



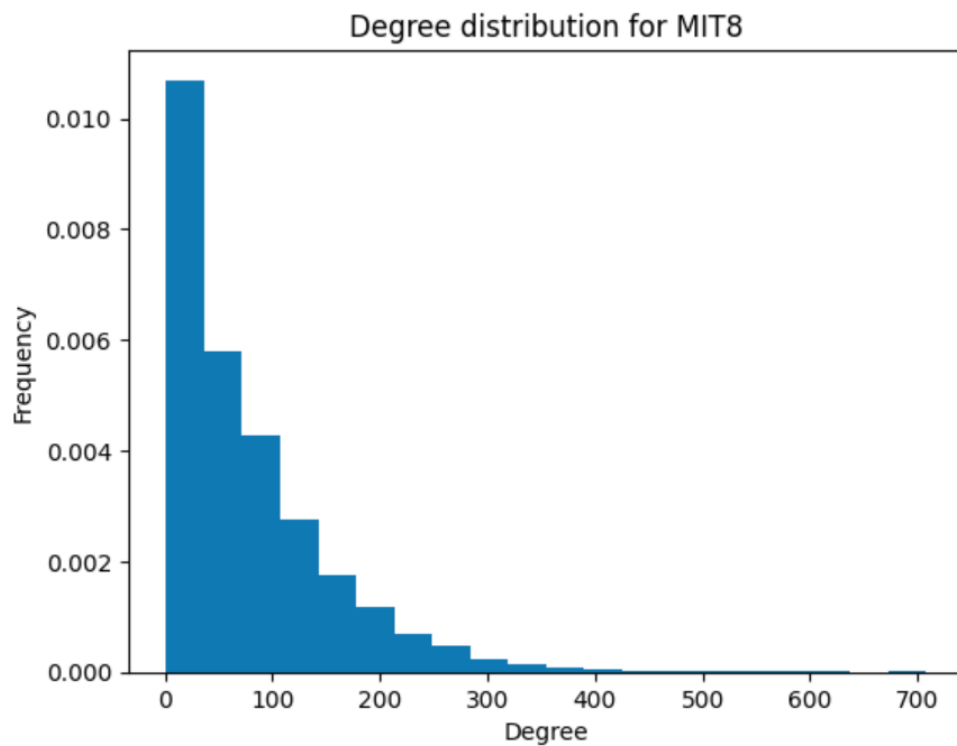
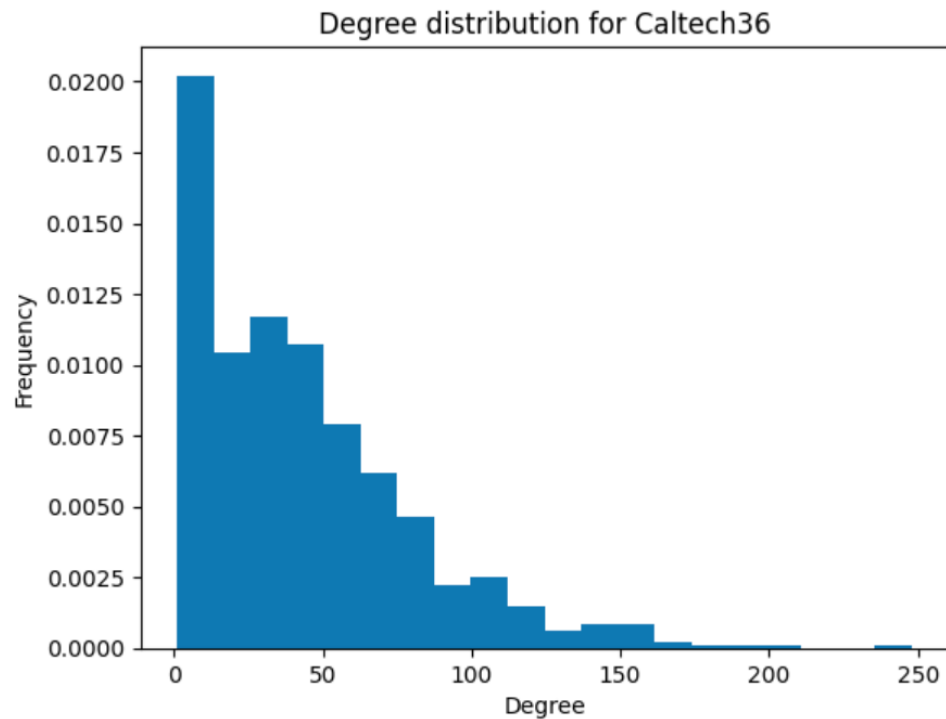
Compte Rendu Net 4103/7431 Réseau complexe

