

Un algorithme de détection de communautés pour des graphes multicouches appliqué à Twitter

Youcef Abdelsadek¹, Kamel Chelghoum¹, Francine Herrmann¹, Imed Kacem¹

¹ Université de Lorraine, LCOMS, F-57045 Metz, France

{youcef.abdelsadek, kamel.chelghoum, francine.herrmann, imed.kacem}@univ-lorraine.fr

Mots-clés : *Détection de communautés, graphes multicouches, algorithme de détection, réseaux sociaux, Twitter.*

1 Introduction

De nos jours, les données générées par les réseaux sociaux, tel que Twitter, croissent de manière exponentielle. Ces données représentent généralement des relations entre des entités modélisées par des graphes [1]. Dans ce contexte, l'objectif de la détection de communautés [2] est de faire émerger les groupes d'entités (i.e., communautés) ayant une forte intra-connectivité comparée au reste du graphe. Ce travail répond au problème de la détection de communauté où plusieurs relations sont considérées simultanément, représentées par un graphe multicouche. Nous proposons un algorithme de détection de communautés dans le contexte multi-relationnel. Afin d'évaluer l'efficacité et l'applicabilité de notre algorithme, nous nous sommes basés sur les données du projet ANR INFO-RSN dont le but de comprendre la circulation de l'information sur twitter.

2 Algorithme de détection de communautés

Un graphe multicouche peut être défini comme étant un graphe simple ayant des arêtes multiples entre ses nœuds. Plus formellement, soit $G_m = (N, A_1, \dots, A_k)$ un graphe multicouche avec k couches, où chaque ensemble d'arrêtes de la $i^{\text{ème}}$ couche est noté A_i , $1 \leq i \leq k$. Les communautés induites par G_m sont notées $C_m = (C_1, \dots, C_h)$ avec h représentant le nombre de communautés détectées. Dans ce contexte, on parle communément de détection de communautés multi-relationnelles [3, 4, 5]. La question est : doit-on appliquer un algorithme de détection de communautés mono-relation [6] pour chacune des couches séparément ou bien appliquer un algorithme de détection de communautés multi-relationnelles ? Toute la difficulté réside dans la façon dont laquelle les relations sont agrégées à un niveau qui permet de préserver les relations sous-jacentes.

Pour le problème décrit précédemment, nous proposons un algorithme de détection de communautés pour des graphes multicouches qui tient compte des contraintes suivantes :

- Des attributs sont assignés aux arêtes des couches du graphe
- Le chevauchement entre les communautés est autorisé, reflétant ainsi des cas réels dans lesquels un individu appartient à plusieurs communautés simultanément par le biais de plusieurs relations différentes
- Le degré d'importance d'une relation par rapport à une autre est considéré

L'idée maitresse de l'algorithme que nous proposons consiste à attribuer des coefficients aux nœuds du graphe permettant ainsi, de refléter l'implication de chaque relation dans l'appartenance d'un individu à une communauté et son degré de participation à sa formation.

3 Application à Twitter

Dans l'optique d'évaluer la performance de notre algorithme, nous avons considéré les données du projet ANR INFO-RSN qui a comme objectif de comprendre comment l'information circule dans Twitter. Pour répondre à cet objectif, une des facettes consiste à détecter les communautés, inhérentes aux données collectées, afin d'analyser les structures communautaires qui régissent ce phénomène de propagation d'informations dans les réseaux sociaux, notamment au sein de Twitter.

Dans ce travail, notre étude s'articule autour de trois relations sociales, à savoir :

- Le *partage* qui est induit par la republication d'un tweet
- La *mention*, définie par l'action de mentionner le nom d'un utilisateur dans le texte du tweet
- Le *suivi*, la relation qui consiste à suivre les activités (i.e., tweets) d'un autre utilisateur

Chacune de ces relations sociales représente une couche du graphe multicouche. Trois jeux de données de tailles différentes ont été pris en compte dans la partie test avec 117 nœuds (3 relations), 1000 nœuds (2 relations) et 10000 nœuds (2 relations) chacun.

4 Conclusions et perspectives

Dans ce travail, nous nous intéressons au domaine de la détection de communautés pour des graphes multicouches dans le contexte des réseaux sociaux. Pour ce faire, nous proposons un algorithme de détection de communautés multi-relationnelles qui tient compte des aspects observés dans notre cas d'application. Dans le but de tester notre algorithme, nous avons conduit une évaluation sur des données issues du projet ANR INFO-RSN.

Comme perspectives, nous projetons d'étendre ce travail pour le contexte multicouche dynamique. En effet, l'évolution dans le temps des différentes relations sociales implique une dynamique structurelle des communautés. Par ailleurs, nous envisageons de traiter la problématique de l'hétérogénéité des types de relations. Effectivement, plusieurs types de relations peuvent coexister dans un graphe multicouche (p.ex., simple, attribué et orienté). Nous envisageons de visualiser des communautés obtenues à travers des représentations interactives adéquates, répondant à des besoins d'analyse bien précis.

Références

- [1] Leskovec, J., Rajaraman, A., & Ullman, J. D. Mining of Massive Datasets, 2nd Ed.. Cambridge University Press, Chapter 10, 2014.
- [2] Fortunato, S. Community detection in graphs. Physics Reports, 486, 75 – 174, 2010.
- [3] Kim, J., & Lee, J.-G. Community detection in multi-layer graphs: A survey. SIGMOD Rec., 44, 37-48, 2015.
- [4] Brodka, P., Filipowski, T., & Kazienko, P. An introduction to community detection in multi-layered social network. In: Information Systems, E-learning, and Knowledge Management Research: 4th World Summit on the Knowledge Society, WSKS 2011, 185-190, 2011.
- [5] Qi, G.-J., Aggarwal, C. C., & Huang, T. Community detection with edge content in social media networks. In Proc. 28th Int'l Conf. on Data Engineering, 534–545, 2012.
- [6] Abdelsadek, Y., Chelghoum, K., Herrmann, F., Kacem, I. & Otjacques, B. Community detection algorithm based on weighted maximum triangle packing. In: CIE45, 2015.