical (direct) basis of a closure system on a set S have a size bounded by  $2^{|S|}$  in the worst case.
- The canonical (direct) basis and the minimal generators of a lattice  $(X, \leq)$  have a size bounded by  $2^{|X|}$  in the worst case.

For this class of problems, a classical worst-case analysis makes them exponential, thus NP-complete. However, a more precise estimation can be obtained from output-sensitive analysis where the time complexity per element – i. e. one implication or one closed set – is estimated either by the amortized complexity or the delay complexity. A polynomial delay algorithm is then an algorithm with a delay cost bounded by a polynomial. In lattice theory, generation problems with an exponential output are mainly resolved:

#### Closed set lattice generation from a closure system: Polynomial delay problem

Numerous polynomial delay algorithms have been proposed in the literature [21, 12, 5, 15, 22]. Ganter's NextClosure [12] is the reference algorithm that computes the closures lecticaly ordered (which can then be ordered to form the lattice) while Bordat's algorithm [5] is the first algorithm that computes directly the Hasse diagram of the lattice.

#### Canonical direct basis generation from a closure system: Open problem

Let us cite the algorithm of Mannila in [18], where the canonical direct basis is generated with an exponential time per implication in the worst case. It is based on the generation of all minimal transverses, a problem known to be an open problem (there actually exists a quasi- polynomial algorithm to solve it). More generally, the canonical direct basis can be obtained from the closed set lattice of the closure system, with the generation of all closed set as an additional cost, thus an exponential delay complexity. In the particular case where the closure system is an implicational system, there exist specific algorithms, always with the same exponential delay complexity. The algorithms described in [27] and [3] compute an intermediate but larger implicational system of exponential size in the worst case. The algorithm with the best known complexity, in logic area, is due to Fredman and Khachiyan ([11]). It generates one implication with a quasi-polynomial time and has been modified in [17] to solve this problem with a first step in deterministic polynomial time, followed by  $O(log^2|S|)$  non deterministic steps.

### 10 14201 – Horn Formulas, Directed Hypergraphs, Lattices and Closure Systems

#### Canonical direct basis generation from a lattice: Polynomial amortized problem

In the area of data mining the algorithm of Taouil and Bastide in [26] generates the canonical direct basis with an exponential complexity per implication. The incremental algorithm of Pfaltz [24] and Jen algorithm [10] both compute the minimal generators – i. e. premises of the canonical direct basis – of a lattice with the same complexity. However, in logic area, the algorithm attributed to Ibaraki et al. ([16]) computes an implication in polynomial time with a set of closures as input. The algorithm in [2] is similar to those in [16]. It computes the dependency graph of a lattice, and then extracts the canonical direct basis from this graph.

Equivalent notions such as lattices (or closed set lattices), closure systems, closure operators (or dual closure operators), (pure) Horn functions have been studied by different authors in different domains (topology, lattice theory, hypergraph theory, choice functions, relational databases, data mining and concept analysis, artificial intelligence and expert systems, knowledge spaces, logic and logic programming, theorem proving...). It is not surprising that one finds the same notions, results or algorithms under various names. One can also find many original results or algorithms only known in a specific domain. It would be very profitable to increase (or create) the communications between the various domains that use the same (or equivalent) notions and tools.

#### References

- 1 M. Barbut and B. Monjardet. Ordres et classifications : Algèbre et combinatoire. Hachette, Paris, 1970. 2 tomes.
- 2 K. Bertet. The dependence graph of a lattice. In Ninth International Conference on Concept Lattices and their Applications (CLA'12), Malaga, Spain, October 2012.
- 3 K. Bertet and M. Nebut. Efficient algorithms on the family associated to an implicationnal system. *Discrete Mathematical and Theoretical Computer Science*, 6:315–338, 2004.
- 4 G. Birkhoff. *Lattice theory*. American Math. Society, 1st edition, 1940.
- 5 J.P. Bordat. Calcul pratique du treillis de Galois d'une correspondance. Math. Sci. Hum., 96:31–47, 1986.
- 6 N. Caspard, B. Leclerc, and B Monjardet. *Ensembles ordonnées finis: concepts, résultats et usages.* Springer, 2007.
- 7 N. Caspard and B. Monjardet. The lattice of closure systems, closure operators and implicational systems on a finite set: a survey. *Discrete Applied Mathematics*, 127(2):241–269, 2003.
- 8 B.A. Davey and H.A. Priestley. *Introduction to lattices and orders*. Cambridge University Press, 2nd edition, 1991.
- 9 S. Ferré and O. Ridoux. An introduction to logical information systems. Information Processing & Management, 40(3):383-419, 2004.
- 10 A. Le Floch, C. Fisette, R. Missaoui, P. Valtchev, and R. Godin. Jen : un algorithme efficace de construction de generateurs minimaux pour l'identication de regles d'association. Nouvelles Technologies de l'Information, 1(1):135–146, 2003.
- 11 M.L. Fredman and L. Khachiyan. On the complexity of dualization of monotone disjunctive normal forms. *Journal of Algorithms*, 21:618–628, 1996.
- 12 B. Ganter. Two basic algorithms in concept analysis. *Technische Hochschule Darmstadt* (*Preprint 831*), 1984.
- 13 B. Ganter and S.O. Kuznetsov. Pattern structures and their projections. In In LNCS of International Conference on Conceptual Structures (ICCS'01), pages 129–142, 2001.
- 14 B. Ganter and R. Wille. Formal concept analysis, Mathematical foundations. Springer Verlag, Berlin, 1999.

- 15 R. Godin, R. Missaoui, and H. Alaoui. Learning algorithms using a Galois lattice structure. *Third International Conference on Tools for Artificial Intelligence, San Jose, California*, pages 22–29, 1991.
- 16 T. Ibaraki, A. Kogan, and K. Makino. Functional dependencies in Horn theories. Artificial Intelligence, 108:1–30, 1999.
- D.J. Kavvadias and E.C. Stravropoulos. Monotone boolean dualization is in co-np[log<sup>2</sup>n]. Information Processing Letter, 85:1–6, 2003.
- 18 H. Mannila and K.J. Räihä. The design of relational databases. Addison-Wesley, 1992.
- 19 G. Markowsky. Some combinatorial aspects on lattice theory. In *Proceedings of Houston lattice theory conference*, pages 36–68, 1973.
- 20 B. Monjardet. La construction des notions d'ordre et de treillis. In Journées nationales de l'Association des Professeurs de Mathématiques de l'Enseignement Public (APMEP), La Rochelle, France, October 2008.
- 21 E. Norris. An algorithm for computing the maximal rectangles in a binary relation. *Revue Roumaine de Math. Pures et Appliquées*, 23(2), 1978.
- 22 L. Nourine and O. Raynaud. A fast algorithm for building lattices. Information Processing Letters, 71:199–204, 1999.
- 23 O. Öre. Galois connexion. Trans. of the Amer. Math. Society, 55, 1944.
- 24 JL. Pfaltz and CM. Taylor. Scientific discovery through iterative transformations of concept lattices. In Workhop on Discrete Applied Mathematics, in conjonction with the 2<sup>nd</sup> SIAM International Conference on Data-Mining, pages 65–74, 2002.
- 25 G. Polaillon. Organisation et interprétation par les treillis de Galois de données de type multivalué, intervalle ou histogramme. PhD thesis, Paris IX-Dauphine, 1998.
- 26 R. Taouil and Y. Bastide. Computing proper implications. In 9<sup>th</sup> International Conference on Conceptual Structures, Stanford, USA, 2002.
- 27 M. Wild. Computations with finite closure systems and implications. In Proceedings of the 1st Annual International Conference on Computing and Combinatorics, volume 959 of LNCS, pages 111–120. Springer, 1995.

