

RAPORT Du PROJET

Analyse des réseaux sociaux : Cas d’étude sur Facebook





1. Introduction

Dans le cadre de ce projet, j’ai fait une étude sur l’Analyse des Réseaux Sociaux, qui a une méthodologie permettant d’examiner les interactions et les influences entre individus ou organisations. Utilisée dans plusieurs domaines, elle aide à comprendre les dynamiques sociales.

Les plateformes comme Facebook offrent des données précieuses pour analyser les liens entre individus, détecter des communautés et mesurer l’influence. Grâce à ces analyses, il devient possible d’observer comment les tendances et les informations se propagent au sein d’un réseau.

Ce projet s’inscrit dans cette démarche en exploitant les données Facebook issues du dataset facebook\_combine.txt, disponible sur le **Stanford Network Analysis Project (SNAP)**, afin d’explorer les structures et interactions qui façonnent un réseau social en ligne.

Etude Du Dataset

Le dataset utilisé dans ce projet est composé de **4 039 nœuds** et **88 234 arêtes**, représentant les relations d’amitié entre les utilisateurs de Facebook. Chaque utilisateur est un **nœud**, et chaque lien d’amitié est une **arête**, formant ainsi un réseau structuré.

L’analyse de cette structure permet d’explorer la **connectivité du réseau**, d’identifier les **communautés** et de comprendre les **dynamiques d’influence** entre les individus. Grâce aux techniques d’**Analyse des Réseaux Sociaux (ARS)**, il devient possible de mieux appréhender l’organisation et l’évolution des interactions au sein du réseau.

1. Méthodologie

Dans cette section de méthodologie, on commencera par l’étude des données, ce qui nécessite tout d’abord d’établir un processus d’étapes pour mener à bien le travail demandé .Ces étapes sont les suivantes :

* **Prétraitement des données et création d’un graphe**

Pour mener cette tâche, on va suivre ses points suivants:

*1. Importation des bibliothèques nécessaires*

Nous importerons les bibliothèques essentielles telles que **pandas**, **networkx**, **numpy**, **community**, etc.

*2. Téléchargent et lecture du dataset*

Le fichier **facebook\_combine.txt** sera téléchargé et lu afin d’explorer son contenu. La lecture de ce fichier nous permettra de visualiser les données.

*3. Affichage des informations de base*

À l’aide de la fonction print(), nous afficherons des informations essentielles telles que le nombre de **nœuds** et d'**arêtes** présentes dans le dataset

*4. Représentation du Réseau*   
Dans notre modèle, chaque utilisateur est un nœud, et les arêtes symbolisent les amitiés entre ces utilisateurs. Ce réseau est donc non dirigé, avec des connexions bidirectionnelles entre les nœuds.

* **Analyse de métriques topologiques globales et structurelles**

Dans cette étape, nous commencerons par l’étude des métriques topologiques globales du réseau, qui incluent les calculs suivants :

* **Calcul de la densité** du graphe
* **Calcul du diamètre** du graphe
* **Calcul du coefficient de clustering**

Ensuite, pour l’analyse structurelle, nous nous concentrerons sur l'étude de la **centralité**. Cette analyse consistera à explorer différentes mesures de centralité afin d’identifier les nœuds les plus influents du réseau. Les mesures de centralité que nous étudierons sont les suivantes :

* **Centralité de degré** : Mesure le nombre d’arêtes connectées à un nœud, indiquant sa connectivité directe.
* **Centralité de proximité** : Évalue la distance moyenne d’un nœud à tous les autres nœuds du réseau, montrant l’accessibilité d’un nœud à l’ensemble du réseau.
* **Centralité d’intermédiarité** : Mesure le nombre de fois qu’un nœud apparaît sur le plus court chemin entre deux autres nœuds, indiquant l'influence d'un nœud dans les flux de communication du réseau.
* **Centralité d’autonomie (ou centralité de vecteur propre)** : Évalue l'importance d'un nœud en fonction de l'importance de ses voisins, ce qui permet de déterminer les nœuds qui ont une influence sur d'autres nœuds influents.
* **Détection de communautés**

Cette étape consiste à :

