

Apprentissage de représentations sur des graphes complexes par GNN pour le débruitage des graphes de connaissances.

Mots-clés : Knowledge Graphs, Graph Neural Networks, Graph Representation learning, Complex Networks.

1 Résumé

Ces dernières années, les réseaux de neuronaux opérant sur des graphes (Graph Neural Networks ou GNN) Kipf et Welling (2017) Veličković et al. (2018) sont devenus une approche populaire pour l'apprentissage de représentations sur des graphes. Ces approches permettent d'apprendre des représentations vectorielles de nœuds d'arêtes ou de graphes afin de capturer leur structure et leurs interactions. En agrégeant de la bonne manière les représentations des nœuds voisins, ces architectures ont montré des résultats très intéressants sur différentes tâches au niveau du nœud, du graphe, ou de l'arête dans de nombreux domaines informatique, chimie, biologie Errica et al. (2019). Cependant, la représentation et l'analyse de graphes complexes présentent des défis importants en raison de leur structure, de leur hétérogénéité, de leur taille. En effet, la majorité des architectures sont basées sur le principe d'homophilie, autrement dit, les nœuds voisins donnent des informations similaires. Ainsi si un graphe modélise des interactions entre nœuds différents, les GNNs se montreraient peu performants Zhu et al. (2020) Ma et al. (2022). De plus, de nombreuses structures de graphe limitent la qualité du passage de message. En effet, il est établi que certains graphes possèdent des goulots d'étranglement (Fig. 1) qui forcent la propagation de l'information à se condenser limitant ainsi son efficacité Topping et al. (2022) Uri Alon (2021).

Le problème des goulots d'étranglement apparaît aussi dans les graphes de connaissances (Knowledge Graphs, KGs) acquis de façon automatique à partir du texte. Le manque d'une précision suffisante dans les outils de Traitement Automatique de la Langue e comporte la présence dans les graphes résultants d'un certain nombre de connexions (dans les KGs les connexions représentent des relations sémantiques) qui sont en fait erronées Dessì et al. (2021). Or, comme on peut voir de l'exemple d'une erreur de ce type en Fig. 2, ces erreurs sont souvent associés à des goulots d'étranglement entre deux sous-reseaux plus complexes. Par conséquent, travailler sur la détection et l'amélioration des réseaux dans le cas de goulots d'étranglement pourrait avoir

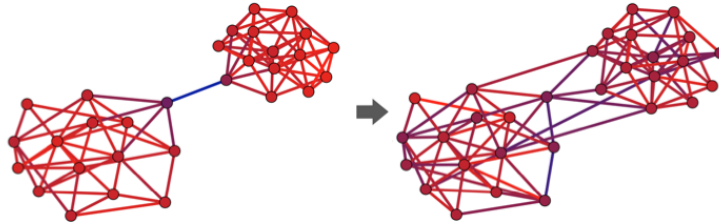


FIG. 1 – Exemple de goulot d'étranglement et d'une structure qui diffuse mieux l'information entre deux sous-réseaux.

un effet positif sur les KGs pour éliminer ou réduire le poids des erreurs dans l'étape d'acquisition.

En plus, certains graphes modélisant des fortes structures hiérarchiques, sont mieux représentés, non pas dans l'espace euclidien, mais dans l'espace hyperbolique dû à la courbure de cet espace Octavian Ganea (2018) Ines Chami (2019). Or, la relation "is_a" ou hyperonymie est une relation hiérarchique très fréquente dans les graphes de connaissances Yuan et al. (2020). Apprendre des représentation de graphes dans un espace hyperbolique pourrait donc apporter des bénéfices à la représentation de graphes de connaissances et taxonomies.

2 Objectif de la thèse

L'objectif de cette thèse est d'apprendre des représentations pour des structures de graphes complexes et d'étudier leur application aux graphes de connaissances principalement pour le débruitage et l'apprentissage de représentations efficaces. Le travail du doctorant se déroulera selon deux grands axes :

- Identifier la topologie des graphes étudiés pour détecter d'éventuelles structures qui pourraient diminuer la qualité du passage de message sur différents graphes de la littérature (graphe de citations, graphe de connaissances, graphe modélisant des molécules..).Évaluer comment la structure du graphe évolue selon ce qu'elle modélise.
- Mettre en place différentes approches pour apprendre des représentations sur ces graphes selon la structure identifiée. Des pistes de travail seraient notamment d'apprendre des représentations dans un autre espace ou les données sont mieux représentés (hyperbolique pour structure hiérarchique par exemple) ou d'apprendre en parallèles des représentations une meilleure structure du



FIG. 2 – Exemple de connexion erronée dans le graphe de connaissances AI-KG Dessì et al. (2021) pour la relation “is_a”.

graphe. Ces architectures devront être appliquées sur divers types de graphes et évaluées sur différentes tâches.

Références

- Dessì, D., F. Osborne, D. R. Recupero, D. Buscaldi, et E. Motta (2021). Generating knowledge graphs by employing natural language processing and machine learning techniques within the scholarly domain. *Future Gener. Comput. Syst.* 116, 253–264.
- Errica, F., M. Podda, D. Bacciu, et A. Micheli (2019). A fair comparison of graph neural networks for graph classification. *ICLR*.
- Ines Chami, Zhitao Ying, C. R. J. L. (2019). Hyperbolic graph convolutional neural networks. In *Neurips*.
- Kipf, T. N. et M. Welling (2017). Semi-Supervised Classification with Graph Convolutional Networks. In *Proceedings of the International Conference on Learning Representations, ICLR*.
- Ma, Y., X. Liu, N. Shah, et J. Tang (2022). Is homophily a necessity for graph neural networks? *ICLR*.
- Octavian Ganea, Gary Becigneul, T. H. (2018). Hyperbolic neural networks. In *Neurips*.
- Topping, J., F. Di Giovanni, B. P. Chamberlain, X. Dong, et M. M. Bronstein (2022). Understanding over-squashing and bottlenecks on graphs via curvature. *Proceedings of the International Conference on Learning Representations*.
- Uri Alon, E. Y. (2021). On the bottleneck of graph neural networks and its practical implications. In *ICLR*.
- Veličković, P., G. Cucurull, A. Casanova, A. Romero, P. Liò, et Y. Bengio (2018). Graph Attention Networks. *ICLR*.
- Yuan, J., Z. Jin, H. Guo, H. Jin, X. Zhang, T. Smith, et J. Luo (2020). Constructing biomedical domain-specific knowledge graph with minimum supervision. *Knowledge and Information Systems* 62, 317–336.
- Zhu, J., Y. Yan, L. Zhao, M. Heimann, L. Akoglu, et D. Koutra (2020). Beyond homophily in graph neural networks : Current limitations and effective designs. *Advances in Neural Information Processing Systems* 33, 7793–7804.