

# Systemes Complexes Avances

## Mini Projet : Analyse de r seaux r els

Winona Pasquier

# Introduction

L'objectif de ce projet est de réaliser des analyses sur deux réseaux provenant de données réelles. Nous avons dans ce projet un petit et un grand réseau.

Le petit réseau [1] contient les données de collaboration dans des articles scientifiques. Si un auteur A a coécrit un article avec un auteur B. Alors le graphe contiendra une arête entre A et B. Donc si l'article est coécrit pas plusieurs auteurs, un sou graph complètement connecté sera généré. Les données couvrent les articles publiés entre janvier 1993 et avril 2003.

Le grand réseau contient les données d'amitié des utilisateurs de Facebook[2]. Un nœud représente un utilisateur et une arête représente une amitié entre deux utilisateurs. Les données correspondent à l'état d'une partie réduite du réseau d'amitié de Facebook en 2009.

Toutes les représentations graphiques ont pu être effectuées à l'aide de la librairie NetworkX[3].

## Traitements des données brutes.

La première étape à effectuer a été de filtrer les données des deux réseaux donnés. Cela comprend la vérification de plusieurs choses :

- Pas d'arêtes dupliquées.
- Pas de boucle (une arête de A vers B et une autre de B vers A)
- Pas de troue dans les données, pas d'index incohérents (17, 18, 20). Pour cette partie il a été nécessaire de modifier le petit réseau. En effet, les index dans ce graph sont sous la forme (3466,10310,5052...). Nous avons donc modifié les index de telle sorte qu'on ait des index sous la forme (1,2,3, 4...).

## Statistiques des réseaux

### Nombre de sommets et arêtes

Réseau	GrQc		Facebook	
Données	Originales	PCC*	Originales	PCC*
Sommets	5242	4158	63731	63392
Arêtes	14496	13422	817035	816831

\*plus grande composante connexe

### Densité et degré moyen

Réseau	GrQc	Facebook
Densité	0.002	0.00040
Degré Moyen	6.5	25.6

Pour le réseau GrQc en moyenne un article est cité 6 fois dans d'autres articles. De plus, la densité est de 0.002 cela veut dire que le graph n'est pas dense. En effet, tous les articles ne peuvent pas tous se citer entre eux. On va donc obtenir des communautés d'articles qui se citent entre eux.

On peut donc dire pour le réseau Facebook qu'en moyenne chaque utilisateur est connecté à environ 26 personnes. Également la densité est de 0,0004 ce qui signifie que le graph n'est pas dense. En effet, une personne ne peut avoir toutes les autres personnes du graph en amis. Il est

donc normal que la densité de ce graph soit faible. Il y aura en effet des groupes de personnes se connaissant mais pas un unique groupe.

## Distance moyenne entre deux sommets du réseau

Pour calculer la fonction de distance et diamètre du réseau le plus grand, il a été nécessaire de créer une fonction pour avoir une valeur approximative de la distance et du diamètre, avec 1000 nœuds à échantillonner. En effet les fonctions `nx.average_shortest_path_length` et `nx.diameter` ont une complexité de  $O(n^3)$ . En calculant une approximation, la complexité passe à  $O(n)$ .