* **Calculer et afficher** le nombre de communautés détectées dans le réseau.
* **Visualiser les communautés** sur le graphe en utilisant l'algorithme de Louvain, qui est un algorithme efficace pour la détection de communautés dans les réseaux complexes.
* **Diffusion d'information**

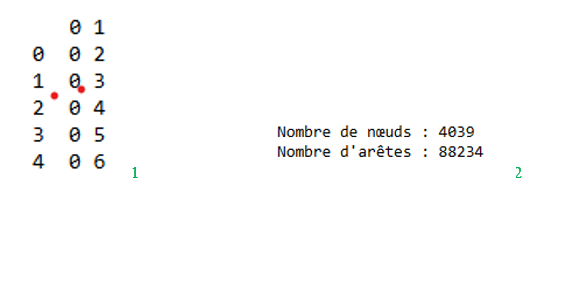
Cette étape consiste à étudier la propagation de l'information en utilisant les nœuds les plus influents identifiés grâce aux différentes mesures de centralité. Nous procéderons ensuite à la **comparaison des résultats** obtenus à partir des différentes simulations de diffusion.

1. les résultats de l’Etude

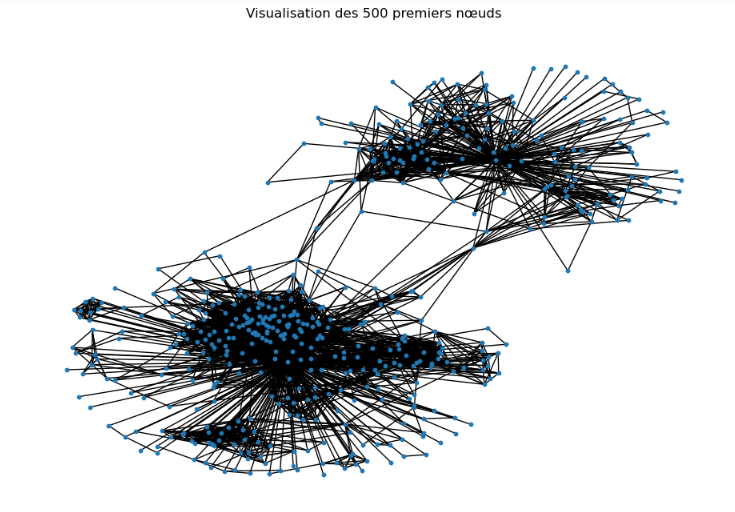
Dans cette section, nous présentons les résultats des analyses effectuées sur le réseau social à partir des données traitées. Les résultats sont organisés en fonction des étapes clés du processus méthodologique.

1. **Prétraitement des données et création d’un graphe**

Cette partie a permis de connaitre après étude des données la structure du dataset (1) ainsi que ces nombres de nœuds et d’arêtes. (2).



Par la suite, on a visualisé une partie du graphe dont 500 nœuds pour avoir une aperçue du graphe.



Cette visualisation montre comment la connectivité entre les nœuds est structurée.

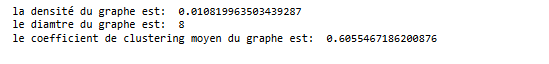
1. **Analyse de métriques topologiques globales et structurelles**

Cette partie d’analyse a montré pour les métriques topologiques que :

-la densité du graphe (0.0108) a de nombreuses connexions manquantes dans le réseau.

-le diamètre du graphe(8) : les utilisateurs sont relativement proches les uns des autres.

-le coefficient de clustering moyen : le réseau est bien structuré en communautés car les utilisateurs sont fortement connectés au sein de leurs groupes sociaux.

