



IN PARTNERSHIP WITH:  
**CNRS**

**Université de Lorraine**

Activity Report 2017

## **Project-Team ORPAILLEUR**

Knowledge discovery, knowledge engineering

IN COLLABORATION WITH: Laboratoire lorrain de recherche en informatique et ses applications (LORIA)

RESEARCH CENTER  
**Nancy - Grand Est**

THEME  
**Data and Knowledge Representation  
and Processing**



## Table of contents

<b>1. Personnel</b>	<b>2</b>
<b>2. Overall Objectives</b>	<b>3</b>
<b>3. Research Program</b>	<b>3</b>
3.1. Knowledge Discovery guided by Domain Knowledge	3
3.2. Text Mining	4
3.3. Knowledge Systems and Web of Data	4
<b>4. Application Domains</b>	<b>5</b>
4.1. Life Sciences: Biology, Chemistry and Medicine	5
4.2. Other Application Domains	6
4.2.1. Cooking	6
4.2.2. Agronomy	6
4.2.3. Digital Humanities	6
<b>5. Highlights of the Year</b>	<b>6</b>
<b>6. New Software and Platforms</b>	<b>7</b>
6.1. ARPEntAge	7
6.2. CarottAge	7
6.3. CORON	7
6.4. Tuurbine	8
6.5. LatViz: Visualization of Concept Lattices	8
6.6. OrphaMine: Data Mining Platform for Orphan Diseases	8
6.7. Siren: Interactive and Visual Redescription Mining	9
<b>7. New Results</b>	<b>9</b>
7.1. Mining of Complex Data	9
7.1.1. FCA and Variations: RCA, Pattern Structures and Biclustering	9
7.1.2. Text Mining	10
7.1.3. Mining Sequences and Trajectories	10
7.1.4. Redescription Mining	10
7.1.5. Mining subgroups as a single-player game	11
7.1.6. Data Privacy: Online link disclosure strategies for social networks	11
7.1.7. Aggregation	11
7.2. Knowledge Discovery in Healthcare and Life Sciences	12
7.2.1. Ontology-based Clustering of Biological Linked Open Data	12
7.2.2. Biological Data Aggregation for Knowledge Discovery	12
7.2.3. Validation of Pharmacogenomics Knowledge	13
7.2.4. Analysis of biomedical data annotated with ontologies	13
7.3. Knowledge Engineering and Web of Data	13
7.3.1. Current Trends in Case-Based Reasoning	14
7.3.2. Exploring and Classifying the Web of Data	14
7.4. Advances in Graph Theory, Clone Theory and Multiple-Valued Logic	15
7.4.1. Graph Theory	15
7.4.2. Multiple-Valued Logic and (Partial) Clone Theory	15
<b>8. Partnerships and Cooperations</b>	<b>16</b>
8.1. Regional Initiatives	16
8.1.1. Hydreos	16
8.1.2. The Smart Knowledge Discovery Project	16
8.2. National Initiatives	16
8.2.1. ANR	16
8.2.1.1. Elker (2017–2020)	16
8.2.1.2. ISTEEX (2014–2017)	17

8.2.1.3. PractiKPharma (2016–2020)	17
8.2.2. CNRS PEPS and Mastodons projects	17
8.2.2.1. Mastodons Projects: from HyQual to HyQualiBio (2016–2018)	17
8.2.2.2. PEPS Decade	17
8.3. European Initiatives	18
8.4. International Initiatives	18
8.4.1. Inria Associate Teams Not Involved in an Inria International Lab	18
8.4.2. Participation in International Programs	19
8.4.2.1. A stay at NASA Frontier Development Lab	19
8.4.2.2. LEA STRUCO	19
8.4.2.3. Research Collaboration with HSE Moscow	19
<b>9. Dissemination</b> .....	<b>20</b>
9.1. Promoting Scientific Activities	20
9.2. Teaching - Supervision - Juries	20
<b>10. Bibliography</b> .....	<b>21</b>

# Project-Team ORPAILLEUR

*Creation of the Project-Team: 2008 January 01*

## **Keywords:**

### **Computer Science and Digital Science:**

- A3. - Data and knowledge
- A3.1.1. - Modeling, representation
- A3.1.7. - Open data
- A3.2. - Knowledge
- A3.2.1. - Knowledge bases
- A3.2.2. - Knowledge extraction, cleaning
- A3.2.3. - Inference
- A3.2.4. - Semantic Web
- A3.2.5. - Ontologies
- A3.3.2. - Data mining
- A3.3.3. - Big data analysis
- A3.4.1. - Supervised learning
- A3.4.2. - Unsupervised learning
- A3.4.5. - Bayesian methods
- A3.4.8. - Deep learning
- A3.5.2. - Recommendation systems
- A4. - Security and privacy
- A4.1. - Threat analysis
- A8.1. - Discrete mathematics, combinatorics
- A8.7. - Graph theory
- A9. - Artificial intelligence
- A9.1. - Knowledge
- A9.2. - Machine learning
- A9.6. - Decision support

### **Other Research Topics and Application Domains:**

- B1.1.2. - Molecular biology
- B2. - Health
- B2.3. - Epidemiology
- B2.4.1. - Pharmacokinetics and dynamics
- B2.4.2. - Drug resistance
- B3.1. - Sustainable development
- B3.5. - Agronomy
- B3.6. - Ecology
- B3.6.1. - Biodiversity
- B6.3.4. - Social Networks
- B6.4. - Internet of things
- B8.5.2. - Crowd sourcing

B9. - Society and Knowledge

B9.4.5. - Data science

## 1. Personnel

### Research Scientists

Amedeo Napoli [Team leader, CNRS, Researcher, HDR]  
Esther Catherine Galbrun [Inria, Researcher, until November 2017]  
Chedy Raïssi [Inria, Researcher]  
Jean-Sébastien Sereni [CNRS, Researcher, until August 2017]

### Faculty Members

Miguel Couceiro [Univ. de Lorraine, Professor]  
Adrien Coulet [Univ. de Lorraine, Associate Professor]  
Nicolas Jay [Univ. de Lorraine, Professor, HDR]  
Jean Lieber [Univ. de Lorraine, Associate Professor, HDR]  
Florence Le Ber [ENGEES Strasbourg, Professor (Team Associate), HDR]  
Jean-François Mari [Univ. de Lorraine, Professor, HDR]  
Emmanuel Nauer [Univ. de Lorraine, Associate Professor]  
Sébastien Da Silva [Univ. de Lorraine, Associate Professor]  
Malika Smaïl-Tabbone [Univ. de Lorraine, Associate Professor, HDR]  
Yannick Toussaint [Univ. de Lorraine, Professor, HDR]

### Post-Doctoral Fellows

Emmanuelle Gaillard [Univ. de Lorraine, ATER, until August 2017]  
Joël Legrand [Univ. de Lorraine]

### PhD Students

Quentin Brabant [Univ. de Lorraine]  
Kévin Dalleau [CNRS]  
Nyoman Juniarta [CNRS]  
Tatiana Makhalova [Inria, from November 2017]  
Pierre Monnin [Univ. de Lorraine]  
François Pirot [Univ. de Lorraine, until August 2017]  
Justine Reynaud [Inria]  
Mohsen Sayed [Univ. de Lorraine, until July 2017]

### Technical staff

Thi Nhu Nguyen Le [Inria, until February 2017]  
Jérémie Nevin [Inria, from July 2017]

### Interns

Ilias Benjelloun [CNRS, from March 2017 until July 2017]  
Colin Flory [Univ. de Lorraine, from April 2017 until June 2017]  
Cyrille Formont [CESI, from September 2017 until October 2017]  
Tristan Gillard [Univ. de Lorraine, from March 2017 until July 2017]  
Romain Gogdemir [Univ. de Lorraine, from June 2017 until August 2017]  
Graziella Husson [Univ. de Lorraine, from June 2017 until August 2017]  
Doryan Kaced [Inria, from Mar 2017 until Jul 2017]  
Laëtitia Lemièrre [Inria, from April 2017 until June 2017]  
Alexandre Mayeux [CESI, from September 2017 until October 2017]  
Nicolas Warnet [Univ. de Lorraine, from April 2017 until October 2017]  
Yelen Per [Univ. de Lorraine, Student (long stay)]

### Administrative Assistants

Emmanuelle Deschamps [Inria]  
Delphine Hubert [Univ. de Lorraine]  
Martine Kuhlmann [CNRS]

#### Visiting Scientist

Maha Thabet [Université de Sousse, Tunisie, from September 2017]

## 2. Overall Objectives

### 2.1. Introduction

Knowledge discovery in databases (KDD) consists in processing large volumes of data in order to discover knowledge units that are significant and reusable. Assimilating knowledge units to gold nuggets, and databases to lands or rivers to be explored, the KDD process can be likened to the process of searching for gold. This explains the name of the research team: in French “orpailleur” denotes a person who is searching for gold in rivers or mountains. The KDD process is based on three main operations: data preparation, data mining and interpretation of the extracted units as knowledge units. Moreover, the KDD process is iterative, interactive, and generally controlled by an expert of the data domain, called the analyst. The analyst selects and interprets a subset of the extracted units for obtaining knowledge units having a certain plausibility. In this view, KDD is an exploratory process similar to “exploratory data analysis”.

As a person searching for gold may have a certain experience about the task and the location, the analyst may use general and domain knowledge for improving the whole KDD process. Accordingly, the KDD process may be related to knowledge bases –or domain ontologies– related to the domain of data for implementing *knowledge discovery guided by domain knowledge* (KDDK). In KDDK, extracted units have “a life” after the interpretation step: they are represented as knowledge units using a knowledge representation formalism and integrated within an ontology to be reused for problem-solving needs. In this way, knowledge discovery extends and updates existing knowledge bases, reifying the complementarity of knowledge discovery and knowledge representation.

## 3. Research Program

### 3.1. Knowledge Discovery guided by Domain Knowledge

**Keywords:** knowledge discovery in databases, knowledge discovery in databases guided by domain knowledge, data mining, data exploration, formal concept analysis, classification, pattern mining, numerical methods in data mining.

Knowledge discovery in databases (KDD) aims at discovering patterns in large databases. These patterns can then be interpreted as knowledge units to be reused in knowledge systems. From an operational point of view, the KDD process is based on three main steps: (i) selection and preparation of the data, (ii) data mining, (iii) interpretation of the discovered patterns. The KDD process –as implemented in the Orpailleur team– is based on data mining methods which are either symbolic or numerical. Symbolic methods are based on pattern mining (e.g. mining frequent itemsets, association rules, sequences...), Formal Concept Analysis (FCA [78]) and extensions of FCA such as Pattern Structures [84] and Relational Concept Analysis (RCA [90]). Numerical methods are based on probabilistic approaches such as second-order Hidden Markov Models (HMM [85]), which are well adapted to the mining of temporal and spatial data. Other numerical methods in data mining which are also of interest for the team are Random Forests, SVM, and neural networks.

Domain knowledge, when available, can improve and guide the KDD process, materializing the idea of *Knowledge Discovery guided by Domain Knowledge* or KDDK. In KDDK, domain knowledge plays a role at each step of KDD: the discovered patterns can be interpreted as knowledge units and reused for problem-solving activities in knowledge systems, implementing the exploratory process “mining, interpreting (modeling), representing, and reasoning”. In this way, knowledge discovery appears as a core task in knowledge engineering, with an impact in various semantic activities, e.g. information retrieval, recommendation and ontology engineering. Usual application domains include agronomy, astronomy, biology, chemistry, and medicine.

One main operation in the research work of Orpailleur on KDDK is *classification*, which is a polymorphic process involved in modeling, mining, representing, and reasoning tasks. Classification problems can be formalized by means of a class of objects (or individuals), a class of attributes (or properties), and a binary correspondence between the two classes, indicating for each individual-property pair whether the property applies to the individual or not. The properties may be features that are present or absent, or the values of a property that have been transformed into binary variables. Formal Concept Analysis (FCA) relies on the analysis of such binary tables and may be considered as a symbolic data mining technique to be used for extracting a set of formal concepts then organized within a concept lattice [78] (concept lattices are also known as “Galois lattices” [71]).

In parallel, the search for frequent itemsets and the extraction of association rules are well-known symbolic data mining methods, related to FCA (actually searching for frequent itemsets can be understood as traversing a concept lattice). Both processes usually produce a large number of items and rules, leading to the associated problems of “mining the sets of extracted items and rules”. Some subsets of itemsets, e.g. frequent closed itemsets (FCIs), allow finding interesting subsets of association rules, e.g. informative association rules. This explains why several algorithms are needed for mining data depending on specific applications [91].

