

La théorie des graphes pour étudier les communautés en ligne : une application d'analyse de données sur Twitch, Twitter et Youtube.

BOUCHAIB, NICOLAS, nicolas.bouchaib@first-link.fr, First Link (*Data Scientist*), Bordeaux

BAHS, NATANIEL, nataniel.bahs@first-link.fr, First Link (*CEO*), Bordeaux

Thématique : Analyse des données.

Résumé : Cette conférence présente trois approches pour comprendre les phénomènes communautaires en ligne à l'aide de la théorie des graphes. Nous utiliserons les exemples de Twitch, Twitter et Youtube pour montrer comment extraire, traiter et visualiser les données sociales pour obtenir des insights sur les communautés en ligne grâce aux network graphs.

Mots clés : cartographie des données sociales, théorie des graphes, Twitch, Twitter, Youtube, extraction, traitement, visualisation, modularité, Page Rank, Betweenness Centrality.

1) Introduction :

La cartographie des données sociales est une approche pour visualiser les données du web à l'aide de la théorie des graphes. Cette approche permet de comprendre les interactions entre les différents acteurs dans les communautés en ligne et d'obtenir des insights sur leur fonctionnement. Dans cette conférence, nous présenterons trois approches pour utiliser la théorie des graphes pour comprendre les phénomènes communautaires en ligne en utilisant l'exemple de Twitch, Twitter et Youtube. Nous expliquerons nos méthodologies, et comment nous extrayons, traitons et visualisons les données pour obtenir des insights sur les communautés en ligne.

2) Méthodologie :

Nous utilisons la partie IRC de Twitch pour en extraire les données, puis les stocker dans une base de données "orientée graphe". Nous utilisons ensuite Gephi pour relier les viewers aux streamers et ainsi visualiser ces données et obtenir des insights sur les communautés sur la plateforme. Avec cet exemple, nous présenterons deux outils mathématiques issus de la théorie des graphes : la *modularité* et le *degré entrant pondéré*, pour segmenter les communautés et mesurer les audiences uniques des streamers.

Pour Twitter, nous utilisons l'API afin d'extraire les followers des comptes. Nous utilisons à nouveau Gephi pour visualiser ces données et obtenir des insights sur les communautés et leurs centres d'intérêt sur Twitter. Avec cet exemple, l'outil de la théorie des graphes que nous présenterons sera le *Page Rank*, un indicateur probabiliste qui permet d'estimer l'influence des individus composant un réseau.

Enfin, pour Youtube, nous utilisons l'API de Google pour extraire les données de vidéos et de chaînes et pour tisser un réseau de recommandations, toujours via Gephi. L'intérêt ici est de suivre le fonctionnement de l'algorithme pour en tirer des insights. Dans l'exposé de cette

approche, la *Betweenness Centrality* sera l'outil de la théorie de graphe que nous présenterons, car il permet de détecter les *bridges* entre les communautés sur Youtube.

3) Originalité / perspective :

Cette présentation tente de montrer que la théorie des graphes est une approche pertinente dès qu'il est question d'analyser les systèmes complexes d'interaction. Cette théorie peut se décliner de manières très diverses, avec des méthodologies axées sur différents types de données et les différentes interprétations de la notion de "lien" et de "communauté".

Références

- Martin Grandjean, Mathieu Jacomy, Paul Girard. *Visual Network Analysis with Gephi*. Digital Humanities, 2016, Krakow, Poland. pp.930-931. (hal-01525569)
- Tommaso Venturini, Mathieu Jacomy, Pablo Jensen, "What do we see when we look at networks: Visual network analysis, relational ambiguity, and force-directed layouts", *Big Data & Society*, vol. 8, issue 1, First published: 21 May 2021, DOI: 10.1177/20539517211018488
- Tommaso Venturini, Anders Kristian Munk, *Controversy Mapping. A Field Guide*, Cambridge ; Medford, Polity, 2021, 385 p., EAN : 9781509544516
- <https://first-link.fr/nos-differents-types-de-cartographies/>