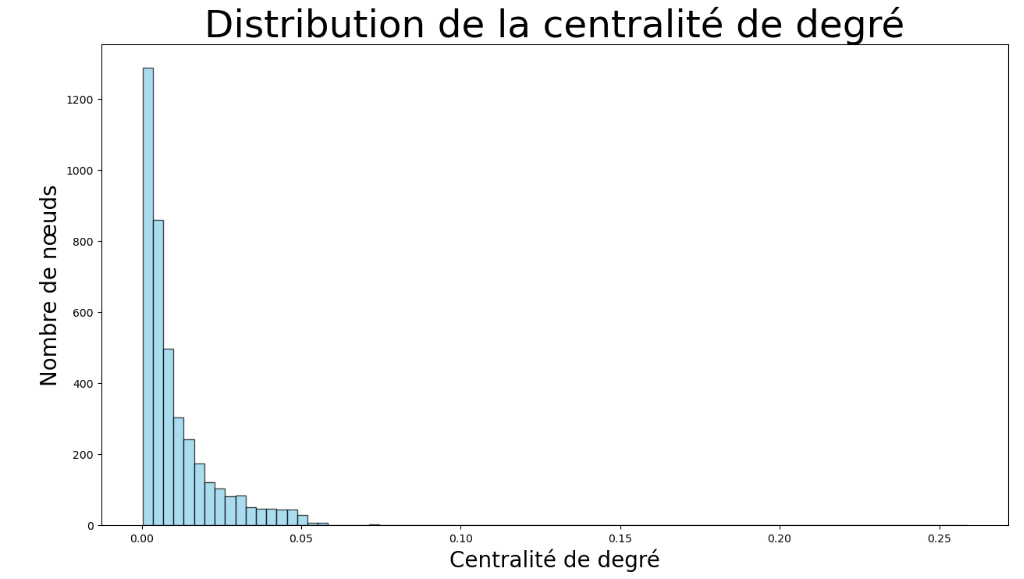


**2) Analyse métrique**

* Centralité de degré

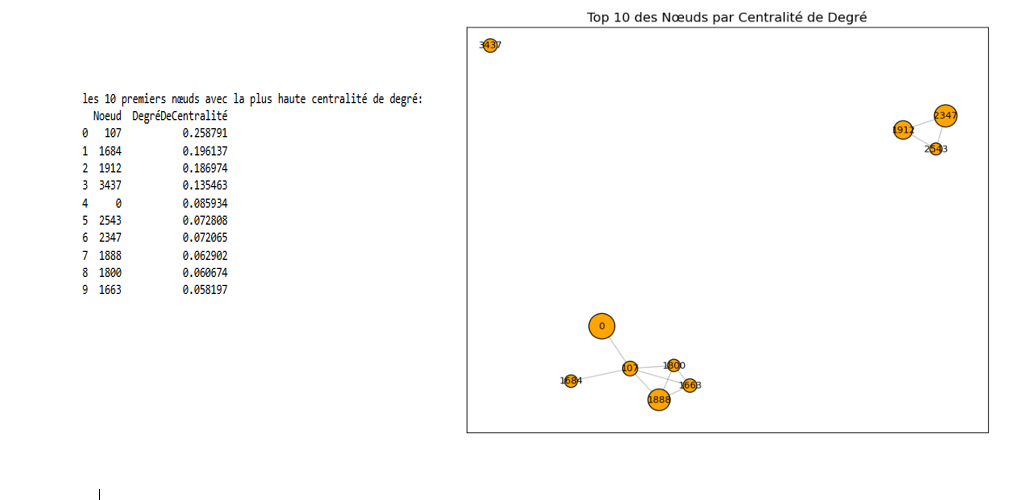
*Analyse de la distribution*

Cette distribution montre que la centralité de degré des nœuds est extrêmement faible tournant autour de 0.00 à 0.05.Cela dû au fait à une interconnexion réduite.



*Analyse des top10 nœuds de centralité de degré*

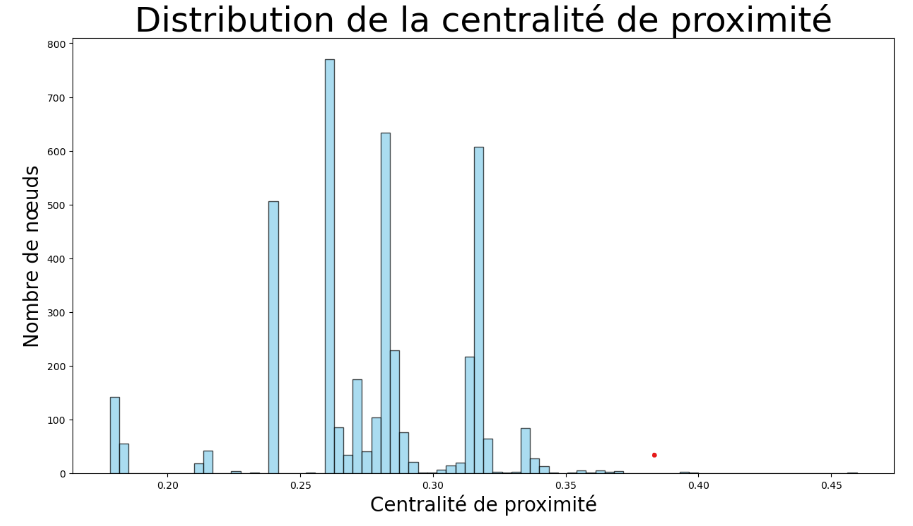
L’analyse des nœuds de centralité de degré élevé montre que le nœud 107 a plus de connectivité avec les autres nœuds soit un degré de centralité de 0.258791.



