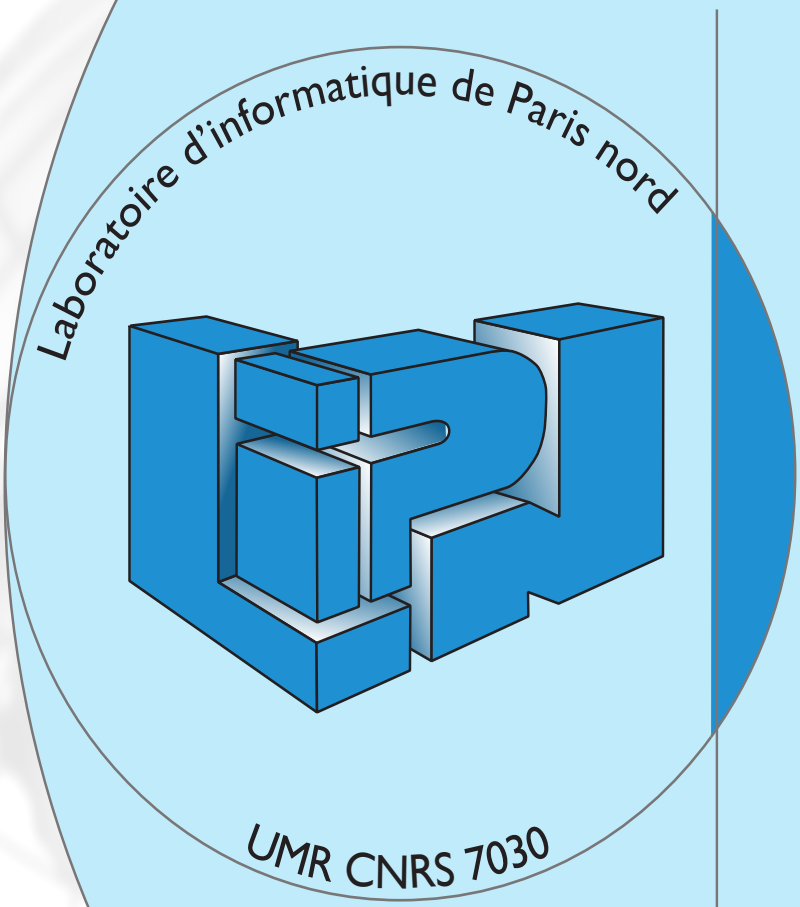


Rapport 2012-2017

Projet 2019-2023

Apprentissage Artificiel et Applications

Partie I **Le LIPN** **page 5**

The LIPN at a glance 7

Scientific activities
Environment
Doctoral studies

Partie II **Equipe A3** **page 11**

Présentation générale 13

Overview
Membres de l'équipe

Activités de recherche 19

Bilan 2012-2017
Analyse AFOM/SWOT
Perspectives et projet

Produits de la recherche 33

Synthèse des publications
Publications
Publications antérieures à l'arrivée au LIPN
Réalisations informatiques (logiciels), Brevets

Activités de recherche et indices de reconnaissance 55

Participation à des comités, animation scientifique
Organisation de colloques et de congrès
Activités d'évaluation
Postdoctorants et chercheurs accueillis
Interactions avec l'environnement
Contrats de recherche publics
Indices de reconnaissance

Formation à la recherche 67

Masters, Écoles doctorale et post-doctorale
Thèses et Habilitations
Stages de master, post-doctorats

Partie III **Acronymes** **page 73**

Partie I
Le LIPN

Chapitre 1
The LIPN at a glance

The LIPN[1] (Laboratoire d'Informatique de Paris-Nord) was created in 1985, has been associated with CNRS since January 1992, before becoming a CNRS UMR in January 2001. The LIPN is the computer science laboratory of University Paris 13 which is part of the COMUE Université Sorbonne Paris Cité.

The laboratory comprises 5 teams with 80 permanent staff (full-time researchers or professors and associate professors), 10 technical and administrative staff, as well as PhD students and postdocs. It has overall more than 170 members. The LIPN has experienced an impressive growth during the past decade, which allowed for strengthening its major areas of expertise.

The scientific activities display many collaborations, be they national, international, or through research projects, in particular with the industry. LIPN members are extremely involved on the national scene, in national boards (at CNRS, CNU — national universities council, HCERES, etc.), and in institutional cooperation projects (ANR). On the international scene, members are widely present in editorial boards, programme committees, expert pools of different countries, international schools, and conduct collaborative research with colleagues worldwide.

1. http://lipn.univ-paris13.fr

1.1 Scientific activities

The laboratory is structured into 5 teams, focussing on the following main areas.

- **A³ (Apprentissage Artificiel et Applications): Machine Learning and Applications**

The A³ team tackles machine learning problems and covers a wide spectrum of issues, ranging from supervised and unsupervised learning to reinforcement learning. Its research is fed, coordinated and evaluated thanks to various applications in the field of pattern recognition and data mining. Research in A³ focuses on the following main topics: algebraic and logical models of learning, collaborative and transfer learning, cluster analysis, dimensionality reduction, link prediction in social networks, recommender systems, and learning structures from complex data.

- **AOC (Algorithmes et Optimisation Combinatoire): Algorithms and Combinatorial Optimisation**

The AOC team develops research in optimisation on graphs, mathematical programming, parallel and distributed computing. Optimisation on graphs is conducted with a particular focus on complexity, polyedral theory and approximation. The team has expertise in design and analysis of mathematical programming approaches and algorithms, and develops both exact and heuristic approaches for solving linear and nonlinear problems. Many distributed environments issues are also considered: distributed middleware and architectures, distributed programming and distributed algorithms.

- **CALIN (Combinatoire, ALgorithmique et INteractions): Combinatorics, ALgorithmics, and INteractions**

The CALIN team brings together researchers with skills in a variety of aspects of combinatorics (analytic, bijective, geometric, and algebraic). They are interested in the complexity of algorithms, with a focus on determining their behaviour on average or in distribution. They also have an interest in the fine-grained analysis of data structures, and study problems of physics with a distinct combinatorial flavour, or apply methods from physics to combinatorial problems. The team is organised into two subgroups: one of them focuses on the analysis of algorithms and combinatorial structures, and the other is devoted to the interactions between combinatorics, geometry and physics.

- **LCR (Logique, Calcul et Raisonnement): Logic, Computation and Reasoning**

The LCR team tackles different aspects of computational and software models, from theory to applications. It is organised into two subgroups. The specification and verification group addresses methodologies for the development of specifications, and verification through modular, compositional, distributed and parameterised approaches. The logic, theory of computation and programming languages group has expertise in computational issues of logical systems and formal semantics of programming languages: use of tools and methodologies issued from Linear Logic to analyse resource sensitive properties of programs, study graph formalisms for representing proofs ; algebraic and geometric foundations of computation.

- **RCLN (Représentation des Connaissances et Langage Naturel): Knowledge Representation and Natural Language**

The RCLN team is interested in natural language for its expressive capacity, and in knowledge representation, in particular for its connection to semantic analysis and text understanding. It tackles corpus analysis and semantic annotation of texts. Search and

structuration of semantic web knowledge is achieved by the design of ontologies and knowledge patterns. Text understanding and knowledge acquisition are integrated in a virtuous circle between language, data and knowledge.

1.2 Environment

The LIPN has many research cooperations with its environment: other laboratories at University Paris 13, but also in the COMUE Université Sorbonne Paris Cité, with industry, etc.

- **MathSTIC research federation, and University Paris 13 laboratories**

Research collaborations with other laboratories of University Paris 13 have been favoured over the years, in particular with LAGA [2] (UMR 7539, mathematics laboratory), L2TI [3] (signal and image processing).

This led to the creation of a new research federation in Mathematics and Information Technology (MathSTIC [4], which became a CNRS Research Federation – FR 3734 – in January 2016), gathering members of LAGA, LIPN, and L2TI laboratories, in order to enhance cross-fertilisation in three of their major areas of expertise:

— Optimisation and learning applied to digital contents;
— High-performance computing, distributed systems;
— Mathematical physics, statistical physics, combinatorics.

- **Regional setting: COMUE Université Sorbonne Paris Cité, and EFL LabEx**

The COMUE Université Sorbonne Paris Cité comprises 8 universities and colleges. It includes University Paris 5, and University Paris 7, which both have computer science laboratories (LIPADE, and IRIF). The COMUE offers support for research projects between these laboratories.

The team RCLN at LIPN is a major actor in the EFL LabEx (Empirical Foundations of Linguistics). Its multidisciplinarity nature is a key to evolution of research topics at the meeting point of linguistics and computer science.

- **International cooperations**

Members of the laboratory have many collaborations worldwide, with exchanges of researchers. Moreover, approximately 15 foreign researchers are invited for a month at our laboratory each year to carry collaborative research. Cooperation projects involve laboratories in Germany, Italy, Netherlands, Norway, Poland, Tunisia, Canada, Argentina, Brazil, Chile, Mexico, Uruguay, Singapore, Taïwan, Vietnam.

- **Industrial collaborations**

Most research areas of the LIPN also participate in collaborations with the industry, mainly through research projects (ANR, FUI, FEDER) and CIFRE PhDs. Some large national or European collaborative projects involve both academic and industrial research laboratories. Some researchers of the LIPN have also created a start-up company, or benefit from a few months industrial experience. Moreover, the LIPN is involved in different boards of 3 business clusters (pôles de compétitivité) which favour technology transfer and research projects with companies at a regional level.

2. http://www.math.univ-paris13.fr/laga/
3. http://www-l2ti.univ-paris13.fr/
4. http://mathstic.univ-paris13.fr/

1.3 Doctoral studies

The doctoral studies are organised within the doctoral school ED146 “Galilée : Santé, Technologie, Santé”, which addresses three major scientific areas, LIPN being part of the first group:

1. mathematics, computer science, signal processing;
2. physics, materials, engineering sciences;
3. health sciences, medicine, human biology, chemistry, ethology.

The doctoral school organises mandatory courses, and allocates grants to doctoral students. Every year, approximately five of them are allocated to the LIPN (out of 25 for the whole doctoral school).

- **PhDs and Habilitations**

Over the 2012–2017 period, 54 PhD defences took place. Half of the doctoral students were financed by a grant from the doctoral school, and the others through research projects, industry, or international support.

During the same period, 13 habilitations (Habilitations à diriger les Recherches) were delivered.

The laboratory organises an oral presentation halfway through the PhD (after 18 months), in front of a jury comprising an external member, who often is to become one of the reviewers. The aim is thus to assess the work that has already been done, and confirm the future directions of the doctoral work. Doctoral students benefit from a broader view of their work as well as advice from the jury.

- **Masters courses**

A masters programme in computer science is delivered at University Paris 13. It currently features two specialities:

— EID^2 (Exploration Informatique des Données et Décisionnel): Machine Learning and Data Science;

— PLS (Programmation et Logiciels Sûrs): Programming and Secure Software.

Both specialities lead to either industrial or research careers, depending on the master thesis.

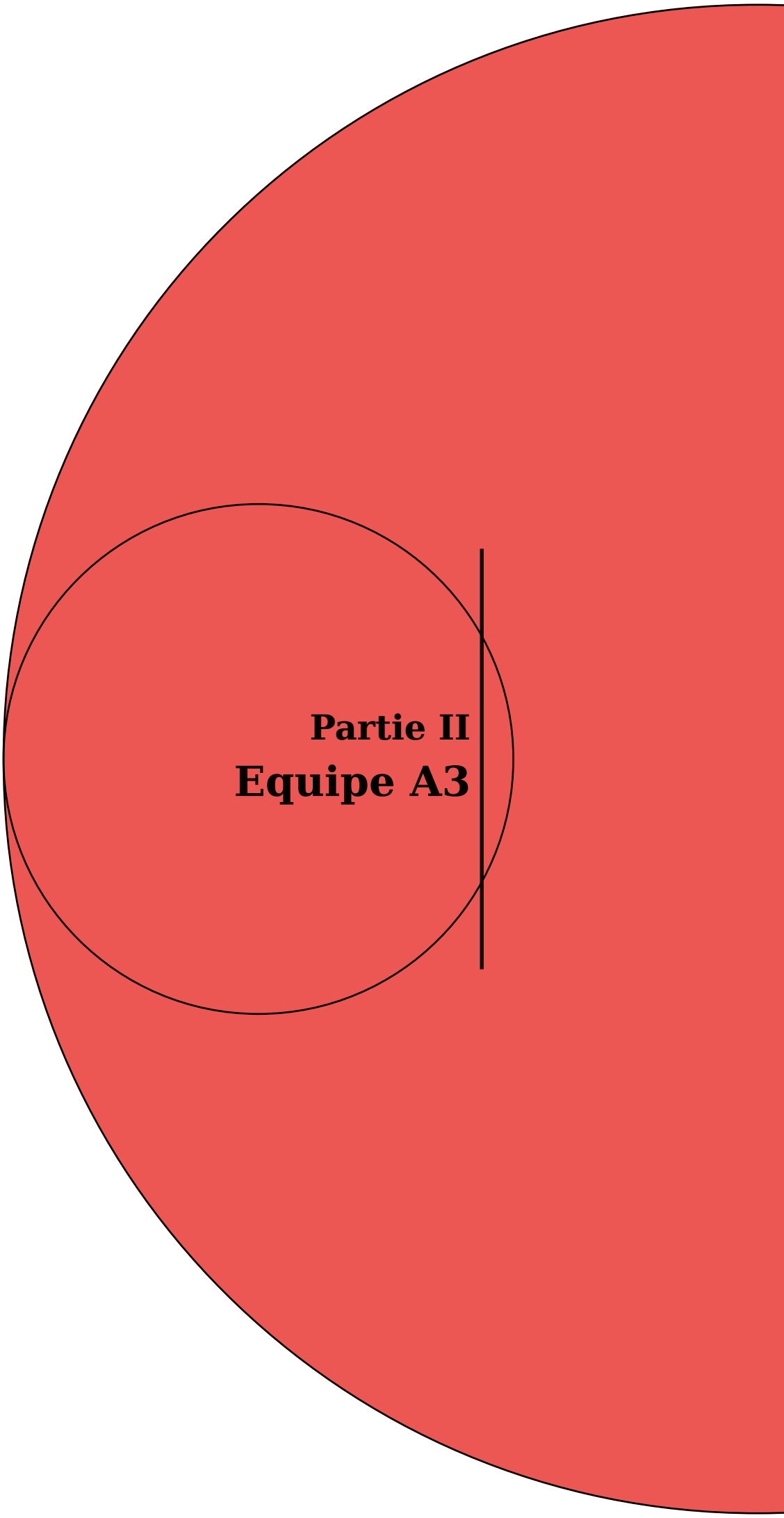
LIPN researchers also deliver courses in other masters programmes of universities in the parisian region.

Furthermore, they are involved in international cooperations for setting up masters programmes in foreign countries, e.g. Madagascar, Vietnam.

- **Summer schools and conferences**

LIPN supports the organisation and participation in thematic schools, by encouraging doctoral students attendance (which is recognised as mandatory courses), talks by members of the laboratory, as well as financial support to the organisation of such events.

In particular, Logics and Interactions 2012, Proof Theory 2012, Alea 2012, AAFD, JPOC/ISCO, Transversal Aspects of Tilings 2016, Tilings and Tesselations 2015, were co-organised by LIPN members. Some of these thematic schools take place in University Paris 13, while others are elsewhere in France or abroad.


Partie II
Equipe A3

Chapitre 1
Présentation générale

Sommaire

1.1 Overview 13
1.2 Membres de l'équipe 15
1.2.1 Membres permanents 15
1.2.2 Doctorants au 1er juillet. 2017 15
1.2.3 Personnel temporaire du 1/9/2012 au 1/9/2017 16
1.2.4 Évolution de l'équipe (permanents) 16
1.2.5 Responsabilités des membres de l'équipe 17

1.1 Overview

Machine Learning is a scientific discipline concerned with the design and development of algorithms that improve their behaviour from experience (observations, labelled or not). It borrows techniques and theoretical background from the field of Artificial Intelligence, Logics, Statistics, to form a highly challenging domain. Machine Learning is now a mature field of computer science, with solid theoretical models and results, and a broad range of applications, both in industry and in multi-disciplinary research. It also

became quite popular to a wide audience, thanks to Deep Learning successes in Game playing and with the emergence of Data Science, following the prominence of massive datasets requiring intelligent analysis.

Our group covers a remarkably broad range of topics, from Statistical Learning to Graph Mining and Reinforcement Learning. This allows us to be quite reactive to cope with new challenges raised by emerging applications of Machine Learning. It also makes it possible to study innovative combinations of learning methods for tackling complex problems.

The team A³ was created in 2005, with the goal of gathering Machine Learning related research of LIPN within a strongly coherent group, highly visible on the national and international research scenes.

The structure of the group has evolved over the years, and now consists of four research axes, for which the group is widely recognized :

— *Machine Learning for Decision Support*, which gathers research that deals with supervised learning related to decision support topics;
— *Collaborative Unsupervised Learning for Knowledge Transfer* which studies non supervised learning in the context of distributed data, and transfer from one learning task to another.
— *Learning from Graphs*, motivated by data mining issues, with a strong focus on mining structures attributed graphs and community detection and link prediction from complex (social) networks.
— *Learning topological models from massive datasets*, which studies unsupervised learning of the hidden topology in the data, with a strong focus on online and massively distributed learning techniques.

In each of these axes, theoretical issues are addressed, and algorithms and software are developed, in the context of collaborative projects, including both academic and industrial partners. The team publishes in main Machine Learning international conferences (IJCNN, ICML, AAMAS to name a few) as well as main international journals (Artificial Intelligence, Machine Learning, Pattern Recognition, . . .).

The strong links of the group with industrial partners is in particular testified by the high number of PhD theses in the group that take place in an industrial context, and its participation in FUI contracts. A member of the group has also created in 2010 a technological spin-off of University of Paris 13, that develops services in Intelligent Data Analysis such as Data Mining, Knowledge Discovery in Databases and Predictive Analytics.

The research topics of the team for the next five years aim at investigating promising areas where the team may emerge as a leader. These promising research topics are the following:

— Unsupervised learning in the framework of optimal transport theory
— Deep Learning and quantum approaches
— Mining augmented and heterogeneous graphs
— New challenges for scalable computational statistics

The group actively participates in national projects, and has several international collaborations, in particular through a CHIST-ERA contract. The researchers of the team are involved in teaching at the undergraduate and graduate levels, in Paris 13 University and in other institutions. In particular, the group has created a highly visible second year master specialisation on Data Mining, Analytics, and Knowledge Discovery (*Exploration Informatique des Données et Décisionnel* (EID^2)) of the Paris 13 master in computer science.

1.2 Membres de l'équipe

1.2.1 Membres permanents

(membres ou, *en italique, anciens membres* au 01/10/2017)

	Nom	*Prénom*	*Situation*	*Institution*
	Alphonse	*Erick*	MCF	IG
	Azzag	Hanane	MCF	IUTV
	Bennani	Younès	PU	IG
	Bouthinon	Dominique	MCF	IUTV
	Champesme	Marc	MCF	IG
	Cabanes	Guénaël	MCF	IG
	Chevaleyre	**Yann**	PU	IG
	Gérard	Pierre	MCF	IUTV
	Grozavu	Nistor	MCF	IG
	Guérif	Sébastien	MCF	IG
	Kanawati	Rushed	MCF	IUTV
	Lebbah	Mustapha	MCF	IUTV
	Matei	Basarab	MCF	IG
	Osmani	Aomar	MCF	IUTV
	Rouveirol	**Céline**	PU	IG
	Santini	Guillaume	MCF	IUTV
	Soldano	Henry	MCF	IG

1.2.2 Doctorants au 1er juillet. 2017

	Nom	*Prénom*	*Situation*	*Financement*
	Beck	Gaël	CIFRE	Kameleoon

	Falih	Issam	AM	Université Paris 13
	Léger	Hippolyte	AM	Université Paris 13
	Masmoudi	Nesrine	AM	Université Paris 13
	Rastin	Parisa	CIFRE	Mindlytix
	Rifi	Mouna	CIFRE	EDF
	Sarazin	Tugdual	CIFRE	Smokewatchers
	Veillon	Lise-Marie	AM	Université Paris 13

1.2.3 Personnel temporaire du 1/9/2012 au 1/9/2017

Nom	*Prénom*	*Situation*	*Début*	*Fin*	*Financement*
Alizadeh	Pegah	Allocataire	01/10/2012	20/09/2015	UP13
		Doctorante	01/10/2016	09/12/2016	
		Postdoc	09/12/2016	31/08/2017	
Benchettara	Nasserin	Postdoc	20/12/2011	10/12/2012	
Chaibi	Amine	CIFRE	29/11/2010	31/08/2014	
		Postdoc	15/12/2010	29/11/2013	
Chebil	Inès	Doctorante	01/11/2010	26/09/2014	
		Postdoc	27/09/2014	31/08/2015	
Chelly	Zeineb	Postdoc	2016	2018	Bourse Marie Sklodowska Curie
Cohen Ganouna	Raphaël	AM	01/10/2014	abandon 2017	UP13
Coutant	Anthony	Postdoc	15/01/2016	30/04/2018	ANR Adalab
Doan	Nhat-Quang	Doctorant	14/12/2010	09/12/2013	Vietnam
Duong	Thanh Buu	Postdoc	01/12/2014	31/08/2016	Square-Predict
Essaidi	Moez	CIFRE	01/12/2008	02/07/2013	
Fenzar Sadki	Jalia	Postdoc	07/07/2011	02/11/2013	
Ghassani	Mohamad	Doctorant	01/09/2010	07/11/2013	
Ghesmoune	Mohammed	Doctorant	01/11/2013	25/11/2016	
		ATER	26/11/2016	31/08/2017	
Hamdi	Fatma	Doctorante	01/11/2009	06/12/2012	
Hindawi	Mohammed	Postdoc	07/05/2013	30/04/2014	
Jaziri	Rakia	CIFRE	15/10/2009	27/06/2013	
		CIFRE	27/06/2013	31/08/2014	
Lachaud	Antoine	Postdoc	01/03/2016	31/12/2016	ANR Coclico
Laclau	Charlotte	Postdoc	01/06/2016	31/10/2016	ANR Coclico
Le	Van Minh	Doctorant	01/10/2012	13/12/2016	Vietnam
Mouhoubi	Karima	Doctorante, ATER	01/10/2009	20/09/2013	
		Postdoc	20/09/2013	31/12/2013	
Pujari	Manisha	Doctorante	01/10/2010	04/03/2015	
		Postdoc	05/03/2015	01/10/2015	
Redko	Ievgen	Doctorant	01/10/2012	26/11/2015	
		Postdoc	26/11/2015	31/03/2016	
Rodrigues	Christophe	Doctorant, ATER	01/10/2007	21/01/2013	
		Postdoc	21/01/2013	30/09/2015	
Yacoubi	Zied	Doctorant, ATER	01/10/2010	04/12/2014	

1.2.4 Évolution de l'équipe (permanents)

Les effectifs de l'équipe A³ sont en légère progression depuis 2012. Il y a eu une arrivée depuis le LAGA, laboratoire de mathématiques de l'Université Paris 13, nous avons recruté un Maître de Conférences, et il y a eu un départ de Professeur en mutation.

Un fait remarquable concernant la composition de l'équipe est qu'elle accueille de nombreux doctorants, en particulier en CIFRE.

Départs

2017 Y. Chevaleyre, PU, IG (mutation PU Univ. Paris-Dauphine)

Arrivées

2012 G. Cabanes, MCF IG (thèse LIPN ; post-doc Canberra, Australie)

2015 B. Matei, MCF IG (MCF au LAGA, Université Paris 13)

1.2.5 Responsabilités des membres de l'équipe

Cette section résume les responsabilités importantes assumées par des membres de l'équipe.

