

PROJET DE TME

PARTIE RESEAUX SOCIAUX

L'objectif du projet est d'explorer les liens entre la structure des réseaux sociaux et les calculs de similarité. Le mini-projet se compose de 2 sous-parties principales et se basera comme pour la première partie sur le jeu de données Jester Jokes. La dernière partie est beaucoup plus ouverte.

Le sujet est volontairement très ouvert pour vous inciter à faire preuve de réflexion.

1. CONSTRUCTION D'UN RESEAU SOCIAL

Les données Jester Jokes ne comportant aucune information sur les liens entre utilisateurs, il est nécessaire de construire un réseau à partir des informations disponibles : les notes. L'approche standard est de supposer que deux individus sont connectés s'ils sont similaires, ce qui revient à construire un graphe de similarité.

Dans un premier temps, il convient donc de définir les structures nécessaires à la construction d'un graphe de similarité à partir des valeurs de similarité. Il faut être capable de :

- construire un graphe pondéré dont les poids sont les similarités ;
- construire un graphe non pondéré dans lequel deux individus sont reliés s'ils sont « similaires à plus de x% ». Le x est un seuil de similarité.

Le format standard de stockage est un fichier texte avec une ligne par lien : « src dest poids ».

2. CALCUL DE COMMUNAUTES

La deuxième étape consiste à appliquer l'algorithme de Louvain vu en cours (cours 2) sur le réseau de similarité construit dans la première partie. L'algorithme fournit une partition et une qualité (modularité) qui permet de valider les résultats. Vous devez étudier les liens entre les différentes méthodes de calcul de similarité, le seuil de similarité utilisé et la qualité finale de la décomposition.

Soyez attentif à :

- la signification de la modularité qui varie grandement selon la structure du graphe (notamment sa densité) ;
- ne pas faire qu'une expérience. L'algorithme étant non déterministe, il peut fournir des résultats variables selon l'exécution.

Les sources de l'algorithme sont disponibles là : <http://sites.google.com/site/findcommunities/> et une version plus récente sera mise en ligne avant la deuxième séance sur la page de l'UE.

3. PREDICTION DE LIENS

La technique de prédiction de liens vue en cours (cours 3) ne permet pas de prédire des notes mais uniquement des apparitions de liens. Comment adapteriez-vous la technique pour faire de la prédiction de notes ? Il n'est pas nécessaire de coder une technique mais de la formuler et de proposer une technique de validation.