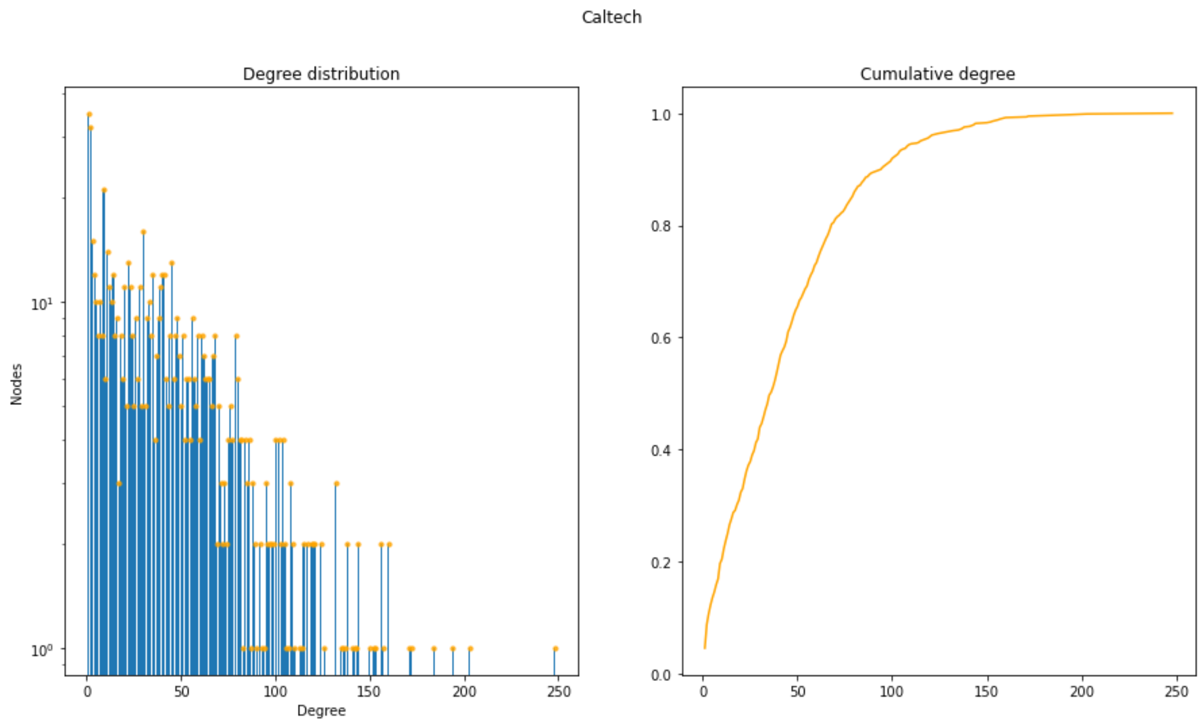
NET4103 - PROJET RÉSEAUX COMPLEXES

SCHNEE LOUISE

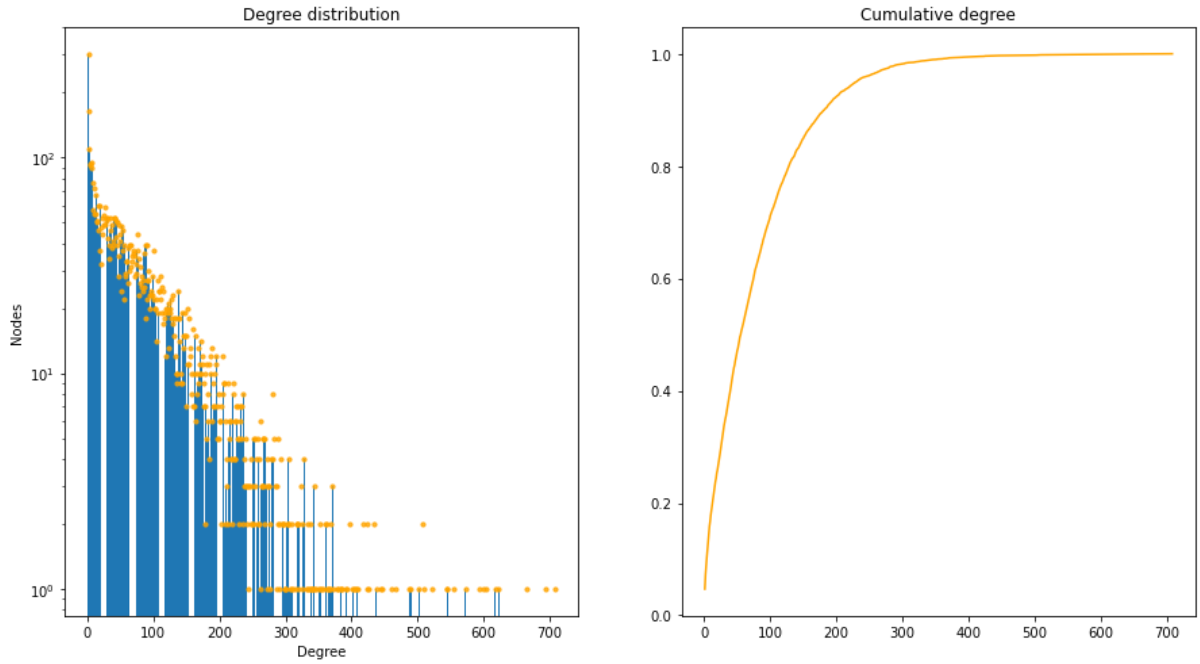
Github repository: https://github.com/Sherlousch/NET4103-fb100-project.git

QUESTION 2: SOCIAL NETWORK ANALYSIS WITH THE FACEBOOK100 DATASET

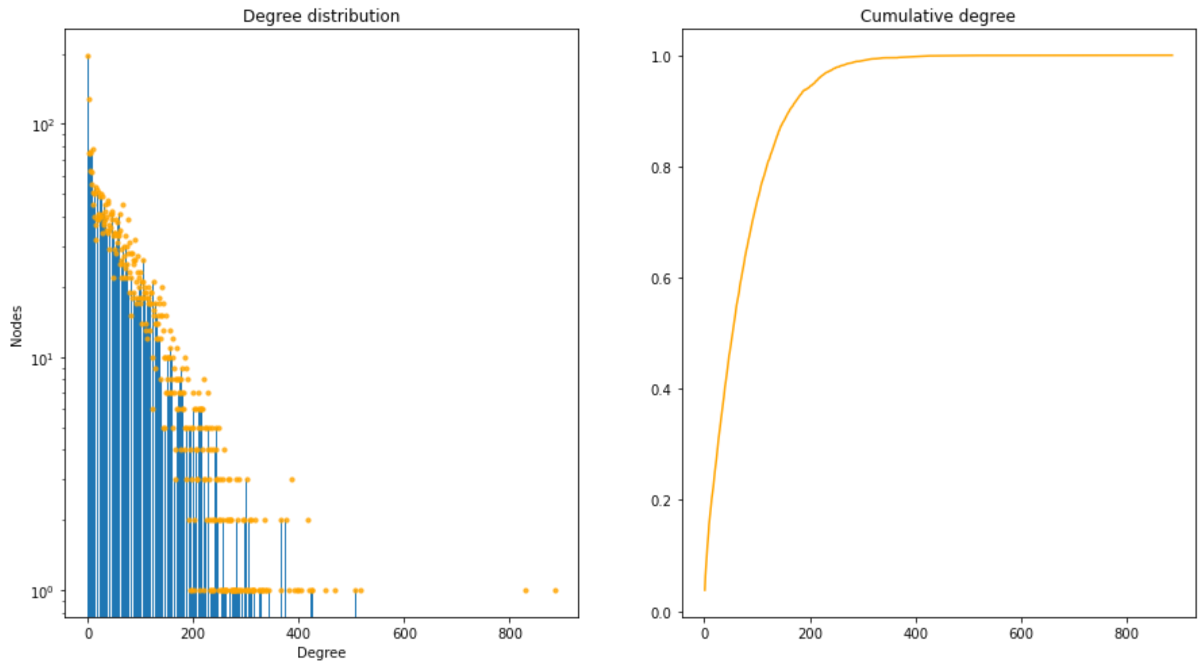
(a)



