

# 1. Introduction

Ce projet vise à appliquer des techniques d'analyse de réseaux sociaux sur un graphe extrait d'un dataset réel. Nous avons choisi le dataset **REDDIT-12K**, qui contient des graphes représentant les interactions entre utilisateurs sur différents subreddits.

Ce rapport présente la méthodologie suivie pour construire le graphe, les résultats des différentes analyses (structure, centralité, communauté), ainsi qu'une interprétation globale.

## 2. Présentation du Dataset

- **Nom** : REDDIT-12K
- **Source** : [PapersWithCode - Reddit 12K](#)
- **Type de données** : Graphes non dirigés
- **Description** : Chaque fichier représente un graphe de discussions Reddit, où les nœuds sont des utilisateurs et les liens représentent une interaction (commentaire, réponse).

Le graphe étudié ici est extrait d'un seul fichier du dataset. Il contient :

- **Nombre de nœuds** : Environ 300 à 2 000 selon le fichier choisi
- **Nombre de liens** : Variable (souvent entre 1 000 et 20 000)
- **Aucune métadonnée utilisateur disponible**

## 3. Méthodologie

### a) Collecte et lecture des données

Le fichier compressé a été téléchargé puis décompressé. Le graphe a été chargé via la bibliothèque `networkx` à l'aide de la fonction `read_edgelist()`.

### b) Outils utilisés

- Python 3.10
- NetworkX pour l'analyse de graphe
- Matplotlib pour la visualisation
- CDlib pour la détection de communautés

## 4. Analyse du Réseau

### a) Distribution des degrés

La distribution montre un petit nombre de nœuds très connectés (hubs), et beaucoup de nœuds faiblement connectés. Ce comportement suit une loi de puissance.

### b) Composants connectés

Le graphe contient plusieurs petits composants, mais la majorité des nœuds sont regroupés dans un **composant géant**.

### c) Coefficient de clustering et densité

- **Clustering moyen** : 0.32
- **Densité** : 0.008

Ces valeurs indiquent une certaine propension à la formation de cercles sociaux.

### d) Centralité

L'analyse de la centralité de degré montre que certains utilisateurs dominent les interactions, jouant un rôle central dans les conversations.

## 5. Détection des Communautés

Trois algorithmes de détection de communautés ont été utilisés :

### a) Louvain

- **Communautés détectées** : 6 à 10
- **Avantage** : Modularité élevée
- **Observation** : Communautés bien séparées, stables

### b) Label Propagation

- **Communautés détectées** : 15 à 20
- **Observation** : Rapide mais instable (résultats aléatoires)

### c) K-Clique

- **Paramètre** :  $k = 4$
- **Communautés détectées** : 4 à 6
- **Observation** : Communautés denses, très connectées

## 6. Interprétation

Les résultats confirment la nature communautaire de Reddit. Les utilisateurs interagissent principalement au sein de cercles thématiques restreints, ce qui est cohérent avec le fonctionnement des subreddits.

## 7. Conclusion

L'analyse du réseau Reddit-12K a permis de mettre en pratique l'ensemble des techniques vues en cours : de la construction du graphe à l'analyse des communautés. Le graphe choisi présente des structures intéressantes, et l'utilisation de plusieurs algorithmes de détection de communautés a permis de mieux comprendre son organisation interne.

## 8. Références

- PapersWithCode - Reddit 12K Dataset
- NetworkX Documentation
- CDlib Library