Complex Network

Project Presentation

Aryan Jaira

Amazon Co-Purchasing Dataset

- Taken from snap.stanford.edu
- In the form (i,j) where ith item is frequently co- bought with jth item
- VERY big dataset, 262111 nodes and 1234877 edges
- VERY bare-bones dataset, no node-attributes
- NOT strongly connected

From the vertex 0, we take all the nodes that are a distance of less than 7.5 from it.

The new network has 1040 nodes and 4197 edges.

As the graph is not strongly connected, We can’t take its diameter directly, so we find the strongly connected component and find their diameter and then choose the maximum of it. The diameter is 21

Resulting Network

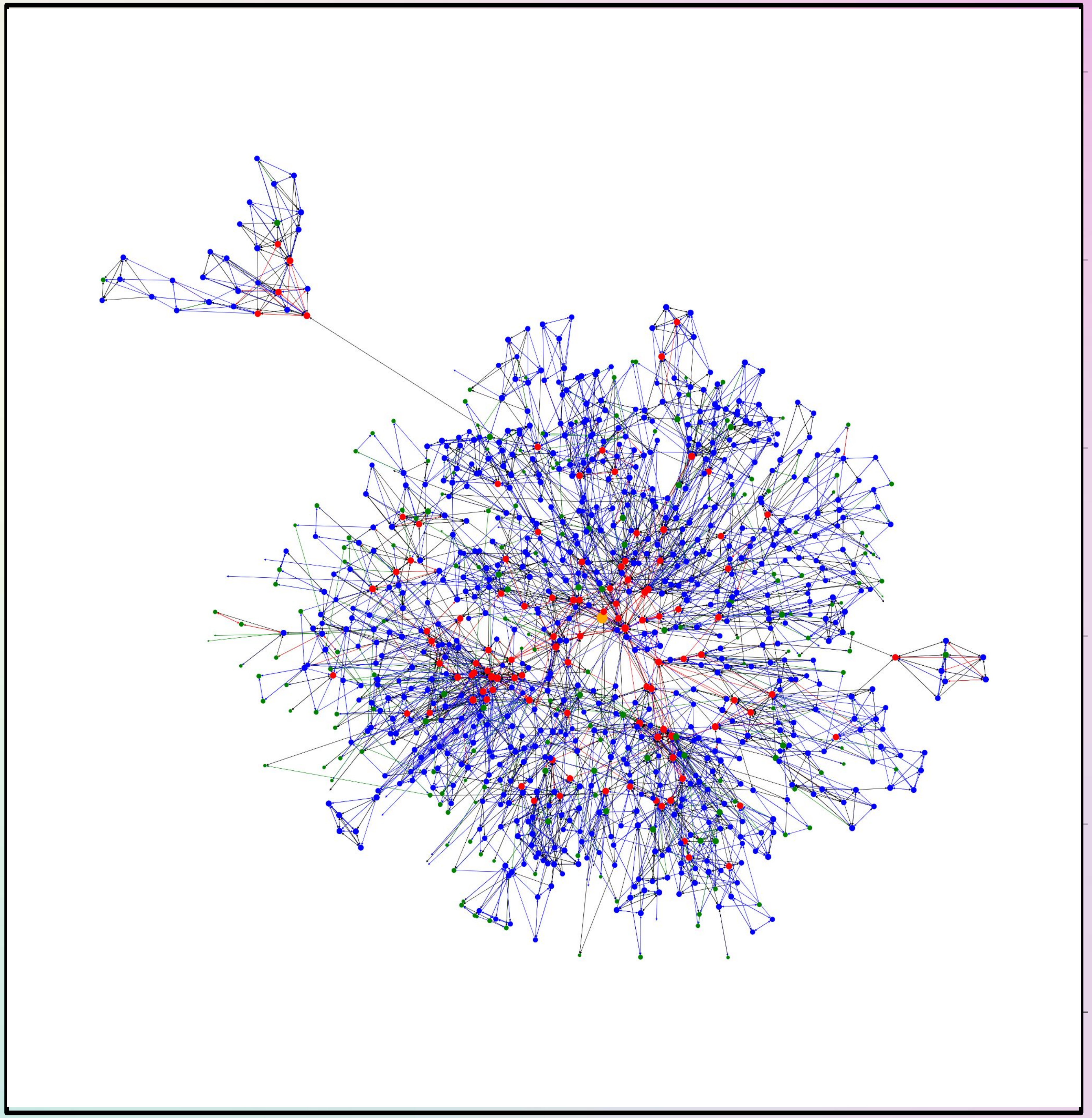
We color vertices with degree more than 12 as Red

