

Kruse Céline

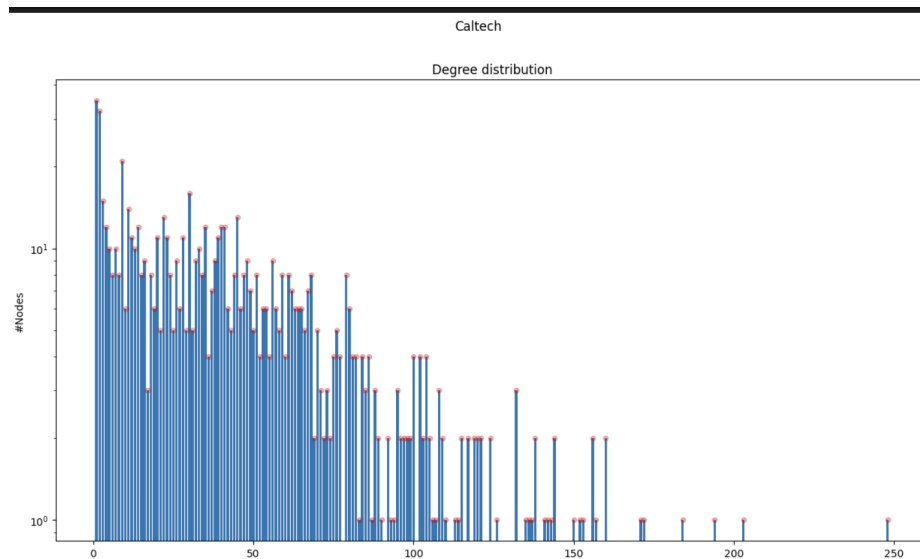
Projet Réseaux Complexes

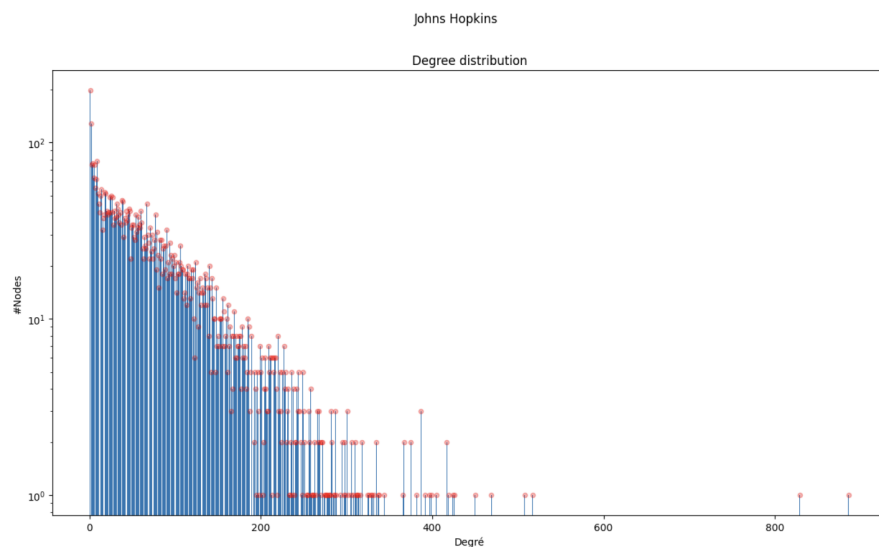
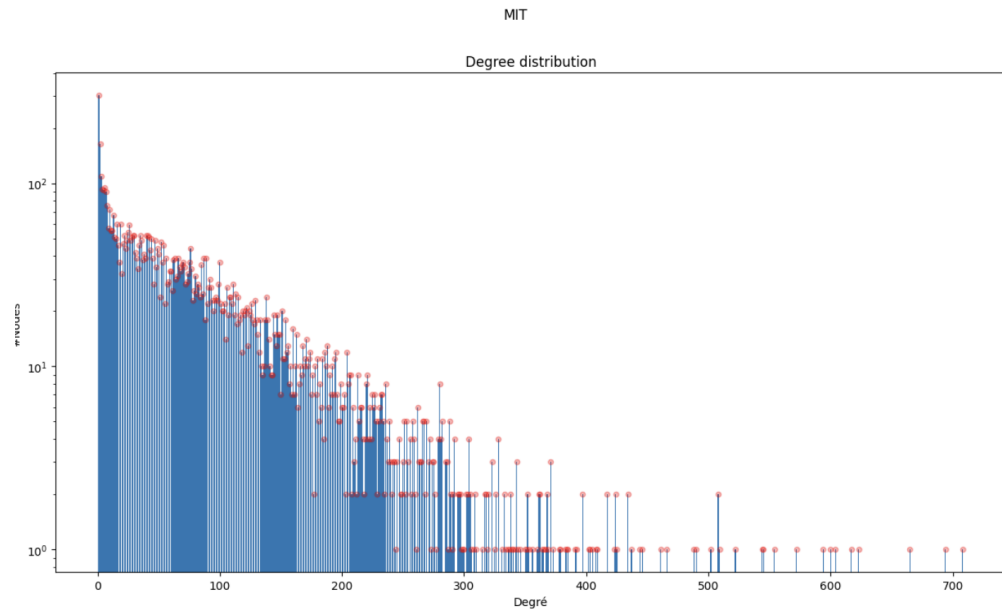
Github : <https://github.com/Ayris2/NET4103-projet>