## 3.2. Text Mining

**Keywords:** text mining, knowledge discovery from texts, text classification, annotation, ontology engineering from texts.

The objective of a text mining process is to extract useful knowledge units from large collections of texts [67]. The text mining process shows specific characteristics due to the fact that texts are complex objects written in natural language. The information in a text is expressed in an informal way, following linguistic rules, making text mining a difficult task. A text mining process has to take into account –as much as possible– paraphrases, ambiguities, specialized vocabulary and terminology. This is why the preparation of texts for text mining is usually dependent on linguistic resources and methods.

From a KDDK perspective, text mining aims at extracting “interesting units” (nouns and relations) from texts with the help of domain knowledge encoded within a knowledge base. The process is roughly similar for text annotation. Text mining is especially useful in the context of semantic web for ontology engineering. In the Orpailleur team, we work on the mining of real-world texts in application domains such as biology and medicine, using mainly symbolic data mining methods, and especially Formal Concept Analysis. Accordingly, the text mining process may be involved in a loop used to enrich and to extend linguistic resources. In turn, linguistic and ontological resources can be exploited to guide a “knowledge-based text mining process”.

## 3.3. Knowledge Systems and Web of Data

**Keywords:** knowledge engineering, web of data, semantic web, ontology, description logics, classification-based reasoning, case-based reasoning, information retrieval.

The web of data constitutes a good platform for experimenting ideas on knowledge engineering and knowledge discovery. Following the principles of semantic web, a software agent may be able to read, understand, and manipulate information on the web, if and only if the knowledge necessary for achieving those tasks is available: this is why knowledge bases (domain ontologies) are of main importance. OWL is the knowledge representation language used to design ontologies and knowledge bases, which is based on description logics (DLs [68]). In OWL, knowledge units are represented by classes (DL concepts) having properties (DL roles)



and instances. Concepts can be organized within a partial order based on a subsumption relation, and the inference services are based on satisfiability, classification-based reasoning and case-based reasoning (CBR).

Actually, there are many interconnections between concept lattices in FCA and ontologies, e.g. the partial order underlying an ontology can be supported by a concept lattice. Moreover, a pair of implications within a concept lattice can be adapted for designing concept definitions in ontologies. Accordingly, we are interested here in two main challenges: how the web of data, as a set of potential knowledge sources (e.g. DBpedia, Wikipedia, Yago, Freebase) can be mined for helping the design of definitions and knowledge bases and how knowledge discovery techniques can be applied for providing a better usage of the web of data (e.g. LOD classification).

Accordingly, a part of the research work in Knowledge Engineering is oriented towards knowledge discovery in the web of data, as, with the increased interest in machine processable data, more and more data is now published in RDF (Resource Description Framework) format. Particularly, we are interested in the completeness of the data and their potential to provide concept definitions in terms of necessary and sufficient conditions [69]. We have proposed a novel technique based on FCA which allows data exploration as well as the discovery of definition (bidirectional implication rules).

## 4. Application Domains

### 4.1. Life Sciences: Biology, Chemistry and Medicine

**Participants:** Adrien Coulet, Kévin Dalleau, Esther Catherine Galbrun, Nicolas Jay, Joël Legrand, Jean Lieber, Pierre Monnin, Amedeo Napoli, Chedy Raïssi, Mohsen Sayed, Malika Smail-Tabbone, Yannick Toussaint.

**Keywords:** knowledge discovery in life sciences, biology, chemistry, medicine, pharmacogenomics and precision medicine.

One major application domain which is currently investigated by the Orpailleur team is related to life sciences, with particular emphasis on biology, medicine, and chemistry. The understanding of biological systems provides complex problems for computer scientists, and the developed solutions bring new research ideas or possibilities for biologists and for computer scientists as well. Indeed, the interactions between researchers in biology and researchers in computer science improve not only knowledge about systems in biology, chemistry, and medicine, but knowledge about computer science as well.

Knowledge discovery is gaining more and more interest and importance in life sciences for mining either homogeneous databases such as protein sequences and structures, or heterogeneous databases for discovering interactions between genes and environment, or between genetic and phenotypic data, especially for public health and precision medicine (pharmacogenomics). Pharmacogenomics is one main challenge for the Orpailleur team as it considers a large panel of complex data ranging from biological to medical data, and various kinds of encoded domain knowledge ranging from texts to formal ontologies.

On the same line as biological data, chemical data are presenting important challenges w.r.t. knowledge discovery, for example for mining collections of molecular structures and collections of chemical reactions in organic chemistry. The mining of such collections is an important task for various reasons among which the challenge of graph mining and the industrial needs (especially in drug design, pharmacology and toxicology). Molecules and chemical reactions are complex data that can be modeled as labeled graphs. Graph mining methods may play an important role in this framework and Formal Concept Analysis can also be used in an efficient and well-founded way [86]. Graph mining as considered in the framework of FCA is an important task on which we are working, whose results can be transferred to text mining as well.

We are working on knowledge management in medicine and analysis of patient trajectories. The Kasimir research project is about decision support and knowledge management for the treatment of cancer. This is a multidisciplinary research project in which researchers in computer science (Orpailleur) and experts in oncology are participating. For a given cancer localization, a treatment is based on a protocol, which is applied in 70% of the cases and provides a treatment. The 30% remaining cases are “out of the protocol”, e.g. contraindication, treatment impossibility, etc. and the protocol have to be adapted, based on discussions among specialists. This adaptation process is modeled in Kasimir thanks to CBR, where semantic web technologies have been used.

The analysis of patient trajectories, i.e. the “path” of a patient during illness (chronic illnesses and cancer), can be considered as an analysis of sequences. It is important to understand such sequential data and sequence mining methods should be adapted for addressing the complex nature of medical events. We are interested in the analysis of trajectories at different levels of granularity and w.r.t. external domain ontologies. In addition, it is also important to be able to compare and classify trajectories according to their content. Then we are also interested in the definition of similarity measures able to take into account the complex nature of trajectories and that can be efficiently implemented for allowing quick and reliable classifications.

## 4.2. Other Application Domains

**Participants:** Emmanuelle Gaillard, Florence Le Ber, Jean Lieber, Jean-François Mari, Amedeo Napoli, Emmanuel Nauer, Sébastien Da Silva.

### 4.2.1. Cooking

**Keywords:** cooking, knowledge engineering, case-based reasoning, semantic web

The origin of the Taaable project is the Computer Cooking Contest (CCC). A contestant to CCC is a system that answers queries about recipes, using a recipe base; if no recipe exactly matches the query, then the system adapts another recipe. Taaable is a case-based reasoning system based on knowledge representation, semantic web and knowledge discovery technologies. The system enables to validate scientific results and to study the complementarity of various research trends in an application domain which is simple to understand and which raises complex issues at the same time.

### 4.2.2. Agronomy

**Keywords:** simulation in agronomy, graph model in agronomy

Research in agronomy was conducted in cooperation between Inria and INRA, within the INRA research network PAYOTE about landscape modeling. The research work was related to the characterization and the simulation of hedgerow structures in agricultural landscapes, based on Hilbert-Peano curves and Markov models [73]. Moreover, an on-going research work about the representation of farmer experience is carried on within a collaboration with IRD in Madagascar [81]. Sketches drawn by farmers were transformed into graphs and compared thanks to Formal Concept Analysis.

### 4.2.3. Digital Humanities

**Keywords:** digital humanities, semantic web, SPARQL, approximate search, case-based reasoning

Members of the Orpailleur team are collaborating with a group of researchers working in history and philosophy of science and technologies (they are located in Brest, Montpellier and Nancy). The idea is to reuse semantic web technologies for better access and better representation of their text corpora.

## 5. Highlights of the Year

### 5.1. Highlights of the Year

Classical properties of functions such as associativity, although algebraically easy to read, are hard to meaningfully interpret. In [60] Miguel Couceiro and colleagues at the University of Luxembourg (Jean-Luc Marichal, Jimmy Devillet) showed that associative and quasi-trivial operations that are non-decreasing are characterized in terms of total and weak orderings through the so-called single-peakedness property introduced in social choice theory by Duncan Black. This enabled visual interpretations of the above mentioned algebraic properties, and the enumeration of such operations led to several, previously unknown, integer sequences in Sloane's On-Line Encyclopedia of Integer Sequences (<http://www.oeis.org>), e.g., A292932, A292933, and A292934.

## 6. New Software and Platforms

### 6.1. ARPEntAge

*Analyse de Régularités dans les Paysages : Environnement, Territoires, Agronomie*

KEYWORDS: Stochastic process - Hidden Markov Models

FUNCTIONAL DESCRIPTION: ARPEntAge is a software based on stochastic models (HMM2 and Markov Field) for analyzing spatio-temporal data-bases. ARPEntAge is built on top of the CarottAge system to fully take into account the spatial dimension of input sequences. It takes as input an array of discrete data in which the columns contain the annual land-uses and the rows are regularly spaced locations of the studied landscape. It performs a Time-Space clustering of a landscape based on its time dynamic Land Uses (LUS). Displaying tools and the generation of Time-dominant shape files have also been defined.

- Partner: INRA
- Contact: Jean-François Mari
- URL: [http://carottage.loria.fr/index\\_in\\_english.html](http://carottage.loria.fr/index_in_english.html)

### 6.2. CarottAge

KEYWORDS: Stochastic process - Hidden Markov Models

FUNCTIONAL DESCRIPTION: The system CarottAge is based on Hidden Markov Models of second order and provides a non supervised temporal clustering algorithm for data mining and a synthetic representation of temporal and spatial data. CarottAge is currently used by INRA researchers interested in mining the changes in territories related to the loss of biodiversity (projects ANR BiodivAgrim and ACI Ecoger) and/or water contamination. CarottAge is also used for mining hydromorphological data. Actually a comparison was performed with three other algorithms classically used for the delineation of river continuum and CarottAge proved to give very interesting results for that purpose.

- Participants: Florence Le Ber and Jean-François Mari
- Partner: INRA
- Contact: Jean-François Mari
- URL: [http://carottage.loria.fr/index\\_in\\_english.html](http://carottage.loria.fr/index_in_english.html)

### 6.3. CORON

KEYWORDS: Data mining - Closed itemset - Frequent itemset - Generator - Association rule - Rare itemset

FUNCTIONAL DESCRIPTION: The Coron platform is a KDD toolkit organized around three main components: (1) Coron-base, (2) AssRuleX, and (3) pre- and post-processing modules.

The Coron-base component includes a complete collection of data mining algorithms for extracting itemsets such as frequent itemsets, closed itemsets, generators and rare itemsets. In this collection we can find APriori, Close, Pascal, Eclat, Charm, and, as well, original algorithms such as ZART, Snow, Touch, and Talky-G. AssRuleX generates different sets of association rules (from itemsets), such as minimal non-redundant association rules, generic basis, and informative basis. In addition, the Coron system supports the whole life-cycle of a data mining task and proposes modules for cleaning the input dataset, and for reducing its size if necessary.

- Participants: Adrien Coulet, Aleksey Buzmakov, Amedeo Napoli, Florent Marcuola, Jérémie Bourseau, Laszlo Szathmary, Mehdi Kaytoue, Victor Codocedo and Yannick Toussaint
- Contact: Amedeo Napoli
- URL: <http://coron.loria.fr/site/index.php>

## 6.4. Tuurbine

KEYWORD: Semantic Web

FUNCTIONAL DESCRIPTION: Tuurbine: a Generic Ontology Guided Case-Based Inference Engine. The experience acquired since 5 years with the Taaable system conducted to the creation of a generic case-based reasoning system, whose reasoning procedure is based on a domain ontology. This new system, called Tuurbine, takes into account the retrieval step, the case base organization, and also an adaptation procedure which is not addressed by other generic case-based reasoning tools. Moreover, Tuurbine is built over semantic web standards that will ensure facilities for being plugged over data available on the web. The domain knowledge is represented in an RDF store, which can be interfaced with a semantic wiki, for collaborative edition and management of the knowledge involved in the reasoning system (cases, ontology, adaptation rules).

- Contact: Emmanuel Nauer
- URL: <http://tuurbine.loria.fr/>

## 6.5. LatViz: Visualization of Concept Lattices

- Contact: Amedeo Napoli
- URL: <http://latviz.loria.fr/>
- KEYWORDS: Formal Concept Analysis, Pattern Structures, Concept Lattice, Implications, Visualization

FUNCTIONAL DESCRIPTION.

