

Grands Réseaux d'Interaction

TP 3 : Calculs dans les graphes non-orientés

Fabien de Montgolfier
fm@irif.fr

18 février 2022

à rendre pour le 25 fevrier 23h59

1 Rappel des règles générales en TP

Les règles générales restent valides en particulier celle-ci : chaque rendu doit être constitué d'une seule archive au **format zip**, contenant un répertoire ayant le nom de votre groupe, de sorte que pour executer le TP il suffise de taper

```
unzip *.zip ; cd groupeXY ; javac TP3.java ; java TP3 graphe.txt 1234
```

2 Graphes non-orientés

Le but de ce TP est de travailler sur des graphes non-orientés : une seule arête maximum peut relier deux sommets **différents** x et y . Une **boucle** est une ligne dans le fichier où *le même sommet apparaît deux fois*, par exemple :

```
123 123
```

De telles boucles sont interdites dans un graphe non-orienté. Un **doublon** est une occurrence multiple de la même arête. Par exemple si l'on a

```
12 34
56 78
34 12
12 34
78 56
```

alors il y a seulement deux arêtes (la première entre 12 et 34, la deuxième entre 56 et 78) et trois doublons (on ne parle pas de triplon mais de deux doublons). En d'autres termes, dans un fichier de Stanford avec L lignes, s'il y a C lignes de commentaires, B boucles et D doublons, alors le nombre d'arêtes du graphes est

$$m = L - (C + B + D)$$

Un sommet x est **isolé** s'il a degré 0 ou, de façon équivalente, si et seulement si il est plus petit que le numéro max de sommet n et n'apparaît pas dans le fichier de Stanford. En effet le

fichier de Stanford utilise, pour le coefficient de clustering, le nombre de sommets **non-isolés** n_N . S'il y a n_I isolés on a

$$n = n_N + n_I$$

3 Travail à faire

En entrée, on prend deux paramètres seulement : le nom du graphe (fichier texte au format de Stanford) et le nombre de lignes de ce fichier, tel que la commande `wc -l` le donne.

En sortie, on affiche neuf nombres, un par ligne. Ce sont, dans l'ordre :

1. Nombre de **sommet non-isolés** du graphe. Il s'appelle *Nodes* dans la base de Stanford
2. **Nombre d'arêtes** du graphe (hors doublons, boucles et commentaires, donc). Attention, dans la base de Stanford c'est *Edges* seulement quand le graphe est de type *Undirected*.
3. Nombre de **boucles** qui ont été lues et ignorées
4. Nombre de **doublons** lus et ignorés (cf définition ci-dessus)
5. **Degré maximum** d'un sommet. Attention, avec le même fichier d'entrée on peut avoir une valeur différente de celle du TP1 car on est en non-orienté maintenant.
6. **Cardialité** du graphe. Attention, c'est la version non-orientée de la cardialité, donc elle peut différer de celle du TP2 pour le fichier.
7. Nombre de **triangles** du graphe. On peut le vérifier dans la base de Stanford (*Number of triangles*)
8. Coefficient de **clustering global**. On rappelle que c'est 3 fois le nombre de triangles du graphe divisé par son nombre de \mathbb{V} . Si un graphe ne contient aucun \mathbb{V} alors $cluG(G) = 0$
9. Coefficient de **clustering local moyen** nommé *Average clustering coefficient* dans la base. On rappelle que si un sommet x a 0 ou 1 voisin son clustering local est 0. Sinon, c'est la densité de son voisinage, c'est-à-dire le nombre d'arêtes reliant deux voisins de x , divisé par le maximum qu'il pourrait y avoir. Puis on calcule en divisant par le **nombre de sommets non-isolés**

$$\frac{\sum_x cluL(x)}{n_e}$$

Des exemples sont donnés dans `resultats.txt`.