Question 2

Dans cette question, on se propose d'étudier les réseaux d'étudiants de 3 universités : Caltech, MIT et Johns Hopkins. On étudie notamment les degrés et les clusterings.

En observant les graphes, on peut dire qu'une majorité de personnes n'a pas beaucoup d'amis sur facebook100. On peut les qualifier de sparse.





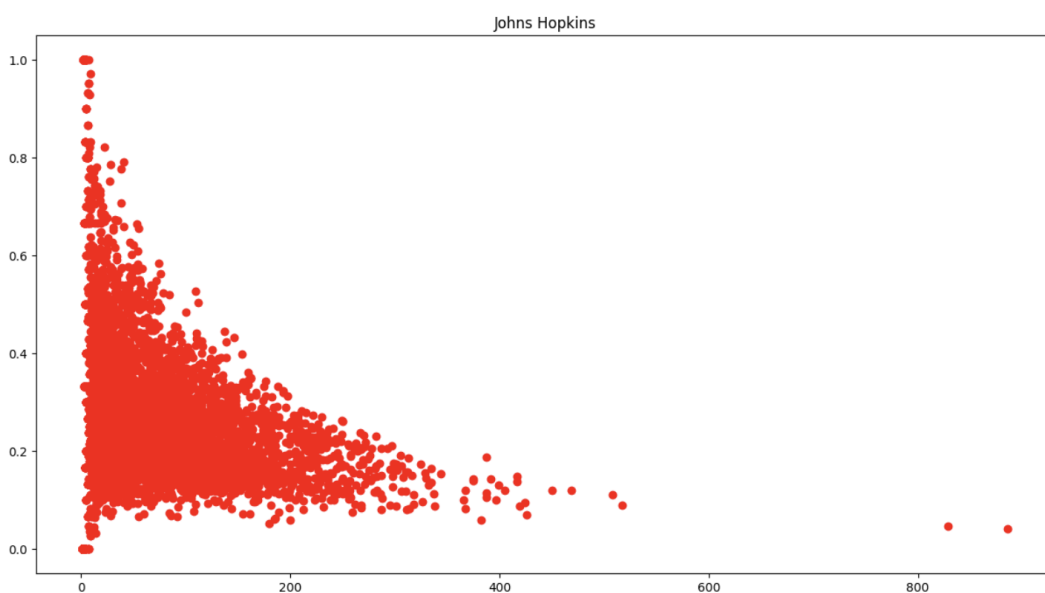
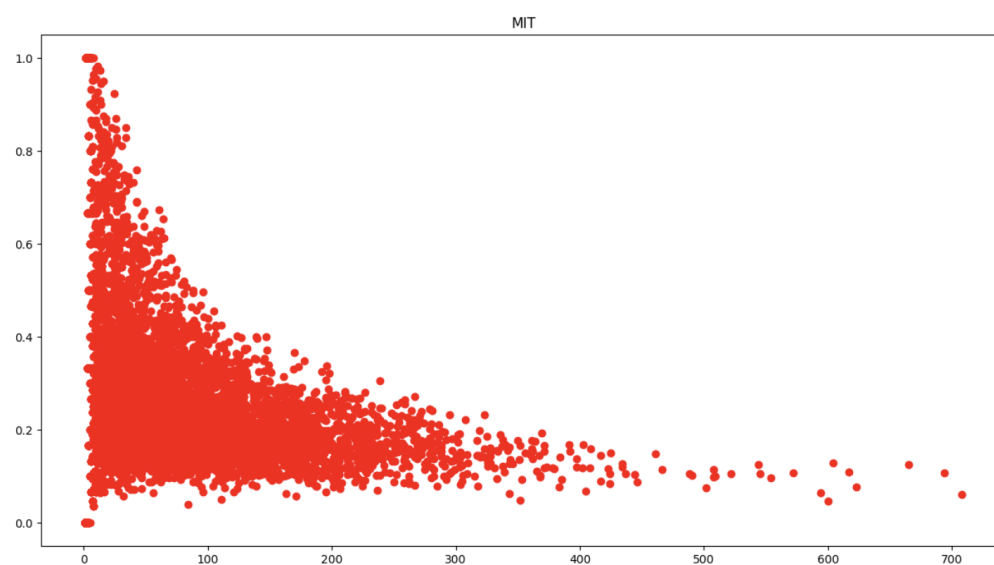
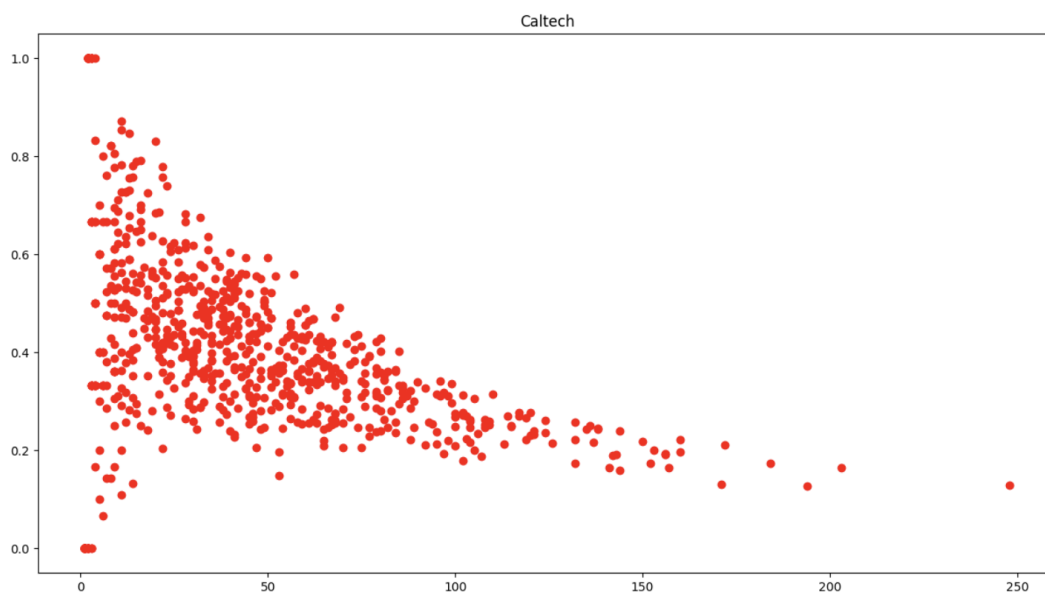
Etudions maintenant le clustering.

