Project-Team DRUID

Declarative & Reliable management of Uncertain, user-generated & Interlinked Data

Lannion - Rennes

Activity Report
2014

1 Team

DRUID is a newly created team at IRISA, starting October 2014. DRUID is supported by 5 active members from distinct IRISA sites, Rennes and Lannion. The team is completed by 4 associated researchers and several Ph.D students. PR means full professor, and MCF means "Maître de conférences", a tenured assistant professor.

Head of the team

David Gross-Amblard, Professor, ISTIC Rennes 1 - Rennes
Arnaud Martin, Professor, IUT Lannion - Lannion

Administrative assistant

Lydie Mabil - Rennes

Université Rennes 1 personnel

Tristan Allard, Assistant Professor, ISTIC Rennes 1 - Rennes
Tassadit Bouadi, Assistant Professor, IUT Lannion - Lannion
Zoltan Miklos, Assistant Professor, ESIR Rennes 1 - Rennes

Associate members

Jean-Christophe Dubois, Assistant Professor, IUT Lannion - Lannion
Mouloud Kharoune, Assistant Professor, IUT Lannion - Lannion
Yolande Le Gall, Assistant Professor, IUT Lannion - Lannion
Virginie Sans, Assistant Professor, ISTIC Rennes 1 - Rennes

PhD students

Dorra Attiaoui, Tunisian grant, since June 2013
Amira Essaid, Tunisian grant and ATER, since June 2012
Siwar Jendoubi, Tunisian MOBIDOC grant, since June 2014
Panagiotis Mavridis, MENRT/Rennes 1 grant, since October 2014
Kuang Zhou, Chineese grant, since October 2013

Master students

Salma Ben Dhaou, IHEC, Tunisia, March-June 2014
Imem Dlala, ISG, Tunisia, March-June 2014
Adam Kammoun, master MOMO, École centrale de Lille, June-August 2014
Wiem Maalel, IHEC, Tunisia, March-June 2014
Maxime Audinot, ENS de Rennes, Oct 2014- June 2015

2 Overall Objectives

2.1 Overview

Our perception of digital information has completely shifted in recent years, in several ways. First, data are no longer isolated, but are now part of distributed, **interlinked networks**. Such networks include web documents (URLs and URIs), communities in a social graph (*e.g.* FOAF), conceptual networks on the Semantic Web or the continuously growing network of Linked Open Data (RDF). Second, data are now dynamic. Obtaining an up-to-date piece of information is as simple as a Web service call or a syndication (as is RSS or Atom). A large diversity of such dynamic data sources is available, including corporate Web services, wireless sensors in the environment, humans in the participative Web, or workers in crowdsourcing platforms[1]. Hence, what becomes important is the **data source** itself.

The openness and liveliness of such interlinked data networks is a great opportunity. Business Intelligence applications no longer restrict their attention to the companies own data sets or sales records, but try to incorporate data collected from the Web (such as opinions from social networks or Web forums). In this way they can extract useful information about their customers and the reception of their products and services. Another domain is the integration of personal information from multiple devices. The same opportunities arise also in the context of non-profit organizations or societal challenges: there is a lot of information available on health problems (Web forums on health, body area networks), environmental issues (environmental sensors) or in administrative domains (smart cities, Open Data initiatives). A new key issue is also to benefit from the growing "digital presence", that allows **interaction** with users at virtually any moment through mobile phone applications such as Twitter. Feedback loops between users and data managers can now be devised[2].

But the diversity and the dynamic of data sources raise several challenges. One can legitimately question if a data source is reliable or malevolent or if two data sources are independent These problems are strengthened by the mutual links between data items or data sources. Hence fact provenance and sources independence are prominent data annotations that shall be taken into account. For user-generated or crowdsourced content, knowing the skills or the social relationships between participants allows for a better understanding of the produced raw data. This calls for a powerful **qualification mechanisms** that would integrate these annotations and help data managers in understanding their data and selecting their sources. Furthermore, even if interaction with participants is technically possible, the **orchestration** of complex data acquisition tasks from a mass still remains a black art.

[1] `https://www.mturk.com`

[2] See for example participative journalism platforms, `https://witness.theguardian.com/moreabout`

The objective of the DRUID team is to provide models and algorithms for the annotation and management of interlinked data and sources at a large scale. We consider three main goals:

1. To propose well-founded models for interlinked data and, more importantly, interlinked data sources (for example, profiling users in a social network, orchestrating users and tasks in a crowdsourcing platform),
2. To develop theories for the qualification of such data and sources in terms of reliability, certainty, provenance, influence, economical value, trust, etc.
3. To implement systems that are proof-of-concepts of these models and theories. In particular we would like to demonstrate that these systems can overcome specific key problems in real-world applications, such as scalable data qualification and data adaptation to the final users.

More concretely, we would like to address the following challenges:

- to develop integrated and scalable analysis tools for participants in social networks, that encompass the semantics of communications between users, computes user influence or user independence for example.
- to extend existing crowdsourcing platforms with fine user profiles, team building or complex task management abilities, with application for e-science or e-government (smart cities).
- to develop reliability assessment techniques for large sensor networks (uncertainty), heterogeneous data sources or Linked Open Data (quality), or microblog conversations (misinformation).
- to adapt data to its use (data visualization, accessibility of information).

2.2 Key issue 1: Well-Founded Models for Interlinked Data and Sources

The Data Management field aims to build pertinent models for information, expressive query languages at a high level of abstraction for computer engineers or basic users, and efficient evaluation methods. The field was successful with a wide acceptance of solutions at the industrial level (banking, electronic commerce, document management, ticketing, etc.).

But classical approaches are not directly suited for nowadays applications. On the one hand, with the spreading of graph data models such as RDF, data are no longer relational (structured into tables) not even tree-based (XML), but graph-based (reminiscent of the semi-structured data model). Furthermore, these graphs are no longer centralized but interlinked through the network. The success of NoSQL graph database for social networks is an illustration[3]. New

[3]Indeed, Facebook is using a NoSQL, graph-oriented database for its core data, Neo4j, `http://www.zdnet.com/facebook-neo4j-7000009866/`.

models are then required to express queries on such graphs in a well-founded manner[BLLW12].

On the other hand, the very structure of data sources should now be investigated. A first example is sensors (in a broad sense, from specialized sensors to smart phones sensors or personal health monitors). Such devices support severe constraints on their connectivity and their ability to provide data. They are mobile and energy-restricted. A more recent and striking example is to consider also humans as data sources. These sources are related to each others (social network) and they produce data with a rich semantics (see for example post contents in a forum). Hence being able to query such data sources with a clean language, while taking into account their relationships and reasoning about their semantics would greatly impact practical applications. A typical relevant query is to find the central user of a social network (structural query), restricted to the subset of users talking about action movies (semantic query).

Finally, it is now possible to interact with data sources, as in large participative, crowd-based systems that gather information (e.g. participative science) or resolve tasks (Human based computing)[4]. Having a clean framework to organize such interactions would also benefit to these applications.

Our first objective is to provide well-funded models for interlinked data sources (social graphs, microblogs, sensors) and complex workflows of human tasks.

The sub-goals of this scientific axis are listed below, ordered by their priority (short terms are already started, mid terms cover a classical Ph.D duration, and long terms target prospective issues).

[short term] Adding semantics Integration of the semantics of the information flow within networks, as many properties in social graphs are not only syntactical or linked-based. Relevant tools are taxonomies, ontologies but also sentiment analysis and controversy analysis.

[mid term] Querying social graph data Few query languages are available to reason about the structure of a (huge) social graph. For example, selecting nodes that are part of distinct communities is hardly expressible without relying on a ad hoc program.

[mid term] Modeling users Modeling a user as a data source with it's profile, opinion, social network, motivations, personal goals, personal strategies, location and available time (e.g. for real-time crowdsourcing applications).

[mid term] Mixing queries about data and sources As a new query type, we have for example "give me the average ranking of this movie from users who are absolutely not connected with me, but have shown a long habit in watching and ranking movies". The first

[4] As an example, the micro-tasking platform AMT (Amazon Mechanical Turk) had in 2013 more than 300 000 users available at any time to resolve a task, `https://requester.mturk.com/tour`.

[BLLW12] P. Barceló, L. Libkin, A. W. Lin, P. T. Wood, "Expressive Languages for Path Queries over Graph-Structured Data", *ACM Trans. Database Syst. 37*, 4, 2012, p. 31.

part of the query is related to the graph underlying the social network, the second part to the properties of the data source.

[mid term] Crowdsourcing complex tasks Most existing crowdsourcing platforms are not generic and the deployment of complex tasks supposes a huge development cost[ABMK11]. Recently, declarative crowdsourcing systems, mixing database approaches with user interaction have emerged[FKK+11,PPP+13]. These first efforts reduce the development cost of simple data curation tasks. We propose to model complex tasks using declarative workflow models[HMT12, AV13] in order to reason about the correctness of complex, human-based computation processes.

