

Stage M2 - 2020

Analyse spectrale de grands graphes

Contacts, lieu et durée

- Bernd AMANN bernd.amann@lip6.fr Hubert NAACKE hubert.naacke@lip6.fr
- Equipe Bases de Données du LIP6
- 6 mois à partir de 1^{er} mars 2021

Contexte

De nombreux réseaux réels ont des arcs pondérés représentant l'intensité de l'interaction entre les paires de sommets. Aussi, certains réseaux pondérés, par exemple les réseaux financiers et les réseaux biologiques, ont une forte densité d'arêtes. Un enjeu majeur est d'interpréter ces réseaux pour comprendre les interactions qu'ils contiennent. Ce stage vise à analyser ces réseaux pour extraire les sous graphes ayant une certaine intensité et dont la structure satisfait certains critères. Par exemple, on cherche tous les sous-graphes qui ont un certain diamètre ou un certain degré moyen quand l'intensité est dans un certain intervalle. Mais lorsque les réseaux sont très denses, ce type d'analyse est souvent impossible à réaliser avec les outils existants. Face à ce problème, une méthode appelée *sparsification* consiste à réduire le nombre d'arêtes.

Le défi de la sparsification

Idéalement, on voudrait élaguer autant d'arcs que possible mais sans modifier radicalement les caractéristiques essentielles du graphe original. Or dans des réseaux complexes, les propriétés sont généralement perturbées à un stade précoce par l'élagage des arêtes dont l'intensité est inférieure à un seuil donné. Cela soulève des questions fondamentales : Quelle est la modification effective de structure du graphe induite par la suppression d'arêtes ? Peut-on mesurer cette modification de manière fiable ? Différentes techniques de sparsification, plus avancées qu'un simple élagage, ont été proposées pour répondre à ces questions difficiles [1,2,3].

La nouvelle approche spectrale

Dans le projet ANR EPIQUE [4], nous avons développé une nouvelle approche d'analyse et de visualisation de réseaux complexes sans élagage. Cette approche consiste à décomposer un réseau en ensemble de *graphes pivots* où chaque graphe pivot représente un sous-graphe avec un nœud central, appelé nœud pivot, et l'ensemble des nœuds connectés par des arêtes ayant une intensité supérieure à un seuil donné. Cette approche présente un atout saillant : elle évite de fixer un seuil au préalable, et génère pour le même nœud *tous les graphes pivots possibles*. L'ensemble de ces graphes pivots est appelé le **spectre pivot** d'un graphe. Ce modèle ouvre des possibilités inédites pour analyser et visualiser la structure de grands réseaux complexes : il permet d'explorer l'ensemble des graphes du spectre pivot en appliquant des filtres spécifiques et dépendant du contexte applicatif.

Une première implantation du modèle et d'un langage de filtrage pour des réseaux phylomémétiques [5] sur la plate-forme Apache Spark (GraphX, Dataframes) a montré la faisabilité et l'utilité de cette approche [6].

Objectifs

Ce stage a deux objectifs principaux :

- Le premier objectif est d'étudier l'application du modèle pivot à d'autres types de réseaux, comme les réseaux sociaux ou les réseaux biologiques. Il s'agira également de définir des nouvelles mesures pour le filtrage des graphes pivot adaptées à ces applications.
- Un deuxième défi important est de calculer d'une manière **efficace** les graphes pivots et des propriétés agrégées, comme par exemple leur diamètre, le degré moyens sortant, etc.

Travail à réaliser :

- Lire l'article [6] Exploring the Evolution of Science with Pivot Topic Graphs. Comprendre et prendre en main le workflow EPIQUE.
- Généraliser le modèle pivot pour d'autres types d'applications (biologie, réseaux sociaux, réseaux financiers...)
- Etudier les algorithmes et l'implantation existante pour calculer les graphes pivots.
- Proposer des nouvelles solutions pour optimiser le calcul et le stockage (compression) des spectres pivot.
- Implémenter un premier prototype du modèle généralise
- Validation expérimentale sur le cluster du laboratoire LIP6.

Prérequis :

Très bonne expérience en programmation : Java/Scala, Apache Spark

Références bibliographiques:

[1] Dianati, N. 2016. Unwinding the hairball graph: Pruning algorithms for weighted complex networks. *Physical Review E*. 93, 1 (Jan. 2016), 012304. DOI:<https://doi.org/10.1103/PhysRevE.93.012304>.

[2] Langer, N. et al. 2013. The Problem of Thresholding in Small-World Network Analysis. *PloS one*. 8, (Jan. 2013), e53199. DOI:<https://doi.org/10.1371/journal.pone.0053199>.

[3] Yan, X. 2018. Weight thresholding on complex networks. *PHYSICAL REVIEW E*. (2018), 9.

[4] EPIQUE Projet ANR <http://www-bd.lip6.fr/wiki/site/recherche/projets/epique/start>

[5] D. Chavalarias, P. Cointet. Phylomemetic Patterns in Science Evolution—The Rise and Fall of Scientific Fields, *Plos One* 2013, <https://doi.org/10.1371/journal.pone.0054847>

[6] Li KE, Bernd Amann, Hubert Naacke, Exploring the Evolution of Science with Pivot Topic Graphs: 3rd International Workshop on Big Data Visual Exploration and Analytics EDBT/ICDT 2020 Lire le [pdf](https://bigvis.imsi.athenarc.gr/bigvis2020/papers/BigVis2020_paper_3.pdf) https://bigvis.imsi.athenarc.gr/bigvis2020/papers/BigVis2020_paper_3.pdf