Réseau	GrQc	Facebook
Distance	6.0	4.3
Temps d'exécution	23s	6min

## Diamètre des réseaux

Réseau	GrQc	Facebook
Diamètre	17	14
Temps d'exécution	20s	6min

## Distribution des degrés des réseaux

### Réseau GrQc

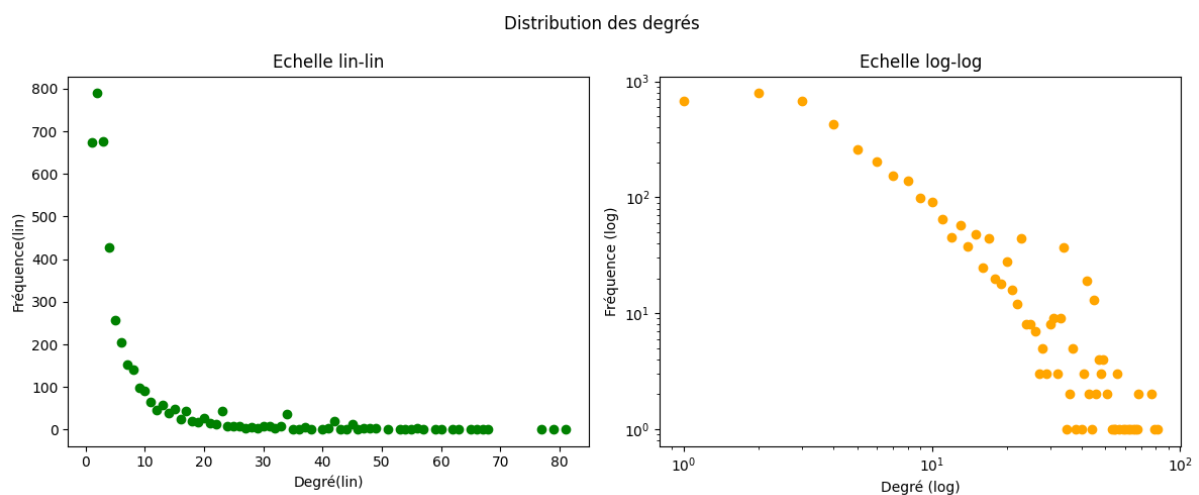


Figure 1 Distribution des degrés pour le réseau GrQc

## Réseau Facebook

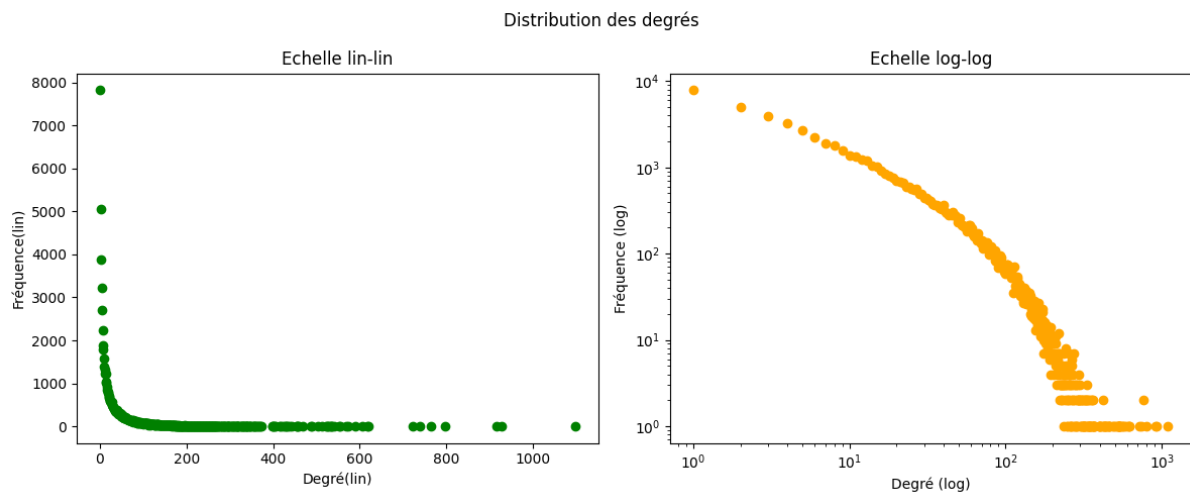


Figure 2 Distribution des degrés pour le réseau Facebook

## Coefficient de clustering global et nombre de triangles

Reseau	GrQc	Facebook
Nombre de triangles	47779	3500534
Clustering Global	0.56	0.22

## Distribution du coef de clustering local

Les premiers graphiques, sur la gauche, représentent la distribution cumulative normale du réseau. Les deuxièmes, à droite, sont la distribution cumulative inversée.

Le coefficient de clustering local est la mesure de la tendance d'un nœud à former des clusters avec ses voisins. On peut l'utiliser pour savoir le degré auquel les nœuds du graph se regroupent le plus. Une valeur élevée indique que les voisins d'un nœud sont également voisins entre eux, ce qui forme un cluster de nœud. A l'inverse une valeur faible indique que les voisins d'un nœud ne sont pas connectés entre eux.

