

Comparaison de la capacité des plongements de graphes à capturer les propriétés des réseaux

Rémi Vaudaine*, Rémy Cazabet**, Christine Largeron*

*Univ Lyon, UJM-Saint-Etienne, CNRS, Institut d'Optique Graduate School
Laboratoire Hubert Curien UMR 5516, F-42023, SAINT-ETIENNE, France
{remi.vaudaine,christine.largeron}@univ-st-etienne.fr

**Univ Lyon, UCBL, CNRS, LIRIS UMR 5205, F-69621, Lyon, France
remy.cazabet@gmail.com

1 Motivations

Les graphes sont utiles pour modéliser des systèmes complexes dans un large éventail de domaines. Parmi les approches conçues pour les étudier, le plongement de graphes consiste à encoder les noeuds, des parties du graphe ou un graphe entier dans un espace de faible dimension tout en préservant certaines propriétés structurelles. Puisqu'il permet à toute la gamme de techniques d'exploration de données et d'apprentissage automatique qui nécessitent des vecteurs d'être appliquée aux données relationnelles, il peut profiter à de nombreuses applications.

Plusieurs articles de l'état de l'art consacrés au plongement de graphes récemment publiés Cui et al. (2019); Goyal et Ferrara (2017); Cai et al. (2018); Zhang et al. (2020); Wu et al. (2020), comportent une étude comparative des performances des méthodes à résoudre des tâches spécifiques. Or, nous considérons qu'une telle évaluation des modèles de plongement de graphes guidée par une tâche (prédiction de liens / classement / détection de communautés) induit un biais lié au choix de l'algorithme employé pour traiter la tâche. Si les performances obtenues sur la tâche sont mauvaises cela est-il dû à l'algorithme choisi pour traiter la tâche ou à l'algorithme d'embedding? En réalité, l'évaluation guidée par la tâche ignore le fait que les algorithmes de plongement sont conçus, explicitement ou implicitement, pour préserver certaines propriétés structurelles particulières. Leur utilité pour une tâche donnée dépend donc de la manière dont ils réussissent à capturer ces propriétés. Par exemple, dans le cas de l'attachement préférentiel, un classifieur aura du mal à prédire de nouveaux liens si les représentations de deux nœuds voisins ne sont pas proches dans l'espace de plongement. Dans cet article, nous adoptons donc un point de vue différent des évaluations précédentes axées sur les tâches en étudiant directement, à travers une étude comparative expérimentale, la capacité des algorithmes de plongement de graphes à capturer des propriétés spécifiques, qui sont la proximité de premier ordre des nœuds, l'équivalence structurelle (proximité de second ordre) ou l'équivalence isomorphe. Pour ce faire, nous proposons différentes mesures permettant d'évaluer dans quelle mesure les méthodes de plongement les encodent.

2 Propriétés structurelles et mesures

Nous définissons des mesures pour évaluer la capacité des méthodes de plongement de graphes à capturer les trois propriétés correspondantes.

Voisinage ou proximité de premier ordre (P1) : un algorithme qui capture la proximité de premier ordre cherche à rapprocher dans l'espace de plongement chaque paire de nœuds liés dans le graphe. La première mesure S que nous avons définie calcule la proportion des voisins d'un nœud du graphe qui sont aussi les plus proches voisins dans l'espace de plongement.

Équivalence structurelle ou proximité de second ordre (P2) : deux nœuds sont structurellement équivalents s'ils ont beaucoup de voisins en commun. Pour quantifier cette propriété, nous calculons, pour chaque paire de nœuds, la distance entre les lignes de la matrice d'adjacence (qui représentent l'ensemble de voisins pour un nœud) et la distance entre les vecteurs représentant les nœuds. Finalement, la corrélation entre ces deux grandeurs nous indiquent si la méthode de plongement considérée est capable de capturer l'équivalence structurelle.

Équivalence isomorphe (P3) : deux nœuds sont équivalents isomorphiquement s'ils ont le même rôle dans le graphe, c'est-à-dire si leur réseaux égo-centrés sont isomorphiques. Le réseau égo-centré d'un nœud est le sous-graphe défini par l'ensemble de ses voisins et l'ensemble des liens entre eux. Pour comparer deux réseaux égo-centrés, nous utilisons la distance d'édition de graphes. De plus, pour chaque paire de nœuds, nous calculons la distance entre leurs représentants dans l'espace de plongement. Finalement, la corrélation entre ces deux distances nous indique si une méthode de plongement peut conserver l'équivalence isomorphe.

Le lecteur intéressé pourra se référer à Vaudaine et al. (2020) pour une présentation des résultats expérimentaux de l'étude comparative que nous avons réalisée sur 10 méthodes de plongement portant sur 12 graphes de nature variée à l'aide des mesures que nous avons définies. Le code et les jeux de données sont également accessibles librement sur notre GitHub ¹.

Nous remercions les projets BITUNAM ANR-18-CE23-0004 et IDEXLYON ACADEMICS ANR-16-IDEX-0005 pour leur soutien.

Références

- Cai, H., V. W. Zheng, et K. C. Chang (2018). A comprehensive survey of graph embedding : Problems, techniques, and applications. *IEEE TKDE* 30(9), 1616–1637.
- Cui, P., X. Wang, J. Pei, et W. Zhu (2019). A survey on network embedding. *IEEE TKDE* 31(5), 833–852.
- Goyal, P. et E. Ferrara (2017). Graph embedding techniques, applications, and performance : A survey. *CoRR abs/1705.02801*.
- Vaudaine, R., R. Cazabet, et C. Langeron (2020). Comparing the preservation of network properties by graph embeddings. In *IDA XVIII*, pp. 522–534. Springer International Publishing.
- Wu, Z., S. Pan, F. Chen, G. Long, C. Zhang, et P. S. Yu (2020). A comprehensive survey on graph neural networks. *IEEE Trans. on Neural Networks and Learning Systems*, 1–21.
- Zhang, D., J. Yin, X. Zhu, et C. Zhang (2020). Network representation learning : A survey. *IEEE Transactions on Big Data* 6(1), 3–28.

1. https://github.com/vaudaine/Comparing_embeddings