

Recherche de motifs fréquents dans des graphes spatio-temporels

Aurélie Leborgne*, Ezriel Steinberg*, Laurine Lafontaine*,**,
Florence Le Ber*, Antoine Vacavant***

*Université de Strasbourg, CNRS, ENGEES, ICube UMR 7357, F-67000 Strasbourg
{aurelie.leborgne,florence.le-ber}@unistra.fr

** Université Gustave Eiffel, ESIPe, F-77400 Champs-sur-Marne

*** Université Clermont Auvergne, CNRS, SIGMA Clermont, Institut Pascal,
F-63000, Clermont-Ferrand

Les progrès technologiques récents entraînent une production massive de données complexes dans de nombreux domaines. Cette complexité est liée en particulier à l'existence de relations multiples entre les objets observés, dont les relations temporelles et spatiales peuvent faire partie. Les multi-graphes sont adaptés pour modéliser de telles relations. Par exemple, Del Mondo et al. (2010) utilisent un graphe spatio-temporel (multi-graphe étiqueté orienté) pour modéliser la propagation de ronces ; Leborgne et al. (2019) utilisent ce même modèle pour représenter l'évolution des cultures de parcelles agricoles. Dans ce dernier type d'applications, les graphes spatio-temporels peuvent atteindre plusieurs dizaines de milliers de nœuds, et sont donc difficilement interprétables. Il faut donc disposer de méthodes pour les résumer, en extrayant les informations pertinentes, d'où notre proposition de méthode de recherche de motifs fréquents dans un multi-graphe étiqueté orienté.

Nous traitons ici des multi-graphes orientés aux sommets étiquetés, définis comme des tuples (V, E, f, g, L, T) où : V est l'ensemble des sommets ; $E \subseteq V \times V$ est l'ensemble des multi-arcs ; L est l'ensemble des étiquettes des sommets ; T est l'ensemble des types d'arcs ; f est une application de l'ensemble des sommets V dans L et pour tout sommet v , $f(v)$ est appelée étiquette de v ; $g : E \rightarrow 2^T$ est une fonction d'étiquetage, qui assigne un sous-ensemble de T à chaque multi-arc de E . L'algorithme LD-MuGraM que nous proposons s'appuie sur MuGraM (Ingalalli et al., 2018) qui effectue la recherche de sous-graphes dans un multi-graphe non orienté. MuGraM effectue un parcours en profondeur, à partir de « graines », *i.e.* des sous-graphes de deux nœuds dont le support est supérieur à un seuil donné. Chaque graine est décrite comme une liste des types d'arêtes reliant les deux nœuds. Chaque graine est ensuite successivement augmentée des autres graines, jusqu'à obtenir un sous-graphe non fréquent. Dans LD-MuGraM ont été apportées des modifications pour tenir compte des spécificités des graphes spatio-temporels : (i) ajout de l'étiquette dans l'indexation des nœuds ; (ii) modification des structures de données : dans LD-MuGraM, une graine est définie comme un tuple comportant l'étiquette du premier nœud, l'étiquette du second nœud, l'énumération des types des arcs allant du premier nœud au second nœud puis l'énumération des types des arcs allant du second nœud au premier nœud. Diverses optimisations ont également été réalisées.

Pour tester l'algorithme, nous avons utilisé deux types de jeux de données : des jeux de données synthétiques, obtenus par un générateur (Leborgne et al., 2021), et des jeux de données

Recherche de motifs fréquents dans un multi-graphe étiqueté orienté

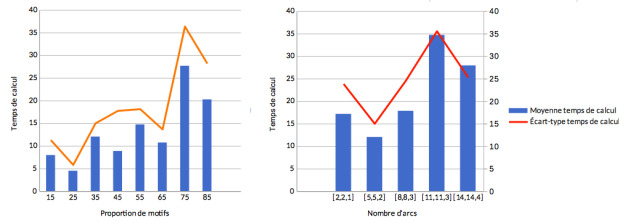


FIG. 1: Temps de recherche (mn) des motifs (a) en fonction de la proportion de motifs (motif de taille 5), et (b) du nombre d'arcs par nœud dans les motifs (variant de 5 à 32)

réelles issues d'une base de données agricole nationale. Le générateur permet de construire des graphes spatio-temporels incluant des motifs en contrôlant différents paramètres (taille des graphes, nombre et types de relations, étiquettes des nœuds, *etc.*); de plus les motifs extraits par LD-Mugram peuvent être comparés aux motifs insérés dans le graphe. Les données réelles permettent de tester LD-MuGraM sur des configurations diversifiées en nombre et répartition des étiquettes. Les premières expériences sur les données synthétiques montrent une bonne capacité de LD-MuGraM à retrouver les motifs insérés mais une augmentation rapide des temps de calcul en relation avec la densité des motifs, la densité des arêtes et le nombre de types d'arêtes et d'étiquettes des nœuds. Les figures 1(a) et 1(b) montrent les temps de calcul pour des graphes de taille 1000 (sur Ubuntu 20.04.2 version LTS, Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz). Les données réelles, où le nombre et les types d'arêtes sont plus faibles, permettent d'analyser des graphes de plus grande taille (10000 nœuds). Aux seuils les plus bas, de cent à plusieurs centaines de motifs de 4 à 7 nœuds peuvent être extraits, mais il s'agit de motifs purement temporels ou purement spatiaux.

Références

- Del Mondo, G., J. G. Stell, C. Claramunt, et R. Thibaud (2010). A graph model for spatio-temporal evolution. *Journal of Univers Comput Sci* 16, 1452–1477.
- Ingalalli, V., D. Ienco, et P. Poncelet (2018). Mining Frequent Subgraphs in Multigraphs. *Information Sciences* 451-452, 50–66.
- Leborgne, A., M. Kirandjiska, et F. Le Ber (2021). Random generation of a locally consistent spatio-temporal graph. In *ICCS 2021*, pp. 1–15.
- Leborgne, A., A. Meyer, H. Giraud, F. Le Ber, et S. Marc-Zwecker (2019). Un graphe spatio-temporel pour modéliser l'évolution de parcelles agricoles. In *SAGEO*, pp. 1–13.

Summary

This paper introduces LD-MuGraM, an algorithm based on MuGraM and designed for extracting frequent sub-graphs from directed labeled multigraphs. We focus on a specific category of multigraphs, namely spatio-temporal graphs. Our algorithm was tested both on synthetic data and on temporal land-use data obtained from French national databases.