## Réseau GrQc

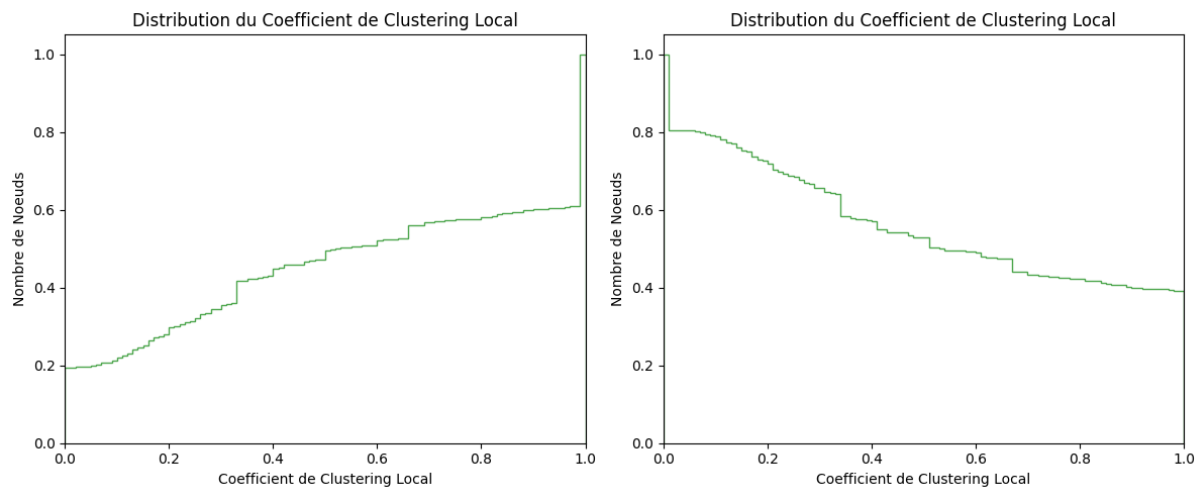


Figure 3 Distribution du coefficient de clustering local du réseau GrQc

## Réseau Facebook

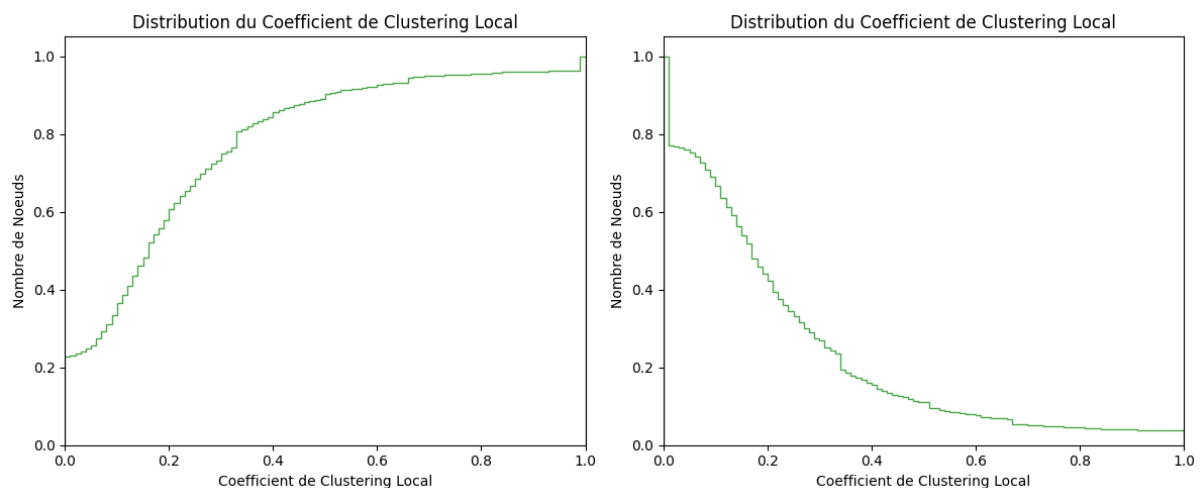


Figure 4 Distribution du coefficient de clustering local du réseau Facebook

Le réseau GrQc montre des distances moyennes plus élevées ainsi qu'un coefficient de clustering plus important. Ce qui suggère que c'est un réseau fortement regroupé. Le réseau Facebook présente des distances moyennes plus courtes et un coefficient de clustering plus faible, indiquant une structure plus dispersée et moins regroupée.