* Centralité de proximité

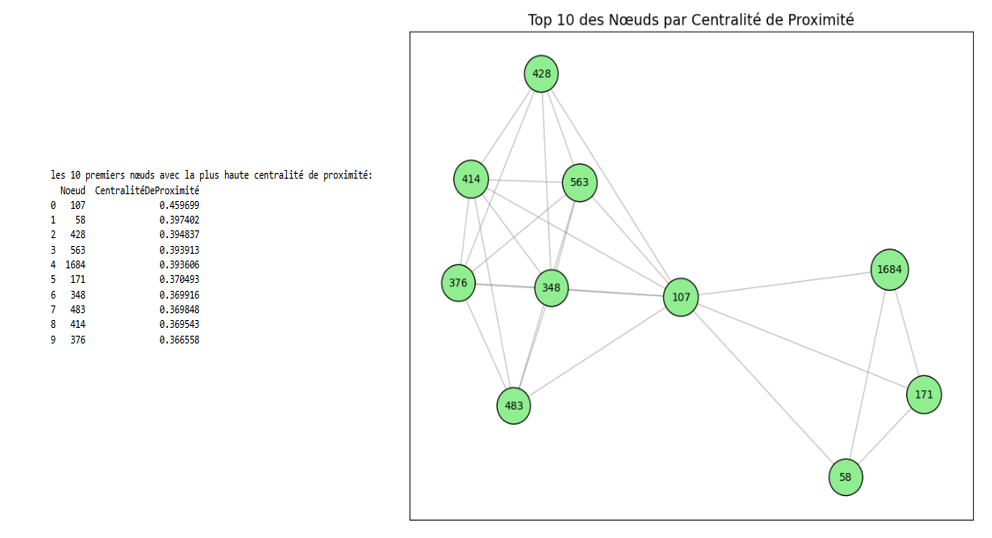
*Analyse de la distribution*

Pour la centralité de proximité, on trouve qu’il a une variabilité des valeurs de la centralité de proximité. Cela montre la proximité de certains nœuds entre eux.



*Analyse des top10 nœuds de centralité de proximité*

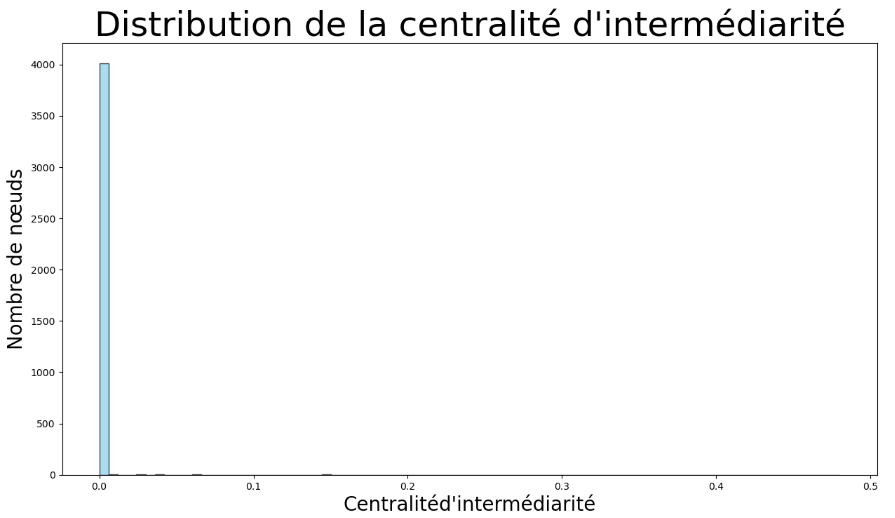
Ce graphe des nœuds par centralité de proximité montre comment les nœuds sont proches entre eux .Cela dit que le nœud 107 est le nœud le plus proche. De plus on peut même apercevoir que leurs valeurs de centralité ne sont pas aussi éloignées.