#### 3.6 Horn functions: a combinatorial view

Endre Boros (Rutgers University – Piscataway, US)

License  $\bigcirc$  Creative Commons BY 3.0 Unported license

© Endre Boros

Joint work of Boros, Endre; Čepek, Ondřej; Crama, Yves; Gruber, Aritanan; Hammer, Peter L.; Kogan,

Alexander; Kučera, Petr

In this talk basic properties of Horn functions and their representations are surveyed. Horn functions are viewed as 0-1 valued real functions, and representations in terms of their implicants (implicates) are considered. The space of implicants and implicates are viewed as abstract sets, in which consensus and resolution provides some order. We recalled results about decomposability of the Horn minimization problem in terms of exclusive sets and components. We provided constructive descriptions via forward chaining and syntactical properties for such exclusive families of implicants.

These properties provided the bases for the efficient (polynomial delay) generation of ALL prime implicants/implicates of a Horn DNF/CNF (equivalent to generating the direct implication basis of an implication system).

Main reference
 E. Boros, "Horn Functions," in Y. Crama and P.L. Hammer, eds., "Boolean Functions: Theory, Algorithms, and Applications", Cambridge University Press, 2011.

#### 12 14201 – Horn Formulas, Directed Hypergraphs, Lattices and Closure Systems

We also recalled an associated graph, the so called implication graph of a Horn function, and showed that the structure of this graph provides additional exclusive subfamilies of exclusive sets. Based on such exclusive sets and the resulting decomposition ideas, one can simplify Horn minimization and also can simplify the proof of correctness of some hardness results.

Finally, we recalled recent hardness results about Horn DNF/CNF minimization and Horn literal minimization, namely that unless P=NP, one cannot approximate within a factor of  $2^{\log^{1-o(1)}n}$  the term/literal minimum of a Horn DNF (and the same applies to clause/literal minimum Horn CNF-s), even if the input is restricted to 'short' cubic Horn DNF/CNF-s.

This talk was based on joint research with O. Čepek, Y. Crama, A. Gruber, P. L. Hammer, A. Kogan, and P. Kučera.

#### References

- 1 Boros, E., *Horn Functions*, In: Boolean Functions: Theory, Algorithms, and Applications (Y. Crama and P. L. Hammer, eds.) Cambridge University Press, 2011.
- 2 E. Boros, Y. Crama, and P. L. Hammer. Polynomial-time inference of all valid implications for Horn and related formulae. Annals of Mathematics and Artificial Intelligence, 1:21–32, 1990.
- 3 E. Boros, O. Čepek, and A. Kogan. Horn minimization by iterative decomposition. Annals of Mathematics and Artificial Intelligence, 23(3-4):321–343, September 1998.
- 4 E. Boros. Maximum renamable Horn sub-CNFs. Discrete Applied Mathematics, 96–97(1– 3):29-40, October 1999.
- 5 E. Boros, O. Čepek, A. Kogan, and P. Kučera: Exclusive and essential sets of implicates of Boolean functions. Discrete Applied Mathematics 158 (2), 81–96, 2010.
- **6** E. Boros, O. Čepek, and P. Kučera: A decomposition method for CNF minimality proofs, Theoretical Computer Science, 510:111–126, 2013.
- 7 E. Boros and A. Gruber: Hardness Results for Approximate Pure Horn CNF Formulae Minimization. Annals of Mathematics and Artificial Intelligence, accepted, 2014.

# 3.7 Shortest CNF representations of Horn functions: complexity issues, decomposition schemes, and lower bounds

Ondřej Čepek (Charles University – Prague, CZ)

License © Creative Commons BY 3.0 Unported license © Ondřej Čepek

Joint work of Boros, Endre; Čepek, Ondřej; Kogan, Alex; Kučera, Petr; Savický, Petr

Main reference E. Boros, O. Čepek, A. Kogan, P. Kučera, "Exclusive and essential sets of implicates of Boolean functions," Discrete Applied Mathematics, Vol. 158, Issue 2, pp. 81–96, 2010.

URL http://dx.doi.org/10.1016/j.dam.2009.08.012

Finding a shortest CNF representation of a given Horn function is a problem with many practical applications in artificial intelligence (knowledge compression) and other areas of computer science. In this talk we briefly survey complexity results known for this problem and then concentrate on the relationships between CNF representations of Horn functions and certain sets of implicates of these functions. We introduce two definitions of sets of implicates which are both based on the properties of resolution. The first type of sets, called exclusive sets of implicates, is shown to have a functional property useful for decompositions which allow to reduce the minimization problem to smaller subCNFs. The second type of sets, called essential sets of implicates, has an orthogonality property, which implies that every CNF representation and every essential set must intersect. This property leads to interesting lower bounds on the CNF size. Given the topic of the Dagstuhl seminar, it would be interesting to see whether concepts similar to exclusive nad essential sets of implicates were studied in the context of directed hypergraphs or closure systems.

## 3.8 Some variations on lexicographic enumeration of closed subsets and canonical bases of implications

Vincent Duquenne (UPMC - Paris, FR)

License <br/>  $\textcircled{\mbox{\scriptsize C}}$  Creative Commons BY 3.0 Unported license<br/>  $\textcircled{\mbox{\scriptsize C}}$  Vincent Duquenne

Since their promotion in 1984 [3], many efforts have been done to overtake the so-called NextCLOSURE algorithms, which are applying natural lexicographic enumeration [7]. More than going faster (who cares getting more than 105/s on a 200E laptop?), a hidden motivation could have been to try escaping their simplicity ... These attempts have been partly successful to scan through intents (closed subsets) with some costs. For the canonical basis of implications based on pseudo-intents [2, 4] the fees are a bit higher and more than often – if not always – involve an exponential explosion in memory [5, 6]. Here we come back to mere lexicographic enumeration with some specific tricks to speed up the process, which can be adapted to other related objects. Hence with a special dedication to ecologically minded users, we will try to give some visions of what could be next in doing with boosted basic algorithms.