LatViz is a tool allowing the construction, the display and the exploration of concept lattices. LatViz proposes some noticeable improvements over existing tools and introduces various functionalities focusing on interaction with experts, such as visualization of pattern structures for dealing with complex non-binary data, AOC-poset which is composed of the core elements of the lattice, concept annotations, filtering based on various criteria and a visualization of implications [70]. This way the user can effectively perform interactive exploratory knowledge discovery as often needed in knowledge engineering.

The Latviz platform can be associated with the Coron platform and extends its visualization capabilities (see <http://coron.loria.fr>). Recall that the Coron platform includes a complete collection of data mining algorithms for extracting itemsets and association rules.

## 6.6. OrphaMine: Data Mining Platform for Orphan Diseases

- Contact: Chedy Raïssi
- URL: <http://orphamine.inria.fr/>
- KEYWORDS: Bioinformatics, data mining, biology, health, data visualization, drug development.

FUNCTIONAL DESCRIPTION.

The OrphaMine platform enables visualization, data integration and in-depth analytics in the domain of “orphan diseases”, where data is extracted from the OrphaData ontology (<http://www.orpha.net/consor/cgi-bin/index.php>). At present, we aim at building a true collaborative portal that will serve different actors: (i) a general visualization of OrphaData data for physicians working, maintaining and developing this knowledge database about orphan diseases. (ii) the integration of analytics (data mining) algorithms developed by the different academic actors. (iii) the use of these algorithms to improve our general knowledge of rare diseases.

## 6.7. Siren: Interactive and Visual Redescription Mining

- Contact: Esther Catherine Galbrun
- URL: <http://siren.gforge.inria.fr/main/>
- KEYWORDS: Redescription mining, Interactivity, Visualization.

### FUNCTIONAL DESCRIPTION.

Siren is a tool for interactive mining and visualization of redescriptions. Redescription mining aims to find distinct common characterizations of the same objects and, vice versa, to identify sets of objects that admit multiple shared descriptions. The goal is to provide domain experts with a tool allowing them to tackle their research questions using redescription mining. Merely being able to find redescriptions is not enough. The expert must also be able to understand the redescriptions found, adjust them to better match his domain knowledge and test alternative hypotheses with them, for instance. Thus, Siren allows mining redescriptions in an anytime fashion through efficient, distributed mining, to examine the results in various linked visualizations, to interact with the results either directly or via the visualizations, and to guide the mining algorithm toward specific redescriptions.

New features, such as a visualization of the contribution of individual literals in the queries and the simplification of queries as a post-processing, have been added to the tool, during the internship of IUT student Laëtitia Lemière.

## 7. New Results

### 7.1. Mining of Complex Data

**Participants:** Quentin Brabant, Miguel Couceiro, Adrien Coulet, Esther Catherine Galbrun, Nyoman Juniarta, Florence Le Ber, Joël Legrand, Pierre Monnin, Tatiana Makhlova, Amedeo Napoli, Justine Reynaud, Chedy Raïssi, Mohsen Sayed, Yannick Toussaint.

**Keywords:** formal concept analysis, relational concept analysis, pattern structures, pattern mining, association rule, redescription mining, graph mining, sequence mining, biclustering, skyline, aggregation

#### 7.1.1. FCA and Variations: RCA, Pattern Structures and Biclustering

Advances in data and knowledge engineering have emphasized the needs for pattern mining tools working on complex data. In particular, FCA, which usually applies to binary data-tables, can be adapted to work on more complex data. In this way, we have contributed to two main extensions of FCA, namely Pattern Structures and Relational Concept Analysis. Pattern Structures (PS [77]) allow building a concept lattice from complex data, e.g. numbers, sequences, trees and graphs. Relational Concept Analysis (RCA) is able to analyze objects described both by binary and relational attributes [90] and can play an important role in text classification and text mining.

Many developments were carried out in pattern mining and FCA for improving data mining algorithms and their applicability, and for solving some specific problems such as information retrieval, discovery of functional dependencies and biclustering. We also worked on a generic framework based on FCA in which we can define the pattern mining process at a formal level [3]. We consider several types of patterns and we are making precise the mining of complex patterns represented as sequences, trees and graphs.

We also worked on a significant extension of previous work on the discovery of skyline patterns (or “skypatterns”) based on the theoretical relationships with condensed representations of patterns. We have shown how these relationships facilitate the computation of skypatterns. Thus we proposed a flexible and efficient approach to mine skypatterns using a dynamic constraint satisfaction problems (CSP) framework [30].

### 7.1.2. Text Mining

In the context of the PractikPharma ANR Project, we study cross-corpus training with Tree-LSTM for the extraction of biomedical relationships from texts, especially, how large annotated corpora developed for alternative tasks may improve the performance on biomedicine related tasks, for which few annotated resources are available [55]. We experiment two deep learning-based models to extract relationships from biomedical texts with high performance. The first one combines locally extracted features using a Convolutional Neural Network (CNN) model, while the second exploits the syntactic structure of sentences using a Recursive Neural Network (RNN) architecture. Our experiments show that the latter benefits from a cross-corpus learning strategy to improve the performance of relationship extraction tasks. Indeed our approach leads to state-of-the-art performances for four biomedical tasks for which few annotated resources are available (less than 400 manually annotated sentences). This may have a particular impact in specialized domains in which training resources are scarce, because they would benefit from the training data of other domains for which large annotated corpora do exist.

In the framework of the Hybride ANR project (terminated at the end of 2016), Mohsen Sayed Hassan proposed an original machine learning approach for identifying in texts about diseases phenotypes that are not yet represented within existing ontologies [9]. The result of the extraction is used to enrich existing ontologies of the considered domain. We studied three research directions: (1) extracting relationships from texts, i.e., extracting Disease-Phenotype (D-P) relationships, (2) identifying new complex entities standing as phenotypes of a rare disease, and (3) enriching an existing rare disease ontology on the basis of the relationships previously extracted.

A collection of abstracts of scientific articles is represented as a collection of dependency graphs used for discovering relevant pieces of biomedical knowledge. We focused on the completion of rare disease descriptions, by extracting Disease-Phenotype relationships. We developed an automatic approach named SPARE $\star$ , for extracting Disease-Phenotype relationships from PubMed abstracts, where phenotypes and rare diseases are previously annotated by a Named Entity Recognizer. SPARE $\star$  is the resulting hybrid approach that combines a graph-pattern based method, called SPARE, and a machine learning method (SVM). It benefits both from the good precision of SPARE and from the good recall of SVM. Finally, we applied pattern structures for classifying rare diseases and enriching an existing ontology about such diseases.

### 7.1.3. Mining Sequences and Trajectories

Nowadays datasets are available in very complex and heterogeneous ways. Mining of such data collections is essential to support many real-world applications ranging from healthcare to marketing. This year we finished a work on the analysis of “complex” sequential data and its usage in video games for the analysis of strategy “balance” in those games [14].

### 7.1.4. Redescription Mining

Among the mining methods developed in the team is redescription mining. Redescription mining aims to find distinct common characterizations of the same objects and, vice versa, to identify sets of objects that admit multiple shared descriptions [89]. It is motivated by the idea that in scientific investigations data oftentimes have different nature. For instance, they might originate from distinct sources or be cast over separate terminologies. In order to gain insight into the phenomenon of interest, a natural task is to identify the correspondences that exist between these different aspects.

A practical example in biology consists in finding geographical areas that admit two characterizations, one in terms of their climatic profile and one in terms of the occupying species. Discovering such redescrptions can contribute to better our understanding of the influence of climate over species distribution. Besides biology, applications of redescription mining can be envisaged in medicine or sociology, among other fields.

In a preceding work [83], we focused on the problem of pattern selection, developing a method for filtering a set of redescription to identify a non-redundant, interesting subset to present to the analyst. Also, we showcased the usability of redescription mining on an application in the political domain [76]. More specifically, we applied redescription mining to the exploratory analysis of the profiles and opinions of candidates to the parliamentary elections in Finland in 2011 and 2015.

We presented an introductory tutorial on redescription mining at SDM in April 2017 to help foster the research on these techniques and widen their use ([http://siren.mpi-inf.mpg.de/tutorial\\_sdm2017/main/](http://siren.mpi-inf.mpg.de/tutorial_sdm2017/main/)).

### ***7.1.5. Mining subgroups as a single-player game***

Discovering patterns that strongly distinguish one class label from another is a challenging data-mining task. The unsupervised discovery of such patterns would enable the construction of intelligible classifiers and to elicit interesting hypotheses from the data. Subgroup Discovery (SD) is one framework that formally defines this pattern mining task. However, SD still faces two major issues: (i) how to define appropriate quality measures to characterize the uniqueness of a pattern; (ii) how to select an accurate heuristic search technique when exhaustive enumeration of the pattern space is unfeasible. The first issue has been tackled by the Exceptional Model Mining (EMM) framework. This general framework aims to find patterns that cover tuples that locally induce a model that substantially differs from the model of the whole dataset. The second issue has been studied in SD and EMM mainly with the use of beam-search strategies and genetic algorithms for discovering a pattern set that is non-redundant, diverse and of high quality.

In [1], we argue that the greedy nature of most of these approaches produce pattern sets that lack of diversity. Consequently, we proposed to formally define pattern mining as a single-player game, as in a puzzle, and to solve it with a Monte Carlo Tree Search (MCTS), a technique mainly used for artificial intelligence and planning problems. The exploitation/exploration trade-off and the power of random search of MCTS lead to an any-time mining approach, in which a solution is always available, and which tends towards an exhaustive search if given enough time and memory. Given a reasonable time and memory budget, MCTS quickly drives the search towards a diverse pattern set of high quality. MCTS does not need any knowledge of the pattern quality measure, and we show to what extent it is agnostic to the pattern language.

### ***7.1.6. Data Privacy: Online link disclosure strategies for social networks***

Online social networks are transforming our culture and world. While online social networks have become an important channel for social interactions, they also raise ethical and privacy issues. A well known fact is that social networks leak information, that may be sensitive, about users. However, performing accurate real world online privacy attacks in a reasonable time frame remains a challenging task. We continued our work on this aspect and we address the problem of rapidly disclosing many friendship links using only legitimate queries (i.e., queries and tools provided by the targeted social network). The results of this joint work with the Pesto Inria Team are published in [31].

### ***7.1.7. Aggregation***

Aggregation and consensus theory study processes dealing with the problem of merging or fusing several objects, e.g., numerical or qualitative data, preferences or other relational structures, into a single or several objects of similar type and that best represents them in some way. Such processes are modeled by so-called aggregation or consensus functions [79], [82]. The need to aggregate objects in a meaningful way appeared naturally in classical topics such as mathematics, statistics, physics and computer science, but it became increasingly emergent in applied areas such as social and decision sciences, artificial intelligence and machine learning, biology and medicine.

We are working on a theoretical basis of a unified theory of consensus and to set up a general machinery for the choice and use of aggregation functions. This choice depends on properties specified by users or decision makers, the nature of the objects to aggregate as well as computational limitations due to prohibitive algorithmic complexity. This problem demands an exhaustive study of aggregation functions that requires an axiomatic treatment and classification of aggregation procedures as well as a deep understanding of their structural behavior. It also requires a representation formalism for knowledge, in our case decision rules and methods for discovering them. Typical approaches include rough-set and FCA approaches, that we aim to extend in order to increase expressivity, applicability and readability of results. Applications of these efforts already appeared and further are expected in the context of three multidisciplinary projects, namely the “Fight Heart Failure” (research project with the Faculty of Medicine in Nancy), the European H2020 “CrossCult” project, and the “ISIPA” (Interpolation, Sugeno Integral, Proportional Analogy) project.

In our recent work, we mainly focused on the utility-based preference model in which preferences are represented as an aggregation of preferences over different attributes, structured or not, both in the numerical and qualitative settings. In the latter case, the Sugeno integral is widely used in multiple criteria decision making and decision under uncertainty, for computing global evaluations of items based on local evaluations (utilities). The combination of a Sugeno integral with local utilities is called a Sugeno utility functional (SUF). A noteworthy property of SUFs is that they represent multi-threshold decision rules. However, not all sets of multi-threshold rules can be represented by a single SUF. We showed how to represent any set of multi-threshold rules as a combination of SUFs and studied their potential advantages as a compact representation of large sets of rules, as well as an intermediary step for extracting rules from empirical datasets [38], [59]. Problems related to feature selection and model elicitation were tackled in [15].

## 7.2. Knowledge Discovery in Healthcare and Life Sciences

**Participants:** Miguel Couceiro, Adrien Coulet, Kévin Dalleau, Nicolas Jay, Joël Legrand, Pierre Monnin, Amedeo Napoli, Chedy Raïssi, Mohsen Sayed, Malika Smaïl-Tabbone, Yannick Toussaint.

