3iA Côte d'Azur
Interdisciplinary Institute for Artificial Intelligence

Ph.D. research topic

- Title of the proposed topic: **AI models hybridizing machine learning and argumentation to counter online misinformation and cyberbullying**
- Research axis of the 3IA: Axis 1, Axis 4
- Supervisor (name, affiliation, email): Charles Bouveyron (UCA), charles.bouveyron@univ-cotedazur.fr
- Potential co-supervisor (name, affiliation): Serena Villata (CNRS), villata@i3s.unice.fr
- The laboratory and/or research group: MASAAI & WIMMICS teams (Université Côte d'Azur, CNRS, Inria)

Apply by sending an email directly to the supervisor.
The application will include:

- Letter of recommendation of the supervisor indicated above
- Curriculum vitæ.
- 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:

Context

With the recent developments of information and communication media, it becomes necessary to actively monitor certain types of activities that could harm individuals and even society. In particular, social media offer new communication means that can be used to misinform or harass. Given the size of these social networks and the huge number of exchanged messages everyday, to effectively fight against misinformation and cyberbullying, it is nowadays necessary to rely on automatic tools capable of analyzing this mass of data. Among those tools, AI algorithms designed for this task can be based on either Machine Learning (ML) [BN06] or Natural Language Processing (NLP) [JM09]. ML approaches based on statistical or deep learning are usually efficient and can scale to the large size of data. However, they do not allow a fine understanding of the reasons behind people's behaviour on social networks. Conversely, NLP approaches, such as argumentation analysis techniques, allow for a fine modeling of the way some online users either spread misinformation or harass

other users. On the downside, these approaches have a significant computational cost and are not easily applied on a large scale.

Research project

This project aims at proposing **new AI models mixing machine learning and argumentation for countering misinformation and cyberbullying**, allowing both for a fine understanding of the misleading strategies and the possibility to be applied at a large scale.

To address this goal, we propose an approach able to cluster the nodes of a network taking into account the argumentation graph connecting the arguments proposed by the different actors to each other through support and attack argumentative relations. This will allow us to detect groups of persons connected on the network and sharing the adoption of common argumentation patterns. To this end, we could extend the (deep) latent variable models (DLVM) [Bis98] designed for network clustering by adding a modeling of the distribution of the argumentation graphs. Latent variable models for network clustering include popular models such as the stochastic block model (SBM) [SN97], known to be particularly efficient to model complex networks, and the latent space model (LPM) [HRH02], which provides a meaningful latent representation of the data. Recent extensions of this class of models use deep learning, in particular graph neural networks (GNN) [SGT+08, KW16], to further describe the network topology and thus allowing an even better clustering performance. The extension of these models could be addressed by introducing the information about the argumentation structure carried out by the texts (e.g., social media messages and posts) and by enforcing the clusters to be representative of both the network connection patterns and the argumentation structures. Thus, the resulting clustering of the nodes of the network should explicit the different roles of people and allow the detection of particular groups that mobilize argumentation for specific (potentially nefastous) actions in the network.

Expected skills

The candidate should be a Master 2 student in a NLP / Statistics / Machine Learning program, with a strong background in computer science and mathematics. This internship could be followed by a Ph.D. thesis, for which the funding is already secured.

References

[Bis98] Christopher M Bishop. Latent variable models. In Learning in graphical models, pages 371–403. Springer, 1998.

[BN06] Christopher M Bishop and Nasser M Nasrabadi. Pattern recognition and machine learning, volume 4. Springer, 2006.

[HRH02] Peter D Hoff, Adrian E Raftery, and Mark S Handcock. Latent space approaches to social network analysis. Journal of the american Statistical association, 97(460):1090–1098, 2002.

[JM09] Dan Jurafsky and James H. Martin. Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition. Pearson Prentice Hall, Upper Saddle River, N.J., 2009.

[KW16] Thomas N Kipf and Max Welling. Variational graph auto-encoders. arXiv preprint arXiv:1611.07308, 2016.

[SGT+08] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. IEEE transactions on neural networks, 20(1):61–80, 2008.

[SN97] Tom AB Snijders and Krzysztof Nowicki. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. Journal of classification, 14(1):75–100, 1997.