

# TP1 - Proprietes standards d'un reseau complexe

SYSTEMES COMPLEXES AVANCES

M1 Informatique - semestre d'automne 2022-2023

UNIVERSITÉ CÔTE D'AZUR

Christophe Crespelle

`christophe.crespelle@univ-cotedazur.fr`

Le but de ce TP est de dresser une sorte de portrait standard d'un reseau complexe, en determinant ses quatre proprietes fondamentales vues en cours : densite, distances, distribution des degres, densite locale. Cela constituera la premiere etape de votre projet de TP. Aujourd'hui, vous menerez ces analyses sur le reseau `moreno` que vous trouverez dans le dossier `/data` accessible depuis la page web des TP. Pour votre projet vous aurez a analyser deux autres reseaux qui vous seront attribues au hasard : un de petite taille (entre 3 000 et 12 000 noeuds) et un de taille plus grande (entre 30 000 et 100 000 noeuds).

## 1 Prise en main et filtrage des donnees

Telechargez les donnees brutes du reseau `moreno` et lisez les informations dans les fichiers de description des donnees.

**Question 1.** Vous commencerez par reecrire les donnees dans un format defini a la question 2 et en les filtrant au prealable ainsi :

- enlevez toutes les informations autres que l'adjacence qui sont attachees aux aretes (poids, date, etc.)
- oubliez l'orientation des aretes
- enlevez les boucles et les aretes multiples
- gardez uniquement la plus grande composante connexe du reseau
- renumerotez les sommets en leur attribuant un identifiant entre 0 et  $n - 1$  (ou  $n$  est le nombre de sommets dans la plus grande composante)

**Suggestion.** Dans tout le TP, vous pouvez utiliser les outils que vous souhaitez. Pour cette question, une option simple est d'utiliser les outils suivants :

- `awk`, en ligne de commande, permet de scanner un fichier texte en faisant un traitement a chaque ligne. Par exemple, la commande

```
cat fichier.txt | awk '{print $2" "$1}'
```

ne garde que les deux premiers champs de chaque ligne, les inverse et les ecrit sur la sortie separees par un espace.

- `sort` et `uniq`, en ligne de commande, permettent de trier les lignes d'un fichier et d'enlever les lignes consecutives en double. Par exemple, la commande

```
cat fichier.txt | sort | uniq
```

enleve les lignes en double dans le fichier d'entree, mais modifie l'ordre des lignes.

- Pour extraire la plus grande composante connexe, vous pouvez faire un programme en C. Pour vous aider, la bibliotheque `graph.h` vous est fournie dans le dossier `/progs` accessible depuis la page web des TP. Elle permet notamment de charger en memoire un graphe lu dans un fichier au format de la question 2 et stocke en memoire dans la structure `graph` decrite dans le fichier `graph.h`.

On rappelle que les pages de manuel d'une commande systeme `cmde` sont accessibles en tapant `man cmde`.

**Question 2.** Une fois ce pre-traitement effectue, ecrivez les donnees filtreees dans un fichier texte qui contiendra :

- sur la premiere ligne : le nombre de sommets
- une arete sur chacune des ligne suivantes, en separant les identifiants des deux sommets par un espace

**Question 3.** Combien y a-t-il de sommets et d'aretes dans les donnees filtreees? combien dans les donnees originales?

Dans toute la suite, toutes les statistiques seront faites sur le reseau filtre, dans lequel, en particulier, on a conserve uniquement la plus grande composante connexe.

**Question 4.** Calculez la densite et le degre moyen du reseau `moreno`.

## 2 Distances

**Question 5.** Calculez la distance moyenne entre deux sommets du reseau `moreno`.

**Question 6.** Calculez le diametre du reseau `moreno`.

**Suggestion.** Comme dans tout le TP, vous pouvez utiliser les outils que vous souhaitez pour repondre aux questions. Neanmoins, il peut etre interessant d'implementer le calcul du diametre et de la distance moyenne par vous meme. Pour cela, vous pourrez effectuer un parcours en largeur depuis chaque sommet du graphe. Attention, cette methode sera sans doute trop couteuse en temps pour des reseaux de plus grande taille.

## 3 Distribution des degres

**Question 7.** Calculez la distribution des degres du reseau `moreno` et tracez la avec une echelle lin-lin puis avec une echelle log-log.

**Suggestion.** Le calcul de la distribution des degres est tres simplement realisable en utilisant `awk`, `sort` et l'option `-c` de `uniq`, qui compte le nombre d'occurences de chaque ligne.

**Suggestion.** Il existe plusieurs outils pour tracer des courbes a partir de fichiers de donnees. Une option qui est modereement ergonomique mais assez complete est celle offerte par `gnuplot`. Par exemple, la commande `gnuplot`

```
plot "data.txt" using 1:2 with points
```

trace la colonne 2 (sur l'axe des  $y$ ) en fonction de la colonne 1 (sur l'axe des  $x$ ) du fichier `data.txt`, dans lequel les deux colonnes sont separees par un espace ou une tabulation.

## 4 Densite locale

**Question 8.** Calculez le nombre de triangles, le coefficient de clustering global et le coefficient de clustering local du reseau `moreno`.

**Suggestion.** La encore, vous avez le choix d'utiliser des outils et implementations deja existants qui font ces calculs (<https://www-complexnetworks.lip6.fr/~latapy/Triangles/>) ou de les faire par vous meme, ce qui est tres interessant mais un peu long.

**Question 9.** Tracez la distribution du coefficient de clustering local de chaque noeud. Est-elle lisible ?

Si non, tracez plutot la *distribution cumulative inverse*, c'est a dire la courbe de la fonction  $x \mapsto P(cc_{loc}(u) \geq x)$  (ou  $cc_{loc}(u)$  est le coefficient de clustering local du noeud  $u$ ).

**Suggestion.** Pour calculer la distribution cumulative inverse, vous pouvez :

1. trier le fichier de donnees selon l'ordre inverse sur la colonne concerne, a l'aide de la commande `sort -r -k1n,1` par exemple, puis
2. faire le cumul a l'aide de `awk`.

En `awk`, les instructions sont separees par des `;` et vous pouvez initialiser des variables grace a `BEGIN{...}`. Exemple : `awk 'BEGIN{var=0;} {var+=$2; print $1" "var;}'`.

## 5 Comparaison aux modeles classiques

**Question 10.** Generez un graphe d'Erdős-Rényi de memes parametres que le reseau `moreno`.

**Suggestion.** Vous pouvez soit implementer l'algorithme vu en TD, soit utiliser le generateur telechargeable ici : <http://jlguillaume.free.fr/www/programs.php>.

**Question 11.** Suivez la meme methode de traitement des donnees et faites les memes statistiques que pour le reseau original.

**Question 12.** Generez un graphe avec le modele de configuration de memes parametres que le reseau `moreno`.

**Suggestion.** Vous pouvez soit implementer l'algorithme vu en TD, soit utiliser le generateur telechargeable ici : <http://jlguillaume.free.fr/www/programs.php>.

**Question 13.** Suivez la meme methode de traitement des donnees et faites les memes statistiques que pour le reseau original.

**Question 14.** Comparez par un tableau et une courbe le reseau `moreno` a ses deux versions aleatoires.