Caltech a la plus grande densité ce qui indique que son graphe a plus de connexions que son nombre total de nœuds. De plus, elle a le plus grand degré de Clustering global (0.291) ce qui suggère que les nœuds se regroupent plus en petit groupes.

MIT et Hopkins ont une densité plus faible donc elles ont moins de connexions que de nœuds.

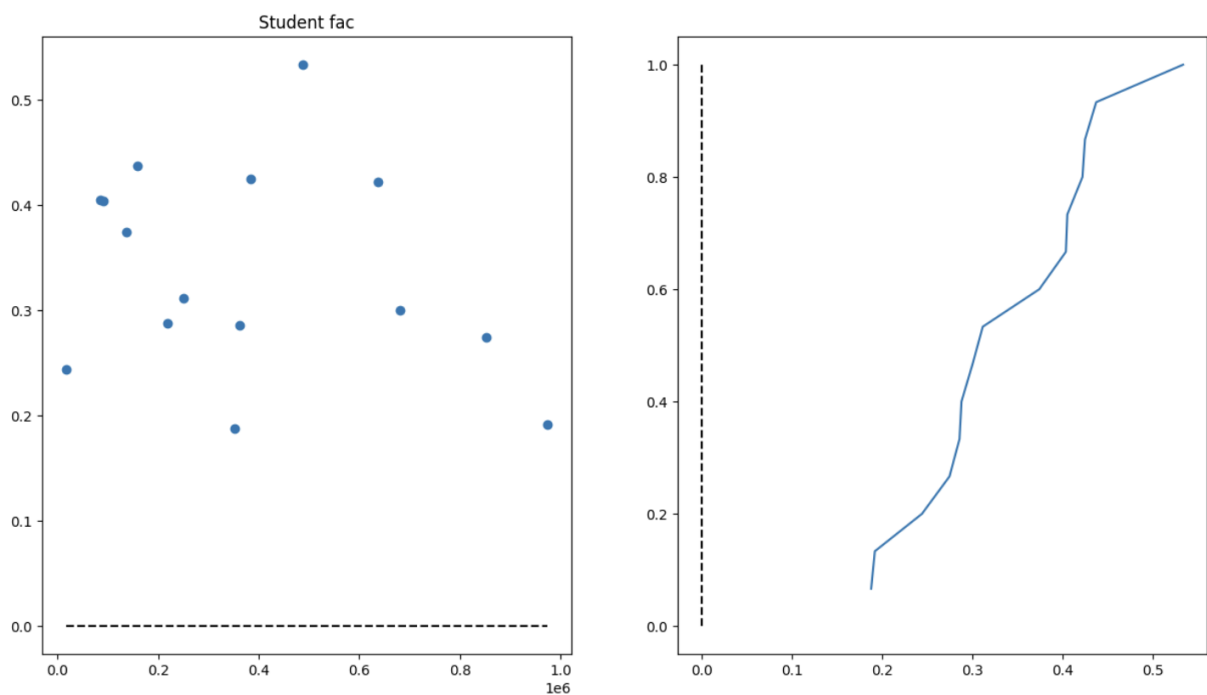
De ce fait, leurs topologies sont différentes. Caltech de son côté semble avoir un réseau plus connecté avec des groupes de nœuds plus resserrés, tandis que ceux du MIT et de Hopkins sont moins denses et les nœuds moins resserrés.