### 7.2.1. Ontology-based Clustering of Biological Linked Open Data

Increasing amounts of biomedical data provided as Linked Open Data (LOD) offer novel opportunities for knowledge discovery in bio-medicine. We proposed an approach for selecting, integrating, and mining LOD with the goal of discovering genes responsible for a disease [87]. We are currently working on the integration of LOD about known phenotypes and genes responsible for diseases along with relevant bio-ontologies. We are also defining a corpus-based semantic distance. One possible application of this work is to build and compare possible “diseaseomes”, i.e. global graphs representing all diseases connected according to their pairwise similarity values.

### 7.2.2. Biological Data Aggregation for Knowledge Discovery

This specific research takes place within two multi-disciplinary projects initiated in 2016, in collaboration with the Capsid Team, with a group of clinicians from the Regional University Hospital (CHU Nancy) and bio-statisticians from the Maths Lab (IECL). The first project is entitled ITM2P<sup>1</sup> and depends on the so-called CPER 2015–2020 framework. We are involved in the design of the SMEC platform as a support for “Simulation, Modeling and Knowledge Extraction from Bio-Medical Data”.

The second project is a RHU<sup>2</sup> project entitled *Fight Heart Failure* (FHF), where we are in charge of a workpackage about entitled “Network-based analysis and integration”. Accordingly, we are working on the definition of multidimensional similarity measure for comparing and clustering sets of patients. Each cluster should correspond to a bioprofile, i.e. a subgroup of patients sharing the same form of the disease and thus the same diagnosis and care strategy. The first results were presented at the “International Symposium on Aggregation and Structures (ISAS 2016)” [74] where we proposed an approach for complex graph aggregation resulting in a similarity graph between a subset of nodes. In a recent work we explored an alternative to define

<sup>1</sup>“Innovations Technologiques, Modélisation et Médecine Personnalisée”

<sup>2</sup>“Recherche Hospitalo-Universitaire”



and efficiently compute pairwise patient similarity thanks to “Unsupervised Extremely Randomized Trees” [62].

The next challenge is to build a prediction model for each bioprofile/subgroup, once validated by clinicians, to be integrated in a decision support system. Currently, we are investigating “Statistical Relational Learning” and analogy-based methods for achieving this goal.

### 7.2.3. Validation of Pharmacogenomics Knowledge

A standard task in pharmacogenomics research is identifying genes that may be involved in drug response variability. Those genes are called “pharmacogenes”. As genomic experiments in this domain tend to generate many false positives, computational approaches based on background knowledge may generate more valuable results. Until now, the latter have only used molecular network databases or biomedical literature. We developed a new method that takes advantage of various linked data sources to evaluate the validity of uncertain drug-gene relationships, i.e. pharmacogenes [5]. One advantage relies on the standard implementation of linked data that facilitates the joint use of various sources and makes easier to consider features of various origins. The second advantage is related to graph mining approaches that we are using, which consider linked data in their original form, i.e. as graphs. We selected, formatted, interconnected and published an initial set of linked data sources relevant to pharmacogenomics, named PGxLOD (for “PharmacoGenomic Linked Open Data”). We applied and compared distinct numerical classification methods on these data and identified candidate pharmacogenes.

This work is a first attempt for validating state-of-the-art knowledge in pharmacogenomics, which is one objective of the ANR project “PractiKPharma” initiated in 2016 (<http://praktikpharma.loria.fr/>). This year, we improved and enriched PGxLOD in various ways. Firstly, we wanted PGxLOD to be able to encompass pharmacogenomic knowledge of various origin, such as scientific literature, specialized databases, or Electronic Health Records (EHRs). To represent the fact that a given knowledge unit may have distinct provenances, we developed a simple ontology named PGxO (“Pharmacogenomic Ontology”) which relies on the Standard Ontology PROV-O to represent provenance. This makes possible to compare similar knowledge units that may have distinct origins [45].

### 7.2.4. Analysis of biomedical data annotated with ontologies

In the context of the Snowflake Inria Associate Team (at present Snowball), we developed an approach based on pattern structures to identify frequently associated ADRs (Adverse Drug Reactions) from patient data either in the form of EHR or ADR spontaneous reports. In this case, pattern structures provide an expressive representation of ADR, taking into account the multiplicity of drugs and phenotypes involved in such reactions. Additionally, pattern structures allow considering diverse biomedical ontologies used to represent or annotate patient data, enabling a “semantic” comparison of ADRs. Up to now, this is the first research work considering such representations to mine rules between frequently associated ADRs. We illustrated the generality of the approach on two patient datasets, each of them linked to distinct biomedical ontologies. The first dataset corresponds to anonymized EHRs, extracted from “STRIDE”, the EHR data warehouse of Stanford Hospital and Clinics. The second dataset is extracted from the U.S. FDA (for Food & Drug Administration) “Adverse Event Reporting System” (FAERS). Several significant association rules have been extracted and analyzed and may be used as a basis of a recommendation system [29].

## 7.3. Knowledge Engineering and Web of Data

**Participants:** Emmanuelle Gaillard, Nicolas Jay, Florence Le Ber, Jean Lieber, Amedeo Napoli, Emmanuel Nauer, Justine Reynaud, Yannick Toussaint.

**Keywords:** knowledge engineering, web of data, definition mining, classification-based reasoning, case-based reasoning, belief revision, semantic web

### 7.3.1. Current Trends in Case-Based Reasoning

The Taaable project was originally created as a challenger of the Computer Cooking Contest (ICCB Conference) [72]. Beyond its participation to the CCC challenges, the Taaable project aims at federating various research themes including case-based reasoning (CBR), knowledge discovery, knowledge engineering and belief change theory [6]. CBR performs adaptation of recipes w.r.t. user constraints. The reasoning process is based on a cooking domain ontology (especially hierarchies of classes) and adaptation rules. The knowledge base is encoded within a semantic wiki containing the recipes, the domain ontology and adaptation rules.

Adaptation rules have been used to manage ingredient adaptation with a restrictive set of available ingredients [43]. Three types of rule have been identified. The first type is about the substitution of ingredients belonging to a same category (e.g. dairy) by the sole available ingredient of this category (e.g. yogurt). The second type of rule is in concern with substitution, according to the role the ingredients play in the recipe, e.g. egg can be replaced by salmon in salad recipes because they are both playing the role of a protein. The last type of rules consists in removing ingredients of original recipes when they are not concerned by a rule of the first nor second type.

FCA allows the classification of objects according to the properties they share into a concept lattice. A lattice has been built from a large set of cocktail recipes according to the ingredients they use, producing a hierarchy of ingredient combinations. For example, when a cocktail recipe  $R$  has to be adapted, this lattice can be used to search the best ingredient combinations in the concepts that are the closest to the concept representing  $R$  [43].

Two main research works were carried out about the application of CBR in medicine. Imaging, in particular in nuclear medicine, is getting more and more complex over the years. Each year, new radiotracers and machines are developed and tested. Despite this rapid evolution, few studies address the issue of image interpretation and imaging report. In [35], we show how nuclear image interpretation is improved by Tetra, a new case-based decision support system.

Cancer registries are important tools in the fight against cancer. At the heart of these registries is the data collection and coding process. Ruled by complex international standards and numerous best practices, operators are easily overwhelmed. In [48], a system is presented to assist operators in the interpretation of best medical coding practices.

Finally, an approach to adaptation based on the principles of analogical transfer applied to the formalism RDFS has been developed. It is based on the problem-solution dependency represented as an RDFS graph: this dependency within the source case is modified so that it fits the context of the target problem [2]. This is implemented within the so-called SQTRL system (for “SPARQL Query Transformation Rule Language” <http://tuurbine.loria.fr/sqtrl/>) [2]. The development of SQTRL is based on a collaboration between Orpailleur team and the Archives Henri Poincaré (<http://poincare.univ-lorraine.fr/>).

### 7.3.2. Exploring and Classifying the Web of Data

A part of the research work in Knowledge Engineering is oriented towards knowledge discovery in the web of data, following the increase of data published in RDF (Resource Description Framework) format and the interest in machine processable data. The quick growth of Linked Open Data (LOD) has led to challenging aspects regarding quality assessment and data exploration of the RDF triples that shape the LOD cloud. In the team, we are particularly interested in the “completeness of the data” viewed as their potential to provide concept definitions in terms of necessary and sufficient conditions [69]. We have proposed a novel technique based on Formal Concept Analysis which classifies subsets of RDF data into a concept lattice [47]. This allows data exploration as well as the discovery of implication rules which are used to automatically detect “possible completions of RDF data” and to provide definitions. Moreover, this is a way of reconciling syntax and semantics in the LOD cloud. Experiments on the DBpedia knowledge base shows that this kind of approach is well-founded and effective.

In the same way, FCA can be used to improve ontologies associated with the Web of data. Accordingly, we proposed a method to build a concept lattice from linked data and compare the structure of this lattice with an

ontology used to type the considered data [46]. The result of this comparison shows which “new axioms” can be proposed to ontology developers for guiding their design work.

## 7.4. Advances in Graph Theory, Clone Theory and Multiple-Valued Logic

**Participants:** Quentin Brabant, Miguel Couceiro, Amedeo Napoli, François Pirot, Chedy Raïssi, Jean-Sébastien Sereni.

**Keywords:** graph theory, graph colouring, extremal graph theory, chromatic number, multiple-valued logic, clone theory

### 7.4.1. Graph Theory

Proper colouring of triangle-free planar graphs is an active research topic with interesting algorithmic ramifications. It has been known for more than fifty years that such graphs can be properly 3-coloured, and Thomassen conjectured in 2007 that they actually admit an exponential number of such colourings. This statement is still wide open, and to bring forward further insight we established [75] it to be equivalent to the following:

there exists a positive real  $\alpha$  such that whenever  $G$  is a planar graph and  $A$  is a subset of its edges whose deletion makes  $G$  triangle-free, there exists a subset  $A'$  of  $A$  of size at least  $\alpha|A|$  such that  $G-(A \setminus A')$  is 3-colourable. This equivalence allows us to study restricted situations, where we can prove the statement to be true.

Still on graph colouring, we demonstrated [93] a conjecture by Zhang and Whu made in 2011, that for every positive integer  $\Delta$ , every  $K_4$ -minor-free graph with maximum degree  $\Delta$  admits an equitable colouring with  $k$  colours whenever  $k \geq \frac{\Delta+3}{2}$ . A key ingredient was to *not* use the discharging method and rather exploit decomposition trees of  $K_4$ -minor-free graphs.

We also considered [88] distance colouring in graphs of maximum degree at most  $d$  and how excluding one fixed cycle of length  $\ell$  affects the number of colours required as  $d \rightarrow \infty$ . For vertex-colouring and  $t \geq 1$ , if any two distinct vertices connected by a path of at most  $t$  edges are required to be coloured differently, then a reduction by a logarithmic (in  $d$ ) factor against the trivial bound  $O(d^t)$  can be obtained by excluding an odd cycle length  $\ell \geq +3t$  if  $t$  is odd or by excluding an even cycle length  $\ell \geq 2t + 2$ . For edge-colouring and  $t \geq 2$ , if any two distinct edges connected by a path of fewer than  $t$  edges are required to be coloured differently, then excluding an even cycle length  $\ell \geq 2t$  is sufficient for a logarithmic factor reduction. For  $t \geq 2$ , neither of the above statements are possible for other parity combinations of  $\ell$  and  $t$ . These results can be considered extensions of results due to Johansson (1996) and Mahdian (2000), and are related to open problems of Alon and Mohar (2002) and Kaiser and Kang (2014).

### 7.4.2. Multiple-Valued Logic and (Partial) Clone Theory

Clone theory was primarily motivated by the study of Boolean logic, and it currently constitutes a major subject in universal algebra, multiple-valued logic, and theoretical computer science. A clone on a set  $A$  is a class of functions  $f : A^n \rightarrow A$ ,  $n \geq 1$ , that contains all projections and that is closed under compositions. Clones on a set  $A$  constitute a closure system, in fact, an algebraic lattice where meet is given by set-intersection. Clones on a 2-element set were completely classified by Emil Post. Since Post’s classification several studies on clone theory have appeared and many variants and generalizations have been proposed.

As a closure system, clones can be specified within a Galois framework, namely, through the well known Pol-Inv Galois connection by the polarity between functions and the relations they preserve. This Galois connection became the main tool in several studies, in particular, in the classification of the complexity classes of CSPs (“Constraint Satisfaction Problems”) [92]. Another, rather surprisingly, application of this Galois framework led to the description of of analogy-preserving Boolean classifiers [4].

