

▼ Accès au(x) document(s)

Accéder au(x) document(s) :

 <https://docassas.u-paris2.fr/nuxeo/site/esupversions/0b8fac0c-9239-4fc4-8046-430a74690e70>

Ce document est protégé en vertu du Code de la Propriété Intellectuelle.

Modalités de diffusion de la thèse :

- **Thèse consultable sur internet, en texte intégral.**

▼ Informations sur les contributeurs

Auteur : [Vimont Guillaume](#)

Date de soutenance : 24-10-2019

Directeur(s) de thèse : [Rougemont Michel de](#)

Etablissement de soutenance : [Paris 2](#)

Laboratoire : [Université Panthéon-Assas \(Paris\). Centre de recherches en économie du droit](#)

Ecole doctorale : [École doctorale des sciences économiques et gestion, sciences de l'information et de la communication \(Paris\)](#)

▼ Informations générales

Discipline : Informatique

Classification : Informatique

Mots-clés libres : Algorithmes de streaming, Graphes dynamiques, Clustering, Approximation

Mots-clés :

- Graphes dynamiques
- Algorithmes
- Approximation, Théorie de l'
- Grilles informatiques
- Analyse des données


Résumé : Nous étudions comment détecter des clusters dans un graphe défini par un flux d'arêtes, sans stocker l'ensemble du graphe. Nous montrons comment détecter de gros clusters de l'ordre de n dans des graphes qui ont $m = O(n \log(n))$ arêtes, tout en stockant $n \cdot \log(n)$ arêtes. Les graphes sociaux suivent le régime où m satisfait cette condition. Nous étendons notre approche aux graphes dynamiques définis par les arêtes les plus récentes du flux et à plusieurs flux. Nous proposons des méthodes simples et robustes afin de détecter ces clusters de manière approchée. Nous définissons la corrélation de contenu de deux flux (t) par la similarité de Jaccard de leurs clusters, dans les fenêtres au temps t . Nous proposons une méthode simple et efficace pour approcher cette corrélation en ligne et montrons que pour les graphes aléatoires dynamiques qui suivent une loi de puissance, nous pouvons garantir une bonne approximation. Une des applications est l'analyse des flux Twitter. Nous calculons les corrélations de contenu de ces flux en ligne. Nous proposons ensuite une recherche par corrélation où les réponses aux ensembles de mots-clés sont entièrement basées sur les petites corrélations des flux. Les réponses sont ordonnées par les corrélations, et les explications peuvent être tracées avec les clusters stockés.

▼ Informations techniques

Type de contenu : Text

Format : PDF

▼ Informations complémentaires

Entrepôt d'origine :  star
Identifiant : 2019PA020007
Type de ressource : Thèse