• Y. Bennani est vice-président de l'Université Paris 13 chargé de la Transformation Numérique depuis 2016.
• C. Rouveirol est directrice adjointe de l'IG en charge de l'école d'ingénieurs Sup Galilée depuis 2011.
• R. Kanawati est chef du département Réseaux & Télécommunications - IUT de Villetaneuse depuis 2016.
• P. Gérard a été directeur du département Informatique - IUT de Villetaneuse jusqu'en 2014, puis directeur du centre de formation continue de 2014 à 2015 et enfin responsable du CREIP (Centre de Relations Extérieures pour l'Insertion Professionelle) de l'IUT de Villetaneuse depuis 2015.
• M. Champesme est trésorier national du SNESUP.

Chapitre 2
Activités de recherche

Sommaire

2.1 Bilan 2012-2017 19
2.1.1 Apprentissage pour l'Aide à la Décision 20
2.1.2 Apprentissage non supervisé de représentations pour le transfert et la collaboration 21
2.1.3 Apprentissage dans les Graphes 23
2.1.4 Apprentissage de modèles topologiques à partir de données massives 25
2.2 Analyse AFOM/SWOT 26
2.3 Perspectives et projet 27
2.3.1 Vers de nouveaux cadres théoriques pour l'apprentissage de représentations 28
2.3.2 Apprentissage et structures 30

2.1 Bilan 2012-2017

L'équipe A³ était structurée en 2012 par types de méthodes d'apprentissage : un axe *Modèles algébriques et logiques de l'apprentissage,* un axe *Apprentissage non super-*

visé collaboratif et évolutif, et un axe *Apprentissage de structures à partir de données hétérogènes*.

La composition de l'équipe et surtout sa structuration ont sensiblement évolué depuis 2012. Le recrutement en 2010 d'un professeur actif dans le domaine de l'*Apprentissage pour l'Aide à la Décision* a motivé la création d'un axe sur ce thème. L'axe *Apprentissage non supervisé de représentations pour le transfert et la collaboration* est une évolution directe d'un ancien axe de l'équipe, qui s'est renforcé avec l'arrivée de deux MCF. L'axe *Apprentissage à partir de Graphes* regroupe des travaux autour de la fouille de données et l'apprentissage de et à partir de graphes, précédemment dispersés dans plusieurs axes. Les travaux sur ce thème sont durant la période de référence plus nombreux et ont contribué à renforcer la visibilité de l'équipe. Enfin, l'axe *Apprentissage de modèles topologiques à partir de données massives*, est issu de l'ancien axe trois et se concentre sur l'apprentissage non supervisé de données massives, et notamment pour des flux de données.

Activités de recherche

Les principaux travaux transversaux aux axes ainsi décrits portent sur des problématiques de recommandation et de clustering de graphes.

2.1.1 Apprentissage pour l'Aide à la Décision

Dans le domaine de l'aide à la décision, de très nombreux modèles ont été développés pour aider un décideur à faire un choix parmi plusieurs alternatives. Selon qu'il s'agisse de prendre une décision sous *incertitude* ou non, que la décision soit *séquentielle* ou non, qu'elle implique un seul individu ou que la décision soit *collective*, les types de modèles envisagés seront différents. Dans certains domaines, il peut être souhaitable de produire des modèles *interprétables* par le décideur, plutôt que des « boîtes noires ».

- **Apprentissage de modèles combinatoires interprétables**

Ainsi, dans le domaine médical, construire des modèles *interprétables* pour aider le médecin ou le chercheur à prendre des décisions est crucial. Nous nous sommes intéressés à la construction de modèles de prédiction du risque de développement de maladies cardio-vasculaires, à partir de données génomiques et métagénomiques. Dans ce type de problème, le nombre d'attributs (le nombre de gènes, de l'ordre de 10^8) est bien plus grand que le nombre d'exemples (le nombre de patients, de l'ordre de la centaine). Jusqu'à présent, la seule approche possible pour éviter le sur-apprentissage était d'utiliser des modèles linéaires parcimonieux, difficiles à interpréter par les médecins.

Pour apprendre des modèles plus facilement interprétables, nous avons introduit une contrainte supplémentaire dans ces modèles : l'algorithme d'apprentissage force les poids à ne prendre comme valeur que -1, 0, ou 1. Le problème d'apprentissage devient alors un problème combinatoire, difficile à résoudre.

> Nous avons montré dans [**CI-18**, CO-15] que ce problème n'admet pas de schéma d'approximation en temps polynomial et proposé un algorithme à garantie de performance pour résoudre ce problème. Nous avons également prouvé en utilisant la *complexité de Rademacher* que ces contraintes permettaient de dépasser la limitation classique des modèles parcimonieux, pour lesquels le nombre d'exemples doit être de l'ordre du logarithme du nombre d'attributs.

- **Apprentissage et Incertitude**

Nous nous sommes intéressés à l'apprentissage supervisé à partir de données incomplètes dans des langages logiques d'ordre 1. Un exemple imparfaitement connu est un ensemble de clauses instanciées, qui sont des représentations candidates de cet exemple. Nous avons développé une méthodologie prenant en compte au mieux

les connaissances disponibles sur de tels exemples, sous la forme de clauses exprimant des contraintes sur les valeurs que peuvent prendre simultanément les attributs des exemples, ainsi qu'un programme [LO-1] pour apprendre un prédicat cible même quand les exemples sont très incomplètement connus, à condition d'en fournir un grand nombre ainsi qu'un ensemble de clauses. Un résultat théorique de complexité d'échantillonnage donne les conditions dans lesquelles la convergence est garantie [CO-41].

- **Apprentissage et Planification**

La planification d'une séquence d'actions peut s'effectuer en utilisant diverses techniques, notamment les Processus de Décision de Markov (PDM) ou, lorsque les applications sont de nature plus combinatoire, les modèles STRIPS.

En se basant sur le formalisme des PDMs, nous avons traité le problème de l'apprentissage des préférences de l'utilisateur dans le cadre de la planification dans l'incertain. Planifier en prenant ces préférences en compte nécessite que l'agent apprenne à la fois les préférences de l'utilisateur et les effets de ses propres actions. Nous avons développé des algorithmes pour traiter des PDMs bien plus grands que ceux de l'état de l'art et de façon plus efficace pour résoudre ces problèmes [**CI-75**, CO-69, CI-60, **CO-70**].

Nous avons également exploré l'apprentissage de modèles d'action exprimés dans un formalisme relationnel (à la STRIPS) à partir de traces d'exécution d'un agent tentant de résoudre un but. Nous avons modélisé ce type d'apprentissage comme un processus de révision de théorie guidé par les données : l'agent déclenche des actions et met à jour son modèle d'actions pour rendre compte *a minima* des effets observés. Nous avons proposé une stratégie d'apprentissage actif [**CI-1**], *i.e.* favorisant l'application de certaines actions dans le but de généraliser plus rapidement ces règles d'action.

- **Apprentissage dans le cadre collectif**

Nous avons travaillé sur un modèle d'apprentissage collectif dans lequel, d'une part, la communauté forme un réseau d'agents dont les liens représentent les communications possibles, et d'autre part, les agents font des observations et communiquent en parallèle. Nous nous sommes intéressés à l'effet de la structure du réseau sur la vitesse d'apprentissage pour une tâche d'apprentissage supervisé [**CI-76***], ainsi qu'à un apprentissage pleinement parallèle pour lequel nous avons obtenu des garanties de consistance globale avec les observations [CI-77*].

Par ailleurs, dans le cadre d'une communauté sans contraintes de communication, nous avons exploré l'apprentissage collectif lorsque la tâche d'apprentissage est la révision d'un modèle d'action relationnel [**CO-30**, CI-27] et étudié différentes informations qu'une communauté d'agents peut apporter à un agent au moment de décider, par exemple, quelle action effectuer pour atteindre son but lorsque sa théorie ne lui permet pas de planifier [CI-63, **CL-5**].

2.1.2 Apprentissage non supervisé de représentations pour le transfert et la collaboration

- **Apprentissage non supervisé par transfert**

L'apprentissage par transfert consiste à utiliser un jeu de tâches pour influencer l'apprentissage et améliorer les performances sur une autre tâche. Cependant, ce paradigme d'apprentissage peut en réalité gêner les performances si les tâches (sources et cibles) sont trop dissemblables. Un défi est donc de développer des approches qui détectent et évitent le transfert négatif des connaissances utilisant très peu d'informations sur la tâche cible. Nos contributions dans cette direction s'articulent autour de deux dimensions complémentaires : théorique et pratique.

Factorisation matricielle non-négative pour l'adaptation de domaine non supervisée Nous avons développé une nouvelle méthode [RI-21, **CO-32**] pour l'adaptation de domaine non supervisée qui vise à aligner deux domaines (distributions de probabilités) en utilisant un ensemble commun de vecteurs de base dérivés de vecteurs propres de chaque domaine.

Alignement de noyaux pour l'apprentissage par transfert Nous avons proposé [CI-64] une approche d'apprentissage par transfert non supervisé qui minimise de manière itérative la distance entre les distributions de probabilités source et cible en optimisant l'alignement des noyaux calculés sur les jeux de données initiaux.

Bornes de généralisation pour l'adaptation de domaine à l'aide des plongements de Hilbert-Schmidt À partir des résultats pour l'adaptation de domaine comprenant les bornes de généralisation de Vapnik-Chervonenkis et celles issues de la théorie de l'apprentissage de Rademacher, nous avons montré comment on peut obtenir des bornes plus intéressantes en utilisant les plongements de Hilbert-Schmidt [**CI-28**].

Activités de recherche

Nous avons ainsi développé un cadre théorique basé sur les plongements Hilbert-Schmidt [CI-29] qui nous a permis d'améliorer les résultats théoriques de l'adaptation au domaine, en introduisant une mesure de distance naturelle et intuitive avec de fortes garanties de calcul pour son estimation. En utilisant l'alignement de noyaux, une contribution théorique et algorithmique pour le problème de « co-clustering » a été proposée dans [**CI-85**].

- **Apprentissage non supervisé collaboratif**

Le clustering collaboratif permet de préserver la confidentialité des données en utilisant d'autres résultats de classification non supervisée, sans avoir recours aux données de ces dernières. À partir d'une collection de bases de données distribuées sur plusieurs sites différents, le problème consiste à partitionner chacune de ces bases en considérant les données locales et les classifications distantes des autres bases collaboratrices, sans partage de données entre les différents centres. Le clustering collaboratif consiste à appliquer les algorithmes de clustering localement sur les différents sites, puis à faire collaborer les sites en partageant les résultats obtenus lors de la phase locale.

Clustering collaboratif basé sur un modèle génératif Nous avons proposé un clustering collaboratif basé sur les cartes topographiques génératives (GTM : Generative Topographic Mapping) dans [CI-16, CI-7]. Il s'agit d'un modèle génératif non linéaire défini de manière entièrement probabiliste. Cette approche a été validée sur des données artificielles et réelles en utilisant des critères internes et externes.

Clustering collaboratif flou des GTM variationnelles L'optimisation des paramètres du modèle GTM par l'agorithme Espérance-Maximisation (EM) ne tient pas compte de la complexité du modèle et, par conséquent, le risque de sur-apprentissage des données est élevé. Une solution élégante pour éviter le sur-apprentissage est d'approximer les GTM avec une vision variationnelle. Pour ce faire, nous avons proposé dans [RI-17] un algorithme qui combine VBGTM (Variational Bayesian Generative Topographic Mapping) et FCM (Fuzzy C-Means) pour effectuer la classification non supervisée et la visualisation des données en même temps.

Apprentissage non supervisé collaboratif entre algorithmes différents Nous avons aussi proposé deux méthodes permettant d'optimiser les liens de collaboration entre les différents algorithmes. La première [CI-42, CI-62, **RI-27**] propose une optimisation des liens de collaboration sous conditions de Karush-Kuhn-Tucker en maximisant la vraisemblance globale du système d'apprentissage ; la seconde [RI-12] se base sur des régressions linéaires effectuées à partir de simulations expérimentales et propose un modèle plus empirique.

Un framework pour l'apprentissage non supervisé collaboratif Nous avons enfin développé un framework [CI-40] permettant à des algorithmes différents de collaborer ensemble, indépendamment du type d'algorithme ou du nombre de clusters. Ce framework tient compte des caractéristiques locales des algorithmes pendant l'étape collaborative. En plus d'une étude sur la complexité et la convergence [CI-21], nous nous sommes intéressés au critère d'arrêt de notre algorithme et avons choisi un critère mesurant l'évolution de l'entropie globale du système d'apprentissage [**RI-27**]. La méthode a été testée sur plusieurs jeux de données réelles, en particulier pour la segmentation d'images satellite à très haute résolution dans le cadre du projet ANR COCLICO. Les résultats ont été validés avec plusieurs critères internes et externes et ont montré la solidité de notre approche ainsi que son efficacité.

- **Apprentissage non supervisé profond de représentations hiérarchiques**

La recherche internationale est actuellement très active autour de l'étude de méthodes d'apprentissage de représentations latentes par extraction de caractéristiques présentant des niveaux d'abstraction de plus en plus élevés. L'apprentissage profond (*Deep Learning*) s'est déjà montré très performant dans différents domaines dont la reconnaissance d'image, le traitement du langage et la robotique.

En utilisant le cadre théorique de [CL-7], nous avons proposé dans [**CO-51**] une première approche d'apprentissage profond basée sur la factorisation matricielle non-négative qui nous a permis d'analyser finement le comportement de ce type d'apprentissage durant la construction de nouvelles représentations hiérarchiques. Ce résultat peut être considéré comme une variante des approches de mise en commun largement utilisées dans les réseaux de neurones profonds. Nous avons observé que pour plusieurs ensembles de données choisis, la *parcimonie* diminue pendant la procédure d'optimisation à chaque couche. Ce résultat est plutôt surprenant car il était courant de supposer que l'amélioration de la *parcimonie* des prototypes rend la *Multilayer-NMF* plus robuste.

2.1.3 Apprentissage dans les Graphes

Cet axe rassemble les travaux autour de la conception d'algorithmes d'apprentissage dédiés aux graphes. Il regroupe des travaux théoriques, comme la formalisation d'opérateurs de fermeture dans les graphes, et plus appliqués, concernant la famille des réseaux attribués ou les réseaux d'interactions en biologie.

- **Opérateurs de Fermeture pour la Fouille de Graphes**

La fouille de motifs fréquents se base traditionnellement sur les notions d'opérateur de fermeture et de motif fermé. À l'aide de l'opérateur de fermeture, l'ensemble des motifs peut être partitionné en classes d'équivalences et les motifs fermés sont les représentants de chacune de ces classes. Ces représentants sont moins nombreux que l'ensemble des motifs fréquents et ils peuvent être énumérés efficacement. Nous avons étendu ces algorithmes aux graphes et à leurs extensions.

Nous nous sommes appuyés sur des travaux récents en fouille de données pour montrer, d'un point de vue théorique, comment préserver l'existence de l'opérateur de fermeture en affaiblissant la structure de treillis en une structure plus faible, appelée une pré-confluence [CI-23], qui est en particulier la structure de l'ensemble des sous-graphes (induits) connexes d'un graphe.

Nous avons généralisé la notion de motif fermé pour que le résultat obtenu respecte des contraintes naturelles dans les graphes [CI-45]; cela permet par exemple d'énumérer les motifs dits locaux, *i.e.* partagés par les sommets de sous-graphes connexes et ayant d'autres propriétés topologiques désirables.

Activités de recherche

- **Fouille de réseaux attribués**

Dans un réseau attribué, chaque sommet est pourvu d'une description sous la forme des valeurs prises par un ensemble d'attributs. En pratique, cette description est ramenée à un sous-ensemble d'*items*, et un motif particulier (un sous-ensemble d'*items*) est associé au sous-graphe induit par les sommets où ce motif est présent. Dans ce cadre, nous avons exploré les *abstractions de graphes* [**CI-22**]. Il s'agit d'une forme particulière de granularité obtenue en considérant la notion de noyau *(core)* d'un graphe. À un motif donné est d'abord associé le sous-graphe induit par le motif, puis ce sous-graphe est réduit en un *sous-graphe noyau* dont les sommets satisfont une certaine propriété topologique. Par exemple, dans le cas du noyau le plus classique, le *k-core*, les sommets du sous-graphe réduit doivent être de degré au moins k. Le motif fermé associé, dit *fermé abstrait*, est alors le plus spécifique porté par les sommets du sous-graphe noyau. Nous avons ainsi proposé une méthode générale d'énumération des motifs fermés abstraits associés à une notion de noyau.

Dans [CI-43, **CL-9***], on décompose de plus ces sous-graphes noyaux en communautés structurales. Dans le cas le plus simple, ces communautés sont les composantes connexes du noyau, mais on peut utiliser des formes plus contraintes, comme les k-communautés proposées par G. Palla et al. [2]. À chaque sous-graphe noyau associé à un motif donné, et à chacune de ses communautés structurales, est alors associé un motif *fermé local* : le motif le plus spécifique partagé par les sommets de la communauté. Nous avons développé et expérimenté un algorithme qui récolte dans un réseau attribué les motifs locaux fréquents et les implications locales associées, c'est-à-dire valides dans la communauté.

- **Analyse de graphes de terrain**

Nous avons poursuivi nos travaux sur l'analyse de grands graphes de terrain démarrée dès l'année 2007. Trois problématiques ont été abordées.

Détection de communautés locales. Nous avons proposé un algorithme efficace basé sur une approche gloutonne d'optimisation multi-objectif [**CI-35**, RI-15]. Pour le partitionnement des grands graphes en communautés, nous avons proposé des algorithmes dits *centrés-graine* où l'idée est de calculer un partitionnement d'un graphe en identifiant des communautés locales d'un ensemble de graines choisies avec des heuristiques adéquates [CI-34] : les graines sont les nœuds d'articulation dans le graphe [CI-33] ou sont des nœuds *leader* identifiés dans le graphe [RE-3].

La prévision de liens. Nous avons proposé une approche originale de prévision de liens fondée sur des techniques *supervisées* d'aggrégation de préférences [CI-4, CI-3], approche qui a été étendue aux réseaux hétérogènes [RI-13].

Analyse de graphes multiplexes. Un réseau multiplexe est un réseau multi-couches où chaque couche contient le même ensemble de nœuds reliés par différents ensembles de liens et les nœuds dans les différentes couches sont reliés entre eux par des liens dits de couplage. Ce nouveau modèle permet de prendre en compte, dans un formalisme unifié, différents cas intéressants issus d'applications réelles [RE-6].

Nous avons proposé une extension des travaux de [RE-3] au cas des réseaux multiplexes [RI-16]. Cet algorithme a été utilisé dans le cadre d'un système original de recommendation de tags dans les folksonomies [**CO-64**, CI-87]. Une biblio-

thèque d'analyse et de fouille de graphes multiplexes a été développée [**CO-55**].

- **Apprentissage de réseaux d'interactions**

D'une manière plus appliquée, nous travaillons depuis une dizaine d'années sur des méthodes d'apprentissage et de fouille de données pour la construction de réseaux d'interactions à partir de données génomiques. Nous avons proposé un modèle hybride de réseau de régulation, permettant de prédire de façon robuste l'expression d'un gène cible en fonction de l'expression de ses régulateurs, ainsi que l'algorithme d'apprentissage construisant ce modèle. Des corégulateurs candidats sont d'abord identifiés par une méthode de fouille de données dans des données discrètes, puis un ensemble de modèles de régression linéaire faisant intervenir ces corégulateurs est ensuite appris à partir de donnés perturbées [CI-19, **RI-11**]. L'algorithme a été validé, en collaboration avec l'Université d'Evry sur des données réelles de l'Institut Curie.

2.1.4 Apprentissage de modèles topologiques à partir de données massives

Avant de construire un système opérationnel à base d'apprentissage statistique, l'exploration des données représente une phase nécessaire, pendant laquelle les concepteurs d'un système analysent et visualisent les données. Ces données peuvent être multidimensionnelles, issues de sources différentes (bases de données, enquêtes, capteurs, télédétection, etc.). Elles peuvent comporter des composantes mixtes (numériques, binaires, catégorielles), structurées en blocs, présentant des dépendances. Parfois, elles sont imprécises et entachées d'erreurs. C'est pourquoi la phase d'exploration et de visualisation des données est essentielle. Cet axe rassemble à la fois les travaux théoriques de modélisation comme le développement de modèles topologiques évolutifs et des travaux plus appliqués concernant par exemple la formalisation des algorithmes en utilisant le paradigme MapReduce.

- **Apprentissage topologique à partir de flux de données**

Un flux de données est une séquence, potentiellement infinie, de données non stationnaires qui arrivent en continu [RI-24]. Ces flux imposent de nouvelles contraintes, en particulier sur la manière d'accéder aux données ; par exemple, accéder individuellement à chaque donnée dans un ordre aléatoire n'est plus envisageable et il est impossible de toutes les conserver en mémoire. L'algorithme Growing Neural Gas (GNG) [1] est une approche auto-organisatrice incrémentale de la même famille que les cartes auto-organisatrices de Kohonen (SOM). Nous décrivons ci-dessous trois de nos contributions qui étendent GNG au traitement des flux de données.

G-Stream [**RI-25**, CI-56] est une méthode « séquentielle » de clustering qui permet de découvrir de manière incrémentale des clusters de formes arbitraires en ne faisant qu'une seule passe sur les données. G-Stream utilise une fonction d'oubli afin de réduire l'impact des anciennes données dont la pertinence diminue au fil du temps.

batchStream est un algorithme de clustering de flux de données qui traite les données en micro-batch et qui passe à l'échelle [CI-55]. Il utilise une nouvelle formalisation de la méthode des nuées dynamiques en utilisant le modèle de traitement en micro-batch : la fonction de coût tient compte des sous-ensembles de données qui arrivent par paquets et une pondération est introduite pour pénaliser des données anciennes.

Dans le cadre du projet ***Square Predict***, nous avons validé l'algorithme batchStream sur des données d'assurance. Un modèle prédictif combinant le résultat du

clustering avec les arbres de décision a été aussi présenté [**RI-28**]. Le logiciel associé est disponible sur notre plateforme de démonstration [**LO-2**].

GH-Stream est un algorithme de visualisation et de clustering de flux de données massives. Cette méthode étend l'algorithme G-Stream en lui ajoutant une composante hiérarchique [**CI-58**] qui permet de décrire l'évolution des flux de données et d'analyser explicitement leur similitude.

- **Apprentissage non supervisé massif**

Activités de recherche

Toujours dans le cadre de l'apprentissage sur des données massives, nous avons proposé une décomposition d'un algorithme de clustering et biclustering topologique en fonctions élémentaires en utilisant le paradigme MapReduce [3] [**CO-29**, CI-25, CI-24]. Les logiciels associés sont disponibles sur notre plateforme de démonstration [**LO-2**].

Nous avons développé une approche originale combinant cartes topologiques et un algorithme de classification hiérarchique biomimétique qui nous a permis d'obtenir des résultats sur le résumé de graphes et la classification hiérarchique incrémentale [RI-7].

Plus précisément en unifiant deux modèles d'apprentissage, celui des cartes topologiques et celui de l'algorithme de classification hiérarchique bioinspirée AntTree, nous avons pu classifier des données simultanément, d'un point de vue topologique et hiérarchique, et ainsi obtenir une partition des données organisées sur une grille 2D avec une organisation hiérarchique sous forme d'arbre au niveau de chaque cellule de la grille. Nous avons proposé l'extension de ce formalisme à la classification et la visualisation de données structurées sous forme de graphe.

En assimilant chaque donnée à un nœud d'un graphe, notre approche permet aussi de décomposer un graphe en une succession de sous-graphes [CI-8, CI-10, **CI-17**]. Nous avons développé une version incrémentale du modèle hiérarchique où la topologie est libre et la relation de voisinage entre les clusters évolue de manière dynamique en s'inspirant du modèle Neural Gas [**CI-58**].

- **Apprentissage par modèles de mélanges à partir de données séquentielles**

Nous avons proposé une approche d'apprentissage statistique non supervisé pour modéliser la dynamique d'un ensemble de séquences. En combinant les points forts des cartes topologiques et des modèles de Markov cachés, nous avons défini des modèles de Markov cachés topologiques et auto-organisés, qui s'adaptent à la nature, la structure et la dynamique des séquences. Ces modèles généralisent les chaînes de Markov en introduisant une relation spatio-séquentielle entre les états cachés [RI-14, CI-68]. Ces travaux s'inscrivaient dans le cadre d'un projet CIFRE avec l'INA.

