



Sujet de thèse

Extraction de connaissances à partir de grands graphes spatio-temporels – application à hydroécologie et à la télédétection

Laboratoire d'accueil : Laboratoire ICube, UMR CNRS -- Université de Strasbourg – INSA -- ENGEES, équipe SDC

Directrice de thèse : Florence Le Ber (HDR, ICube)

Encadrante : Aurélie Leborgne (Maîtresse de Conférences, ICube) et Stella Marc-Zwecker (Maîtresse de Conférences, ICube)

Contact : florence.leber@engees.unistra.fr, aurelie.leborgne@unistra.fr, stella@unistra.fr

Contexte : Valoriser les grandes masses de données spatio-temporelles disponibles en Sciences de l'environnement est crucial. Ceci nécessite de concevoir et développer de nouvelles approches innovantes aptes à traiter conjointement les aspects spatiaux et les aspects temporels, ce qui n'est que peu le cas avec les méthodes actuelles. Si les graphes, outils puissants à la fois théoriquement et méthodologiquement, sont utilisés pour modéliser des phénomènes temporels ou spatiaux, les méthodes actuelles de fouille, d'analyse et d'extraction de connaissances n'exploitent, en général, qu'une seule dimension de l'information, spatiale versus temporelle. Cela implique souvent une perte de précision et possibilité d'interprétation des résultats [1]. Notre équipe travaille depuis de nombreuses années en collaboration avec des hydro-écologistes sur l'évaluation de la qualité des hydro-écosystèmes. Ce travail a été l'objet de plusieurs projets, dont le projet ANR Fresqueau 2011-2015 (engees-fresqueau.unistra.fr). Nous travaillons également en étroite collaboration avec le SERTIT (Service Régional de Traitement d'Image et de Télédétection, sertit.u-strasbg.fr), dont le but est de produire de l'information géographique à partir des images satellites. Cette collaboration a également fait l'objet de plusieurs projets comme par exemple le projet R&T A2CNES (2017-2018) financé par le CNES. Au travers de ces différentes collaborations, nous avons pu manipuler et fouiller des données nombreuses et variées, tant en termes d'informations apportées que de formats de données. Afin d'extraire des connaissances pertinentes de ces données, en combinant à la fois les aspects spatiaux et temporels, nous nous intéressons maintenant au modèle des graphes spatio-temporels [2].

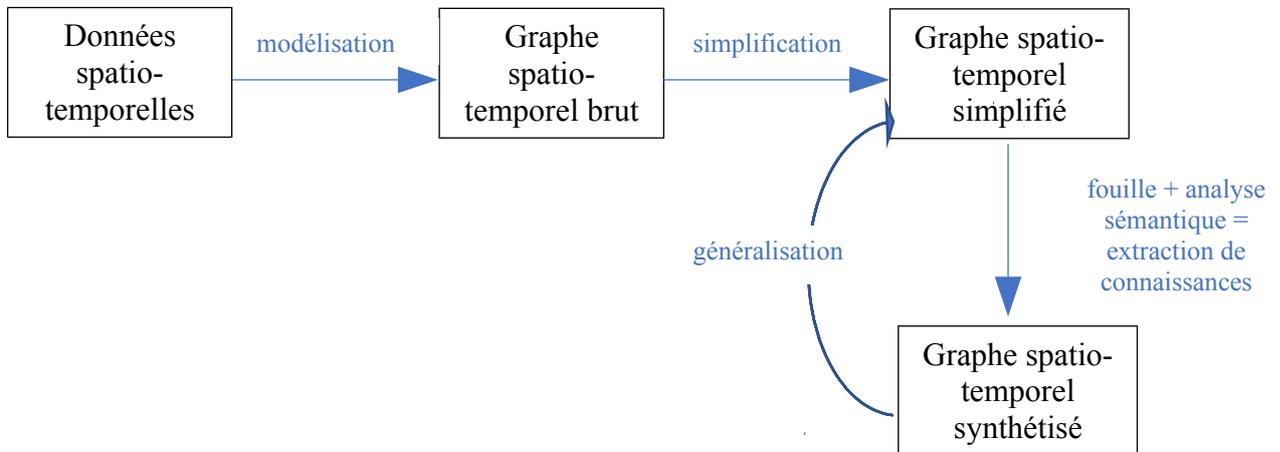
Objectifs de la thèse : L'objectif général de la thèse est d'étudier et développer des techniques pour extraire des connaissances à partir de grands graphes spatio-temporels. Différentes questions devront être abordées de manière itérative.

