**Rendu Network Analysis & Modeling**

Lien du repository Github : <https://github.com/Fcrtt/NSGLRendu>

**Question 2 :**

**a)**

**Une image contenant texte, capture d’écran, Tracé

Description générée automatiquementUne image contenant texte, capture d’écran, Tracé, ligne

Description générée automatiquement**Une image contenant texte, capture d’écran, Tracé

Description générée automatiquement

1. Caltech :
   * La distribution de degrés pour Caltech montre une diminution rapide de la fréquence à mesure que le degré augmente.
   * La plupart des nœuds ont un degré compris entre 1 et 50, avec très peu de nœuds ayant des degrés supérieurs à 100.
   * Le degré le plus élevé observé est d'environ 241.
2. MIT :
   * La distribution de degrés pour le MIT montre également une diminution rapide de la fréquence à mesure que le degré augmente, mais la diminution est moins abrupte par rapport à Caltech.
   * La plupart des nœuds ont un degré compris entre 1 et 100, avec un nombre significatif de nœuds ayant des degrés allant jusqu'à 200.
   * Le degré le plus élevé observé est d'environ 683.
3. Johns Hopkins :
   * La distribution de degrés pour Johns Hopkins est similaire à celle du MIT, mais avec une diminution légèrement plus abrupte.
   * La plupart des nœuds ont un degré compris entre 1 et 150, avec un nombre significatif de nœuds ayant des degrés allant jusqu'à 250.
   * Le degré le plus élevé observé est d'environ 868.

Conclusions

1. Caractéristiques des réseaux :
   * Les trois réseaux présentent quelques nœuds (hubs) de degré très élevé, tandis que la plupart des nœuds ont un faible degré. Cela est évident à partir de la longue queue dans les distributions de degrés.
   * La présence de hubs (nœuds avec des degrés très élevés) est plus prononcée au MIT et à Johns Hopkins par rapport à Caltech.
2. Taille et Connectivité du Réseau :

* Le MIT et Johns Hopkins ont des réseaux plus grands (en termes de nombre de nœuds dans la LCC) par rapport à Caltech. Cela se reflète dans les degrés maximaux plus élevés observés au MIT et à Johns Hopkins.
* Les réseaux plus grands (MIT et Johns Hopkins) montrent une gamme plus large de degrés, indiquant des schémas de connectivité plus variés.

1. Densité du Réseau :

Le réseau de Caltech semble être moins dense par rapport à ceux du MIT et de Johns Hopkins, comme l'indique la diminution plus abrupte de la distribution de degrés. Cela suggère que les nœuds du réseau de Caltech ont en moyenne moins de connexions.

**b)**

****

Caltech :

* Coefficient de Clustering Global : 0.2913
* Coefficient de Clustering Local Moyen : 0.4093
* Densité des Arêtes : 0.0564

MIT :

* Coefficient de Clustering Global : 0.1803
* Coefficient de Clustering Local Moyen : 0.2712
* Densité des Arêtes : 0.0121

Johns Hopkins :

* Coefficient de Clustering Global : 0.1932
* Coefficient de Clustering Local Moyen : 0.2684
* Densité des Arêtes : 0.0139

Analyse

1. Densité des Arêtes :
   * Caltech : La densité des arêtes est de 0.0564, ce qui est relativement plus élevé par rapport aux autres réseaux.
   * MIT : La densité des arêtes est de 0.0121, indiquant un réseau plus sparse.
   * Johns Hopkins : La densité des arêtes est de 0.0139, également indiquant un réseau sparse.

Conclusion : Les réseaux de MIT et Johns Hopkins peuvent être considérés comme sparse en raison de leur faible densité des arêtes. Caltech, bien que plus dense, reste relativement sparse avec une densité des arêtes de 0.0564.

1. Coefficients de Clustering :
   * Caltech : Le coefficient de clustering global (0.2913) et le coefficient de clustering local moyen (0.4093) sont les plus élevés parmi les trois réseaux. Cela indique une forte tendance à la formation de triangles (groupes d'amis) dans le réseau.
   * MIT : Les coefficients de clustering global (0.1803) et local moyen (0.2712) sont plus faibles, indiquant une tendance modérée à la formation de triangles.
   * Johns Hopkins : Les coefficients de clustering global (0.1932) et local moyen (0.2684) sont similaires à ceux de MIT, indiquant également une tendance modérée à la formation de triangles.

Conclusion : Caltech présente une structure de réseau plus fortement clusterisée, ce qui suggère des communautés plus serrées et interconnectées. MIT et Johns Hopkins ont des structures de réseau moins clusterisées, indiquant des communautés moins serrées.

Topologie du Graphe

* Caltech : La densité des arêtes plus élevée et les coefficients de clustering plus élevés suggèrent un réseau avec des communautés plus denses et interconnectées. Cela pourrait indiquer une structure de réseau plus modulaire avec des groupes bien définis.
* MIT et Johns Hopkins : Les densités des arêtes plus faibles et les coefficients de clustering modérés suggèrent des réseaux plus sparse avec des communautés moins denses. Cela pourrait indiquer une structure de réseau plus étendue avec des connexions plus dispersées.