MIT



Johns Hopkins



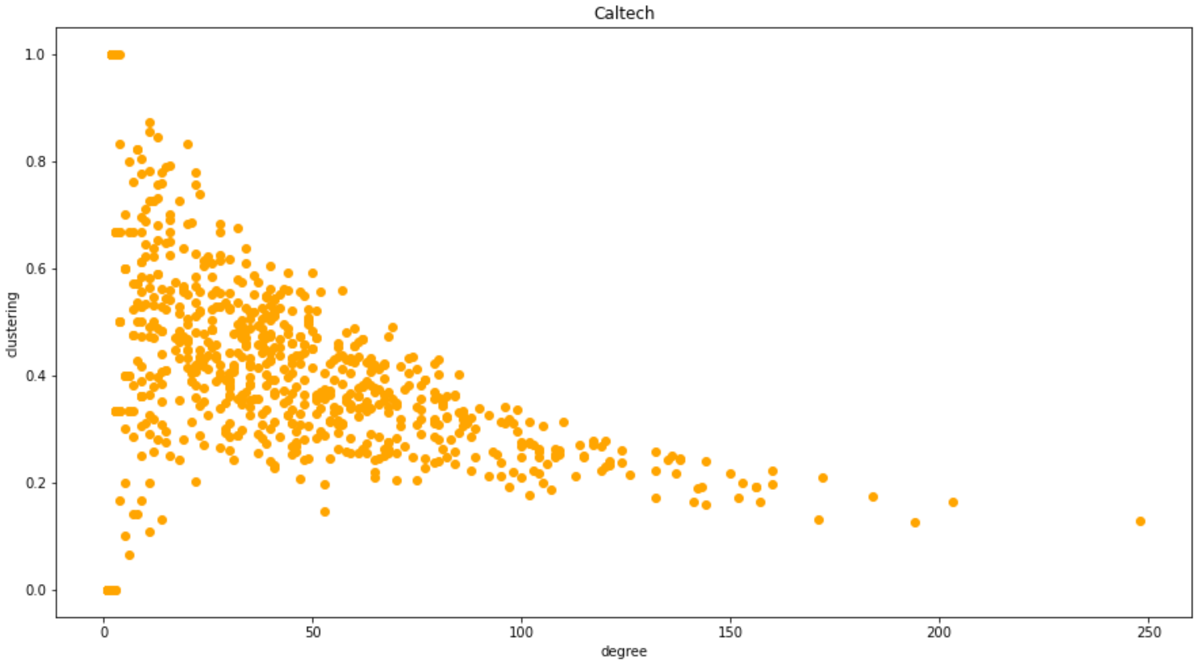
The degree distribution shows that many people have between 0 and 200 friends on Facebook100. There are only a few people with more than 300 friends.