[long term] Integrated interaction model Our goal is to break the discrepancy between the content on the one side, and the users generating this content on the other side. We would like to achieve a generic model where one can reason both on the graph structure of data (paths, clusters), the social structure of the users (skills, friendship, team structure, centrality), and on the social workflows that connect them.

2.3 Key issue 2: Interlinked Data and Sources Qualification

By qualification we mean any type of qualitative and quantitative indicators on data and sources of data. We can evoke as examples data uncertainty, imprecision, economical and strategic value, privacy, accessibility, data provenance and also reliability, expertise, independence, conflict of sources of data.

Our second objective is to provide qualification mechanisms for interlinked data and sources, taking into account their mutual interactions and the available information, even if this information is unsure and imprecise.

Data and sources qualification is a great social need for the contemporary Web. Users shall be enlighten by the provenance, quality and accessibility of their information sources. We mention three important directions:

[ABMK11] S. Ahmad, A. Battle, Z. Malkani, S. Kamvar, "The jabberwocky programming environment for structured social computing", *in : Proceedings of the 24th annual ACM symposium on User interface software and technology, UIST '11*, p. 53–64, 2011.

[FKK+11] M. J. Franklin, D. Kossmann, T. Kraska, S. Ramesh, R. Xin, "CrowdDB: answering queries with crowdsourcing", *in : Proceedings of the 2011 ACM SIGMOD International Conference on Management of data, SIGMOD '11*, p. 61–72, 2011.

[PPP+13] H. Park, R. Pang, A. Parameswaran, H. Garcia-Molina, N. Polyzotis, J. Widom, "An overview of the deco system: data model and query language; query processing and optimization", *SIGMOD Rec. 41*, 4, January 2013, p. 22–27.

[HMT12] R. Hull, J. Mendling, S. Tai, "Business process management", *Inf. Syst. 37*, 6, 2012, p. 517.

[AV13] S. Abiteboul, V. Vianu, "Models for Data-Centric Workflows", *in : In Search of Elegance in the Theory and Practice of Computation*, V. Tannen, L. Wong, L. Libkin, W. Fan, W.-C. Tan, M. P. Fourman (editors), *Lecture Notes in Computer Science, 8000*, Springer, p. 1–12, 2013.

[short term] Assessing social network reliability Using social networks can exhibit some risks users are not aware of. Indeed, erroneous information can be send deliberately or involuntarily (by lack of scrutiny, of from hacked accounts). Information in social networks can easily be distorted and amplified according to relationships between relaying users. Even if information is corroborated by several contacts, its source can be unique and erroneous. There is a crucial need for tools to evaluate social networks reliability and weaknesses, in order to take valuable decisions. The relationships between users has to be taken into account, along with the quality and amount of independent data sources. A great challenge is to identify relevant information in a mass of data exchange and to predict real events from purely electronic activities.

[short term] Interlinked data integration Schema integration is a long standing problem in information systems. This problematic is amplified by the relationship between data sets. We propose the notion of schema networks to model this situation, and techniques to provide schema mappings as an equilibrium within this network.

[mid term] Interlinked data fusion More and more information systems gather information from network-organized sensors. The vanishing price of such sensors allows their use in everyday life and tools. They are also used in dedicated applications such as military watching, aerial, terrestrial or oceanic missions (sensor swarms). In such complex networks, sensors do not play an equal role: some may be dedicated to observation, others to positioning, or communication. The flow of information can be altered because of sensor deficiency or the structure of the network itself. In such scenarios, data fusion must be preceded by a correct qualification of data and sources. Such qualification also leverages reliability [68], independence [36] and conflict measurement [77]. While mature approaches for data fusion already exist, the network structure is rarely studied. One of our goal is then to propose efficient methods that incorporate this structure.

[mid term] Crowdsourcing quality optimization In crowdsourcing applications, data quality is a central concern. Many techniques have been envision to enhance this quality, by, for example, performing majority voting between redundant tasks. Since our first goal is to go beyond simple query-answer tasks, that is to encompass their composition, adapted quality enhancement mechanisms have to be designed accordingly. As an example, we will consider models of user motivation to select which part of a complex task is more suited to a given user. Other directions concern designing incentive to motivate users, and taking the reputation of users into account. The mixing of users skills and the knowledge of their social network is also a natural direction. We also would like to allow the user to provide a self estimation of his input accuracy. This feedback would aim to estimate the imprecision and the level of certainty of his answers, in order to optimize the decision process.

[long term] Integrated annotation model Our vision is a transparent data model that accepts and triggers any kind of source contribution in a non-blocking way, while offering a

coherent, qualified view of the data set at any time and from any user perspective. Our goal is to keep the model simple in order to promote its adoption by industry.

2.4 Key issue 3: Data & Sources Management: Large Scale, High Rate, Ease of Use

The two previous goals we just introduced will provide models for interlinked data and sources, along with rich qualification mechanisms.

> Our third objective is to provide fully integrated systems that allow for the manipulation of interlinked data and interlinked sources, along with rich qualification indicators, while being efficient and adapted to users.

Two ingredients are needed for the success of such systems: scalability and ease of use. We discuss these two issues in the sequel.

Optimization and scalability From the efficiency point of view, many problems arise. Qualification indicators may appear as meta-data in the core of a data management system, with the difficulty of their storage. But the main challenge is the algorithmic complexity of their computation. In order to deal with high volume or rate of data, the proposed algorithms should be designed for scalability. Several directions are envisioned:

- **[short term]** Using distributed computation paradigms such as MapReduce, and iterative computations such as PageRank and variants.
- **[mid term]** Relying on controlled approximation algorithms: only an estimate of the correct data qualification will be obtained, but with a small error (say 5%), with a small failure probability (say one chance over 1 billion), but with a rapid computation time.
- **[mid term]** Filtering relevant information with coarse-grain qualification estimate, in order to reduce the amount of data (for example, to reduce the number of focal elements for belief functions approaches).
- **[long term]** Using streaming algorithms, where computations are done on the fly, also with a controlled error.

Security The crowdsourcing litterature is growing at a fast pace. However, security and privacy have been ignored until now in crowdsourcing contexts despite their importance. Indeed, crowdsourcing processes involve (1) exporting data and workflows to the crowd, and/or (2) collecting data and results from the crowd. We plan to study privacy and security issues that arise in these contexts.

- **[mid term]** Exporting Data and Workflows to the Crowd. Most works have focused on exploiting the crowd by delegating specific tasks to workers. Usually, the task specification involves sending data to workers together with the task specification. For example,

matching pairs of similar items [WLK+13] requires sending the items to workers, or planning a schedule [KLMN13] requires sending the objects or actions to be planned and their constraints. However, sensitive data cannot be sent in the clear to participants. In a traditional context, where only machines participate to the computation, strong cryptographic protocols can let machines participate without accessing non-encrypted sensitive data. In a crowdsourcing context, such cryptographic protocols cannot be used anymore because they would simply preclude humans to participate. Data has to be disclosed to humans. How can we disclose sensitive data to humans in crowdsourcing processes while still guaranteeing its privacy ? Similarly, involving the crowd in a complex workflows implies disclosing each task to the human workers to which it is assigned. How can we guarantee the confidentiality of workflows, *e.g.*, for intellectual property reasons, while still allowing the participation of workers ?

- **[mid term]** Collecting Data and Results from the Crowd. The crowd can be viewed as a specific database that can be, *e.g.*, queried [PW14], indexed [ADM+14], or mined [AAM14,ADM+14]. However, data that is collected by such algorithms is individual data and may be consequently identifying or sensitive. How can we guarantee the privacy of individual data in crowdsourcing data-oriented processes ? Protecting such obviously-sensitive data is however not sufficient. Indeed, covert channels may exist and lead to the disclosure of sensitive data. For example, a worker may answer intriguingly fastly to questions related to a given disease or to a given place. This may reveal a surprising strong connection between the worker and this disease or place. How can we protect workers from covert channel attacks in crowdsourcing processes ?

Data presentation Beside efficiency, there is a tremendous need from end-user for an adapted presentation of information. The amount of available data along with the rich annotations we will add are certainly overwhelming for any user. We will consider in this axis also

[WLK+13] J. Wang, G. Li, T. Kraska, M. J. Franklin, J. Feng, "Leveraging Transitive Relations for Crowdsourced Joins", *in : Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data, SIGMOD '13*, ACM, p. 229–240, New York, NY, USA, 2013, http://doi.acm.org/10.1145/2463676.2465280.

[KLMN13] H. Kaplan, I. Lotosh, T. Milo, S. Novgorodov, "Answering Planning Queries with the Crowd", *Proc. VLDB Endow. 6*, 9, July 2013, p. 697–708, http://dx.doi.org/10.14778/2536360.2536369.