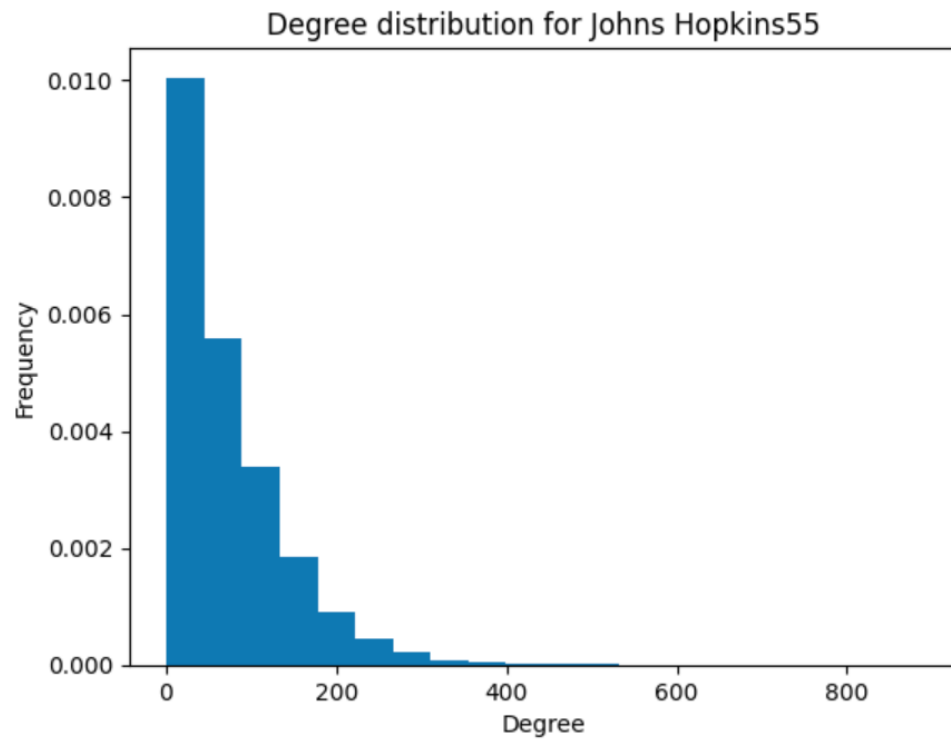
Col Jean-Baptiste

<https://github.com/jbcol/complexnetworksProject>

Question 2: Social Network Analysis with the Facebook100 Dataset

(a) For these three networks plot the degree distribution for each of the three networks that you downloaded. What are you able to conclude from these degree distributions?





On remarque que pour les trois réseaux que plus le degré augmente, plus la densité diminue, ce qui veut dire qu'il y a beaucoup de nœuds avec un degré faible et peu de nœuds avec un degré élevé.

(b) Compute the global clustering coefficient and mean local clustering coefficient for each of the 3 networks. In addition, compute the edge density of each network. Should either of these networks be construed as sparse? Based on the density information and the clustering information what can you say about the graph topology?

Tous les ont une densité de liens très faible (c'est à dire une très petite par rapport à 1) ce qui signifie que les réseaux sont très clairsemés.