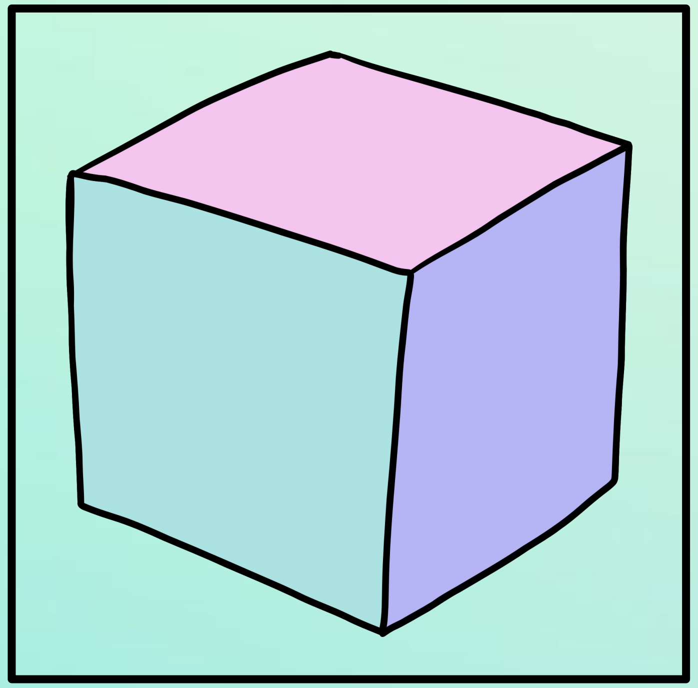
and vertices with degree <12 but >4 as blue and

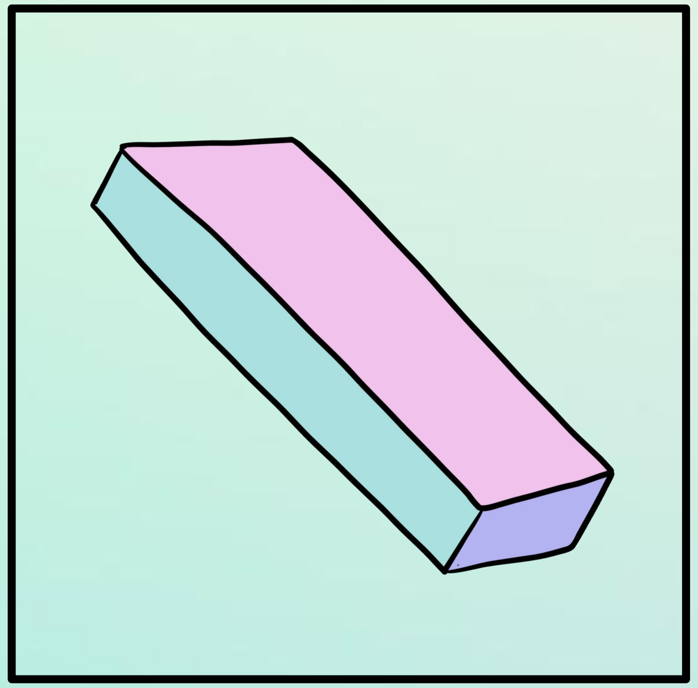
vertices < 4 as green.