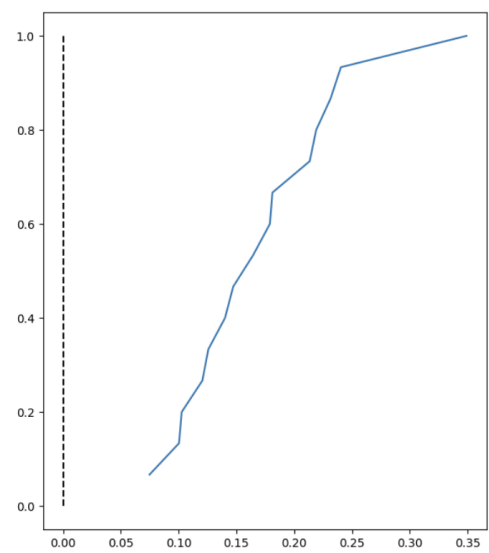
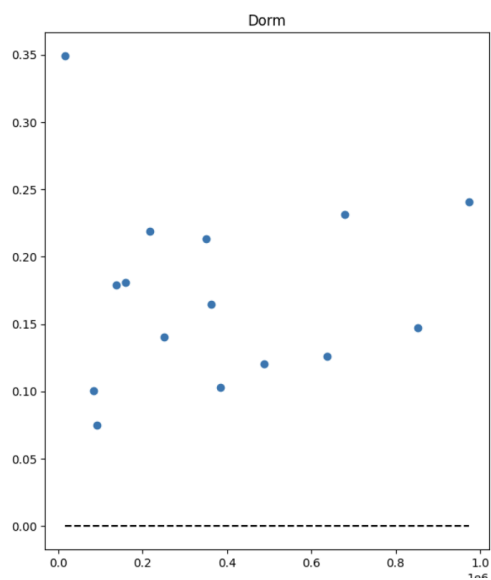
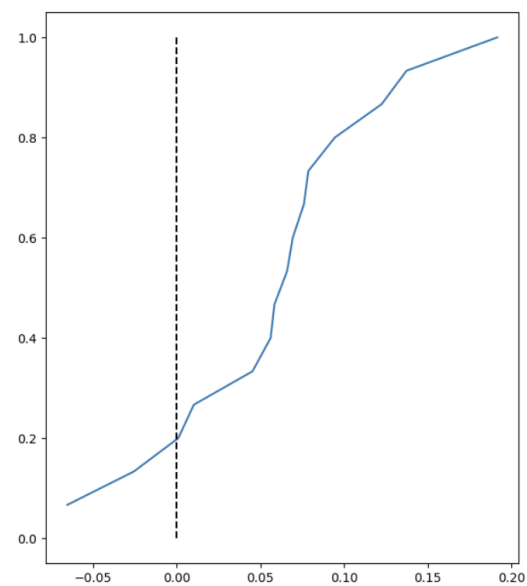
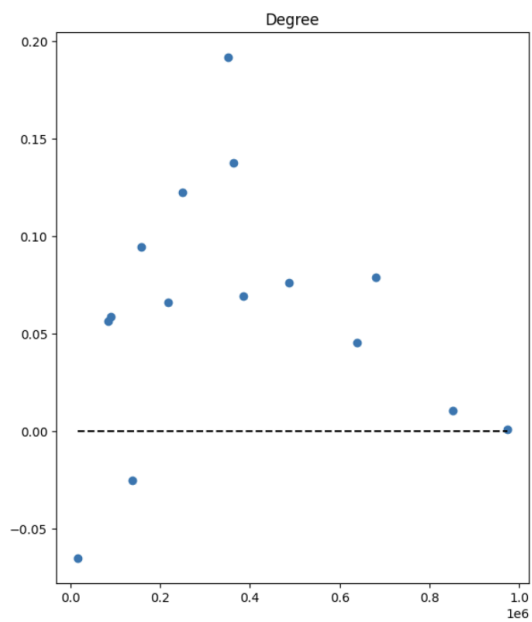
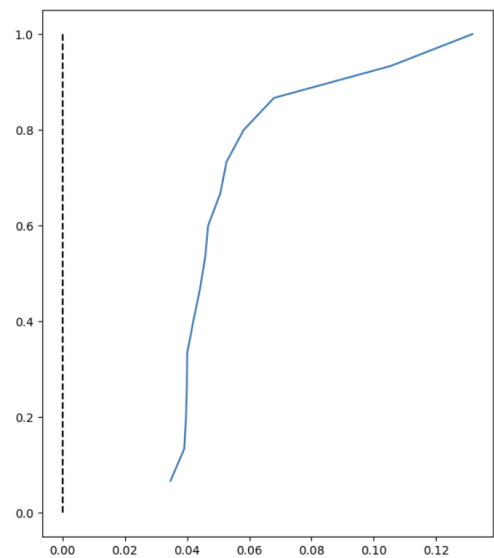
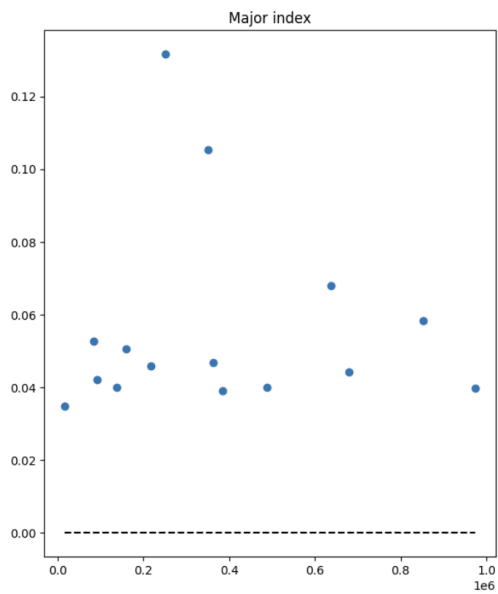
Pour ce qui est du degré par rapport au clustering :