## Analyse des réseaux

Pour commencer j'ai décidé de représenter les graphs en fonction de leur communauté. La couleur du nœud correspond à la communauté à laquelle il appartient. Plus le nœud est grand plus ça betweenness centrality est important.

Community structure of the graph

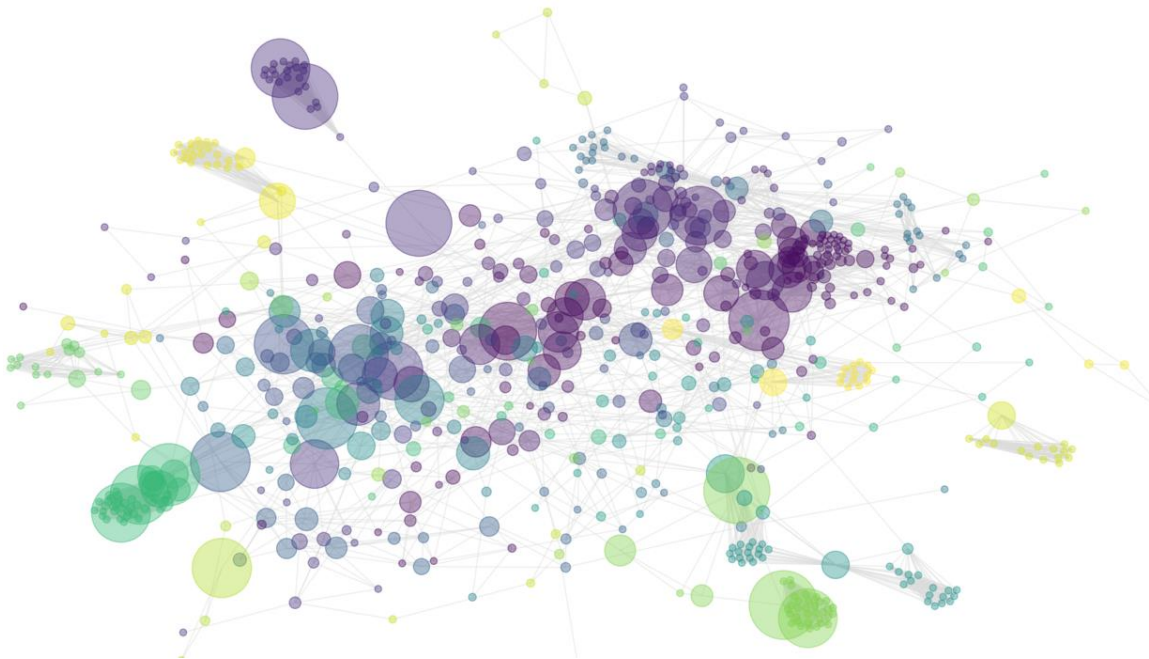


Figure 5 Représentation des communautés pour le réseau GrQc

On peut supposer que les nœuds les plus grands sont les articles les plus cités étant donné qu'ils ont une betweenness centrality plus importante.

