

Workshop sur la dynamique des communautés sur Twitter en période électorale : analyse par graphes aléatoires – 26-27 avril 2018

Auditorium bâtiment IMAG, 700 avenue centrale, 38400 Saint Martin d'Hères

- [Alisa Kirichenko](#) :

FUNCTION ESTIMATION ON LARGE GRAPHS USING BAYESIAN LAPLACIAN REGULARIZATION

Abstract : In recent years there has been substantial interest in high-dimensional estimation and prediction problems on large graphs. These can in many cases be viewed as high-dimensional or nonparametric regression or classification problems in which the goal is to learn a “smooth” function on a given graph. We present a mathematical framework that allows to study the performance of nonparametric function estimation methods on large graphs and we derive minimax convergence rates within the framework. We consider simple undirected graphs that satisfy an assumption on their “asymptotic geometry”, formulated in terms of the graph Laplacian. We also introduce a Sobolev-type smoothness condition on the target function using the graph Laplacian to quantify smoothness. Then we develop Bayesian procedures for problems at hand and we show how asymptotically optimal Bayesian regularization can be achieved under these conditions. The priors we study are randomly scaled Gaussians with precision operators involving the Laplacian of the graph.

- [Marco Corneli](#) :

THE DYNAMIC STOCHASTIC TOPIC BLOCK MODEL FOR DYNAMIC NETWORKS WITH TEXTUAL EDGES

Abstract : We develop a probabilistic approach for the clustering of nodes of a dynamic graph, accounting for the content of textual edges as well as their frequency. Vertices are clustered in groups which are homogeneous both in terms of interaction frequency and discussed topics. The dynamic graph is considered stationary on a time interval if the proportions of topics discussed between each pair of nodes do not change in time during that interval. A classification variational expectation-maximization (C-DEM) algorithm is adopted to perform inference. A model selection criterion is also derived to select the number of node groups, time clusters and topics. Experiments on both simulated and real data are carried out to assess the proposed methodology.

- [Floriana Gargiulo](#) :

THE TOPOLOGY OF A DISCUSSION : THE #OCCUPY CASE

Abstract : I will present the analysis of a large sample of the Twitter activity that developed around the social movement ‘Occupy Wall Street’, to study the complex interactions between the human communication activity and the semantic content of a debate.

I will start analyze a network approach based on the study of the bipartite graph @Users-#Hashtags and of its projections : the ‘semantic network’, whose nodes are hashtags, and the ‘users interest network’, whose nodes are users. In the first instance, I will show that discussion topics (#hashtags) present a high structural heterogeneity, with a relevant role played by the semantic hubs that are responsible to guarantee the continuity of the debate. In the users’ case, the self-organisation process of users’ activity, leads to the emergence of two classes of communicators : the ‘professionals’ and the ‘amateurs’. Both the networks present a strong community structure, based on the differentiation of the semantic topics, and a high level of structural robustness when certain sets of topics are censored and/or accounts are removed.

By analysing the characteristics of the dynamical networks I will show that three different phases of the discussion about the movement can be identified. Each phase corresponds to a specific moment of the movement : from declaration of intent, organisation and development and the final phase of political reactions. Each phase is characterised by the presence of prototypical #hashtags in the discussion.

- [Guillaume Deffuant, Ilaria Bertazzi, Sylvie Huet](#) :

A SIMPLE OPINION DYNAMICS MODEL SUGGESTING AN INTRINSICALLY NEGATIVE EFFECT OF GOSSIPS

Abstract : We consider a simple model of interacting agents, each holding an opinion about herself and the others. During random encounters by pairs, agents modify their opinions under the noisy influence of others. The influence is attractive and agents opinions are more strongly attracted by the opinions of whom they value higher than themselves and vice versa. We focus on two unexpected emerging patterns : Starting from zero, agents opinions tend to grow and then their average stabilises at a significantly positive value. When introducing gossips, this pattern is inverted ; the opinions tend to decrease and stabilise on average at a negative value. With the aim to disclose the mechanisms behind the emergence of these patterns, we study simplified settings where the opinions about only one agent change whereas the self-opinions of the others are fixed. This allows us to show that the patterns’ emergence is related to both the amplitude of the fluctuations on the agent’s self-opinion, responsible for a positive bias, and the amplitude of the fluctuations on the others’ opinions on her, responsible for a negative bias. The gossips increase the negative bias on the opinions about others and can overcome the positive bias on self-opinions, which otherwise dominates.

- [Georgios Balikas](#) :

MACHINE LEARNING FOR PREDICTING THE SENTIMENT OF TWEETS

Abstract : Being able to automatically analyze the sentiment conveyed by tweets is important for various applications, for instance when assessing consumer satisfaction. While Twitter is a rich source of user-generated content, there are major challenges in analyzing its content. For instance, tweets are very short messages where punctuation and language are used in very creative ways.

In this talk, I will present machine learning approaches for predicting the sentiment of tweets. After introducing the task and the difficulties it poses, I will describe the steps taken for building a state-of-the-art sentiment classifier for tweets. I will conclude the talk with empirical evaluation of the proposed method and with a discussion on the challenges and limitations of sentiment analysis.

- Jean-Philippe Magué :

APPROCHES SOCIO-

LINGUISTIQUES ET COMPUTATIONNELLES DE LA VARIABILITÉ DU FRANÇAIS SUR TWITTER

Abstract : Les langues varient. Que ce soit aux niveaux phonologique, lexical, syntaxique ..., les manières de dire sont toujours multiples, les locuteurs sont toujours confrontés à des choix entre variantes linguistiques en compétition. Depuis une cinquantaine d'années, la sociolinguistique variationniste a montré comment cette variation linguistique est conditionnée d'une part par la structure sociodémographique des populations à grande échelle et, d'autre part, par la topologie des réseaux sociaux à plus petite échelle.

Les médias sociaux tels que Twitter proposent de nouveaux canaux de communication à travers lesquels les utilisateurs utilisent leur langue dans des formes qui diffèrent des formes écrites et des formes orales standard, et ce à travers un réseau de contacts sociaux. Puisque ces nouveaux usages ne sont régis par aucune instance prescriptive (enseignement, manuels, etc.), les réseaux sociaux sont le lieu d'une forte variabilité. Ils constituent de ce fait un observatoire particulièrement intéressant pour comprendre comment se structure cette variabilité, d'autant plus que la quantité et le format numérique des données permettent de renouveler en profondeur les méthodologies traditionnelles de la sociolinguistique.