Keywords: closure operator, lattice, combinatorial exhaustive enumeration, NextCLOSURE, (quasi / pseudo) intents, canonical basis of implications.

#### References

- Duquenne, V.: Some Variations on Alan Day's Algorithm for Calculating Canonical Basis of Implications. In Diatta, J., Ecklund, P., Liquière, M. (eds.), CLA-2007, 17–25.
- 2 Duquenne, V.: Contextual implications between attributes and some representation properties for finite lattices. In Ganter, B., Wille, R., Wolf, K. (eds.), Beitrage zur Begriffsanalyse, 213-239, Wissenschaftsverlag, Mannheim: (1987), reprinted in ICFCA-2013.
- 3 Ganter, B.: Algorithmen zur Formalen Begriffsanalyse. In Ganter, B., Wille, R., Wolf, K. (eds.), Beitrage zur Begriffsanalyse, Wissenschaftsverlag, Mannheim: 241–255 (1987).
- 4 Guigues J.L., Duquenne.V.: Familles minimales d'implications informatives résultant d'un tableau de données binaires. Mathématiques & Sciences Humaines 95, 5–18, (1986).
- 5 Obiedkov S., Duquenne, V.: Attribute incremental construction of the canonical basis, Annals of Mathematics and Artificial Intelligence 49, 77–99, (2007).
- 6 Valtchev, P., Duquenne, V.: On the merge of factor canonical bases. In Medina, R., Obiedkov, S. (eds), ICFCA 2008, LNAI 4933, 182–198 (2008).
- 7 Nijenhuis, A., Wilf, H.S.: *Combinatorial algorithms*, Academic Press, second ed., 1978. http://www.math.upenn.edu/~wilf/website/CombinatorialAlgorithms.pdf

#### 14201 – Horn Formulas, Directed Hypergraphs, Lattices and Closure Systems

## **3.9** Random Horn satisfiability. Some (old) results and some new(er) surprises.

Gabriel Istrate (West University of Timisoara, RO)

14

We will overview results on thresholds in random versions of satisfiability, applications to random tree automata emptiness, and the way random Horn satisfiability is important in a classification (similar in spirit to Schaefer's dichotomy theorem) of thresholds of generalized satisfiability problems in a model based on clausal formulas. We will also discuss the connection of this classification to recently investigated the problem of random mixed Horn satisfiability.

The talk is based on several papers, some of them authored solely, one of them in cooperation with Cristopher Moore and Moshe Vardi, as well as some unpublished results.

## 3.10 GFODDs – a representation and calculus for structured functions over possible worlds

Roni Khardon (Tufts University, US)

License 
 © Creative Commons BY 3.0 Unported license
 © Roni Khardon

 Joint work of Khardon, Roni; Saket, Joshi; Kersting, Kristian; Wang, Chenggang; Schermerhorn, Paul; Scheutz, Matthias; Raghavan, Aswin; Fern, Alan; Tadepalli, Prasad
 URL http://www.cs.tufts.edu/research/ml/index.php?op=res\_full&id=5

The Generalized first order decision diagrams (GFODD) representation, captures numerical generalizations of first order logic where truth values are generalized to real values and logical quantifiers are generalized to aggregation functions over such values. GFODDs can be seen as expressions representing functions over possible worlds (interpretations) and can therefore represent probability functions, value functions etc. A calculus of expressions allows us to perform operations over the functions they represent and perform inference in this way. A heuristic, using reasoning from examples, allows us to implement inference reasonably efficiently and therefore support complex patterns of inference. The talk will give an overview of this work motivating the representation, discussing complexity, explaining heuristics, and illustrating our main application in solving structured Markov decision problems.

Based on joint work in multiple papers with Saket Joshi, Kristian Kersting, Chenggang Wang, Paul Schermerhorn, Matthias Scheutz, Aswin Raghavan, Alan Fern, Prasad Tadepalli.

A complexity based treatment can be found in [1]. For other papers please see our project site (see URL above) for online versions of the papers.

#### References

1 Benjamin J. Hescott and Roni Khardon, *The Complexity of Reasoning with FODD and GFODD*, Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2014 http://www.cs.tufts.edu/~roni/PUB/aaai2014gfcomplexity.pdf

## 3.11 Closed Sets in Relational Data

Roni Khardon (Tufts University, US)

License 
 © Creative Commons BY 3.0 Unported license
 © Roni Khardon

 Joint work of Khardon, Roni; Garriga, Gemma ; De Raedt, Luc

 Main reference G. C. Carriga, R. Khardon, L. De Raedt, "Mining closed patterns in relational, graph and network
 data," Annals of Mathematics and Artificial Intelligence, 69(4):315–342, 2013.

 URL http://dx.doi.org/10.1007/s10472-012-9324-8

I this talk I reviewed our work from IJCAI 2007 and AMAI 2013 that investigates semantic properties of closed sets in multi-relational datasets, and the implications for algorithmic properties. Prior literature has multiple definitions for the notions of occurrence of a pattern and the notion of closure relative to such occurrences. Our work shows that these implicit choices are important because they yield dramatically different properties, resulting closed sets, and algorithms.

Based on joint work with Gemma Garriga and Luc De Raedt.

Note: Given the strong interest and focus on closed sets in the workshop I added this second brief talk to complement the discussion.

#### References

- 1 G. Garriga, R. Khardon and L. De Raedt, Mining Closed Patterns in Relational, Graph and Network Data, Annals of Mathematics and Artificial Intelligence, Vol 69, Issue 4, pp. 315-342, 2013
- 2 G. Garriga, R. Khardon and L. De Raedt, On Mining Closed Sets in Multi-Relational Data, In the proceedings of the International Joint Conference on Artificial Intelligence., 2007

### 3.12 Expressive Power of Quantified Horn Formulas

Hans Kleine Büning (Universität Paderborn, DE)

License © Creative Commons BY 3.0 Unported license © Hans Kleine Büning Joint work of Kleine Büning, Hans; Bubeck, Uwe; Zhao, Xishun

In this talk we give an overview of the expressive power of quantified Horn formulas and related quantified Boolean formulas. Additionally, we present the computational complexities of the satisfiability problem for various classes and the structure of satisfying models. These are Skolem functions for the existential quantifiers satisfying the formula. Another application is the linear time transformation of propositional formulas into satisfiability equivalent formulas in CNF by means of existentially quantified CNF formulas with free variables where the bounded part of the clauses are 2-Horn clauses.