Community structure of the graph

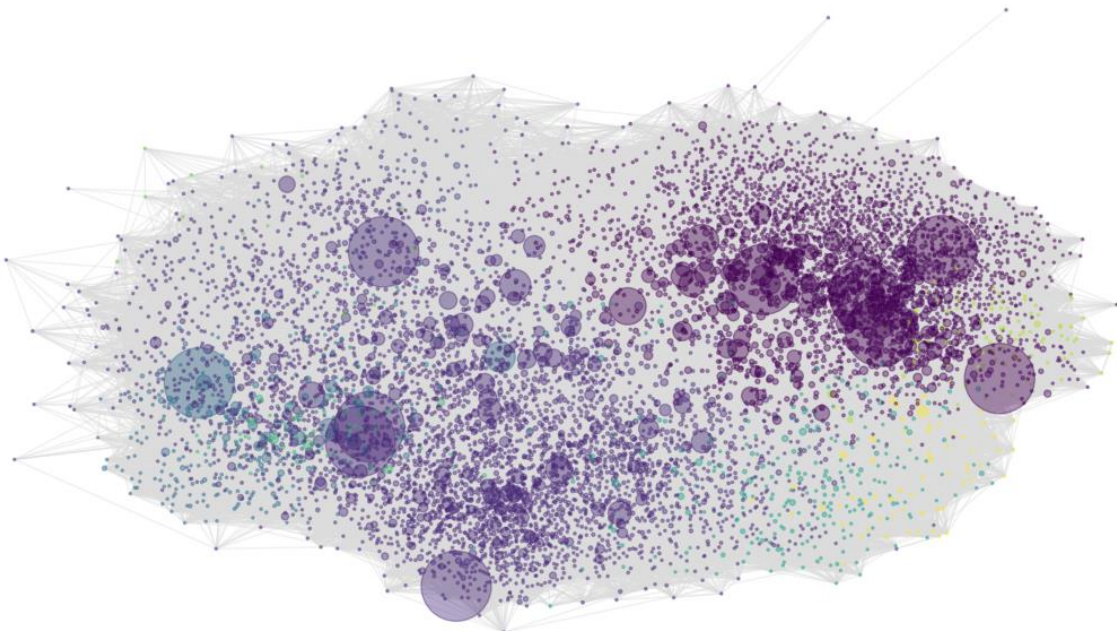


Figure 6 Représentation des communautés pour le réseau Facebook

On peut voir que des nœuds dans ce graph ont une très forte betweenness ce qui suggère que ce sont les personnes les plus suivies.

### 3. Centralités

#### Question 6 :

Pour cette question j'ai choisi d'étudier deux centralités la *closeness centrality* et la *betweenness centrality*.

- La *closeness centrality* permet de savoir à quel point un nœud est proche de tous les autres nœuds du réseau. Un nœud avec une grande closeness peut interagir rapidement avec les autres nœuds du réseau.
- La *betweenness centrality* mesure le nombre de fois où un nœud apparaît dans les chemins les plus courts. Un nœud avec une grande betweenness est un nœud central dans le réseau.

