Activity Report 2015

Team MAGNET

Machine Learning in Information Networks

Inria teams are typically groups of researchers working on the definition of a common project, and objectives, with the goal to arrive at the creation of a project-team. Such project-teams may include other partners (universities or research institutions).

RESEARCH CENTER
Lille - Nord Europe

THEME
Data and Knowledge Representation and Processing

Table of contents

1. Members **1**
2. Overall Objectives **2**
3. Research Program **2**
3.1. Introduction 2
3.2. Beyond Vectorial Models for NLP 2
3.3. Adaptive Graph Construction 4
3.4. Prediction on Graphs and Scalability 5
3.5. Beyond Homophilic Relationships 6
4. Application Domains **6**
5. Highlights of the Year **7**
6. New Software and Platforms **7**
7. New Results **8**
7.1. Decentralized Estimation in Networks 8
7.2. Large-Scale Learning with Higher-Order Risk Functionals 8
7.3. Natural Language Processing 8
7.4. Some Ongoing Work 9
7.4.1. Metric Learning for Graph-based Label Propagation 9
7.4.2. Link Classification in Signed Graphs 9
7.4.3. Graph-based Learning for Dependency Parsing 9
8. Bilateral Contracts and Grants with Industry **9**
8.1. Bilateral Contracts with Industry 9
8.2. Bilateral Grants with Industry 10
8.2.1. Cifre Clic and Walk (2013-2016) 10
8.2.2. Cifre SAP (2011-2014) 10
9. Partnerships and Cooperations **10**
9.1. Regional Initiatives 10
9.2. National Initiatives 10
9.2.1. Competitivity Clusters 10
9.2.2. EFL 11
9.2.3. SCAGLIA 11
9.3. European Initiatives 11
9.4. International Initiatives 12
9.4.1. Inria Associate Teams not involved in an Inria International Labs 12
9.4.2. Inria International Partners 12
9.5. International Research Visitors 12
9.5.1. Visits of International Scientists 12
9.5.2. Visits to International Teams 12
10. Dissemination **13**
10.1. Promoting Scientific Activities 13
10.1.1. Scientific events organisation 13
10.1.2. Scientific events selection 13
10.1.2.1. General chair, scientific chair 13
10.1.2.2. Chair of conference program committees 13
10.1.2.3. Member of the conference program committees 13
10.1.2.4. Other 13
10.1.3. Journal 13
10.1.3.1. Member of the editorial boards 13
10.1.3.2. Reviewer - Reviewing activities 13
10.1.4. Invited talks 13

10.1.5. Scientific expertise 14
10.2. Teaching - Supervision - Juries 14
10.2.1. Teaching 14
10.2.2. Supervision 15
10.2.3. Juries 15
10.3. Popularization 16
11. Bibliography **16**

Team MAGNET

Creation of the Team: 2013 January 01

Keywords:

Computer Science and Digital Science:

3.3. - Data and knowledge analysis
3.4.1. - Supervised learning
3.4.2. - Unsupervised learning
3.4.4. - Optimization and learning
3.5.2. - Recommendation systems
7.2. - Discrete mathematics, combinatorics
8.2. - Machine learning
8.4. - Natural language processing

Other Research Topics and Application Domains:

6.3.1. - Web
6.3.4. - Social Networks
6.5. - Information systems
9.4.5. - Data science

1. Members

Research Scientists

Aurélien Bellet [Inria, Researcher, from Nov 2015]
Pascal Denis [Inria, Researcher]
Jan Ramon [Inria, Senior Researcher, from Oct 2015]

Faculty Members

Marc Tommasi [Team leader, Univ. Lille III, Professor, HdR]
Rémi Gilleron [Univ. Lille III, Professor, HdR]
Mikaela Keller [Univ. Lille III, Associate Professor]
Fabien Torre [Univ. Lille III, Associate Professor]
Fabio Vitale [Univ. Lille III, Associate Professor]

PhD Students

David Chatel [Conseil Régional du Nord-Pas de Calais, Inria]
Mathieu Dehouck [Univ. Lille I, from Oct 2015]
Géraud Le Falher [Inria]
Pauline Wauquier [Cifre Clic and Walk]

Visiting Scientists

Claudio Gentile [Università dell'Insubria, Jul 2015]
Mark Herbster [University College London, Jan and May 2015]

Administrative Assistant

Julie Jonas [Inria]

2. Overall Objectives

2.1. Presentation

MAGNET is a research group that aims to design new machine learning based methods geared towards mining information networks. Information networks are large collections of interconnected data and documents like citation networks and blog networks among others. Our goal is to propose new prediction methods for texts and networks of texts based on machine learning algorithms in graphs. Such algorithms include node and link classification, link prediction, clustering and probabilistic modeling of graphs. We aim to tackle real-world problems such as browsing, monitoring and recommender systems, and more broadly information extraction in information networks. Application domains cover natural language processing, social networks for cultural data and e-commerce, and biomedical informatics.

3. Research Program

3.1. Introduction

The main objective of MAGNET is to develop original machine learning methods for networked data in order to build applications like browsing, monitoring and recommender systems, and more broadly information extraction in information networks. We consider information networks in which the data consist of both feature vectors and texts. We model such networks as (multiple) (hyper)graphs wherein nodes correspond to entities (documents, spans of text, users, ...) and edges correspond to relations between entities (similarity, answer, co-authoring, friendship, ...). Our main research goal is to propose new on-line and batch learning algorithms for various problems (node classification / clustering, link classification / prediction) which exploit the relationships between data entities and, overall, the graph topology. We are also interested in searching for the best hidden graph structure to be generated for solving a given learning task. Our research will be based on generative models for graphs, on machine learning for graphs and on machine learning for texts. The challenges are the dimensionality of the input space, possibly the dimensionality of the output space, the high level of dependencies between the data, the inherent ambiguity of textual data and the limited amount of human labeling. An additional challenge will be to design scalable methods for large information networks. Hence, we will explore how sampling, randomization and active learning can be leveraged to improve the scalability of the proposed algorithms.

Our research program is organized according to the following questions:

1. How to go beyond vectorial classification models in Natural Language Processing (NLP) tasks?
2. How to adaptively build graphs with respect to the given tasks? How to create networks from observations of information diffusion processes?
3. How to design methods able to achieve a good trade-off between predictive accuracy and computational complexity?
4. How to go beyond strict node homophilic/similarity assumptions in graph-based learning methods?

3.2. Beyond Vectorial Models for NLP

One of our overall research objectives is to derive graph-based machine learning algorithms for natural language and text information extraction tasks. This section discusses the motivations behind the use of graph-based ML approaches for these tasks, the main challenges associated with it, as well as some concrete projects. Some of the challenges go beyond NLP problems and will be further developed in the next sections. An interesting aspect of the project is that we anticipate some important cross-fertilizations between NLP and ML graph-based techniques, with NLP not only benefiting from but also pushing ML graph-based approaches into new directions.