2.2 Analyse AFOM/SWOT

- **Forces et Atouts**

  — *Large spectre scientifique de l'équipe*, lui assurant une bonne visibilité extérieure sur ses thèmes de recherches ;
  — *Attractivité du thème et de l'équipe dans le monde académique et industriel* : nombreux doctorants, bon taux d'insertion académique et industriel à l'issue de la thèse, nombreux partenariats académiques (ANR) et industriels (FUI, PIA, CIFRE) ;
  — *Formation à la recherche* : adossement à un master très visible (EID^2) ;
  — *Nombreuses initiatives d'animation de la recherche* : organisation de conférences françaises, d'écoles thématiques, de tutoriels, de sessions spéciales dans des conférences internationales ;
  — *Valorisation de la recherche* : transfert technologique par création de start-ups.

- **Faiblesses**
  - *Absence de chercheurs permanents* : la politique de recrutement de chercheurs CNRS doit être dynamisée, notamment à travers le recrutement de postdocs et un recours plus fréquent à des mois de chercheurs invités ;
  - *Relations internationales à intensifier*, notamment au niveau des contrats internationaux, même s'il y a progression par rapport à la période précédente ;
  - *Diffusion de logiciels à renforcer* : des logiciels sont développés dans l'équipe, mais leur diffusion est encore trop restreinte, faute de personnel dédié au développement. L'équipe a fait une demande d'ingénieur CNRS en ce sens.

- **Opportunités**
  - *Renforcement de notre position à l'international :* les liens avec d'autres équipes et chercheurs à l'international s'intensifient, ce qui se concrétise par davantage de collaborations, de professeurs invités, de workshops et tutoriels organisés à l'étranger, et sans doute à terme de projets internationaux ;
  - *Visibilité accrue de l'apprentissage (et de l'Intelligence Artificielle en général)* dans le grand public, ce qui peut susciter des vocations d'étudiants.
  - *Large couverture des thèmes de l'apprentissage*, ce qui est original dans le paysage français d'où la possibilité de recherches transversales en apprentissage ;
  - *Insertion dans la COMUE* : fortes collaborations notamment avec le LIPADE, sur des thèmes déjà existants dans l'équipe.

- **Menaces**
  - Les départs possibles de membres HDR pourraient déstabiliser certains thèmes de recherche, majoritairement animés par des MCF ;
  - *Surcharge administrative* : les enseignants-chercheurs de l'équipe assument des responsabilités pédagogiques/administratives lourdes (vice-président UP13 chargé de la Transformation Numérique, directeur du CREIP et responsable de département à l'IUTV, directrice adjointe de l'IG chargée de l'École d'Ingénieurs).
  - *Malgré l'adossement à un M2 très visible*, les entreprises attirent de plus en plus de bons étudiants qui se détournent des carrières universitaires sur le domaine de l'apprentissage artificiel et de la fouille de données.

Références complémentaires

[1] B. Fritzke. A growing neural gas network learns topologies. In *7th International Conference on Neural Information Processing Systems (NIPS'94), Denver, Colorado*, pages 625–632. MIT Press, 1994.

[2] G. Palla, I. Derényi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435 :814–818, 2005.

[3] M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauley, M. J. Franklin, S. Shenker, and I. Stoica. Resilient distributed datasets : a fault-tolerant abstraction for in-memory cluster computing. In *9th USENIX conference on Networked Systems Design and Implementation (NSDI'12), San Jose, CA, USA*, pages 2–2. USENIX Association, 2012.

2.3 Perspectives et projet

Notre projet est de contribuer significativement sur des thèmes de recherche pour lesquels notre expertise est reconnue. Suite au départ d'un PU, et à l'évolution des questions scientifiques autour de l'apprentissage, le projet présenté est resserré autour de deux axes, ce qui permettra d'animer plus efficacement l'équipe. Un premier axe de recherche abordera des problèmes théoriques pour l'apprentissage de représentations, dans la continuité des résultats obtenus de l'axe *Apprentissage non supervisé de représentations pour le transfert et la collaboration* présenté dans le bilan. Un

deuxième axe s'intéressera à l'apprentissage de et à partir de structures (principalement des graphes) d'une part, et à la conception d'algorithmes d'apprentissage non supervisés distribués pour traiter des données massives et le plus souvent dynamiques. Dans chacun de ces axes, des problèmes fondamentaux sont étudiés en lien avec des domaines d'applications privilégiés : bioinformatique, données bibliographiques et de réseaux sociaux, images, ... L'équipe renforcera ses collaborations avec les équipes du laboratoire : RCLN, notamment à travers un FUI qui débute, AOC sur le développement de méthodes hybrides Apprentissage - Optimisation Combinatoire, CALIN sur la thématique des approches quantiques pour l'apprentissage profond. Des collaborations nationales et internationales sont mentionnées dans chacun des thèmes.

Activités de recherche

2.3.1 Vers de nouveaux cadres théoriques pour l'apprentissage de représentations

- **Apprentissage non supervisé dans le cadre de la théorie du transport optimal**

Le transport optimal (TO) définit un ensemble d'outils géométriques pour comparer les mesures de probabilité prises sur un espace métrique. Ces outils utilisent des techniques d'optimisation et de calcul différentiel. Ils ont trouvé de multiples applications dans des domaines tels que l'économie, l'imagerie et la mécanique des fluides pour modéliser les transferts, les morphings et les déplacements. La géométrie du TO, connue sous différents noms, p.e. Wasserstein ou Monge-Kantorovich, a des propriétés qui la rendent particulièrement adaptée pour résoudre des problèmes d'apprentissage artificiel, pour lesquels la capacité de comparer des mesures de probabilité empirique en grande dimension est souvent cruciale. En effet, contrairement aux divergences d'information (Kullback-Leiber, Jensen-Shannon, Hellinger, etc.), qui ne sont définies que pour des mesures absolument continues, la distance TO est bien définie pour des mesures discrètes (p.e. empiriques) et continues. Malgré l'intérêt de cette théorie, les applications pratiques du TO en apprentissage artificiel sont en retard par rapport aux réalisations théoriques. Cette situation peut être essentiellement attribuable au coût élevé généralement associé au calcul du TO.

Cette direction de recherche vise à établir un pont entre les problèmes de TO et le fondement de la science des données. Nous cherchons à développer de nouveaux outils méthodologiques pour l'analyse et la représentation de données accessibles sous la forme de distributions empiriques ainsi qu'à revisiter plusieurs cadres classiques de l'apprentissage statistique (apprentissage profond, factorisation matricielle, ...) avec le point de vue du TO. Nous proposons d'étudier deux aspects liés à l'apprentissage non supervisé :

1. Le premier concerne le clustering de données distributionnelles, pour lequel nous proposons d'approfondir le concept pour mettre au point des approches originales d'apprentissage adaptées à la classification non supervisée et à la visualisation de données distributionnelles. Pour ce type de données, chaque observation est représentée par un ensemble de distributions de valeurs numériques, par exemple sous la forme d'histogrammes ou de fonction de probabilités. Nous étudierons ainsi des distances adaptées dans le cadre du TO, telle que la distance de Wasserstein, qui sont nécessaires à l'exploration de ce type de données.
2. Le deuxième aspect porte sur le co-clustering vu comme un problème de transport entre la mesure empirique définie sur les observations d'une matrice de données et la mesure empirique définie sur les variables de cette même matrice. Ainsi nous proposons de développer des algorithmes de co-clustering qui reposent sur la théorie du TO en s'intéressant à la solution sous la forme d'une matrice de couplage qui peut être vue comme une approximation de la densité jointe des données entre exemples et caractéristiques.

Les membres impliqués dans ce travail sont Younès Bennani , Guénael Cabanes, et Basarab Matei en collaboration avec notamment Ievgen Redko (INSA Lyon), Rosanna Verde et Antonio Irpino (Universita della Campania à « Luigi Vanvitellià », Italie).

- **Apprentissage profond et approches quantiques**

Bien que l'informatique quantique et l'apprentissage artificiel puissent sembler a priori très différents, ils ont déjà interagi de multiples façons par le passé. L'apprentissage quantique est le domaine à l'intersection entre l'informatique quantique et l'apprentissage automatique. Il s'agit d'une discipline émergente présentant de nombreuses perspectives pertinentes pour la science.

Par ailleurs, ces dernières années, est apparu dans la littérature scientifique une panoplie d'articles traitant de l'apprentissage profond (*Deep Learning*), traduisant un regain d'activité de recherche en intelligence artificielle, notamment autour des réseaux de neurones profonds qui avaient été éclipsés par les machines à noyaux. Ce regain s'explique en grande partie par leur capacité à obtenir d'excellentes performances pour des tâches de classification et de réduction de dimension, et par leur capacité à extraire dans leurs couches cachées successives des informations de plus en plus haut niveau sur l'entrée du système. Les applications en image sont impressionnantes avec l'extraction de primitives visuellement très pertinentes et appropriées pour les types d'images utilisées en apprentissage.

La recherche internationale est actuellement très active autour de l'étude de méthodes d'apprentissage permettant de mieux contrôler la construction, dans les couches cachées successives, de représentations internes par extraction de caractéristiques présentant des niveaux d'abstraction de plus en plus élevés. Des algorithmes pour les ordinateurs quantiques ont également été proposés pour résoudre efficacement certains problèmes qui ne peuvent être traités par des ordinateurs conventionnels classiques. Ces propositions montrent que le calcul quantique non seulement réduit le temps nécessaire à la conception d'un système d'apprentissage artificiel, mais fournit également un cadre plus riche et plus complet pour l'apprentissage profond que dans le cas de l'informatique classique et conduit à des améliorations significatives dans l'optimisation.

Nous envisageons d'étudier deux aspects liées à l'apprentissage non supervisé :

— Nous proposons de développer une première direction concernant la quantisation d'algorithmes d'apprentissage non-supervisé, où certaines parties d'algorithmes classiques sont remplacées par des sous-routines quantiques, afin d'obtenir une accélération du temps de calcul ou encore de réduire le coût de communication dans le cas d'une situation d'apprentissage distribué.
— Une deuxième direction que nous souhaitons explorer consiste à définir l'analogue de l'apprentissage artificiel dans un monde où l'information et l'ensemble des données sont quantiques, ce qui influence directement le processus d'apprentissage et ses limites. Ainsi, en reformulant certains problèmes de la théorie quantique comme étant des tâches d'apprentissage, on peut utiliser des notions d'apprentissage artificiel classique, pour résoudre ces problèmes.

Les membres impliqués dans ce travail sont Younès Bennani, Guénaël Cabanes, Basarab Matei et Jérémie Sublime (ISEP Paris), en collaboration avec Joseph Ben Geloun et Gérard Duchamp de l'équipe CALIN. Par ailleurs, nous souhaitons appliquer ces principes de l'apprentissage profond quantique aux problématiques de la sécurité de grandes masses de données notamment dans le cadre de l'Open data. Les membres impliqués dans ce travail sont Younès Bennani et Nistor Grozavu en collaboration, notamment, avec Nicoleta Rogovschi (Paris Descartes) et Abdelouahid Lyahyaoui (ENSA Tanger).

2.3.2 Apprentissage et structures

- **Apprentissage dans les graphes enrichis et hétérogènes**

Les données structurées sous forme de graphes sont des objets devenus prédominants ces dernières années, en particulier en sciences de la vie (réseaux biologiques) et sur internet (réseaux sociaux en ligne), auxquels s'ajoutent les réseaux sociaux à proprement parler, objets d'étude de la sociologie. L'étude de ces objets s'est longtemps limitée à des questions topologiques, mesurant des indices et modélisant ces réseaux pour identifier leurs caractéristiques, leur structure, en extraire des communautés. Depuis quelques années, on s'intéresse à des réseaux enrichis, réseaux attribués dans lesquels les sommets ou les liens sont étiquetés par des descriptions, réseaux multiplexes lorsque les liens qui relient les sommets sont de différentes natures, réseaux multimodaux lorsque les sommets sont de différents types, réseaux orientés lorsque les liens sont des arcs, réseaux géographiques dont les sommets sont localisés sur un espace physique, ces différents enrichissements pouvant être de plus combinés en un réseau hétérogène. Ceci conduit d'une part à techniquement adapter et étendre les méthodes existantes spécifiques aux réseaux, d'autre part à importer les techniques provenant de l'apprentissage et de la fouille de données, en considérant que le réseau est une structure des données que l'on examine. Nous allons de ce fait vers la possibilité d'exploiter de la manière la plus large la nature relationnelle des données.

Activités de recherche

Nous avons récemment développé différentes approches méthodologiques concernant l'apprentissage et la fouille de données sur les réseaux complexes/sociaux multiplexes ou attribués. En particulier, nous nous sommes intéressés à la détection de communautés, à l'énumération et à la sélection de motifs portant sur les attributs, via la réduction de graphes à des graphes dits *abstraits*, et à l'apprentissage de réseaux biologiques. Dans cette partie de l'axe apprentissage et structure, nos prochains travaux s'orientent comme suit :

Méthodes pour les réseaux enrichis et hétérogènes : Nous étudierons, formellement et d'un point vue algorithmique et expérimental, l'extension des méthodes de détection de communauté et d'énumération de motifs à ces différents types de réseaux enrichis. Chaque forme de réseau suppose une manière particulière de l'aborder, que ce soit pour l'analyser, le construire, le réviser, en extraire des communautés ou des sous-réseaux *abstraits*. En particulier nous nous intéresserons aux types et aux rôles des différents éléments constituant un réseau hétérogène, avec l'idée de travailler sur des motifs multiformes, chaque composante du motif contraignant un type d'élément dans le réseau.

Analyse des réseaux multiplexes : Nous prévoyons de développer les travaux les plus récents entamés pendant la période de référence autour de l'analyse des réseaux multiplexes. Le lien avec les travaux de fouille de graphes attribués mentionnés ci-dessus sera tout particulièrement exploré.

Caractérisation des réseaux, choix d'abstraction, sélection de motifs inattendus : Nous travaillerons également sur les questions suivantes : comment évaluer, sélectionner en fonction de leur intérêt, les communautés, motifs, sous-graphes extraits ? Quels types d'abstraction, avec quels paramètres, sont adéquats pour analyser un réseau donné ? Peut-on considérer un ensemble de réseaux et y trouver une structure, peut-on les classifier selon leurs similarités ? En changeant d'échelle, on change ainsi la nature des connaissances à extraire.

Un premier résultat a été obtenu sur la définition d'une forme d'abstraction de graphe dédiée aux réseaux dirigés. Elle a permis de chercher des motifs caractérisant les échanges de conseils dans un grand cabinet d'avocats, les motifs étant pour cela

sélectionnés par une mesure d'intérêt comparant la taille des sous-réseaux abstraits à leur taille attendue [CI-79*].

Réseaux biologiques : Dans le cadre du projet CHIST-ERA Adalab, nous avons abordé la révision de réseaux d'interactions orientés, hétérogènes (régulation et métabolique), dynamiques et incertains. Nous avons obtenu de premiers résultats sur la révision de réseaux de régulation, modélisés comme des réseaux bayésiens dynamiques, à partir de données temporelles d'expression de gènes [CO-75*]. Nous poursuivrons nos travaux sur la révision de réseaux plus complexes (hétérogènes) à partir de données (très) partiellement observées. Ceci requiert la conception et le développement de méthodes de révision de réseaux efficaces et robustes. Nous nous orientons vers la révision de réseaux bayésiens dynamiques, afin de pouvoir adapter des méthodes d'inférence bayésiennes capables de calculer les états des nœuds du réseau non observés à partir de données (très) incomplètes.

Dans un contexte de sciences expérimentales, de nouvelles données (résultats d'expériences) arrivent incrémentalement, la révision de modèle doit donc se faire en ligne. L'apprentissage actif aborde la question de sélectionner parmi toutes les expériences possibles celles dont les résultats permettraient de produire des révisions pertinentes (*experimental design*). Les données observées étant clairement non indépendamment et identiquement distribuées, de nombreuses questions théoriques se posent notamment pour évaluer l'intérêt d'une nouvelle observation pour évaluer un modèle candidat. Ces interrogations, autant théoriques que pratiques, dépassent le cadre de l'apprentissage de réseaux biologiques et peuvent également s'instancier, par exemple, dans le domaine des objets connectés.

Les membres impliqués sur ce sujet sont D. Bouthinon, R. Kanawati, A. Osmani, C. Rouveirol, G. Santini et H. Soldano. Des collaborations sur ce thème ont commencé avec une thèse co-dirigée avec l'équipe RCLN dans le cadre d'un projet FUI, avec M. Atzmüller (Tilburg, Pays-Bas) sur la fouille de réseaux hétérogènes, et avec E. Lazega (Sciences Po, Paris) sur l'analyse de réseaux sociaux. Sur la question de l'apprentissage de réseaux biologiques, nous poursuivrons une collaboration ancienne avec M. Elati (Université de Lille) et les collaborations nationales et internationales (Univ. Manchester et Brunel, UK) établies au sein du contrat Adalab peuvent se prolonger autour de l'étude de problèmes liés à l'automatisation/reproductibilité des processus de découverte scientifique, sujet suscitant un intérêt grandissant dans la communauté bioinformatique.

- **Les nouveaux défis du passage à l'échelle de l'apprentissage statistique**

Depuis une dizaine d'années, le volume des données disponibles s'est considérablement accru et la nature de ces données s'est enrichie (binaires, catégorielles, mixtes, séquentielles, fonctionnelles, *etc*). Pour traiter ces données massives, la programmation parallèle ou distribuée est souvent présentée comme la solution rapide et pragmatique. Il existe des paradigmes maintenant bien établis (*Map-Reduce*, *Actor*, *etc*) qui permettent de travailler à un niveau conceptuel et de s'affranchir des aspects d'ingénierie pour se concentrer sur les aspects théoriques de l'apprentissage statistique dans un domaine souvent pluridisciplinaire (*la statistique computationnelle*). Notre axe de recherche a pour ambition de développer de nouveaux algorithmes de *clustering* et de visualisation passant à l'échelle, à base de modèles non paramétriques, ainsi que des modèles de régression. Nous développerons en particulier les thèmes ci-dessous.

Algorithme *Mean-shift* et modèles topologiques : Ce thème se situe dans le cadre des travaux sur l'utilisation des modèles non paramétriques comme le *Mean-shift* pour le *clustering* de données multi-dimensionnelles, binaires et mixtes. L'approche *Mean-shift* forme des *clusters* qui correspondent aux régions les plus denses dans l'espace des données. Nous souhaitons introduire une structure topologique entre les *clusters*

pour enrichir ce modèle et faire le lien avec les modèles de mélanges. Nous envisageons également d'exploiter cette structure topologique pour accélérer la montée du gradient qui permet d'associer à chaque donnée un *cluster* en suivant une intuition similaire à celles des modèles topologiques. Des collaborations autour du calcul massivement distribué et de l'apprentissage statistique (*la statistique computationnelle*) ont été initiées dans le cadre d'un projet de recherche avec Christophe Cérin (équipe AOC) et Jean-Luc Gaudiot (University of California, Irvine and 2017 IEEE Computer Society President). Un des objectifs visés au travers de ces collaborations est de proposer un espace de représentation des données qui facilitera l'adaptation de l'algorithme du *Mean-shift* pour des architectures matérielles hétérogènes.

Activités de recherche

Visualisation analytique : Une autre aspect sur lequel nous voulons travailler est la visualisation de grandes masses de données. De nombreux travaux ont été menés sur l'extraction de paramètres pertinents à visualiser à l'aide du clustering. Il s'agit d'un problème important puisque la quantité d'information ne cesse de croître. Ainsi, les techniques de visualisation doivent être repensées pour manipuler des données gigantesques. Un des objectifs est de proposer de nouvelles approches qui permettent d'utiliser l'information visuelle sur une zone de densité à l'aide des algorithmes décrits précédemment afin de visualiser des données continues, catégorielles et mixtes. Nous envisageons de travailler sur l'adaptation des modèles de mélanges et des KDE (Kernel Density Estimation) pour extraire des paramètres à visualiser à grande échelle [1].

Régression typologique (Clusterwise) : Nous souhaitons également étudier finement la combinaison entre les méthodes supervisées et non supervisées pour la classification, plus précisément le clusterwise ou la régression typologique. L'idée principale du clusterwise est que pour chaque classe on ajuste un modèle linéaire, de manière à minimiser la somme des carrés des résidus. Une première piste de recherche dans le cadre de données massives a été lancée en collaboration avec Mme Ndeye Niang membre de l'équipe MSDMA (Méthodes Statistiques de Data-Mining et Apprentissage) du CEDRIC au CNAM, et Stéphanie Bougeard de l'ANSES [1].

Nouveaux algorithmes d'apprentissage distribués Nous avons enfin initié des travaux et collaborations prometteurs qui déboucheront sur la conception de nouveaux algorithmes d'apprentissage distribués. Nous nous intéresserons à l'apprentissage relationnel distribué d'une part, en nous appuyant sur de premiers résultats obtenus dans une thèse en cours [CI-84*] concernant la distribution d'une étape cruciale en apprentissage relationnel, le test de subsomption entre une hypothèse et un exemple. D'autre part, en collaboration avec l'équipe AOC, nous étudierons une approche mixte apprentissage et optimisation combinatoire pour la résolution de problèmes d'allocation dynamique de ressources.

Les membres impliqués dans ces travaux seront H. Azzag, D. Bouthinon, S. Guérif, M. Lebbah et C. Rouveirol.

Références

[1] A. Gramacki and J. Gramacki. FFT-based fast computation of multivariate kernel density estimators with unconstrained bandwidth matrices. *Journal of Computational and Graphical Statistics*, 26(2) :459–462, 2017.

1. Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail

Chapitre 3
Produits de la recherche

Sommaire

3.1 Synthèse des publications 34
3.2 Publications 34
Articles dans des revues avec comité de lecture et chapitres de livres (ACL) 35
Conférences invitées (INV) 38
Communications avec actes (ACT) 39
Logiciels et Brevets (LO) 50
Thèses et Habilitations (TH) 51
3.3 Publications antérieures à l'arrivée au LIPN 52
Articles dans des revues avec comité de lecture et chapitres de livres (ACL) 53
3.4 Réalisations informatiques (logiciels), Brevets 53
3.4.1 Logiciels 53
3.4.2 Start-up 53

3.1 Synthèse des publications

	2012	2013	2014	2015	2016	2017	*Total*
Livres	0	0	0	0	0	0	*0*
Chapitres de livre	0	2	2	0	4	2	*10*
Revues internationales	6	3	2	7	8	4	*30*
Autres revues	1	1	1	4	3	0	*10*
Conférences internationales	13	7	19	21	15	13	*88*
Autres communications	17	8	17	15	13	5	*75*
Total	*37*	*21*	*41*	*47*	*43*	*24*	*213*

3.2 Publications

Les références

- comportant « * » sont acceptées et à paraître.
- indiquées sur fond gris sont les productions phare de chaque catégorie.
- en gras constituent les 20% les meilleures de la production scientifique de chaque catégorie.

Signification des soulignements :

- *simple* : enseignant-chercheur ou chercheur titulaire,
- *double* : doctorant,
- *pointillé* : post-doc, ingénieur ou autre situation.

Articles dans des revues avec comité de lecture et chapitres de livres (ACL)

Articles dans des revues internationales avec comité de lecture sélectif

2017 [**RI-30**] Y. Chevaleyre, U. Endriss et N. Maudet. Distributed fair allocation of indivisible goods. *Artificial Intelligence*, 242 :1–22, 2017.

[RI-29*] A. Cornuejols, C. Wemmert, P. Gançarski et Y. Bennani. Collaborative clustering : Why, when, what and how. *Information Fusion*, 39 :81–95, 2017.