#### GrQc

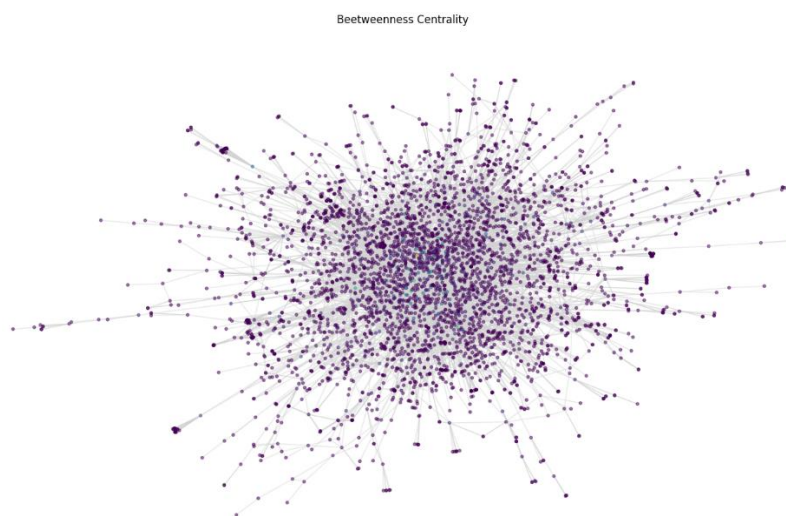


Figure 7 Représentation de la Betweenness Centrality pour le réseau GrQc

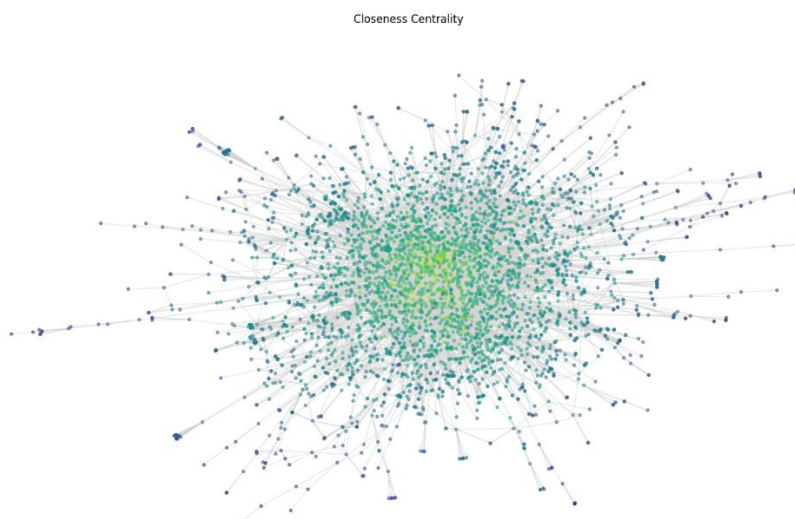


Figure 8 Représentation de la Closeness Centrality pour le réseau GrQc

#### Facebook

Pour afficher la représentation et les calculs de betweenness et closeness centrality j'ai choisi de conserver une partie réduite du réseau Facebook. Les représentations ci-dessous sont des



versions zommés des résultats obtenus. Des nœuds n'étaient reliés à aucun autre nœud et n'avaient donc pas d'intérêt à être conservés pour la représentation de la betweenness et closeness. La betweenness et closeness sans l'application du zoom se trouve dans [\[Annexe 1\]](#).