* Centralité d’intermédiarité

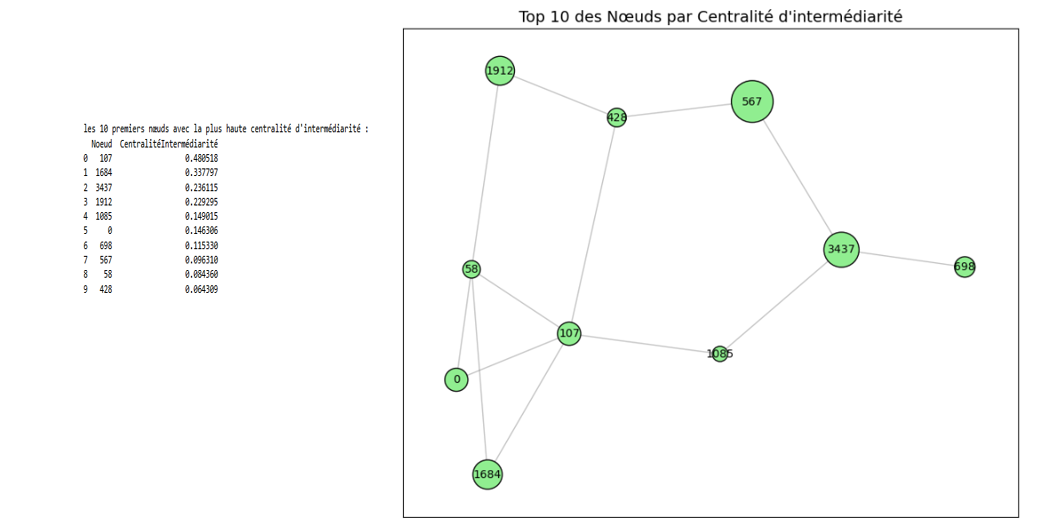
*Analyse de la distribution*

Cette distribution montre que la plupart des nœuds leurs centralités d’intermédiarité est de 0.Cela due au fait que la plupart des nœuds n’agissent pas comme des ponts dans les chemins les plus courts.



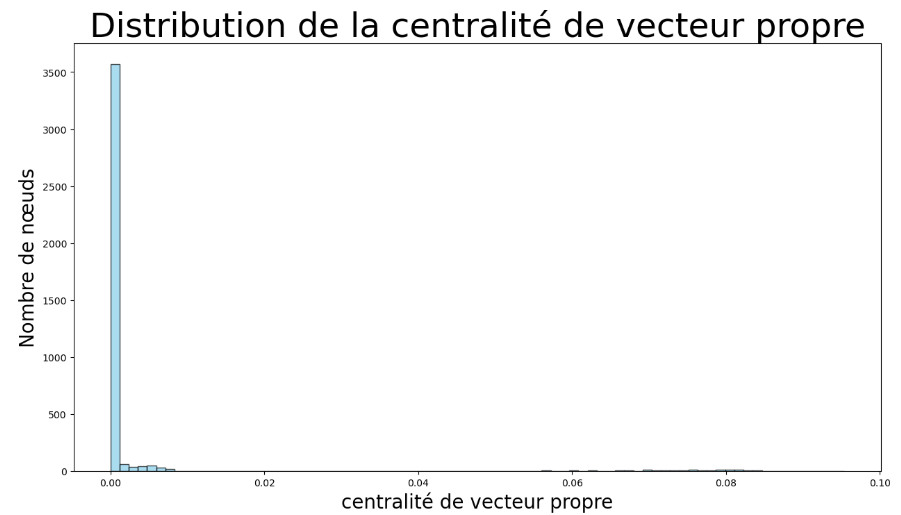
*Analyse des top10 nœuds de centralité d’intermédiarité*

On constate ici que la centralité d’intermédiarité du nœud 107 est plus élevée montrant ainsi comme la connexion est présente sur lui. Cela montre par lui les informations peuvent aboutir aux autres nœuds vu qu’il est des plus influents.