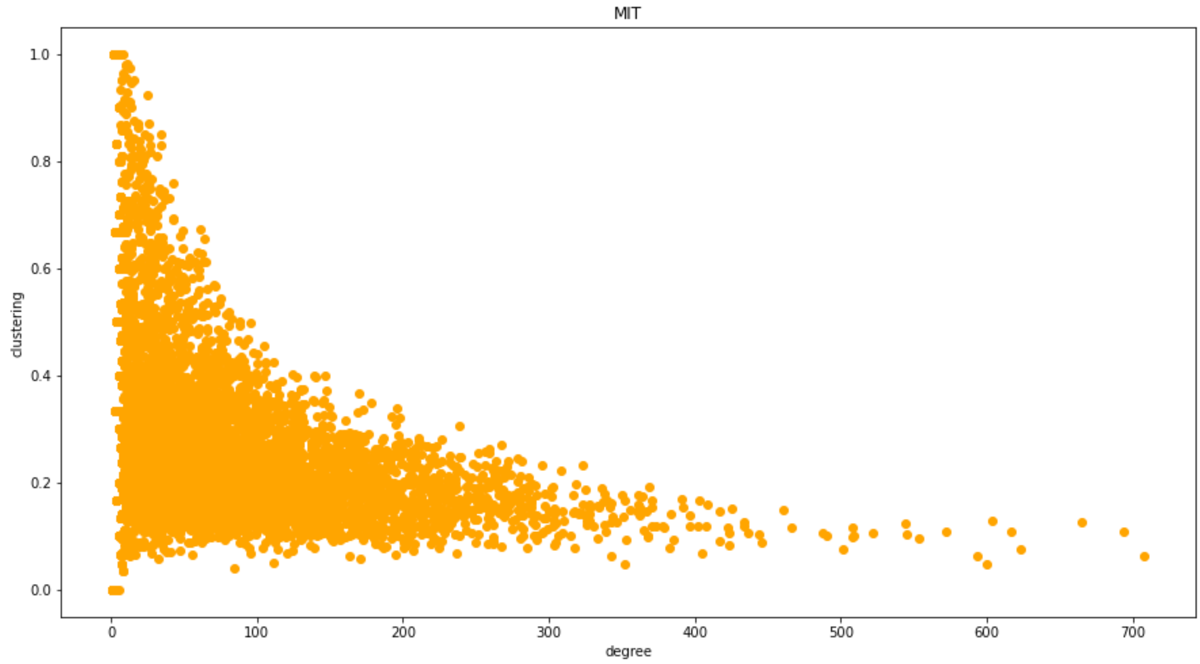
(b)

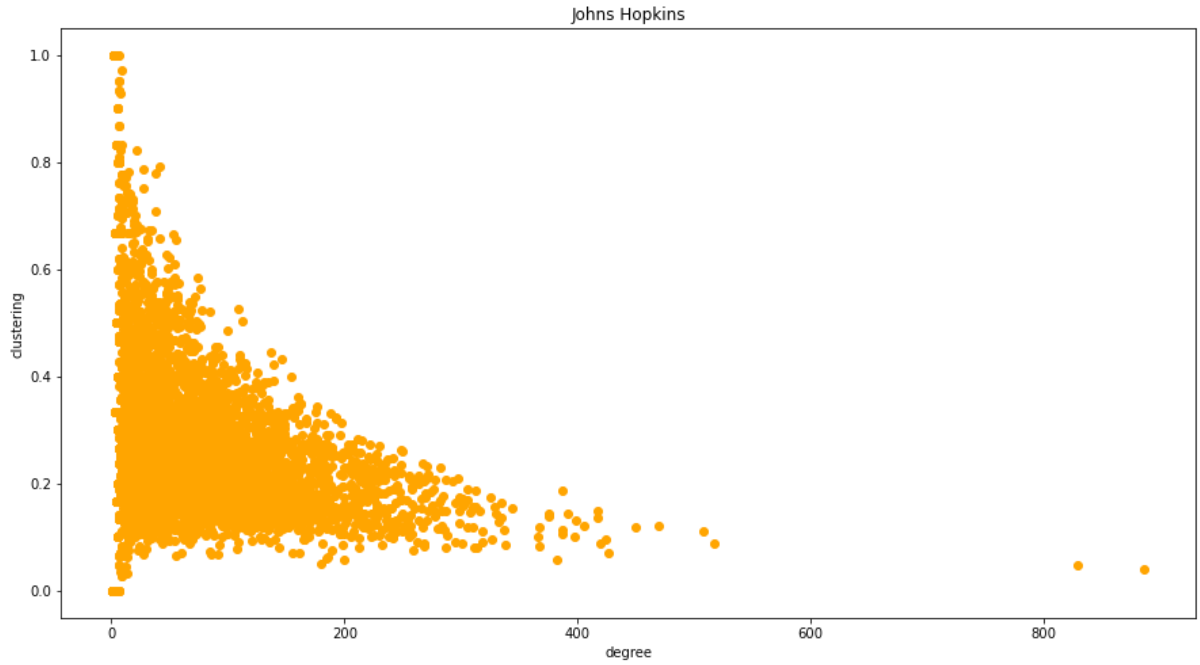
```
-------------------- Caltech --------------------
Global clustering coefficient:      0.2912826901150874
Mean local clustering coefficient:  0.40929439048517247
Edge density:                       0.05640442132639792
-------------------- MIT --------------------
Global clustering coefficient:      0.18028845093502427
Mean local clustering coefficient:  0.2712187419501315
Edge density:                       0.012118119495041378
-------------------- Johns Hopkins --------------------
Global clustering coefficient:      0.19316123901594015
Mean local clustering coefficient:  0.26839307371293525
Edge density:                       0.013910200162372396
```

