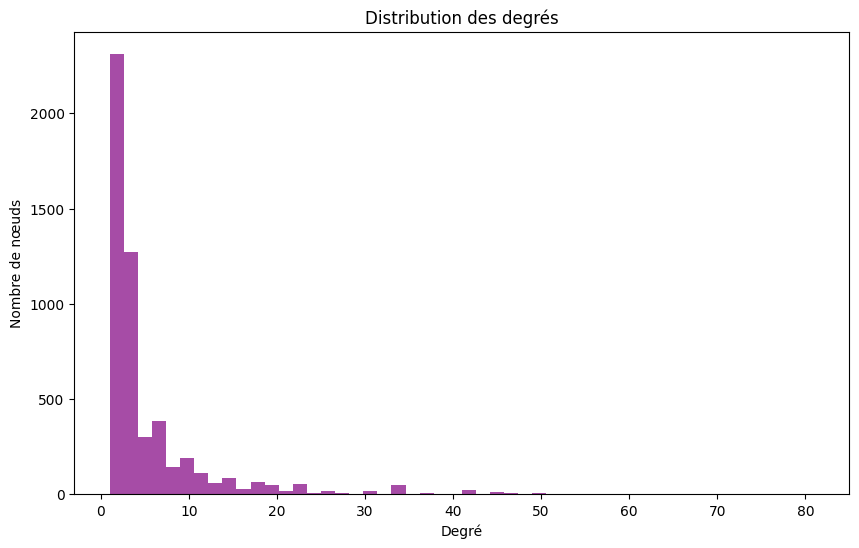
**Analyse des réseaux sociaux**

Élaboré par : Mariem Gatfaoui .

**1- Introduction :**

Pour ce projet, j’ai choisi une source de données en ligne : le réseau de collaborations scientifiques de la catégorie "General Relativity and Quantum Cosmology" de l'arXiv. Dans ce réseau, les nœuds représentent les auteurs d'articles scientifiques et les liens représentent les collaborations entre eux, c’est-à-dire lorsqu'ils ont coécrit un article ensemble. Les informations supplémentaires disponibles sont : pour chaque auteur, son identité (nom ou identifiant), et pour chaque lien, il n'y a pas de poids (toutes les collaborations sont considérées égales). Les données couvrent la période de janvier 1993 à avril 2003 et comportent 5242 auteurs et 14496 collaborations. Nous avons obtenu ces données depuis un dépôt en ligne lié à arXiv. Ensuite, nous avons construit un réseau non orienté, où chaque nœud est un auteur et chaque lien est une collaboration scientifique. Si plusieurs auteurs ont co-écrit le même article, ils sont tous reliés entre eux dans un sous-graph complet.

**2- Analyse de la distribution des données :**



L'histogramme montre que beaucoup de nœuds ont peu de connexions (degré faible), tandis que très peu de nœuds ont beaucoup de connexions (degré élevé). Le degré moyen est faible (environ 5.5), mais le degré maximum est de 81, ce qui confirme qu'il y a quelques nœuds très connectés dans le réseau.

**3- Analyse des composantes connectées :**

Le réseau est fragmenté en 355 groupes distincts de nœuds connectés. La plupart de ces groupes sont petits, car la plus grande composante ne contient que 4158 nœuds. Cela suggère un réseau peu interconnecté dans son ensemble.

4- Analyse des chemins :

Au sein de la plus grande partie connectée du réseau, il faut en moyenne traverser 6 liens pour relier deux nœuds. La distance maximale entre deux nœuds quelconques de cette partie est de 17 liens. Ces valeurs indiquent une relative proximité et efficacité de connexion au sein de cette composante majeure, malgré la fragmentation globale du réseau.

5- Coefficient de clustering et analyse de la densité :

Ces deux métriques fournissent des informations cruciales sur la structure locale et globale de la plus grande composante connectée :

* **Coefficient de clustering moyen : 0.53.** En moyenne, les voisins d'un nœud ont une probabilité de 53% d'être également connectés entre eux. Cela indique une forte tendance à la formation de "triangles" ou de groupes denses au sein du réseau, suggérant une structure locale cohésive.
* **Densité du graphe : 0.0011.** La densité, très faible (environ 0.1%), révèle que seulement une petite fraction des connexions possibles entre les nœuds de la plus grande composante existe réellement. Cela contraste avec le coefficient de clustering élevé, indiquant que si les nœuds ont tendance à former des groupes denses localement, le réseau dans son ensemble est plutôt clairsemé.

6- **Analyse de la centralité:**

Les nœuds les plus importants en termes de connexions directes sont 21012, 21281, 22691, 12365 et 6610. Les nœuds les plus centraux en termes de proximité avec les autres sont 13801, 14485, 9572, 17655 et 2654. On observe une redondance dans les résultats de la centralité de proximité.

**7- Algorithmes de détection de communautés :**

La modularité mesure la qualité de la division d'un réseau en communautés (groupes denses de nœuds avec peu de liens vers d'autres groupes). Un score élevé indique des communautés bien définies.

Les résultats de modularité indiquent une structure communautaire très forte selon l'algorithme de Louvain (0.86), suggérant des groupes de nœuds très bien interconnectés en interne et faiblement connectés entre eux. L'algorithme de Label Propagation donne également une modularité élevée (0.74), confirmant une structure communautaire significative, bien que légèrement moins prononcée qu'avec Louvain. La méthode K-Clique révèle une modularité plus faible (0.47), ce qui peut indiquer qu'elle identifie des communautés basées sur des critères de cliques plus stricts, potentiellement révélant une structure communautaire moins globale ou plus imbriquée. En conclusion, le réseau présente une organisation communautaire robuste, particulièrement bien capturée par les algorithmes de Louvain et de Label Propagat

**8- Explorer et interpréter :**

Avec Louvain, on voit 394 groupes de nœuds. La plupart sont petits (environ 13 nœuds en moyenne), mais un grand groupe a 245 nœuds.

Avec Label Propagation, on trouve beaucoup plus de groupes (1018), mais ils sont très petits (environ 5 nœuds en moyenne, le plus grand en a 153). Le réseau semble divisé en de nombreux petits morceaux.

Avec K-Clique, il y a 835 groupes, plutôt petits en général (environ 6 nœuds). Par contre, un groupe est beaucoup plus grand (952 nœuds). Cette méthode montre beaucoup de petits groupes et un gros groupe important.

En résumé, les trois façons de regarder les groupes dans le réseau donnent des idées un peu différentes. Louvain et K-Clique montrent qu'il y a des groupes assez grands, alors que Label Propagation voit surtout beaucoup de petits groupes.

**9- Visualisation de la plus grande communauté Louvain :**

L'image représente le plus grand groupe identifié par Louvain. On observe un noyau central dense de nœuds fortement connectés. Quelques nœuds périphériques sont moins intégrés à ce groupe principal.

