



## Explicabilité pour l'analyse d'opinions dans les médias sociaux

**Mots clés :** Explicabilité, analyse des médias sociaux, fouille d'opinions, GNN, Graphes de terrain

**Encadrant :** Maria Malek

**Laboratoire :** ETIS UMR 8051

**Équipe :** MIDI

**Durée :** 6 mois - stage rémunéré

### Contexte

Il existe aujourd'hui un besoin émergent pour comprendre les décisions rendues par des méthodes d'IA lorsque celles-ci affectent la vie des humains. Les paradigmes sous-jacents à ce problème révèlent du domaine dit de l'IA explicable (XAI) reconnu comme une caractéristique cruciale pour le déploiement pratique des modèles d'IA [1]. L'explicabilité a pour objectif de faciliter la compréhension de divers aspects d'un modèle conduisant à des informations qui peuvent être utilisées par plusieurs acteurs. L'explicabilité peut être considérée comme une caractéristique active d'un modèle, désignant toute action prise par un modèle dans le but de clarifier ou de détailler ses fonctions internes [1,2].

D'un autre côté, l'étude des réseaux complexes nommés aussi graphes de terrain est un domaine de recherche scientifique jeune et actif largement inspiré des découvertes empiriques sur les réseaux réels tels que les réseaux informatiques, les réseaux biologiques, les réseaux technologiques, les réseaux cérébraux, les réseaux climatiques et les réseaux sociaux [3].

Nous avons exploré dans des travaux récents [5,6] portant sur l'analyse des médias sociaux, la combinaison des méthodes classiques d'exploration d'opinion avec celles de l'analyse des réseaux sociaux ainsi que leur impact sur la formation et la propagation d'opinion. Afin d'étudier l'impact des utilisateurs influents (nœuds influents), plusieurs facteurs d'influence extraits du réseau (graphe) ont été intégrés dans le processus d'exploration d'opinions. Ces facteurs sont généralement calculés en utilisant différentes mesures de centralité comme le degré, la proximité, l'intermédiarité, la centralité PageRank, etc.

En intégrant une méthode d'explicabilité adéquate, nous souhaitons rendre plus compréhensible également les résultats concernant la polarité de l'opinion trouvée au niveau des utilisateurs et au niveau des groupes (communautés). De même, le modèle doit être capable d'expliquer les changements d'opinion détectés en prenant en compte les

informations extraites du réseau de propagation et les séquences d'actions entreprises (par exemple : tweets, retweets, réponses) menant à ce changement.

Un modèle transparent basé sur l'apprentissage automatique pour la détection de la modification d'opinions au sein des réseaux égocentriques autour des influenceurs, a été proposé [5,6]. Ainsi, des caractéristiques (attributs) de différentes natures (textuelle, contextuelle et topologique) qui expliquent la modification de l'opinion, ont pu être identifiées. Nous souhaitons généraliser cette approche afin de pouvoir proposer un cadre général de l'explicabilité pour l'analyse des opinions dans les médias sociaux.

## Objectif

Le but de stage est de proposer et d'intégrer une (ou plusieurs) méthode(s) d'explicabilité dans les algorithmes d'analyse d'opinions afin de produire des explications émergentes qui combinent des informations nodales (comme le profil d'utilisateur et les données textuelles) et topologiques extraites de la structure du graphe de propagation des opinions.

Dans la plupart des méthodes d'apprentissage automatique les entités traitées dans une méthode d'explicabilité sont liées aux caractéristiques (attributs), aux données, aux prototypes ainsi qu'à des connaissances extraites ayant souvent la forme de règles ou d'arbre décision. Lorsqu'il s'agit des réseaux complexes, des nouvelles entités doivent être prises en compte comme la nature des liens dans le graphe ainsi que d'autres informations topologiques.

L'objectif est d'étudier et de comparer deux principales approches de l'explicabilité dans les réseaux complexes à travers l'exemple de l'étude de l'analyse d'opinions [7] : la première consiste à l'utilisation des méthodes XAI lors de l'application des méthodes d'apprentissage automatique pour l'analyse des réseaux, comme par exemple, les méthodes de network embedding [8] ainsi que les réseaux de neurones en graphes (Graph Neural networks)[9], la deuxième approche consiste à définir des méthodes d'explicabilité propres au domaine de l'analyse des réseaux. Ces méthodes seront par nature spécifiques et en lien étroit avec les algorithmes d'analyse du graphe du terrain social utilisés lors de l'analyse d'opinions.

**Profil du candidat :** dernière année d'un Master ou d'un diplôme d'ingénieur dans un domaine en lien avec l'informatique ou les mathématiques appliquées.

**Compétences requises :** connaissances en apprentissage automatique et compétences en Python. Une compétence en analyse des réseaux complexes serait appréciée.

**Candidature :** lettre de motivation, CV et relevés de notes M1/M2

**Contact :** maria.malek@cyu.fr

## Références

- [1] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, Francisco Herrera, Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Information Fusion, Volume 58, 2020, Pages 82-115, ISSN 1566-2535, <https://doi.org/10.1016/j.inffus.2019.12.012>
- [2] Andrea Morichetta, Pedro Casas, and Marco Mellia. 2019. EXPLAIN-IT: Towards Explainable AI for Unsupervised Network Traffic Analysis. In Proceedings of the 3rd ACM CoNEXT Workshop on Big DATA, Machine Learning and Artificial Intelligence for Data Communication Networks (Big-DAMA '19). Association for Computing Machinery, New York, NY, USA, 22–28. DOI:<https://doi.org/10.1145/3359992.3366639>
- [3] J. Newman, The structure and function of complex networks, SIAM Reviews, 45(2): 167-256, 2003
- [4] S. Fortunato, Community detection in graphs, Physics Reports 486, 75-174 (2010), <http://dx.doi.org/10.1016/j.physrep.2009.11.002>
- [5] Kossi Folly, Maria Malek, Dimitris Kotzinos, Social networks analysis for opinion model extraction. Networks 2021: first combined meeting of the International Network for Social Network Analysis (Sunbelt XLI), and the Network Science Society (NetSci 2021), Indiana, United States, Jul 2021.
- [6] Kossi Folly, Youssef Boughaba, Maria Malek, leveraging nodal and topological information for studying the interaction between two opposite ego networks, In *HCI 2023: Social Computing and Social Media*, volume 14026, of *Lecture Notes in Computer Science*, pages 295-307, Copenague, Denmark, July 2023. Springer Nature Switzerland.
- [7] Maria Malek, Towards Explainability for Interaction Networks Analysis, ACONTA'22: First European Conference on Augmented Complex Networks - Trustworthy Analysis, Université Sorbonne Paris Nord, 1-2 December 2022 (Invited Talk).
- [8] P. Cui, X. Wang, J. Pei and W. Zhu, "A Survey on Network Embedding, IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 5, pp. 833-852, 1 May 2019,
- [9] Yuan, Hao and Yu, Haiyang and Gui, Shurui and Ji, Shuiwang, Explainability in Graph Neural Networks: A Taxonomic Survey, arXiv, 2020.