[**RI-28**] M. Ghesmoune, H. Azzag, S. Benbernou, M. Lebbah, T. Duong et M. Ouziri. Big data : From collection to visualization. *Machine Learning. Special issue "Discovery Science"*, 106(6) :837–862, 2017. doi :10.1007/s10994-016-5622-4, http ://rdcu.be/ourr.

[**RI-27**] J. Sublime, B. Matei, G. Cabanes, N. Grozavu, Y. Bennani et A. Cornuéjols. Entropy based probabilistic collaborative clustering. *Pattern Recognition*, 72 :144 – 157, 2017.

2016 [RI-26] T. Duong, G. Beck, H. Azzag et M. Lebbah. Nearest neighbour estimators of density derivatives, with application to mean shift clustering. *Pattern Recognition Letters*, 80 :224–230, 2016.

[**RI-25**] M. Ghesmoune, M. Lebbah et H. Azzag. A new growing neural gas for clustering data streams. *Neural Networks*, 78 :36–50, 2016.

[RI-24] M. Ghesmoune, M. Lebbah et H. Azzag. State-of-the-art on clustering data streams. *Big Data Analytics*, 1, 2016. #13 - 27 pages.

[RI-23] B. Matei. Exact reconstruction of a class of nonnegative measures using model sets. *Asymptotic Analysis*, (97) :291–299, 2016.

[RI-22] B. Matei et S. Meignen. Nonlinear and nonseparable bidimensional multiscales representation based on cell-average representation. *IEEE Transactions on Image Processing*, (17) :4570–4580, 2016.

[RI-21] I. Redko et Y. Bennani. Non-negative embedding for fully unsupervised domain adaptation. *Pattern Recognition Letters*, 77 :35–41, 2016.

[RI-20] I. Ronai, B. P. Oldroyd, D. A. Barton, G. Cabanes, J. Lim et V. Vergoz. Anarchy Is a Molecular Signature of Worker Sterility in the Honey Bee. *Molecular Biology and Evolution*, 33(1) :134–142, 2016.

[RI-19] M.-A. Torres-Arredondo, J. Sierra-Pérez et G. Cabanes. An optimal baseline selection methodology for data-driven damage detection and temperature compensation in acousto-ultrasonics. *Smart Materials and Structures*, 25(5) :055034 – 15 pages, 2016.

2015 [RI-18] G. Cabanes, E. van Wilgenburg, M. Beekman et T. Latty. Ants build transportation networks that optimize cost and efficiency at the expense of robustness. *Behavioral Ecology*, 26(1) :223–231, 2015.

[RI-17] M. Ghassany et Y. Bennani. Collaborative fuzzy clustering of variational bayesian generative topographic mapping. *International Journal of Computational Intelligence and Applications*, 14(1) :1550001 – 16 pages, 2015.

[RI-16] M. Hmimida et R. Kanawati. Community detection in multiplex networks : a seed-centric approach. *Networks and Heterogeneous Media*, 10(1) :71–85, 2015.

[RI-15] R. Kanawati. Empirical evaluation of applying ensemble methods to ego-centered community identification in complex networks. *Neurocomputing*, 150, B :417–427, 2015.

[RI-14] M. Lebbah, R. Jaziri, Y. Bennani et J.-H. Chenot. Probabilistic self-organizing map for clustering and visualizing non-i.i.d data. *International Journal of Computational Intelligence and Applications*, 14(2) :1550007 – 29 pages, 2015.

[RI-13] M. Pujari et R. Kanawati. Link prediction in multiplex networks. *Networks and Heterogeneous Media*, 10(1) :17–35, 2015.

[RI-12] J. Sublime, N. Grozavu, G. Cabanes, Y. Bennani et A. Cornuéjols. From horizontal to vertical collaborative clustering using generative topographic maps. *International Journal of Hybrid Intelligent Systems*, 12(4) :245–256, 2015.

2014 [**RI-11**] I. Chebil, R. Nicolle, G. Santini, C. Rouveirol et M. Elati. Hybrid method inference for the construction of cooperative regulatory network in human. *IEEE Transactions on NanoBioscience*, 13(2) :97–103, 2014.

[RI-10] B. Roche, H. Broutin, M. Choisy, S. Godreuil, C. G. de Magny, Y. Chevaleyre, J.-D. Zucker, R. Breban, B. Cazelles et F. Simard. The niche reduction approach : an opportunity for optimal control of infectious diseases in low-income countries. *BMC Public Health*, 14 :753–761, 2014.

2013 [**RI-9**] G. Cabanes, Y. Bennani, R. Destenay et A. Hardy. A new topological clustering algorithm for interval data. *Pattern Recognition*, 46(11) :3030–3039, 2013.

[RI-8] A. Chaibi, M. Lebbah et H. Azzag. Group outlier factor : a new score using self-organising map for group-outlier and novelty detection. *International Journal of Computational Intelligence and Applications*, 12(2), 2013. 18 pages.

[RI-7] N.-Q. Doan, H. Azzag et M. Lebbah. Growing self-organizing trees for autonomous hierarchical clustering. *Neural Networks - Special Issue on Autonomous Learning*, 41 :85–95, 2013.

2012 [RI-6] H. Azzag, C. Guinot et G. Venturini. An artificial ants model for fast construction and approximation of proximity graphs. *Adaptive Behaviour*, 20(6) :443–459, 2012.

[RI-5] G. Cabanes, Y. Bennani et D. Fresneau. Enriched topological learning for cluster detection and visualization. *Neural Networks*, 32 :186–195, 2012.

[RI-4] Y. Chevaleyre, J. Lang, N. Maudet, J. Monnot et L. Xia. New candidates welcome ! possible winners with respect to the addition of new candidates. *Mathematical Social Sciences*, 64(1) :74–88, 2012.

[RI-3] M. Ghassany, N. Grozavu et Y. Bennani. Collaborative clustering using prototype-based techniques. *International Journal of Computational Intelligence and Applications*, 11(3) :17 pages, 2012.

[**RI-2**] G. Santini, H. Soldano et J. Pothier. Automatic classification of protein structures relying on similarities between alignments. *BMC Bioinformatics*, 13(1), 2012. #233 (13 pages).

[RI-1] V. Vergoz, J. Lim, M. Duncan, G. Cabanes et B. P. Oldroyd. Effects of natural mating and CO2 narcosis on biogenic amine receptor gene expression in the ovaries and brain of queen honey bees, Apis mellifera. *Insect Molecular Biology*, 21(6) :558–567, Dec 2012.

Articles dans des revues nationales ou internationales

2016 [RE-10] R. Kanawati. Détection de communautés dans les grands graphes d'interactions (multiplexes) : état de l'art. *Apprentissage Articiel et Fouille de Données (AAFD 2016), Villetaneuse, France*, volume RNTI-A-8 de *Revue des Nouvelles Technologies de l'Information*, pages 67–100, 2016.

[RE-9] N. Masmoudi, H. Azzag, M. Lebbah, C. Bertelle et M. B. Jemaa. Clustering numerical data using artificial ants. *International Journal of Computer Science and Information Security*, 14(8) :1219–1225, 2016.

[RE-8] H. Soldano, G. Santini et D. Bouthinon. Motifs abstraits et sous-communautés dans les réseaux attribués. *Revue d'Intelligence Artificielle*, 29(4) :441–468, 2016.

2015 [RE-7] K. Benabdeslem, C. Biernacki et M. Lebbah. Les trois défis du big data - eléments de réflexion. *Statistique et Société*, 3(1), 2015. 4 pages.

[RE-6] R. Kanawati. Multiplex network mining : a brief survey. *IEEE Intelligent Informatics Bulletin*, 16 :24–28, 2015.

[RE-5] J. Sierra-Perez, M. A. Torres-Arredondo, G. Cabanes, A. Geme et L. E. Mujica. Structural health monitoring by means of strain field pattern recognition on the basis of PCA and automatic clustering techniques based on SOM. *IFAC-PapersOnLine*, 48(28) :987–992, 2015.

[**RE-4**] H. Soldano. Motifs fermés et abstraction : au delà des treillis. *Revue d'Intelligence Artificielle*, 29(3-4) :321–348, 2015.

2014 [RE-3] Z. Yakoubi et R. Kanawati. Licod : Leader-driven approaches for community detection. *Vietnam Journal of Computer Science*, 1(4) :241–256, 2014.

2013 [RE-2] M. Pujari et R. Kanawati. Link prediction in multiplex bibliographical networks. *International Journal of Complex Systems in Science*, 2 :77–82, 2013.

2012 [RE-1] T. N. A. Nguyen, Y. Chevaleyre et J.-D. Zucker. Optimizing sign placements for crowd evacuation on road network in case of tsunami alert. *International Journal of Computer Science and Artificial Intelligence*, 2(4) :33–39, 2012.

Chapitres de livre

2017 [CL-10] M. Pujari et R. Kanawati. Link prediction in large-scale multiplex networks. S. Cordier, N. Debarsy, C. Ertur, F. Nemo, D. Nourrit-Lucas, G. Poisson et C. Vrain, éditeurs, *Understanding Interactions in complex systems*, Chapitre 7, pages 107–136. Cambridge Scholar Publishing, 2017.

[**CL-9***] H. Soldano, G. Santini et D. Bouthinon. Formal concept analysis of attributed networks. R. Missaoui, S. Obiedkov et S. Kuznetsov, éditeurs, *Formal Concept Analysis of Social Networks*, Lecture Notes in Social Network. Springer, 2017.

2016 [CL-8] S. Bouveret, Y. Chevaleyre et N. Maudet. Fair allocation of indivisible goods. F. Brandt, V. Conitzer, U. Endriss, J. Lang et A. D. Procaccia, éditeurs, *Handbook of Computational Social Choice*, Chapitre 12. Cambridge University Press, 2016. 36 pages.

[CL-7] B. Matei. Model sets and new versions of Shannon theorem. A. Aldroubi et al., éditeur, *New Trends in Harmonical Analysis*, pages 215–279. Birkhauser, 2016.

[CL-6] M. Pujari et R. Kanawati. Link prediction in complex networks. N. Meghanathan, éditeur, *Advanced Methods for Complex Network Analysis*, Chapitre 3, pages 58–97. IGI Global, 2016.

[**CL-5**] C. Rodrigues, H. Soldano, G. Bourgne et C. Rouveirol. Collaborative online learning of an action model. *Solving Large Scale Learning Tasks. Challenges and Algorithms*, volume 9580 de *Lecture Notes in Computer Science*, pages 300–319. Springer, 2016.

2014 [CL-4] M. Hmimida et R. Kanawati. Système de recommandation par niveau dans la classification à facettes pour la recherche documentaire. G. Kembellec, G. Chartron et I. Saleh, éditeurs, *Les moteurs et systèmes de recommandation*, Chapitre 6, pages 127–140. Hermès, 2014.

[CL-3] M. Hmimida et R. Kanawati. A two-level recommendation approach for document search. G. Kembellec, G. Chartron et I. Saleh, éditeurs, *Recommander Systems*, Chapitre 6, pages 119–134. Wiley, 2014.

2013 [CL-2] L. Amgoud, Y. Chevaleyre et N. Maudet. Systèmes multi-agents : interaction, négociation, persuasion. P. Marquis, éditeur, *Panorama de l'intelligence artificielle*, pages 273–296. Cepadues, 2013.

[CL-1] H. Azzag et M. Lebbah. A new way for hierarchical and topological clustering. F. Guillet, B. Pinaud, G. Venturini et D. A. Zighed, éditeurs, *Advances in Knowledge Discovery and Management*, volume 471 de *Studies in Computational Intelligence*, pages 85–97. Springer, 2013.

Conférences invitées (INV)

Communications invitées dans une conférence internationale

2014 [IN-1] R. Kanawati. Seed-centric algorithms for high performance community detection in complex networks. *Analysing and Mining Large Graphs and Networks : Algorithms and Performance - AMLagap Workshop, Orléans, France*, 2014.

Tutoriaux dans des conférences internationales

2017 [TU-5] N. Grozavu et R. Kanawati. Topological & graph-based clustering : Recent algorithmic advances. *2017 International Joint Conference on Neural Networks (IJCNN 2017), Anchorage, Alaska, USA*, page 156, 2017.

[TU-4*] R. Kanawati et M. Atzmueller. Mining attributed networks (tutorial). *4th IEEE International conference on Data Science and Advanced Analytics (DSAA 2017), Tokyo, Japon*, 2017.

2016 [TU-3] R. Kanawati et N. Grozavu. Topological & graph-based clustering : Recent algorithmic advances : A tutorial. *23th International Conference On Neural Information Processing (ICONIP 2016), Kyoto, Japan*, page 120, 2016.

2015 [TU-2] R. Kanawati. Multiplex network mining (a tutorial). *5th International Conference on Web Intelligence, Mining and Semantics, Singapore*, 2015.

2014 [TU-1] Y. Chevaleyre. Preference learning problems. *European Conference on Machine Learning 2014 (ECML 2014), Nancy, france*, 2014.

Communications avec actes (ACT)

Communications dans des conférences internationales avec comité de lecture

2017 [**CI-88**] S. Bouveret, Y. Chevaleyre, F. Durand et J. Lang. Voting by sequential elimination with few voters. *Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI 2017), Melbourne, Australia*, pages 128–134, 2017.

[CI-87] M. Hmimida et R. Kanawati. A graph-based meta-approach for tag recommendation. H. Cherifi, S. Gaito, W. Quattrociocchi et A. Sala, éditeurs, *Complex Networks & Their Applications V (2016), Milano, Italy*, volume 693 de *Studies in Computational Intelligence*, pages 309–320. Springer, 2017.

[CI-86] A. Lachaud, N. Grozavu, B. Matei et Y. Bennani. Collaborative clustering between different topological partitions. *2017 International Joint Conference on Neural Networks (IJCNN 2017), Anchorage, Alaska, USA*, pages 4111–4117, 2017.

[**CI-85**] C. Laclau, I. Redko, B. Matei, Y. Bennani et V. Brault. Co-clustering through optimal transport. *2017 International Conference on Machine Learning (ICML 2017), Sidney, Australia*, volume 70 de *Proceedings of Machine Learning Research (PMLR)*, pages 1955–1964, 2017.

[CI-84*] H. Léger, D. Bouthinon, M. Lebbah et H. Azzag. An instance based model for scalable theta-subsumption. *29th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2017), Boston, MA, USA*, 2017.

[CI-83*] A. Osmani, M. Hamidi et A. Chibani. Machine learning approach for infant cry interpretation. *29th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2017), Boston, MA, USA*, 2017.

[**CI-82**] P. Rastin, B. Matei, G. Cabanes et I. El Baghdadi. Signal-based autonomous clustering for relational data. *2017 International Joint Conference on Neural Networks (IJCNN 2017), Anchorage, Alaska, USA*, pages 4103–4110, 2017.

[CI-81] N. Rogovschi, J. Kitazono, N. Grozavu, T. Omori et S. Ozawa. t-distributed stochastic neighbor embedding spectral clustering. *2017 International Joint Conference on Neural Networks (IJCNN 2017), Anchorage, Alaska, USA*, pages 1628–1632, 2017.

[**CI-80**] N. Sokolovska, Y. Chevaleyre et J.-D. Zucker. The fused lasso penalty for learning interpretable medical scoring systems. *2017 International Joint Conference on Neural Networks (IJCNN 2017), Anchorage, Alaska, USA*, pages 4504–4511, 2017.

[CI-79*] H. Soldano, G. Santini, D. Bouthinon et E. Lazega. Hub-authority cores and attributed directed network mining. *IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI 2017), Boston, MA, USA*. IEEE Computer Society, 2017.

[**CI-78**] J. Sublime, B. Matei et P. A. Murena. Analysis of the influence of diversity in collaborative and multi-view clustering. *2017 International Joint Conference on Neural Networks (IJCNN 2017), Anchorage, Alaska, USA*, pages 4126–4133, 2017.

[CI-77*] L.-M. Veillon, G. Bourgne et H. Soldano. Effect of network topology on neighbourhood-aided collective learning. *9th International Conference on Computational Collective Intelligence (ICCCI), Nicosia, Cyprus*, Lecture Notes in Artificial Intelligence. Springer, 2017.

[**CI-76***] L.-M. Veillon, G. Bourgne et H. Soldano. Waves : a model of collective learning. *2017 IEEE/WIC/ACM International Conference on Web Intelligence (WI'17), Leipzig, Germany*. ACM, 2017.

2016 [**CI-75**] P. Alizadeh, Y. Chevaleyre et F. Lévy. Advantage based value iteration for Markov decision processes with unknown rewards. *2016 International Joint Conference on Neural Networks (IJCNN 2016), Vancouver, Canada*, pages 3837–3844, 2016.

[**CI-74**] G. Beck, T. Duong, H. Azzag et M. Lebbah. Distributed mean shift clustering with approximate nearest neighbours. *2016 International Joint Conference on Neural Networks (IJCNN 2016), Vancouver, Canada*, pages 3110–3115, 2016.

[CI-73] H. Chahdi, N. Grozavu, I. Mougenot, Y. Bennani et L. Berti-Equille. Towards ontology reasoning for topological cluster labeling. *23rd International Conference On Neural Information Processing (ICONIP 2016), Kyoto, Japan*, volume 9949 de *Lecture Notes in Computer Science*, pages 156–164, 2016.

[CI-72] H. Chahdi, N. Grozavu, I. Mougenot, L. Berti-Equille et Y. Bennani. On the use of ontology as a priori knowledge into constrained clustering. *3rd IEEE International Conference on Data Science and Advanced Analytics (DSAA 2016), Montreal, Canada*, 2016.

[CI-71] S. Cherki, P. Rastin, G. Cabanes et B. Matei. Improved sparse prototyping for relational k-means. *2016 IEEE Symposium Series on Computational Intelligence (SSCI 2016), Athens, Greece*, pages 1–8, 2016.

[CI-70] N. Grozavu, G. Cabanes, H. Chahdi et N. Rogovschi. Automated topological co-clustering using fuzzy features partition. *2016 International Joint Conference on Neural Networks (IJCNN 2016), Vancouver, Canada*, pages 2638–2645, 2016.

[CI-69] E. Janvier, T. Couronne et N. Grozavu. Features learning and transformation based on deep autoencoders. *23rd International Conference On Neural Information Processing (ICONIP 2016), Kyoto, Japan*, volume 9949 de *Lecture Notes in Computer Science*, pages 111–118, 2016.

[CI-68] R. Jaziri, F. Chamroukhi, M. Lebbah et Y. Bennani. GTM mixture through time for sequential data. *2016 International Joint Conference on Neural Networks (IJCNN 2016), Vancouver, Canada*, pages 2805–2810, 2016.

[CI-67] J. Kitazono, N. Grozavu, N. Rogovschi, T. Omori et S. Ozawa. t-distributed stochastic neighbor embedding with inhomogeneous degrees of freedom. *23rd International Conference On Neural Information Processing (ICONIP 2016), Kyoto, Japan*, volume 9949 de *Lecture Notes in Computer Science*, pages 119–128, 2016.

[CI-66] N. Masmoudi, H. Azzag, M. Lebbah, C. Bertelle et M. B. Jemaa. CL-AntInc algorithm for clustering binary data streams using the ants behavior. *20th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2016), York, UK*, Procedia Computer Science, pages 187–196, 2016.

[CI-65] P. Rastin, T. Zhang et G. Cabanes. A new clustering algorithm for dynamic data. *23rd International Conference On Neural Information Processing (ICONIP 2016), Kyoto, Japan*, volume 9949 de *Lecture Notes in Computer Science*, pages 175–182, 2016.

[CI-64] I. Redko et Y. Bennani. Kernel alignment for unsupervised transfer learning. *23rd International Conference on Pattern Recognition (ICPR 2016), Cancún, Mexico*, pages 525–530, 2016.

[CI-63] C. Rodrigues, H. Soldano, G. Bourgne et C. Rouveirol. Collaborative decision in multi-agent learning of action models. *28th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2016), San Jose, California, USA*, pages 640–647, 2016.

[CI-62] J. Sublime, A. Cornuéjols et Y. Bennani. Collaborative-based multi-scale clustering in very high resolution satellite images. *23rd International Conference On Neural Information Processing (ICONIP 2016), Kyoto, Japan*, volume 9949 de *Lecture Notes in Computer Science*, pages 148–155, 2016.

[CI-61] B. C. Thai, A. Mokraoui et B. Matei. Performance evaluation of high dynamic range image tone mapping operators based on separable non-linear multiresolution families. *24th European Signal Processing Conference (EUSIPCO 2016), Budapest, Hungary*, pages 1891–1895, 2016.

2015 [CI-60] P. Alizadeh, Y. Chevaleyre et J.-D. Zucker. Approximate regret based elicitation in markov decision process. *11th IEEE International Conference on Computing and Communication Technologies (ICCCT), Chennai, India*, 2015. 6 pages.

[**CI-59**] A. Damamme, A. Beynier, Y. Chevaleyre et N. Maudet. The power of swap deals in distributed resource allocation. *2015 International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Istanbul, Turkey*, pages 625–633, 2015.

[**CI-58**] N.-Q. Doan, M. Ghesmoune, H. Azzag et M. Lebbah. Growing hierarchical trees for data stream clustering and visualization. *2015 International Joint Conference on Neural Networks (IJCNN 2015), Killarney, Ireland*. IEEE, 2015. 8 pages.

[CI-57] I. Falih, N. Grozavu, R. Kanawati et Y. Bennani. A recommendation system based on unsupervised topological learning. *22nd International Conference On Neural Information Processing (ICONIP 2015), Istambul, Turkey*, volume 9490 de *Lecture Notes in Computer Science*, pages 224–232, 2015.

[CI-56] M. Ghesmoune, M. Lebbah et H. Azzag. Clustering over data streams based on growing neural gas. *19th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining - (PAKDD 2015), Ho Chi Min, Vientnam*, volume 9078 de *Lecture Notes in Computer Science*, pages 134–145. Springer, 2015.

[CI-55] M. Ghesmoune, M. Lebbah et H. Azzag. Micro-batching growing neural gas for clustering data streams using spark streaming. *INNS Conference on Big Data 2015 Program, San Francisco, California, USA*, volume 53 de *Procedia Computer Science*, pages 158–166, 2015.

[CI-54] N. Grozavu, N. Rogovschi, G. Cabanes, A. Troya-Galvis et P. Gançarski. VHR satellite image segmentation based on topological unsupervised learning. *14th IAPR International Conference on Machine Vision Applications (MVA 2015), Tokyo, Japan*, pages 543–546, 2015.

[CI-53] M. Kaaniche, B. Matei et S. Meignen. Optimized lifting schemes based on eno stencils for image approximation. *2015 IEEE International Conference on Image Processing (ICIP 2015), Québec, Canada*, pages 4308–4312, 2015.

[CI-52] R. Kanawati. Ensemble selection for community detection in complex networks. G. Meiselwitz, éditeur, *7th international conference on Social Computing and Social Media (SCSM 2015), Los Angeles, California, USA*, volume 9182 de *Lecture Notes in Computer Science*, pages 138–147. Springer, 2015.

[CI-51] N. Masmoudi, H. Azzag, M. Lebbah, C. Bertelle et M. B. Jemaa. Clustering of binary data sets using artificial ants algorithm. *22nd International Conference On Neural Information Processing (ICONIP 2015), Istambul, Turkey*, volume 9489 de *Lecture Notes in Computer Science*, pages 716–723. Springer, 2015.

[CI-50] N. Masmoudi, H. Azzag, M. Lebbah, C. Bertelle et M. B. Jemaa. How to use ants for data stream clustering. *IEEE Congress on Evolutionary Computation (CEC 2015), Sendai, Japan*, pages 656–663, 2015.

[CI-49] B. Matei. Exact reconstruction of a class of nonnegative measures using model sets. *Proceedings of 11th IEEE International Conference on Sampling Theory and Applications, Wahington DC, USA*, pages 578–582, 2015.

[CI-48] P. Rastin, G. Cabanes, N. Grozavu et Y. Bennani. Collaborative clustering : How to select the optimal collaborators ? *2015 IEEE Symposium Series on Computational Intelligence, Capetown, South Africa*, pages 787–794, 2015.