Similarly, clones of partial functions (i.e., functions  $f : D \rightarrow A$  for  $D \subseteq A^n$ ) can be described by the relations its members preserve. Unlike the lattice of clones, the lattice of partial clones is of continuum cardinality even in the case of 2-element underlying sets. This shows that a complete description of this lattice is hard to attain. However, many efforts have been made towards local descriptions of this lattice, for instance, concerning the classification of its intervals that has entailed a long lasting open problem. This was settled [20] in the form of a dichotomy theorem showing that such intervals are either finite or of continuum cardinality, and we presented precise descriptions of the structure some challenging intervals in [21]. Further developments and related problems were also tackled in [11], [24], [40], [39].

## 8. Partnerships and Cooperations

### 8.1. Regional Initiatives

#### 8.1.1. *Hydreos*

**Participant:** Jean-François Mari.

Hydreos is a state organization -a so-called “Pôle de compétitivité”- aimed at evaluating the delivering and the quality of water (<http://www.hydreos.fr/fr>). Actually, data about water resources rely on many agronomic variables, including land use successions. The data to be analyzed are obtained by surveys or by satellite images and describe the land use at the level of the agricultural parcel. Then there is a search for detecting changes in land use and for correlating these changes to groundwater quality. Accordingly, one main challenge in our participation in Hydreos is to process and analyze space-time data for reaching a better understanding of the changes in the organization of a territory. The systems ARPEnTAge and CarottAge are used in this context, especially by agronomists of INRA (ASTER Mirecourt <http://www6.nancy.inra.fr/sad-aster>).

#### 8.1.2. *The Smart Knowledge Discovery Project*

**Participants:** Jérémie Nevin, Amedeo Napoli, Chedy Raïssi.

The SKD project for “Smart Knowledge Discovery” aims at analyzing complex industrial data for troubleshooting and decision making, and is funded by “Grand Est Region”. We are working with the Vize company –which is based in Nancy and specialized in visualization-based data mining– on exploratory knowledge discovery. The data which are under study are provided by the Arcelor-Mittal Steel Company and are related to the monitoring of rolling mills. Data are complex time series and the problem can be set as follows: problem statement, data access and preparation, design of adapted knowledge discovery methods based on symbolic and numerical methods, interaction with analysts, tests and validation. For the Orpailleur team, one main objective of SKD is to combine sequence mining and visualization tools for recognizing and then preventing the occurrences of defects in the outputs of the rolling mills.

### 8.2. National Initiatives

#### 8.2.1. *ANR*

##### 8.2.1.1. *Elker (2017–2020)*

**Participants:** Miguel Couceiro, Esther Catherine Galbrun, Amedeo Napoli, Chedy Raïssi.

The objectives of the new ELKER ANR Research Project is to study, formalize and implement the search for link keys in RDF data. Link keys generalize database keys in two independent directions, i.e. they deal with RDF data and they apply across two datasets. The goal of ELKER is to study the automatic discovery of link keys and reasoning with link keys, especially in taking an FCA point of view. One main idea is to rely on the competencies of Orpailleur in FCA for solving the problem using FCA and pattern structures algorithms, especially those related to the discovery of functional dependencies. This project involves the EPI Orpailleur at Inria Nancy Grand Est, the EPI MOEX at Inria Rhône Alpes, and LIASD at Université Paris 8.

### 8.2.1.2. ISTEEX (2014–2017)

**Participant:** Yannick Toussaint.

ISTEX is a so-called “Initiative d’excellence” managed by CNRS and DIST (“Direction de l’Information Scientifique et Technique”). ISTEEX aims at providing the research and teaching community an on-line access to scientific publications in all domains (<http://www.istex.fr/istex-excellence-initiative-of-scientific-and-technical-information/>). In this way, ISTEEX requires a massive acquisition of documents such as journals, proceedings, corpora, and databases. The Orpailleur team was especially involved in the development of facilities for querying full-text documentation, analyzing content and extracting information. The project was carried out in collaboration with the ATILF laboratory and the INIST Institute (both located in Nancy).

### 8.2.1.3. PractiKPharma (2016–2020)

**Participants:** Adrien Coulet, Joël Legrand, Pierre Monnin, Amedeo Napoli, Malika Smail-Tabbone, Yannick Toussaint.

PractiKPharma for “Practice-based evidences for actioning Knowledge in Pharmacogenomics” is an ANR research project (<http://praktikpharma.loria.fr/>) about the validation of domain knowledge in pharmacogenomics. Pharmacogenomics is interested in understanding how genomic variations related to patients have an impact on drug responses. Most of the available knowledge in pharmacogenomics (state of the art) lies in biomedical literature, with various levels of validation. An originality of PractiKPharma is to use Electronic Health Records (EHRs) to constitute cohorts of patients. These cohorts are then mined for extracting potential pharmacogenomics patterns to be then validated w.r.t. literature knowledge for becoming actionable knowledge units. More precisely, firstly we should extract pharmacogenomic patterns from the literature and secondly we should confirm or moderate the interpretation and validation of these units by mining EHRs. Comparing knowledge patterns extracted from the literature with facts extracted from EHRs is a complex task depending on the EHR language –literature is in English whereas EHRs are in French– and on knowledge level, as EHRs represent observations at the patient level whereas literature is related to sets of patients. The PractiKPharma involves three other laboratories, namely LIRMM in Montpellier, SSPIM in St-Etienne and CRC in Paris.

## 8.2.2. CNRS PEPS and Mastodons projects

### 8.2.2.1. Mastodons Projects: from HyQual to HyQualiBio (2016–2018)

**Participants:** Miguel Couceiro, Esther Catherine Galbrun, Tatiana Makhalova, Amedeo Napoli, Chedy Raïssi, Justine Reynaud.

The HyQual project was proposed in 2016 in response to the Mastodons CNRS Call about data quality in data mining (see <http://www.cnrs.fr/mi/spip.php?article819&lang=fr>). This project is interested in the mining of nutritional data for discovering predictive biomarkers of diabetes and metabolic syndrome in elder populations. The considered data mining methods are hybrid, and they combine symbolic and numerical methods for mining complex and noisy metabolic data [80]. Regarding the mining process, we are interested in the quality of the data at hand and in the discovered patterns. In particular, we check the incompleteness of the data, the quality of the extracted rules and the possible existence of re-descriptions.

Initially, the project involved researchers from the EPI Orpailleur, with researchers from LIRIS Lyon, ICube Strasbourg, and INRA Clermont-Ferrand. This year, we were merged with another Mastodons project, namely QualiBioConsensus, about the “ranking of biological data using consensus ranking techniques”. The joint Mastodons project is now called “HyQualiBio”. The topics of interest for the participants are the mining of complex biological data, rankings and ties in rankings, and the search of dependencies in the web of data.

### 8.2.2.2. PEPS Decade

**Participants:** Miguel Couceiro, Esther Catherine Galbrun, Nyoman Juniarta, Amedeo Napoli, Justine Reynaud, Chedy Raïssi.

Decade stands for “Découverte et exploitation des connaissances pour l’aide à la décision en chimie thérapeutique”. The objective of the CNRS PEPS Decade project is to study the basis of knowledge system for analyzing the so-called PAINS (“Pan Assay Interference Compounds”) in chemistry. The system should rely on the knowledge possibly discovered in the data and domain knowledge and expertise. The members of the projects are interested in data mining techniques guided by constraints and preferences, “instant data mining”, subgroup discovery and exceptional model mining. All these topics were already of interest in the PEPS Prefute (2015-2016) which was about interaction and iteration in the knowledge discovery process.

The members of the Decade project are from Greyc Caen, LIFO Orléans LIRIS Lyon, Université de Tours-Blois, EPI Lacodam in Rennes and EPI Orpailleur (in association with chemists based in Caen and Orléans)

## 8.3. European Initiatives

### 8.3.1. FP7 & H2020 Projects

#### 8.3.1.1. CrossCult (H2020 Project, 2016-2020)

**Participants:** Miguel Couceiro, Nyoman Juniarta, Amedeo Napoli, Chedy Raïssi.

CrossCult (<http://www.crosscult.eu/>) aims to make reflective history a reality in the European cultural context, by enabling the re-interpretation of European (hi)stories through cross-border interconnections among cultural digital resources, citizen viewpoints and physical venues. The project has two main goals. The first goal is to lower cultural EU barriers and create unique cross-border perspectives, by connecting existing digital historical resources and by creating new ones through the participation of the public. The second goal is to provide long-lasting experiences of social learning and entertainment that will help for achieving a better understanding and re-interpretation of European history. To achieve these goals, CrossCult will use cutting-edge technology to connect existing digital cultural assets and to combine them with interactive experiences that all together are intended to increase retention, stimulate reflection and help European citizens appreciate their past and present in a holistic manner. CrossCult will be implemented on four real-world flagship pilots involving a total of 8 sites across Europe.

The role of the Orpailleur Team (in conjunction with the LORIA Kiwi Team) is to work on knowledge discovery and recommendation. The focus is on the mining of visitor trajectories for analysis purposes, and on the definition of a visitor profile in connection with domain knowledge for recommendation.

The numerous partners of the Orpailleur team in the CrossCult project are: Luxembourg Institute for Science and Technology and Centre Virtuel de la Connaissance sur l’Europe (Luxembourg, leader of the project), University College London (England), University of Malta (Malta), University of Peloponnese and Technological Educational Institute of Athens (Greece), Università degli Studi di Padova (Italy), University of Vigo (Spain), National Gallery (London, England), and GVAM Guías Interactivas (Spain).

## 8.4. International Initiatives

### 8.4.1. Inria Associate Teams Not Involved in an Inria International Lab

#### 8.4.1.1. Snowball

Title: Discovering knowledge on drug response variability by mining electronic health records

International Partner (Institution - Laboratory - Researcher):

Stanford (United States) - Department of Medicine, Stanford Center for Biomedical Informatics Research (BMIR) - Nigam Shah

Start year: 2017

See also: <http://snowball.loria.fr/>

Snowball (2017-2019) is an Inria Associate Team and the continuation of the preceding Associate Team called Snowflake (2014-2016). The objective of Snowball is to study drug response variability through the lens of Electronic Health Records (EHRs) data. This is motivated by the fact that many factors, genetic as well as environmental, imply different responses from people to the same drug. The mining of EHRs can bring substantial elements for understanding and explaining drug response variability.

Accordingly the objectives of Snowball are to identify in EHR repositories groups of patients which are responding differently to similar treatments, and then to characterize these groups and predict patient drug sensitivity. These objectives are complementary to those of the PractiKPharma ANR project. Moreover, it should be noticed that Adrien Coulet has started a one-year sabbatical stay in the lab of Nigam Shah at Stanford University since September 2017.

## 8.4.2. Participation in International Programs

### 8.4.2.1. A stay at NASA Frontier Development Lab

In July 2017, Chedy Raïssi visited NASA Ames and SETI Institute as part of the Frontier Development Lab. He worked on mentoring teams and developing meaningful research opportunities, as well as supporting the work of the planetary defense community and showing the potential of this kind of applied research methodology to deliver breakthrough of significant value.

Delay-Doppler radar imaging is a powerful technique to characterize the trajectories, shapes, and spin states of near-Earth asteroids and has yielded detailed models of dozens of objects. Since the 1990s, delay-Doppler data has been analyzed using the SHAPE software developed originally by Steven J. Ostro. SHAPE performs sequential single-parameter fitting, and requires considerable computation runtime and human intervention. Recently, multiple-parameter fitting algorithms have been shown to more efficiently invert delay-Doppler datasets thus decreasing runtime while improving accuracy. However, reconstructing asteroid shapes and spins from delay-Doppler data is, like many inverse problems, computationally intensive and requires extensive human oversight of the shape modeling process.

Thus we have explored two new techniques to better automate delay-Doppler shape modeling: Bayesian optimization and deep generative models. Firstly we have implemented a Bayesian optimization routine that uses SHAPE to autonomously search the space of spin-state parameters. Bayesian optimization yielded similar spin state constraints with computer runtime reduced by a factor of 3. Secondly, the shape modeling process could be further accelerated using a deep “*generative model*” to replace or complete iterative fitting. Accordingly, we have implemented and trained a deep generative model based on different architectures of deep convolutional networks. Results are currently under analysis and future publications are in preparation.

