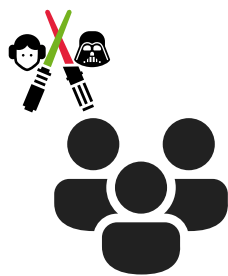

Communautés dynamiques et FCA

Apport de l'Analyse Formelle de Concepts à la détection de communautés dynamiques

Christophe Demko, Jean-Loup Guillaume, Karel Bertet



lundi 10 avril 2023

*Pour les candidatures, envoyez un courriel en utilisant [ce lien](#) en y joignant votre **CV** détaillé ainsi qu'une lettre de motivation. Nous reviendrons vers vous lorsque les calendriers d'audition auront été établis.*

Titre de la thèse

La détection de communautés dynamiques vue par le prisme de l'**Analyse Formelle de Concepts**¹

Direction de la thèse directeur·trice·s (grade, HDR) et éventuels co-directeur·trice·s

- Christophe Demko MCF (dérogation d'HDR demandée)

Co-encadrants

- Jean-Loup Guillaume PU
- Karell Bertet PU

Adéquation scientifique avec les priorités de l'établissement

Ce sujet vise à créer une synergie entre le projet **GALACTIC** actuellement financé par la SATT AST pour une maturation logicielle, le projet SmartFCA financé par l'ANR et l'étude des communautés dynamiques dont Jean-Loup Guillaume est l'un des experts. Les applications visées pourraient concerner des outils d'aide à la décision pour des problématiques d'aménagement du territoire en considérant le réseau des transports en commun comme un graphe entre les quartiers d'une ville étiquetés par des informations symboliques et numériques comme la fréquentation, l'accessibilité ou encore les mesures de la pollution.

Descriptif du sujet (enjeux scientifiques, applicatifs, sociétaux...)

La détection de communauté consiste à essayer de répartir les nœuds d'un graphe dans des groupes homogènes en utilisant sa structure. L'algorithme le plus connu est la méthode dite de Louvain [1] permettant de partitionner un réseau en optimisant une mesure appelée la modularité représentant la différence entre la densité des arêtes intra-communautaires et celle des arêtes inter-communautaires.

Les communautés peuvent être recouvrantes et le résultat obtenu n'est alors plus une partition. Elles peuvent également évoluer au fil du temps si la structure du réseau change. On parle alors de communautés dynamiques.

¹la validation finale du financement est prévue avant le 15 avril.

L'**Analyse Formelle de Concepts** peut être un regard différent sur ces communautés puisqu'elle permet de construire une hiérarchie de concepts appelée treillis des concepts où chaque concept est un couple stable (éléments, attributs). De manière intrinsèque, l'**AFC** produit une partition hiérarchique des données qui possède naturellement la capacité d'être recouvrante. En soi, la combinaison des algorithmes efficaces de détection de communauté et de l'**AFC** permettrait certainement de pallier les problèmes computationnels liés aux algorithmes existants de partitionnement de graphes de type gSpan [2].

En considérant que l'affectation d'un nœud du graphe à une communauté représente un attribut caché mais partagé avec tous les nœuds de cette communauté, on peut considérer qu'une communauté représente un concept (nœuds, attributs). De plus, dans certains cas, les nœuds du graphe peuvent être étiquetés et les attributs associés représentent une information complémentaire qui pourraient guider la détection de communautés. L'analyse relationnelle de concepts [3] est une des branches de l'**Analyse Formelle de Concepts** permettant de construire des concepts à partir de données relationnelles ; elles a certainement un lien avec la détection de communautés. L'analyse temporelle de concepts [4] est une tentative pour introduire la notion d'attributs temporels dont la valeur de vérité dépend du temps. Cette théorie peut être le point de départ pour l'analyse de communautés dynamiques.

Contexte partenarial (cotutelle internationale, EU-CONEXUS, partenariat avec un autre laboratoire, une entreprise...)