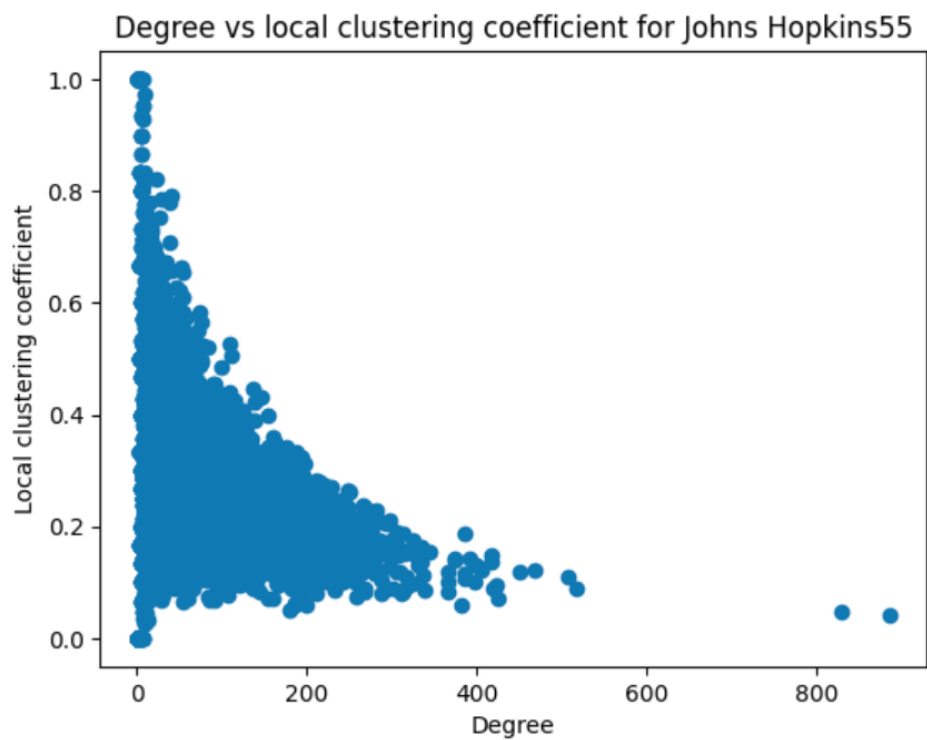
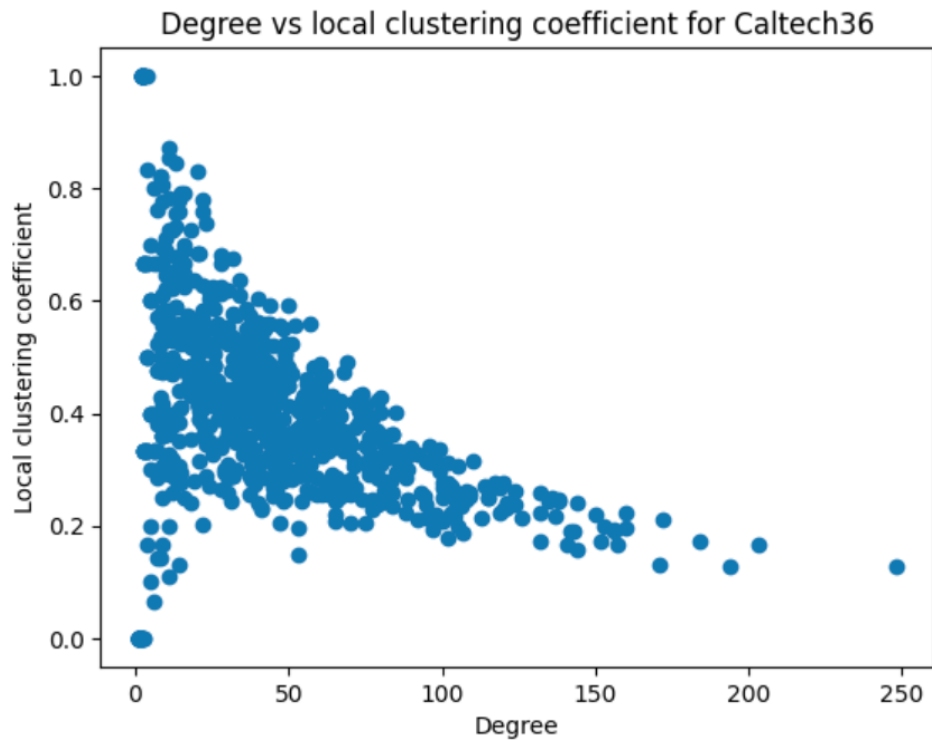
```
- caltech_edge_density = 0.05640442132639792  
- mit_edge_density = 0.012118119495041378  
- johnsHopkins_edge_density = 0.013910200162372396
```