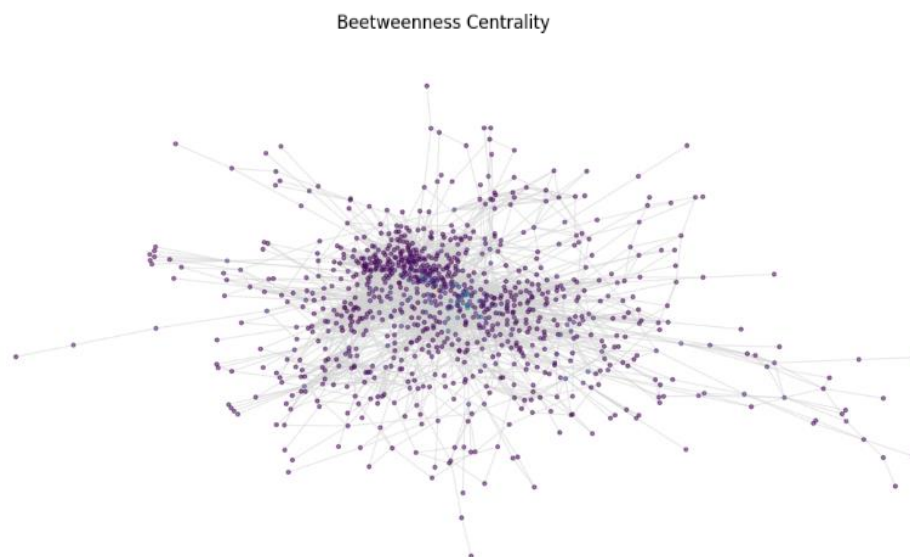


Figure 9 Représentation de la Betweenness Centrality pour le réseau Facebook

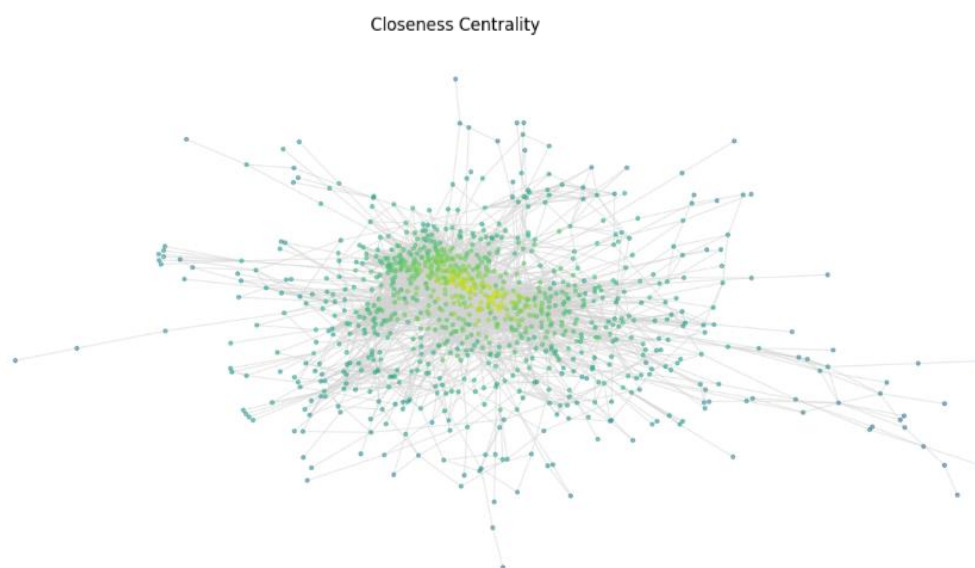


Figure 10 Représentation de la Closeness Centrality pour le réseau Facebook

Pour calculer la corrélation entre les mesures de closeness et de betweenness j'ai décidé d'utiliser la corrélation de Spearman. Elle est utile pour savoir si notre betweenness et closeness sont souvent élevées en même temps ou si elles sont souvent faibles en même temps. Spearman est utile pour détecter ces tendances même si les données ne suivent pas une ligne droite.

Réseau	GrQc	Facebook
Correlation de Spearman	0.52	0.62
Temps d'exécution	1min	45min



## Question 7

Pour ces deux réseaux nous avons une corrélation de classement de 0.52 pour le réseau GrQc et de 0.62 pour Facebook.

Dans Facebook il y a un grand nombre de connexion entre les personnes. Les utilisateurs les plus populaires sont les nœuds centraux. Ils sont susceptibles d'être perçus de manière identique par une grande partie du réseau, contribuant ainsi à une corrélation plus élevée des classements des nœuds.

Dans le réseau GrQc, les interactions sont plus spécifiques et limitées. Les liens entre les articles sont basés sur des collaborations entre auteurs ou des citations. De ce fait, les critères de classement peuvent varier davantage, ce qui contribue à une corrélation plus faible entre les classements des nœuds.

Reseau	GrQc	Facebook
Corrélation de Spearman	0.52	0.62

## Question 8

Les types de nœuds où les deux mesures sont d'accord sont les nœuds avec une :

- **Faible** betweenness centrality et une **Faible** closeness centrality.
- **Forte** betweenness centrality et une **Forte** closeness centrality.

A l'inverse les types de nœuds où les deux mesures ne seront pas d'accord sont les nœuds avec une :

- **Faible** betweenness centrality et une **Forte** closeness centrality.
- **Forte** betweenness centrality et une **Faible** closeness centrality.

## Conclusion

En conclusion, ce projet d'analyse a permis d'étudier le réseau GrQc et le réseau de Facebook.

Le réseau GrQc possède un nombre modéré de sommets et d'arêtes, avec une densité faible (tous les articles ne se citent pas entre eux). Le réseau Facebook présente un grand nombre de sommets et d'arêtes avec une densité très faible. Cela illustre la nature dispersée du réseau.

L'étude des centralités de betweenness/closeness a permis d'identifier les nœuds centraux dans chaque réseau. Dans le réseau GrQc, les nœuds les plus centraux sont les articles les plus cités, tandis que dans le réseau Facebook, ce sont les utilisateurs les plus populaires qui ont une position centrale.

Également, on a pu observer une corrélation plus élevée dans le réseau Facebook. Cela peut être dû à la connectivité importante entre les utilisateurs. En revanche, la corrélation est plus faible dans le réseau GrQc sûrement en raison de la nature plus spécifique des interactions entre chaque article.