Motivations for resorting to graph-based algorithms for texts are at least threefold. First, online texts are organized in networks. With the advent of the web, and the development of forums, blogs, and micro-blogging, and other forms of social media, text productions have become strongly connected. Interestingly, NLP research has been rather slow in coming to terms with this situation, and most of the literature still focus on document-based or sentence-based predictions (wherein inter-document or inter-sentence structure is not exploited). Furthermore, several multi-document tasks exist in NLP (such as multi-document summarization and cross-document coreference resolution), but most existing work typically ignore document boundaries and simply apply a document-based approach, therefore failing to take advantage of the multi-document dimension [33], [36].

A second motivation comes from the fact that most (if not all) NLP problems can be naturally conceived as graph problems. Thus, NLP tasks often involve discovering a relational structure over a set of text spans (words, phrases, clauses, sentences, etc.). Furthermore, the *input* of numerous NLP tasks is also a graph; indeed, most end-to-end NLP systems are conceived as pipelines wherein the output of one processor is in the input of the next. For instance, several tasks take POS tagged sequences or dependency trees as input. But this structured input is often converted to a vectorial form, which inevitably involves a loss of information.

Finally, graph-based representations and learning methods appear to address some core problems faced by NLP, such as the fact that textual data are typically not independent and identically distributed, they often live on a manifold, they involve very high dimensionality, and their annotations is costly and scarce. As such, graph-based methods represent an interesting alternative to, or at least complement, structured prediction methods (such as CRFs or structured SVMs) commonly used within NLP. Graph-based methods, like label propagation, have also been shown to be very effective in semi-supervised settings, and have already given some positive results on a few NLP tasks [12], [38].

Given the above motivations, our first line of research will be to investigate how one can leverage an underlying network structure (e.g., hyperlinks, user links) between documents, or text spans in general, to enhance prediction performance for several NLP tasks. We think that a “network effect”, similar to the one that took place in Information Retrieval (with the Page Rank algorithm), could also positively impact NLP research. A few recent papers have already opened the way, for instance in attempting to exploit Twitter follower graph to improve sentiment classification [37].

Part of the challenge here will be to investigate how adequately and efficiently one can model these problems as instances of more general graph-based problems, such as node clustering/classification or link prediction discussed in the next sections. In a few cases, like text classification or sentiment analysis, graph modeling appears to be straightforward: nodes correspond to texts (and potentially users), and edges are given by relationships like hyperlinks, co-authorship, friendship, or thread membership. Unfortunately, modeling NLP problems as networks is not always that obvious. From the one hand, the right level of representation will probably vary depending on the task at hand: the nodes will be sentences, phrases, words, etc. From the other hand, the underlying graph will typically not be given a priori, which in turn raises the question of how we construct it. A preliminary discussion of the issue of optimal graph construction for semi-supervised learning in NLP is given in [12], [41]. We identify the issue of adaptive graph construction as an important scientific challenge for machine learning on graphs in general, and we will discuss it further in Section 3.3.

As noted above, many NLP tasks have been recast as structure prediction problems, allowing to capture (some of the) output dependencies. How to best combine structured output and graph-based ML approaches is another challenge that we intend to address. We will initially investigate this question within a semi-supervised context, concentrating on graph regularization and graph propagation methods. Within such approaches, labels are typically binary or they correspond to small finite set. Our objective is to explore how one propagates an exponential number of *structured labels* (like a sequence of tags or a dependency tree) through graphs. Recent attempts at blending structured output models with graph-based models are investigated in [38], [23]. Another related question that we will address in this context is how does one learn with *partial labels* (like partially specified tag sequence or tree) and use the graph structure to complete the output structure. This last question is very relevant to NLP problems where human annotations are costly; being able to learn from partial annotations could therefore allow for more targeted annotations and in turn reduced costs [25].

The NLP tasks we will mostly focus on are coreference resolution and entity linking, temporal structure prediction, and discourse parsing. These tasks will be envisioned in both document and cross-document settings, although we expect to exploit inter-document links either way. Choices for these particular tasks is guided by the fact that are still open problems for the NLP community, they potentially have a high impact for industrial applications (like information retrieval, question answering, etc.), and we already have some expertise on these tasks in the team (see for instance [24], [19], [22]). As a midterm goal, we also plan to work on tasks more directly relating to micro-blogging, such sentiment analysis and the automatic thread structuring of technical forums; the latter task is in fact an instance of rhetorical structure prediction [40]. We have already initiated some work on the coreference resolution with graph-based learning, by casting the problem as an instance of spectral clustering [22].

3.3. Adaptive Graph Construction

In most applications, edge weights are computed through a complex data modeling process and convey crucially important information for classifying nodes, making it possible to infer information related to each data sample even exploiting the graph topology solely. In fact, a widespread approach to several classification problems is to represent the data through an undirected weighted graph in which edge weights quantify the similarity between data points. This technique for coding input data has been applied to several domains, including classification of genomic data [35], face recognition [21], and text categorization [28].

In some cases, the full adjacency matrix is generated by employing suitable similarity functions chosen through a deep understanding of the problem structure. For example TF-IDF representation of documents, the affinity between pairs of samples is often estimated through the cosine measure or the χ^2 distance. After the generation of the full adjacency matrix, the second phase for obtaining the final graph consists in an edge sparsification/reweighting operation. Some of the edges of the clique obtained in the first step are pruned and the remaining ones can be reweighted to meet the specific requirements of the given classification problem. Constructing a graph with these methods obviously entails various kinds of loss of information. However, in problems like node classification, the use of graphs generated from several datasets can lead to an improvement in accuracy ([42], [13], [14]). Hence, the transformation of a dataset into a graph may, at least in some cases, partially remove various kinds of irregularities present in the original datasets, while keeping some of the most useful information for classifying the data samples. Moreover, it is often possible to accomplish classification tasks on the obtained graph using a running time remarkably lower than is needed by algorithms exploiting the initial datasets, and a suitable sparse graph representation can be seen as a compressed version of the original data. This holds even when input data are provided in a online/stream fashion, so that the resulting graph evolves over time.

In this project we will address the problem of adaptive graph construction towards several directions. The first one is about how to choose the best similarity measure given the objective learning task. This question is related to the question of metric and similarity learning ([15], [16]) which has not been considered in the context of graph-based learning. In the context of structured prediction, we will develop approaches where output structures are organized in graphs whose similarity is given by top-k outcomes of greedy algorithms.

A different way we envision adaptive graph construction is in the context of semi-supervised learning. Partial supervision can take various forms and an interesting and original setting is governed by two currently studied applications: detection of brain anomaly from connectome data and polls recommendation in marketing. Indeed, for these two applications, a partial knowledge of the information diffusion process can be observed while the network is unknown or only partially known. An objective is to construct (or complete) the network structure from some local diffusion information. The problem can be formalized as a graph construction problem from partially observed diffusion processes. It has been studied very recently in [30]. In our case, the originality comes either from the existence of different sources of observations or from the large impact of node contents in the network.

We will study how to combine graphs defined by networked data and graphs built from flat data to solve a given task. This is of major importance for information networks because, as said above, we will have to deal with multiple relations between entities (texts, spans of texts, ...) and also use textual data and vectorial data.