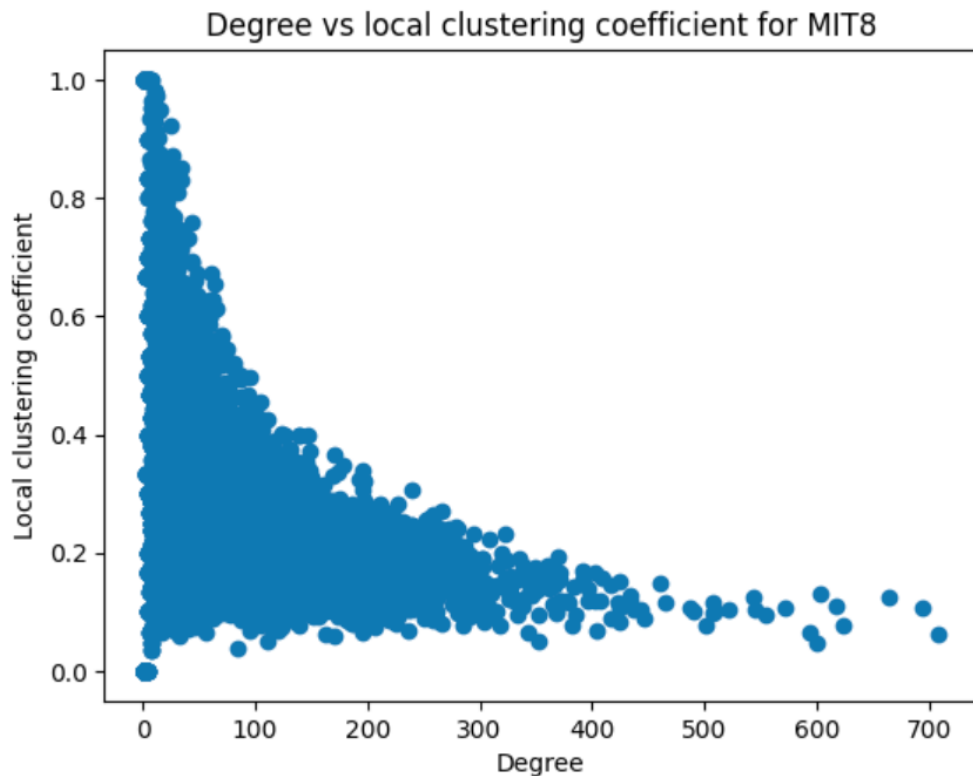
On remarque aussi que le coefficient de clustering est très faible, ce qui signifie que les réseaux sont très peu connectés.

```
- caltech_global_clustering_coefficient = 0.40929439048517247  
- mit_global_clustering_coefficient = 0.2712187419501315  
- johnsHopkins_global_clustering_coefficient = 0.26839307371293525
```

On peut donc conclure que les réseaux ont une topologie très dispersée et peu connectés.

(c) For each network, also draw a scatter plot of the degree versus local clustering coefficient. Based on these calculations as well as your previous ones, are you able to draw any conclusions about any similarities or differences between the tree networks? What other observations can you make?