- Simplification : les graphes étant de grande taille, il sera nécessaire de les simplifier (par exemple, regrouper des sommets dans un super-sommet représentant l'évolution globale de cet ensemble de sommets [3]) ou de les segmenter, sur une base experte, pour réduire leur taille et les exploiter ou les visualiser plus facilement.
- Recherche de motifs : il s'agira de rechercher des motifs spatio-temporels définis par un expert (par exemple une évolution temporelle typique de l'urbanisation pour un ensemble de parcelles voisines) ou faire apparaître des répétitions fréquentes dans un ou plusieurs graphes. La recherche de motifs pourra se faire de manière inexacte, en s'appuyant sur des distances d'édition de graphes [4] et/ou sur des méthodes de recherche récursives [5].
- Synthèse interprétable : il s'agira de coupler recherche de motifs et analyse sémantique afin de synthétiser un graphe spatio-temporel par un graphe de motifs étiquetés, représentatifs de

phénomènes intéressant l'expert.

- Généralisation : dans le graphe initialement simplifié, chaque sous-graphe correspondant à un motif pourra être analysé par les méthodes développées pour finalement obtenir une hiérarchie de graphes-motifs permettant de visualiser de plus en plus finement les phénomènes ayant eu lieu.

En conclusion, le pipeline générique développé visera à répondre à des problématiques actuelles liées à la valorisation de grandes masses de données spatio-temporelles en télédétection [6] et hydroécologie [7].



Apports attendus :

- Avancées théoriques sur les méthodes de représentation de données spatio-temporelles pour l'extraction de connaissances
- Développement et mise en œuvre d'un prototype basé sur les graphes
- Interprétation et validation sur des données réelles

Profil recherché :

- Master 2 en Informatique ou équivalent
- Formation en logique, graphes et programmation
- Curiosité, capacité à appréhender différents domaines et à interagir avec les experts de ces domaines

Références :

- [1] ATLURI, Gowtham, KARPATNE, Anuj, et KUMAR, Vipin. Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys (CSUR)*, 2018, vol. 51, no 4, p. 83.
- [2] DEL MONDO, Géraldine, RODRÍGUEZ, M. Andrea, CLARAMUNT, Christophe, et al. Modeling consistency of spatio-temporal graphs. *Data & Knowledge Engineering*, 2013, vol. 84, p. 59-80.
- [3] NABTI, Chems Edinne. Subgraph Isomorphism Search In Massive Graph Data. 2017. Thèse de doctorat. Lyon.
- [4] GAO, Xinbo, XIAO, Bing, TAO, Dacheng, et al. A survey of graph edit distance. *Pattern Analysis and applications*, 2010, vol. 13, no 1, p. 113-129.
- [5] SHOKOUFANDEH, Ali, BRETZNER, Lars, MACRINI, Diego, et al. The representation and matching of categorical shape. *Computer Vision and Image Understanding*, 2006, vol. 103, no 2, p. 139-154.
- [6] BHATT, Meul et WALLGRÜN, Jan Oliver. Geospatial narratives and their spatio-temporal dynamics: Commonsense reasoning for high-level analyses in geographic information systems. *ISPRS International Journal of Geo-Information*, 2014, vol. 3, no 1, p. 166-205.
- [7] NICA, Cristina, BRAUD, Agnès, LE BER, Florence. Exploring Heterogeneous Sequential Data on River Networks with Relational Concept Analysis. In *23rd International Conference on Conceptual Structures, Proceedings*. Jun 2018, Edimbourg, United Kingdom.