* Centralité de Vecteur Propre

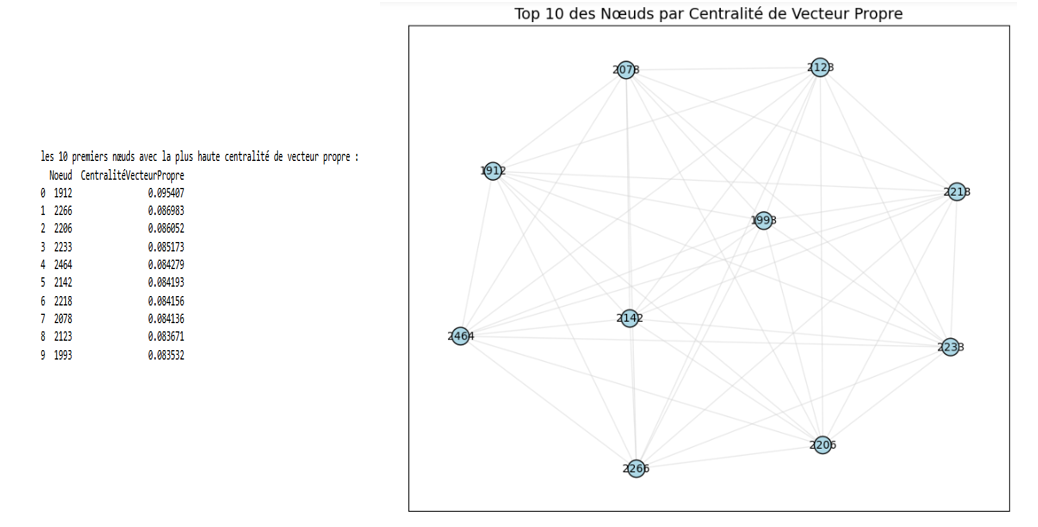
*Analyse de la distribution*

La distribution montre que les nœuds sont à des valeurs à 0.Cette mesure permet d’identifier les nœuds ayant le plus d’influence sur l’ensemble du réseau. 

*Analyse des top10 nœuds de centralité de vecteur propre*

 D’après les valeurs obtenues le nœud 1912 est considéré comme le nœud le plus important de ce réseau en termes d’influence globale sur l’ensemble du réseau. Elle possède également certaines centralité de plus haut degré et des centralités d’intermédiarité ce qui rend l’utilisateur très populaire et influent sur d’autres nœuds

.



**3) Détection de Communauté**

Sur cette partie, l’analyse de communauté a montré l’existence de seize communautés

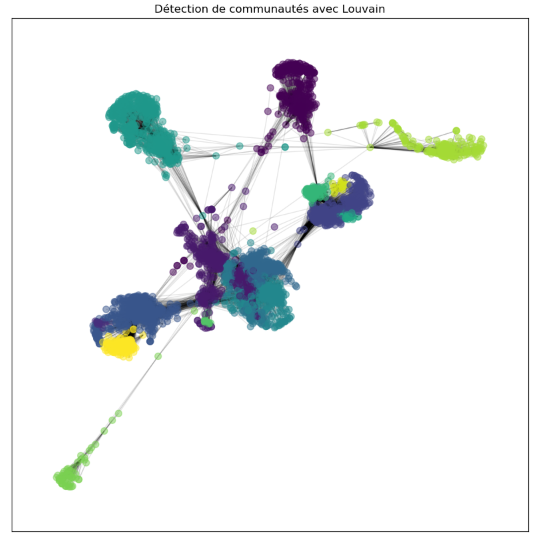


Cette valeur va être utilisée pour pouvoir visualiser ces communautés sur le graphe.

Et pour faire cette visualisation a utilisé l’algorithme de Louvain.