[**CI-47**] I. Redko et Y. Bennani. Sparsity analysis of learned factors in multilayer NMF. *2015 International Joint Conference on Neural Networks (IJCNN 2015), Killarney, Ireland*, 2015. 7 pages.

[CI-46] N. Rogovschi, N. Grozavu et L. Labiod. Spectral clustering through topological learning for large datasets. *22nd International Conference On Neural Information Processing (ICONIP 2015), Istambul, Turkey*, volume 9490 de *Lecture Notes in Computer Science*, pages 216–223, 2015.

[**CI-45**] H. Soldano. Extensional confluences and local closure operators. *13th International Conference on Formal Concept Analysis (ICFCA 2015), Nerja, Spain*, volume 9113 de *Lecture Notes in Computer Science*, pages 128–144. Springer, 2015.

[CI-44] H. Soldano, G. Santini et D. Bouthinon. Abstract and local rule learning in attributed networks. F. Esposito, M.-S. Hacid, O. Pivert et Z. Ras, éditeurs, *22nd International Symposium on Foundations of Intelligent Systems (ISMIS 2015), Lyon, France*, volume 9384 de *Lecture Notes in Computer Science*, pages 313–323. Springer, 2015.

[CI-43] H. Soldano, G. Santini et D. Bouthinon. Local knowledge discovery in attributed graphs. A. Esposito, éditeur, *27th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2015), Vietri sul Mare, Italy*, pages 250–257. IEEE Computer Society, 2015.

[CI-42] J. Sublime, N. Grozavu, Y. Bennani et A. Cornuéjols. Collaborative clustering with heterogeneous algorithms. *2015 International Joint Conference on Neural Networks (IJCNN 2015), Killarney, Ireland*, 2015. 8 pages.

[CI-41] J. Sublime, N. Grozavu, Y. Bennani et A. Cornuéjols. Vertical collaborative clustering using generative topographic maps. *7th International Conference of Soft Computing and Pattern Recognition (SoCPaR 2015), Fukuoka, Japan*, pages 199–204, 2015.

[CI-40] J. Sublime, A. Troya-Galvis, Y. Bennani, A. Cornuéjols et P. Gançarski. Semantic rich ICM algorithm for VHR satellite images segmentation. *14th IAPR International Conference on Machine Vision Applications (MVA 2015), Tokyo, Japan*, pages 45–48, 2015.

2014

[CI-39] D. Bouthinon et H. Soldano. Learning first order rules from ambiguous examples. G. A. Papadopoulos, éditeur, *International Conference on Tools with Artificial Intelligence (ICTAI 2014), Limassol, Cyprus*, pages 39–46. IEEE, 2014.

[**CI-38**] N. Grozavu, G. Cabanes et Y. Bennani. Diversity analysis in collaborative clustering. *2014 International Joint Conference on Neural Networks (IJCNN 2014), Beijing, China*, pages 1754–1761, 2014.

[CI-37] R. Kanawati. Community detection in social networks : the power of ensemble methods. *ACM/IEEE International Conference on Data Science & Advanced Analytics (DSAA 2014), Shangai, China*, pages 46–52. IEEE, 2014.

[CI-36] R. Kanawati. Empirical evaluation of applying ensemble ranking to ego-centered communities identification in complex networks. *4th International Conference on Multimedia Computing and Systems (ICMCS), Marrakech, Morocco*, pages 536–541. IEEE, 2014.

[**CI-35**] R. Kanawati. Multi-objective approach for local community computation. *26th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'2014), Limassol, Cyprus*, pages 367–373. IEEE, 2014.

[CI-34] R. Kanawati. Seed-centric approaches for community detection in complex networks. G. Meiselwitz, éditeur, *6th international conference on Social Computing and Social Media (SCSM 2014), Heraklion, Greece*, volume 8531 de *Lecture Notes in Computer Science*, pages 197–208. Springer, 2014.

[CI-33] R. Kanawati. YASCA : An ensemble-based approach for community detection in complex networks. *20th International Conference on Computing and Combinatorics (COCOON 2014), Altanta, Georgia, USA*, volume 8591 de *Lecture Notes in Computer Science*, pages 657–666, 2014.

[CI-32] V.-M. Le, Y. Chevaleyre, J.-D. Zucker et T. V. Ho. Approaches to optimize local evacuation maps for helping evacuation in case of tsunami. *First International Conference on Information Systems for Crisis Response And Management in Mediterranean countries (ISCRAM-med 2014), Toulouse, France*, volume 196 de *Lecture Notes in Business Information Proceedings*, pages 21–31. Springer, 2014.

[CI-31] N. Masmoudi, H. Azzag, M. Lebbah et C. Bertelle. Incremental clustering of data stream using real ants behavior. *Sixth World Congress on Nature and Biologically Inspired Computing (NaBIC 2014), Porto, Portugal*, pages 262–268, 2014.

[CI-30] I. Redko et Y. Bennani. Controlling orthogonality constraints for better NMF clustering. *2014 International Joint Conference on Neural Networks (IJCNN 2014), Beijing, China*, pages 3894–3900, 2014.

[CI-29] I. Redko et Y. Bennani. Non-negative matrix factorization with schatten p-norms reguralization. *21st International Conference On Neural Information Processing (ICONIP 2014), Kuching, Malaysia*, volume 8835 de *Lecture Notes in Computer Science*, pages 52–59, 2014.

[**CI-28**] I. Redko et Y. Bennani. Random subspaces nmf for unsupervised transfer learning. *2014 International Joint Conference on Neural Networks (IJCNN 2014), Beijing, China*, pages 3901–3908, 2014.

[CI-27] C. Rodrigues, H. Soldano, G. Bourgne et C. Rouveirol. Multi agent learning of relational action models. T. Schaub, G. Friedrich et B. O'Sullivan, éditeurs, *21st European Conference on Artificial Intelligence (ECAI 2014), Prague, Czech Republic*, volume 263 de *Frontiers in Artificial Intelligence and Applications*, pages 1087–1088. IOS Press, 2014.

[CI-26] N. Rogovschi et N. Grozavu. Opinion retrieval through unsupervised topological learning. *2014 International Joint Conference on Neural Networks (IJCNN 2014), Beijing, China*, pages 3130–3134, 2014.

[CI-25] T. Sarazin, M. Lebbah et H. Azzag. Biclustering using Spark-MapReduce. *2014 IEEE International Conference on Big Data (Big Data 2014), Washington DC, USA*, pages 58–60, 2014.

[CI-24] T. Sarazin, M. Lebbah, H. Azzag et A. Chaibi. Feature group weighting and topological biclustering. *21st International Conference On Neural Information Processing (ICONIP 2014), Kuching, Malaysia*, volume 8835 de *Lecture Notes in Computer Science*, pages 369–376. Springer, 2014.

[CI-23] H. Soldano. Closed patterns and abstraction beyond lattices. Springer, éditeur, *12th International Conference on Formal Concept Analysis (ICFCA 2014), Kluj-Napoca, Romania*, volume 8478 de *Lecture Notes in Computer Science*, pages 203–218, 2014.

[**CI-22**] H. Soldano et G. Santini. Graph abstraction for closed pattern mining in attributed network. T. Schaub, G. Friedrich et B. O'Sullivan, éditeurs, *21st European Conference on Artificial Intelligence (ECAI 2014), Prague, Czech Republic*, volume 263 de *Frontiers in Artificial Intelligence and Applications*, pages 849–854. IOS Press, 2014.

[CI-21] J. Sublime, A. Cornuéjols et Y. Bennani. A new energy model for the hidden markov random fields. *21st International Conference On Neural Information Processing (ICONIP 2014), Kuching, Malaysia*, volume 8835 de *Lecture Notes in Computer Science*, pages 60–67, 2014.

2013

[CI-20] I. Chebil, M. Elati, C. Rouveirol et G. Santini. SetNet : Ensemble method techniques for learning regulatory networks. *12th International Conference on Machine Learning and Applications (ICMLA 2013), Miami, Floride, USA*, volume 1, pages 34–39, 2013.

[CI-19] I. Chebil, R. Nicolle, G. Santini, C. Rouveirol et M. Elati. Hybrid method inference for the construction of cooperative regulatory network in human. *2013 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2013), Shanghai, China*, pages 121–126, 2013.

[**CI-18**] Y. Chevaleyre, F. Koriche et J.-D. Zucker. Rounding methods for discrete linear classification. *30th International Conference on Machine Learning (ICML 2013), Atlanta, Georgia, USA*, volume 28 de *Proceedings of Machine Learning Research (PMLR)*, pages 651–659, 2013.

[**CI-17**] N.-Q. Doan, H. Azzag, M. Lebbah et G. Santini. Self-organizing trees for visualizing protein dataset. *2013 International Joint Conference on Neural Networks (IJCNN 2013), Dallas, Texas, USA*, 2013. 8 pages.

[CI-16] M. Ghassany, N. Grozavu et Y. Bennani. Collaborative multi-view clustering. *2013 International Joint Conference on Neural Networks (IJCNN 2013), Dallas, Texas, USA*, 2013. 8 pages.

[CI-15] V. Nikulin, N. Rogovschi et N. Grozavu. Incremental learning from several different microarrays. *2013 International Joint Conference on Neural Networks (IJCNN 2013), Dallas, Texas, USA*, 2013. 7 pages.

[CI-14] Z. Yakoubi et R. Kanawati. Leader-driven approach for community detection in complex network. *International Conference on Interactions in Complex Systems, Orléans, France*, 2013. 12 pages.

2012

[CI-13] G. Cabanes et Y. Bennani. Change detection in data streams through unsupervised learning. *2012 International Joint Conference on Neural Networks (IJCNN 2012), Brisbane, Australia*, pages 1–6, 2012.

[CI-12] A. Chaibi, H. Azzag et M. Lebbah. Automatic group-outlier detection. *20th European Symposium on Artificial Neural Networks (ESANN 2012), Bruges, Belgium*, 2012. 6 pages.

[CI-11] Y. Chevaleyre et A. Machado Pamponet. Adaptive probabilistic policy reuse. *19th International Conference On Neural Information Processing (ICONIP 2012), Doha, Qatar*, volume 7665 de *Lecture Notes in Computer Science*, pages 603–611. Springer, 2012.

[CI-10] N.-Q. Doan, H. Azzag et M. Lebbah. Graph decomposition using self-organizing trees. *16th International Conference on Information Visualisation (IV 2012), Montpellier, France*, pages 246–251, 2012.

[**CI-9**] N.-Q. Doan, H. Azzag et M. Lebbah. Growing self-organizing trees for knowledge discovery from data. *2012 International Joint Conference on Neural Networks (IJCNN 2012), Brisbane, Australia*, pages 1–8, 2012.

[CI-8] N.-Q. Doan, H. Azzag et M. Lebbah. Self-organizing map and tree topology for graph summarization. *22nd International Conference on Artificial Neural Networks and Machine Learning (ICANN 2012), Lausanne, Suisse*, volume 7553 de *Lecture Notes in Computer Science*, pages 363–370, 2012.

[CI-7] M. Ghassany, N. Grozavu et Y. Bennani. Collaborative generative topographic mapping. *19th International Conference On Neural Information Processing (ICONIP 2012), Doha, Qatar*, volume 7664 de *Lecture Notes in Computer Science*, pages 591–598, 2012.

[**CI-6**] N. Grozavu, Y. Bennani et L. Labiod. Feature space transformation for transfer learning. *2012 International Joint Conference on Neural Networks (IJCNN 2012), Brisbane, Australia*, pages 2637–2642. IEEE, 2012.

[CI-5] K. Mouhoubi, L. Létocart et C. Rouveirol. A knowledge-driven bi-clustering method for mining noisy datasets. *19th International Conference On Neural Information Processing (ICONIP 2012), Doha, Qatar*, volume 7665 de *Lecture Notes in Computer Science*, pages 585–593. Springer, 2012.

[CI-4] M. Pujari et R. Kanawati. Link prediction in complex networks by supervised rank aggregation. *Proceedings of the 24th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2012), Athens, Greece*, pages 782–789. IEEE, 2012.

[CI-3] M. Pujari et R. Kanawati. Supervised rank aggregation approach for link prediction in complex networks. *21st World Wide Web Conference (WWW 2012) - MSND'12 workshop - Companion Volume, Lyon, France*, pages 1189–1196, 2012.

[CI-2] M. Pujari et R. Kanawati. Tag recommendation by link prediction based on supervised machine learning. *Sixth International AAAI Conference on Weblogs and Social Media (ICWSM'2012), Dublin, Ireland*, pages 547–550, 2012.

[**CI-1**] C. Rodrigues, P. Gérard, C. Rouveirol et H. Soldano. Active learning of relational action models. S. Muggleton, A. Tamaddoni-Nezhad et F. Lisi, éditeurs, *21st International Conference on Inductive Logic Programming (ILP 2011), Windsor Great Park, UK*, volume 7207 de *Lecture Notes in Computer Science*, pages 302–316. Springer, 2012.

Autres articles dans des conférences ou workshops à comité de lecture

2017 [CO-75*] A. Coutant et C. Rouveirol. Network inference of dynamic models by the combination of spanning arborescences. *Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM 2017), Lille, France*, 2017.

[CO-74] I. Falih, N. Grozavu, R. Kanawati et Y. Bennani. Multiplex network clustering based collaborative filtering. *61th World Statistics Conference (WSC 2017), Marrakech, Morocco*, 2017.

[CO-73] H. Léger, D. Bouthinon, M. Lebbah et H. Azzag. Nouveau modèle pour un passage à l'échelle de la 0-subsomption. *Extraction et Gestion des Connaissances (EGC 2017), Grenoble, France*, volume RNTI-E-33, pages 339–344. Hermann, 2017.

[**CO-72***] A. Osmani, M. Hamidi et A. Chibani. Platform for assessment and monitoring of infant comfort. *2017 AAAI Fall Symposium Series*, 2017.

[CO-71*] M. Rifi, M. Hibti et R. Kanawati. Exploring network metrics for accident scenarios : a case of study of the uncontrolled level drop. *2017 International Topical Meeting on Probabilistic Safety Assessment and Analysis (PSA 2017), Pittsburgh, Pennsylvanie, USA*, 2017. #6.

2016 [**CO-70**] P. Alizadeh, Y. Chevaleyre et F. Lévy. Solving MDPs with unknown rewards using nondominated vector-valued functions. *The 8th European Starting AI Researcher Symposium (STAIRS 2016), The Hague, The Netherlands*, pages 15–26, 2016.

[CO-69] P. Alizadeh, Y. Chevaleyre et F. Lévy. Itération de la valeur basée sur des avantages pour des processus de décision de Markov avec récompenses inconnues. *Conférence Conjointe Francophone sur la Science des Données (AAFD et SFC'16), Marrakech, Morocco*, 2016. 8 pages.

[CO-68] S. Benbernou, M. Bentounsi, P. Bourdoncle, M. Lebbah, M. Ouziri et S. Sahri. Towards Big Data in Medical Imaging. Symposium IDV - Imageries du Vivant, 2016. (poster).

[CO-67] H. Chahdi, N. Grozavu, I. Mougenot, L. Berti-Equille et Y. Bennani. Génération de contraintes pour le clustering à partir d'une ontologie - application à la classification d'images satellites. *16ème Journées Francophones Extraction et Gestion des Connaissances (EGC 2016), Reims, France*, pages 81–92, 2016.

[CO-66] I. Falih, N. Grozavu, R. Kanawati et Y. Bennani. Attributed network clustering based recommendation approach. *Actes de la 7ème Conférence sur les modèles et l'analyse des réseaux : Approches mathématiques et informatiques (MARAMI/JFGG 2016), Cergy, France*, 2016. 6 pages.

[CO-65] I. Falih, N. Grozavu, R. Kanawati et Y. Bennani. Clustering dans les graphes attribués. *Conférence Conjointe Francophone sur la Science des Données (AAFD & SFC 2016), Marrakech, Morocco*, 2016.

[**CO-64**] M. Hmimida et R. Kanawati. A graph-coarsening approach for tag recommendation. *25th International Conference Companion on World Wide Web (WWW 2016) - Companion Volume, Montréal, Canada*, pages 43–44, 2016. (poster).

[**CO-63**] M. Hmimida et R. Kanawati. A multiplex network based tag recommendation approach. *2nd European Conference on Social networks, Paris, France*, 2016. 2 pages.

[CO-62] M. Hmimida et R. Kanawati. Une approche topologique pour la recommandation de tags dans les folksonomies. *Conférence Conjointe Francophone sur la Science des Données (AAFD & SFC 2016), Marrakech, Marocco*, 2016.

[**CO-61**] H. Leger, D. Bouthinon, M. Lebbah et H. Azzag. A new model for scalable θ-sumbsumption. *International Conference on Inductive Logic Programming (ILP 2016), London, UK*, volume 1865 de *CEUR Workshop Proceedings*, pages 41–47, 2016. 12 pages.

[CO-60] S. Sellami, I. Falih, T. Auvray et R. Kanawati. Analyse des réseaux d'interaction entre entreprises cotées en bourses. *Actes de la 7ème Conférence sur les modèles et l'analyse des réseaux : Approches mathématiques et informatiques (MARAMI/JFGG 2016), Cergy, France*, 2016. 2 pages.

[**CO-59**] N. Sokolovska, Y. Chevaleyre et J.-D. Zucker. Interpretable score learning by fused lasso and integer linear programming. *From Multiple Criteria Decision Aid to Preference Learning (DA2PL), Paris, France*. EURO, 2016. 8 pages.

[CO-58] H. Soldano, G. Santini et D. Bouthinon. Efficient pattern mining in attributed graphs. G. Melançon, éditeur, *Modèles et Analyse des Réseaux : Approches Mathématiques et Informatiques (MARAMI 2016), Cergy-Pontoise, France*, pages 1–12, 2016.

2015 [CO-57] V. S. Dao, Y. Chevaleyre et J.-D. Zucker. Terda : An interpretable rule learner for high-dimensional data. *Bioinformatics and Genomic Medicine, Ho Chi Minh, Viet Nam*, 2015. 1 page.

[CO-56] I. Falih, M. Hmimida et R. Kanawati. Une approche centrée graine pour la détection de communautés dans les réseaux multiplexes. J. Darmont, éditeur, *Conférence francophone sur l'extraction et la gestion des connissances (EGC 2015), Luxembourg*, pages 377–382, 2015.

[**CO-55**] I. Falih et R. Kanawati. Muna : A multiplex network analysis library. *DemoPapers@ASONAM2015 - The 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Paris, France*, pages 757–760, 2015.

[CO-54] M. Ghesmoune, M. Lebbah et H. Azzag. Clustering topologique pour le flux de données. *Extraction et Gestion des Connaissances (EGC 2015), Luxembourg*, volume RNTI-E-28 de *Revue des Nouvelles Technologies de l'Information*, pages 137–142. Hermann, 2015.

[CO-53] R. Kanawati. Ensemble selection for enhancing graph coarsening quality. *5th international workshop on Social network analysis, Capri, Italy*, 2015. 10 pages.

[CO-52] V.-M. Le, Y. Chevaleyre, J.-D. Zucker et T. V. Ho. Hybrid of linear programming and genetic algorithm for optimizing agent-based simulation. application to optimization of sign placement for tsunami evacuation. *11th IEEE International Conference on Computing and Communication Technologies (ICCCT 2015), Chennai, India*, 2015. 6 pages.

[**CO-51**] B. Matei et Y. Bennani. Pruned simple model sets for fast exact recovery of image. *IEEE International Conference on Data Mining Workshop (ICDMW 2015), Atlantic City, New Jersey, USA*, pages 863–870, 2015.

[CO-50] P. Rastin et R. Kanawati. A multiplex-network based approach for clustering ensemble selection. *MANEM@ASONAM 15, First International workshop on Multiplex and Attributed Network Mining, Paris, France*, pages 1332–1339. ACM Press, 2015.

[CO-49] P. Rastin et R. Kanawati. A multiplex-network based approach for clustering ensemble selection. *International workshop on emerging trends in clustering, Orléans, France*, 2015. 8 pages.

[CO-48] T. Sarazin, H. Azzag et M. Lebbah. Modèle de biclustering dans un paradigme "MapReduce". *15èmes Journées Francophones Extraction et Gestion des Connaissances (EGC 2015), Luxembourg*, Revue des Nouvelles Technologies de l'Information, pages 467–468. Hermann, 2015.

[CO-47] H. Soldano, G. Santini et D. Bouthinon. Abstract and local concepts in attributed networks. *Workshop on Social Network Analysis SNAFCA in conjunction with the 13th International Conference on Formal Concept Analysis (ICFCA 2015), Nerja, Spain*, volume 1534 de *CEUR Workshop Proceedings*. CEUR-WS.org, 2015.

[CO-46] H. Soldano, G. Santini et D. Bouthinon. Local rules associated to k-communities in an attributed graph. *MANEM@ASONAM 15, First International workshop on Multiplex and Attributed Network Mining, Paris, France*, pages 1340–1347. ACM Press, 2015.

[CO-45] J. Sublime, Y. Bennani et A. Cornuéjols. Un algorithme ICM basé sur la compacité pour la segmentation des images satellites à très haute résolution. *15èmes Journées Francophones Extraction et Gestion des Connaissances (EGC 2015), Luxembourg*, pages 191–196, 2015.

[CO-44] L.-M. Veillon, G. Bourgne et H. Soldano. Apprentissage collaboratif de proximité. *Journée Réseaux Sociaux et Intelligence Artificielle, PFIA, Rennes, France*, 2015.

[CO-43] J.-D. Zucker, Y. Chevaleyre et V. S. Dao. Experimental analysis of new learning algorithms for ternary classifiers. *11th IEEE International Conference on Computing and Communication Technologies (ICCCT 2015), Chennai, India*, 2015. 6 pages.

2014

[CO-42] M. A. T. Arredondo, J. Sierra-Pérez, E. Zenuni, G. Cabanes, J. Rodellar, A. Guemes, C.-P. Fritzen et al. A pattern recognition approach for damage detection and temperature compensation in acousto-ultrasonics. *EWSHM-7th European Workshop on Structural Health Monitoring, Nantes, France*, pages 1093–1100, 2014.

[CO-41] D. Bouthinon et H. Soldano. Relational learning from ambiguous examples. J. Davis et J. Ramon, éditeurs, *Inductive Logic Programming (ILP 2014), Nancy, France*, 2014. 12 pages.

[CO-40] D. Bouthinon et H. Soldano. Relational learning from ambiguous examples. M.-O. Berger et S. Desprès, éditeurs, *Reconnaissance de Formes et Intelligence Artificielle (RFIA 2014), Rouen, France*, 2014.

[CO-39] A. Chaibi, H. Azzag et M. Lebbah. Pondération de blocs de variables en bi-partitionnement topologique. *Extraction et Gestion des Connaissances (EGC'2014), Rennes, France*, Revue des Nouvelles Technologies de l'Information, pages 317–328. Hermann, 2014.

[CO-38] I. Falih, M. Hmimida et R. Kanawati. Community detection in multiplex network : a comparative study. *Multiplex networks, Satelitte workshop at European conference on complex systems, Lucca,Italy*, Lucca, Italy, 2014.

[CO-37] M. Hmimida et R. Kanawati. A seed-centric algorithm for community detection in multiplex networks. *First european conference on social network analysis, Barcelona, Spain*, 2014. 8 pages.

[CO-36] R. Kanawati. Combinaison de modularités locales pour l'identification de communautéss égo-centrées. L. Boudjeloud, B. Le Grand et R. Kanawati, éditeurs, *Actes de 2ème atelier de fouille de grands graphes (EGC'14), Rennes, France*, 2014.

[CO-35] R. Kanawati. A multi-objective algorithm for local community identification in complex networks. *International Workshop on Computational Social Networks, Thessaloniki, Greece*. Springer, 2014.

[CO-34] F. Koriche et Y. Chevaleyre. Apprentissage online de modeles d'ising parcimonieux. *Conference d'Apprentissage Automatique (CAp'2014), Saint-Étienne, France*, pages 81–88. 2014.