ANR smartFCA

Le laboratoire **L3i**² est lauréat, avec plusieurs laboratoires français partenaires, d'un projet financé par l'**ANR**³ visant à produire notamment, à l'horizon 2025-2026, un système de déploiement des algorithmes de l'**Analyse Formelle des Concepts**⁴ au travers d'**API RESTfull**⁵. Leur mise en œuvre est issue des travaux des équipes françaises et mondiales de ce domaine et implémente entre autres des algorithmes de la plateforme **GALACTIC**⁶ écrite en python dont certaines extensions sont le fruit d'une allocation doctorale régionale. **GALACTIC** est basée sur un nouvel algorithme NEXTPRIORITYCONCEPT [5] permettant d'étendre la génération des treillis à des contextes décrivant des données complexes et hétérogènes. Cette plateforme a aussi été étendue par un nouvel algorithme de complétion de contexte REDUCEDCONTEXTCOMPLETION dont la publication est prévue en 2023 après la soutenance de la thèse de Jérémy Richard, coencadré par Cyril Faucher et Karell Bertet.

Environnement

Ces problématiques s'inscrivent dans une logique de continuité (voir Fig. 1) au sein de l'Université de La Rochelle. Nous avons déjà évoqué le soutien doctoral ayant permis à **GALACTIC** d'étendre ses

²<https://l3i.univ-larochelle.fr>

³<https://www.smartfca.org>

⁴https://fr.wikipedia.org/wiki/Analyse_formelle_de_concepts

⁵https://fr.wikipedia.org/wiki/Representational_state_transfer

⁶<https://galactic.univ-lr.fr>

capacités aux données séquentielles et nous pouvons y ajouter la valorisation initiée par la **SATT**⁷ Nouvelle-Aquitaine au travers d'un soutien financier pour la réalisation d'une interface graphique pour l'éco-système **GALACTIC** ainsi que l'utilisation de ces outils dans le cadre d'un projet déposé pour le plan France-Relance au bénéfice de l'**OiP**-Atlantique⁸.