The 0th vertex is colored orange.

Also the vertex are scaled to size.





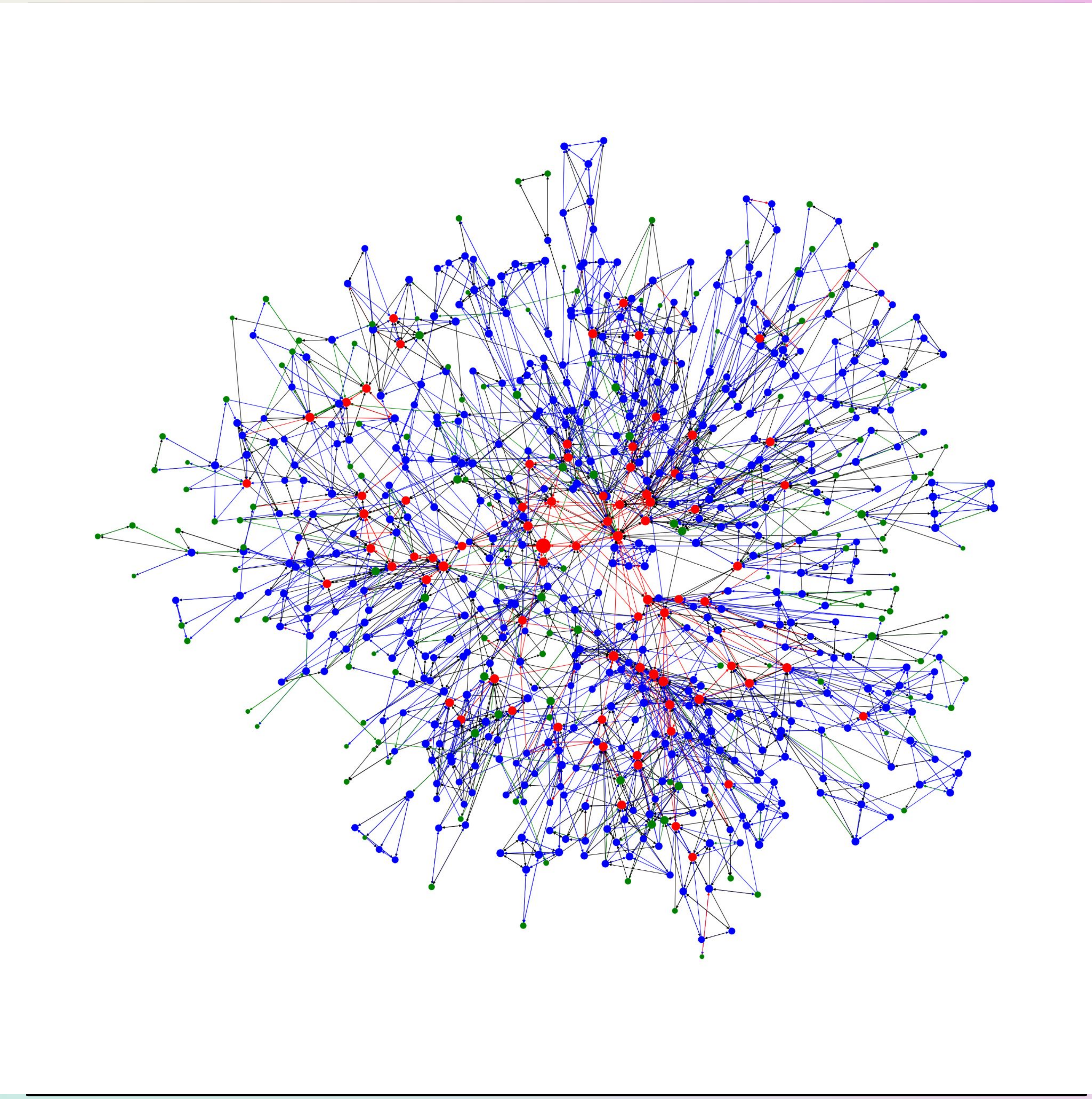


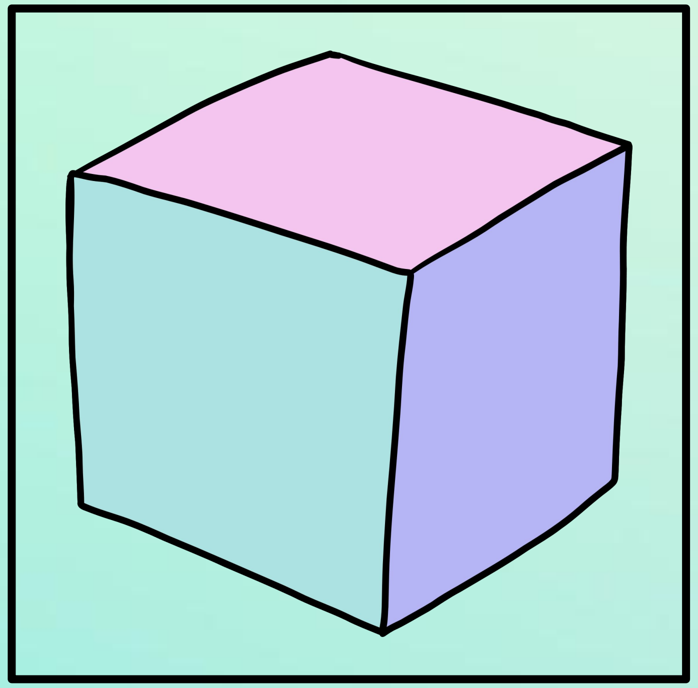
Largest Connected Component

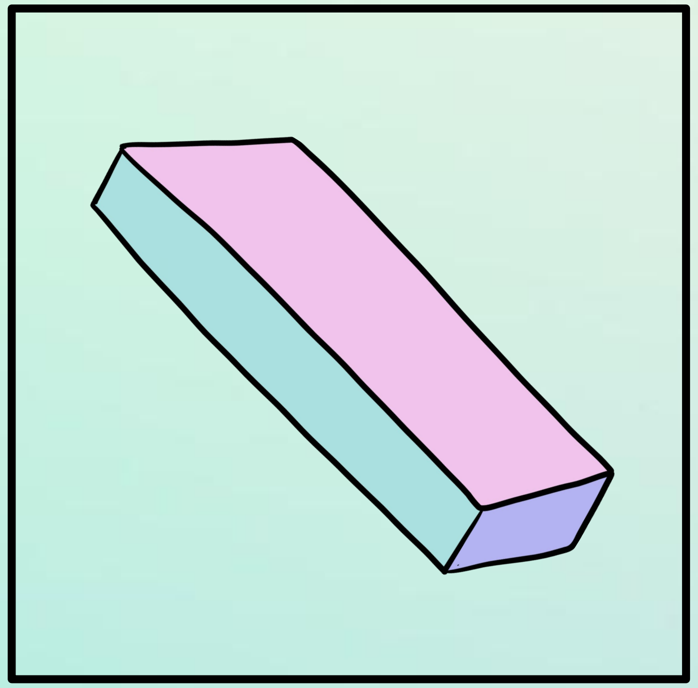
We visualize the largest connected component of a graph.

It has 720 nodes and 2929 edges.

Its radius is 9 and Diameter is 21(The same as the original graph which makes sense when you considered the methodology we used to calculate diameter)







Comparison of our network with random graph with same number of nodes and (approx) number of edges

Local Clustering Coefficient

Average Shortest Path

Degree Distribution

Generating Random Network

Using Erdos-Renyi Model

We use the Erdos-Renyi model which is of the form G(n,p)