*Visualisation des communautés sur le graphe*

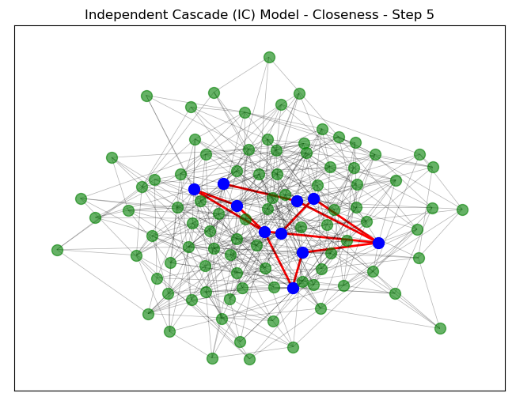
Cette visualisation montre comment est faite la partition des communautés. On peut voir que chaque communauté est représentée en couleur, montrant ainsi les similarités pour chaque groupe que ce soit sur la stratégie de communication ou pour les messages de diffusion.

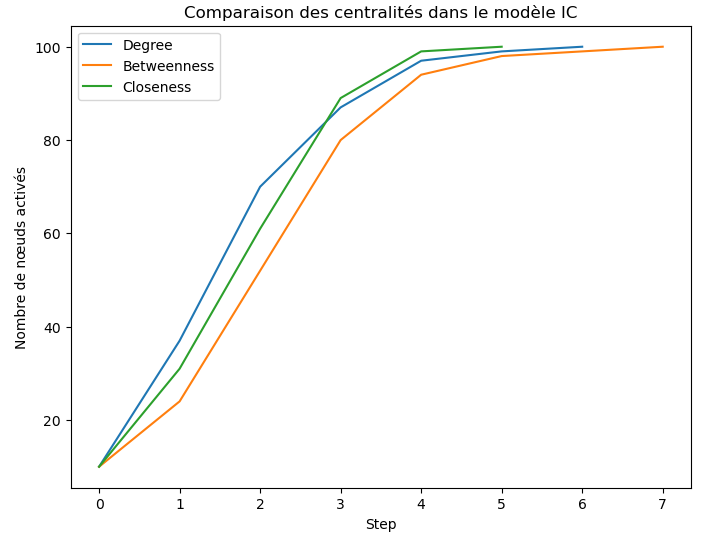


**4) Diffusion d’information**

Dans section on a effectué la simulation de la propagation d’information pour les différents types de centralités. L’algorithme utilisé pour faire cette simulation est : Independent Cascade (IC).

Le graphe (ci-dessous) montre la propagation de l'information à partir des nœuds les plus influents, choisis selon différentes mesures de centralité (degré, intermédiaire, proximité). Au fur et à mesure des étapes, des nœuds sont activés, montrant comment l'information se diffuse à travers le réseau. Les arêtes entre les nœuds influents sont rouges, tandis que celles des autres nœuds sont fines et noires. Cela permet de visualiser comment l'information se propage différemment selon la centralité utilisée.

****



Cette comparaison évolutive permet de mieux comprendre l’évolution de simulation de l’information selon chaque type de centralité. Et cela varie en fonction de l’exécution du code de simulation.

1. Conclusion

L’étude de ce projet a permis de comprendre comment l’analyse des réseaux sociaux est importante. Son étude a permis de savoir les utilisateurs (nœuds) les plus influents sur le réseau mais de connaître la structure de connectivité entre les utilisateurs (nœuds).

REFERENCE

1. [Analyse du réseau Facebook — Carnets NetworkX](https://networkx.org/nx-guides/content/exploratory_notebooks/facebook_notebook.html" \l "network-communities)

2. [Network Analysis with Node Degree Distribution](https://www.youtube.com/watch?v=XMDxpIXh-as&t=439s)

3. [Identification of the most influential nodes... - Abbas Salavaty - NetBio - Poster - ISMB/ECCB 2021](https://www.youtube.com/watch?v=NqMtdiVR2ho)

4. [Working with Graphs in Social Network Analysis.| Degree distibustion,plots using NetworkX librarary](https://www.youtube.com/watch?v=PuGtpst08Qo&t=244s)

5. [read\_edgelist — NetworkX 3.4.2 documentation](https://networkx.org/documentation/stable/reference/readwrite/generated/networkx.readwrite.edgelist.read_edgelist.html)

6. [Graph Algorithm for Social Media Network Analysis | by Bobby Kawade | All About Data | Medium](https://medium.com/socialpages/graph-algorithm-for-social-media-network-analysis-cbba87e28587)