3.4. Prediction on Graphs and Scalability

As stated in the previous sections, graphs as complex objects provide a rich representation of data. Often enough the data is only partially available and the graph representation is very helpful in predicting the unobserved elements. We are interested in problems where the complete structure of the graph needs to be recovered and only a fraction of the links is observed. The link prediction problem falls into this category. We are also interested in the recommendation and link classification problems which can be seen as graphs where the structure is complete but some labels on the links (weights or signs) are missing. Finally we are also interested in labeling the nodes of the graph, with class or cluster memberships or with a real value, provided that we have (some information about) the labels for some of the nodes.

The semi-supervised framework will be also considered. A midterm research plan is to study how graph regularization models help for structured prediction problems. This question will be studied in the context of NLP tasks, as noted in Section 3.2, but we also plan to develop original machine learning algorithms that have a more general applicability. Inputs are networks whose nodes (texts) have to be labeled by structures. We assume that structures lie in some manifold and we want to study how labels can propagate in the network. One approach is to find a smooth labeling function corresponding to an harmonic function on both manifolds in input and output.

Scalability is one of the main issues in the design of new prediction algorithms working on networked data. It has gained more and more importance in recent years, because of the growing size of the most popular networked data that are now used by millions of people. In such contexts, learning algorithms whose computational complexity scales quadratically, or slower, in the number of considered data objects (usually nodes or edges, depending on the task) should be considered impractical.

These observations lead to the idea of using graph sparsification techniques in order to work on a part of the original network for getting results that can be easily extended and used for the whole original input. A sparsified version of the original graph can often be seen as a subset of the initial input, i.e. a suitably selected input subgraph which forms the training set (or, more in general, it is included in the training set). This holds even for the active setting. A simple example could be to find a spanning tree of the input graph, possibly using randomization techniques, with properties such that we are allowed to obtain interesting results for the initial graph dataset. We have started to explore this research direction for instance in [39].

At the level of mathematical foundations, the key issue to be addressed in the study of (large-scale) random networks also concerns the segmentation of network data into sets of independent and identically distributed observations. If we identify the data sample with the whole network, as it has been done in previous approaches [29], we typically end up with a set of observations (such as nodes or edges) which are highly interdependent and hence overly violate the classic i.i.d. assumption. In this case, the data scale can be so large and the range of correlations can be so wide, that the cost of taking into account the whole data and their dependencies is typically prohibitive. On the contrary, if we focus instead on a set of subgraphs independently drawn from a (virtually infinite) target network, we come up with a set of independent and identically distributed observations—namely the subgraphs themselves, where subgraph sampling is the underlying ergodic process [17]. Such an approach is one principled direction for giving novel statistical foundations to random network modeling. At the same time, because one shifts the focus from the whole network to a set of subgraphs, complexity issues can be restricted to the number of subgraphs and their size. The latter quantities can be controlled much more easily than the overall network size and dependence relationships, thus allowing to tackle scalability challenges through a radically redesigned approach.

We intend to develop new learning models for link prediction problems. We have already proposed a conditional model in [27] with statistics based on Fiedler values computed on small subgraphs. We will investigate the use of such a conditional model for link prediction. We will also extend the conditional probabilistic models to the case of graphs with textual and vectorial data by defining joint conditional models. Indeed, an important challenge for information networks is to introduce node contents in link ranking and link prediction methods that usually rely solely on the graph structure. A first step in this direction was already proposed in [26] where we learn a mapping of node content to a new representation constrained by the existing link structure and applied it for link recommendation. This approach opens a different view on recommendation by means of link ranking, for which we think nonparametric approaches should be fruitful.

Regarding link classification problems, we plan to devise a whole family of active learning strategies, which could be based on spanning trees or sparse input subgraphs, that exploit randomization and the structure of the graph in order to offset the adversarial label assignment. We expect these active strategies to exhibit good accuracies with a remarkably small number of queried edges, where passive learning methods typically break down. The theoretical findings can be supported by experiments run on both synthetic and real-world (Slashdot, Epinions, Wikipedia, and others) datasets. We are also interested in studying generative models for graph labeling, exploiting the results obtained in p-stochastic model for link classification (see [20]) and statistical model for node label assignment which can be related to tree-structured Markov random fields [31].

3.5. Beyond Homophilic Relationships

In many cases, algorithms for solving node classification problems are driven by the following assumption: linked entities tend to be assigned to the same class. This assumption, in the context of social networks, is known as homophily ([18], [32]) and involves ties of every type, including friendship, work, marriage, age, gender, and so on. In social networks, homophily naturally implies that a set of individuals can be parted into subpopulations that are more cohesive. In fact, the presence of homogeneous groups sharing common interests is a key reason for affinity among interconnected individuals, which suggests that, in spite of its simplicity, this principle turns out to be very powerful for node classification problems in general networks.

Recently, however, researchers have started to consider networked data where connections may also carry a negative meaning. For instance, disapproval or distrust in social networks, negative endorsements on the Web. Although the introduction of signs on graph edges appears like a small change from standard weighted graphs, the resulting mathematical object, called signed graph, has an unexpectedly rich additional complexity. For example, their spectral properties, which essentially all sophisticated node classification algorithms rely on, are different and less known than those of their unsigned counterparts. Signed graphs naturally lead to a specific inference problem that we have discussed in previous sections: link classification. This is the problem of predicting the sign of links in a given graph. In online social networks, this may be viewed as a form of sentiment analysis, since we would like to semantically categorize the relationship between individuals.

Another way to go beyond homophily between entities will be studied using our recent model of hypergraphs with bipartite hyperedges [34]. A bipartite hyperedge connects two ends which are disjoint subsets of nodes. Bipartite hyperedges is a way to relate two collections of (possibly heterogeneous) entities represented by nodes. In the NLP setting, while hyperedges can be used to model bags of words, bipartite hyperedges are associated with relationships between bags of words. But each end of bipartite hyperedges is also a way to represent complex entities, gathering several attribute values (nodes) into hyperedges viewed as records. Our hypergraph notion naturally extends directed and undirected weighted graph. We have defined a spectral theory for this new class of hypergraphs and opened a way to smooth labeling on sets of nodes. The weighting scheme allows to weigh the participation of each node to the relationship modeled by bipartite hyperedges accordingly to an equilibrium condition. This condition provides a competition between nodes in hyperedges and allows interesting modeling properties that go beyond homophily and similarity over nodes (the theoretical analysis of our hypergraphs exhibits tight relationships with signed graphs). Following this competition idea, bipartite hyperedges are like matches between two teams and examples of applications are team creation. The basic tasks we are interested in are hyperedge classification, hyperedge prediction, node weight prediction. Finally, hypergraphs also represent a way to summarize or compress large graphs in which there exists highly connected couples of (large) subsets of nodes.

