

RAPPORT

Mini-projet de TP : analyse de réseaux réels

Janvier 2023

Nicolas Guiblin

Étudiant en Master 1 Informatique

Enseignant – Mr. Christophe Crespelle

Année scolaire 2022-2023

Table des matières

1	Graphe US Power Grid	3
1.1	Analyses standards	3
1.2	Distribution des degrés du réseau US Power Grid	4
1.3	Distribution des degrés du réseau d'Erdos Renyi associé à US Power Grid.....	5
1.4	Distribution des degrés du réseau aléatoire associé à US power Grid	5
2	Graphe Flickr.....	6
2.1	Analyses standards	6
2.2	Distribution des degrés du réseau Flickr	7
2.3	Distribution des degrés du réseau d'Erdos Renyi associé à Flickr.....	8
2.4	Distribution des degrés du réseau aléatoire associé à Flickr	8
3	Annexes	9

1 Graphe US Power Grid

Ce réseau non dirigé contient des informations sur le réseau électrique des États de l'Ouest des États-Unis d'Amérique. Une arête représente une ligne d'alimentation. Un nœud est soit un générateur, un transformateur ou une sous-station.

1.1 Analyses standards

Après formatage des données nous obtenons les informations suivantes sur le graphe ainsi que sur le graphe de Erdos-Renyi associé et au graphe aléatoire associé en respectant la configuration du graphe initiale.

	US Power Grid	Graphe Erdos-Renyi	Graphe aléatoire
Sommets	4941	4520	4533
Arêtes	6594	6547	6671
Densité	0.000540	0.000641	0.000649
Degré moyen	2.669095	0.000641	2.943304
Distance moyenne entre 2 sommets	18.989185	2.896902	8.297107
Diamètre	46	8.511283	19
Triangles	651	7	3
Coefficient de clustering global	0.103153	0.000515	0.000487
Coefficient de clustering local	0.080103	0.000431	0.000801

Similitudes :

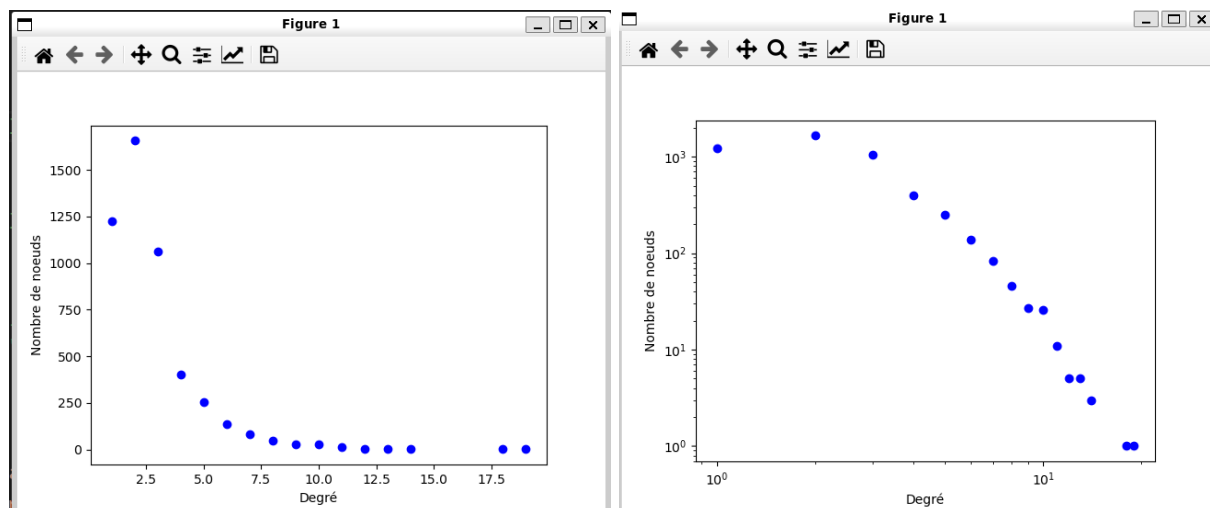
- Tous les trois ont un nombre de sommets et d'arêtes relativement similaires.
- Tous les trois ont des densités relativement proches, indiquant qu'ils ont un nombre similaire d'arêtes par rapport à leur nombre de sommets.
- Le degré moyen (nombre moyen de liens par sommet) est similaire dans les trois graphes.