Based on joint work with Uwe Bubeck and Xishun Zhao.

## 3.13 Hydras: Complexity on general graphs and a subclass of trees

Petr Kučera (Charles University – Prague, CZ)

License O Creative Commons BY 3.0 Unported license

© Petr Kučera

Main reference P. Kučera, "Hydras: Complexity on general graphs and a subclass of trees," in Proc. of the Int'l Symp. on Artificial Intelligence and Mathematics (ISAIM'14), 2014.
 URL http://www.cs.uic.edu/pub/Isaim2014/WebPreferences/ISAIM2014\_Boolean\_Kucera.pdf

Hydra formulas were introduced in [1]. A hydra formula is a Horn formula consisting of definite Horn clauses of size 3 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. The minimal formula size for hydras is then called the hydra number of the underlying graph. In this paper we aim to answer some open questions regarding complexity of determining the hydra number of a graph which were left open in [1]. In particular we show that the problem of checking, whether a graph G = (V, E) is single-headed, i. e. whether the hydra number of trees and we describe a family of trees for which the hydra number can be determined in polynomial time.

#### References

 R. H. Sloan, D. Stasi, G. Turán, Hydras: Directed hypergraphs and horn formulas, in: M. C. Golumbic, M. Stern, A. Levy, G. Morgenstern (Eds.), WG, Vol. 7551 of Lecture Notes in Computer Science, Springer, 2012, pp. 237–248.

## 3.14 Good SAT representations of boolean functions, with applications to XORs

Oliver Kullmann (Swansea University, GB)

License Creative Commons BY 3.0 Unported license

© Oliver Kullmann Joint work of Kullmann, Oliver; Gwynne, Matthew

- Main reference M. Gwynne, O. Kullmann, "On SAT representations of XOR constraints," in Proc. of the 8th Int'l Conf. on Language and Automata Theory and Applications (LATA'14), LNCS, Vol. 8370, pp. 409–420, Springer, 2014.
  - URL http://dx.doi.org/10.1007/978-3-319-04921-2 33
  - URL http://www.cs.swan.ac.uk/ csoliver/papers.html#DAGSTUHL2014a

An approach towards measuring the quality of SAT representation is outlined, via various 'hardness measurements'. The application to XOR-constraints is considered, showing that there is no polysize 'good' representation in general, while finding such a "good" representation is fixed-parameter tractable in the number of constraints. 'Good' here means 'UR-representation', that is, every partial instantiation to the variables of the boolean function represented, which destroys all satisfying assignments, is recognised by unit-clause propagation. We consider also stronger positive results, using unit-propagation completeness.

## 3.15 On Complexity of Computing Concept Lattices, Implication Bases and Dualization

Sergei O. Kuznetsov (NRU Higher School of Economics – Moscow, RU)

License 

 © Creative Commons BY 3.0 Unported license
 © Sergei O. Kuznetsov

 Joint work of Kuznetsov, Sergei O.; Babin, Mikhail A.; Obiedkov, Sergei A.
 Main reference M. .A. Babin, S. O. Kuznetsov, "Computing premises of a minimal cover of functional dependencies is intractable," Discrete Applied Mathmatics, Vol. 161, Issue 6, pp. 742–749, 2013.
 URL http://dx.doi.org/10.1016/j.dam.2012.10.026

Most important computation problems of lattices and related implication systems are discussed: computing a concept lattice from context (i. e., computing a lattice from the sets of its irreducible elements), computing cardinality-minimal implication base (i. e., minimum Horn theory or minimum cover of functional dependencies) related to the lattice. Learning from positive and negative examples in a concept lattice is shown to be equivalent to a dualization problem on a lattice given implicitly by the ordered set of its irreducible elements. Efficient lazy classification is proposed as an alternative to the direct usage of implications as classification rules. The positive results are shown to be extendible to lattices of arbitrary closed descriptions.

#### References

- S. O.Kuznetsov, Mathematical aspects of concept analysis. J.Math. Sci., vol. 80(2), pp. 1654– 1698 (1996).
- 2 S. O. Kuznetsov, On Computing the Size of a Lattice and Related Decision Problems. Order, vol. 18, no. 4, pp. 313-321 (2001).
- 3 S.O. Kuznetsov, On the Intractability of Computing the Duquenne- Guigues Base. J. Univ. Comp. Sci., vol. 10, no. 8, pp. 927–933 (2004).
- 4 B. Ganter, P.A. Grigoriev, S.O. Kuznetsov, and M.V. Samokhin, *Concept-based Data Mining with Scaled Labeled Graphs.* In: K.E. Wolff, H. D. Pfeiffer, H. S. Delugach, Eds., Proc. 12th International Conference on Conceptual Structures (ICCS 2004), Lecture Notes in Artificial Intelligence (Springer), Vol. 3127, pp. 94–108, 2004.
- 5 S.O. Kuznetsov, Complexity of Learning in Concept Lattices from Positive and Negative Examples. Discr. Appl. Math., vol. 142, pp. 111–125 (2004).
- 6 S. O. Kuznetsov, S. A. Obiedkov, Some Decision and Counting Problems of the Duquenne-Guigues Basis of Implications. Discr. Appl. Math., vol. 156, no. 11, pp. 1994–2003 (2008).
- 7 F. Distel, B. Sertkaya, On the complexity of enumerating pseudo- intents. Discr. Appl. Math. vol. 159(6), pp. 450–466 (2011).
- 8 M. A. Babin, S. O. Kuznetsov, Enumeration Minimal Hypotheses and Dualizing Monotone Boolean Functions on Lattices. In: Proc. 9th ICFCA, LNAI, Vol. 6628, pp. 42–48, 2011.
- 9 M. A. Babin, S. O. Kuznetsov, Computing premises of a minimal cover of functional dependencies is intractable. Discr. Appl. Math., vol. 161(6), pp. 742–749 (2013).
- 10 Sergei O. Kuznetsov, Scalable Knowledge Discovery in Complex Data with Pattern Structures. In: P. Maji, A. Ghosh, M. N. Murty, K. Ghosh, S. K. Pal, Eds., Proc. 5th International Conference Pattern Recognition and Machine Intelligence (PReMI'2013), Lecture Notes in Lecture Notes in Computer Science (Springer), Vol. 8251, pp. 30–41, 2013.
- 11 Sergei O. Kuznetsov, Fitting Pattern Structures for Knowledge Discovery in Big Data. In: Cellier, Peggy; Distel, Felix; Ganter, Bernhard, Eds., Proc. 11th International Conference on Formal Concept Analysis (ICFCA 2013), Lecture Notes in Artificial Intelligence (Springer), Vol. 7880, pp. 254–266, 2013.