4. Application Domains

4.1. Overview

The real-world problems we target include browsing, monitoring and mining in information networks. The discovered structures would also be beneficial to predicting links between users and texts which is at the core of recommender systems. More generally, all the learning tasks considered in the project such as node clustering, node and link classification and link prediction are likely to yield important improvements in these applications. Application domains cover natural language processing, social networks for cultural data and e-commerce, and biomedical informatics.

5. Highlights of the Year

5.1. Highlights of the Year

We have published two papers at NIPS [3], [6], the leading conference in machine learning. The first paper presents novel results on large-scale learning with higher-order risk functionals, which has applications in link prediction, graph inference and metric learning (among others). The second paper proposes new gossip algorithms for decentralized estimation of pairwise statistics in networks.

We have published a paper at AAAI [4], one of the top conferences in Artificial Intelligence. The contribution is a new structured model for learning anaphoricity detection and coreference resolution, which achieved the best score to date on the popular CoNLL benchmark with gold mentions.

We have published a paper at EMNLP [2], a leading conference in Natural Language Processing. The work presents a detailed comparative framework for assessing the usefulness of popular unsupervised word representations for identifying so-called implicit discourse relations.

6. New Software and Platforms

6.1. CoRTex

Python library for noun phrase COreference Resolution in natural language TEXts
FUNCTIONAL DESCRIPTION

CoRTex is a LGPL-licensed Python library for Noun Phrase coreference resolution in natural language texts. This library contains implementations of various state-of-the-art coreference resolution algorithms, including those developed in our research. In addition, it provides a set of APIs and utilities for text preprocessing, reading the main annotation formats (ACE, CoNLL and MUC), and performing evaluation based on the main evaluation metrics (MUC, B-CUBED, and CEAF). As such, CoRTex provides benchmarks for researchers working on coreference resolution, but it is also of interest for developers who want to integrate a coreference resolution within a larger platform.

- Participants: Pascal Denis and David Chatel
- Contact: Pascal Denis
- URL: https://gforge.inria.fr/projects/cortex/

7. New Results

7.1. Decentralized Estimation in Networks

In [3], we studied the problem of decentralized estimation in networks, where each node of the network holds a data point and the goal is to estimate some statistics on the entire data under communication constraints imposed by the graph topology of the network. This generic problem has many applications in Internet of Things as well as for extracting knowledge from massive information graphs such as interlinked Web documents and online social media. In this work, we focused on estimating pairwise mean statistics. Popular examples of such statistics include the sample variance, the average distance and the Area Under the ROC Curve, among others. We proposed new synchronous and asynchronous randomized gossip algorithms which simultaneously propagate data across the network and maintain local estimates of the quantity of interest. We establish convergence rate bounds of $O(1/t)$ and $O(\log t/t)$ for the synchronous and asynchronous cases respectively, where t is the number of iterations, with explicit data and network dependent terms. Beyond favorable comparisons in terms of rate analysis, numerical experiments provide empirical evidence the proposed algorithms surpasses the previously introduced approach.

7.2. Large-Scale Learning with Higher-Order Risk Functionals

In [6], we studied learning problems where the performance criterion consists of an average over tuples (e.g., pairs or triplets) of observations rather than over individual observations, as in many learning problems involving networked data (e.g., link prediction), but also in metric learning and ranking. In this setting, the empirical risk to be optimized takes the form of a U-statistic, and its terms are highly dependent and thus violate the classic i.i.d. assumption. In this work, we focused on how to best implement a stochastic approximation approach to solve such risk minimization problems in the large-scale setting. We argue that gradient estimates should be obtained by sampling tuples of data points with replacement (incomplete U-statistics) rather than sampling data points without replacement (complete U-statistics based on subsamples). We develop a theoretical framework accounting for the substantial impact of this strategy on the generalization ability of the prediction model returned by the Stochastic Gradient Descent (SGD) algorithm. It reveals that the method we promote achieves a much better trade-off between statistical accuracy and computational cost. Beyond the rate bound analysis, we provide strong empirical evidence of the superiority of the proposed approach on metric learning and ranking problems.

7.3. Natural Language Processing

In [4], we introduce a new structured model for learning anaphoricity detection and coreference resolution in a joint fashion. Specifically, we use a latent tree to represent the full coreference and anaphoric structure of a document at a global level, and we jointly learn the parameters of the two models using a version of the structured perceptron algorithm. Our joint structured model is further refined by the use of pairwise constraints which help the model to capture accurately certain patterns of coreference. Our experiments on the CoNLL-2012 English datasets show large improvements in both coreference resolution and anaphoricity detection, compared to various competing architectures. Our best coreference system obtains a CoNLL score of 81.97 on gold mentions, which is to date the best score reported on this setting.

In [2], we present a detailed comparative framework for assessing the usefulness of unsupervised word representations for identifying so-called implicit discourse relations. Specifically, we compare standard one-hot word pair representations against low-dimensional ones based on Brown clusters and word embeddings. We also consider various word vector combination schemes for deriving discourse segment representations from word vectors, and compare representations based either on all words or limited to head words. Our main finding is that denser representations systematically outperform sparser ones and give state-of-the-art performance or above without the need for additional hand-crafted features.

7.4. Some Ongoing Work

7.4.1. Metric Learning for Graph-based Label Propagation

The efficiency of graph-based semi-supervised algorithms depends on the graph of instances on which they are applied. The instances are often in a vectorial form before a graph linking them is built. The construction of the graph relies on a metric over the vectorial space that helps define the weight of the connection between entities. The typical choice for this metric is usually a distance or a similarity measure based on the Euclidean norm. We claim that in some cases the Euclidean norm on the initial vectorial space might not be the most appropriate to solve the task efficiently.

In a paper currently under review, we proposed an algorithm that aims at learning the most appropriate vectorial representation for building a graph on which label propagation is solved efficiently, with theoretical guarantees on the classification performance.

7.4.2. Link Classification in Signed Graphs

We worked on active link classification in signed graphs. Namely, the idea is to build a spanning tree of the graph and query all its edge signs. In the two clusters case, this allows to predict the sign of an edge between nodes u and v as the product of the signs of edge along the path in the spanning tree from u to v. It turns out that ensuring low error rate amounts to minimizing the stretch, a long open standing problem known as Low Stretch Spanning Tree [11]. While we are still working on the theoretical analysis, experimental results showed that our construction is generally competitive with a simple yet efficient baseline and outperforms it for specific graph geometry like grid graphs.

Moreover, based on experimental observations, we will also analyze a heuristic which exhibits good performance at a very low computational cost and is therefore well suited for large-scale graphs. In a nutshell, it predicts the sign of an edge from u to v based on the fraction of u negative outgoing edges and v negative incoming edges, exploiting a behavioral consistency bias from signed social network users.

Going further in link classification, we believe that the notion of sign can be extended, going from one binary label per edge to a more holistic approach where the similarity between two nodes is measured across different contexts. These contexts are represented by vectors whose dimension matches the dimension of unknown feature vectors associated with each node. The goal is to answer queries of the form: how similar are nodes u and v along a specific context? We first plan to validate the relevance of this modeling on real-world problems, then test baseline methods on synthetic and real data before looking for a more effective, online prediction method.