[CO-33] V.-M. Le, Y. Chevaleyre, J.-D. Zucker et T. V. Ho. A general approach to solve decomposable optimization problems in multiagent simulations settings : application to tsunami evacuation. *International Conference on Swarm Intelligence Based Optimization, Mulhouse, France*, pages 20–27, 2014.

[**CO-32**] I. Redko et Y. Bennani. Universal unsupervised transfer learning through non-negative matrix factorization. *AutoML workshop, International Conference on Machine Learning (ICML), Beijing, China*, 2014.

[CO-31] C. Rodrigues, H. Soldano, G. Bourgne et C. Rouveirol. Collaborative on line learning of an action model. J. Davis et J. Ramon, éditeurs, *Inductive Logic Programming (ILP 2014), Nancy, France*, 2014. 6 pages.

[**CO-30**] C. Rodrigues, H. Soldano, G. Bourgne et C. Rouveirol. A consistency based approach of action model learning in a community of agents. *2014 International Conference on Autonomous Agents and Multi-agent Systems (AAMAS '14), Paris, France*, pages 1557–1558. International Foundation for Autonomous Agents and Multiagent Systems, 2014. (ext. abstract).

[**CO-29**] T. Sarazin, H. Azzag et M. Lebbah. SOM clustering using Spark-MapReduce. *28th IEEE International Parallel and Distributed Processing Symposium - International Workshop on High Performance Data Intensive Computing (IPDPS-HPDIC2014), Phoenix, Arizona, France*, pages 1727–1734. IEEE Computer Society, 2014.

[CO-28] J. Sierra-Pérez, M. A. Torres-Arredondo, G. Cabanes, A. Güemes, L. E. Mujica et C.-P. Fritzen. Damage detection in metallic beams from dynamic strain measurements under different load cases by using automatic clustering and pattern recognition techniques. *EWSHM-7th European Workshop on Structural Health Monitoring, Nantes, France*, pages 348–354, 2014.

[CO-27] H. Soldano. Closed patterns and abstraction beyond lattices. M.-O. B. e. S. Desprès, éditeur, *Reconnaissance de Formes et Intelligence Artificielle (RFIA 2014), Rouen, France*, 2014.

[CO-26] H. Soldano et G. Santini. Graph abstraction for closed pattern mining in attributed networks. M. Sebban et L. Denoyer, éditeurs, *Conférence d'Apprentissage (CAP 2014), Saint-Étienne, France*, 2014.

2013 [CO-25] S. Bannour, L. Audibert et H. Soldano. Ontology-based semantic annotation : an automatic hybrid rule-based method. *The BioNLP Shared Task 2013 Workshop, Sofia, Bulgaria*, pages 139–143. Association for Computational Linguistics, 2013. (poster).

[**CO-24**] G. Cabanes, Y. Bennani et N. Grozavu. Unsupervised learning for analyzing the dynamic behavior of online banking fraud. *13th IEEE International Conference on Data Mining Workshops (ICDM Workshops), Dallas, Texas, USA*, pages 513–520, 2013.

[CO-23] A. Chaibi, H. Azzag et M. Lebbah. Nouvelle approche de bi-partitionnement topologique. *Extraction et Gestion des Connaissances (EGC'2013), Toulouse, France*, volume RNTI-E-24 de *Revue des Nouvelles Technologies de l'Information*, pages 67–78. Hermann, 2013.

[CO-22] N.-Q. Doan, H. Azzag et M. Lebbah. Sélection de variables non supervisée sous contraintes hiérarchiques. *Extraction et Gestion des Connaissances (EGC'2013), Toulouse, France*, volume RNTI-E-24 de *Revue des Nouvelles Technologies de l'Information*, pages 37–48. Hermann, 2013.

[CO-21] R. Kanawati. Co-authorship link prediction in multiplex bibliographical networks. *Multiplex network workshop - European conference on complex systems (ECCS'13), Paris, France*, 2013.

[CO-20] R. Kanawati. A complex network approach for evaluating query similarity metrics. *International workshop on Web mining (Webi), Angers, France*, 2013.

[CO-19] R. Kanawati. Seed-centric approaches for community detection in complex interaction networks : A comparative review. *Complexity in social systems : from data to models, Cergy, France*, 2013. 2 pages.

[CO-18] V.-M. Le, Y. Chevaleyre, J.-D. Zucker et T. V. Ho. Speeding up the evaluation of casualties in multi-agent simulations with linear programming : application to optimization of sign placement for tsunami evacuation. *IEEE International Conference on Computing and Communication Technologies (ICCCT 2015), Chennai, India*, pages 54–62. 2013.

2012 [CO-17] A. Chaibi, H. Azzag et M. Lebbah. Détection de groupes outliers en classification non supervisée. Y. Lechevallier, G. Melançon et B. Pinaud, éditeurs, *Extraction et Gestion des Connaissances (EGC'2012), Bordeaux, France*, volume RNTI-E-23 de *Revue des Nouvelles Technologies de l'Information*, pages 119–125. Hermann, 2012.

[CO-16] I. Chebil, M. Elati, R. Nicolle, C. Rodrigues et C. Rouveirol. LICORN* : construction de réseaux de régulation chez l'homme. *Conférence Francophone d'Apprentissage (CAp'2012), Nancy, France*, page 16 pages, 2012.

[CO-15] Y. Chevaleyre, F. Koriche et J.-D. Zucker. Learning linear classifiers with ternary weights from metagenomic data. *Conference francophone sur l'Apprentissage automatique (CAp'2012), Nancy, France*, 2012. 17 pages.

[CO-14] N.-Q. Doan, H. Azzag et M. Lebbah. Clustering multi-niveaux de graphes : hiérarchique et topologique. Y. Lechevallier, G. Melançon et B. Pinaud, éditeurs, *Extraction et Gestion des Connaissances (EGC'2012), Bordeaux, France*, volume RNTI-E-23 de *Revue des Nouvelles Technologies de l'Information*, pages 567–569. Hermann, 2012.

[CO-13] N. Grozavu, Y. Bennani et L. Labiod. Transformation de l'espace de description pour l'apprentissage par transfert. Y. Lechevallier, G. Melançon et B. Pinaud, éditeurs, *Extraction et Gestion des Connaissances (EGC'2012), Bordeaux, France*, volume RNTI-E-23 de *Revue des Nouvelles Technologies de l'Information*, pages 35–40. Hermann, 2012.

[CO-12] F. Hamdi et Y. Bennani. Apprentissage d'ensemble d'opérateurs de projection orthogonale pour la détection de nouveauté. Y. Lechevallier, G. Melançon et B. Pinaud, éditeurs, *Extraction et Gestion des Connaissances (EGC'2012), Bordeaux, France*, volume RNTI-E-23 de *Revue des Nouvelles Technologies de l'Information*, pages 41–52. Hermann, 2012.

[CO-11] M. Hmimida, M. Ankoud et R. Kanawati. Nouveau modèle de recommandation pour la classification à facettes. *15éme édition du Colloque International sur le Document Électronique (CiDE'15), Montpellier, France*, 2012. 12 pages.

[CO-10] R. Jaziri, M. Lebbah et Y. Bennani. Classification probabiliste non supervisée et visualisation des données séquentielles. Y. Lechevallier, G. Melançon et B. Pinaud, éditeurs, *Extraction et Gestion des Connaissances (EGC'2012), Bordeaux, France*, volume RNTI-E-23 de *Revue des Nouvelles Technologies de l'Information*, pages 137–148. Hermann, 2012.

[CO-9] R. Kanawati. CaSE : A CBR approach for re-ranking web search results. *International workshop on case-based reasoning (CBR-MD'12), Berlin, Germany*. IBAI, 2012. 10 pages.

[CO-8] R. Kanawati. Mining the dynamics of scientific publication networks for collaboration recommendation. *Second international workshop on mining communities and people recommender (COMMPER@ ECML2012), Bristol, UK*, 2012. 10 pages.

[CO-7] L. Labiod et Y. Bennani. Classification des données catégorielles via la maximisation spectral de la modularité. Y. Lechevallier, G. Melançon et B. Pinaud, éditeurs, *Extraction et Gestion des Connaissances (EGC'2012), Bordeaux, France*, volume RNTI-E-23 de *Revue des Nouvelles Technologies de l'Information*, pages 149–154. Hermann, 2012.

[CO-6] K. Mouhoubi, L. Létocart et C. Rouveirol. Extraction de biclusters contraints dans des contextes bruités. *Conférence Francophone d'Apprentissage (CAp 2012), Nancy, France*, 2012. 17 pages.

[CO-5] N. A. Nguyen Thi, Y. Chevaleyre et J.-D. Zucker. Optimizing sign placements for crowd evacuation on road network in case of tsunami alert. *21st IEEE International Workshop on Enabling Technologies : Infrastructure for Collaborative Enterprises (WETICE 2012), Toulouse, France*, pages 394–396, 2012.

[CO-4] M. Pujari et R. Kanawati. Applying supervised rank aggregation to link prediction in large scale complex networks. *Journée : Big data mining and visualization, Tours, France*, 2012.

[CO-3] M. Pujari et R. Kanawati. Supervised rank aggregation approach for link prediction in complex networks. *3ème Conférence sur l'analyse et la modélisation des réseaux : approches mathématiques et informatiques (MARAMI 2012), Villetaneuse, France*, 2012.

[CO-2] Z. Yakoubi et R. Kanawati. Applying leaders driven community detection algorithms to data clustering. *The 36th Annual Conference of the German Classification Society on Data Analysis, Machine Learning and Knowledge Discovery (GfKI'12), Hildesheim, Germany*, 2012. 2 pages.

[CO-1] Z. Yakoubi et R. Kanawati. Classification non-supervisée par application d'un algorithme de détection de communautés dans les réseaux complexes. *XIX journée de la société Francaise de Classification (SFC'12), Marseille, France*, 2012. 4 pages.

Édition d'ouvrages collectifs

2017 [**ED-12**] C. Rouveirol, R. G. Pensa et R. Kanawati, éditeurs. *Special issue on Dynamic Networks and Knowledge Discovery*, volume 106 de *Machine Learning Journal*, 2017.

2015 [ED-11] G. Cleuziou, C. D. Runz, M. Lebbah et G. Forestier, éditeurs. *Fouille de données complexes (FDC@EGC2015), Luxembourg*, volume RNTI-E-25 de *RNTI*. Hermann, 2015.

2014 [ED-10] Y. Bennani et E. Viennet, éditeurs. *Apprentissage Artificiel et Fouille de Données, AAFD 2012, Villetaneuse, France,*, volume A-6 de *Revue des Nouvelles Technologies de l'Information*. Hermann-Éditions, 2014, ISBN 9782705688387.

[ED-9] L. Boudjeloud, B. L. Grand et R. Kanawati, éditeurs. *Actes du 2ème Atelier EGC sur la fouille de grands graphes (FGG-EGC 2014), Rennes, France*, 2014.

[ED-8] R. Kanawati et B. Le Grand, éditeurs. *Actes de 5ème journée de fouille de grands graphes, Paris, France*, 2014.

[ED-7] R. Kanawati, R. G. Pensa et C. Rouveirol, éditeurs. *Proceedings of the 2nd International Workshop on Dynamic Networks and Knowledge Discovery, DyNaK 2014, co-located with the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2014), Nancy, France*, volume 1229 de *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.

2013 [ED-6] E. Banissi, H. Azzag, M. W. M. Bannatyne, S. Bertschi, F. Bouali, R. Burkhard, J. Counsell, A. Cuzzocrea, M. J. Eppler, B. Hammer, M. Lebbah, F. T. Marchese, M. Sarfraz, A. Ursyn, G. Venturini et T. G. Wyeld, éditeurs. *17th International Conference on Information Visualisation (IV 2013), London, UK*. IEEE Computer Society, 2013, ISBN 978-0-7695-5049-7.

[ED-5] L. Boudjeloud, B. L. Grand et R. Kanawati, éditeurs. *Actes du 1er Atelier EGC sur la fouille de Grands Graphes (FGG-EGC 2013), Toulouse, France*, 2013.

[ED-4] R. Kanawati et C. Roth, éditeurs. *Actes de la 4ème conférence sur les modéles et l'analyse des réseaux : approches informatiques et mathématiques, Saint-Étienne, France*, 2013.

[ED-3] R. Pensa, F. Cordero, C. Rouveirol et R. Kanawati, éditeurs. *Special Issue on Knowledge Discovery in Dynamic Networks*, volume 17 de *Intelligent Data Analysis*, 2013.

2012 [ED-2] E. Banissi, S. Bertschi, C. Forsell, J. Johansson, S. Kenderdine, F. T. Marchese, M. Sarfraz, L. J. Stuart, A. Ursyn, T. G. Wyeld, H. Azzag, M. Lebbah et G. Venturini, éditeurs. *16th International Conference on Information Visualisation (IV 2012), Montpellier, France*. IEEE Computer Society, 2012, ISBN 978-1-4673-2260-7.

[ED-1] R. Kanawati et G. Santini, éditeurs. *Actes de la troisième journée de fouille de grands graphes (JFGG'12), Villeteneuse, France*, 2012. 20 pages.

Logiciels et Brevets (LO)

2017 [**LO-2**] M. Ghesmoune, T. Sarazin, H. Azzag, S. Benbernou, M. Lebbah, T. Duong et M. Ouziri. Spark-clustering-notebook. Logiciel LIPN, 2017. https://github.com/Spark-clustering-notebook/coliseum.

2012 [LO-1] D. Bouthinon et H. Soldano. Lear, 2012. http://lipn.univ-paris13.fr/~bouthinon/softwares/lear/.

Thèses et Habilitations (TH)

2016 [TH-20] P. Alizadeh. Elicitation and Planning in Markov Decision Processes with Unknown Rewards. Thèse, Université Paris 13 - Sorbonne Paris Cité, 2016.

[TH-19] M. Ghesmoune. Apprentissage non supervisé de flux de données massives : Application aux big data d'assurance. Thèse, Université Paris 13 - Sorbonne Paris Cité, 2016.

[TH-18] V. M. Le. Machine learning methods for optimization in multi-agent decision support system : application to sign placement for tsunami evacuation. Thèse, Université Paris 13 - Sorbonne Paris Cité, 2016.

2015 [TH-17] S. Bannour. Apprentissage interactif de règles d'extraction d'information textuelle. Thèse, Université Paris 13 - Sorbonne Paris Cité, 2015.

[TH-16] M. Pujari. Prévision de liens dans les grands graphes de terrain (application aux réseaux bibliographiques). Thèse, Université Paris 13 - Sorbonne Paris Cité, 2015.

[TH-15] I. Redko. Nonnegative matrix factorization for transfer learning. Thèse, Université Paris 13 - Sorbonne Paris Cité, 2015.

[TH-14] J. Sublime. Contributions to collaborative clustering and its potential applications on very high resolution satellite images. Thèse, Paris Saclay, 2015.

2014 [TH-13] I. Chebil. Méthodes d'ensemble pour l'inférence de réseaux de régulation coopératifs. Thèse, Université Paris 13 - Sorbonne Paris Cité, 2014.

[TH-12] Z. Yakoubi. Détection et évaluation des communautés dans les réseaux complexes. Thèse, Université Paris 13 - Sorbonne Paris Cité, 2014.

2013 [TH-11] H. Azzag. Classification non supervisée et visualisation de données : approches topologiques, hiérarchiques. HdR, Université Paris 13 - Sorbonne Paris Cité, 2013.

[TH-10] A. Chaibi. Contribution en apprentissage topologique non supervisé pour la fouille de données. Thèse, Université Paris 13 - Sorbonne Paris Cité, 2013.

[TH-9] N.-Q. Doan. Hierarchical and topological models for clustering and visualization (modèles hiérarchiques et topologiques pour le partitionnement et la visualisation des données). Thèse, Université Paris 13 - Sorbonne Paris Cité, 2013.

[TH-8] M. Essaidi. Model-driven data warehouse and its automation using machine learning techniques. Thèse, Université Paris 13 - Sorbonne Paris Cité, 2013.

[TH-7] M. Ghassany. Contributions to collaborative clustering. Thèse, Université Paris 13 - Sorbonne Paris Cité, 2013.

[TH-6] R. Jaziri. Modèles de mélanges topologiques pour la classification de données structurées en séquences. Thèse, Université Paris 13 - Sorbonne Paris Cité, 2013.

[TH-5] K. Mouhoubi. Extraction de motifs contraints dans des données bruitées. Thèse, Université Paris 13 - Sorbonne Paris Cité, 2013.

[TH-4] C. Rodrigues. Apprentissage incrémental de modèles d'action relationnels. Thèse, Université Paris 13 - Sorbonne Paris Cité, 2013.

2012 [TH-3] F. Hamdi. Apprentissage en distributions déséquilibrées. Thèse, Université Paris 13 - Sorbonne Paris Cité, 2012.

[TH-2] M. Lebbah. Contributions en apprentissage non supervisé à partir de données complexes. HdR, Université Paris 13 - Sorbonne Paris Cité, 2012.

[TH-1] A. Osmani. Modélisation et raisonnement sur des données relationnelles. HdR, Université Paris 13 - Sorbonne Paris Cité, 2012.

Rapports de contrats

2014 [RC-3] S. Benbernou et M. Lebbah. Workflows and scientific big data preservation. Rapport Technique, CNRS - Scientific Data Preservation initiative, 2014. édité par Cristinel Diaconu.

[RC-2] C. Cérin, M. Lebbah et H. Azzag. Workflows and scientific big data preservation. Rapport Technique, CNRS - Scientific Data Preservation initiative, 2014. édité par Cristinel Diaconu.

[RC-1] C. Diaconu, S. Kraml, C. Surace, D. Chateigner, T. Libourel Rouge, A. Laurent, Y. Lin, M. Schaming, S. Benbernou, M. Lebbah, D. Boucon, C. Cérin, H. Azzag, P. Mouron, J.-Y. Nief, S. Coutin et V. Beckmann. PREDON Scientific Data Preservation 2014. Rapport Technique, LIRMM, 2014.

3.3 Publications antérieures à l’arrivée au LIPN

Les références comportent « ↑ » et correspondent à des publications antérieures à l’arrivée au LIPN, non référencées par ailleurs.

Signification des soulignements :
- *simple* : enseignant-chercheur ou chercheur titulaire,
- *double* : doctorant,
- *pointillé* : post-doc, ingénieur ou autre situation.

Articles dans des revues avec comité de lecture et chapitres de livres (ACL)

Articles dans des revues internationales avec comité de lecture sélectif

2014 [↑RI-2] B. Matei, Y. Meyer et J. Ortega-Cerdà. Stable sampling and Fourier multipliers. *Publicacions Matematiques*, (58) :341–351, 2014.

2012 [↑RI-1] B. Matei et S. Meignen. Nonlinear cell-average multiscales signal representations : Application to signal denoising. *Signal Processing*, (92) :2738–2746, 2012.

3.4 Réalisations informatiques (logiciels), Brevets

3.4.1 Logiciels

- **Spartakus**

Environnement packagé (avec Docker) : https://hub.docker.com/r/spartakus/coliseum/
Code ouvert : https://github.com/Spark-clustering-notebook/coliseum/wiki

Mise à disposition de 3 algorithmes de clustering *scalable* : cartes auto-organisatrices (SOM-MR) [**CO-29**], Mean-shift-LSH [RI-26], batchStream [RI-28]. Le prototype des algorithmes est écrit en Scala en utilisant la plateforme Spark.

3.4.2 Start-up

Erick Alphonse est actuellement (et depuis 2010) en disponibilité pour création d'entreprise. L'origine de ce projet d'essaimage est la volonté d'Erick Alphonse de valoriser ses travaux de recherche sur la fouille de données relationnelles. L'objet même de la société (IDAaaS), créée en décembre 2009, est de développer et commercialiser des solutions innovantes d'Extraction de Connaissances à partir de Données. Le projet est de développer des outils, suivant le modèle SaaS, de valorisation de masses de données, issus de la recherche en Intelligence Artificielle : Intelligent Data Analysis as a Service (IDAaaS). L'utilisateur loue des capacités d'acquisition, de stockage et de sécurisation de données dans un SGBDR (Système de gestion de base de données relationnelles) inductif et de la puissance de traitement et calcul sur ces données. L'outil développé s'appuie sur les recherches des laboratoires internationaux en Apprentissage Automatique (Machine Learning), dont fait partie A³.

De même, en 2012, A. Osmani a obtenu de l'Université Paris 13 une délégation pour création d'entreprise sur le thème de l'Internet des objets. En 2017, il a réintégré l'équipe A³, dans laquelle il va poursuivre des recherches sur l'internet des objets.

Chapitre 4

Activités de recherche et indices de reconnaissance

Sommaire

4.1 Participation à des comités, animation scientifique 56
4.1.1 Comités éditoriaux 56
4.1.2 Comités de programme de conférences 56
4.2 Organisation de colloques et de congrès 57
4.3 Activités d'évaluation 58
4.4 Postdoctorants et chercheurs accueillis 58
4.4.1 Post-doctorants 58
4.4.2 Chercheurs invités 59
4.5 Interactions avec l'environnement 59
4.6 Contrats de recherche publics 60
4.7 Indices de reconnaissance 65
4.7.1 Invitations des membres du LIPN 65

L'équipe Apprentissage Artificiel et Applications (A³) s'implique fortement au niveau national dans l'animation de la communauté scientifique de l'apprentissage automatique et de la fouille de données et des statistiques, ainsi qu'au niveau international notamment dans l'organisation de sessions spéciales dans des conférences internationales de premier plan (ICONIP, IJCNN, IV, ...).

Des membres de l'équipe organisent tous les deux ans et depuis maintenant une dizaine d'années à l'Université Paris 13 les Journées d'Apprentissage Automatique et de Fouille de Données (AAFD). Ces journées ont eu lieu pour la première fois à l'étranger en 2016 (Maroc), couplées aux journées SFC'2016.

4.1 Participation à des comités, animation scientifique

4.1.1 Comités éditoriaux

- **Membres de comités éditoriaux**

  — **International journal on Social Network Mining**, InderScience Publisher, depuis 2011 : R. Kanawati
  — Special issue of **Intelligent Data Analysis journal on Knowledge Discovery in dynamic networks** (2013) : R. Kanawati, C. Rouveirol
  — Special issue of **Machine Learning journal on Dynamic networks and Knowledge Discovery** (2017) : R. Kanawati, C. Rouveirol
  — Special issue of **International journal on Web-based communities** : Communities in complex networks (2012) : R. Kanawati

- **Autres activités de relecture**

Activités de recherche et indices de reconnaissance

Les membres de l'équipe ont participé à des relectures, en particulier dans des numéros spéciaux, pour les revues suivantes :
— Technique et Science Informatiques
— Computational Statistics and Data Analysis
— Intelligent Data Analysis
— Machine Learning
— Journal of Machine Learning Research
— Journal of Artitifical Intelligence Research
— Journal of Artitifical Intelligence
— IEEE/ACM Transactions on Computational Biology and Bioinformatics
— Briefings in Bioinformatics
— Computational Intelligence
— Neurocomputing journal
— Neural Computing & Applications

4.1.2 Comités de programme de conférences

- **Présidence de comités de programme**

  AAFD Journées thématiques Apprentissage Artificiel et Fouille de Données. Y. Bennani, 2012, 2014.

- **Comités de programme de conférences internationales importantes**

  AAAI Conference on Artificial Intelligence. Y. Chevaleyre, 2012, 2014, 2016.

  ECAI European Conference on Artificial Intelligence. Y. Bennani, Y. Chevaleyre, C. Rouveirol, 2012.

  ECML PKDD European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. C. Rouveirol, 2012, 2013, 2014, 2015, 2016, 2017.

IJCAI International Joint Conference on Artificial Intelligence. Y. Chevaleyre, 2013, 2015, 2017.

IJCNN IEEE International Joint Conference on Neural Networks. Y Bennani, 2012, 2013, 2015, 2017.