# References

[1] [General Relativity and Quantum Cosmology collaboration network](#)

[2] [Facebook](#)

[3] [NetworkX](#)

## Annexes

### Annexe 1 : Betweenness et Closeness Centrality pour une partie du réseau Facebook

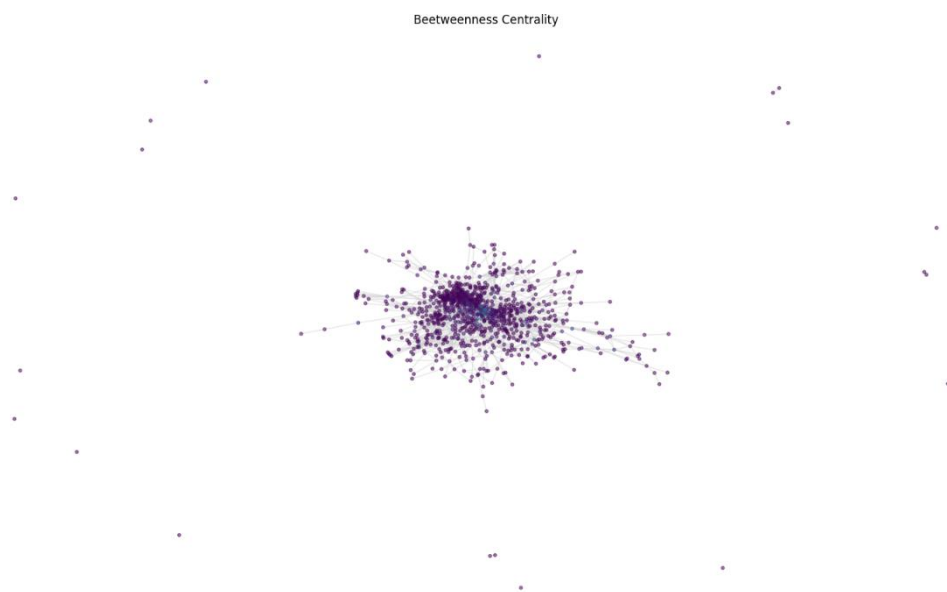


Figure 11 Représentation de la Betweenness Centrality pour le réseau Facebook

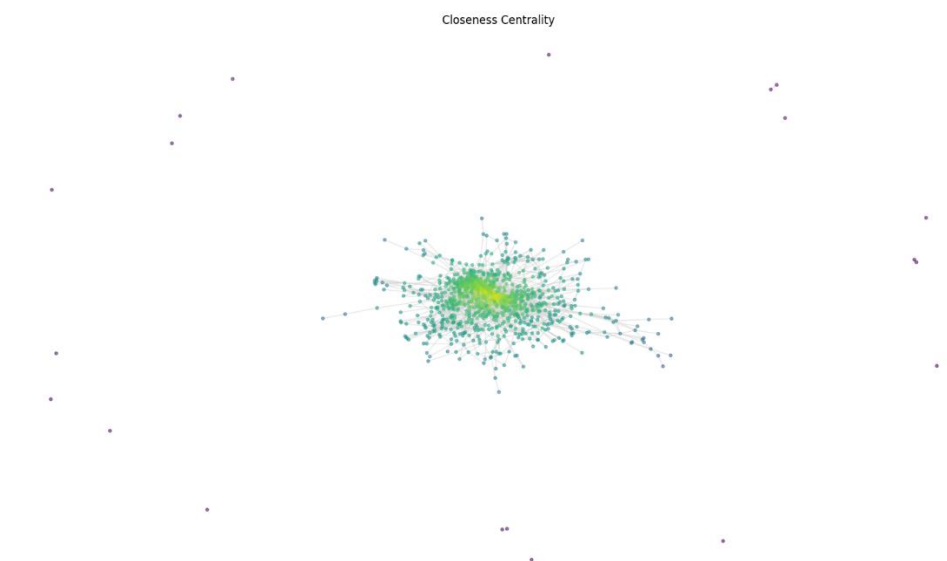


Figure 12 Représentation de la Closeness Centrality pour le réseau Facebook