7.4.3. Graph-based Learning for Dependency Parsing

We are investigating the use of different graph-based learning techniques such as k-nearest neighbors classification and label propagation for the problem of dependency parsing. While most of current approaches rely on learning a single scoring model (through SVM, MIRA, neural networks) from a large set of hand annotated training data (usually thousands of sentences), we are interested in using the sentence space geometry (approximated via a similarity graph over some labeled and unlabeled sentences) to tune the model to better fit a given sentence. This amounts to learning a slightly different model for each unlabeled sentence.

In order to successfully parse sentences in this setting, we need to propagate parsing information from labeled sentences to unlabeled ones through the graph. In order to build a similarity graph well suited to dependency parsing, we worked on learning a similarity function between pairs of sentences, based on the idea that two sentences are similar if they have similar parse trees. We will then investigate how to propagate the trees (which may be of varying sizes) through the graph and consider several propagation schemes.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. KeyCoopt (2015)

Participants: Rémi Gilleron [correspondent], François Noyez, Fabien Torre.

We have a bilateral contract with the KEYCOOPT company. The goal of the company is to suggest candidates for job offers. For this, the company has a large pool of referrers, also named coopters. The process is: given a job offer, some coopters are selected, each coopter may suggest a candidate, the proposed candidates are selected by KEYCOOPT and some candidates are proposed in answer to the job offer. We propose a machine learning based method for selecting coopters given a job offer. The method is a ranking algorithm using support vector machines (SVMRank). It has been developed and tested and can be integrated in the information system of KEYCOOPT. Possible improvements are to use natural language processing methods in order to use texts as texts for job offers, and to use the network of coopters.

8.2. Bilateral Grants with Industry

8.2.1. Cifre Clic and Walk (2013-2016)

Participants: Mikaela Keller [correspondent], Pauline Wauquier, Marc Tommasi.

We have a one to one cooperation with the CLIC AND WALK company that makes marketing surveys by consumers (called clicwalkers). The goal of the company is to understand the community of clicwalkers (40 thousands in one year) and its evolution with two objectives: the first one is to optimize the attribution of surveys to clicwalkers, and the second is to expand company's market to foreign countries. Social data can be obtained from social networks (G+, Facebook, ...) but there is no explicit network to describe the clicwalkers community. But users activity in answering surveys as well as server logs can provide traces of information diffusion, geolocation data, temporal data, sponsorship, etc. We will study the problem of adaptive graph construction from the clicwalkers network. Node (users) classification and clustering algorithms will be applied. For the problem of survey recommendations, the problem of teams constitution in a bipartite graphs of users and surveys will be studied. Random graph modeling and generative models of random graphs will be one step towards the prediction of the evolution of clicwalkers community.

8.2.2. Cifre SAP (2011-2014)

Participants: Rémi Gilleron [correspondent], Marc Tommasi, Thomas Ricatte.

The PhD defense of Thomas Ricatte was held in Lille on January 23th 2015.

9. Partnerships and Cooperations

9.1. Regional Initiatives

MIKAELA KELLER participated in the joint Inria Campus-Institut Pasteur workshop whose goal was to reinforce the collaboration between both institutes.

MARC TOMMASI belongs to the drafting committee of the Lille IDEX project, and is a representative for the COMUE in the DAS commission “Ubiquitaire et Internet des Objets”.

MARC TOMMASI and PASCAL DENIS supervise the PhD thesis of DAVID CHATEL on semi-supervised spectral clustering. The PhD is funded by Inria and the “Région Nord - Pas de Calais”.

9.2. National Initiatives

9.2.1. Competitivity Clusters

We are part of FUI HERMES (2012-2015), a joint project in collaboration with many companies (Auchan, KeyneSoft, Cylande, ...). The main objective is to develop a platform for contextual customer relation management. The project started in November 2012.

9.2.2. EFL

PASCAL DENIS is an associate member of the Laboratoire d'Excellence *Empirical Foundations of Linguistics* (EFL), http://www.labex-efl.org/.

9.2.3. SCAGLIA

The project SCAGLIA (Scalable Graph Algorithms for Learning in Networked Data) of FABIO VITALE was accepted at the JCJC INS2I 2015 call.

9.3. European Initiatives

9.3.1. Collaborations in European Programs, except FP7 & H2020

Program: ERC Advanced Grant

Project acronym: STAC

Project title: Strategic conversation

Duration: Sep. 2011 - Aug. 2016

Coordinator: Nicholas Asher, CNRS, Université Paul Sabatier, IRIT (France)

Other partners: School of Informatics, Edinburgh University; Heriot Watt University, Edinburgh

Abstract: STAC is a five year interdisciplinary project that aims to develop a new, formal and robust model of conversation, drawing from ideas in linguistics, philosophy, computer science and economics. The project brings a state of the art, linguistic theory of discourse interpretation together with a sophisticated view of agent interaction and strategic decision making, taking advantage of work on game theory.

Program: COST Action

Project acronym: TextLink

Project title: Structuring Discourse in Multilingual Europe

Duration: Apr. 2014 - Apr. 2018

Coordinator: Prof. Liesbeth Degand, Université Catholique de Louvain, Belgium

Other partners: 26 EU countries and 3 international partner countries (Argentina, Brazil, Canada)

Abstract: Effective discourse in any language is characterized by clear relations between sentences and coherent structure. But languages vary in how relations and structure are signaled. While monolingual dictionaries and grammars can characterize the words and sentences of a language and bilingual dictionaries can do the same between languages, there is nothing similar for discourse. For discourse, however, discourse-annotated corpora are becoming available in individual languages. The Action will facilitate European multilingualism by (1) identifying and creating a portal into such resources within Europe - including annotation tools, search tools, and discourse-annotated corpora; (2) delineating the dimensions and properties of discourse annotation across corpora; (3) organizing these properties into a sharable taxonomy; (4) encouraging the use of this taxonomy in subsequent discourse annotation and in cross-lingual search and studies of devices that relate and structure discourse; and (5) promoting use of the portal, its resources and sharable taxonomy. With partners from across Europe, TextLink will unify numerous but scattered linguistic resources on discourse structure. With its resources searchable by form and/or meaning and a source of valuable correspondences, TextLink will enhance the experience and performance of human translators, lexicographers, language technology and language learners alike.

9.4. International Initiatives

9.4.1. Inria Associate Teams not involved in an Inria International Labs

Program: Inria North-European Labs

Project acronym: RSS

Project title: Rankings and Similarities in Signed graphs

Duration: late 2015 to late 2017

Partners: Aristides Gionis (Data Mining Group, Aalto University, Finland) and Mark Herbster (Centre for Computational Statistics and Machine Learning, University College London, UK)

Abstract: The project focuses on predictive analysis of networked data represented as signed graphs, where connections can carry either a positive or a negative semantic. The goal of this associate team is to derive novel formal methods and machine learning algorithms towards link classification and link ranking in signed graphs and assess their performance in both theoretical and practical terms.