- **Autres comités de programme**

2012 ESIC, FOSINT-SI, ICMCS, ICMLA, ICPRAM, IWCNA, NaBIC

2013 IClaNov, ICMLA

2014 IClaNov, ICMLA

2015 ICPRAM

2016 ICPRAM, INNS

2017 ICPRAM

- **Principaux comités de programme de conférences nationales**

Les membres de l'équipe participent aux conférences françaises principales du domaine (CAp, EGC, SFC, . . .) ainsi que des ateliers associés.

4.2 Organisation de colloques et de congrès

- **Conférences internationales**

2017 *International Workshop on Advances in Learning from/with Multiple Learners (ALML)*, affiliated to IJCNN. 2015 et WCCI'14. G. Cabanes, N. Grozavu.

2016 Special Session on *Topological and Graph Based Clustering Methods* at ICONIP. N. Grozavu.

2015 *International Workshop on Advances in Learning from/with Multiple Learners (ALML)*, affiliated to IJCNN. 2015 et WCCI'14. G. Cabanes, N. Grozavu.

2015 Special Session on *Incremental Machine Learning (IML)* at SoCPaR. G. Cabanes.

2014 *International Workshop on Advances in Learning from/with Multiple Learners (ALML)*, affiliated to WCCI. G. Cabanes, N. Grozavu.

2014 Special Session on *Incremental Machine Learning (IML)* at WCCI. G. Cabanes.

2014 *18th International Conference Information Visualisation (IV)*. M. Lebbah.

2014 *2nd International workshop on Dynamic Networks and Knowledge Discovery*, workshop associé à ECML-PKDD. R. Kanawati, C. Rouveirol.

2013 Special Session on *Active Learning and Experimental Design (ALED)* at IJCNN. N. Grozavu.

2013 Special Session on *Incremental Machine Learning : Methods and Applications (IML)* at IJCNN. N. Grozavu.

Special Session on *Co-Clustering of Large and High Dimensional Data* at ICONIP. Y. Bennani, N. Grozavu.

2012 Special Session on *Nonnegative Matrix factorization paradigm for unsupervised learning* at IJCNN. Y. Bennani, N. Grozavu.

2012 *Symposium on Advances in Interactive and Visual Data Clustering* at IV. H. Azzag, M. Lebbah.

- **Conférences nationales**

  à venir En 2018, 18ème édition de la conférence *Extraction et Gestion des Connaissances (EGC 2018)*, MSH, Paris 13 : H. Azzag, M. Lebbah.

  2015 Rencontres *Calcul intensif et Sciences des données* à Villetaneuse. M. Lebbah.

  2015 Journée thématique de la SFdS-DMA3, *Big-Data : Une vision globale : Gestion, Analyse, Ethique et Logiciels*. M. Lebbah.

  2014 Rencontres *Calcul intensif et Sciences des données* à Vichy. M. Lebbah.

  2014 3e édition des journées *Big Data Mining and Visualization* EGC et AFIHM. (GT : EGC-FDC-FGG- VIF), à Lille. M. Lebbah.

  2013 Journées *Big Data Mining and Visualization* EGC et AFIHM. (GT : EGC-FDC-FGG-VIF), au CNAM-Paris. M. Lebbah.

  2013 10e édition de l'atelier associé à la conférence EGC, *Fouille de données complexes - complexité liée aux données multiples*. M. Lebbah.

  2012 Colloque *Big Data Mining and Visualization*. (EGC- AFHIM), à Polytech Tours. M. Lebbah.

  2012 Troisième *Journée EGC Fouille de Grands Graphes (JFGG'2012)* associé à la conférence *MARAMI'2012*, Université Paris 13. R. Kanawati, G. Santini.

  2012 Colloque *MARAMI'2012*, Université Paris 13. R. Kanawati, C. Rouveirol, G. Santini.

  2012 9e édition de l'atelier associé à la conférence EGC, *Fouille de données complexes - complexité liée aux données multiples*. M. Lebbah.

  2012 Journées *Big Data Mining and Visualization* EGC et AFIHM. (GT : EGC-FDC-FGG-VIF), à Polytech Tours. H. Azzag, M. Lebbah, R. Kanawati.

4.3 Activités d'évaluation

AERES Évaluation de laboratoire pour l'AERES en 2014 (LIRIS, Lyon) : C. Rouveirol.

HCERES Évaluation de laboratoire pour l'HCERES en 2016 (LI, Tours et LIFO, Orléans) : C. Rouveirol.

HCERES Évaluation de Masters pour l'HCERES : Y. Bennani.

ANR Évaluation de projets pour l'ANR : H. Azzag, Y. Bennani, Y. Chevaleyre, M. Lebbah, B. Matei.

PHC Évaluation des projets PHC : B. Matei.

Pôle Competitivité Cap Digital Expert : Y. Bennani.

Prix Membre du jury d'attribution du prix de thèse *AFIA* en 2012 : H. Soldano.

4.4 Postdoctorants et chercheurs accueillis

4.4.1 Post-doctorants

L'équipe A³ a accueilli 16 post-doctorants depuis 2012.

Nom	*Prénom*	*Situation*	*Début*	*Fin*	*Financement*
Alizadeh	Pegah	Postdoc	09/12/2016	31/08/2017	
Benchettara	Nasserin	Postdoc	20/12/2011	10/12/2012	
Chaibi	Amine	Postdoc	15/12/2010	29/11/2013	
Chebil	Inès	Postdoc	27/09/2014	31/08/2015	
Chelly	Zeineb	Postdoc	2016	2018	Bourse Marie Sklodowska Curie
Coutant	Anthony	Postdoc	15/01/2016	30/04/2018	ANR Adalab
Duong	Thanh Buu	Postdoc	01/12/2014	31/08/2016	PIASquare-Predict

Fenzar Sadki	Jalia	Postdoc	07/07/2011	02/11/2013	
Ghesmoune	Mohammed	ATER	26/11/2016	31/08/2017	
Hindawi	Mohammed	Postdoc	07/05/2013	30/04/2014	
Lachaud	Antoine	Postdoc	01/03/2016	31/12/2016	ANR Coclico
Laclau	Charlotte	Postdoc	01/06/2016	31/10/2016	ANR Coclico
Mouhoubi	Karima	Postdoc	20/09/2013	31/12/2013	
Pujari	Manisha	Postdoc	05/03/2015	01/10/2015	
Redko	Ievgen	Postdoc	26/11/2015	31/03/2016	
Rodrigues	Christophe	Postdoc	21/01/2013	30/09/2015	

4.4.2 Chercheurs invités

L'équipe A³ accueille chaque année des chercheurs invités.

Prénom	*Nom*	*Institution*	*Pays*	*Durée*	*Financement*
2017					
Martin	Atzmueller	Univ. Tilburg	Pays-Bas	1 mois	UP13
2016					
Seiichi	Ozawa	Univ. Kobe	Japon	2 sem.	UP13
Rosanna	Verde	Second Univ. of Naples	Italie	1 sem.	UP13
2015					
Rosanna	Verde	Second Univ. of Naples	Italie	1 mois	UP13
2014					
Michel	Verleysen	Univ. catholique de Louvain	Belgique	1 mois	UP13
2013					
Vladimir	Nikulin	Binghamton Univ	États-Unis	1 mois	UP13
Vladimir	Radevski	South East European Univ.	Macédoine	1 mois	UP13
2012					
Abdelouahid	Lyhyaoui	ENSAT Tanger	Maroc	1 mois	UP13
Nabil	Keskes	ESI Sidi Bel Abbès	Algérie	1 mois	UP13

Activités d recherche e indices de re connaissanc

4.5 Interactions avec l'environnement

Plusieurs thèses CIFRE se sont déroulées en partenariat entre des entreprises et l'équipe A³ :

- INA (2009–2012), *Apprentissage non supervisé de données structurées. Application à la structuration automatique des flux télévisuels par analyse des répétitions.*
- IntelligencePower (2009–2012), *Automatisation du processus de génération d'un entrepôt de données.*
- Anticipeo (2010–2013), *Approches prédictives d'apprentissage statistique.*
- Altic, puis SMokeWatchers (2013–2016), *Apprentissage massivement distribué dans un environnement Big Data.*
- MindLytix (2015–2018), *Apprentissage automatique et adaptatif pour le clustering de flux de données hétérogènes.*
- Kameleoon, (2016–2019), *Classification massive : Application à la personnalisation web en temps-réel.*
- EDF R&D, (2016–2019), *Analyse des réseaux multiplexes pour l'étude de sûreté.*

4.6 Contrats de recherche publics

Les projets en italiques sont en cours. Ceux entre parenthèses sont des contrats de membres du laboratoire avant leur arrivée au LIPN.

Type	Coordinateur	Resp. scient.	Nom	Financeur	Période	Durée (mois)	Budget global (K€)	Part LIPN (K€)
International								
PHC Amadeus	•	B. Matei		MAE	*2015–2017*		13	
CHIST-ERA		C. Rouveirol	*ADALAB*	ANR	*2014–2017*	36	1357	203
ANR - National								
(ANR)		Y. Chevaleyre	*(COCORICO)*	MENESR	*2013–2017*	48	435	
ANR		Y. Bennani	COCLICO	MENESR	2012–2016	48	1 015	231
(ANR)		Y. Chevaleyre	(LARDONS)	MENESR	2010–2014	24	308	
ANR		Y. Bennani	E-FRAUD	MENESR	2010–2013	36	1 092	200
(ANR)		Y. Chevaleyre	(COMSOC)	MENESR	2010–2012	24	338	
ANR JCJC		P. Gérard	HARRI	MENESR	2009–2013	48		126,2
PEPS JCJC	•	N. Grozavu	A3FD	CNRS	2015–2016			
PIA		M. Lebbah	Square Predict	FSN	2013–2016	36	2700	595
Rech. Biomed.		C. Rouveirol	INSIGht	INCa	2010–2013	36	464	117,3
Régional et Local								
BQR	•	C. Rouveirol	ONTO-APP	UP13	2012–2013	12	36	18,5
FUI		R. Kanawati	*PCU*	Region IdF	*2017–2020*	36	1 500	200
FUI		Y. Bennani	Hermès	MENESR	2012–2015	36	1 907	141,5
FUI		R. Kanawati	Urban-D	Cons. Gén. 93	2010–2012	18	1 179	80

Activités de recherche et indices de reconnaissance

• PHC Amadeus

Partenaires : Institut de Mathématiques de Bordeaux, Département de Mathématiques de l'Université de Vienne (Autriche)

Le PHC Amadeus vise à étudier et proposer des méthodes optimales d'échantillonnage irrégulier en utilisant les pavages quasi-périodiques du plan.

Publications associées : [CI-49, RI-22, RI-23].

• CHIST-ERA ADALAB

Partenaires : *Académiques* : Univ. Brunel (UK) (coord.), Univ. Manchester (UK), Kath. Univ. Leuven (B), Univ. Evry (F)

Un « Robot Scientifique » est un système d'automatisation de laboratoire qui exploite des techniques du domaine de l'intelligence artificielle pour exécuter des cycles d'expérimentation scientifique. Ce projet vise à développer le cadre AdaLab (Adaptive, Automated Scientific Laboratory) pour la découverte semi-automatisée et automatisée de connaissances par des équipes de scientifiques humains et robots. Nous évaluerons le cadre AdaLab sur une application importante du monde réel en biologie cellulaire ayant une pertinence biomédicale pour le cancer et le vieillissement. Dans le cadre d'Adalab, le LIPN conçoit et développe des algorithmes de révision de réseaux biologiques incrémentaux, de fouille de données de graphes attribués ainsi que des algorithmes d'apprentissage actif pour la sélection d'expérimentations.

Publications associées : [CO-58, **CL-9***, CO-75*].

- **COCORICO-CODEC : Computation, Communication, Rationality and Incentives in Collective and Cooperative Decision Making**

Partenaires : LAMSADE-CNRS, LIP6, CREM-CNRS, LIG, PSE

Ce projet vise à étudier des classes de problèmes de décision collective sous trois angles différents : (a) l'impact computationel des mécanismes impliqués ; (b) la complexité de communication de ces mécanismes ; (c) et leur vulnérabilité et robustesse.

- **COCLICO : COllaboration, CLassification, Incrémentalité et COnnaissances)**

Partenaires : *Académiques* : LSIIT, LIPN, AgroParisTech/INRA. *Industriels* : LIVE, ESPACE DEV/IRD

La fouille de données est un maillon important dans la chaîne de traitements des données vers les connaissances. Ainsi, par exemple, appréhender et comprendre les processus de fonctionnement et de développement des systèmes plus ou moins anthropisés à diverses échelles spatiales et temporelles (urbanisation et pression sur les terres, érosion de la biodiversité, etc.) à partir de données satellites ou autres devient un élément majeur dans différents domaines tels l'étude de l'environnement ou les politiques publiques d'aménagement du territoire. Or les techniques d'analyse actuelles sont de plus en plus limitées face à l'avalanche actuelle de données hétérogènes souvent incomplètes voire imprécises et de plus en plus souvent fournies en continu au fil de l'eau. Coclico est un projet de recherche visant à étudier et proposer une méthode générique innovante permettant une analyse multi-échelle de grands volumes de données spatio-temporelles fournies en continu de qualité très variable, mettant en œuvre une approche multistratégie incrémentale dans laquelle la collaboration entre les différentes méthodes de fouille de données est guidée par des connaissances à la fois du domaine thématique (géosciences, géographie) formalisées en ontologies et du domaine de l'analyse (connaissances sur les méthodes), et garantissant un objectif de qualité finale prenant en compte à la fois la qualité des données et celles des connaissances.

Activités d recherche e indices de re connaissanc

Publications associées : [RI-12, RI-29*, **CI-78**, **RI-27**]

- **Lardons : Learning and Reasoning for Deciding Optimally using Numerical and Symbolic information**

Partenaires : LIP6-CNRS, GREYC-CNRS, LAMSADE, INRA

LARDONS est un projet de recherche fondamentale s'intéressant aux capacités d'apprentissage, de raisonnement et de prise de décision d'un agent artificiel autonome. Les objectifs de ce projet sont :

— Proposer de nouveaux modèles tenant compte du fait qu'un agent dispose typiquement de connaissances et de perceptions à la fois symboliques et numériques.
— Concevoir de nouvelles approches et de nouveaux algorithmes pour les problèmes en question, qui soient à la fois capables de traiter, mais aussi de tirer parti de cette dualité symbolique/numérique.

- **E-FRAUD BOX : Détection et Investigation de la fraude à la carte bancaire sur Internet**

Partenaires : *Académiques* : LIPN-CNRS, LIP6-CNRS. *Industriels* : ALTIC, KXEN, THALES Communications France (TCF), GIE Cartes Bancaires « CB », Gendarmerie Nationale, Police Nationale

Le projet E-fraud Box vise à développer une boîte à outils intégrée, dédiée à la détection et à l'investigation de la fraude à la carte bancaire sur Internet, et basée sur :

— des algorithmes d'apprentissage capables de modéliser les comportements frauduleux, de suivre la dynamique de la fraude dans le temps et d'étudier les interactions entre les transactions.
— des moteurs de détection et d'investigation de la fraude combinant les résultats des modèles.

Les moteurs seront utilisés pour :

— détecter le plus rapidement possible les cartes utilisées frauduleusement,
— construire des faisceaux d'indice caractérisant des comportements frauduleux et des scénarios de fraude,
— identifier les nouveaux modes opératoires des fraudeurs.

Publications associées : [**RI-9**, RI-5].

- **COMSOC : Choix Social Computationnel**

Partenaires : LAMSADE-CNRS, CRIL-CNRS, PREG-CNRS, CREM-CNRS

La théorie du choix social est une branche de l'économie qui s'intéresse à la prise de décision collective, par l'intermédiaire entre autres de procédures de vote, de négociation, d'enchères. Ce projet explore les aspects computationnels de la prise de décision collective.

Activités de recherche et indices de reconnaissance

- **HARRI**

Le projet HARRI avait pour objectif de résoudre des problèmes d'apprentissage par renforcement et plus généralement d'adaptation d'un agent informatique dans des contextes où les représentations relationnelles sont plus naturelles (objets discrets liés par des relations plutôt que vecteurs propositionnels). C'est souvent le cas pour les agents simulés et moins pour les robots et autres agents physiques. Plus particulièrement, nous avons porté nos efforts sur la question de l'apprentissage incrémental d'un modèle relationnel d'action (règles d'ordre 1) pour permettre à l'agent d'anticiper les conséquences de ses actions et accélérer son adaptation en ligne à un environnement a priori inconnu. Le problème d'apprentissage incrémental est traité comme un problème de révisions successives de modèles [CI-27]. Les algorithmes correspondants ont la particularité d'être ascendants et guidés par les données. Ils convergent dans le cas déterministe et ont montré empiriquement leur efficacité, y compris dans le cas de données bruitées.Les modèles d'action appris peuvent être utilisés en apprentissage par renforcement indirect ou en planification. Des mécanismes originaux d'apprentissage actif ont été proposés, adaptés au caractère ascendant de notre approche [**CI-1**].

- **A3FD**

A3FD est un projet de recherche visant à proposer des méthodes innovantes pour l'analyse autonome de grands volumes de données temporelles en rassemblant une équipe constituée de deux jeunes chercheurs. Ces méthodes doivent permettre la détection et la caractérisation de l'évolution temporelle des clusters en choisissant de manière autonome les valeurs optimales des paramètres des algorithmes. Ce type de méthodes n'existait pas à l'époque malgré un besoin important. Nous avons validé les méthodes proposées sur des données applicatives réelles pour l'identification de groupes d'opinions et d'expressions émotionnelles à partir de données de types tweets et l'étude de leur évolution dans le temps.

Publications associées : [CI-57, CI-46, CI-48].

- **Square Predict**

Partenaires : LIPN, LIPADE, LARIS, Asteck, Isthma-Viseo , Digital & Ethics, AXA

Square Predict est un Projet Investissement d'Avenir 2013 Cloud Computing numéro 3. Big Data (Fonds national pour la société numérique (FSN))

L'enjeu dans ce projet est de développer une plateforme Big Data clé en main dédiée aux producteurs d'assurance afin de leur permettre de croiser, fusionner et d'exploiter l'avalanche de données locales et externes (ouvertes et/ou payantes). Le LIPN avait pour objectif de développer des modèles d'apprentissage statistique et algorithmes pour une meilleure personnalisation des produits permettant aux fournisseurs de données de réaliser des visualisations et prédictions à partir des données croisées avec celles disponibles sur le web (réseaux sociaux, open data, . . .) [CI-55, **RI-25**, **CO-29**]. La plateforme Square Predict avait donc pour but de valoriser et monétiser le patrimoine de données qu'elles ont accumulées depuis des années, tout en respectant la vie privée des personnes [**LO-2** , **RI-28**].

- **INSIGHT : Identification de cibles thérapeutiques par la recherche de réseaux spécifiquement altérés dans les cancers**

Partenaires : Institut Curie ; LIPN-CNRS ; Equipe Recherche translationnelle en oncogénèse urogénitale, INSERM, Créteil ; Institut de Biologie Systémique et Synthétique (iSSB), Université d'Evry

À partir des données du transcriptome, de l'expression des miARNs et des protéines des cancers de vessie et d'urothélium normal à différents stades de prolifération et de différenciation, le but du projet est de construire des réseaux de régulation — facteurs de transcription, miARN et gènes cibles — observés dans le tissu normal et pathologique, et par l'analyse de ces réseaux, de suggérer des régulations significativement perturbées dans les réseaux actifs dans le tissu pathologique.

Le LIPN s'intéresse à des problèmes ouverts en inférence de réseaux. Il a notamment i) proposé des heuristiques robustes pour le problème NP-difficile d'inférence de graphes [CO-16] ii) étudié d'intégration des sources de données hétérogènes dans le mécanisme d'inférence de réseaux. Nous avons développé une méthode de classification croisée [CO-16] capable d'engendrer des régions éventuellement recouvrantes à partir de données expérimentales bruitées, puis nous avons étendu cette méthode pour prendre en compte une classification préalablement connue pour guider la construction de ces régions [CI-5].

- **ONTO-APP : Ontologie et Apprentissage pour un système de suggestions nutritionnelles personnalisées**

Partenaires : Lim&Bio (porteur), équipe A³ (LIPN), équipe déterminants, UREN (Unité de Recherche en Epidémiologie Nutritionnelle, UMR U557 Inserm/U1125 Inra/Cnam)

Le contexte général de ce travail est le développement d'une plateforme d'interrogation à base ontologique dédiée à la nutrition. Parmi les fonctionnalités de cette plateforme figure la possibilité d'obtenir, à partir d'une base de recettes, un ensemble de suggestions de recettes adaptées à un individu. Une suggestion doit être établie à partir du profil de l'individu, du bloc de description des recettes (titre, ingrédients, étapes de préparation, etc.) et de leurs caractéristiques nutritionnelles. La finalité de ce BQR etait i) la construction d'une ontologie du domaine de la nutrition, ii) son exploitation par des techniques d'apprentissage automatique symbolique dans le but de construire des catégories de recettes, et les descriptions symboliques de ces catégories.

Activités d recherche e indices de re connaissanc

- **PCU : Plate-forme de connaissance unifiée**

Partenaires : *Académiques* : LIPN, ESILV. *Industriels* : SMILE, ARMADILLO, QWAZR, PROXEM, OW2

Le projet PCU a pour but de simplifier et faciliter de bout en bout la valorisation des données existantes de l'entreprise, quel que soit son contexte d'origine (comportement client, gestion de connaissances) et d'exploitation (e-commerce mais aussi in-store, dans l'entreprise mais aussi mobile). Il s'appuie pour cela sur des innovations comme un traitement en flux de la donnée sur l'ensemble de son cycle (collecte, enrichissement, exploitation), de nouveaux algorithmes analysant le comportement utilisateur, une « usine » à moteurs de recherche multicanaux, et une architecture modulaire et extensible basée sur l'intégration des meilleures briques disponibles en Open Source.

Si dans le domaine visé de nombreuses briques Open Source de première qualité existent, aucune brique ne couvre l'intégralité de ce cycle de la donnée. Et à côté de cette fragmentation technologique, les solutions existantes du marché, d'ailleurs à forte dominante US, restent de plus réservées aux entreprises d'une certaine taille, du fait de leur coût mais aussi de la complexité de leur mise en œuvre.

Le projet « Plate-forme de connaissance unifiée » (PCU) vise donc à produire une plateforme open source industrielle de valorisation des données de l'entreprise afin d'en optimiser la valeur en tant qu'actif de l'entreprise mais aussi directement faciliter son exploitation. D'une manière générale, la plateforme PCU vise à répondre aux problèmes induits par la multiplication des silos de données au sein des entreprises et à permettre d'utiliser et de valoriser ces données, une fois unifiées et enrichies via des traitements sémantiques et d'apprentissage automatisé.

Activités de recherche et indices de reconnaissance

- **Hermès : Relation Client Personnalisée et Contextualisée**

Partenaires : *Académiques* : INRIA, LEM, LIFL, LIPN, LITIS. *Industriels* : Auchan, Blogbang, Brand Alley, Cylande, Keyneosoft, Leroy Merlin, Norsys, Numsight

Le projet Hermès adresse le champ du marketing one-to-one cross-canal et veut développer autant que possible la pertinence de l'information communiquée au client, dans le strict respect des règles d'éthique et de privacy. A partir des multiples sources hétérogènes susceptibles de délivrer de très nombreuses données sur le client (canaux traditionnels, appareils connectés, réseaux sociaux, etc.), le but du projet est d'en extraire les connaissances utiles au fil de l'eau. À partir de ces connaissances, les moteurs de recommandations permettent de déterminer le contexte d'achat dans lequel se situe le client (l'étape de son parcours d'achat, sa motivation, son projet, son état d'esprit, ses contraintes, son comportement d'achat, . . .), les propositions à faire au client, le contexte dans lequel elles doivent lui être délivrées (sur quel canal, à quel moment, . . .). Hermès a pour objectif de construire une plate forme de marketing contextuel générique, modulaire, industrialisée et automatisée capable :

— d'extraire des données provenant de tous les canaux (caisse, web, email, centre d'appel, réseaux sociaux, devices, etc.) ;
— de déterminer les contextes d'achat du client ainsi que les canaux pertinents d'interaction avec ce client ;
— de communiquer avec les systèmes opérationnels sur tous les canaux (notamment avec les devices en magasin et les réseaux sociaux, etc.).