[PW14] H. Park, J. Widom, "CrowdFill: Collecting Structured Data from the Crowd", *in : Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, SIGMOD '14*, ACM, p. 577–588, New York, NY, USA, 2014, http://doi.acm.org/10.1145/2588555.2610503.

[ADM+14] Y. Amsterdamer, S. B. Davidson, T. Milo, S. Novgorodov, A. Somech, "OASSIS: Query Driven Crowd Mining", *in : Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, SIGMOD '14*, ACM, p. 589–600, New York, NY, USA, 2014, http://doi.acm.org/10.1145/2588555.2610514.

[AAM14] A. Amarilli, Y. Amsterdamer, T. Milo, "On the Complexity of Mining Itemsets from the Crowd Using Taxonomies", *in : Proc. 17th International Conference on Database Theory (ICDT), Athens, Greece, March 24-28, 2014.*, p. 15–25, 2014, http://dx.doi.org/10.5441/002/icdt.2014.06.

- **[short term]** Data adaptation methods, that filter information according to the user's needs and capabilities: on a static or mobile environment, on-line or off-line, with or without real-time needs, disabled persons, seniors, and so on.
- **[long term]** Data visualization methods, that present a visual and navigational summary in qualified data.

3 Scientific Foundations

3.1 Data management

To achieve our goals we will rely on techniques of two scientific domains: data management and data qualification. For data management we will naturally elaborate on classical techniques: finite model theory, complexity theory, approximation algorithms, declarative or algebraic languages, execution plans, costs models, indexing. We intend to explore new models such as schema networks for data integration, user modeling for crowdsourcing application[RLT+13], and game theory for the study of strategic aspects in crowdsourcing, data pricing[LLMS13] and data publication[JP13]. For the modeling of complex tasks in crowdsourcing, we envision to extend declarative approaches for business processes[DM12], such as the collaborative business artifact model[AV13].

3.2 Data qualification

For data qualification, our first focus will be on uncertainty. Many frameworks are available, but all are based on the theories of uncertainty that are able to model imperfect data. Two main aspects of imperfection are classically distinguished: uncertainty and imprecision[Sme97].

In particular, the theory of belief functions[Dem67,Sha76] (also commonly referred to as evidence theory or Dempster-Shafer theory) allows to take simultaneously into account both

[RLT+13] S. B. Roy, I. Lykourentzou, S. Thirumuruganathan, S. Amer-Yahia, G. Das, "Crowds, not Drones: Modeling Human Factors in Interactive Crowdsourcing", *in: DBCrowd*, p. 39–42, 2013.

[LLMS13] C. Li, D. Y. Li, G. Miklau, D. Suciu, "A theory of pricing private data", *in: ICDT*, W.-C. Tan, G. Guerrini, B. Catania, A. Gounaris (editors), ACM, p. 33–44, 2013.

[JP13] S. Jain, D. C. Parkes, "A game-theoretic analysis of the ESP game", *ACM Trans. Econ. Comput. 1*, 1, January 2013, p. 3:1–3:35.

[DM12] D. Deutch, T. Milo, *Business Processes: A Database Perspective*, *Synthesis Lectures on Data Management*, Morgan & Claypool Publishers, 2012.

[AV13] S. Abiteboul, V. Vianu, "Models for Data-Centric Workflows", *in: In Search of Elegance in the Theory and Practice of Computation*, V. Tannen, L. Wong, L. Libkin, W. Fan, W.-C. Tan, M. P. Fourman (editors), *Lecture Notes in Computer Science, 8000*, Springer, p. 1–12, 2013.

[Sme97] P. Smets, "Imperfect information: Imprecision - Uncertainty", *in: Uncertainty Management in Information Systems*, A. Motro and P. Smets (editors), Kluwer Academic Publishers, 1997, p. 225–254.

[Dem67] A. P. Dempster, "Upper and Lower probabilities induced by a multivalued mapping", *Annals of Mathematical Statistics 38*, 1967, p. 325–339.

[Sha76] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.

uncertainty and imprecision. This theory is one of the most popular one among the quantitative approaches because it can be seen as a generalization of both classical probabilities and possibilities theories[DPS96]. Its strength lies in (1) its richer representation of uncertainty and imprecision compared to probability theory and (2) its higher ability to combine pieces of information. In particular, a crucial task in information fusion is the management of conflict between different (partially or totally) disagreeing sources. The origins of conflict can come from the source reliability, disinformation, truthfulness, etc. For interlinked data such as posts flowing through a social network, we also have to consider the quality of the data, especially its uncertainty and imprecision. We can also see each node of the network as a node of information fusion. The framework of belief functions is therefore well adapted.

Two main difficulties are to be underlined: first, to find a correct definition of data quality (in order to encompass reliability, truthfulness, disinformation) combined with source quality (reliable experts, liars, collusion,

trollers) [PDD12,Sme93] ;

second to possibly resolve the problem of scalability associated with belief function approaches (still, the corresponding complexity is lower than for other approaches such as imprecise probability theories or random set theories).

4 Application Domains

4.1 Generic Crowdsourcing Plateform, Data annotatation and sensing

Participants: Tristan Allard, Tassadit Bouadi, David Gross-Amblard, Jean-Christophe Dubois, Yolande Le Gall, Panagiotis Mavridis, Arnaud Martin, Zoltan Miklos, Virginie Sans.

The models we develop for crowdsourcing provide a strong basis for the development of a generic crowdsourcing platform that can be adapted to various uses. We envision for now to target two kinds of applications: data annotations and data sensing. In data annotation, the crowd is asked to tag a set of resources (images, videos, locations, etc.) using a free or controlled vocabulary. In data sensing, the crowd is consulted to obtain any kind of data, say for example environmental measurements (temperature, weather, water quality, etc.) or personal information (location, speed, feelings about a place, etc.). The role of the crowdsourcing platform is to orchestrate crowd interactions and to protect (sanitize) the collection of private information.

4.2 Social Network Analysis for Humanities and Marketing

Participants: Dorra Attiaoui, Tassadit Bouadi, David Gross-Amblard, Siwar Jendoubi,

[DPS96] D. Dubois, H. Prade, P. Smets, "Representing Partial Ignorance", *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 26*, 3, 1996, p. 361–377.

[PDD12] F. Pichon, D. Dubois, T. Denoeux, "Relevance and truthfulness in information correction and fusion", *International Journal of Approximate Reasoning 53*, 2, 2012, p. 159 – 175, Theory of Belief Functions (BELIEF 2010).

[Sme93] P. Smets, "Belief Functions: the Disjunctive Rule of Combination and the Generalized Bayesian Theorem", *International Journal of Approximate Reasoning 9*, 1993, p. 1–35.

Mouloud Kharoune, Arnaud Martin, Zoltan Miklos, Virginie Sans, Kuang Zhou.

We consider social network analysis by the way of heterogeneous social networks where we integrate the models of imperfect linked data. Therefore, we consider several problems for social network analysis such as the community detection, experts and trolls identification and message's propagation for example for viral marketing applications. Hence, we consider different kinds of social network such as Twiter and dblp. We also test our models on generated networks.

5 Software

5.1 ibelief

Participants: Kuang Zhou, Arnaud Martin [contact point].

The R package ibelief aims to provide some basic functions to implement the theory of belief functions, and it has included many features such as:

1. Fast Mobius Transformation to convert any of the belief measures (such as basic belief assignment, credibility, plausibility and so on) to another type;
2. Some commonly used combination rules including DS rule, Smets' rule, Yager's rule, DP rule, PCR6 and so on;
3. Some rules for making decisions;
4. The discounting rules in the theory of belief functions;
5. Different ways to generate random masses.

The stable version of package ibelief could be found on CRAN (common R code repository).

5.2 Crowd

Participants: Tristan Allard, Tassadit Bouadi, David Gross-Amblard [contact point], Panagiotis Mavridis, Zoltan Miklos, Virginie Sans.

We have realized a crowdsourcing platform[5] that can execute complex tasks that one can obtain as a composition of simple human intelligence tasks. The platform uses a skill model to affect the tasks. We have presented the software at the BDA'2014 conference[38].

6 New Results

6.1 Belief Social Network

Participants: Salma Ben Dhaou, Siwar Jendoubi, Mouloud Kharoune, Arnaud Martin.

[5] http://craft.irisa.fr

Most existing research works focus on the analysis of homogeneous social networks, *i.e.* we have a single type of node and link in the network. However, in the real world, social networks offer several types of nodes and links. Hence, with a view to preserve as much information as possible, it is important to consider social networks as heterogeneous and uncertain.