Différences :

- Le graphe US power grid a une distance moyenne entre deux sommets beaucoup plus élevés que les deux autres graphes, indiquant qu'il faut passer par un nombre plus élevé de sommets pour aller d'un sommet à un autre dans ce graphe.
- Le diamètre (longueur du plus long chemin entre deux sommets) est plus élevé dans le graphe US power grid que dans les deux autres graphes.
- Le coefficient de clustering global (mesure de la tendance des sommets d'un graphe à former des triangles) est beaucoup plus élevé dans le graphe US power grid que dans les deux autres graphes.
- Le coefficient de clustering local (mesure de la tendance d'un sommet à être connecté à d'autres sommets qui sont eux-mêmes connectés) est plus élevé dans le graphe aléatoire que dans les deux autres graphes.

Cela est dû à la caractéristique du réseau électrique, qui est conçu pour être étendu afin de couvrir de vastes territoires et de fournir de l'électricité à de nombreuses personnes. Le fait que le graphe US power grid ait un nombre de sommets et d'arêtes plus élevé, une densité inférieure et un degré moyen inférieur peut être dû au fait que le réseau électrique comprend de nombreuses lignes d'alimentation et de nœuds (générateurs, transformateurs et sous-stations), qui sont espacés de manière à couvrir une grande zone. Cela peut expliquer pourquoi la distance moyenne entre deux sommets du graphe US power grid est plus élevée et pourquoi le diamètre du graphe est plus grand que ceux des deux autres graphes. Le fait que le graphe US power grid ait des coefficients de clustering global et local plus élevés peut être dû au fait que le réseau électrique comprend de nombreux groupes de nœuds qui sont reliés entre eux et qui sont conçus pour fonctionner de manière autonome, ce qui peut contribuer à former des clusters ou des groupes de sommets dans le graphe.

1.2 Distribution des degrés du réseau US Power Grid



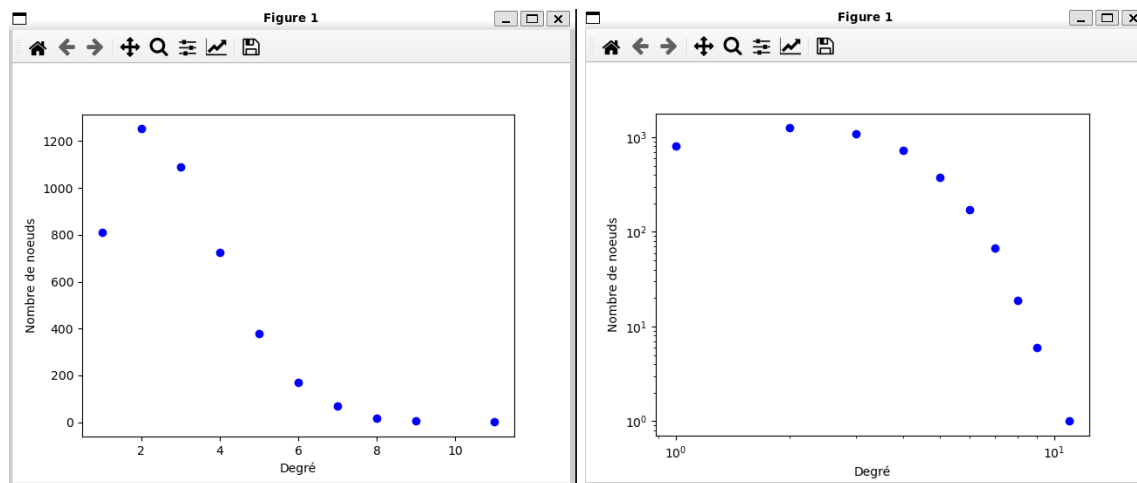
En analysant la distribution des degrés du graphe US power grid, on peut émettre les hypothèses suivantes :

La forte concentration de sommets avec un degré relativement faible (2 et 3) peut être due au fait que la plupart des générateurs, transformateurs et sous-stations du réseau électrique ne sont connectés qu'à un ou deux autres éléments du réseau (par exemple, un générateur peut être connecté à un transformateur et à une sous-station, mais pas à d'autres générateurs).

Le faible nombre de sommets avec un degré élevé peut être dû au fait que seuls quelques éléments du réseau (par exemple, certaines sous-stations ou transformateurs) sont connectés à un grand nombre d'autres éléments du réseau.