- **UrbanD**

Partenaires : *Académiques* : Télécom ParisTech. *Industriels* : LAA, UfO, Thinkout, Preview

Le projet UrbanD consiste à développer un système de notation et de suivi (scoring and monitoring) destiné à augmenter la qualité de vie au sein des territoires urbains en vérifiant l'amélioration conjointe de réalisation des objectifs du développement durable. L'objectif innovant d'UrbanD est de lier qualité d'usages et développement durable en éclairant la compréhension des interactions positives et négatives au sein des éco-systèmes urbains.

Dans le cadre de ce projet, le LIPN a travaillé plus particulièrement sur le développement des approches automatiques pour l'analyse des rythmes des territoires. L'approche proposée consiste à modéliser la dynamique des territoires par un réseau complexe dynamique. Des approches de détection automatique de communautés sont utilisées afin de découvrir des rythmes dans des vastes territoires difficiles à analyser avec les approches manuelles habituellement employées par les urbanistes et les anthropologues [CO-2].

4.7 Indices de reconnaissance

4.7.1 Invitations des membres du LIPN

- **Tutoriels et tables rondes**

Tutoriels

— *Mining Attributed Networks*, 4th IEEE International conference on Data Science and Advanced Analytics, 19-21 October 2017,Tokyo : R. Kanawati, M. Atzmueller
— *Introduction à l'Apprentissage Profond*, Conference de Modélisation Mathématique et Informatique des Systemes Complexes, 2016 : Y. Chevaleyre
— *Topological and Graph-based clustering : Recent algorithmic advances*, ICONIP 2016, 19 October 2016, Kyoto, Japan : R. Kanawati, N. Grozavu
— *Topological and graph based Clustering : recent algorithmic advances*, ICONIP 2016, 19 October 2016, Kyoto, Japan : N. Grozavu, N. Kanawati
— *Topological Collaborative Learning*, 2nd ALML workshop, 2015, Kilarney, Ireland : N. Grozavu
— *Analyse de graphes de terrain pour la recherche d'information*, Journée IA et RI CNRS, Paris 1 Décembre 2015 : R. Kanawati
— *Multiplex network mining*, WI'2015, 6 December 2015, Singapore : R. Kanawati
— *Multiplex network mining*, WIMS'2015, 14 July 2015, Limassol, Cyprus : R. Kanawati
— *École de Printemps en Apprentissage arTificiel* (EPAT-2014) : N. Grozavu
— *Thematic Summer School FOCOLISE* : N. Grozavu
— *Preference Learning Problems*, ECML/PKDD 2014, Nancy, France : Yann Chevaleyre
— Workshop & Tutorial *Big Data Mining*, 16-18 Decembre 2014, Université Paris 6 : M. Lebbah
— *Prévision de liens dans les grands graphes de terrain : approches et applications*. RFIA'12, 24 Janvier 2012, Lyon, France : R. Kanawati, C. Rouveirol

Exposés invités

— *Découverte dans les réseaux biologiques hétérogènes : l'expérience Adalab*, Journées BIOSS-IA : Modèles logiques pour la représentation formelle des systèmes vivants, 22-23 juin 2017, Gif/Yvette : C. Rouveirol
— *Big Data as a challenge for Data Science*, Journée BIG DATA & DATA SCIENCE, 28 octobre 2016 à Tunis : Mustapha Lebbah
— *Le Deep Learning : un mystère mathématique révolutionnant l'intelligence artificielle*, juillet 2016, Centre CURI - USMBA, Fès : Y. Bennani
— *Big Data : sky is the limit*, BusIT Week 2016, 7-11 Mars 2016 : Mustapha Lebbah

— *Extraction de règles d'un graphe attribué*, mini-symposium on Instant data mining, interactive data mining, preference-based pattern mining, PEPS Préfute, à Rennes, octobre 2016 : Henry Soldano
— *Fouille de réseaux attribués*, colloque sur les Media Sociaux du PEPS EXIA, à l'université Paris Est Marne la vallée en octobre 2016 : Henry Soldano
— *Mythes et réalités sur l'exploitation du Big Data*, Ministère de l'enseignement supérieur, de la recherche scientifique et de la formation des cadres, 19-20 mars 2015, Rabat : Y. Bennani
— *Multiplex network analysis*, MARAMI/JFGG, 15 October 2015, Nimes, France : R. Kanawati
— *Multiplex network Analysis*, FRUMAM, Marseilles, France, 4 Feb. 2015 : R. Kanawati
— *Fouille de graphes attribués*, Journée sur les Grands Graphes (joint avec MARAMI 2014), octobre 2014, lip6 : H. Soldano
— *Abstraction et la connectivité en analyse formelle de concepts*, atelier FCA4AI de ECAI 2014, aout 2014 : H. Soldano
— *Data Science et IA*, Laboratoire des Technologies Innovantes de l'ENSA, Septembre 2014, Tanger, Morocco : Y. Bennani
— *Seed-centric algorithms for high performance community detection in complex networks*, Invited talk at AMLagap Workshop, 19 May 2014, Orléans, France : R. Kanawati
— *Détection de communautés dans les grands graphes de terrain* RIIMA Days , 16 avril 2014, USTHB, Alger : R. Kanawati
— *Analyse de graphes de terrain : application au calcul de recommendation*, Hermès workshop on recommendation systems, 27 mars 2014, Lille : R. Kanawati
— *Applying ensemble methods to community detection in complex networks*, Journée du pôle Math-Stic : Apprentissage et optimisation appliqués aux contenus numériques, 16 janvier 2014 Université Paris Nord, Villetaneuse : R. Kanawati
— *Apprentissage symbolique et statistique à l'ère du mariage pour tous*, RFIA 2014, Caen France : C. Rouveirol (avec S. Canu)
— *Prévision de liens dans des grands réseaux hétérogènes*, Séminaire AMISC, 21 mars 2013, Orléans : R. Kanawati
— *Programmation Logique Inductive*, JIAF/JFPC'12, Toulouse, France, 24 mai 2012 : Céline Rouveirol (avec Christel Vrain)
— Workshop on Scientific Data Preservation. Marseille 19-21 Novembre 2012 : Mustapha Lebbah

Activités de recherche et indices de reconnaissance

- **Invitations dans des laboratoires étrangers**

Prénom	*Nom*	*Lieu*	*Pays*	*Durée*
2016				
Nistor	Grozavu	Univ. Kobe	Japon	2 sem.
Younès	Bennani	UCA - Marrakech	Maroc	2 sem.
Basarab	Matei	Univ. Munich	Allemagne	2 sem.
Younès	Bennani	Centre CURI - USMBA - Fès	Maroc	2 sem.
2015				
Basarab	Matei	Univ. Buenos Aires	Argentine	2 sem.
Basarab	Matei	Univ. Nashville	États-Unis	2 sem.
Basarab	Matei	Univ. Trondheim	Norvège	2 sem.
2013				
Nistor	Grozavu	Vyatka State University, Kirov	Russie	1 mois
2012				
Younès	Bennani	ENSA, Tanger	Maroc	1 mois

Chapitre 5
Formation à la recherche

Sommaire

5.1 Masters, Écoles doctorale et post-doctorale 68

5.1.1 Cours de master 68

5.1.2 Ecoles thématiques - Organisations et Interventions 68

5.1.3 Autres cours 68

5.2 Thèses et Habilitations 68

5.2.1 Thèses en cours 68

5.2.2 Thèses soutenues 69

5.2.3 Habilitations à Diriger des Recherches soutenues 71

5.3 Stages de master, post-doctorats 71

5.3.1 Jurys de thèse et d'HDR 72

5.3.2 Post-doctorats 72

5.1 Masters, Écoles doctorale et post-doctorale

5.1.1 Cours de master

Le travail de recherche d'une grande partie des membres de l'équipe a débouché sur la création de nouveaux enseignements au niveau du Master d'Informatique en particulier dans la spécialité « Exploration Informatique des Données et Décisionnel » EID^2.

- *Factorisation matricielle*, master 1 Informatique, Paris 13 : Y. Bennani
- *Traitement Numérique des Données*, master 2 Informatique, Paris 13 : Y. Bennani
- *Informatique Décisionnelle et Data Mining*, Paris 13 : Y. Bennani
- *Apprentissage connexionniste (option)*, master 2 Informatique, spécialité EID2, Paris 13 : Y. Bennani
- Traitement numérique des données (tronc commun), master 2 Informatique, Paris 13 : Y. Bennani
- *Aide à la décision (option)*, master 2 Informatique, Paris 13 : Y. Chevaleyre
- *Réseaux sociaux (option)*, master 2 Informatique, Paris 13 : R. Kanawati
- *Apprentissage, Contraintes, Planification (option)*, master 2 Informatique, Paris 13 : C. Rouveirol, Y. Chevaleyre

- **Autres masters**

- M1 informatique à l'Université Paris 8, (option Big Data) : M. Lebbah
- Master 2 TRIED-Paris Saclay (M2 Traitement de l'Information et Exploitation des Données) pour la partie Big Data : M. Lebbah

5.1.2 Ecoles thématiques - Organisations et Interventions

Formation à la recherche

- ETH'17 : École Thematique d'Hiver « Machine Learning and Distributional Data », 13-16 Février 2017, Faculté des Sciences et Techniques de Marrakech.
- AAFD et SFC'16 : Conférence Internationale Francophone sur la Science des Données, 22-26 mai 2016 Marrakech (Maroc). Présidence du comité scientifique : Y. Bennani
- H2DM'16 : International Workshop on High Dimensional Data Mining 2016, 7 juin 2016 Naples (Italy). Co-organisateur : Y. Bennani
- ETA'16 : École thématique d'automne « Deep Learning and Data Science », 19-22 décembre 2016, USMBA-Centre CURI, Fès (Maroc). Co-organisateur : Y. Bennani
- EPAT'14 : École de Printemps sur l'Apprentissage artificiel, 7-12 juin 2014, Carry-le-Rouet (France). Co-présidence du comité scientifique : Y. Bennani

5.1.3 Autres cours

- Apprentissage Profond avec TensorFlow (Cours de 6h au centre IRD de Marrakech, 2017) : Y. Chevaleyre

5.2 Thèses et Habilitations

5.2.1 Thèses en cours

Nom Prénom	1ère insc.	Financement
Titre	*Directeur*	*Co-encadrant*
BECK Gael	01/03/16	CIFRE
Classification massivement distribuée. Application à la personnalisation web en temps-réel	*H. Azzag*	*M. Lebbah*

FALIH Issam		01/10/14	contrat doctoral Paris 13
	Détection de communautés dans les réseaux multiplexes	*Y. Bennani*	*R. Kanawati*
LÉGER Hippolyte		01/09/15	contrat doctoral Paris 13
	Apprentissage Relationnel Massif	*M. Lebbah*	*D. Bouthinon*
RASTIN Parisa		01/09/15	CIFRE
	Apprentissage non-supervisé à partir de données évolutives	*B. Matei*	*G. Cabanes*
RIFI Mouna		01/02/16	CIFRE EDF
	Analyse des réseaux multiplexes pour les études de sûreté.	*Y. Chevaleyre*	*R. Kanawati*
SARAZIN Tugdual		01/09/12	CIFRE Smokewatchers
		M. Lebbah	*H. Azzag*
VEILLON Lise-Marie		01/09/14	contrat doctoral Paris 13
	Apprentissage Artificiel Collectif	*H. Soldano*	*G. Bourgne*
ZEVIO Stella		15/09/2017	FUI Projet PCU
	Découverte et enrichissement de connaissances pour la recherche d'experts	*T. Charnois*	*G. Santini et H. Zargayouna*

5.2.2 Thèses soutenues

(depuis 2012)

Nom Prénom	Situation ultérieure	Soutenance	Mention
Titre		*Directeur*	*Co-encadrant*
Jury : Jury			
ALIZADEH Pegah	ATER	09/12/2016	
Elicitation et Planification dans les Processus de Decision de Markov avec Recompenses Incertaines		*Y. Chevaleyre*	
Jury : Jérôme Lang, Nicolas Maudet (rapp), Henry Soldano, Paolo Viappiani, Bruno Zanuttini (rapp)			
DOAN Nhat-Quang	MCF USTH, Vietnam	09/12/2013	
Modèles hiérarchiques et topologiques pour le partitionnement et la visualisation des données		*M. Lebbah*	*H. Azzag*
Jury : Cyrille Bertelle, , Yann Chevaleyre, Cyril De Runz, Gilles Venturini (rapp), Cédric Wemmert (rapp)			
BANNOUR Sondes	Ingénieur, GEOLSemantics	16/06/2015	
Apprentissage interactif de règles d'extraction d'information textuelle		*H. Soldano*	*L. Audibert*
Jury : Jean-Gabriel Ganascia, Adeline Nazarenko, Claire Nédellec (rapp), Pascale Sébillot (rapp)			
CHAIBI Amine	Ingénieur	29/11/2013	
Contribution en apprentissage topologique non supervisé pour la fouille de données		*M. Lebbah*	*H. Azzag*

5.2 Thèses et Habilitations

Jury : Gilles Bisson, Ndeye Niang Keita, Pascale Kuntz (rapp), Mohammed Nadif, Céline Rouveirol, Christel Vrain (rapp)

CHEBIL Ines — Data Scientist à Pixit.IO — 26/09/2014

Méthodes d'ensemble pour l'inférence de réseaux de régulation coopératifs — *C. Rouveirol*

Jury : Florence d'Alché-Buc (rapp), Jérôme Azé (rapp), Younes Bennani, Antoine Cornuéjols, Mohamed Elati

ESSAIDI Moez — Ingénieur — 02/07/2013

Model-Driven Data Warehouse and its Automation using Machine Learning Techniques — *C. Rouveirol* — *A. Osmani*

Jury : Jacky Akoka (rapp), Yann Chevaleyre (pdt), Nicolas Prat, Jean-Marc Vanel, Christel Vrain (rapp)

GHASSANY Mohamad — Ingénieur à Télécom Sud Paris — 07/11/2013

Contributions à l'apprentissage collaboratif non supervisé — *Y. Bennani*

Jury : Baydaa Al-Ayoubi (rapp), Antoine Cornuéjols (rapp), Christophe Fouqueré, Marc Gelgon, Nistor Grozavu, Cédric Wemmer

GHESMOUNE Mohammed — ATER — 25/11/2016

Apprentissage non supervisé de flux de données massives : Application aux Big Data d'assurance — *M. Lebbah* — *H. Azzag*

Jury : Salima Benbernou, Christophe Cérin, Marcin Detyniecki, Marc Gelgon (rapp), Céline Rouveirol, Allou Samé (rapp)

HAMDI Fatma — Ingénieur chez Sarenza — 06/12/2012

Apprentissage en distributions déséquilibrées — *Y. Bennani*

Jury : Khalid Benabdeslem, Christophe Fouqueré, Pascale Kuntz (rapp), Stéphane Lallich, Jean-Charles Lamirel (rapp), Vincent Lemaire, Vladimir Radevski (rapp)

JAZIRI Rakia — MCF à Paris 8 — 27/06/2013

Modèles de mélanges topologiques pour la classification de données structurées en séquences — *Y. Bennani* — *J-H. Chenot*

Jury : Faicel Chamroukhi, Antoine Cornuéjols, Haythem Elghazel, Pierre Gançarski (rapp), Mustapha Lebbah, Sylvie Thiria (rapp), Jean Luc Zarader (rapp)

Formation à la recherche

LE Van Minh — 13/12/2016

Machine Learning methods for optimization in Multi-Agent Decision Support System : application to Sign Placement for Tsunami Evacuation — *Y. Chevaleyre* — *J-D. Zucker*

Jury : Nicolas Brédéché, Alexis Drogoul, Benoît Gaudou, Mathieu Lacroix

MOUHOUBI Karima — 20/09/2013

Extraction de motifs dans des données bruitées — *C. Rouveirol* — *L. Létocart*

Jury : Younès Bennani, Jean-François Boulicaut, Clarisse Dhaenens (rapp), Mohamed Nadif (rapp)

PUJARI Manisha — 04/03/2015

Prévision de liens dans les grands graphes de terrain (Application aux réseaux bibliographiques) — *C. Rouveirol* — *R. Kanawati*

Jury : Aldo Gangemi, Bénédicte Le Grand (rapp), Christophe Prieur, Céline Robardet (rapp)

REDKO Ievgen — MCF INSA Lyon — 26/11/2015

Nonnegative Matrix Factorization for Transfer learning — *Y. Bennani*

Jury : Stéphane Canu (rapp), Christophe Fouqueré, Patrick Gallinari, Vincent Lemaire, Basarab Matei, Marc Sebban (rapp)

RODRIGUES Christophe — 21/01/2013

Apprentissage incrémental de modèles d'action relationnels — *C. Rouveirol* — *P. Gérard*

Jury : Yann Chevaleyre, Amal El Fallah Seghrouchni, Stuart Russell, Henry Soldano, Christel Vrain (rapp), Bruno Zanuttini (rapp)

YAKOUBI Zied — Data Scientist chez Influans — 04/12/2014

Détection et évaluation des communautés dans les réseaux complexes — *H. Soldano* — *R. Kanawati*

Jury : Hocine Cherifi, Michel Crampes (rapp), Lynda Tamine-Lechani (rapp), Roberto Wolfler Calvo

5.2.3 Habilitations à Diriger des Recherches soutenues

(depuis 2012)

Nom Prénom — Soutenance

Titre

Jury :

AZZAG Hanene

Classification non supervisée et visualisation de données : approches topologiques, hiérarchiques

Jury : Pierre Gancarski (rapp), Pacale Kuntz (pdt), Mustapha Lebbah, Guy Mélançon (rapp), Monique Noirhomme (rapp), Céline Rouveirol, Gilles Venturini

LEBBAH Mustapha — 27/01/12

Contributions en apprentissage non supervisé à partir de données complexes

Jury : Djamel Boucheffra (rapp), Marc Gelgon (rapp), Gérard Govaert (rapp), Djamel Abdelkader Zighed (rapp), Fouad Badran, Younès Benanni, Gérard Duchamp, Gilles Venturini, Céline Rouveirol

OSMANI Aomar — 04/06/12

Modélisation et raisonnement sur des données relationnelles

Jury : Yann Chevaleyre, Antoine Cornuéjols (rapp), Philippe Dague, Christophe Fouqueré, Gérard Ligozat, Debasis Mitra (rapp), Lorenza Saitta (rapp), Céline Rouveirol, Christel Vrain (rapp)

5.3 Stages de master, post-doctorats

- **Stages de master**

2017 Tien Tai Doan, Master 2 USTH, *Apprentissage profond pour le débruitage d'images médicales* : Y. Chevaleyre, S. Guerif

2017 Foutse Yuehgoh, Master 2 AIMS-Sénégal (Institut africain des sciences mathématiques) : C. Cérin, M. Lebbah, J-L Gaudiot

2017 Trong Le Phan, Master 2 USTH, *Modèle de mélanges pour l'extraction d'information massive* : M. Luong, M. Lebbah, H. Azzag

2016 Soumay Sellami, Master 2 Paris 1-CNAM, *Fouille de réseaux multiplexes pour l'analyse des interactions entre entreprises* : R. Kanawati et T. Auvery (CEPN)

2015 Parisa Rastin, Master EID2, *Sélection d'ensemble pour la détection de communauté* : R.Kanawati

2015 Gaël Beck, Master 2 informatique, Univ. Paris-Dauphine, *Modèles de mélanges et visualisation de données massives* : H. Azzag, M. Lebbah

2014 Issam Falih, Master ISI, Univ. Paris-Dauphine, *Détection de communautés dans des réseaux multiplexes* : R. Kanawati
2012 Tugdual Sarazin, Master EID2, *Biclustering et visualisation de grandes masses de données multidimensionnelles* : H. Azzag, M. Lebbah

- **Stages ingénieurs**

2017 Robin Lucas, 3ème année Sup Galilée, *Deep-Gene, un pipeline de traitement profond des séquences métagénomiques* : Y. Chevaleyre
2017 Naïm Kissi, 2ème année Sup-Galilée, *Implémentation et évaluation d'un algorithme d'apprentissage automatique pour la découverte automatique de connaissances biologiques* : A. Coutant
2016 Robin Lucas, 2eme année Sup-Galilée, *Fusion d'algorithmes d'apprentissage de modèles linéaires discrets* : Y. Chevaleyre
2016 Nesrine Bchini, 2eme année Sup Galilée, *Implémentation et comparaison d'algorithmes d'apprentissage automatique pour la découverte de connaissances biologiques* : A. Coutant, C. Rouveirol
2014 Gaël Beck, 2e année ingénieur à l'EISTI, *Visualisation de grandes masses de données : utilisation de la bibliothèque D3JS* : H. Azzag, M. Lebbah
2012 Y. Sonko, 2ème année Sup Galilée, *Évaluation de méthodes de classification automatique de recettes de cuisine*.

- **Autres stages et encadrements**

2016 Quan Cao Anh, L3 USTH, *Modèle de mélanges scalable* : H. Azzag, M. Lebbah, N-Q Doan

5.3.1 Jurys de thèse et d'HDR

Les membres de l'équipe A³ ont participé à 8 jurys d'HDR et 33 jurys de thèse depuis 2012.

5.3.2 Post-doctorats

L'équipe A³ a accueilli 16 post-doctorants durant la période. Le détail est donné dans la section 4.4.1.

Partie III
Acronymes

A^3 Apprentissage Artificiel et Applications

AERES Agence d'Évaluation de la Recherche et de l'Enseignement Supérieur

AM Allocataire Moniteur

ANR Agence Nationale de la Recherche

AOC Algorithmes et Optimisation Combinatoire

ATER Attaché Temporaire d'Enseignement et de Recherche

BQR Bonus Qualité Recherche (Université Paris 13)

CALIN Combinatoire, Algorithmique et Interactions

CEDRIC Centre d'Etudes et de Recherche en Informatique du CNAM - EA 1395

CIFRE Conventions Industrielles de Formation par la Recherche

CNAM Conservatoire National des Arts et Métiers

CNRS Centre National de la Recherche Scientifique

COMUE Communauté d'Universités et Établissements

EA Equipe d'Accueil

EDF Electricité de France

EFL Labex *Empirical Foundations of Linguistics*

FEDER Fonds Européen de Développement Régional

FR Fédération de Recherche

FUI Fonds Unique Interministériel

HCERES Haut Conseil de l'Évaluation de la Recherche et de l'Enseignement Supérieur

HDR Habilitation à Diriger des Recherches

IG Institut Galilée

INA Institut National de l'Audiovisuel

INCa Institut National du Cancer

IRIF Institut de Recherche en Informatique Fondamentale - UMR 8243

IUTV IUT de Villetaneuse

JCJC Jeunes Chercheuses et Jeunes Chercheurs

L2TI Laboratoire de traitement et transport de l'information, Université Paris 13 - EA 3043

LabEx Laboratoire d'Excellence (appel Grand Emprunt)

LAGA Laboratoire Analyse, Géométrie et Applications, Université Paris 13 - UMR 7539

LCR Logique, Calcul et Raisonnement

Lim&Bio Laboratoire d'Informatique Médicale et Bio-Informatique, université Paris 13 - EA

LIPADE Laboratoire d'Informatique Paris Descartes

LIPN Laboratoire d'Informatique de Paris-Nord, Université Paris 13 - UMR 7030

MAE Ministère des Affaires Etrangères

MathSTIC Fédération de Recherche MathSTIC, Université Paris 13 - FR 3734

MCF Maître de Conférences

MENESR Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche

PEPS Projet Exploratoire de Premier Soutien (CNRS)

PHC Programme Hubert Curien

PIA Programme d'Investissements d'Avenir

PU Professeur des Universités

RCLN Représentation des Connaissances et Langage Naturel

UMR Unité Mixte de Recherche

UP13 Université Paris 13

USTH Université des Sciences et des Technologies de Hanoi, http://usth.edu.vn/fr/