## 3.16 Databases and lattice theory

Leonid Libkin (University of Edinburgh, GB)

In this talk we surveyed recent database applications that used lattice-theoretic techniques, specifically the lattice of graph cores ordered by homomorphisms between them. We explained why these are important in the study of conjunctive queries, and concentrated on two applications: approximation of queries by efficient ones, and evaluation of queries over databases with incomplete information.

### 3.17 Knowledge compilation and Horn formulae

Pierre Marquis (Artois University – Lens, FR)

License 
 © Creative Commons BY 3.0 Unported license
 © Pierre Marquis

 Main reference P. Marquis, "Existential Closures for Knowledge Compilation," in Proc. of the 22nd Int'l Joint Conference on Artificial Intelligence (IJCAI'11), pp. 996–1001, IJCAI/AAAI, 2011.
 URL https://www.aaai.org/ocs/index.php/IJCAI/IJCAI11/paper/view/3026/

In this talk I will focus on the concept of disjunctive closure of the language of (renamable) Horn CNF formulae, i.e., the class of existentially quantified disjunctions of (renamable) propositional Horn CNF formulae. I will analyze the corresponding languages L along the lines of the knowledge compilation map, i.e. I will make precise the queries and transformations of interest feasible in polynomial time (and those which are not unless P = NP) when the input is in L and I will also compare the succinctness of L with those of other propositional languages.

# 3.18 Directed hypergraphs: a leverage to investigate properties of Petri nets, workflow nets, and more

Umberto Nanni (University of Rome "La Sapienza", IT)

License O Creative Commons BY 3.0 Unported license

© Umberto Nanni

Joint work of Alimonti, Paola; Feuerstein, Esteban; Laura, Luigi; Nanni, Umberto

Main reference P. Alimonti, E. Feuerstein, L. Laura, U. Nanni, "Linear time analysis of properties of conflict-free and general Petri nets," Theoretical Computer Science, Vol. 412, Issues 4–5, pp. 320–338, 2011.

URL http://dx.doi.org/10.1016/j.tcs.2010.09.030

Petri nets have been introduced in 1962 by Carl Adam Petri to model the dynamic properties of a concurrent system, and have enjoyed a wide success, with application in a number of realms – far beyond computing. A Petri net is a 4-tuple  $\langle P, T, A, M0 \rangle$  where: P is the set of places, T is the set of transitions, A is the set of arcs (from places to transitions, or vice-versa, i. e.,  $\langle P, T, A \rangle$  is a bipartite graph), and M0 is the initial marking which can be expressed as a function from places to natural numbers; you can figure out the marking of a net as a distribution of "tokens" over the places. The marking of a Petri net, representing the state of the net, can evolve from the initial state by firing transitions. Most problems on Petri nets (a bit less expressive than Turing machines) are decidable, although with high computational complexity. We introduce the notion of T-path within Petri nets, and adopt directed hypergraphs in order to determine structural properties of nets. In particular, we study the relationships between T-paths and firable sequences of transitions. Let us consider a Petri net P, with initial marking M0, and the set of places with a positive marking in M0, i. e., P0 = p|M0(p) > 0. It can be shown that the existence of a T-path from any subset of P0 to a transition t is a necessary condition for the potential firability of t and, furthermore: (a) if P is a Conflict Free Petri net, this is also a sufficient condition, (b) if P is a general Petri net, t is potentially firable "by augmentation" of the initial marking (i. e., by increasing the number of tokens in P0). For the class of Conflict-Free nets (CFPN), using the notion of T-path, we show that there exist simple algorithms requiring linear space and time for the following problems: (a) determining the set of firable transitions, (b) determining the set of reachable places, (c) determining the set of live transitions, (d) deciding the boundedness of the net. All these problems, untractable on general Petri nets, have been solved previously in CF-nets by matrix operations in quadratic time in several papers by Howell, Rosier, Yen.

Given a Petri net and a marking M, the well-known coverability problem consists in finding a reachable marking M' such that M' >= M. We introduce a relaxed form of this problem: we say that a marking M is "coverable by augmentation" if it is coverable from an augmented marking M0+ of the initial marking M0: M0+ >= M0 and, for any place p, M0 + (p) = 0 if M0(p) = 0. The coverability problem is known to be EXPSPACE-hard [Rackoff, 1978]; we solve the relaxed problem of coverability by augmentation in linear time in general Petri nets.

Workflow nets are the results of adopting Petri nets for modeling Workflow Management [van der Aalst, 1997]. This seminal approach has boosted the adoption of these structures within the context of Business Process Management, with impact on methodologies with high industrial relevance (UML, BPMN). We show as adopting the notion of augmented marking allow us to investigate properties of Workflow nets.

More recent applications of directed hypergraphs in the context of eLearning show that it is possible to build personalized learning paths in linear time within a collection of learning components: this solution tackle the problem of managing and reusing learning objects stored in huge reportiories where they have been accumulated in the range of millions.

### 3.19 Formal Concept Analysis, Variations and Applications

Amedeo Napoli (LORIA & INRIA Nancy, FR)

 $\mbox{License} \ensuremath{\textcircled{\mbox{\scriptsize \mbox{$\odot$}}}}\xspace$ Creative Commons BY 3.0 Unported license  $\ensuremath{\textcircled{\mbox{$\odot$}}}\xspace$ Amedeo Napoli

In Formal Concept Analysis (FCA [8]), the formalization of a classification problem relies on a formal context K = (G, M, I) where G is a set of objects, M a set of attributes and  $I \subseteq G \times M$  a binary relation describing links between objects and attributes. Then a formal concept corresponds to a maximal set of objects – the extent – associated with a maximal set of attributes – the intent. Formal concepts are ordered within a complete lattice thanks to a subsumption relation based on extent inclusion.

The standard FCA formalism can be extended to deal with relational and complex data. Relational Concept Analysis (RCA [12]) is able to take into account relational data, i.e. relations between objects while the basic data structure in FCA relies on object and attribute relationships. Pattern Structures [7] provide an extension of Formal Concept Analysis for dealing with complex data such as numbers, intervals, sequences, trees and graphs. Pattern

#### 20 14201 – Horn Formulas, Directed Hypergraphs, Lattices and Closure Systems

Structures are based on a triple  $(G, (D, \sqcap), \delta)$ , where G is a set of objects,  $(D, \sqcap)$  is a semilattice of descriptions, and  $\delta$  is a mapping associating an object with a description. The similarity operation  $\sqcap$  induces a subsumption relation in  $(D, \sqcap)$  such as  $c \sqcap d = c$  iff  $c \sqsubseteq d$ . In addition, a similarity between objects based on the closeness of attribute values can be considered and formalized as a tolerance relation, i.e. reflexive and symmetric [9].

