

Multi-SPMiner : un framework d'apprentissage profond pour l'extraction de motifs fréquents dans des multi-graphes avec application aux graphes spatio-temporels

Assaad Zeghina*, Aurélie Leborgne*, Florence Le Ber*, Antoine Vacavant**

*Université de Strasbourg, CNRS, ENGEES, ICube UMR 7357, F67000 Strasbourg
(assaad-oussama.zeghina, aurelie.leborgne, florence.le-ber)@icube.unistra.fr,

**Université Clermont Auvergne, CNRS, SIGMA Clermont, Institut Pascal
antoine.vacavant@uca.fr

La disponibilité et la diversité croissantes des données relationnelles, souvent représentées par des graphes ou des multigraphes, exigent des algorithmes efficaces pour l'extraction de motifs fréquents. Les algorithmes actuels concernent principalement les graphes relationnels simples, mais des tentatives récentes visent à adapter ces algorithmes à la recherche de motifs dans les multigraphes. Notre approche est inspirée de l'utilisation novatrice de l'apprentissage profond par SPMiner (Ying et al., 2020) pour la recherche de motifs de graphes. Mais, si SPMiner excelle dans l'extraction de motifs de graphes simples, son application à dans d'autres formats reste largement inexploré. Par exemple, dans notre étude, nous considérons des multigraphes particuliers, les graphes spatio-temporels (st-graphes), constitués d'un ensemble d'entités spatiales et de leurs relations. Ce modèle peut être considéré comme l'union de trois sous-graphes : le sous-graphe spatial représente les interactions spatiales entre entités à un temps donné, le sous-graphe spatio-temporel représente les interactions spatiales entre entités à des temps successifs. Le sous-graphe de filiation représente la transmission d'identité entre les entités à des temps successifs (Del Mondo et al., 2013).

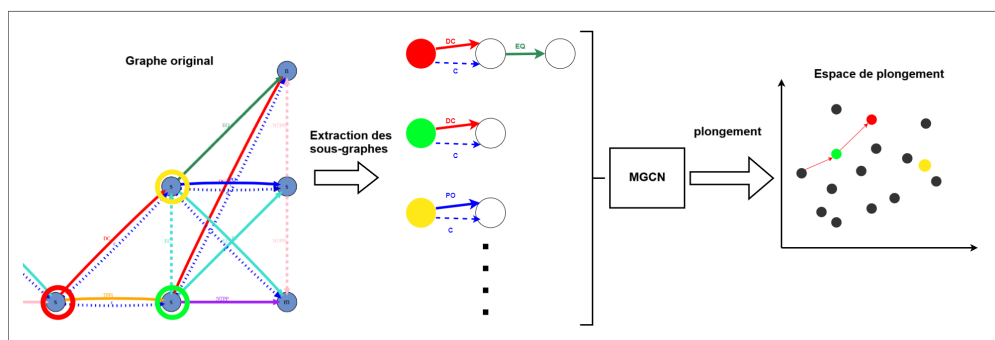


FIG. 1 – Etape de projection par le MGCN dans MultiSPMiner

Pour traiter ce type de multigraphes, l'approche MultiSPminer, comme SPMiner, décompose les graphes en voisinages ancrés à des nœuds, mais utilise un modèle d'apprentissage profond adapté aux multigraphes pour projeter efficacement ces voisinages dans un espace de dimension réduite. L'accent est ainsi mis sur la découverte de motifs ancrés à des nœuds, avec un support dépassant la valeur s . MultiSPMiner utilise pour cela un processus en deux étapes (figure 1). Dans la première étape, un réseau neuronal convolutionnel multigraphe (MGCN) opère une projection des voisinages ancrés à des nœuds dans un espace ordonné. Le MGCN permet de préserver l'ordre lors de la projection : ainsi si le voisinage A est un sous-graphe du voisinage B, alors A est intégré en bas à gauche de B (pour un espace en deux dimensions). Dans la deuxième étape, une recherche itérative est mise en œuvre pour trouver les motifs ancrés à des nœuds apparaissant fréquemment dans le graphe projeté. L'ajout itératif de nœuds et des arêtes vise à minimiser la marge totale $m(G)$ du graphe G , définie comme suit, où \mathcal{N} représente l'ensemble des voisinages de tous les nœuds dans G , et ϕ est la fonction de projection utilisée par le MGCN :

$$m(G) = \sum_{N \in \mathcal{N}} |\max(0, \phi(G) - \phi(N))|^2 \quad (1)$$

L'évaluation expérimentale est réalisée sur des graphes synthétiques, obéissant à différents paramètres ; 200 graphes d'entraînement et 10 graphes de test sont générés par expérience, pour un total de 14 tests. Les paramètres suivants sont étudiés : la taille des motifs, leur fréquence et la distribution des nœuds. Dans ces expériences, Multi-SPMiner a obtenu des performances stables, avec une précision allant de 60% à 68%, pour des distributions différentes en nombre de nœuds et de relations. La précision la plus élevée (68,3%) a été observée pour avec avec des graphes contenant des motifs d'une taille unique.

Afin d'étudier la capacité de généralisation de notre modèle et sa capacité à apprendre des relations structurelles entre les nœuds facilitant la découverte de motifs, nous avons mené une deuxième série d'expériences sur des graphes ayant des motifs de taille variable. Les résultats obtenus ont été comparés avec les tests précédents, concluant sur la capacité de la méthode à apprendre l'extraction de motifs indépendamment de la taille. De plus, MultiSPMiner montre une résilience à la suppression de nœuds, maintenant une stabilité à 2% et 5% de suppression mais connaissant une légère diminution à 10% et 15% en raison de la fréquence réduite des motifs. Ces résultats soulignent l'adaptabilité du modèle à des distributions diverses de motifs.

Les orientations futures de nos travaux visent à affiner la précision de MultiSPMiner, à étendre les techniques d'apprentissage, à incorporer des architectures avancées telles que les transformateurs de graphes, et à explorer son application à des données réelles.

Références

- Del Mondo, G., M. Rodríguez, C. Claramunt, L. Bravo, et R. Thibaud (2013). Modeling consistency of spatio-temporal graphs. *Data & Knowledge Engineering* 84, 59–80.
- Ying, R., A. Wang, J. You, et J. Leskovec (2020). Frequent subgraph mining by walking in order embedding space. *The International Conference on Machine Learning (ICML) workshops*.