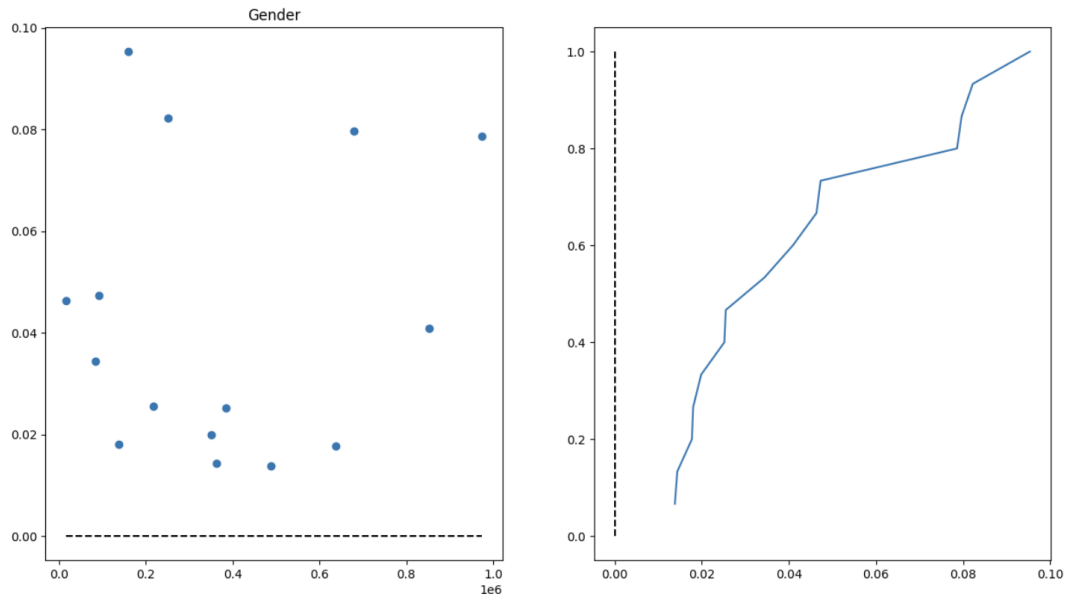


En observant ces graphes, on observe que plus le degré est élevé, moins le coefficient de clustering est élevé. On peut l'expliquer par le fait que plus les nœuds ont de connexions, moins l'apparition de communautés est susceptible.