Belief Social Network structure

Several works have focused on the representation of social networks with graphs. A classical graph is represented by $G = \{V; E\}$ with: V a set of type's nodes and E a set of type's edges. This representation does not take into account the uncertainty of the nodes and edges.

In fact, graphical models combine the graph theory with any theory dealing with uncertainty like probability[PGPB14,KBGG14] or possibility or theory of belief functions to provide a general framework for an intuitive and a clear graphical representation of real-world problems[LbYS12]. The propagation of messages in networks has been modeled using the theory of belief functions combined with other theories such as hidden Markov chains[Ram09].

In this context, we introduce in [39]: the belief social network which has the role of representing a social network using the theory of belief functions. Indeed, we will associate to each node, link and message an a priori mass and observe the interaction in the network to determine the mass of the message obtained in a well-defined node. To do this, we consider an evidential graph $G = \{V^b; E^b\}$ with: V^b a set of nodes and E^b a set of edges. We attribute to every node i of V^b a mass $m_i^{\Omega_N}$ defined on the frame of discernment Ω_N of the nodes. Moreover, we attribute also to every edge (i, j) of E^b a mass $m_{ij}^{\Omega_L}$ defined on the frame of discernment Ω_L of the edges. Therefore, we have:

$$V^b = \{V_i, m_i^{\Omega_N}\} \tag{1}$$

and

$$E^b = \{(V_i^b, V_j^b), m_{ij}^{\Omega_L}\} \tag{2}$$

In social network, we can have for example the frame of the nodes given by the classes, *e.g.* Person, Company, Association and Place. The frame of discernment of the edges can be for example Friendly, Professional or Family. Moreover we note: $\Omega_N = \{\omega_{n_1}, \ldots, \omega_{n_N}\}$ and $\Omega_L = \{\omega_{l_1}, \ldots, \omega_{n_L}\}$.

In social network, many messages can transit in the network. They can be categorized as commercial, personal, and so on.The class of the message is also full of uncertainty. Therefore to each message, we add a mass function in the considered frame of discernment $\Omega_{\text{Mess}} = \{\omega_{M_1}, \ldots, \omega_{M_k}\}$.

[PGPB14] P. Parchas, F. Gullo, D. Papadias, F. Bonchi, "The pursuit of a good possibe world: Extracting representative instances of uncertain graphs", *in: SIGMOD'14 , Snowbird*, ACM press, 2014.

[KBGG14] A. Khan, F. Bonchi, A. Gionis, F. Gullo, "Fast Reliability Search in uncertain graphs", *in: International Conference on Extending Database Technology (EDBT)*, 2014.

[LbYS12] W. Laamari, B. ben Yaghlane, C. Simon, "Dynamic Directed Evidential Networks with conditional belief functions:Application to system reliability", *IPMU*, 1, 2012, p. 481–490.

[Ram09] E. Ramasso, "Contribution of belief functions to hidden markov models with an application to fault diagnosis", *IEEE*, 1, 2009, p. 1–6.

Propagation analysis

The goal of our work published in [65] is to classify the social message based on its spreading in the network and the theory of belief functions. The proposed classifier interprets the spread of messages on the network, crossed paths and types of links. We tested our classifier on a real word network that we collected from Twitter, and our experiments show the performance of our belief classifier.

6.2 Social Network Analysis

Participants: Dorra Attiaoui, Arnaud Martin, Kuang Zhou.

The web plays an important role in people's social lives since the emergence of Web 2.0. It facilitates the interaction between users, gives them the possibility to freely interact, share and collaborate through social networks, online communities forums, blogs, wikis and other online collaborative media.

Experts and trolls detection

For the last few years, Question Answering Communities (Q&A C) have changed the way people seek for information. Easy access, quick responses, topics well organized, these web sites became more and more popular. Among them we can cite Yahoo!Answers, Stackoverflow, Quora, Wikianswers, etc.

These web sites are based on a collaborative method: A person asks a question seeking for the best answers he could have, and anyone can respond. This kind of system gives every user the opportunity to contribute to a discussion, sharing her (his) knowledge and/or expertise.

However, an other side of the web is negatively taken such as posting inflammatory messages. Thus, when dealing with the online communities forums, the managers seek to always enhance the performance of such platforms. In fact, to keep the serenity and prohibit the disturbance of the normal atmosphere, managers always try to novice users against these malicious persons by posting such message (DO NOT FEED TROLLS). But, this kind of warning is not enough to reduce this phenomenon. In this context we propose a new approach for detecting malicious people also called 'Trolls' in order to allow community managers to take their ability to post online. To be more realistic, we propose in [81] to define within an uncertain framework. Based on the assumption consisting on the trolls' integration in the successful discussion threads, we try to detect the presence of such malicious users. Indeed, this method is based on a conflict measure of the belief function theory applied between the different messages of the thread. In order to show the feasibility and the result of our approach, we test it in different simulated data.

Community detection

We have proposed in [21] a new prototype-based clustering method, called Median Evidential C-Means (MECM), which is an extension of median c-means and median fuzzy c-means on the theoretical framework of belief functions is proposed. The median variant relaxes the restriction of a metric space embedding for the objects but constrains the prototypes to be in

the original data set. Due to these properties, MECM could be applied to graph clustering problems. A community detection scheme for social networks based on MECM is investigated and the obtained credal partitions of graphs, which are more refined than crisp and fuzzy ones, enable us to have a better understanding of the graph structures. An initial prototype-selection scheme based on evidential semi-centrality is presented to avoid local premature convergence and an evidential modularity function is defined to choose the optimal number of communities. Finally, experiments in synthetic and real data sets illustrate the performance of MECM and show its difference to other methods.

6.3 Uncertainty in Ontology Matching: a Decision Rule-based Approach

Participants: Amira Essaid, Arnaud Martin.

Considering the high heterogeneity of the ontologies published on the web, ontology matching is a crucial issue whose aim is to establish links between an entity of a source ontology and entities from a target ontology. Perfectible similarity measures, considered as sources of information, are combined to establish these links. The theory of belief functions is a powerful mathematical tool for combining such uncertain information. In [49, 48], we introduce a decision process based on a distance measure to identify the best possible matching entities for a given source entity.

Ontology matching is the process for finding for each entity of an ontology source O_1 its corresponding entity in an ontology target O_2. This process can focus on finding simple mappings (1:1) or complex mappings (1:n or n:1). The first consists in matching only one entity of O_1 with only one entity of O_2 whereas the second consists in finding either for one entity of O_1 its multiple correspondences of entities in O_2 or matching multiple entities of O_1 with only one entity of O_2. We are interested in this paper in finding simple mappings as well as the complex one of the form (1:n). The matching process is performed through the application of matching techniques which are mainly based on the use of similarity measures. Since no similarity measure applied individually is able to give a perfect alignment, we suggest that with the exploitation of the complementarity of different similarity measures, a better alignment would be achieved. Combining these similarity measures can lead to a conflict between the different results obtained which should be modeled and resolved.

We suggest the theory of belief functions as a tool for modeling the ontology matching and especially for combining the results of the different similarity measures. Due to the fact that we are working on an uncertain aspect and we are interested in finding complex matching which can be viewed as finding composite hypotheses formed from entities of two ontologies, we suggest to apply our proposed decision rule on the combined information and to choose for each entity of the ontology source, the entities of the target ontology with the lowest distance.

6.4 A belief function-based accessibility indicator to improve web browsing for disabled people

Participants: Jean-Christophe Dubois, Yolande Le Gall, Arnaud Martin.

The Web constitutes today an essential source of information and communication. While

users have a growing interest in terms of social, cultural and economic value, and in spite of legislations and recommendations of the W3C community for making websites more accessible, its accessibility remains hardly efficient for some disabled or ageing users. Actually, making websites accessible and usable by disabled people is a challenge[20110] that society needs to overcome[JMI+04].

To measure the accessibility of a webpage, several accessibility metrics have been developed[MG11]. Evaluations are based on the failure to comply with the recommendations of standards, using automatic evaluation tools. They often give a final value, continuous or discrete, to represent content accessibility. However, the fact remains that tests on accessibility criteria are far from being trivial[G.04]. Evaluation reports of automatic assessors contain errors considered as certain, but also warnings or potential problems which are uncertain. Moreover there are differences between assessor evaluations, even for errors considered as certain.

We propose in [98, 45] a study to provide an accessibility measure of webpages, in order to draw disabled users to the pages that have been designed to be accessible to them. This work provides a new measure of accessibility and an information fusion framework to fuse information coming from the reports of automatic assessors allowing search engines to re-rank their results according to an accessibility level, as some users would like[IMYK04]. This accessibility indicator considers several categories of deficiencies. Our approach is based on the theory of belief functions, using data which are supplied by reports produced by automatic web content assessors that test the validity of criteria defined by the WCAG 2.0 guidelines proposed by the World Wide Web Consortium (W3C) organization. These tools detect errors with gradual degrees of certainty and their results do not always converge. For these reasons, to fuse information coming from the reports, we choose to use an information fusion framework which can take into account the uncertainty and imprecision of information as well as divergences between sources. Our accessibility indicator covers four categories of deficiencies. To validate the theoretical approach in this context, we propose an evaluation completed on a corpus of 100 most visited French news websites, and 2 evaluation tools. The results obtained illustrate the interest of our accessibility indicator.