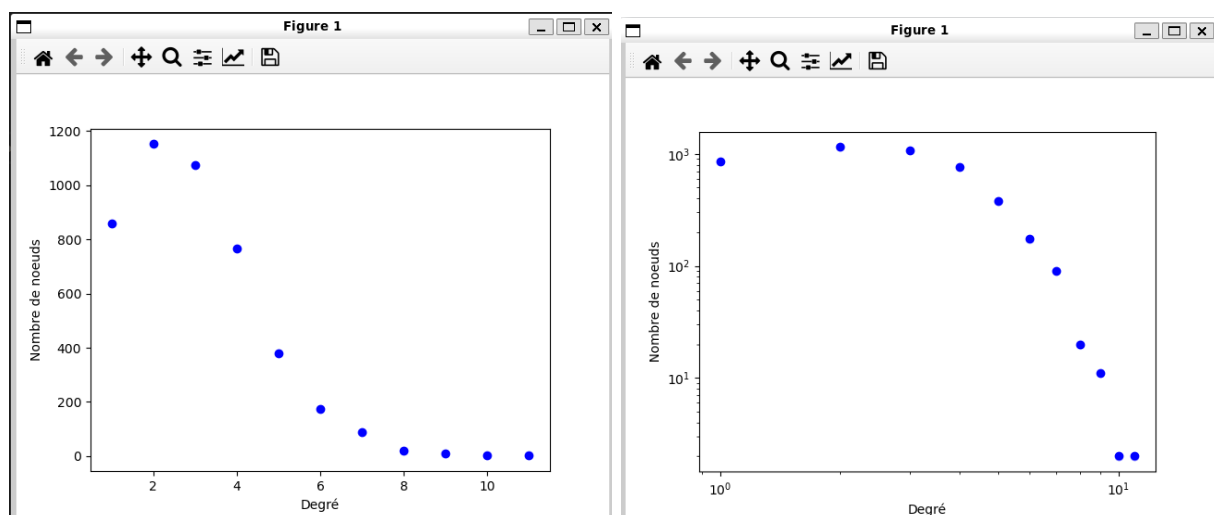
Le sommet avec un degré très élevé (18) peut être un transformateur ou une sous-station qui est connecté à un grand nombre de générateurs, de transformateurs et de sous-stations.

1.3 Distribution des degrés du réseau d'Erdos Renyi associé à US Power Grid



La distribution des degrés du graphe d'Erdos-Renyi qu'il y a une forte concentration de sommets avec un degré relativement faible (2 et 3), ainsi qu'un nombre relativement faible de sommets avec un degré élevé (supérieur à 6). Cela peut indiquer que la plupart des sommets du graphe sont peu connectés, et qu'il y a peu de sommets qui ont un nombre élevé de liens.

1.4 Distribution des degrés du réseau aléatoire associé à US power Grid



La distribution des degrés du graphe aléatoire comporte une forte concentration de sommets avec un degré relativement faible (2 et 3), ainsi qu'un faible nombre de sommets avec un degré élevé (supérieur à 6). Le nombre de sommets avec un degré élevé (8, 9, 10, etc.) est similaire à celui observé dans la distribution des degrés du graphe US power grid, mais inférieur à celui observé dans la distribution des degrés du graphe d'Erdos-Renyi. Aucun sommet n'a un degré très élevé (supérieur à 10).

2 Graphe Flickr

Il s'agit du réseau non dirigé d'images Flickr partageant des métadonnées communes telles que des balises, des groupes, des emplacements, etc. Un nœud représente une image et un bord indique que deux images partagent les mêmes métadonnées.

2.1 Analyses standards

Après formatage des données nous obtenons les informations suivantes sur le graphe ainsi que sur le graphe de Erdos-Renyi associé et au graphe aléatoire associé en respectant la configuration du graphe initiale.

	Flickr	Graphe Erdos-Renyi	Graphe aléatoire
Sommets	105 722	105 722	105 722
Arêtes	2 316 668	2 318 754	2 318 886
Densité	0.000415	0.000415	0.000415
Degré moyen	43.825656	43.865118	43.867615
Distance moyenne entre 2 sommets	4.33771	Inconnu	Inconnu
Diamètre	9	Inconnu	Inconnu
Triangles	323 960 853	14 034	13 989
Coefficient de clustering global	0.401599	0.000414	0.000413
Coefficient de clustering local	0.088402	0.000414	0.000413

Similitudes :

- Nombre de sommets et d'arêtes relativement similaires.
- Densités relativement proches.
- Degré moyen similaire.