These 3 networks are sparse.

(c)



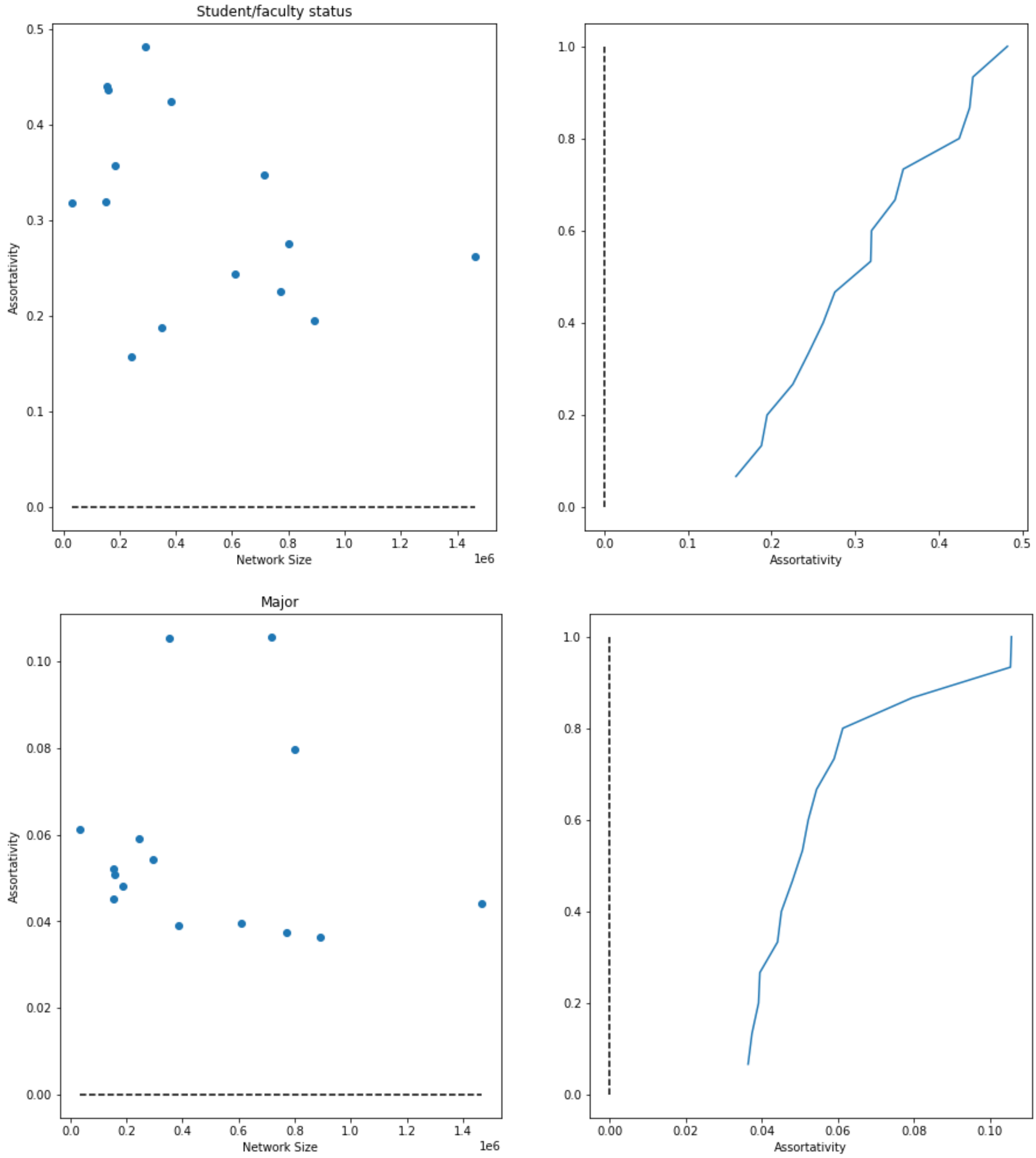


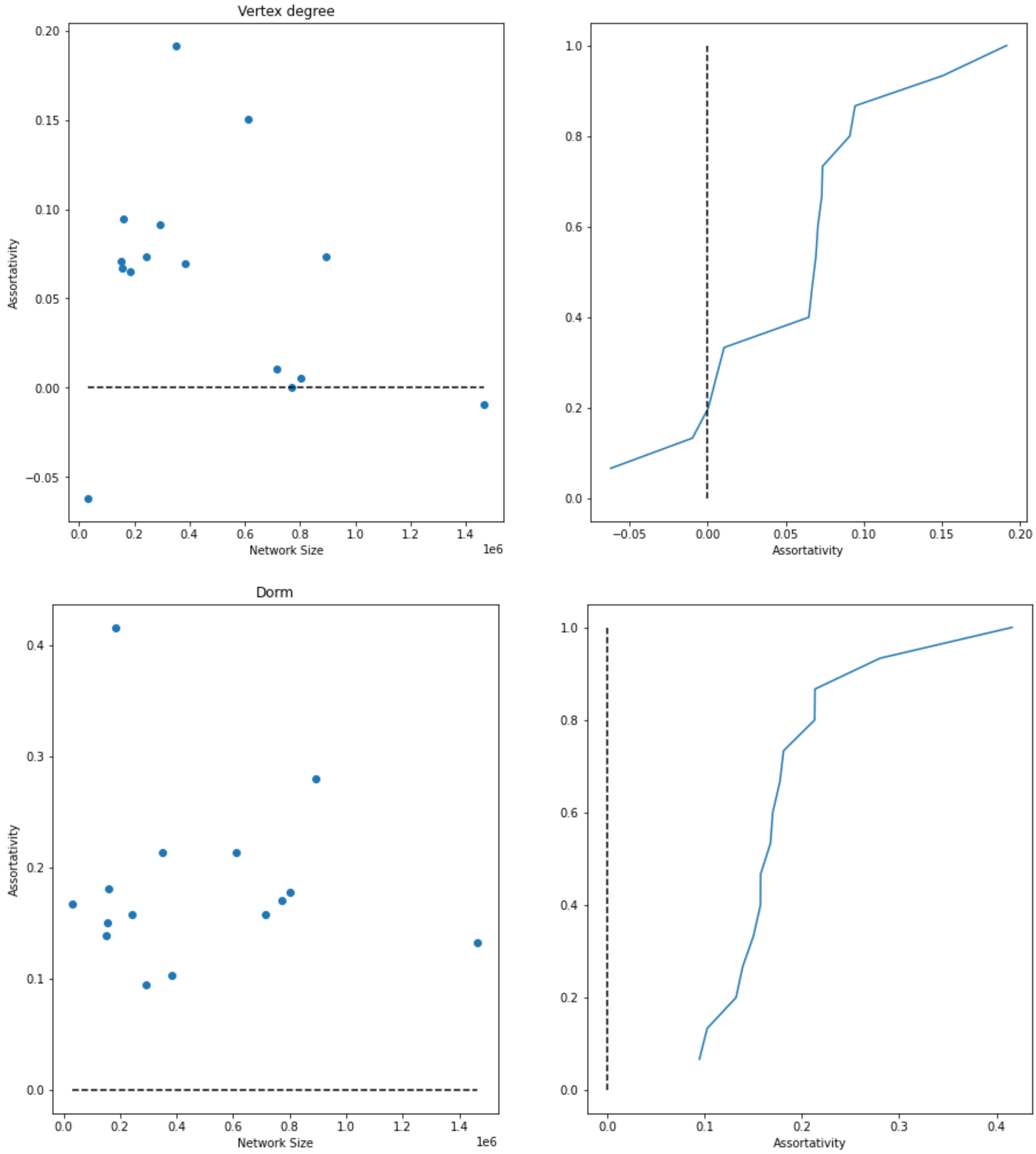