6.5 Schema matching networks and crowdsourcing

Participants: Zoltan Miklos.

In a collaboration with researchers from EPFL we have improved our results on schema matching networks, in particular on the use of crowdsouring techniques. We have published

[20110] E. D. S. 2010-20, "A Renewed Commitment to a Barrier-Free Europe", *Universal Access in the Information Society COM 636 final*, 2010.

[JMI+04] A. J., A. M., F. I., G. N., T. J., "The use of guidelines to automatically verify web accessibility", *Universal Access in the Information Society 3*, 1, 2004, p. 71–79.

[MG11] V. M., B. G., "Automatic web accessibility metrics: where we are and where we can go", *Interacting with Computers 23*, 2, 2011, p. 137–155.

[G.04] B. G., "Comparing accessibility evaluation tools: a method for tool effectiveness", *Universal Access in the Information Society 3*, 3-4, 2004, p. 252–263.

[IMYK04] Y. S. Ivory M. Y., G. K., "Search result exploration: a preliminary study of blind and sighted users' decision making and performance", *CHI Extended Abstracts*, 2004, p. 1453–1456.

an extended version of our DASFAA'2013 paper in the EAI Endorsed Transactions on Collaborative Computing[16]. The schema matching network model can capture situations where one would like to establish semantic interoperability among databases, connected in a network. We have conceived a system that enables that a large group of non-experts can create or improve attribute correspondences between the attributes of the involved schemas. By exploiting network-level constraints we can improve the (expected) quality of the results and reduce the necessary human efforts.

6.6 Securing outsourced data

Participants: Tristan Allard, David Gross-Amblard.

Within the POSEIDON project (see below), we investigate security issues related to data outsourcing (in e.g. a cloud). In [5] we propose a model where data fragmentation plus encryption are used to shatter data on different independent cloud services in order to protect sensitive data against malicious join operations. In [26], we deal with the general problem of security policies enforcement for sensitive data. We provide a logical framework for the definition of security constraints (say anonymity, precision, intellectual property protection). This formalism allows for the detection of incompatible constraints.

6.7 Hierarchical Skyline queries for multidimensional decision support

Participants: Tassadit Bouadi.

Conventional skyline queries retrieve the skyline points in a context of dimensions with a single hierarchical level. However, in some applications with multidimensional and hierarchical data structure (e.g. data warehouses), skyline points may be associated with dimensions having multiple hierarchical levels. Thus, we have proposed an efficient approach reproducing the effect of the OLAP operators "drill-down" and "roll-up" on the computation of skyline queries. It allows the user to navigate along the dimensions hierarchies (i.e. specialize / generalize) while ensuring an online calculation of the associated skyline. The method is described in [29, 31] and Tassadit Bouadi's thesis [3]. An extended version of this paper is currently under submission to the "Transactions on Large Scale Data and Knowledge Centered Systems" (TLDKS) Journal.

7 Contracts and Grants with Industry

7.1 DGA

- Data & source qualification for sensor networks: collaboration with the French National Defense Agency (Direction générale de l'armement – DGA[6] Cassidian[7]).

[6] http://www.defense.gouv.fr/dga
[7] http://www.cassidian.com/fr

8 Other Grants and Activities

8.1 International Collaborations

- Regular collaboration with LSIR/EPFL (Switzerland), LARODEC (Tunisia) and Northwestern Polytechnical University (Xi'an, China).
- Our team is an associated member of the European Network of Excellence on Large-scale Data Management PlanetData[8] (completed in September 2014).

8.2 National Collaborations

- We are part of the POSEIDON project on outsourced data security funded by the LABEX COMINLABS[9].
- We have obtained a grant from the University of Rennes 1, "Defis scientifique emergents". This enabled us to organize a workshop (2-4 July 2014), where we have invited a number of researchers (Telecom ParisTech, INRIA Nancy, INRIA Lille) in order to build an ANR proposal.
- We are currently involved in the CNRS Big Data (MASTODONS) ARESOS[10] project on semantic social networks analysis, particularly with the database team at LIP6.
- We have regular informal collaborations with the following teams: Vertigo/CEDRIC/Cnam-Paris, Hadas/LIG-Grenoble, DBWeb/Telecom Paristech-Paris, DAHU/ENS-Cahan, OAK-LRI/Orsay, ONERA, LABSTICC-Telecom Bretagne.

9 Dissemination

9.1 Scientific Responsabilities

Phd defense in DRUID in 2014

- Mouna Chebbah, co-directed by Arnaud Martin and Boutheina Ben Yaghlane defences her Phd entitled "Source independence in the theory of belief functions" June, 25, 2014 behind the jury members: W. Liu, E. Lefèvre, Z. Elouedi, L. Liétard

Jury of HDR defense in 2014

- D. Gross-Amblard:
  - Nicolas Anciaux (Université de Versailles-Saint Quentin & Inria, 2014)(External reviewer)

[8] http://www.planet-data.eu/
[9] http://www.poseidon.cominlabs.ueb.eu/
[10] https://mastodons.lip6.fr

Jury of Phd defense in 2014

- A. Martin:
  - Z. Liu (Northwestern Polytechnical University, Xian, Chine, 2014) (External reviewer)
  - N. Sutton-Charani (Université Technologique de Compiègne, 2014) (External reviewer)
  - A. Samet (Université d'Artois, 2014) (External reviewer)
  - M. Loudahi (Lille 1, 2014) (External reviewer)
  - D. Zhou (Université Technologique de Compiègne, 2014) (External reviewer)
  - T.N. Hoang (Université Technologique de Compiègne, 2014) (External reviewer)
  - S. Prigent (Rennes 1, 2014) (member)
  - M. Bou Farah (Université d'Artois, 2014) (member)

Steering committees in 2014

- D. Gross-Amblard: national conference on data management (BDA).
- A. Martin: international conference on belief functions (Belief), Extraction et Gestion de Connaissances (EGC) national conference.

Organizing committees in 2014

- A. Martin:
  - Organizer of EGC 2014[11].
  - Organization of the registrations to Belief 2014[12], Oxford, UK.
- D. Gross-Amblard:
  - PC chair of BDA 2014.

Program committees in 2014

- D. Gross-Amblard: PC member of EDBT 2014, Scientific committee of database summer school MDD 2014[13]
- A. Martin: PC member of Belief 2014, IPMU 2014, IJCNN 2014, IGARSS 2014, EGC 2014, RFIA 2014, ITAT 2014
- Z. Miklos: PC member of BDA'2014, OnToContent'2014

[11] http://egc2014.irisa.fr/
[12] http://cms.brookes.ac.uk/staff/FabioCuzzolin/BELIEF2014/
[13] http://webdam.inria.fr/SummerSchool-2014/

Journals Reviews in 2014

- Arnaud Martin: Aerospace Science and Technology, Fuzzy Sets and Systems, Transactions on Fuzzy Systems,Information Fusion, Infromation Sciences, International Journal of Approximate Reasoning, Journal of Applied Logic, International Journal of Pattern Recognition and Artificial Intelligence, Pattern Recognition
- Zoltan Miklos: Future Generation Computing Systems

9.2 Involvement in the Scientific Community

- Arnaud Martin: treasurer of BFAS society[14]
- Arnaud Martin: in charge of economical and business relations for EGC society[15]
- Zoltan Miklos: member of the ACM and SIGMOD

9.3 Teaching

- Our team is in charge of most of the database-oriented courses at University of Rennes 1 (ISTIC department and ESIR Engineering school), with courses ranging from classical databases to business intelligence, database theory or MapReduce paradigm.
- Database course (theory and practice) for ENS Rennes (one of the major French "grande école").
- Database course at INSA Rennes (also a "grande école").
- Arnaud Martin is in charg of a M2 research module on learning and data fusion at ENSSAT.
- Arnaud Martin teachs at the GDR Magis school 2014 a research course on the theory of belief functions.

10 Bibliography

Major publications by the team in recent years

[1] D. Baelde, P. Courtieu, D. Gross-Amblard, C. Paulin-Mohring, "Towards Provably Robust Watermarking", *in: Interactive Theorem Proving (ITP)*, Princeton, USA, August 2012.

[2] A. Bkakria, F. Cuppens, N. Cuppens-Boulahia, J. M. Fernandez, D. Gross-Amblard, "Preserving Multi-relational Outsourced Databases Confidentiality using Fragmentation and Encryption", *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA) 4*, 2, June 2013, p. 39–62.