9.4.2. Inria International Partners

9.4.2.1. Informal International Partners

We have started a collaboration with Fei Sha (University of California, Los Angeles) on the topic of representation learning for Natural Language Processing, materialized by the submission of a proposal to the 2016 call of the Inria Associate Teams program.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

We invited Prof. Claudio Gentile (Università dell'Insubria, Italy) in July, collaborating with MARC TOMMASI and FABIO VITALE on contextual node classification and bipartite graph matching problems on social network with user binary feedback.

Prof. Mark Herbster (University College London, UK) was invited for the PhD dissertation defense of THOMAS RICATTE in January and for Amir Sani's thesis in May 2015. He also collaborated with FABIO VITALE.

Several international researchers have also been invited to give a talk at the MAGNET seminar:

- Jan Ramon (KU Leuven, Belgium): "Learning theory for network-structured data" (January)
- Borja Balle (University of McGill, Canada): "A General Framework for Learning Weighted Automata" (February)
- Tiago P. Peixoto (Universität Bremen, Germany): "Inferring the large-scale structure of networks" (April)
- Dan Roth (University of Illinois at Urbana/Champaign, USA): "Learning, Inference and Supervision for Structured Prediction Tasks" (May)
- Michael Mathioudakis (Helsinki Institute for Information Technology, Finland): "Absorbing random-walk centrality – theory and algorithms" (June)
- Andre Martins (Priberam Labs and Instituto Superior Técnico Lisbon, Portugal): "Advances in Structured Regularization" (December)

9.5.2. Visits to International Teams

In July and in August, FABIO VITALE visited Aalto University (Helsinki, Finland), collaborating with Prof. Aristides Gionis on learning influence processes in social networks and graph reconstruction with queries.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific events organisation

10.1.1.1. Member of the organizing committees

MIKAELA KELLER, GÉRAUD LE FALHER and PAULINE WAUQUIER were members of the organizing committee of CAp 2015.

10.1.2. Scientific events selection

10.1.2.1. General chair, scientific chair

PASCAL DENIS served as an Area Chair (Discourse, Coreference, Pragmatics) for ACL 2015.

10.1.2.2. Chair of conference program committees

MARC TOMMASI was the program chair of CAp 2015.

10.1.2.3. Member of the conference program committees

PASCAL DENIS was member of the program committees of EMNLP 2015 and NAACL 2015.

PASCAL DENIS, RÉMI GILLERON, MIKAELA KELLER and MARC TOMMASI were members of the program committee of CAp 2015.

PASCAL DENIS and MIKAELA KELLER were members of the program committee of IJCAI 2015.

MIKAELA KELLER was member of the program committee of ICLR 2016.

MIKAELA KELLER and MARC TOMMASI were members of the program committee of NIPS 2015.

JAN RAMON was member of the program committee of BNAIC 2015.

JAN RAMON and MARC TOMMASI were members of the program committee of ICML 2015.

MARC TOMMASI was member of the program committee of CRI 2015.

FABIO VITALE was member of the program committees of COLT 2015 and WWW 2016.

10.1.2.4. Other

PASCAL DENIS is the program co-chair of the Polaris Colloquium.

10.1.3. Journal

10.1.3.1. Member of the editorial boards

JAN RAMON is member of the editorial boards of Machine Learning Journal (MLJ) and Data Mining and Knowledge Discovery (DMKD).

10.1.3.2. Reviewer - Reviewing activities

AURÉLIEN BELLET was reviewer for IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), IEEE Transactions on Cybernetics (TCYB) and Signal Processing (SIGPRO).

MARC TOMMASI was reviewer for Computer Communications (COMCOM).

FABIO VITALE was reviewer for the European Journal of Operational Research (EJOR).

10.1.4. Invited talks

MIKAELA KELLER was invited to give a talk at "Journée scientifique sur gestion de la connaissance à l'aide des Méthodes formelles pour la prévision des crises en épilepsie" (CRAN, Nancy).

JAN RAMON was invited to give a talk at the International Conference on Concept Lattices and their Applications (CLA 2015).

MARC TOMMASI was invited to give a talk at Data, Learning and Inference (DALI 2015): Networks – Processes and Causality.

10.1.5. Scientific expertise

PASCAL DENIS was member of the PhD Award Committee of ATALA (French association for NLP).

RÉMI GILLERON was member of the AERES evaluation committee of the LITIS Computer Science research laboratory (Rouen, France).

RÉMI GILLERON was also head of the selection committee for PhD and postdoctoral researchers at Inria Lille.

RÉMI GILLERON, MIKAELA KELLER and MARC TOMMASI were members of evaluation committees for the French Research Agency (ANR).

MARC TOMMASI was president of the jury for the recruitment of Junior Research Scientists (CR1/CR2) at Inria Lille.

FABIEN TORRE is elected for "CNU section 27 (informatique)" since Oct 2011 (reelected in Oct 2015) and is also member of the bureau since Nov 2015.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

University courses

Licence Informatique: DAVID CHATEL, Initiation à l'informatique, 45.5h, L1, Université de Lille, France

Master MOCAD: PASCAL DENIS, Apprentissage artificiel et aide à la décision, 37.5h, M2, Université de Lille, France

Master MIASHS: RÉMI GILLERON, NoSQL databases, 24h, M1, Université de Lille, France

Master MIASHS: RÉMI GILLERON, Programming in R, 24h, M1, Université de Lille, France

Master MIASHS: RÉMI GILLERON, Search Engine Optimization, 24h, M2, Université de Lille, France

Licence SID: RÉMI GILLERON, Information Digital Representation, 24h, L2, Université de Lille, France

Licence MIASHS: RÉMI GILLERON, Data Processing with Spreadsheet Program, 24h, L1, Université de Lille, France

Licence: MIKAELA KELLER, C2i, 25h, L2, Université de Lille, France

Licence MIASHS: MIKAELA KELLER, Information digital representation, 42h, L1, Université de Lille, France

Licence Sociologie: MIKAELA KELLER, Programming and algorithms, 28h, L2, Université de Lille, France

Licence Digital Humanities: MIKAELA KELLER, Information digital representation, 24h, L2, Université de Lille, France

Licence MIASHS: MARC TOMMASI, Réseaux, 64h, L1, Université de Lille, France

Licence MIASHS: MARC TOMMASI, Programmation client, 24h, L2, Université de Lille, France

Licence: MARC TOMMASI, C2i, 25h, L2, Université de Lille, France

Licence: MARC TOMMASI, Culture numérique, 30h, L2, Université de Lille, France. Online courses for all students in Lille 3 university L1/L2/L3.

