Ph.D. research topic

- Title of the proposed topic: Statistical analysis of dynamic bipartite networks with heterogeneous edges, with applications in computational medicine
- Research axis of the 3iA: Axis 1
- **Supervisor (name, affiliation, email):** Pr. Charles Bouveyron, UCA & Inria, charles.bouveyron@univ-cotedazur.fr
- Potential co-supervisor (name, affiliation): Dr. Marco Corneli, UCA
- The laboratory and/or research group: Maasai team, Inria Sophia-Antipolis, LJAD, Université Côte d’Azur

Apply by sending an email directly to the supervisor.
The application will include:
- Letter of recommendation of the supervisor indicated above
- Curriculum vitae.
- Motivation Letter.
- Academic transcripts of a master’s degree(s) or equivalent.
- At least, one letter of recommendation.
- Internship report, if possible.

- Description of the topic:
  
  Statistical learning on dynamic bipartite networks with heterogeneous edges, with applications in computational medicine

Context and project: With the increasing capacity of measurement in all domains, it is now possible to record a large variety of data. In the context of computational medicine, hospitals and health agencies are recording a large amount of information involving two entities (doctors/patients, drugs/adverse reactions, ...) that can be recorded along the time and with different types of data (quantitative, textual, functional data). Such data may be viewed as a dynamic bipartite network with heterogeneous data. It is of course of great interest for doctors and health system managers to have access to a summary of those massive data and to detect weak signals among the data. The understanding of the dynamic of the recovered signals is also meaningful when facing a specific phenomenon, such an epidemic.

We propose in this Ph.D. project to focus on unsupervised statistical learning for the co-clustering of dynamic bipartite networks with heterogeneous edges. We want to consider the following types of data for the heterogeneous edges: texts, ordinal data and time series. To this end, we will have to introduce a new generative co-clustering model which either
supposes a latent representation of the nodes, or directly models the generation process of each data type. One of the best-known approaches used for unsupervised text classification is latent Dirichlet allocation (LDA, Blei et al., 2003). However, this method does not take into account the order of the words and, from a generative point of view, it can produce texts that have no sense (the words are drawn by chance in a dictionary). Other more recent methods are related to natural language processing (NLP), an area of artificial intelligence that aims at making computers capable of understanding and manipulating human language. For statistical NLP, deep-learning neural networks are used for textual analysis and prediction (see, for example, Collobert and Weston, 2008). In the context of network analysis, the LDA model has been extended recently by Bouveyron et al. (2017) to allow the clustering of networks with textual edges. The stochastic topic block model (STBM) model is a probabilistic model for networks with textual edges allowing to discover 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 model inference. STBM has also been adapted to the dynamic framework through non-homogeneous Poisson processes (dSTBM, Corneli et al, 2019). Extending those models to the case of dynamic bipartite networks with heterogeneous edges will allow to both monitor and understand the role of individuals in the existing system, and to propose an automatic way to characterize the interventions in the future on the basis of the recorded elements.

**Expected skills:** The candidate should have a graduate degree (Master 2 degree) in Statistical or Machine Learning. Him/her scholar background should include:

- statistical/machine learning, statistical inference, clustering, classification,
- mixture models, time series modeling, state-space models, Gaussian processes,
- good knowledge of R, Python and/or C++.

**Application:** Application files should contain a CV, an application letter and grade records of the 2 last years (M1 & M2). Applications should be sent by email to charles.bouveyron@univ-cotedazur.fr and marco.corneli@univ-cotedazur.fr before 15th May 2020.

**Advisor and location:**

- Team: Maasai project-team, Université Côte d'Azur & Inria
- Advisor: Pr. Charles Bouveyron (Université Côte d’Azur & Inria)
- Co-advisor: Dr. Marco Corneli (Université Côte d’Azur)
- Localization: Maasai project-team, Inria Sophia-Antipolis, 2004 Route des Lucioles, 06902 Valbonne

**References:**