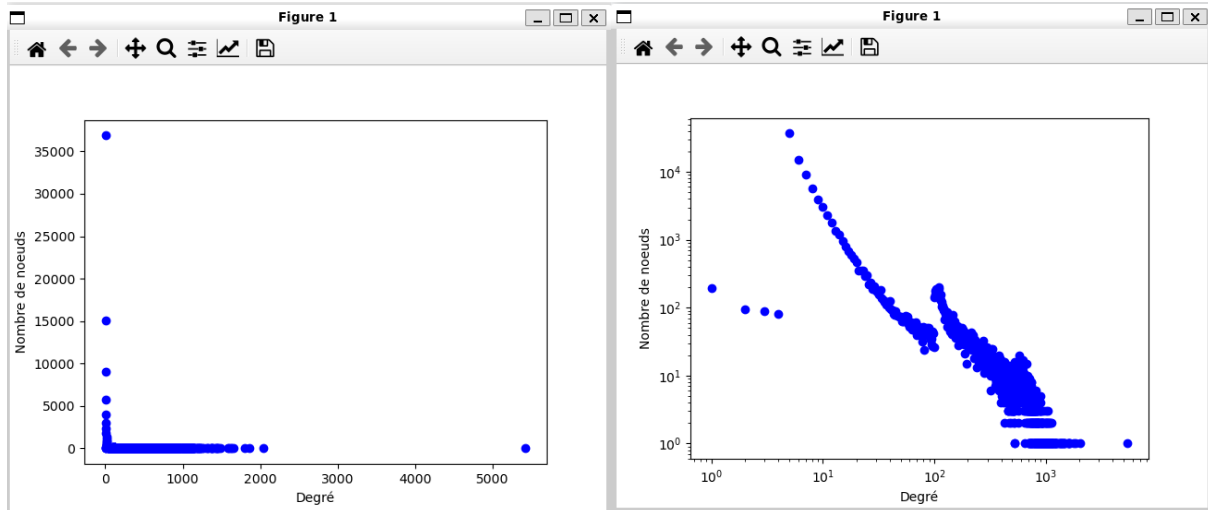
Différences :

- Flickr :
 - Distance moyenne entre deux sommets la plus faible.
 - Diamètre le plus petit.
 - Triangles et coefficient de clustering global les plus élevés.
- Erdos-Renyi :
 - Triangles et coefficient de clustering global les plus faibles.
 - Distance moyenne entre deux sommets et diamètre inconnus.

- Aléatoire :
 - Triangles et coefficient de clustering global légèrement inférieurs à ceux du graphe Erdos-Renyi.
 - Distance moyenne entre deux sommets et diamètre inconnus.
 - Le coefficient de clustering local (mesure de la tendance d'un sommet à être connecté à d'autres sommets qui sont eux-mêmes connectés) est plus élevé dans le graphe aléatoire que dans les deux autres graphes.

Cela est dû à la nature du réseau de partage d'images sur Flickr, qui permet aux utilisateurs de partager des images et de les étiqueter avec des métadonnées communes telles que des balises, des groupes et des emplacements. Le fait que le graphe Flickr ait un nombre de sommets et d'arêtes supérieur, une densité égale et un degré moyen légèrement supérieur peut être dû au fait que de nombreuses images sur Flickr sont étiquetées avec des métadonnées communes et que de nombreux utilisateurs partagent des images sur le site. Cela peut expliquer pourquoi la distance moyenne entre deux sommets du graphe Flickr est faible et pourquoi le diamètre du graphe est petit. Le fait que le graphe Flickr ait des coefficients de clustering global et local plus élevés peut être dû au fait que de nombreuses images sur Flickr partagent des métadonnées communes et que de nombreux utilisateurs forment des groupes pour partager des images avec des intérêts communs, ce qui peut contribuer à former des clusters ou des groupes de sommets dans le graphe.

