Detection d'anomalies contextuelles dans un graphe attribué

Rémi Vaudaine, Baptiste Jeudy, Christine Largeron

Univ Lyon, UJM-Saint-Etienne, CNRS, Institut d Optique Graduate School
Laboratoire Hubert Curien UMR 5516, F-42023, SAINT-ETIENNE, France
{remi.vaudaine,baptiste.jeudy,christine.largeron}@univ-st-etienne.fr

Résumé. La détection d'anomalies dans des données relationnelles modélisées par un graphe s'est avérée très utile dans un large éventail de domaines, par exemple pour détecter des comportements frauduleux sur des plateformes en ligne ou des intrusions sur des réseaux de télécommunication. Cependant, la plupart des méthodes existantes utilisent des prédicteurs pré-construits à partir du graphe et n'exploitent pas nécessairement des informations locales. Pour surmonter ces limites, nous proposons CoBaGAD, un détecteur d'anomalies, basé sur le contexte, qui exploite les informations locales pour détecter les noeuds anormaux d'un graphe de manière semi-supervisée. Ce modèle de réseaux de neurones, inspiré du Graph Attention Network (GAT), avec un mécanisme d'attention personnalisé permet de créer des représentations des noeuds, de les agréger et de classer les noeuds non étiquetés en normal ou anormal. Les résultats expérimentaux ont montré que CoBaGad surpasse les méthodes de pointe en terme de rappel et de précision. [1]

1 Introduction

La détection d'anomalies est un domaine de recherche intense ces dernières décennies, aussi bien pour les données relationnelles (Akoglu et al. (2015)) que vectorielles (Mehrotra et al. (2017)). Même s'il n'existe pas de définition consensuelle de ce qu'est une anomalie, elle est souvent décrite comme un écart par rapport à la norme. Dans un ensemble de graphes, l'anomalie peut concerner un des graphes alors que dans un seul graphe, elle peut se situer au niveau d'un sous-graphe, des liens ou encore des sommets. C'est ce dernier cas qui sera considéré dans la suite de cet article.

Les méthodes proposées dans la littérature visent rarement à trouver le même type d'anomalies mais elles peuvent être classées en différentes catégories selon le type de graphe qu'elles traitent. Le premier type est le graphe simple dans lequel des algorithmes de clustering tels que Scan (Xu et al. (2007)) ou PAICAN (Bojchevski et Günnemann (2018)) identifient les classes tout en détectant les anomalies. Ces anomalies sont généralement des ponts entre les communautés comme dans (Wang et Davidson (2009)). De son côté, OddBall (Akoglu et al. (2010)) est basé sur des lois de puissance et détecte des noeuds du graphe dont les caractéristiques

1. Cet article est une traduction de Vaudaine et al. (2021)