[14] `http://www.bfasociety.org`
[15] `http://www.egc.asso.fr`

[3] T. Bouadi, M.-O. Cordier, R. Quiniou, "Computing Skyline Incrementally in Response to Online Preference Modification", *T. Large-Scale Data- and Knowledge-Centered Systems 10*, 2013, p. 34–59.

[4] P.-E. Doré, A. Martin, I. Abi-Zeid, A.-L. Jousselme, P. Maupin, "Belief functions induced by multimodal probability density functions, an application to search and rescue problem", *RAIRO - Operations Research 44*, 4, October 2010, p. 323–343.

[5] J.-C. Dubois, Y. Le Gall, A. Martin, "Brevet : Procédé de traitement de données d'accessibilité, dispositif et programme correspondant", France, 1451562, February 26 2014.

[6] A. Fiche, J.-C. Cexus, A. Martin, A. Khenchaf, "Features modeling with an alpha-stable distribution: Application to pattern recognition based on continuous belief functions", *Information fusion 14*, 2013, p. 504–520.

[7] D. Gross-Amblard, "Query-Preserving Watermarking of Relational Databases and XML Documents", *ACM Transactions on Database Systems (TODS) 36*, 1, March 2011, p. 1–24.

[8] H. Laanaya, A. Martin, D. Aboutajdine, A. Khenchaf, "Support Vector Regression of membership functions and belief functions - Application for pattern recognition", *Information fusion 11*, 4, October 2010, p. 338–350.

[9] Z. Lu, Z. Miklós, L. He, S. M. Cai, J. Z. Gu, "A novel multi-aspect consistency measurement for ontologies", *Journal of Web Engineering 10*, 1, 2011.

[10] H. Nguyen Quoc Viet, T. Nguyen Thanh, Z. Miklos, K. Aberer, "Reconciling Schema Matching Networks Through Crowdsourcing", *EAI Endorsed Transactions on Collaborative Computing 14*, 1, October 2014, p. e2.

[11] F. M. Suchanek, D. Gross-Amblard, S. Abiteboul, "Watermarking for Ontologies", *in : International Semantic Web Conference (ISWC)*, Bonn, Germany, October 2011.

[12] S. R. Yerva, Z. Miklós, K. Aberer, "What have fruits to do with technology? The case of Orange, Blackberry and Apple.", *in : International Conference on Web Intelligence, Mining and Semantics (WIMS'2011)*, ACM press, 2011.

[13] S. R. Yerva, Z. Miklós, K. Aberer, "Entity-based Classification of Twitter Messages", *International Journal of Computer Science & Applications 9*, 1, 2012, p. 88–115.

[14] S. R. Yerva, Z. Miklós, K. Aberer, "Quality-aware similarity assessment for entity matching in Web data", *Information Systems 37*, 2012, p. 336–351.

[15] K. Zhou, A. Martin, Q. Pan, Z.-G. LIU, "Median evidential c-means algorithm and its application to community detection", *Knowledge-Based Systems 74*, 2015, p. 69 – 88.

Books and Monographs

[1] B. Ben Yaghlane, G. Cleuziou, M. Lebbah, A. Martin (editors), *Fouille de données complexes. Complexité liée aux données multiples*, *Revue des Nouvelles Technologies de l'Information*, *RNTI-E-21*, 2011.

[2] C. Reynaud, A. Martin, R. Quiniou (editors), *Extraction et Gestion des Connaissances*, *Revue des Nouvelles Technologies de l'Information*, *RNTI-E-26*, 2014.

Doctoral dissertations and "Habilitation" theses

[3] T. Bouadi, *Analyse multidimensionnelle interactive de résultats de simulation. Aide à la décision dans le domaine de l'agroécologie*, PhD Thesis, 2013.

[4] M. Chebbah, *Source independence in the theory of belief functions*, PhD Thesis, June 2014.

Articles in referred journals and book chapters

[5] A. Bkakria, F. Cuppens, N. Cuppens-Boulahia, J. M. Fernandez, D. Gross-Amblard, "Preserving Multi-relational Outsourced Databases Confidentiality using Fragmentation and Encryption", *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA) 4*, 2, June 2013, p. 39–62.

[6] T. Bouadi, M.-O. Cordier, R. Quiniou, "Computing Skyline Incrementally in Response to Online Preference Modification", *T. Large-Scale Data- and Knowledge-Centered Systems 10*, 2013, p. 34–59.

[7] M. Chebbah, M. Kharoune, A. Martin, B. Ben Yaghlane, "Considérant la dépendance dans la théorie des fonctions de croyance", *Revue des Nouvelles Technologies Informatiques (RNTI) Fouille de données complexes, RNTI-E-27*, December 2014, p. 43–64.

[8] M. Chebbah, A. Martin, B. Ben Yaghlane, "Estimation de la fiabilité des sources des bases de données évidentielles", *Revue des Nouvelles Technologies de l'Information Fouille de données complexes. Complexité liée aux données multiples*, RNTI-E-21, 2011, p. 193–210.

[9] P.-E. Doré, A. Martin, I. Abi-Zeid, A.-L. Jousselme, P. Maupin, "Belief functions induced by multimodal probability density functions, an application to search and rescue problem", *RAIRO - Operations Research 44*, 4, October 2010, p. 323–343.

[10] A. Fiche, J.-C. Cexus, A. Martin, A. Khenchaf, "Features modeling with an alpha-stable distribution: Application to pattern recognition based on continuous belief functions", *Information fusion 14*, 2013, p. 504–520.

[11] A. Fiche, A. Khenchaf, J.-C. Cexus, A. Martin, M. Rochdi, "Étude statistique du fouillis de mer à partir de lois alpha-stables", *Traitement du signal 30*, 3-4, 2013, p. 243–271.

[12] D. Gross-Amblard, "Query-Preserving Watermarking of Relational Databases and XML Documents", p. 1–24.

[13] H. Laanaya, A. Martin, D. Aboutajdine, A. Khenchaf, "Support Vector Regression of membership functions and belief functions - Application for pattern recognition", *Information fusion 11*, 4, October 2010, p. 338–350.

[14] Z. Lu, Z. Miklós, L. He, S. M. Cai, J. Z. Gu, "A novel multi-aspect consistency measurement for ontologies", *Journal of Web Engineering 10*, 1, 2011.

[15] A. Martin, M.-H. Masson, "Editorial", *International Journal of Approximate Reasoning 53*, 2, 2012, p. 117, Theory of Belief Functions (BELIEF 2010).

[16] H. Nguyen Quoc Viet, T. Nguyen Thanh, Z. Miklos, K. Aberer, "Reconciling Schema Matching Networks Through Crowdsourcing", *EAI Endorsed Transactions on Collaborative Computing 14*, 1, October 2014, p. e2.

[17] H. Park, R. Pang, A. G. Parameswaran, H. Garcia-Molina, N. Polyzotis, J. Widom, "Deco: A System for Declarative Crowdsourcing", *PVLDB 5*, 12, 2012, p. 1990–1993.

[18] C. Rominger, A. Martin, "Recalage et fusion d'images sonar multivues : critère de dissimilarité et ignorance", *Revue des Nouvelles Technologies de l'Information Fouille de données complexes. Complexité liée aux données multiples*, RNTI-E-21, 2011, p. 233–248.

[19] S. R. Yerva, Z. Miklós, K. Aberer, "Entity-based Classification of Twitter Messages", *International Journal of Computer Science & Applications 9*, 1, 2012, p. 88–115.

[20] S. R. Yerva, Z. Miklós, K. Aberer, "Quality-aware similarity assessment for entity matching in Web data", *Information Systems 37*, 2012, p. 336–351.

[21] K. Zhou, A. Martin, Q. Pan, Z.-G. LIU, "Median evidential c-means algorithm and its application to community detection", *Knowledge-Based Systems 74*, 2015, p. 69 – 88.

Publications in Conferences and Workshops

[22] D. Attiaoui, P.-E. Doré, A. Martin, B. Ben Yaghlane, "A Distance Between Continuous Belief Functions", *in : International Conference on Scalable Uncertainty Management (SUM)*, Marburg, Germany, September 2012.

[23] D. Attiaoui, P.-E. Doré, A. Martin, B. Ben Yaghlane, "Inclusion within Continuous Belief Functions", *in : International Conference on Information Fusion*, Turkey, 9-12 July 2013.

[24] D. Baelde, P. Courtieu, D. Gross-Amblard, C. Paulin-Mohring, "Towards Provably Robust Watermarking", *in : Interactive Theorem Proving (ITP)*, Princeton, USA, August 2012.

[25] S. Ben Harrath, A. Martin, B. Ben Yaghlane, "Traitement des requêtes Top-K sur une base de données crédibiliste", *in : EGC-M: The 3rd International Conference on the Extraction and Management of Knowledge - Maghreb*, Tunisia, November 2013.

