Internship offer: deep generative models for the joint analysis of text and network data

Supervisors and location

- Teams: MASSAI (INRIA, Sophia-Antipolis) + PAS (UCA, Aubière)
- Supervisors: Charles Bouveyron (PR, UniCA, INRIA), Pierre Latouche (PR, UCA, X, IUF, INRIA)
- Location: Centre INRIA Université Côte d'Azur, 2004 Rte des Lucioles, 06902 Sophia-Antipolis

Contacts

charles.bouveyron@inria.fr, pierre.latouche@uca.fr

Environment

Maasai is a research project-team at Inria Sophia-Antipolis, working on the models and algorithms of Artificial Intelligence. This is a joint research team with the laboratories LJAD (Mathematics, UMR 7351) and I3S (Computer Science, UMR 7271) of Université Côte d'Azur. The team is made of both mathematicians and computer scientists in order to propose innovative learning methodologies, addressing real-world problems, that are both theoretically sound, scalable and affordable. The INRIA center is part of the 3IA Côte d'Azur institute which is an AI cluster in France.

Context

Since the seminal approach on variational graph auto-encoders of Kipf and Welling (2016), graph neural networks (GNN, Scarselli et al., 2008) and more specifically graph convolution networks (GCN, Chen et al., 2020) are now widely used in the unsupervised context, for network analysis. The deep latent variable models (DLVM) based strategies are usually used to build node embeddings characterizing the network topology. The supervision team for this internship has been working on networks using computational statistics and machine learning for the last 15 years. Applications of their work in this context to analyze social networks have recieved strong attention during the last French presidential election, with four papers written in LeMonde journal (Mestre, 2022; Laurent, 2022; Auffret, 2022; Goar, 2022).

Project

The STBM model was introduced in Bouveyron et al. (2018). Relying on variational approximations and optimization techniques from computational statistics, the corresponding methodology was the first to allow the joint analysis of network and text data when texts are present on the edges. Contrary to graph or NLP (natural language processing) techniques focusing solely on the presence of edges or words in documents, it enabled the complete analysis of social networks and more generally of communication networks, without discarding any information. The methodology was used in a series of applications. In particular, it was employed during the last French presidential election to evaluate the weights of the different political parties and to assess the spread of fake news. The original work of Bouveyron et al. (2018) led to a series of extensions in Corneli et al. (2019); Bergé et al. (2019); Liang et al. (2021); Boutin et al. (2023, 2025b,a) relying on deep latent variable models,

graph convolution networks, and autoencodings strategies. In this internship, we aim at proposing the deepest extension to the original paper by considering the most advanced approaches from NLP and graph analysis. The core motivation is to rely on the most advanced methods from the two domains, using graph attention networks, transformers, and attention mechanisms in a generative framework, to define a new standard for the joint analysis or network and text data.

Objectives

- Define a new generative model for the joint analysis of network and text data
- Build an inference procedure for this new model which is scalable in the size of the network and in the number of documents
- Implement the corresponding new methodology

Key words

Computational statistics, machine learning, mathematics of AI, deep generative models, optimization, graph neural networks, graph attention networks, transformers, attention mechanism, NLP, graph analysis, social networks, communication networks.

Expected skills

The candidate should be a master 2 student in a statistics / machine learning program, with a particularly strong background in mathematics and computer science.

References

Auffret, S. (2022). Brigitte macron et jean-michel trogneux, itinéraire d'une infox délirant. *LeMonde*.

Bergé, L. R., Bouveyron, C., Corneli, M., and Latouche, P. (2019). The latent topic block model for the co-clustering of textual interaction data. *Computational Statistics & Data Analysis*, 137:247–270.

Boutin, R., Bouveyron, C., and Latouche, P. (2023). Embedded topics in the stochastic block model. *Statistics and Computing*, 33(5):95.

Boutin, R., Latouche, P., and Bouveyron, C. (2025a). The deep latent position block model for block clustering and latent representation of nodes in networks. *Statistics and Computing*, 35(5):151.

Boutin, R., Latouche, P., and Bouveyron, C. (2025b). The deep latent position topic model for clustering and representation of networks with textual edges. *Scandinavian Journal of Statistics*. arXiv:2304.08242 [cs].

Bouveyron, C., Latouche, P., and Zreik, R. (2018). The stochastic topic block model for the clustering of vertices in networks with textual edges. *Statistics and Computing*, 28(1):11–31.

Chen, M., Wei, Z., Huang, Z., Ding, B., and Li, Y. (2020). Simple and deep graph convolutional networks. In *International conference on machine learning*, pages 1725–1735. PMLR.

Corneli, M., Bouveyron, C., Latouche, P., and Rossi, F. (2019). The dynamic stochastic topic block model for dynamic networks with textual edges. *Statistics and Computing*, 29(4):677–695.

Goar, M., C. N. (2022). Présidentielle 2022 : faut-il se couper de twitter, huis clos politique devenu hostile ? *LeMonde*.

Kipf, T. N. and Welling, M. (2016). Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*.

Laurent, S. (2022). Comment la gauche sociale-démocrate a perdu la bataille des réseaux sociaux. *LeMonde*.

Liang, D., Corneli, M., Bouveyron, C., and Latouche, P. (2021). DeepLTRS: A deep latent recommender system based on user ratings and reviews. *Pattern Recognition Letters*, 152:267–274.

Mestre, A. (2022). Eric zemmour, nouveau président de la fachosphère ? *LeMonde*.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. (2008). The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80.