In this presentation, we will introduce basic elements on Formal Concept Analysis (FCA), Relational Concept Analysis and Pattern Structures. Then we will discuss the capabilities of RCA and Pattern Structures in various applications. RCA is used in text mining while Pattern Structures are used in situations involving real-world data for dealing with numbers and intervals [11, 10], with strings and sequences [3], for discovering functional dependencies [2, 1], for information retrieval [5] and for ontology engineering [6, 4]. Actually, FCA, RCA and Pattern Structures are become an efficient and well-founded mathematical support for knowledge discovery in real-world applications.

*Keywords:* Formal Concept Analysis, pattern structures, relational concept analysis, classication, knowledge discovery, complex data mining.

#### References

- Jaume Baixeries, Mehdi Kaytoue, and Amedeo Napoli. Computing Similarity Dependencies with Pattern Structures. In Manuel Ojeda-Aciego and Jan Outrata, editors, The Tenth International Conference on Concept Lattices and their Applications – CLA 2013, La Rochelle, France, pages 33–44. Université de La Rochelle, 2013. CEUR Workshop Proceedings 1062.
- 2 Jaume Baixeries, Amedeo Napoli, and Mehdi Kaytoue. Functional Dependencies and Pattern Structures. In Uta Priss and Laszlo Szathmary, editors, The Nineth International Conference on Concept Lattices and their Applications – CLA 2012, pages 175–186, Fuengirola, Spain, 2012. University of Malaga.
- 3 Aleksey Buzmakov, Elias Egho, Nicolas Jay, Sergei O. Kuznetsov, Amedeo Napoli, and Chedy Raïssi. On Projections of Sequential Pattern Structures (with an application on care trajectories). In Manuel Ojeda-Aciego and Jan Outrata, editors, The Tenth International Conference on Concept Lattices and their Applications – CLA 2013, La Rochelle, France, pages 199–208. Université de La Rochelle, 2013. CEUR Workshop Proceedings 1062.
- 4 Melisachew Wudage Chekol, Mehwish Alam, and Amedeo Napoli. A Study on the Correspondence between FCA and ELI Ontologies. In Manuel Ojeda-Aciego and Jan Outrata, editors, The Tenth International Conference on Concept Lattices and their Applications CLA 2013, La Rochelle, France, pages 237–248. Université de La Rochelle, 2013. CEUR Workshop Proceedings 1062.
- 5 Víctor Codocedo, Ioanna Lykourentzou, and Amedeo Napoli. A semantic approach to Concept Lattice-based Information Retrieval. Annals of Mathematics and Artificial Intelligence, 2014. To be published.
- 6 Adrien Coulet, Florent Domenach, Mehdi Kaytoue, and Amedeo Napoli. Using pattern structures for analyzing ontology-based annotations. In Peggy Cellier, Felix Distel, and Bernhard Ganter, editors, 11h International Conference on Formal Concept Analysis (ICFCA 2013), Lecture Notes in Artificial Intelligence LNAI 7880, pages 76–91. Springer, Berlin, 2013.
- 7 Bernhard Ganter and Sergei O. Kuznetsov. Pattern structures and their projections. In H.S. Delugach and G. Stumme, editors, Conceptual Structures: Broadening the Base, Proceedings of the 9th International Conference on Conceptual Structures, ICCS 2001, Stanford, CA, Lecture Notes in Computer Science 2120, pages 129–142. Springer, 2001.
- 8 Bernhard Ganter and Rudolf Wille. Formal Concept Analysis. Springer, Berlin, 1999.
- **9** Mehdi Kaytoue, Zainab Assaghir, Amedeo Napoli, and Sergei O. Kuznetsov. *Embedding* Tolerance Relations in Formal Concept Analysis – An Application in Information Fusion.

In Jimmy Huang, Nick Koudas, Gareth Jones, Xindong Wu, Kevyn Collins-Thompson, and Aijun An, editors, *Proceedings of the 19th ACM Conference on Information and Knowledge Management, CIKM 2010, Toronto, Ontario, Canada, October 26-30, 2010*, pages 1689–1692. ACM, 2010.

- 10 Mehdi Kaytoue, Sergei O. Kuznetsov, and Amedeo Napoli. Revisiting Numerical Pattern Mining with Formal Concept Analysis. In Toby Walsh, editor, IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011, pages 1342–1347, Barcelona, Spain, 2011. IJCAI/AAAI.
- 11 Mehdi Kaytoue, Sergei O. Kuznetsov, Amedeo Napoli, and Sébastien Duplessis. Mining Gene Expression Data with Pattern Structures in Formal Concept Analysis. Information Science, 181(10):1989–2001, 2011.
- 12 Mohamed Rouane-Hacene, Marianne Huchard, Amedeo Napoli, and Petko Valtchev. Relational Concept Analysis: Mining Concept Lattices From Multi-Relational Data. Annals of Mathematics and Artificial Intelligence, 67(1):81–108, 2013.

## 3.20 Mathematical Methods for Analyzing Genomic Data

James B. Nation (University of Hawaii at Manoa, US)

License 
 © Creative Commons BY 3.0 Unported license
 © James B. Nation

 Joint work of Nation, James B.; Adarichev, Vyacheslav; Adaricheva, Kira; Okimoto, Gordon; Wenska, Tom; Zeinalzadeh, Ashkan

In this talk we show that reductions based on the singular value decomposition can be combined with order-theoretic methods to produce a useful analysis of biological data for cancer patients. An iterated rank-one decomposition of matrices for gene expression, microRNA expression, and methylation loci identifies a genetic signature for a particular type of cancer. The signature can be correlated with clinical data, with the goal of guiding detection, classification and treatment of cancers.

This method is applied to find a signal based on the expression of five microRNA that is a predictor for survival (response to treatment) of patients with ovarian cancer. In a related study, an order-theoretic method is applied to immunological data from brain tumor patients to again find a predictor for response to treatment.

This is based on joint work with Vyacheslav Adarichev, Kira Adaricheva, Gordon Okimoto, Tom Wenska and Ashkan Zienalzadeh.

# 3.21 Enumeration of minimal dominating sets of a graph and related notions

Lhouari Nourine (University Blaise Pascal – Aubiere, FR)

License 

 © Creative Commons BY 3.0 Unported license
 © Lhouari Nourine

 Joint work of Kante, Mamadou Moustapha; Limouzy, Vincent; Mary, Arnaud; Nourine, Lhouari
 Main reference M. M. Kanté, V. Limouzy, A. Mary, L. Nourine, "On the Enumeration of Minimal Dominating Sets and Related Notions," . arXiv:1407.2053v1 [cs.DM], 2014.
 URL http://arxiv.org/abs/1407.2053v1

We consider the problem of enumerating (inclusion-wise) minimal dominating sets in graphs. The enumeration of minimal dominating sets of graphs is closely related to the well-known

## 22 14201 – Horn Formulas, Directed Hypergraphs, Lattices and Closure Systems

transversal hypergraph problem in hypergraphs, which consists in enumerating the set of minimal transversals (or hitting sets) of a hypergraph. It is well known that the set of minimal dominating sets of a graph is in bijection with the set of minimal transversals of its *closed neighbourhood hypergraph*. We show that the problem of enumerating minimal transversals of a hypergraph can be polynomially reduced to the enumeration of minimal dominating sets of co-bipartite graph, and therefore they are equivalent.