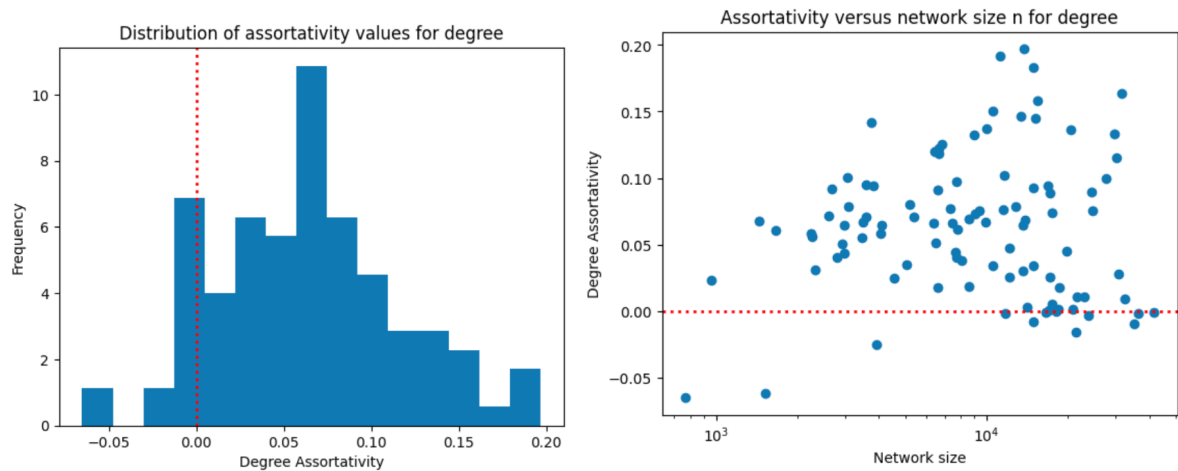


Les réseaux ont tous une allure similaire, c'est à dire que plus le degré augmente, plus le coefficient de clustering converge vers une valeur proche de 0,2 pour caltech et mit et 0,1 pour johns Hopkins. C'est-à-dire que les nœuds de grand degré ont une probabilité moyenne de 0,2 (0,1 pour johns Hopkins) d'avoir des voisins qui sont eux-mêmes connectés entre eux.

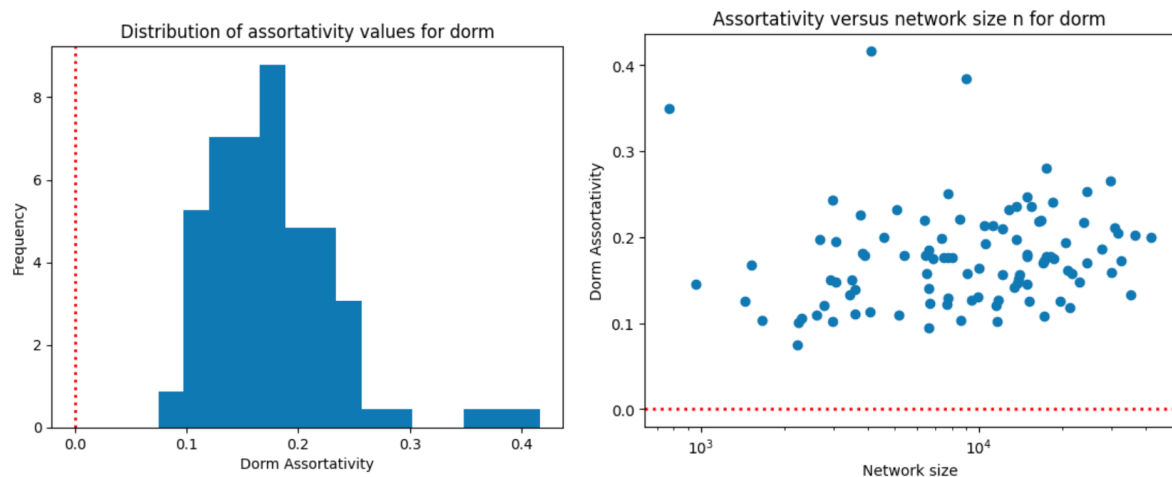
Plus les degrés ont une valeur basse, plus les coefficients de clustering peuvent prendre des valeurs éparées entre 0 et 1. C'est-à-dire que les nœuds de faible degré ont une forte probabilité d'avoir soit beaucoup de voisins qui sont eux-mêmes connectés entre eux soit presque aucun.

Humainement cela se comprend très bien, il y a souvent des petits groupes d'amis qui se connaissent très bien les uns les autres et donc qui sont eux-mêmes connectés entre eux. Il y a aussi parmi ces personnes des gens qui suivent d'autres gens très connus, ce qui fait augmenter le degré des gens connus mais qui ne sont pas eux-mêmes connectés entre eux (ce qui explique pourquoi les nœuds de fort degré ont un coefficient de clustering plus faible).