### 8.4.2.2. LEA STRUCO

**Participant:** Jean-Sébastien Sereni.

LEA STRUCO is an “Associated International Laboratory” of CNRS between IÚUK, Prague, and LIAFA, Paris. It focuses on high-level study of fundamental combinatorial objects, with a particular emphasis on comprehending and disseminating the state-of-the-art theories and techniques developed. The obtained insights shall be applied to obtain new results on existing problems as well as to identify directions and questions for future work. Jean-Sébastien Sereni is the founder and previous director of LEA STRUCO, which was initiated when Jean-Sébastien was a member of LIAFA, and he is now a member of its scientific committee.

### 8.4.2.3. Research Collaboration with HSE Moscow

**Participants:** Miguel Couceiro, Adrien Coulet, Tatiana Makhalova, Amedeo Napoli, Chedy Raïssi, Justine Reynaud.

An on-going collaboration involves the Orpailleur team and Sergei O. Kuznetsov at Higher School of Economics in Moscow (HSE). Amedeo Napoli visited HSE laboratory several times while Sergei O. Kuznetsov visits Inria Nancy Grand Est every year. The collaboration is materialized by the joint supervision of students (such as the thesis of Aleksey Buzmakov defended in 2015 and the on-going thesis of Tatiana Makhalova), and the organization of scientific events, as the workshop FCA4AI with five editions between 2012 and 2016 (see <http://www.fca4ai.hse.ru>).

This year, we participated in the organization of two main events: a special session about Knowledge Discovery and Formal Concept Analysis at the ISMIS Conference in Warsaw (Poland) in June 2017 ([http://ismis2017.ii.pw.edu.pl/s\\_kd\\_fca.php](http://ismis2017.ii.pw.edu.pl/s_kd_fca.php)), and the chairing of the track “General Topics of Data Analysis” at the AIST Conference in Moscow in July 2017 (6th International Conference on Analysis of Images, Social Networks, and Texts <http://aistconf.org/>). Finally a next edition of the seventh edition of the FCA4AI workshop is planned in July 2018 at the ECAI-IJCAI Conference to be held in Stockholm Sweden.

## 9. Dissemination

### 9.1. Promoting Scientific Activities

#### 9.1.1. Scientific Events Organization, General Chairs, Scientific Chairs

- The 25<sup>th</sup> international conference on case-based reasoning (ICCBR-2017) has taken place in Trondheim (Norway) from 26 to 28 June (<http://www.iccbr.org/iccbr17/>). The program chairs were David W. Aha and Jean Lieber.
- Amedeo Napoli was the co-chair with Davide Ciucci and Sergei Kuznetsov of the special session “Special Session on Knowledge Discovery with Formal Concept Analysis and related formalisms (FCA4KD++)”, held between June 26-29 at ISMIS Warsaw (23rd International Symposium on Methodologies for Intelligent Systems [http://ismis2017.ii.pw.edu.pl/s\\_kd\\_fca.php](http://ismis2017.ii.pw.edu.pl/s_kd_fca.php)).
- Amedeo Napoli was the co-chair with Sergei Kuznetsov of the track “General Topics of Data Analysis” at the AIST Conference in Moscow in July 2017 (6th International Conference on Analysis of Images, Social Networks, and Texts <http://aistconf.org/>).
- Amedeo Napoli was the general chair of BDA 2017 held between November 14-17 2017 at Inria Nancy Grand Est/LORIA, “33ième conférence sur la Gestion de Données, Principes, Technologies et Applications” <https://project.inria.fr/bda2017/>.

##### 9.1.1.1. Scientific Animation

- The scientific animation in the Orpailleur team is based on the Team Seminar which is called the “Malotec” seminar (<http://malotec.loria.fr/?p=1>). The Malotec seminar is held in general twice a month and is used either for general presentations of members of the team or for invited presentations of external researchers.
- Members of the Orpailleur team are all involved, as members or as head persons, in various national research groups.
- The members of the Orpailleur team are involved in the organization of conferences and workshops, as members of conference program committees (AAAI, ECAI, ECML-PKDD, ESWC, ICCBR, ICDM, ICFA, IJCAI, ISWC, KDD, SDM...), as members of editorial boards, and finally in the organization of journal special issues.

### 9.2. Teaching - Supervision - Juries

- All the permanent members of the Orpailleur team are involved in teaching at all levels and mainly at University of Lorraine. Actually, most of the members of the Orpailleur team are employed on “Université de Lorraine” positions.
- The members of the Orpailleur team are also involved in student supervision, at all university levels, from under-graduate until post-graduate students, engineers, PhD, postdoc students.
- Finally, the permanent members of the Orpailleur team are involved in HDR and thesis defenses, being thesis referees or thesis committee members.



## 10. Bibliography

### Major publications by the team in recent years

- [1] G. BOSC, J.-F. BOULICAUT, C. RAÏSSI, M. KAYTOUE. *Anytime Discovery of a Diverse Set of Patterns with Monte Carlo Tree Search*, in "Data Mining and Knowledge Discovery", 2017 [DOI : 10.1007/s10618-017-0547-5], <https://hal.archives-ouvertes.fr/hal-01662857>
- [2] O. BRUNEAU, E. GAILLARD, N. LASOLLE, J. LIEBER, E. NAUER, J. REYNAUD. *A SPARQL Query Transformation Rule Language - Application to Retrieval and Adaptation in Case-Based Reasoning*, in "ICCBR 2017 - Case-Based Reasoning Research and Development, 25th International Conference on Case-Based Reasoning", Trondheim, Norway, June 2017, <https://hal.inria.fr/hal-01661651>
- [3] A. BUZMAKOV, S. O. KUZNETSOV, A. NAPOLI. *Efficient Mining of Subsample-Stable Graph Patterns*, in "ICDM 2017 - 17th IEEE International Conference on Data Mining", New Orleans, United States, November 2017, pp. 1-6, <https://hal.inria.fr/hal-01668663>
- [4] M. COUCEIRO, N. HUG, H. PRADE, G. RICHARD. *Analogy-preserving Functions: A Way to Extend Boolean Samples*, in "Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017", Melbourne, Australia, August 2017, <https://hal.inria.fr/hal-01668230>
- [5] K. DALLEAU, Y. MARZOUGUI, S. DA SILVA, P. RINGOT, N. C. NDIAYE, A. COULET. *Learning from biomedical linked data to suggest valid pharmacogenes*, in "Journal of Biomedical Semantics", 2017, vol. 75 [DOI : 10.1007/s10994-008-5089-z], <https://hal.inria.fr/hal-01511773>
- [6] E. GAILLARD, J. LIEBER, E. NAUER. *TAAABLE: a case-based reasoning system which adapts cooking recipes*, in "Revue des Sciences et Technologies de l'Information - Série RIA : Revue d'Intelligence Artificielle", April 2017, vol. 31, n<sup>o</sup> 1-2, pp. 207 - 235 [DOI : 10.3166/RIA.31.207-235], <https://hal.inria.fr/hal-01657387>
- [7] E. GALBRUN, B. GOLSHAN, A. GIONIS, E. TERZI. *Finding low-tension communities*, in "Proceedings of the 17th SIAM International Conference on Data Mining, SDM'17", Houston, TX, United States, April 2017, <https://arxiv.org/abs/1701.05352> , <https://hal.archives-ouvertes.fr/hal-01446461>
- [8] J.-F. MARI, A. GOBILLOT, M. BENOÎT. *Time Space Simulation of Land Use changes by stochastic modeling*, in "Revue Internationale de Géomatique", 2017, forthcoming, <https://hal.inria.fr/hal-01662140>

### Publications of the year

#### Doctoral Dissertations and Habilitation Theses

- [9] M. HASSAN. *Knowledge Discovery Considering Domain Literature and Ontologies : Application to rare Diseases*, Université de Lorraine, July 2017, <https://hal.inria.fr/tel-01678860>

#### Articles in International Peer-Reviewed Journals

- [10] M. ALAM, A. BUZMAKOV, A. NAPOLI. *Exploratory Knowledge Discovery over Web of Data*, in "Discrete Applied Mathematics", 2017, pp. 1-25, forthcoming, <https://hal.inria.fr/hal-01673439>

- 
- [11] J. ALMEIDA, M. COUCEIRO, T. WALDHAUSER. *On the topological semigroup of equational classes of finite functions under composition*, in "Journal of Multiple-Valued Logic and Soft Computing", 2017, vol. 28, n<sup>o</sup> 1, pp. 5-28, <https://hal.archives-ouvertes.fr/hal-01090645>
- [12] J. BAIXERIES, V. CODOCEDO, M. KAYTOUE, A. NAPOLI. *Characterizing Approximate-Matching Dependencies in Formal Concept Analysis with Pattern Structures*, in "Discrete Applied Mathematics", 2017, pp. 1-26, forthcoming, <https://hal.inria.fr/hal-01673441>
- [13] G. BOSC, J.-F. BOULICAUT, C. RAÏSSI, M. KAYTOUE. *Anytime Discovery of a Diverse Set of Patterns with Monte Carlo Tree Search*, in "Data Mining and Knowledge Discovery", 2017 [DOI : 10.1007/s10618-017-0547-5], <https://hal.archives-ouvertes.fr/hal-01662857>
- [14] G. BOSC, P. TAN, J.-F. BOULICAUT, C. RAÏSSI, M. KAYTOUE. *A Pattern Mining Approach to Study Strategy Balance in RTS Games*, in "IEEE Transactions on Computational Intelligence and AI in games", June 2017, vol. 9, n<sup>o</sup> 2, pp. 123-132 [DOI : 10.1109/TCIAIG.2015.2511819], <https://hal.archives-ouvertes.fr/hal-01252728>
- [15] Q. BRABANT, M. COUCEIRO. *k-maxitive Sugeno integrals as aggregation models for ordinal preferences*, in "Fuzzy Sets and Systems", June 2017 [DOI : 10.1016/j.fss.2017.06.005], <https://hal.archives-ouvertes.fr/hal-01657107>
- [16] C. CHAMARD-JOVENIN, C. THIEBAUT, A. CHESNEL, E. BRESSO, C. MOREL, M. SMAÏL-TABBONE, M.-D. DEVIGNES, T. BOUKHOBZA, H. DUMOND. *Low-dose alkylphenol exposure promotes mammary epithelium alterations and transgenerational developmental defects, but does not enhance tumorigenic behaviour of breast cancer cells*, in "Frontiers in Endocrinology", October 2017, vol. 8, 272 p. [DOI : 10.3389/fendo.2017.00272], <https://hal.archives-ouvertes.fr/hal-01609240>
- [17] M. COUCEIRO, J. DEVILLET, J.-L. MARICHAL. *Characterizations of idempotent discrete uninorms*, in "Fuzzy Sets and Systems", 2017, <https://arxiv.org/abs/1701.07253v1> , <https://hal.inria.fr/hal-01447513>
- [18] M. COUCEIRO, S. FOLDES, G. C. MELETIOU. *On homomorphisms between products of median algebras*, in "Algebra Universalis", 2017, 8 p. , <https://hal.inria.fr/hal-01500223>
- [19] M. COUCEIRO, M. GRABISCH. *On integer-valued means and the symmetric maximum*, in "Aequationes Mathematicae", April 2017, vol. 91, n<sup>o</sup> 2, pp. 353–371 [DOI : 10.1007/s00010-016-0460-9], <https://hal.inria.fr/hal-01404593>
- [20] M. COUCEIRO, L. HADDAD, K. SCHÖLZEL, T. WALDHAUSER. *A Solution to a Problem of D. Lau: Complete Classification of Intervals in the Lattice of Partial Boolean Clones*, in "Journal of Multiple-Valued Logic and Soft Computing", 2017, vol. 28, n<sup>o</sup> 1, pp. 47-58, <https://hal.inria.fr/hal-01183004>
- [21] M. COUCEIRO, L. HADDAD, K. SCHÖLZEL, T. WALDHAUSER. *On the interval of strong partial clones of Boolean functions containing  $Pol((0,0),(0,1),(1,0))$* , in "Algebra Universalis", 2017 [DOI : 10.1007/s00012-016-0418-8], <https://hal.archives-ouvertes.fr/hal-01184404>
- [22] M. COUCEIRO, E. LEHTONEN. *Majors of functions*, in "Order", 2017, 14 p. , <https://hal.inria.fr/hal-01519377>
- [23] M. COUCEIRO, G. C. MELETIOU. *On a special class of median algebras*, in "Miskolc Mathematical Notes", 2017, vol. 18, n<sup>o</sup> 1, pp. 167–171, <https://hal.inria.fr/hal-01248142>

