The Stochastic Topic Block Model for the Clustering of Vertices in Networks with Textual Edges

Charles Bouveyron

Université Paris Descartes & Sorbonne Paris Cité

Thursday, 9th March 2017
12:30pm Room 3-E4-SR03 Via Rontgen 1 Milano

Abstract

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, net-work analysis has become a 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 paper introduces the stochastic topic block model (STBM), 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 (C-VEM) algorithm is proposed to perform inference. Simulated data sets 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-word data sets: a directed communication network and an undirected co-authorship network.

Joint work with P. Latouche and R. Zreik

Department of Decision Sciences

Via Röntgen 1 - 20136 Milano

Tel. 02 5836.5632
Fax 02 5836.5630