[26] A. Bkakria, F. Cuppens, N. Cuppens-Boulahia, D. Gross-Amblard, "Specification and Deployment of Integrated Security Policies for Outsourced Data", *in : 28th Annual IFIP WG 11.3 Working Conference on Data and Applications Security and Privacy (DBSec)*, 2014.

[27] T. Bouadi, S. Bringay, P. Poncelet, M. Teisseire, "RequÍtes skyline avec prise en compte des prÈfÈrences utilisateurs pour des donnÈes volumineuses", *in : EGC'10*, p. 399–404, 2010.

[28] T. Bouadi, M. Cordier, R. Quiniou, "Incremental Computation of Skyline Queries with Dynamic Preferences", *in : DEXA (1)*, p. 219–233, 2012.

[29] T. Bouadi, M. Cordier, R. Quiniou, "Computing Hierarchical Skyline Queries "On-the-Fly" in a Data Warehouse", *in : Data Warehousing and Knowledge Discovery - 16th International Conference, DaWaK 2014, Munich, Germany, September 2-4, 2014. Proceedings*, p. 146–158, 2014, `http://dx.doi.org/10.1007/978-3-319-10160-6_14`.

[30] T. Bouadi, M.-O. Cordier, R. Quiniou, C. Gascuel, "Analyse multidimensionnelle interactive de résultats de simulation d'un agro-système", *in : INFORSID*, 2011.

[31] T. Bouadi, M.-O. Cordier, R. Quiniou, "Requêtes skyline hiérarchiques", *in : 10e journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 2014), Vichy, RNTI, B-10*, Hermann, p. 51–68, Paris, Juin 2014.

[32] T. Bouadi, C. Gascuel-Odoux, M.-O. Cordier, R. Quiniou, P. Moreau, "Using Data Warehouses to extract knowledge from Agro-Hydrological simulations", *in: EGU General Assembly Conference Abstracts*, *15*, p. 9322, 2013.

[33] M. Chebbah, B. Ben Yaghlane, A. Martin, "Reliability estimation based on conflict for evidential database enrichment", *in: International Conference on Belief Functions*, France, 1-2 April 2010.

[34] M. Chebbah, A. Martin, B. Ben Yaghlane, "Mesure de concordance pour les bases de données évidentielles", *in: EGC*, Brest, France, January 2011.

[35] M. Chebbah, A. Martin, B. Ben Yaghlane, "About sources dependence in the theory of belief functions", *in: International Conference on Belief Functions*, France, 8-10 May 2012.

[36] M. Chebbah, A. Martin, B. Ben Yaghlane, "Positive and negative dependence for evidential database enrichment", *in: IPMU*, p. 575–584, Italy, July 2012.

[37] M. Chebbah, A. Martin, B. Ben Yaghlane, "Dépendance et indépendance des sources imparfaites", *in: EGC*, Toulouse, France, January 2013.

[38] A. Chettih, D. Gross-Amblard, D. Guyon, E. Legeay, Z. Miklos, "Crowd, une plateforme pour le crowdsourcing complexe (demo)", *in: Bases de données avancées (BDA)*, 2014. sous presse.

[39] S. B. Dhaou, M. Kharoune, A. Martin, B. Ben Yaghlane, "Belief Approach for Social Networks", *in: Belief 2014, Belief Functions: Theory and Applications*, *Lecture Notes in Computer Science, Vol. 8764*, p. 115–123, Oxford, United Kingdom, September 2014.

[40] P. Djiknavorian, A. Martin, D. Grenier, P. Valin, "Etude comparative d'approximation de fonctions de croyances généralisées", *in: Rencontres Francophones sur la Logique Floue et ses Applications (LFA)*, Lannion, France, 18-19 November 2010.

[41] P.-E. Doré, A. Fiche, A. Martin, "Models of belief functions? Impacts for patterns recognition", *in: International Conference on Information Fusion*, Edinburgh, UK, 26-29 July 2010.

[42] P.-E. Doré, A. Martin, "About using beliefs induced by probabilities", *in: International Conference on Belief Functions*, France, 1-2 April 2010.

[43] P.-E. Doré, A. Martin, "Approximation du théorème de Bayes généeralisé pour les fonctions de croyance continues", *in: Rencontres Francophones sur la Logique Floue et ses Applications (LFA)*, Lannion, France, 18-19 November 2010.

[44] P.-E. Doré, C. Osswald, A. Martin, A.-L. Jousselme, P. Maupin, "Continuous belief functions to qualify sensors performances", *in: European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, Belfast, UK, 29 June - 1 July 2011.

[45] J.-C. Dubois, Y. Le Gall, A. Martin, "Designing a Belief Function-Based Accessibility Indicator to Improve Web Browsing for Disabled People", *in: Belief 2014, Belief Functions: Theory and Applications*, *Lecture Notes in Computer Science, Vol. 8764*, p. 134 – 142, Oxford, United Kingdom, September 2014.

[46] A. Essaid, B. Ben Yaghlane, A. Martin, "Gestion du conflit dans l'appariement des ontologies", *in: EGC - Atelier Graphes et Appariement d'Objets Complexes*, Brest, France, January 2011.

[47] A. Essaid, A. Martin, G. Smits, B. Ben Yaghlane, "Processus de Décision Crédibiliste pour l'Alignement des Ontologies", *in : Rencontres Francophones sur la Logique Floue et ses Applications (LFA)*, Reims, France, October 2013.

[48] A. Essaid, A. Martin, G. Smits, B. Ben Yaghlane, "A Distance-Based Decision in the Credal Level", *in : International Conference on Artificial Intelligence and Symbolic Computation (AISC 2014)*, p. 147 – 156, Sevilla, Spain, December 2014.

[49] A. Essaid, A. Martin, G. Smits, B. Ben Yaghlane, "Uncertainty in Ontology Matching: A Decision Rule-Based Approach", *in : International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, p. 46 – 55, Montpellier, France, July 2014.

[50] Z. Faget, P. Rigaux, D. Gross-Amblard, V. Thion-Goasdoué, "Modeling synchronized time series", *in : IDEAS*, p. 82–89, 2010.

[51] A. Fiche, J.-C. Ceuxs, A. Martin, A. Khenchaf, "Intérêt des distributions alpha-stables pour la classification de données complexes", *in : EGC - Atelier Fouille de données complexes - Complexité liée aux données multiples*, Brest, France, January 2011.

[52] A. Fiche, J.-C. Cexus, A. Khenchaf, M. Rochdi, A. Martin, "Characterization of EM sea clutter with alpha-stable distribution", *in : International Geoscience and Remote Sensing Symposium (IGARSS)*, Munich, Germany, July 2012.

[53] A. Fiche, J.-C. Cexus, A. Martin, A. Khenchaf, "Estimation d'un mélange de distributions alpha-stables à partir de l'algorithme EM", *in : Rencontres Francophones sur la Logique Floue et ses Applications (LFA)*, Lannion, France, 18-19 November 2010.

[54] A. Fiche, A. Khenchaf, J.-C. Cexus, M. Rochdi, A. Martin, "A comparative statistical analysis of sea clutter", *in : Radar conference*, Glasgow, UK, October 2013.

[55] A. Fiche, A. Martin, J.-C. Cexus, A. Khenchaf, "Continuous belief functions and alpha-stable distributions", *in : International Conference on Information Fusion*, Edinburgh, UK, 26-29 July 2010.

[56] A. Fiche, A. Martin, J.-C. Cexus, A. Khenchaf, "A comparison between a Bayesian approach and a method based on continuous belief functions for pattern recognition", *in : International Conference on Belief Functions*, France, 8-10 May 2012.

[57] A. Gal, M. Katz, T. Sagi, M. Weidlich, K. Aberer, Z. Miklos, N. Q. V. Hung, E. Levy, V. Shafran., "Completeness and Ambiguity of Schema Cover", *in : 21st International Conference on Cooperative Information Systems (CoopIS 2013)*, 2013.

[58] A. Gal, T. Sagi, M. Weidlich, E. Levy, V. Shafran, Z. Miklós, N. Q. V. Hung, "Making Sense of Top-K Matchings - A Unified Match Graph for Schema Matching", *in : Ninth International Workshop on Information Integration on the Web (IIWeb 2012)*, 2012.

[59] N. Q. V. Hung, X. H. Luong, Z. Miklos, T. T. Quan, K. Aberer, "Collaborative Schema Matching Reconciliation", *in : 21st International Conference on Cooperative Information Systems (CoopIS 2013)*, 2013.

[60] N. Q. V. Hung, X. H. Luong, Z. Miklos, T. Q. Thanh, K. Aberer, " An MAS Negotiation Support Tool for Schema Matching (Demonstration)", *in : Twelfth International Conference on Autonomous Agents and Multiagent Systems (AAMAS'2013)*, 2013.

