

# Extraction d'un réseau d'affinité à partir d'un fichier de Log

Un site web marchand propose de la documentation thématique pour ses clients (jardinage, informatique, ..). Le fichier de log du site nous permet d'identifier le type de documentation qu'a téléchargé chaque usager. Après « nettoyage », le contenu utile de ce fichier est joint en annexe (log.txt).

On vous demande de réaliser les travaux suivants :

1. Construire la matrice (usager, thème) représentant le nombre d'accès de chaque usager à un thème donné.
2. Calculez la matrice des distances\* entre usagers.
3. Transformer la matrice des distances en fichier utilisable par Pajek vous permettant de visualiser le réseau social sous-jacent. La règle établissant une relation entre deux usagers est la suivante : deux usagers sont considérés comme connectés uniquement si leur distance est inférieure à 0,5.
4. Vous visualiserez la matrice de connexion sous Pajek

Les différentes matrices seront extraites par programme (langage de votre choix)

\* On utilisera la distance de Jacquard qui correspond au complément du rapport entre la taille de l'intersection sur la taille de l'union entre deux ensembles.

$$D = 1 - \frac{(U1 \cap U2)}{(U1 \cup U2)}$$