- [24] M. COUCEIRO, B. TEHEUX. *Pivotal decomposition schemes inducing clones of operations*, in "Beiträge zur Algebra und Geometrie / Contributions to Algebra and Geometry", 2017, 16 p. , forthcoming, <https://hal.inria.fr/hal-01450835>
- [25] P. COUVINEAU, H. DE ALMEIDA, B. MAIGRET, C. LLORENS-CORTES, X. ITURRIOZ. *Involvement of arginine 878 together with Ca<sup>2+</sup> in aminopeptidase A substrate specificity for N-terminal acidic amino-acid residues*, in "PLoS ONE", September 2017, <https://hal.inria.fr/hal-01580832>
- [26] K. DALLEAU, Y. MARZOUGUI, S. DA SILVA, P. RINGOT, N. C. NDIAYE, A. COULET. *Learning from biomedical linked data to suggest valid pharmacogenes*, in "Journal of Biomedical Semantics", 2017, vol. 75 [DOI : 10.1007/s10994-008-5089-z], <https://hal.inria.fr/hal-01511773>
- [27] E. GAILLARD, J. LIEBER, E. NAUER. *TAAABLE: a case-based reasoning system which adapts cooking recipes*, in "Revue des Sciences et Technologies de l'Information - Série RIA : Revue d'Intelligence Artificielle", April 2017, vol. 31, n<sup>o</sup> 1-2, pp. 207 - 235 [DOI : 10.3166/RIA.31.207-235], <https://hal.inria.fr/hal-01657387>
- [28] J.-F. MARI, A. GOBILLOT, M. BENOÎT. *Time Space Simulation of Land Use changes by stochastic modeling*, in "Revue Internationale de Géomatique", 2017, forthcoming, <https://hal.inria.fr/hal-01662140>
- [29] G. PERSONENI, E. BRESSO, M.-D. DEVIGNES, M. DUMONTIER, M. SMAÏL-TABBONE, A. COULET. *Discovering associations between adverse drug events using pattern structures and ontologies*, in "Journal of Biomedical Semantics", 2017, vol. 93, pp. 539 - 546 [DOI : 10.1038/CLPT.2013.24], <https://hal.inria.fr/hal-01576341>
- [30] W. UGARTE ROJAS, P. BOIZUMAULT, B. CRÉMILLEUX, A. LEPAILLEUR, S. LOUDNI, M. PLANTEVIT, C. RAÏSSI, A. SOULET. *Skypattern mining: From pattern condensed representations to dynamic constraint satisfaction problems*, in "Artificial Intelligence", March 2017, vol. 244, pp. 48–69 [DOI : 10.1016/J.ARTINT.2015.04.003], <https://hal.inria.fr/hal-01188928>

### International Conferences with Proceedings

- [31] Y. ABID, A. IMINE, A. NAPOLI, C. RAÏSSI, M. RUSINOWITCH. *Two-phase preference disclosure in attributed social networks*, in "DEXA 2017 - 28th International Conference on Database and Expert Systems Applications", Lyon, France, LNCS, Springer, August 2017, vol. 10438, pp. 249-263 [DOI : 10.1007/978-3-319-64468-4\_19], <https://hal.inria.fr/hal-01649246>
- [32] Q. BRABANT, M. COUCEIRO, A. NAPOLI, J. REYNAUD. *From Meaningful Orderings in the Web of Data to Multi-level Pattern Structures*, in "International Symposium on Methodologies for Intelligent Systems", Warsaw, Poland, Foundations of Intelligent Systems: 23rd International Symposium, June 2017, pp. 622-631, <https://hal.archives-ouvertes.fr/hal-01544123>
- [33] O. BRUNEAU, E. GAILLARD, N. LASOLLE, J. LIEBER, E. NAUER, J. REYNAUD. *A SPARQL Query Transformation Rule Language — Application to Retrieval and Adaptation in Case-Based Reasoning*, in "ICCBR 2017 - Case-Based Reasoning Research and Development, 25th International Conference on Case-Based Reasoning", Trondheim, Norway, June 2017, <https://hal.inria.fr/hal-01661651>

- [34] A. BUZMAKOV, S. O. KUZNETSOV, A. NAPOLI. *Efficient Mining of Subsample-Stable Graph Patterns*, in "ICDM 2017 - 17th IEEE International Conference on Data Mining", New Orleans, United States, November 2017, pp. 1-6, <https://hal.inria.fr/hal-01668663>
- [35] M. B. CHAWKI, E. NAUER, N. JAY, J. LIEBER. *Tetra: A Case-Based Decision Support System for Assisting Nuclear Physicians with Image Interpretation*, in "ICCBR 2017 - International Conference on Case-Based Reasoning", Trondheim, Norway, Lecture Notes in Computer Science, Springer, June 2017, vol. 10339, pp. 108-122, <https://hal.inria.fr/hal-01657414>
- [36] V. CODOCEDO, G. BOSC, M. KAYTOUE, J.-F. BOULICAUT, A. NAPOLI. *A Proposition for Sequence Mining Using Pattern Structures*, in "ICFCA 2017 - 14th International Conference on Formal Concept Analysis", Rennes, France, K. BERTET, D. BORCHMANN, P. CELLIER, S. FERRÉ (editors), Lecture Notes in Computer Science, Springer, June 2017, vol. 10308, pp. 106-121 [DOI : 10.1007/978-3-319-59271-8\_7], <https://hal.archives-ouvertes.fr/hal-01549107>
- [37] M. COUCEIRO, J. DEVILLET, J.-L. MARICHAL. *On Idempotent Discrete Uninorms*, in "International Summer School on Aggregation Operators", Skövde, Sweden, Aggregation Functions in Theory and in Practice (Springer), June 2017, vol. 308, pp. 147 - 153 [DOI : 10.1007/978-3-319-59306-7\_15], <https://hal.inria.fr/hal-01551089>
- [38] M. COUCEIRO, D. DUBOIS, H. PRADE, A. RICO. *Enhancing the expressive power of Sugeno integrals for qualitative data analysis*, in "EUSFLAT 2017 - 10th Conference of the European Society for Fuzzy Logic and Technology", Warsaw, Poland, IWIFSGN 2017, EUSFLAT 2017: Advances in Fuzzy Logic and Technology 2017, Springer, Cham, September 2017, vol. 641, pp. 534-547 [DOI : 10.1007/978-3-319-66830-7\_48], <https://hal.inria.fr/hal-01668229>
- [39] M. COUCEIRO, L. HADDAD, V. LAGERKVIST, B. ROY. *On the Interval of Boolean Strong Partial Clones Containing Only Projections as Total Operations*, in "47th IEEE International Symposium on Multiple-Valued Logic (ISMVL)", Novi Sad, Serbia, IEEE Computer Society, May 2017, <https://hal.inria.fr/hal-01504012>
- [40] M. COUCEIRO, L. HADDAD, K. SCHÖLZEL. *On the nonexistence of minimal strong partial clones*, in "ISMVL 2017 - 47th IEEE International Symposium on Multiple-Valued Logic", Novi Sad, Serbia, IEEE Computer Society, May 2017, 6 p. , <https://hal.inria.fr/hal-01504011>
- [41] M. COUCEIRO, N. HUG, H. PRADE, G. RICHARD. *Analogy-preserving Functions: A Way to Extend Boolean Samples*, in "IJCAI 2017 - 26th International Joint Conference on Artificial Intelligence", Melbourne, Australia, August 2017, pp. 1-7, <https://hal.inria.fr/hal-01668230>
- [42] M. COUCEIRO, P. MERCURIALI, R. PÉCHOUX, A. SAFFIDINE. *Median based calculus for lattice polynomials and monotone Boolean functions*, in "ISMVL 2017 - 47th IEEE International Symposium on Multiple-Valued Logic", Novi Sad, Serbia, IEEE Computer Society, May 2017, 6 p. , <https://hal.inria.fr/hal-01504010>
- [43] E. GAILLARD, J. LIEBER, E. NAUER. *Adaptation of TAAABLE to the CCC'2017 Mixology and Salad Challenges, Adaptation of the Cocktail Names*, in "International Conference on Case-Based Reasoning (ICCBR) Computer Cooking Contest Workshop", Trondheim, Norway, June 2017, pp. 253-268, <https://hal.inria.fr/hal-01667183>

- [44] E. GALBRUN, B. GOLSHAN, A. GIONIS, E. TERZI. *Finding low-tension communities*, in "Proceedings of the 17th SIAM International Conference on Data Mining, SDM'17", Houston, TX, United States, April 2017, <https://arxiv.org/abs/1701.05352> , <https://hal.archives-ouvertes.fr/hal-01446461>
- [45] P. MONNIN, C. JONQUET, J. LEGRAND, A. NAPOLI, A. COULET. *PGxO: A very lite ontology to reconcile pharmacogenomic knowledge units*, in "Methods, tools & platforms for Personalized Medicine in the Big Data Era", Palermo, Italy, NETTAB 2017 Workshop Collection, October 2017 [DOI : 10.7287/PEERJ.PREPRINTS.3140v1], <https://hal.inria.fr/hal-01593184>
- [46] P. MONNIN, M. LEZOCHÉ, A. NAPOLI, A. COULET. *Using formal concept analysis for checking the structure of an ontology in LOD: the example of DBpedia*, in "23rd International Symposium on Methodologies for Intelligent Systems, ISMIS 2017", Warsaw, Poland, June 2017, <https://hal.inria.fr/hal-01511909>
- [47] J. REYNAUD, M. ALAM, Y. TOUSSAINT, A. NAPOLI. *A Proposal for Classifying the Content of the Web of Data Based on FCA and Pattern Structures*, in "International Symposium on Methodologies for Intelligent Systems", Warsaw, Poland, June 2017, <https://hal.inria.fr/hal-01667437>
- [48] M. SCHNELL, S. COUFFIGNAL, J. LIEBER, S. SALEH, N. JAY. *Case-Based Interpretation of Best Medical Coding Practices — Application to Data Collection for Cancer Registries*, in "ICCBR 2017 - 25th International Conference on Case-Based Reasoning", Trondheim, Norway, ICCBR 2017: Case-Based Reasoning Research and Development, Springer, June 2017, vol. 10339, pp. 345-359 [DOI : 10.1007/978-3-319-61030-6\_24], <https://hal.inria.fr/hal-01661107>

### National Conferences with Proceedings

- [49] G. BOSC, J.-F. BOULICAUT, C. RAÏSSI, M. KAYTOUE. *Découverte de sous-groupes avec les arbres de recherche de Monte Carlo*, in "Extraction et Gestion de Connaissances EGC 2017", Grenoble, France, January 2017, <https://hal.archives-ouvertes.fr/hal-01433054>
- [50] M. COUCEIRO, J. DEVILLET, J.-L. MARICHAL. *On idempotent discrete uninorms*, in "LFA 2017 – 26èmes Rencontres Francophones sur la Logique Floue et ses Applications, Oct 2017," , Amiens, France, October 2017, <https://hal.inria.fr/hal-01671011>
- [51] M. COUCEIRO, D. DUBOIS, H. PRADE, A. RICO. *Generalized Sugeno integrals in qualitative data analysis*, in "LFA 2017 - 26èmes Rencontres Francophones sur la Logique Floue et ses Applications", Amiens, France, October 2017, <https://hal.inria.fr/hal-01668232>
- [52] M. COUCEIRO, P. MERCURIALI, R. PÉCHOUX. *On the efficiency of normal form systems of Boolean functions*, in "LFA 2017 - 26èmes Rencontres Francophones sur la Logique Floue et ses Applications", Amiens, France, October 2017, pp. 1-8, <https://hal.inria.fr/hal-01656033>
- [53] Y. PER, K. DALLEAU, M. SMAÏL-TABBONE. *Multi-class classification for Prediction of tree diseases*, in "EGC 2017 - Extraction et Gestion des Connaissances", Grenoble, France, Revue des Nouvelles Technologies de l'Information, January 2017, vol. RNTI-E-33, pp. 237-248, <https://hal.inria.fr/hal-01611504>

### Conferences without Proceedings

- [54] M. COUCEIRO, N. HUG, H. PRADE, G. RICHARD. *Analogy-preserving fonctions: A way to extend Boolean samples*, in "11e Journées d'Intelligence Artificielle Fondamentale", Caen, France, July 2017, <https://hal.archives-ouvertes.fr/hal-01577828>

- [55] J. LEGRAND, Y. TOUSSAINT, C. RAÏSSI, A. COULET. *Tree-LSTM and Cross-Corpus Training for Extracting Biomedical Relationships from Text*, in "DLPM2017 Workshop - 2nd International Workshop on Deep Learning for Precision Medicine", Skopje, Macedonia, September 2017, Held in conjunction with ECML-PKDD 2017, <https://hal.inria.fr/hal-01592802>