Question 3







On remarque que les personnes ayant beaucoup d'amis (beaucoup de connexions) ne sont pas forcément reliées entre elles.

De plus, on peut noter que certains attributs sont plus favorables que d'autres pour la formation de liens d'amitié. En effet, faire partie de la même promotion et partager le même dortoir semble être beaucoup plus déterminant que d'avoir la même majeure (peut-être parce que les étudiants se connaissent déjà avant de choisir leur majeure, ou qu'ils préfèrent sympathiser avec des gens ayant fait d'autres choix scolaires). Il semble aussi qu'il y ait plus de liens entre les étudiants du même genre.

Question 4

Dans cette question, on étudie l'efficacité des différentes métriques concernant les prédictions de liens entre les nœuds qui se ressemblent le plus. Les métriques entrant en jeu sont :

- Common neighbors qui regarde les voisins sont communs.
- Jaccard qui regarde la proportion de voisins communs.
- Adamic/Adar qui calcule la somme de l'inverse du log du nombre de voisins communs.

Le graphe choisi est celui de Caltech (pour des raisons de temps d'exécution, je n'ai fait qu'un graphe).

```
evaluate_link_predictor(G1, 20)
```

✓ 1m 5.9s

```
Fraction 0.05 | Nb of Jaccard right predictions : 1
Fraction 0.05 | Nb of Adamic/Adar right predictions : 4
Fraction 0.1 | Nb of Jaccard right predictions : 1
Fraction 0.1 | Nb of Adamic/Adar right predictions : 10
Fraction 0.15 | Nb of Jaccard right predictions : 1
Fraction 0.15 | Nb of Adamic/Adar right predictions : 7
```

On remarque avec ces résultats que la métrique avec le plus de bonnes prédictions est Adamic/Adar donc il semblerait qu'elle soit la plus efficace. Son exécution prend cependant plus de temps car son calcul est plus complexe.

Question 5

Il est question, ici, d'étudier un algorithme de propagation de labels qui consiste alors à prévoir la valeur d'un attribut d'un nœud en fonction de celui de ses voisins. L'algorithme en question (LPA) consiste ici à regarder les fréquences d'apparition des valeurs d'un attribut spécifique et d'attribuer celui apparaissant le plus.

On se focalise ici sur l'université Caltech :

Caltech :

```
dorm | Accuracy : 0.5657894736842105 (10.0% removed labels)
dorm | Accuracy : 0.6078431372549019 (20.0% removed labels)
dorm | Accuracy : 0.5782608695652174 (30.0% removed labels)
major_index | Accuracy : 0.07894736842105263 (10.0% removed labels)
major_index | Accuracy : 0.09803921568627451 (20.0% removed labels)
major_index | Accuracy : 0.11304347826086956 (30.0% removed labels)
year | Accuracy : 0.17105263157894737 (10.0% removed labels)
year | Accuracy : 0.2679738562091503 (20.0% removed labels)
year | Accuracy : 0.19130434782608696 (30.0% removed labels)
gender | Accuracy : 0.5526315789473685 (10.0% removed labels)
gender | Accuracy : 0.6274509803921569 (20.0% removed labels)
gender | Accuracy : 0.6 (30.0% removed labels)
```

On remarque qu'avec un label de 20%, l'accuracy est plus élevée pour les attributs sauf pour major index. Les attributs dorm et gender ont les accuracy les plus élevés. Ceci peut se justifier par le fait que les personnes partageant un même dortoir seront proches plus facilement que si ces personnes sont dans la même majeure. Il en est de même que les personnes du même genre ont tendance à rester entre elles.