En résumé, Caltech présente une topologie de réseau plus dense et clusterisée, tandis que MIT et Johns Hopkins présentent des topologies de réseau plus sparse et moins clusterisées. Ces différences peuvent refléter des dynamiques sociales et des structures organisationnelles différentes au sein de ces institutions.

**c)**

**Une image contenant texte, capture d’écran

Description générée automatiquementUne image contenant texte, capture d’écran, sapin

Description générée automatiquementUne image contenant texte, capture d’écran

Description générée automatiquement**

Observations et Conclusions Supplémentaires:

Observations à partir des Nuages de Points:

1. Corrélation Négative:
   * Les trois réseaux (Johns Hopkins, MIT et Caltech) montrent une corrélation négative claire entre le degré et le coefficient de clustering local. À mesure que le degré d'un nœud augmente, son coefficient de clustering local tend à diminuer.
2. Nœuds de Haut Degré:
   * Johns Hopkins et MIT: Les deux réseaux ont des nœuds avec des degrés dépassant 400, mais ces nœuds présentent des coefficients de clustering locaux très faibles. Cela suggère la présence de hubs qui connectent différentes parties du réseau sans former de clusters serrés.
   * Caltech: Les nœuds de degré le plus élevé sont en dessous de 250, indiquant un réseau moins centralisé avec moins de hubs de haut degré.
3. Comportement de Clustering:
   * Caltech: Malgré un degré maximal plus bas, Caltech montre un coefficient de clustering local moyen plus élevé, indiquant une connectivité locale plus forte au sein de petits groupes.
   * MIT et Johns Hopkins: Ces réseaux ont des coefficients de clustering locaux moyens plus bas, suggérant des communautés locales moins serrées par rapport à Caltech.

Conclusions Basées sur les Analyses Précédentes:

1. Taille et Connectivité du Réseau:
   * MIT et Johns Hopkins: Ces réseaux sont plus grands en termes de nombre de nœuds et de connexions, comme l'indiquent les degrés maximaux plus élevés et une gamme plus large de degrés. Cela suggère des schémas de connectivité plus variés.
   * Caltech: Le réseau est plus petit et moins centralisé, avec une distribution plus uniforme des degrés et un clustering local plus fort, indiquant des communautés plus serrées et interconnectées.
2. Densité et Clustering:
   * Caltech: Avec une densité d'arêtes plus élevée (0,0564) et des coefficients de clustering global (0,2913) et local moyen (0,4093) plus élevés, le réseau de Caltech est plus dense et plus clusterisé, suggérant des groupes bien définis et interconnectés.
   * MIT et Johns Hopkins: Les deux réseaux ont des densités d'arêtes plus faibles (0,0121 et 0,0139, respectivement) et des coefficients de clustering modérés, indiquant des réseaux plus épars avec des communautés moins serrées.
3. Différences Topologiques:
   * Caltech: La densité plus élevée et les coefficients de clustering plus élevés suggèrent une structure plus modulaire avec des groupes bien définis. Cela pourrait refléter une structure sociale ou organisationnelle plus orientée vers la communauté.
   * MIT et Johns Hopkins: Les densités plus faibles et les coefficients de clustering modérés suggèrent des réseaux plus étendus avec des connexions plus dispersées, reflétant une structure plus diverse et moins orientée vers la communauté.

Résumé:

* Caltech présente une topologie de réseau plus dense et plus clusterisée, indiquant des communautés plus fortes et interconnectées.
* MIT et Johns Hopkins présentent des topologies de réseau plus éparses et moins clusterisées, indiquant des schémas de connectivité plus divers et dispersés.
* Ces différences dans la structure du réseau peuvent refléter des dynamiques sociales et des structures organisationnelles variées au sein de ces institutions.

**Question 3 :**

Les calculs ont été effectués seulement sur les 25 premiers graphes (out of memory error au-delà).

Une image contenant texte, capture d’écran

Description générée automatiquementUne image contenant capture d’écran, texte, Tracé, diagramme

Description générée automatiquement

Une image contenant texte, capture d’écran, espace, astronomie

Description générée automatiquementUne image contenant texte, capture d’écran, Tracé, diagramme

Description générée automatiquement

Une image contenant capture d’écran, texte, espace, astronomie

Description générée automatiquementUne image contenant texte, capture d’écran, Tracé, diagramme

Description générée automatiquement

Une image contenant texte, capture d’écran, espace, astronomie

Description générée automatiquementUne image contenant capture d’écran, texte, Tracé, diagramme

Description générée automatiquement

Une image contenant texte, capture d’écran, espace, astronomie

Description générée automatiquementUne image contenant capture d’écran, texte, Tracé, ligne

Description générée automatiquement

1. Statut Étudiant/Professeur

Les valeurs d'assortativité s'étendent à la fois dans les plages positives et négatives, mais la distribution est légèrement décalée vers les valeurs positives. Cela suggère une légère tendance à l'homophilie, où les individus tendent à être amis avec d'autres du même statut (par exemple, les étudiants de premier cycle avec d'autres étudiants de premier cycle, les professeurs avec d'autres professeurs, etc.). La moyenne légèrement positive indique un mélange assortatif faible mais notable par statut étudiant/faculté.