Calculating Probability

We calculate the probability by using the fact that

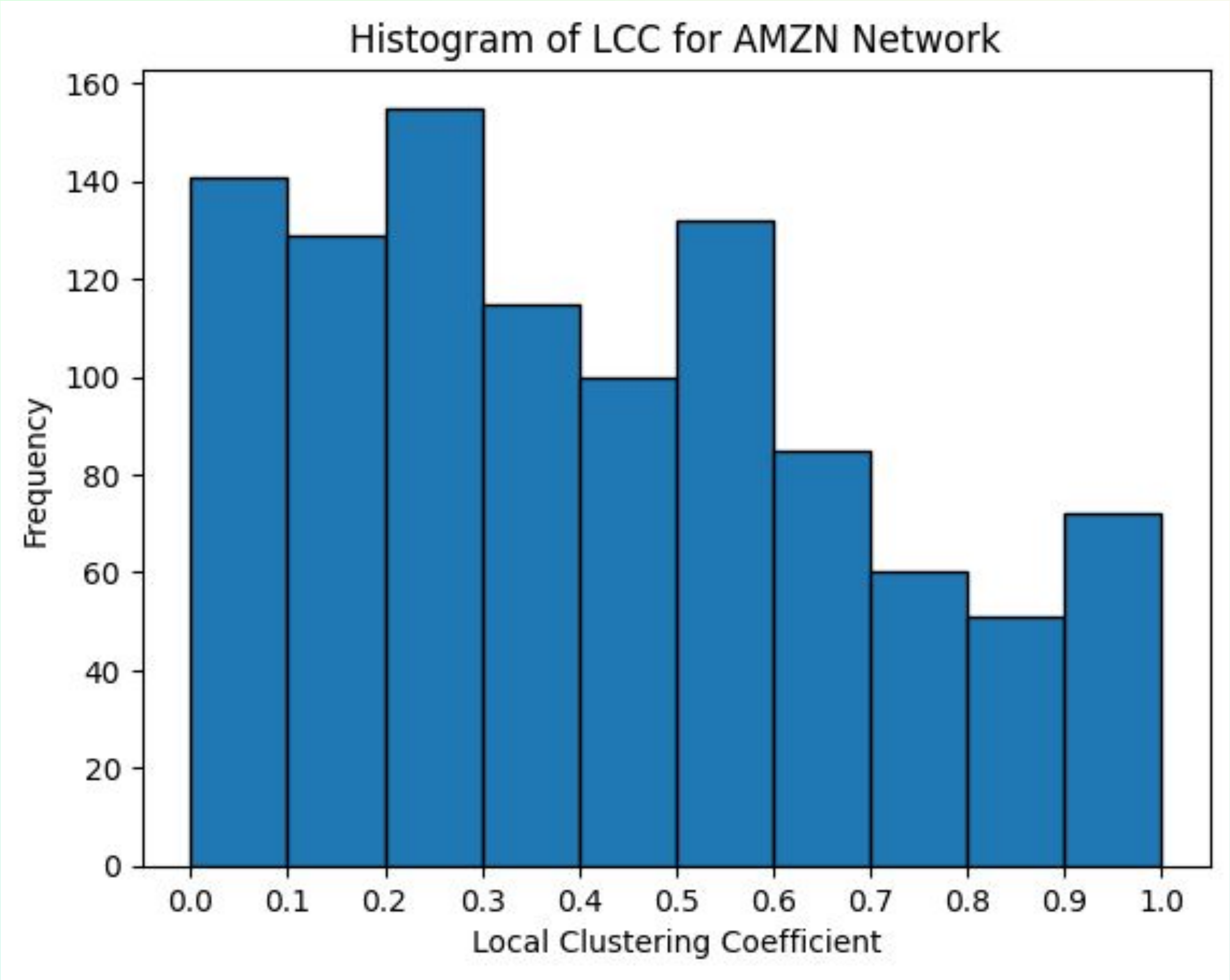
p ~edges/nodes(nodes-1)