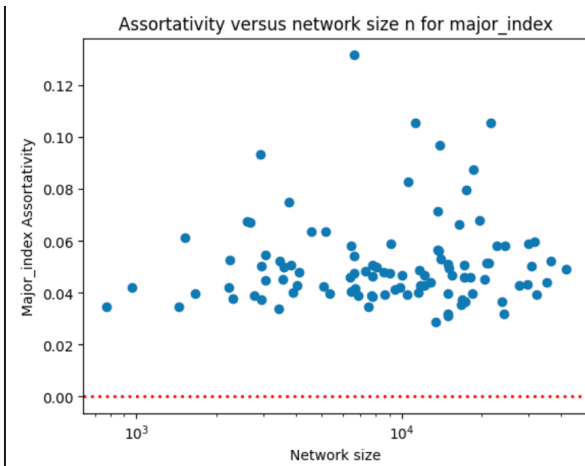
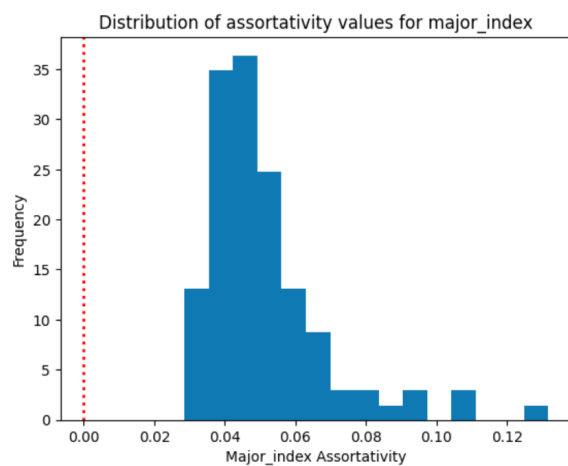
Question 3: Assortativity Analysis with the Facebook100 Dataset



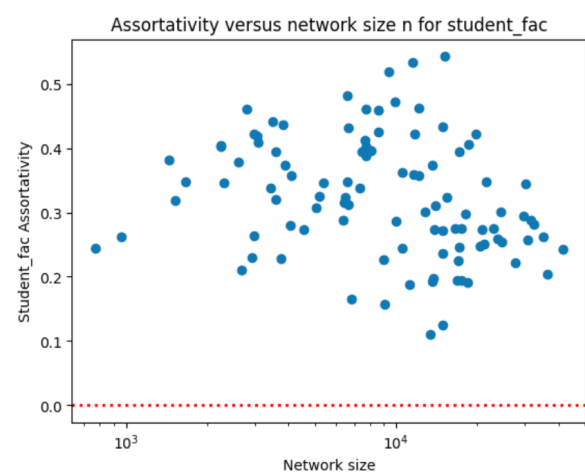
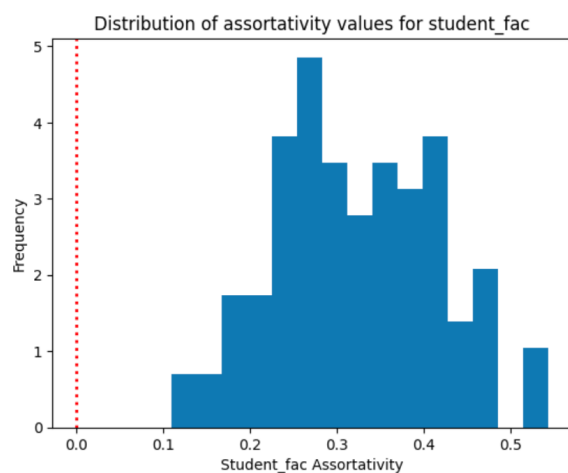
On observe une assortativité moyenne environ égale à 5%, ce qui est relativement faible et ce qui est normal, les gens ne créent pas forcément des connexions avec d'autres personnes en fonction de leur nombre d'amis mais plutôt en fonction de mode de vie similaires.



Pour les dortoirs on observe une assortativité moyenne égale à 20%, ce qui est relativement logique puisque par facteur géographique les personnes de même dortoirs auront tendance à se rencontrer physiquement et donc de créer plus facilement des liens.