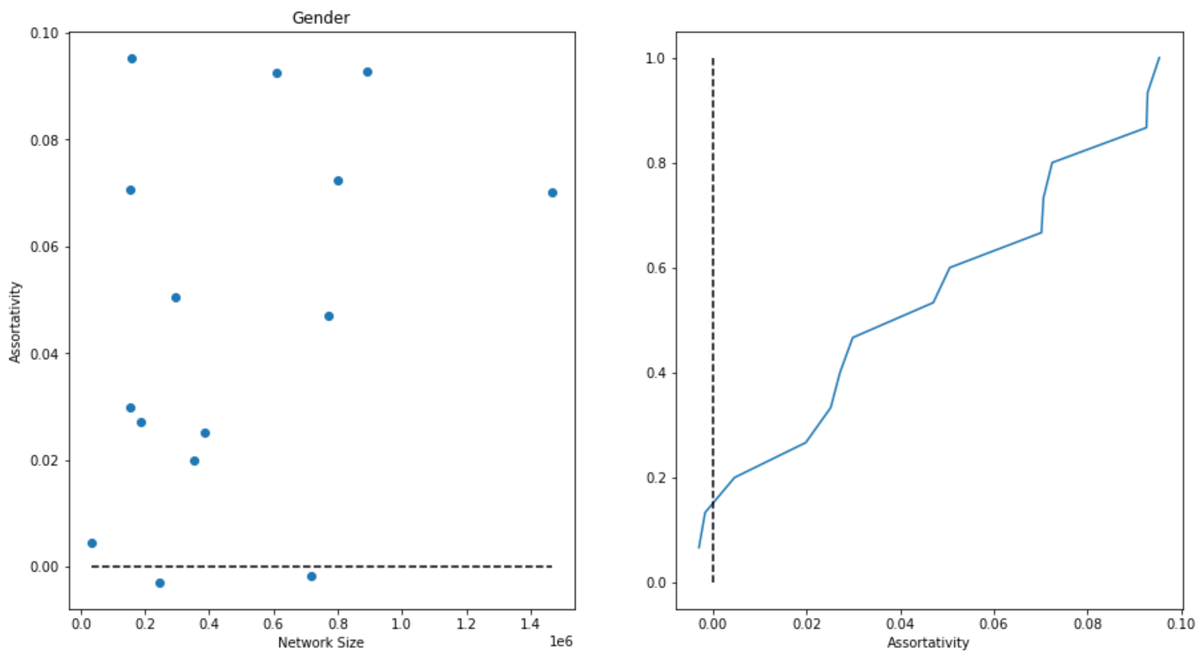
We can observe that for all three networks, when the degree gets higher, the clustering is low. This shows that the users of the networks have a lot of connections, but they don't belong to a large global community.