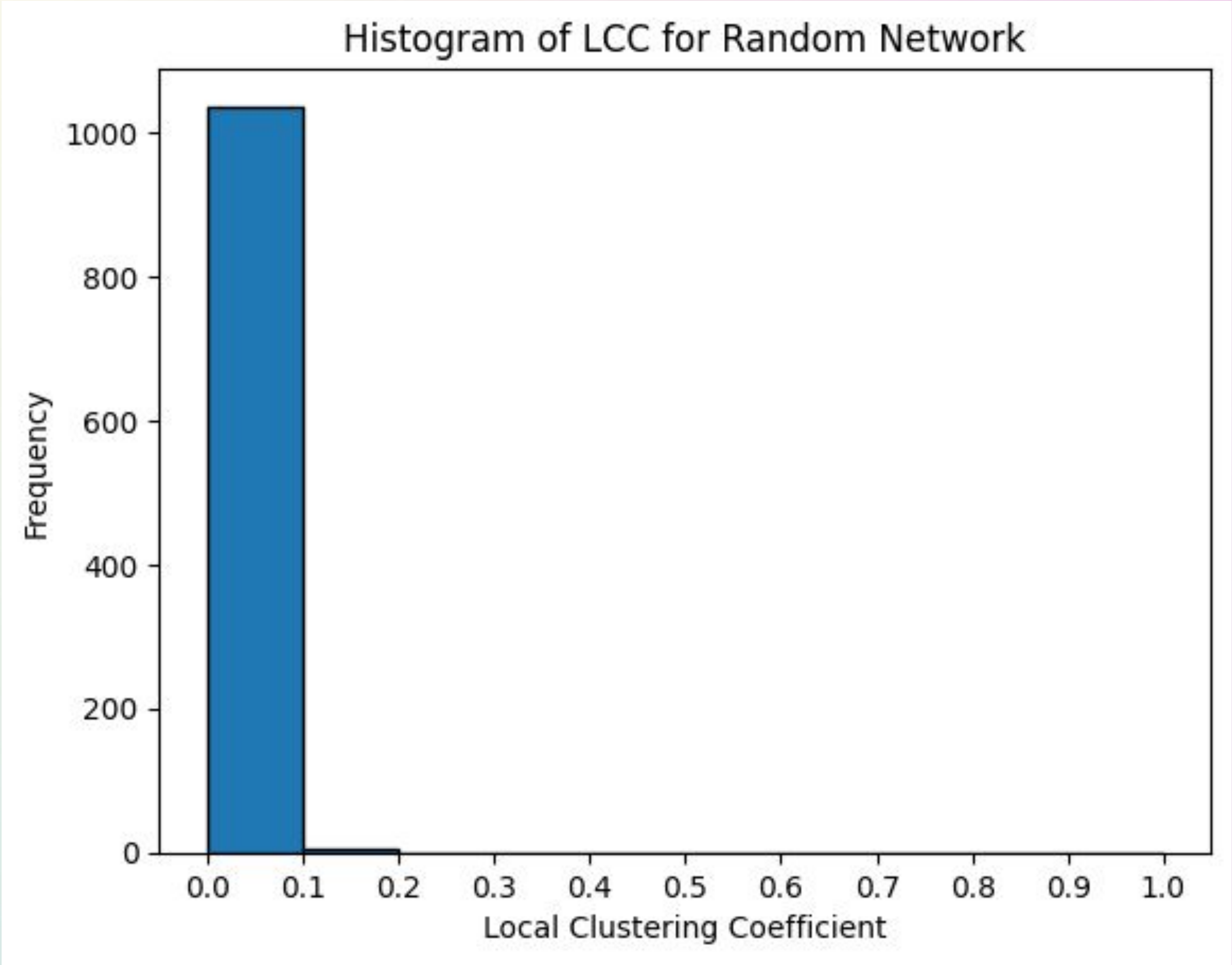
We find that the probability is 0.003884097134818983

Local Clustering Coefficient

As we can see that while the histogram for our graph is distributed all over the ranges the random network is not.