We also formulate the enumeration of minimal dominating sets of a graph as the enumeration of i-keys of some implicational system, where a set is a i-key if its i-closure is the set of vertices of the graph. Whenever the implicational system is direct then i-keys correspond to classical notion of keys.

#### References

- 1 Thomas Eiter and Georg Gottlob. Identifying the minimal transversals of a hypergraph and related problems. SIAM J. Comput., 24(6):1278–1304, 1995.
- 2 Mamadou Moustapha Kanté, Vincent Limouzy, Arnaud Mary, and Lhouari Nourine. Enumeration of minimal dominating sets and variants. In: Olaf Owe, Martin Steffen, and Jan Arne Telle, editors, *FCT*, volume 6914 of *Lecture Notes in Computer Science*, pages 298–309. Springer, 2011.
- 3 Mamadou Moustapha Kanté, Vincent Limouzy, Arnaud Mary, and Lhouari Nourine. On the neighbourhood helly of some graph classes and applications to the enumeration of minimal dominating sets. ISAAC, pages 289–298, 2012.
- 4 Mamadou Moustapha Kanté, Vincent Limouzy, Arnaud Mary, Lhouari Nourine, and Takeaki Uno. On the Enumeration and Counting of Minimal Dominating sets in Interval and Permutation Graphs. ISAAC, pages 339–349, 2013.

## 3.22 Parameterized Ceteris Paribus Preferences over Atomic Conjunctions

Sergei Obiedkov (NRU Higher School of Economics – Moscow, RU)

License O Creative Commons BY 3.0 Unported license

© Sergei Obiedkov

Main reference S. Obiedkov, "Modeling ceteris paribus preferences in formal concept analysis," in Proc. of the 11th Int'l Conf. on Formal Concept Analysis (ICFCA'13), LNCS, Vol. 7880, pp. 188–202, Springer, 2013. URL http://dx.doi.org/10.1007/978-3-642-38317-5\_12

We consider a propositional language for describing parameterised ceteris paribus preferences [4] over atomic conjunctions. Such preferences are only required to hold when the alternatives being compared agree on a specified subset of propositional variables. We propose an approach based on formal concept analysis [1] to learning preferences from data by showing that ceteris paribus preferences valid in a particular model correspond to implications of a special formal context derived from this model. Regarding the expressivity of the language in question, we show that a parameterised preference statement is equivalent to a conjunction of an exponential number of classical, non-parameterized, ceteris paribus statements [3]. Next, we present an inference system for parameterized statements and prove that the problem of checking the semantic consequence relation for such statements is coNP-complete. Finally, we adapt a polynomial-time algorithm for abduction using Horn clauses represented by their characteristic models [2] to the problem of determining preferences over new alternatives from preferences over given alternatives (with ceteris paribus preferences as the underlying model).

#### References

- 1 Bernhard Ganter and Rudolf Wille. Formal Concept Analysis: Mathematical Foundations. Springer, Berlin/Heidelberg, 1999
- 2 H. Kautz, M. Kearns, B. Selman. Horn Approximations of Empirical Data. Artificial Intelligence, 74(1):129–145, 1995.
- 3 Sergei Obiedkov. Modeling ceteris paribus preferences in formal concept analysis. In: Peggy Cellier, Felix Distel, and Bernhard Ganter (eds.) Formal Concept Analysis, volume 7880 of Lecture Notes in Computer Science, pages 188–202. Springer Berlin Heidelberg, 2013
- 4 Johan van Benthem, Patrick Girard, and Olivier Roy. Everything else being equal: A modal logic for ceteris paribus preferences. J. Philosophical Logic, 38(1):83–125, 2009

## 3.23 RQL: a Query Language for Implications

Jean-Marc Petit (INSA - Lyon, FR)

in Databases," in Proc. of the Languages for Data Mining and Machine Learning Workshop (LML'13), pp. 1–16, 2013. URL http://liris.cnrs.fr/publis/?id=6311

RQL is a concrete SQL-like language derived from a well-founded logical query language, called SafeRL, allowing to express a wide variety of rules to be satisfied against the data. RQL extends and generalizes functional dependencies in databases to new and unexpected implications easily expressed with a SQL-like language. In this setting, every rule mining problem turns out to be seen as a query processing problem. We provide a query rewriting technique and a constructive proof of the main query equivalence theorem, leading to an efficient query processing technique. From a practical point of view, we show how a tight integration of RQL can be performed on top of any DataBase Management Systems. The approach has been implemented and experimented on different datasets (http://rql.insa-lyon.fr).

#### References

 B. Chardin, E. Coquery, B. Gouriou, M. Pailloux, and J.-M. Petit. Query Rewriting for Rule Mining in Databases. In: Languages for Data Mining and Machine Learning (LML) Workshop@ECML/PKDD 2013, B. Crémilleux, L. De Raedt, P. Frasconi, and T. Guns, Eds., Sep. 2013, pp. 1–16.

## 3.24 On the Succinctness of Closure Operator Representations

Sebastian Rudolph (TU Dresden, DE)

License 
Creative Commons BY 3.0 Unported license
Creative Commons BY 3.0 Unported license
Creative Common Section Representations," in Proc. of the 12th Int'l
Main reference
Creative Common Common Analysis (ICECA214), LNCS, Vol. 8478, no. 15, 26, Sectioner 2014

Conf. on Formal Concept Analysis (ICFCA'14), LNCS, Vol. 8478, pp. 15–36, Springer, 2014. URL http://dx.doi.org/10.1007/978-3-319-07248-7\_2

It is widely known that closure operators on finite sets can be represented by sets of implications (also known as inclusion dependencies) as well as by formal contexts. In this paper, we consider these two representation types, as well as generalizations of them:

extended implications and context families. We discuss the mutual succinctness of these four representations and the tractability of certain operations used to compare and modify closure operators.

