

---

## Analyse des réseaux sociaux - ARS

### Liste de projets proposés

#### P0 Détection d'intrusions cyber

La détection d'intrusion réseau (IDS) peut se faire d'une manière supervisé ou non supervisé. Nous nous intéressons ici à la mise en place d'un IDS supervisé. Etant donné un ensemble de paquets réseaux labélisé par des experts en paquets d'attaques ou paquets normaux, nous cherchons à apprendre un modèle de classification automatique. Pour chaque paquet nous disposons outre du label, d'un ensemble d'attributs comprenant : l'adresse source, le port source, l'adresse destination, le port destination, le protocole utilisé, les drapeaux positionnées, le nombre d'octets envoyé par la source et la destination, etc. Ayant les adresses et les ports sources et destination, nous sommes en mesure de construire un réseau de connexion qui modélise les interactions entre les machines. L'objectif de ce projet est de mobiliser les techniques d'analyse des réseaux afin d'enrichir l'ensemble d'attributs d'un paquet par des mesures calculés à partir de la topologie du réseau d'interaction afin d'améliorer les performances de la classification automatique des paquets. Les données sont disponibles dans le fichier `UNSW_NB15.tar`

#### P1 Système de recommandation

Nous nous intéressons ici à l'utilisation des approches d'analyse des réseaux complexes pour le calcul de recommandations. La plupart des systèmes de recommandation se basent sur l'analyse d'historique d'évaluation (implique ou explicite) des produits par les utilisateurs afin de prédire l'intérêt d'un produit à un utilisateur donné. L'historique d'évaluation est souvent représenté sous forme d'une matrice  $A$  de dimensions  $|U| \times |P|$  où  $U$  est l'ensemble des utilisateurs et  $P$  est l'ensemble de produits. La matrice  $A$  est naturellement une matrice sparse. Un élément  $A[u_i, p_j]$  donne l'évaluation donnée par l'utilisateur  $u_i$  au produit  $p_j$ . Un algorithme de recommandation a pour objectif de prévoir les valeurs manquante dans cette matrice. La matrice  $A$  peut être vue comme la matrice d'adjacence d'un graphe bipartite qui relie les deux ensemble  $U$  et  $P$ . La problématique de recommandation se réduite alors à un problème de prévision de lien dans ce graphe bipartite. Proposer une approche de prévision de liens pour le calcul de recommandation.

On propose d'évaluer les performances de l'approche proposée sur le jeu de données d'évaluation de films MovieLens disponible sur : <https://grouplens.org/datasets/movielens/> (le jeu 100 K).

Trois approches peuvent être explorées :

- 1 Extension des approches de prévision de liens aux graphes bipartites.

- 
- 2 Application des algorithmes de détection de communautés locales.
  - 3 Application des algorithmes de détection de communautés.

## P2 : Approche de sélection de la modularité locale pour le calcul de communauté égo-centré

Nous avons vu en cours que la qualité d'une communauté égo-centré sur un noeud dépend fortement du choix de la fonction qualité (i.e modularité locale à optimiser). Nous cherchons dans ce projet à proposer un système d'aide à la sélection de la modularité locale à utiliser en fonction des caractéristiques topologiques du noeud cible. Proposer une approche d'apprentissage de la modularité à employer. On utilise les réseaux classiques de benchmark (Karaté, polblogs, dolphins, etc.) pour l'apprentissage et la validation de l'approche proposée.

## P3 : Caractérisation des réseaux complexes

Au delà des caractéristiques topologiques partagée par l'ensemble des réseaux complexes (faible densité globale, coefficient de clustering élevé, distribution de degrés selon une loi de puissance, ...), nous cherchons à caractériser la structure interne d'un réseau complexe. L'intérêt est de fournir un système d'aide au choix de l'algorithme de détection de communauté le plus adapté au réseau. Une approche de caractérisation consiste à étudier les relations entre les noeuds les plus centraux dans le réseau (selon différents centralités). Valider l'approche proposée dans le contexte d'aide à la sélection d'un algorithme de détection de communauté.

## P4 : Recommandation de liens entre pages Wikipédia

Une approche simple de recommandation de liens entre page wikipédia consiste à calculer la communauté égo-centré d'une page puis recommander de la lier la page la plus proche dans la communauté et qui n'est pas connectée à la page cible. Les résultats de la recommandation dépendent de la fonction qualité à utiliser pour le calcul de la communauté locale. Proposer une version multi-critère de l'algorithme classique de calcule de communauté locale. Valider l'approche sur le jeu de données Wikipedia disponible sur le site du cours.