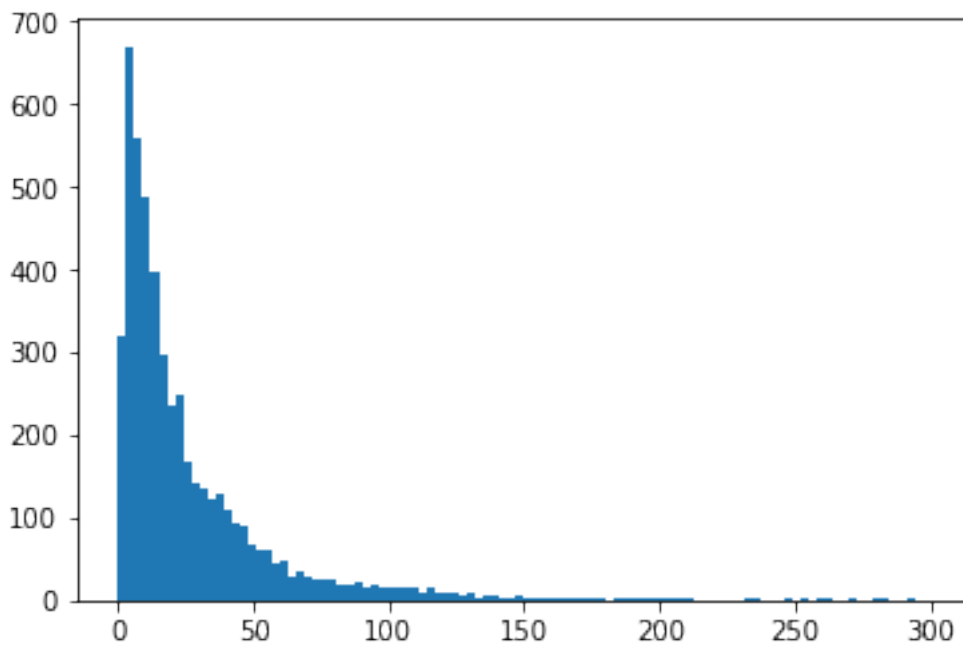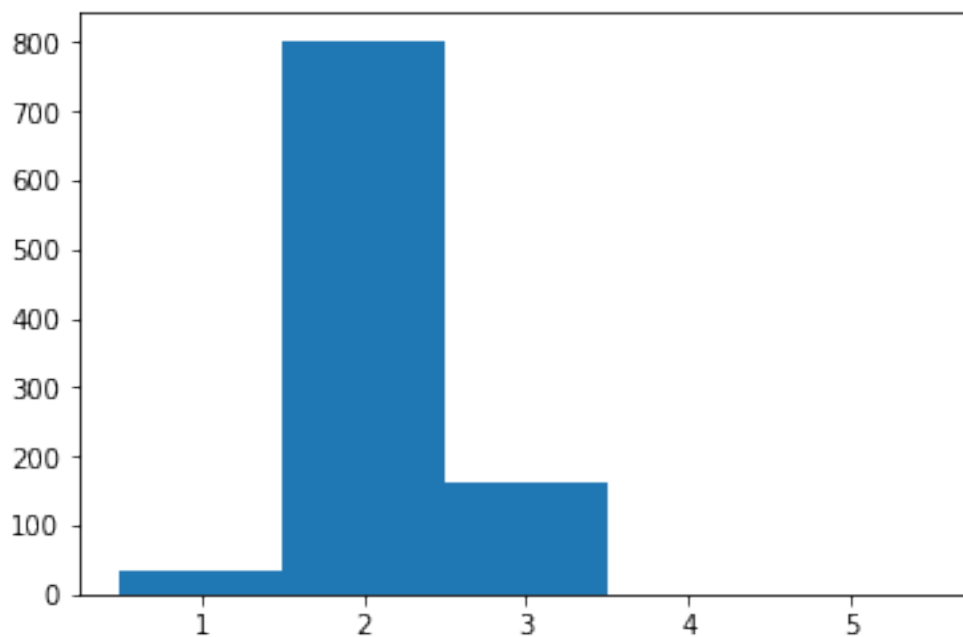1.(a)

(b) YGR296W has clustering coefficient 0.1, YPL098C has clustering coefficient 0.8.
A higher clustering coefficient means most of the neighbors of 'YPL098C' interact with each other(forms an edge)

(c)There are 2127084 3-cliques. The global clustering coefficient is 0.065211

(d)

The shortest path distances is far smaller than I expected
I was thinking the nodes would be quite separate from each other since there are a huge number of nodes
But the experiment result show that there is no two nodes separated by more than 3 edges.

(e) My estimate is 3, since in last question we sampled 1000 pairs of nodes and we don't find a pair with distance more than 3.

2.(b) Instead of majority vote algorithm, I just let each neighbouring node votes in order. If some vote occurs twice, we will take that as our label, otherwise, we will choose the first vote as the label.

(c) I do the leave one out cross validation in 5 groups.
Out of 5001 proteins, majority vote algorithm get 2287 correct while my algorithm get 1467 correct.

(d) Take the largest k result from majority vote instead of only one.
To measure the performance, give each correct answer a partial credit depending on how many labels it have.