[61] N. Q. V. Hung, N. T. Tam, Z. Miklos, K. Aberer, A. Gal, M. Weidlich, "Pay-as-you-go Reconciliation in Schema Matching Networks", *in : 30th International Conference on Data Engineering (ICDE 2014)*, 2014.

[62] N. Q. V. Hung, N. T. Tam, Z. Miklós, K. Aberer, "On Leveraging Crowdsourcing Techniques for Schema Matching Networks", *in : DASFAA 2013*, 2013.

[63] A. Jammali, M. A. Bach Tobji, A. Martin, B. Ben Yaghlane, "Indexing Evidential Data", *in : Third World Conference on Complex Systems*, Marrakech, Morocco, November 2014.

[64] S. Jendoubi, B. Ben Yaghlane, A. Martin, "Belief Hidden Markov Model for speech recognition", *in : 5th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO)*, Tunisia, April 2013.

[65] S. Jendoubi, A. Martin, L. Liétard, B. Ben Yaghlane, "Classification of Message Spreading in a Heterogeneous Social Network", *in : International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, p. 66 – 75, Montpellier, France, July 2014.

[66] F. Karem, M. Dhibi, A. Martin, "Combinaison crédibiliste de classification supervisée et non-supervisée", *in : EGC*, Bordeaux, France, January 2012.

[67] F. Karem, M. Dhibi, A. Martin, "Combination of supervised and unsupervised classification using the theory of belief functions", *in : International Conference on Belief Functions*, France, 8-10 May 2012.

[68] M. Kharoune, A. Martin, "Mesure de dépendance positive et négative de sources crédibilistes", *in : EGC - Atelier Fouille de données complexes*, Toulouse, France, January 2013.

[69] E. Le Chenadec, Gillesand Cantéro, T. Landeau, A. Martin, J.-C. Cexus, Y. Dupas, R. Courtis, A. Bertholom, "Development of a generic data fusion tool for Rapid Environmental Assessment", *in : OCOSS*, France, June 2010.

[70] J. Lengrand-Lambert, A. Martin, R. Courtis, "Sonar images segmentation and classification fusion", *in : European Conference on Undewater Acoustics*, Turkey, July 2010.

[71] D. Leprovost, L. Abrouk, N. Cullot, D. Gross-Amblard, "Temporal Semantic Centrality for the Analysis of Communication Networks", *in : International Conference on Web Engineering (ICWE)*, Berlin, Germany, July 2012.

[72] Z. Lu, Z. Miklós, S. Cai, J. Gu, "Measuring Taxonomic Consistency of Ontologies Using Lexical Semantic Relatedness", *in : The Fifth International Conference on Digital Information Management (ICDIM'2010)*, IEEE, 2010.

[73] W. Maalel, K. Zhou, A. Martin, Z. Elouedi, "Belief Hierarchical Clustering", *in : 3rd International Conference on Belief Functions*, p. 68 – 76, Oxford, United Kingdom, September 2014.

[74] A. Martin, J.-C. Cexus, G. Le Chenadec, E. Cantéro, T. Landeau, Y. Dupas, R. Courtis, "Autonomous Underwater Vehicle sensors fusion by the theory of belief functions for Rapid Environment Assessment", *in : European Conference on Undewater Acoustics*, Turkey, July 2010.

[75] A. Martin, "Le conflit dans la théorie des fonctions de croyance", *in : EGC*, Hammamet, Tunisie, January 2010.

[76] A. Martin, "Multi-view fusion based on belief functions for Autonomous Underwater Vehicle", *in : OCOSS*, France, June 2010.

[77] A. Martin, "About conflict in the theory of belief functions", *in : International Conference on Belief Functions*, France, 8-10 May 2012.

[78] Z. Miklós, N. Bonvin, P. Bouquet, M. Catasta, D. Cordioli, P. Fankhauser, J. Gaugaz, E. Ioannou, H. Koshutanski, A. Mana, C. Niederée, T. Palpanas, H. Stoermer, "From Web Data to Entities and Back", *in : The 22nd International Conference on Advanced Information Systems Engineering (CAiSE'10)*, *LNCS*, *6051*, Springer, p. 302–316, 2010.

[79] R. Narendula, T. G. Papaioannou, Z. Miklós, K. Aberer, "Tunable Privacy for Access Controlled Data in Peer-to-Peer systems", *in : Proceedings of the 22nd International Teletraffic Congress (ITC 22)*, 2010.

[80] Q. V. H. Nguyen, T. K. Wijaya, Z. Miklos, K. Aberer, E. Levy, V. Shafran, A. Gal, M. Weidlich, "Minimizing Human Effort in Reconciling Match Networks", *in : 32nd International Conference on Conceptual Modeling (ER 2013)*, 2013.

[81] I. Ouled Dlala, D. Attiaoui, A. Martin, B. Ben Yaghlane, "Trolls Identification within an Uncertain Framework", *in : International Conference on Tools with Artificial Intelligence - ICTAI*, Limassol, Cyprus, November 2014.

[82] J. Park, M. Chebbah, S. Jendoubi, A. Martin, "Second-Order Belief Hidden Markov Models", *in : Belief 2014*, p. 284 – 293, Oxford, United Kingdom, September 2014.

[83] J. Park, M. Kharoune, A. Martin, "Etude de cas sur DBpédia en français", *in : EGC - Atelier Fouille de données complexes*, Rennes, France, January 2014.

[84] I. Rolewicz, M. Catasta, H. Jeung, Z. Miklós, K. Aberer, "Building a Front End for a Sensor Data Cloud", *in : The First International Workshop on Cloud for High Performance Computing (C4HPC), Special Session of the 11th International Conference on Computational Science and Its Applications (ICCSA'2011)*, 2011.

[85] C. Rominger, A. Martin, "Landmarks on sonar image classification for registration", *in : European Conference on Undewater Acoustics*, Turkey, July 2010.

[86] C. Rominger, A. Martin, "Using the conflict: An application to sonar image registration", *in : International Conference on Belief Functions*, France, 1-2 April 2010.

[87] S. B. Roy, I. Lykourentzou, S. Thirumuruganathan, S. Amer-Yahia, G. Das, "Crowds, not Drones: Modeling Human Factors in Interactive Crowdsourcing", *in : DBCrowd*, p. 39–42, 2013.

[88] F. Smarandache, D. Han, A. Martin, "Comparative study of contradiction measures in the theory of belief functions", *in : International Conference on Information Fusion*, Singapore, 9-12 July 2012.

[89] F. Smarandache, A. Martin, C. Osswald, "Contradiction measures and specificity degrees of basic belief assignments", *in : International Conference on Information Fusion*, Chicago, USA, 5-8 July 2011.

[90] F. M. Suchanek, D. Gross-Amblard, S. Abiteboul, "Watermarking for Ontologies", *in: International Semantic Web Conference (ISWC)*, Bonn, Germany, October 2011.

[91] G. Vakili, T. G. Papaioannou, Z. Miklós, K. Aberer, S. Khorsandi, "Analyzing the Emergence of Semantic Agreement among Rational Agents", *in: 13th IEEE Conference on Commerce and Enterprise Computing (short paper)*, 2011.

[92] S. R. Yerva, Z. Miklós, K. Aberer, "It was easy, when apples and blackberries were only fruits", *in: Third WePS Evaluation Workshop: Searching Information about Entities in the Web*, 2010.

[93] S. R. Yerva, Z. Miklós, K. Aberer, "Towards better entity resolution techniques for Web document collections", *in: 1st International Workshop on Data Engineering meets the Semantic Web (DESWeb'2010) (co-located with ICDE'2010)*, 2010.

[94] S. R. Yerva, Z. Miklós, K. Aberer, "What have fruits to do with technology? The case of Orange, Blackberry and Apple.", *in: International Conference on Web Intelligence, Mining and Semantics (WIMS'2011)*, ACM press, 2011.

[95] S. R. Yerva, Z. Miklós, F. Grosan, T. Alexandru, K. Aberer, "TweetSpector: Entity-based retrieval of Tweets", *in: SIGIR'2012*, 2012. (demo paper).

[96] K. Zhou, A. Martin, Q. Pan, "Evidential Communities for Complex Networks", *in: 15th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, p. 557 – 566, Montpellier, France, July 2014.

[97] K. Zhou, A. Martin, Q. Pan, "Evidential-EM Algorithm Applied to Progressively Censored Observations", *in: 15th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, p. 180 – 189, Montpellier, France, July 2014.

Patents

[98] J.-C. Dubois, Y. Le Gall, A. Martin, "Procédé de traitement de données d'accessibilité, dispositif et programme correspondant", France, 1451562, February 26 2014.