Master Information Documentation: FABIEN TORRE, Langages statiques du web, 37h, M1, Université de Lille, France

Master Information Documentation: FABIEN TORRE, Algorithmique et programmation PHP pour le web, 75h, M1, Université de Lille, France

Master SdL: FABIEN TORRE, Algorithmique et programmation pour l'extraction d'information, 55h, M2, Université de Lille, France

Master Information Documentation: FABIEN TORRE, Javascript langage dynamique du web, 38h, M2, Université de Lille, France

Master MIASHS: FABIEN TORRE, Informatique pour le référencement, 12h, M2, Université de Lille, France

Licence MIASHS: FABIO VITALE, Introduction à l'algorithmique, 66h, L2, Université de Lille, France

Licence Sociologie: FABIO VITALE, Algorithmique des graphes, 28h, L3, Université de Lille, France

Licence SID: FABIO VITALE, Information Coding, 24h, L1, Université de Lille, France

Master MIASHS: FABIO VITALE, Unsupervised Classification, 30h, M1, Université de Lille, France

Licence MIASHS: FABIO VITALE, Unsupervised Classification, 28h, L3, Université de Lille, France

Invited lectures

Séminaire de rentrée: MARC TOMMASI, ENS Cachan, France

Course on Graph-based Machine Learning: MARC TOMMASI, University of Yaoundé, Cameroon.

Administration

MARC TOMMASI is a council member of the UFR MIME.

10.2.2. Supervision

Master: MATHIEU DEHOUCK, Graph-based Semi-Supervised Learning for Dependency Parsing, Université de Lille, Aug 2015, PASCAL DENIS and MARC TOMMASI

PhD: CHLOÉ BRAUD, Discourse Relation Identification from Labeled and Unlabeled Data, Université Paris-Diderot, Dec 2015, PASCAL DENIS (co-supervision with Laurence Danlos, Université Paris-Diderot)

PhD: EMMANUEL LASSALLE, Structured Learning with Latent Trees: a Joint Approach to Coreference Resolution, Université Paris-Diderot, May 2015, PASCAL DENIS (co-supervision with Laurence Danlos, Université Paris-Diderot)

PhD: THOMAS RICATTE, Hypernode graphs for learning from binary relations between sets of objects, Université de Lille, Jan 2015, RÉMI GILLERON

PhD in progress: DAVID CHATEL, Supervised Spectral Clustering and Information Diffusion in Graphs of Texts, since Sep 2012, PASCAL DENIS and MARC TOMMASI

PhD in progress: MATHIEU DEHOUCK, Graph-based Learning for Multi-lingual and Multi-domain Dependency Parsing, since Oct 2015, PASCAL DENIS and MARC TOMMASI

PhD in progress: GÉRAUD LE FALHER, Machine Learning in Signed Graphs, since Oct 2014, MARC TOMMASI, FABIO VITALE and CLAUDIO GENTILE (Università dell'Insubria, Italy)

PhD in progress: PAULINE WAUQUIER, Recommendation in Information Networks, since Dec 2013, MARC TOMMASI and MIKAELA KELLER

10.2.3. Juries

PASCAL DENIS was member of the PhD committee of Juliette Conrath (IRIT, Université Paul Sabatier) in Toulouse.

RÉMI GILLERON was member of the following committees: PhD committee of Marta Soare (Lille), Habilitation committee of Jérémie Mary (Lille), head of the selection committee for assistant professor (Lille).

MIKAELA KELLER was member of the following selection committees: assistant professor in Université de Lille 1, in Université de Lille 3 and at INSA Rouen.

MARC TOMMASI was member of the following committees: selection committee for assistant professor at UPMC (Paris 6), PhD committee (and reviewer) of Nadia Ouali Sebti (LITIS, Rouen), PhD committee of Fragkiskos D. Malliaros (École polytechnique, Paris) and PhD commitee of Thomas Ricatte (Lille).

10.3. Popularization

MIKAELA KELLER was involved in the Fête de la Science as a "chercheur itinérant".

MARC TOMMASI was co-author of the article “L’apprentissage automatique : le diable n’est pas dans l’algorithme” in the blog Binaire of the newspaper Le Monde.

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] T. RICATTE. *Hypernode graphs for learning from binary relations between sets of objects*, Université de Lille, January 2015, https://hal.archives-ouvertes.fr/tel-01246240

International Conferences with Proceedings

[2] C. BRAUD, P. DENIS. *Comparing Word Representations for Implicit Discourse Relation Classification*, in "Empirical Methods in Natural Language Processing (EMNLP 2015)", Lisbonne, Portugal, September 2015, https://hal.inria.fr/hal-01185927

[3] I. COLIN, A. BELLET, J. SALMON, S. CLÉMENÇON. *Extending Gossip Algorithms to Distributed Estimation of U-statistics*, in "Annual Conference on Neural Information Processing Systems (NIPS)", Montréal, Canada, December 2015, https://hal.inria.fr/hal-01214665

[4] E. LASSALLE, P. DENIS. *Joint Anaphoricity Detection and Coreference Resolution with Constrained Latent Structures*, in "AAAI Conference on Artificial Intelligence (AAAI 2015)", Austin, Texas, United States, Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI 2015), January 2015, https://hal.inria.fr/hal-01205189

[5] G. LE FALHER, A. GIONIS, M. MATHIOUDAKIS. *Where is the Soho of Rome? Measures and algorithms for finding similar neighborhoods in cities*, in "9th AAAI Conference on Web and Social Media - ICWSM 2015", Oxford, United Kingdom, May 2015, https://hal.archives-ouvertes.fr/hal-01134117

[6] G. PAPA, S. CLÉMENÇON, A. BELLET. *SGD Algorithms based on Incomplete U-statistics: Large-Scale Minimization of Empirical Risk*, in "Annual Conference on Neural Information Processing Systems (NIPS)", Montréal, Canada, December 2015, https://hal.inria.fr/hal-01214667

Conferences without Proceedings

[7] N. CESA-BIANCHI, C. GENTILE, F. VITALE, G. ZAPPELLA. *Efficient Link Classification in Social Networks*, in "International Conference on Computational Social Science", Helsinki, Finland, June 2015, https://hal.inria.fr/hal-01245747

Research Reports

[8] T. Ricatte, R. Gilleron, M. Tommasi. *Hypernode Graphs for Learning from Binary Relations between Groups in Networks*, Inria Lille, January 2015, https://hal.inria.fr/hal-01247103

Scientific Popularization

[9] P. Philippe, M. Tommasi, T. Vieville, C. De La Higuera. *L'apprentissage automatique : le diable n'est pas dans l'algorithme*, June 2015, Article sur http://binaire.blog.lemonde.fr, https://hal.inria.fr/hal-01246178

Other Publications

[10] G. Laurence, A. Lemay, J. Niehren, S. Staworko, M. Tommasi. *Sequential Tree-to-Word Transducers: Normalization, Minimization, and Learning*, August 2015, Long version of Lata 14 paper, https://hal.archives-ouvertes.fr/hal-01186993

References in notes

[11] I. Abraham, O. Neiman. *Using petal-decompositions to build a low stretch spanning tree*, in "Proceedings of the 44th symposium on Theory of Computing - STOC '12", 2012, http://dx.doi.org/10.1145/2213977.2214015

[12] A. Alexandrescu, K. Kirchhoff. *Graph-based learning for phonetic classification*, in "IEEE Workshop on Automatic Speech Recognition & Understanding, ASRU 2007, Kyoto, Japan, December 9-13, 2007", 2007, pp. 359-364