QUESTION 3: ASSORTATIVITY ANALYSIS WITH THE FACEBOOK100 DATASET

(a)




Vertex degree
Assortativity
Network Size
1e6
Assortativity
Dorm
Assortativity
Network Size
1e6
Assortativity



I examined the data of 15 graphs. Although I couldn't examine all of them (the code was very long to execute), we can observe a certain tendency thanks to the results.
It seems that people who share the same characteristics are easily connected to each other.

QUESTION 4: LINK PREDICTION

Results of the evaluation:

```
--------------- k = 50 ---------------
      Common Neighbors:
Caltech   top 50 = 23 ; precision = 0.48
Reed        top 50 = 22 ; precision = 0.46
Simmons   top 50 = 22 ; precision = 0.42
      Jaccard:
Caltech   top 50 = 22 ; precision = 0.36
Reed        top 50 = 11 ; precision = 0.1
Simmons   top 50 = 18 ; precision = 0.24
      Admic Adar:
Caltech   top 50 = 24 ; precision = 0.48
Reed        top 50 = 22 ; precision = 0.42
Simmons   top 50 = 24 ; precision = 0.46


--------------- k = 100 ---------------
      Common Neighbors:
Caltech   top 100 = 43 ; precision = 0.43
Reed        top 100 = 42 ; precision = 0.4
Simmons   top 100 = 42 ; precision = 0.4
```

```
      Jaccard:
Caltech   top 100 = 28 ; precision = 0.26
Reed        top 100 = 22 ; precision = 0.22
Simmons   top 100 = 19 ; precision = 0.14
      Admic Adar:
Caltech   top 100 = 43 ; precision = 0.41
Reed        top 100 = 35 ; precision = 0.35
Simmons   top 100 = 43 ; precision = 0.4


--------------- k = 200 ---------------
      Common Neighbors:
Caltech   top 200 = 81 ; precision = 0.395
Reed        top 200 = 72 ; precision = 0.37
Simmons   top 200 = 74 ; precision = 0.37
      Jaccard:
Caltech   top 200 = 53 ; precision = 0.255
Reed        top 200 = 45 ; precision = 0.205
Simmons   top 200 = 33 ; precision = 0.16
      Admic Adar:
Caltech   top 200 = 79 ; precision = 0.355
Reed        top 200 = 64 ; precision = 0.31
Simmons   top 200 = 71 ; precision = 0.35


--------------- k = 400 ---------------
      Common Neighbors:
Caltech   top 400 = 154 ; precision = 0.335
Reed        top 400 = 138 ; precision = 0.305
Simmons   top 400 = 134 ; precision = 0.31
      Jaccard:
Caltech   top 400 = 92 ; precision = 0.185
Reed        top 400 = 87 ; precision = 0.2
Simmons   top 400 = 51 ; precision = 0.145
      Admic Adar:
Caltech   top 400 = 139 ; precision = 0.32
Reed        top 400 = 127 ; precision = 0.285
Simmons   top 400 = 137 ; precision = 0.32
```

All 3 metrics have a very low precision, less than 0.5.
If we compare them to each other, the metrics common neighbors and admic adar seem to be pretty efficient for a small portion of edges (very slightly more precise for the common neighbor metric).
But we can notice that the precision decreases when we evaluate them on a larger portion of edges.
The metric jaccard is much less efficient, but it seems that the precision stays the same regardless of the size of the edge set evaluated. It might then be more efficient for a very large portion of edges.