### 3.25 The hydra number of a graph

Despina Stasi (Illinois Inst. of Technology, US)

License 

 © Creative Commons BY 3.0 Unported license
 © Despina Stasi

 Joint work of Sloan, Robert H.; Stasi, Despina; Turán, György
 Main reference R. H. Sloan, D. Stasi, G. Turán, "Hydras: Directed hypergraphs and horn formulas," in Proc. of the 38th Int'l Workshop on Graph-Theoretic Concepts in Computer Science (WG'12), LNCS, Vol. 7551, pp. 237–248, Springer, 2012.
 URL http://dx.doi.org/10.1007/978-3-642-34611-8\_25

We consider a graph parameter, the hydra number, arising from a restricted version of Horn minimization. The hydra number of a graph G=(V,E) is the minimum number of hyperarcs of the form (u,v)->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) is an edge 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. We also give a lower bound for trees which is sharp for spider trees.

# 3.26 The Complexity of Robust Satisfiability of the Constraint Satisfaction Problem (Survey)

Suguru Tamaki (Kyoto University, JP)

License ☺ Creative Commons BY 3.0 Unported license © Suguru Tamaki

In the Constraint Satisfaction Problem (CSP), an instance consists of a set of variables and a set of relations, and the objective is to find an assignment of values to the variables that satisfies all the constraints. In Max-CSP, the objective is to find an assignment that satisfies as many constraints as possible. Robust satisfiability of CSP is the notion between (standard) CSP and Max-CSP.

A CSP instance is  $(1 - \epsilon)$ -satisfiable if some assignment satisfies at least a  $(1 - \epsilon)$ -fraction of constraints. For a class C of instances of CSP, an approximation algorithm is robust if there exists  $f : [0, 1] \rightarrow [0, \infty)$ , where  $\lim_{\epsilon \to 0} f(\epsilon) = 0$ , such that for every  $\epsilon \ge 0$  and every  $(1 - \epsilon)$ -satisfiable instance of C, the algorithm outputs an assignment that satisfies at least a  $(1 - f(\epsilon))$ -fraction of constraints.

Since robust satisfiability of CSP is NP-hard in general, we are naturally interested in tractable subclasses of CSP. A constraint language  $\Gamma$  is a set of relations over some domain. Then a subclass of CSP, denoted by  $\text{CSP}(\Gamma)$ , is defined as a set of CSP instances with relations from  $\Gamma$ .

#### Kira Adaricheva, Giuseppe F. Italiano, Hans Kleine Büning, and György Turán

The complexity classification problem of robust satisfiability of  $\text{CSP}(\Gamma)$  for each  $\Gamma$  was initiated by Zwick, and people have continued the study to reach the ultimate goal, i.e., the complete classification of constraint languages with respect to polynomial time robust approximability. Finally, Barto and Kozik complete the classification for CSP with finite domains.

In this talk, I will give an overviw on the complexity of robust satisfiability of CSP. This talk is in part based on joint work with Gábor Kun, Ryan O'Donnell, Yuichi Yoshida, and Yuan Zhou.

## 3.27 Some combinatorial problems for directed hypergraphs

György Turán (University of Illinois – Chicago, US)

License ☺ Creative Commons BY 3.0 Unported license ☺ György Turán

We consider directed hypergraphs with size-3 edges of the form  $a, b \rightarrow c$ , and discuss some extremal and probabilistic problems corresponding to analogous problems for graphs.

How many edges can a directed hypergraph on n vertices have if it doesn't contain two edges of the form  $a, b \to c$  and  $c, d \to e$ ? It is shown that the maximum is given by the following directed hypergraph: divide the vertices into two parts A and B, and include all triples  $a, b \to c$  with  $a, b \in A$  and  $c \in B$ , choosing sizes to maximize the number of edges. The similar problem for triples  $a, b \to c$  and  $b, c \to d$  is solved only asymptotically.

We also discuss a question about random directed hypergraphs. where every edge is included with probability p. Given a directed hypergraph, a pair (u, v) is good, if forward chaining started from u and v marks every vertex. The probability of having at least one good pair changes from almost 0 to almost 1 over the interval between  $1/n \ln n$  and  $\ln \ln n/n \ln n$ . The probability of every pair being good changes from almost 0 to almost 1 around  $\ln n/n$ . The question of how the number of good pairs grows in between is open.

We also mention an  $n/\ln n$  approximation algorithm for the minimization of Horn formulas in the Steiner version, where it is allowed to introduce new variables in a restricted manner. The algorithm is based on covering bipartite graphs with bicliques.

Joint works with Amitava Bhattacharya, Bhaskar DasGupta, Marina Langlois, Dhruv Mubayi, Despina Stasi and Robert Sloan.

## Participants

Kira V. Adaricheva Yeshiva Univ. – New York, US Marta Arias UPC – Barcelona, ES Giorgio Ausiello University of Rome "La Sapienza", IT Jose Luis Balcazar UPC – Barcelona, ES Laurent Beaudou University Blaise Pascal – Aubiere, FR Kristof Berczi Eötvös Lorand University -Budapest, HU Karell Bertet University of La Rochelle, FR Endre Boros Rutgers Univ. - Piscataway, US Ondřej Čepek Charles University - Prague, CZ Vincent Duquenne UPMC - Paris, FR Thomas Eiter TU Wien, AT Donatella Firmani University of Rome "Tor Vergata", IT John Franco University of Cincinnati, US

Loukas Georgiadis University of Ioannina, GR Amélie Gheerbrant University Paris-Diderot, FR Gabriel Istrate West Univ. of Timisoara, RO Giuseppe F. Italiano University of Rome "Tor Vergata", IT Roni Khardon Tufts University, US Hans Kleine Büning Universität Paderborn, DE Petr Kučera Charles University - Prague, CZ Oliver Kullmann Swansea University, GB Sergei O. Kuznetsov NRU Higher School of Economics - Moscow, RU Luigi Laura University of Rome "La Sapienza", IT Leonid Libkin University of Edinburgh, GB Kazuhisa Makino University of Tokyo, JP Pierre Marquis Artois University - Lens, FR Umberto Nanni University of Rome "La Sapienza", IT

Amedeo Napoli LORIA & INRIA Nancy, FR

■ James B. Nation Univ. of Hawaii at Manoa, US

Lhouari Nourine
 University Blaise Pascal –
 Aubiere, FR

Sergei Obiedkov
 NRU Higher School of
 Economics – Moscow, RU

Jean-Marc Petit INSA – Lyon, FR

 Sebastian Rudolph TU Dresden, DE

Petr Savicky Academy of Science – Prague, CZ

■ Robert H. Sloan Univ. of Illinois – Chicago, US

Ewald Speckenmeyer Universität Köln, DE

Despina Stasi
 Illinois Inst. of Technology, US

Suguru Tamaki
 Kyoto University, JP

György Turán Univ. of Illinois – Chicago, US

Marcel Wild Stellenbosch University – Matieland, ZA