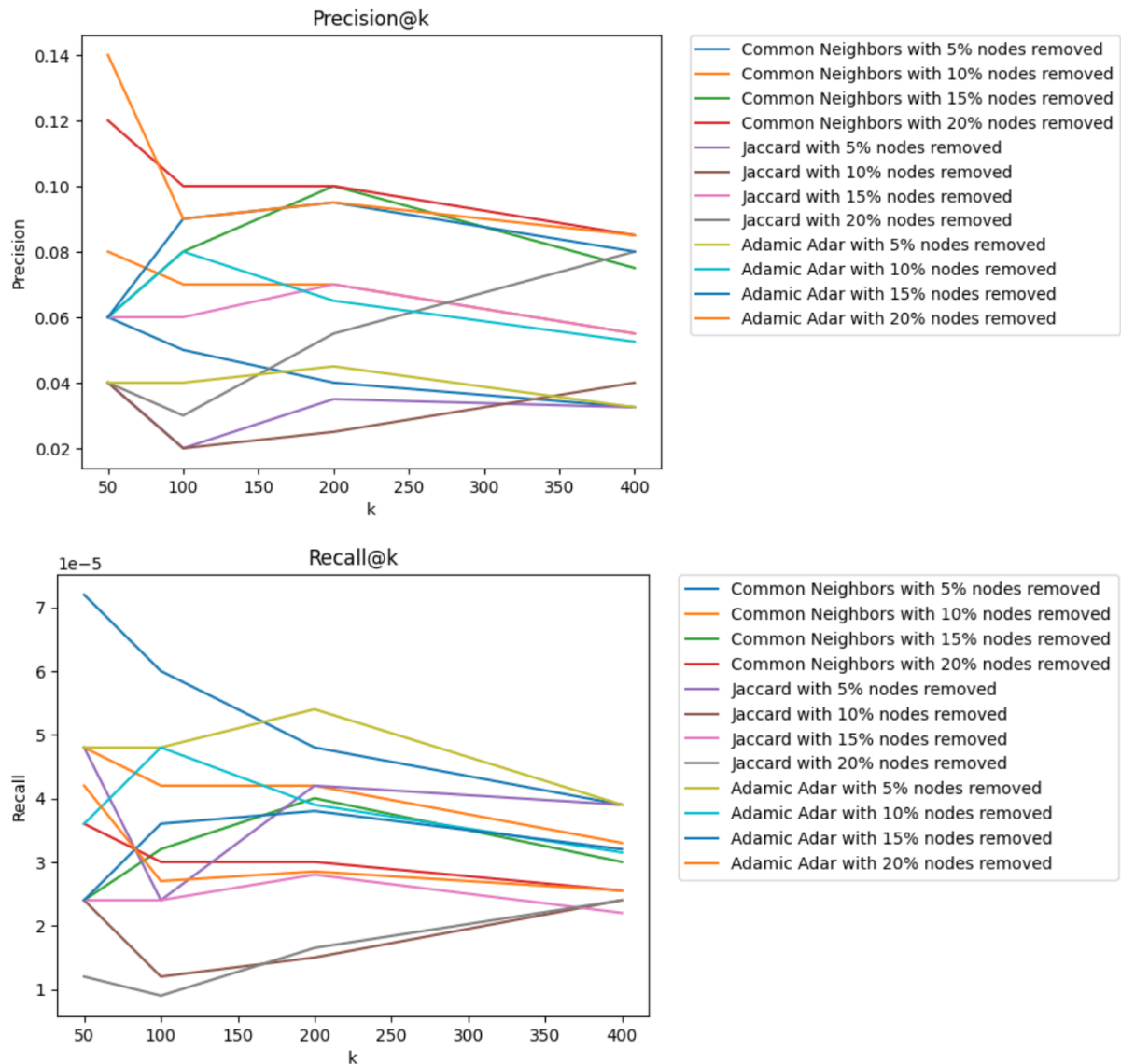


On observe pour les majeurs une assortativité très forte qui est d'environ 50%. Ce qui est normal puisque les personnes d'une même promo sont souvent des amis qui passent beaucoup de temps ensemble et ont des groupes d'amis qui rencontrent d'autres groupes d'amis. Ce qui favorise grandement les liens sociaux.



Pour les mêmes raisons que précédemment, les étudiants au sein d'une même fac ont tendance à régulièrement se côtoyer et créer de nouveaux liens sociaux.

Question 4: Link prediction



On peut tirer une tendance globale de ces graphes. Pour $k \in [50;400]$ on peut remarquer que plus la proportion du nombre de noeuds que l'on retire aléatoirement est importante, plus la précision, le recall et top ont des valeurs importantes (il faut regarder en fonction des couleurs de courbes). Ceci s'intuité facilement puisque la probabilité d'intersection augmente.

Les 3 méthodes de prédiction de liens avec Adamic, Adar, common Neighbors et Jaccard semblent avoir des résultats plutôt similaires. Adamic, Adar prédit légèrement mieux que les autres méthodes.