2.2 Distribution des degrés du réseau Flickr

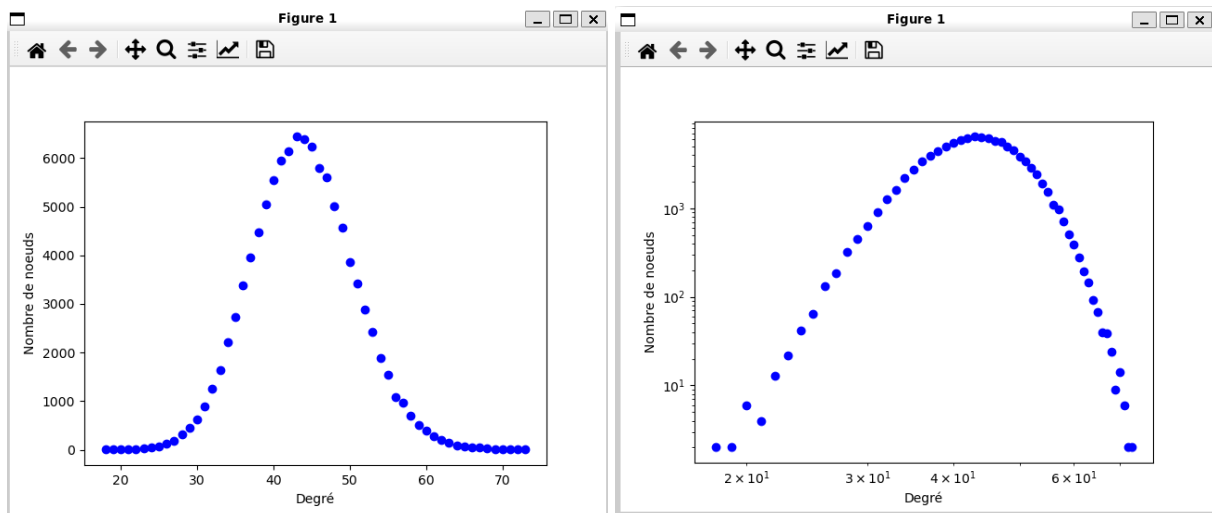


En analysant la distribution des degrés du graphe Flickr, on peut émettre l'hypothèse que la distribution des degrés du graphe reflète la manière dont les images sur Flickr sont liées entre elles par des métadonnées communes.

Le fait que la plupart des sommets du graphe aient un degré relativement élevé (supérieur à 100) indique que de nombreuses images partagent un grand nombre de métadonnées communes. Le fait qu'il y ait quelques sommets avec un degré exceptionnellement élevé (supérieur à 1 500) peut être expliqué par le fait que ces images ont été particulièrement populaires et ont été partagées avec de nombreuses autres images ayant des métadonnées communes.

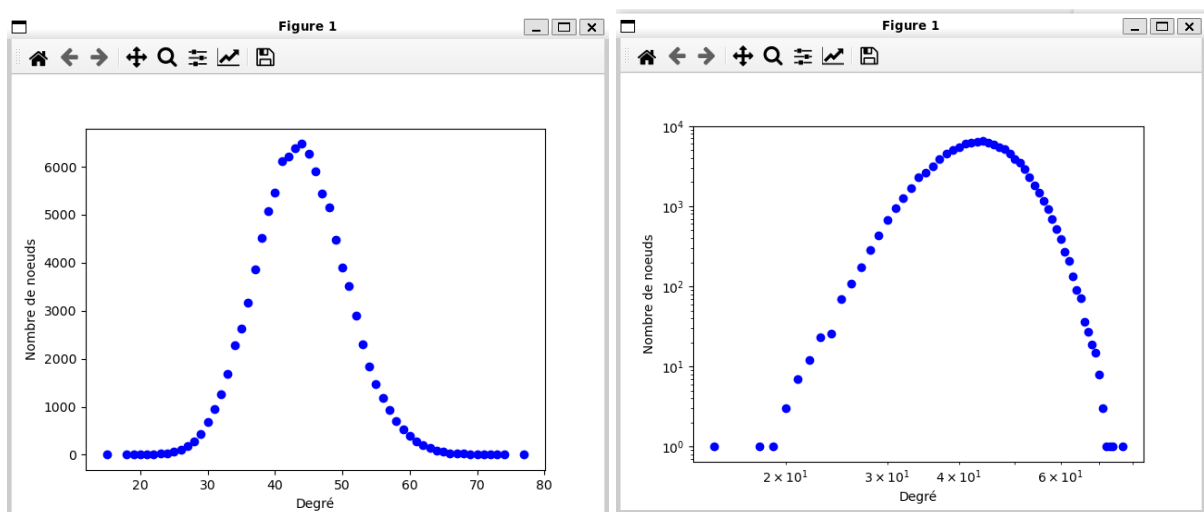
Le fait qu'il y ait également un petit nombre de sommets avec un degré relativement faible (inférieur à 50) peut être expliqué par le fait que ces images ont été partagées avec moins d'autres images ayant des métadonnées communes, ou qu'elles ont été associées à des métadonnées moins courantes.

2.3 Distribution des degrés du réseau d'Erdos Renyi associé à Flickr



On peut voir que la plupart des nœuds ont entre 39 et 49 liens, mais il y a également quelques nœuds ayant très peu de liens (moins de 10) ou beaucoup de liens (plus de 50). Cela indique que le graphe a une distribution de degré plutôt "plate", c'est-à-dire qu'il n'y a pas de nœuds qui dominent largement en termes de nombre de liens.

2.4 Distribution des degrés du réseau aléatoire associé à Flickr



Les distributions sont presque voire totalement identiques, on en conclut la même analyse.

3 Annexes

Vous trouverez le code source comportant toutes les fonctions utiliser pour traiter les réseaux dans le fichier `tp.py` du dossier `src` ainsi que les données des graphes au format `txt` dans le dossier `data`.