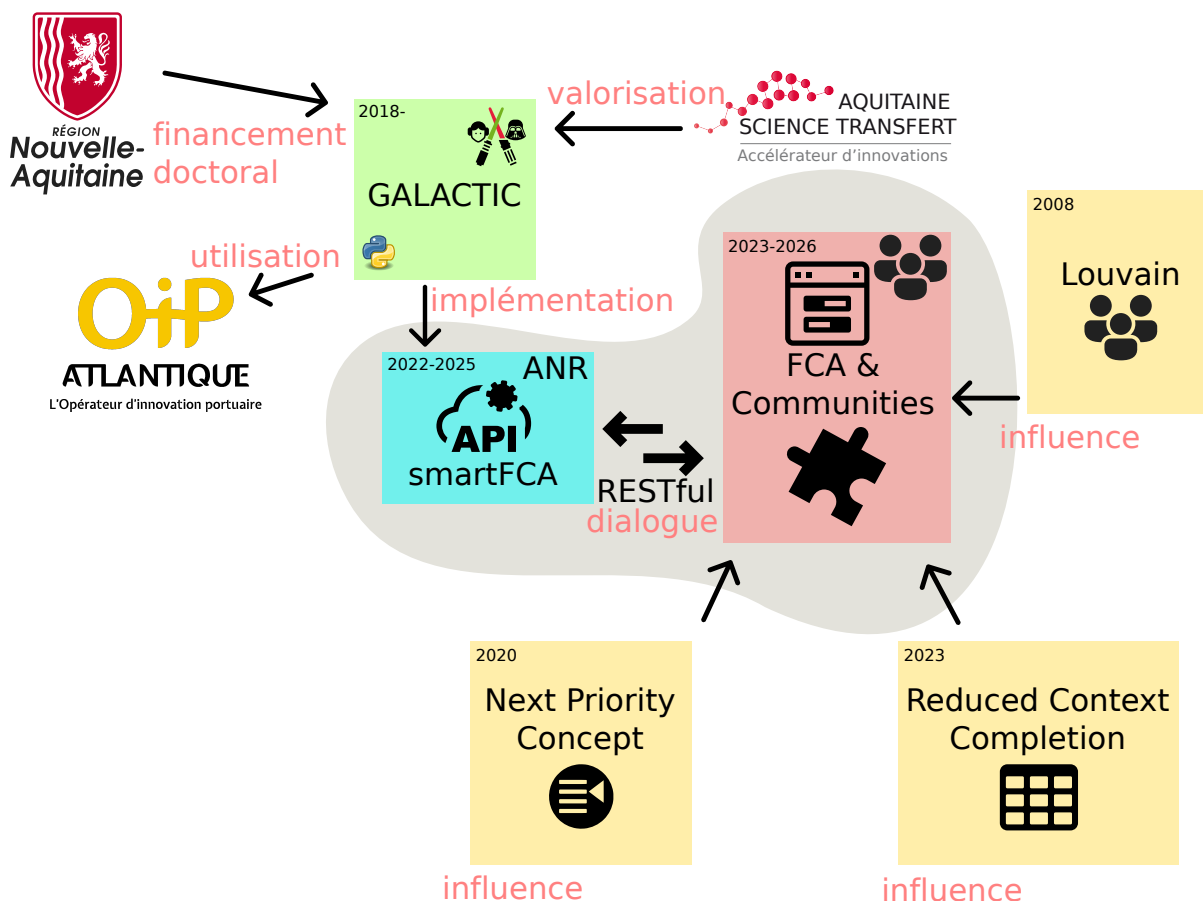


Figure 1: Contexte du projet

Ce sujet s'intègre dans la plateforme **GALACTIC** développée au L3i et pourra utiliser les résultats de l'ANR SmartFCA dont Christophe Demko coordonne le Work Package 3.

Impacts (scientifiques, technologiques, socio-économiques, environnementaux, sociétaux...)

La mise en place d'une plateforme logicielle proposant des outils novateurs d'IA qui soient à la fois explicables, interactifs, non énergivores, et adaptés à tous types de données, est une avancée scientifique majeure. L'apport conjoint de l'Analyse Formelle de Concepts et la détection de communautés permettra de l'enrichir en utilisant la structure topologique du ou des graphes en entrée.

⁷<https://www.ast-innovations.com>

⁸<https://www.oip-atlantique.fr>

Programme de travail du doctorant (tâches confiées au doctorant)

- compréhension des théories évoquées dans la description du sujet
- développement d'algorithmes utilisant conjointement l'Analyse Formelle de Concepts et la détection de communautés dynamiques
- implémentation des algorithmes sur la plateforme **GALACTIC** et la plateforme **smartFCA**

Calendrier de réalisation

- première année: analyse bibliographique et compréhension scientifique des théories évoquées
- deuxième année: formalisation et développement d'algorithmes utilisant conjointement l'Analyse Formelle de Concepts et la détection de communautés dynamiques afin de modéliser leurs évolution au fil du temps. Ces algorithmes pourront utiliser la plateforme **GALACTIC** pour prouver leur fonctionnement
- troisième année: implémentation des algorithmes sur la plateforme **SmartFCA** et rédaction du manuscrit

Accompagnement du doctorant / Fonctionnement de la thèse (accompagnement humain, matériel, financier, en particulier pour la prise en charge du fonctionnement de la thèse et des dépenses associées)

Accompagnement standard du L3i (2k€ par doctorant pour conférence internationale, cofinancement de la formation doctorale, fourniture d'un ordinateur portable neuf et puissant et ses périphériques et prise en charge des frais de soutenance de thèse).

Le doctorant sera aussi impliqué dans le projet ANR SmartFCA pour tous les travaux en lien avec ce projet. D'éventuelles missions en lien avec le projet pourront aussi être prises en charge.

Références

- [1] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment*. 2008 (2008) P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>.
- [2] X. Yan, J. Han, gSpan: graph-based substructure pattern mining, in: 2002 IEEE International Conference on Data Mining, 2002. Proceedings., 2002: p. 721-724. <https://doi.org/10.1109/ICDM.2002.1184038>.
- [3] M. Rouane-Hacene, M. Huchard, A. Napoli, P. Valtchev, Relational Concept Analysis: Mining Concept Lattices from Multi-Relational Data, *Annals of Mathematics and Artificial Intelligence*. 67 (2013) 81-108. <https://doi.org/10.1007/s10472-012-9329-3>.

- [4] K.E. Wolff, Temporal concept analysis, in: ICCS-2001 International Workshop on Concept Lattices-Based Theory, Methods and Tools for Knowledge Discovery in Databases, Stanford University, Palo Alto (CA), 2001: p. 107.
- [5] Ch. Demko, K. Bertet, C. Faucher, J.-F. Viaud, S.O. Kuznetsov, NEXTPRIORITYCONCEPT: A new and generic algorithm computing concepts from complex and heterogeneous data, Theoretical Computer Science. 845 (2020) 1-20. <https://doi.org/10.1016/j.tcs.2020.08.026>.