

# Séminaire commun Probabilités et Statistiques

**Mardi 07 Mars à 14h00**

Laboratoire Dieudonné

Salle de Conférences

**Charles Bouveyron**

(Université Paris-Descartes)

*The Stochastic Topic Block Model for the Clustering of Networks with Textual Edges : Application to Medical Scientific Watch on PubMed*

Due to the significant increase of communications between individuals via social media (Facebook, Twitter, LinkedIn) or electronic formats (email, web, e-publication) in the past two decades, network analysis has become an unavoidable discipline. Many random graph models have been proposed to extract information from networks based on person-to-person links only, without taking into account information on the contents. This talk introduces the stochastic topic block model, a probabilistic model for networks with textual edges. We address here the problem of discovering meaningful clusters of vertices that are coherent from both the network interactions and the text contents. A classification variational expectation-maximization algorithm is proposed to perform inference. Simulated datasets are considered in order to assess the proposed approach and to highlight its main features. Finally, we demonstrate the effectiveness of our methodology on two real-world applications : the analysis of the Enron scandal and Medical Scientific Watch on PubMed.

*Référence* : C. Bouveyron, P. Latouche and R. Zreik, The Stochastic Topic Block Model for the Clustering of Networks with Textual Edges, Statistics and Computing, in press, 2017.