

# Graphe-clustering pour la détection de communautés dynamiques dans les réseaux sociaux

## Mots clés :

- **Directeur de thèse** : Marcin Detyniecki
- **Co-encadrant(s)** :
- **Unité de recherche** : Laboratoire d'informatique de Paris 6
- **Ecole doctorale** : École Doctorale Informatique, Télécommunications, Électronique de Paris
- **Domaine scientifique principal**: Divers

## Résumé du projet de recherche (Langue 1)

On s'intéresse à la problématique du clustering avec des graphes avec pour cadre applicatif principal la détection de communautés dynamiques dans les réseaux sociaux. Dans la modélisation initiale, un réseau social sera représenté par un graphe orienté et multi-varié représentant les relations entre individus. A la différence des algorithmes classiques comme l'algorithme de Louvain [De Meo et al, 2011] [Blondel et al., 2008] qui considère les profils utilisateurs dans leur intégralité pour construire les communautés, on s'intéressera dans cette thèse à la détermination de communautés recouvrantes, pour lesquelles on déterminera à la fois l'ensemble des individus concernés mais également leur sous-ensemble d'attributs caractéristiques. Ainsi un même utilisateur devra pouvoir appartenir pleinement à plusieurs communautés en fonction des attributs considérés. Cette problématique connue pour l'espaces vectoriels classiques sous l'appellation sub-space clustering [Kriegel et al., 2009, 2012] sera étudié dans le cadre de données structurées par des graphes. Les travaux conduits devraient conduire à la proposition de nouveaux modèles de clustering pour répondre à besoins énoncés, sans pour autant oublier la dynamique du réseau, à savoir l'évolution des données au cours du temps pour prendre en compte les évolutions des utilisateurs (arrivée, départ, changement de profil). Dans un deuxième temps, on s'intéressera à confronter les approches théoriques avec des jeux de données réels actuels. On sera particulièrement attentifs à l'efficacité des méthodes pour gérer la complexité observée dans le cadre des grandes masses de données, comme par exemple les données de grande dimensionnalité. Aussi, on travaillera sur le modèle initial pour qu'il soit suffisamment général pour être applicable sur tous types de données dont les relations peuvent être représentées sous forme de graphe. Finalement, l'ensemble des développements devra nous amener naturellement à la réalisation d'un démonstrateur fonctionnel et applicable à de données réelles.

## Résumé du projet de recherche (Langue 2)

clustering avec des graphes, détection de communautés dynamiques recouvrantes, passage à l'échelle, subspace clustering

## Informations complémentaires (Langue 2)

{{Références :}} [Kriegel et al., 2009] Hans-Peter Kriegel, Peer Kröger, Arthur Zimek: Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. TKDD 3(1) (2009) [Kriegel et al., 2009] Hans-Peter Kriegel, Peer Kröger, Arthur Zimek: Subspace clustering. Wiley Interdisc. Rev.: Data Mining and Knowledge Discovery 2(4): 351-364 (2012) [De Meo et al, 2011] Pasquale De Meo, Emilio Ferrara, Giacomo Fiumara, Alessandro Provetti, Generalized Louvain Method for Community Detection in Large Networks, In Proceedings of the 11th International Conference on Intelligent Systems Design and Applications ISDA'11, 2011 [Blondel et al., 2008] V. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," Journal of Statistical Mechanics: Theory and Experiment, p. P10008, 2008.