### Scientific Books (or Scientific Book chapters)

- [56] A. NAPOLI, A. TERMIER. *La fouille de données*, in "Les Big Data à découvert", M. BOUZEGHOUB, R. MOSSERI (editors), CNRS Editions, 2017, pp. 1-3, <https://hal.inria.fr/hal-01673437>

### Books or Proceedings Editing

- [57] D. W. AHA, J. LIEBER (editors). *Case-Based Reasoning Research and Development, 25th International Conference on Case-Based Reasoning (ICCBR 2017)*, June 2017, <https://hal.inria.fr/hal-01661874>

### Other Publications

- [58] S. ARIDHI, S. Z. ALBORZI, M. SMAÏL-TABBONE, M.-D. DEVIGNES, D. RITCHIE. *Neighborhood-Based Label Propagation in Large Protein Graphs*, July 2017, 2 p. , Function SIG @ ISMB/ECCB 2017, Poster, <https://hal.inria.fr/hal-01573381>
- [59] Q. BRABANT, M. COUCEIRO, D. DUBOIS, H. PRADE, A. RICO. *Extracting Decision Rules from Qualitative Data via Sugeno Utility Functionals*, 2017, working paper or preprint, <https://hal.inria.fr/hal-01670924>
- [60] M. COUCEIRO, J. DEVILLET, J.-L. MARICHAL. *Quasitrivial semigroups: characterizations and enumerations*, December 2017, <https://arxiv.org/abs/1709.09162> - working paper or preprint, <https://hal.inria.fr/hal-01668222>
- [61] M. COUCEIRO, P. MERCURIALI, R. PÉCHOUX. *Comparing the efficiency of normal form systems to represent Boolean functions*, June 2017, working paper or preprint, <https://hal.inria.fr/hal-01551761>
- [62] K. DALLEAU, M. COUCEIRO, M. SMAÏL-TABBONE. *Unsupervised Extremely Randomized Trees*, December 2017, working paper or preprint, <https://hal.inria.fr/hal-01667317>
- [63] N. JUNIARTA, A. NAPOLI. *Application of sequential pattern mining to the analysis of visitor trajectories*, November 2017, 33ème conférence sur la Gestion de Données — Principes, Technologies et Applications, Poster, <https://hal.inria.fr/hal-01667442>
- [64] R. J. KANG, F. PIROT. *Distance colouring without one cycle length*, December 2017, <https://arxiv.org/abs/1701.07639> - 14 pages, 1 figure, <https://hal.archives-ouvertes.fr/hal-01668633>
- [65] M. LOEBL, J.-S. SERENI. *Isomorphism of Weighted Trees and Stanley's Conjecture for Caterpillars*, March 2017, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-00992104>
- [66] P. MONNIN, A. NAPOLI, A. COULET. *Discovering Subsumption Axioms with Concept Annotation*, Gestion de Données — Principes, Technologies et Applications (BDA 2017), November 2017, Gestion de Données — Principes, Technologies et Applications (BDA 2017), Poster, <https://hal.inria.fr/hal-01671454>

## References in notes

- [67] C. C. AGGARWAL, C. ZHAI (editors). *Mining Text Data*, Springer, 2012
- [68] F. BAADER, D. CALVANESE, D. MCGUINNESS, D. NARDI, P. PATEL-SCHNEIDER (editors). *The Description Logic Handbook*, Cambridge University Press, Cambridge, UK, 2003
- [69] M. ALAM, A. BUZMAKOV, V. CODOCEDO, A. NAPOLI. *Mining Definitions from RDF Annotations Using Formal Concept Analysis*, in "International Joint Conference in Artificial Intelligence", Buenos Aires, Argentina, Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, July 2015, <https://hal.archives-ouvertes.fr/hal-01186204>
- [70] M. ALAM, T. N. N. LE, A. NAPOLI. *LatViz: A New Practical Tool for Performing Interactive Exploration over Concept Lattices*, in "CLA 2016 - Thirteenth International Conference on Concept Lattices and Their Applications", Moscow, Russia, July 2016, <https://hal.inria.fr/hal-01420751>
- [71] M. BARBUT, B. MONJARDET. *Ordre et classification – Algèbre et combinatoire (2 tomes)*, Hachette, Paris, 1970
- [72] A. CORDIER, V. DUFOUR-LUSSIER, J. LIEBER, E. NAUER, F. BADRA, J. COJAN, E. GAILLARD, L. INFANTE-BLANCO, P. MOLLI, A. NAPOLI, H. SKAF-MOLLI. *Taaable: a Case-Based System for personalized Cooking*, in "Successful Case-based Reasoning Applications-2", S. MONTANI, L. C. JAIN (editors), Studies in Computational Intelligence, Springer, January 2014, vol. 494, pp. 121-162 [DOI : 10.1007/978-3-642-38736-4\_7], <https://hal.inria.fr/hal-00912767>
- [73] S. DA SILVA, F. LE BER, C. LAVIGNE. *Structures de haies dans un paysage agricole : une étude par chemin de Hilbert adaptatif et chaînes de Markov*, in "EGC 2016 – 16èmes Journées Francophones "Extraction et Gestion des Connaissances"", Reims, France, Revue des Nouvelles Technologies de l'Information, January 2016, vol. RNTI-E-30, pp. 279–290, <https://hal.archives-ouvertes.fr/hal-01266344>
- [74] K. DALLEAU, M. COUCEIRO, M.-D. DEVIGNES, C. RAÏSSI, M. SMAÏL-TABBONE. *Using aggregation functions on structured data: a use case in the FIGHT-HF project*, in "International Symposium on Aggregation and Structures (ISAS 2016)", Luxembourg, Luxembourg, G. KISS, J.-L. MARICHAL, B. TEHEUX (editors), International Symposium on Aggregation and Structures (ISAS 2016) - Book of abstracts, July 2016, <https://hal.inria.fr/hal-01399232>
- [75] Z. DVOŘÁK, J.-S. SERENI. *Do triangle-free planar graphs have an exponential number of 3-colorings?*, in "Electronic Journal of Combinatorics", 2017, vol. 24, n<sup>o</sup> 3, #P3.47 p. , <https://hal.archives-ouvertes.fr/hal-01433349v2>
- [76] E. GALBRUN, P. MIETTINEN. *Analysing Political Opinions Using Redescription Mining*, in "Proceedings of the Data Mining in Politics workshop at ICDM 2016, DMiP'16", Barcelona, Spain, December 2016, <https://hal.archives-ouvertes.fr/hal-01399254>
- [77] B. GANTER, S. O. KUZNETSOV. *Pattern Structures and Their Projections*, in "Proceedings of ICCS 2001", LNCS 2120, Springer, 2001, pp. 129–142
- [78] B. GANTER, R. WILLE. *Formal Concept Analysis*, Springer, Berlin, 1999

- [79] M. GRABISCH, J.-L. MARICHAL, R. MESIAR, E. PAP. *Aggregation Functions*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 2009
- [80] D. GRISSA, B. COMTE, E. PUJOS-GUILLOT, A. NAPOLI. *A Hybrid Knowledge Discovery Approach for Mining Predictive Biomarkers in Metabolomic Data*, in "ECML PKDD", Riva del garda, Italy, September 2016, pp. 572 - 587 [DOI : 10.1007/978-3-319-46128-1\_36], <https://hal.archives-ouvertes.fr/hal-01421011>
- [81] D. HERVÉ, J.-H. RAMAROSON, A. RANDRIANARISON, F. LE BER. *Comment les paysans du corridor forestier de Fianarantsoa (Madagascar) dessinent-ils leur territoire ? Des cartes individuelles pour confronter les points de vue*, in "Cybergeographie : Revue européenne de géographie / European journal of geography", 2014, 681 p. , <https://hal.archives-ouvertes.fr/hal-01057112>
- [82] O. HUDRY, B. MONJARDET. *Consensus Theories. An oriented survey*, in "Mathématiques et Sciences Humaines", 2010, vol. 190, n<sup>o</sup> 2, pp. 139–167
- [83] J. KALOFOLIAS, E. GALBRUN, P. MIETTINEN. *From Sets of Good Redescriptions to Good Sets of Redescriptions*, in "Proceedings of the 16th IEEE International Conference on Data Mining, ICDM'16", Barcelona, Spain, December 2016, <https://hal.archives-ouvertes.fr/hal-01399294>
- [84] M. KAYTOUE, V. CODOCEDO, A. BUZMAKOV, J. BAIXERIES, S. O. KUZNETSOV, A. NAPOLI. *Pattern Structures and Concept Lattices for Data Mining and Knowledge Processing*, in "Machine Learning and Knowledge Discovery in Databases", Porto, Portugal, A. BIFET, M. MAY, B. ZADROZNY, R. GALVALDA, D. PEDRESCHI, F. BONCHI, J. CARDOSO, M. SPILIOPOULOU (editors), Lecture Notes in Computer Science, Springer International Publishing, 2015, vol. 9286, pp. 227-231 [DOI : 10.1007/978-3-319-23461-8\_19], <https://hal.archives-ouvertes.fr/hal-01188637>
- [85] J.-F. MARI, F. LE BER, E.-G. LAZRAK, M. BENOÎT, C. ENG, A. THIBESSARD, P. LEBLOND. *Using Markov Models to Mine Temporal and Spatial Data*, in "New Fundamental Technologies in Data Mining", K. FUNATSU, K. HASEGAWA (editors), Intech, 2011, pp. 561–584, <http://hal.inria.fr/inria-00566801/en>
- [86] J.-P. METIVIER, A. LEPAILLEUR, A. BUZMAKOV, G. POEZEVARA, B. CRÉMILLEUX, S. O. KUZNETSOV, J. LE GOFF, A. NAPOLI, R. BUREAU, B. CUISSART. *Discovering structural alerts for mutagenicity using stable emerging molecular patterns*, in "Journal of Chemical Information and Modeling", 2015, vol. 55, n<sup>o</sup> 5, pp. 925–940 [DOI : 10.1021/CI500611v], <https://hal.archives-ouvertes.fr/hal-01186716>
- [87] G. PERSONENI, S. DAGET, C. BONNET, P. JONVEAUX, M.-D. DEVIGNES, M. SMAÏL-TABBONE, A. COULET. *Mining Linked Open Data: A Case Study with Genes Responsible for Intellectual Disability*, in "Data Integration in the Life Sciences - 10th International Conference, DILS 2014", Lisbon, Portugal, H. GALHARDAS, E. RAHM (editors), Lecture Notes in Computer Science, Springer, July 2014, vol. 8574, pp. 16 - 31 [DOI : 10.1007/978-3-319-08590-6\_2], <https://hal.inria.fr/hal-01095591>
- [88] F. PIROT, R. J. KANG. *Distance-coloring without one cycle length*, in "SIAM Journal on Discrete Mathematics", 2017, Submitted for publication, <https://hal.archives-ouvertes.fr/hal-01668633>
- [89] N. RAMAKRISHNAN, D. KUMAR, B. MISHRA, M. POTTS, R. F. HELM. *Turning CARTwheels: An Alternating Algorithm for Mining Redescriptions*, in "Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining", New York, NY, USA, KDD '04, ACM, 2004, pp. 266–275



- 
- [90] M. ROUANE-HACENE, M. HUCHARD, A. NAPOLI, P. VALTCHEV. *Relational Concept Analysis: Mining Concept Lattices From Multi-Relational Data*, in "Annals of Mathematics and Artificial Intelligence", January 2013, vol. 67, n<sup>o</sup> 1, pp. 81-108 [DOI : 10.1007/s10472-012-9329-3], <http://hal.inria.fr/lirmm-00816300>
- [91] L. SZATHMARY, P. VALTCHEV, A. NAPOLI, R. GODIN, A. BOC, V. MAKARENKOV. *A fast compound algorithm for mining generators, closed itemsets, and computing links between equivalence classes*, in "Annals of Mathematics and Artificial Intelligence", 2014, vol. 70, pp. 81 - 105 [DOI : 10.1007/s10472-013-9372-8], <https://hal.inria.fr/hal-01101140>
- [92] D. ZHUK. *A Proof of CSP Dichotomy Conjecture*, in "58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017", 2017, pp. 331–342, <https://doi.org/10.1109/FOCS.2017.38>
- [93] R. DE JOANNIS DE VERCLOS, J.-S. SERENI. *Equitable Colorings of  $K_4$ -minor-free Graphs*, in "Journal of Graph Algorithms and Applications", 2017, vol. 21, n<sup>o</sup> 6, pp. 1091–1105, <https://hal.archives-ouvertes.fr/hal-01483972>