2. Major

Discussion : Les valeurs d'assortativité sont principalement positives, indiquant une tendance à l'homophilie par spécialité. Les étudiants de la même spécialité tendent à être amis entre eux plus qu'avec des étudiants de spécialités différentes. Cela pourrait être dû à des cours communs, des groupes d'étude et des intérêts académiques partagés.

3. Degree

Les valeurs d'assortativité sont distribuées autour de zéro, avec une distribution bimodale suggérant deux groupes distincts : un avec une assortativité positive et un autre avec une assortativité négative. Cela indique que dans certains réseaux, les sommets de haut degré tendent à se connecter avec d'autres sommets de haut degré, tandis que dans d'autres, les sommets de haut degré se connectent avec des sommets de bas degré. Cela pourrait être dû à la présence d'individus populaires qui se connectent avec de nombreux autres, indépendamment de leur degré.

4. Dorm

Les valeurs d'assortativité sont principalement positives, indiquant une tendance à l'homophilie par dortoir. Les étudiants vivant dans le même dortoir tendent à être amis entre eux plus qu'avec des étudiants de dortoirs différents (vraisembblablement dû à la proximité et aux expériences de vie partagées).

5. Genre

Les valeurs d'assortativité s'étendent à la fois dans les plages positives et négatives, mais la distribution est légèrement décalée vers les valeurs positives. Cela suggère une légère tendance à l'homophilie par genre, où les individus tendent à être amis avec d'autres du même genre. La moyenne légèrement positive indique un mélange assortatif faible mais notable par genre.

**Question 5 :**

**d)**

L'algorithme de propagation de labels (LPA) montre des différences significatives en termes d'exactitude pour les différents types de labels (dorm, major\_index, gender). Voici une analyse des résultats et des explications possibles pour ces différences :

Analyse des Résultats

1. Dorm :
   * Fraction de 10% : MAE = 24.51, Accuracy = 0.776
   * Fraction de 20% : MAE = 41.08, Accuracy = 0.686
   * Fraction de 30% : MAE = 28.92, Accuracy = 0.726
2. Major\_index :
   * Fraction de 10% : MAE = 31.33, Accuracy = 0.25
   * Fraction de 20% : MAE = 34.74, Accuracy = 0.222
   * Fraction de 30% : MAE = 132.75, Accuracy = 0.109
3. Gender :
   * Fraction de 10% : MAE = 0.51, Accuracy = 0.579
   * Fraction de 20% : MAE = 0.56, Accuracy = 0.556
   * Fraction de 30% : MAE = 0.496, Accuracy = 0.591

Explications des Différences d'Exactitude

1. Nature des Labels :
   * Dorm : Les labels de dortoir (dorm) sont probablement plus homogènes et localisés géographiquement. Les étudiants vivant dans le même dortoir ont plus de chances d'être connectés entre eux, ce qui facilite la propagation correcte des labels.
   * Major\_index : Les labels de spécialité (major\_index) sont plus diversifiés et moins localisés. Les étudiants d'une même spécialité peuvent être dispersés dans différents dortoirs et avoir des connexions plus variées, ce qui rend la propagation des labels plus difficile.
   * Gender : Les labels de genre (gender) sont généralement bien répartis dans le réseau, mais ils peuvent être influencés par des biais de genre dans les connexions sociales. Les étudiants tendent à avoir des amis du même genre, ce qui peut aider à la propagation des labels, mais les connexions inter-genres peuvent introduire des erreurs.
2. Homophilie :
   * Dorm : Il y a une forte homophilie basée sur le dortoir. Les étudiants vivant dans le même dortoir sont plus susceptibles d'être amis, ce qui facilite la propagation correcte des labels.
   * Major\_index : Il y a une homophilie modérée basée sur la spécialité. Les étudiants de la même spécialité peuvent être amis, mais ils sont également susceptibles d'avoir des amis d'autres spécialités, ce qui complique la propagation des labels.
   * Gender : Il y a une homophilie modérée basée sur le genre. Les étudiants tendent à avoir des amis du même genre, mais il y a aussi des connexions inter-genres, ce qui peut introduire des erreurs dans la propagation des labels.
3. Complexité des Labels :
   * Dorm : Les labels de dortoir sont probablement moins nombreux et plus distincts, ce qui facilite leur propagation.
   * Major\_index : Les labels de spécialité sont plus nombreux et plus diversifiés, ce qui rend leur propagation plus complexe.
   * Gender : Les labels de genre sont binaires (ou limités à quelques catégories), ce qui facilite leur propagation, mais les connexions inter-genres peuvent introduire des erreurs.