[13] M.-F. Balcan, A. Blum, P. P. Choi, J. Lafferty, B. Pantano, M. R. Rwebangira, X. Zhu. *Person Identification in Webcam Images: An Application of Semi-Supervised Learning*, in "ICML2005 Workshop on Learning with Partially Classified Training Data", 2005

[14] M. Belkin, P. Niyogi. *Towards a Theoretical Foundation for Laplacian-Based Manifold Methods*, in "Journal of Computer and System Sciences", 2008, vol. 74, n^o 8, pp. 1289-1308

[15] A. Bellet, A. Habrard, M. Sebban. *A Survey on Metric Learning for Feature Vectors and Structured Data*, in "CoRR", 2013, vol. abs/1306.6709

[16] A. Bellet, A. Habrard, M. Sebban. *Metric Learning*, Morgan & Claypool Publishers, 2015

[17] P. J. Bickel, A. Chen. *A nonparametric view of network models and Newman–Girvan and other modularities*, in "Proceedings of the National Academy of Sciences", 2009, vol. 106, pp. 21068–21073

[18] P. Blau. *Inequality and Heterogeneity: A Primitive Theory of Social Structure*, MACMILLAN Company, 1977, http://books.google.fr/books?id=jvq2AAAAIAAJ

[19] C. Braud, P. Denis. *Combining Natural and Artificial Examples to Improve Implicit Discourse Relation Identification*, in "coling", Dublin, Ireland, August 2014, https://hal.inria.fr/hal-01017151

[20] N. Cesa-Bianchi, C. Gentile, F. Vitale, G. Zappella. *A Linear Time Active Learning Algorithm for Link Classification*, in "Proc of NIPS", 2012, pp. 1619-1627

[21] H. CHANG, D.-Y. YEUNG. *Graph Laplacian Kernels for Object Classification from a Single Example*, in "Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2", Washington, DC, USA, CVPR '06, IEEE Computer Society, 2006, pp. 2011–2016, http://dx.doi.org/10.1109/CVPR.2006.128

[22] D. CHATEL, P. DENIS, M. TOMMASI. *Fast Gaussian Pairwise Constrained Spectral Clustering*, in "ECML/PKDD - 7th European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases", Nancy, France, September 2014, pp. 242 - 257 [*DOI :* 10.1007/978-3-662-44848-9_16], https://hal.inria.fr/hal-01017269

[23] D. DAS, S. PETROV. *Unsupervised Part-of-Speech Tagging with Bilingual Graph-Based Projections*, in "ACL", 2011, pp. 600-609

[24] P. DENIS, P. MULLER. *Predicting globally-coherent temporal structures from texts via endpoint inference and graph decomposition*, in "IJCAI-11 - International Joint Conference on Artificial Intelligence", Barcelone, Espagne, 2011, http://hal.inria.fr/inria-00614765

[25] E. R. FERNANDES, U. BREFELD. *Learning from Partially Annotated Sequences*, in "ECML/PKDD", 2011, pp. 407-422

[26] A. FRENO, G. C. GARRIGA, M. KELLER. *Learning to Recommend Links Using Graph Structure and Node Content*, in "NIPS Workshop on Choice Models and Preference Learning", 2011

[27] A. FRENO, M. KELLER, G. C. GARRIGA, M. TOMMASI. *Spectral Estimation of Conditional Random Graph Models for Large-Scale Network data*, in "Proc. of UAI 2012", Avalon, États-Unis, 2012, http://hal.inria.fr/hal-00714446

[28] A. B. GOLDBERG, X. ZHU. *Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization*, in "Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing", Stroudsburg, PA, USA, TextGraphs-1, Association for Computational Linguistics, 2006, pp. 45–52, http://dl.acm.org/citation.cfm?id=1654758.1654769

[29] A. GOLDENBERG, A. X. ZHENG, S. E. FIENBERG. *A Survey of Statistical Network Models*, Foundations and trends in machine learning, Now Publishers, 2010, http://books.google.fr/books?id=gPGgcOf95moC

[30] M. GOMEZ-RODRIGUEZ, J. LESKOVEC, A. KRAUSE. *Inferring networks of diffusion and influence*, in "Proc. of KDD", 2010, pp. 1019-1028

[31] M. HERBSTER, S. PASTERIS, F. VITALE. *Online Sum-Product Computation Over Trees*, in "Proc. of NIPS", 2012, pp. 2879-2887

[32] M. MCPHERSON, L. S. LOVIN, J. M. COOK. *Birds of a Feather: Homophily in Social Networks*, in "Annual Review of Sociology", 2001, vol. 27, n^{o} 1, pp. 415–444, http://dx.doi.org/10.1146/annurev.soc.27.1.415

[33] A. NENKOVA, K. MCKEOWN. *A Survey of Text Summarization Techniques*, in "Mining Text Data", Springer, 2012, pp. 43-76

[34] T. RICATTE, R. GILLERON, M. TOMMASI. *Hypernode Graphs for Spectral Learning on Binary Relations over Sets*, in "ECML/PKDD - 7th European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases", Nancy, France, Machine Learning and Knowledge Discovery in Databases, September 2014, https://hal.inria.fr/hal-01017025

[35] H. SHIN, K. TSUDA, B. SCHÖLKOPF. *Protein functional class prediction with a combined graph*, in "Expert Syst. Appl.", March 2009, vol. 36, n^{o} 2, pp. 3284–3292, http://dx.doi.org/10.1016/j.eswa.2008.01.006

[36] S. SINGH, A. SUBRAMANYA, F. C. N. PEREIRA, A. MCCALLUM. *Large-Scale Cross-Document Coreference Using Distributed Inference and Hierarchical Models*, in "ACL", 2011, pp. 793-803

[37] M. SPERIOSU, N. SUDAN, S. UPADHYAY, J. BALDRIDGE. *Twitter Polarity Classification with Label Propagation over Lexical Links and the Follower Graph*, in "Proceedings of the First Workshop on Unsupervised Methods in NLP", Edinburgh, Scotland, 2011

[38] A. SUBRAMANYA, S. PETROV, F. C. N. PEREIRA. *Efficient Graph-Based Semi-Supervised Learning of Structured Tagging Models*, in "EMNLP", 2010, pp. 167-176

[39] F. VITALE, N. CESA-BIANCHI, C. GENTILE, G. ZAPPELLA. *See the Tree Through the Lines: The Shazoo Algorithm*, in "Proc of NIPS", 2011, pp. 1584-1592

[40] L. WANG, S. N. KIM, T. BALDWIN. *The Utility of Discourse Structure in Identifying Resolved Threads in Technical User Forums*, in "COLING", 2012, pp. 2739-2756

[41] K. K. YUZONG LIU. *Graph-Based Semi-Supervised Learning for Phone and Segment Classification*, in "Proceedings of Interspeech", Lyon, France, 2013

[42] X. ZHU, Z. GHAHRAMANI, J. LAFFERTY. *Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions*, in "Proc. of ICML", 2003, pp. 912-919