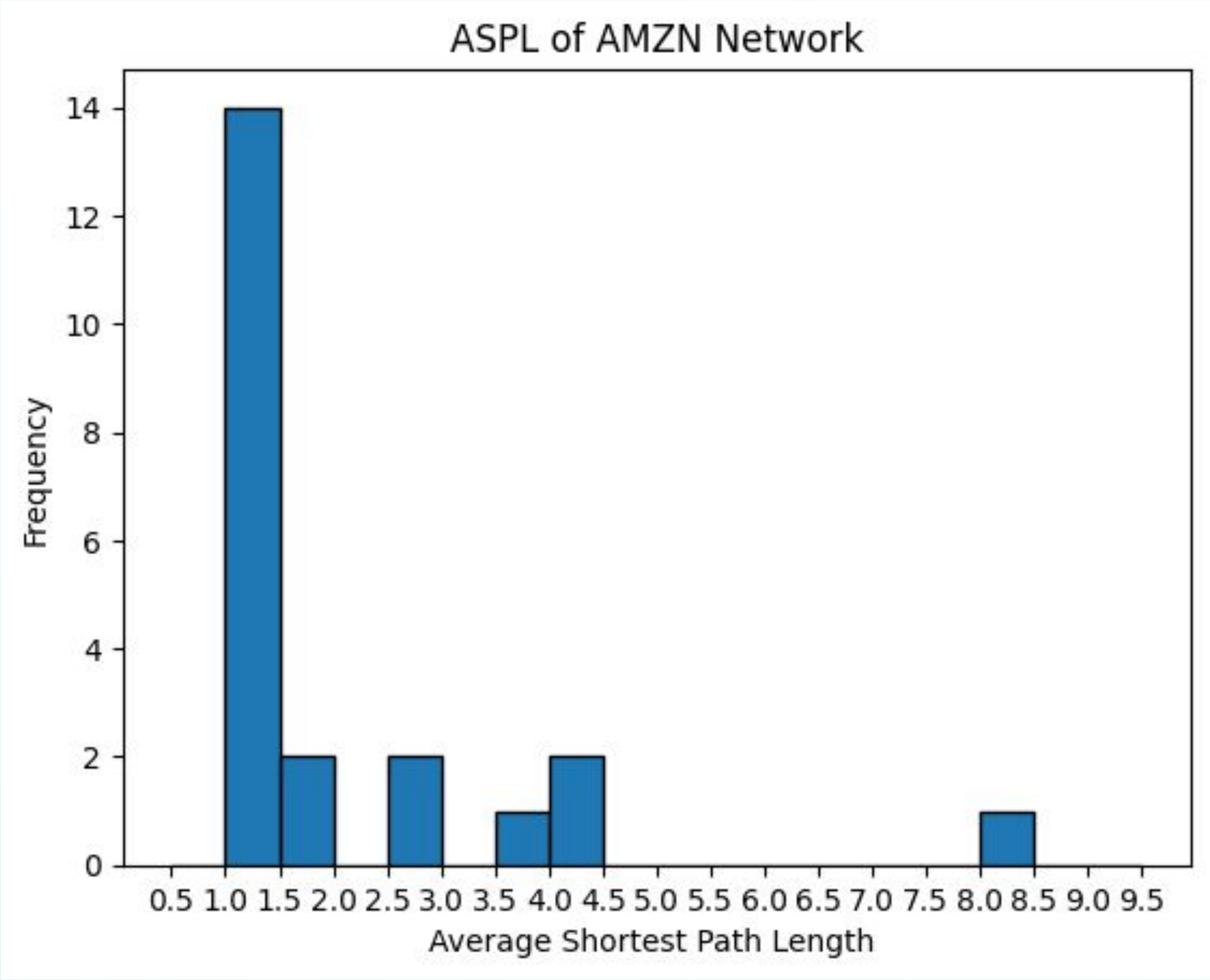
This is due to the simple fact that if you buy a product say a pencil then you are likely to also buy a eraser and a sharpener.

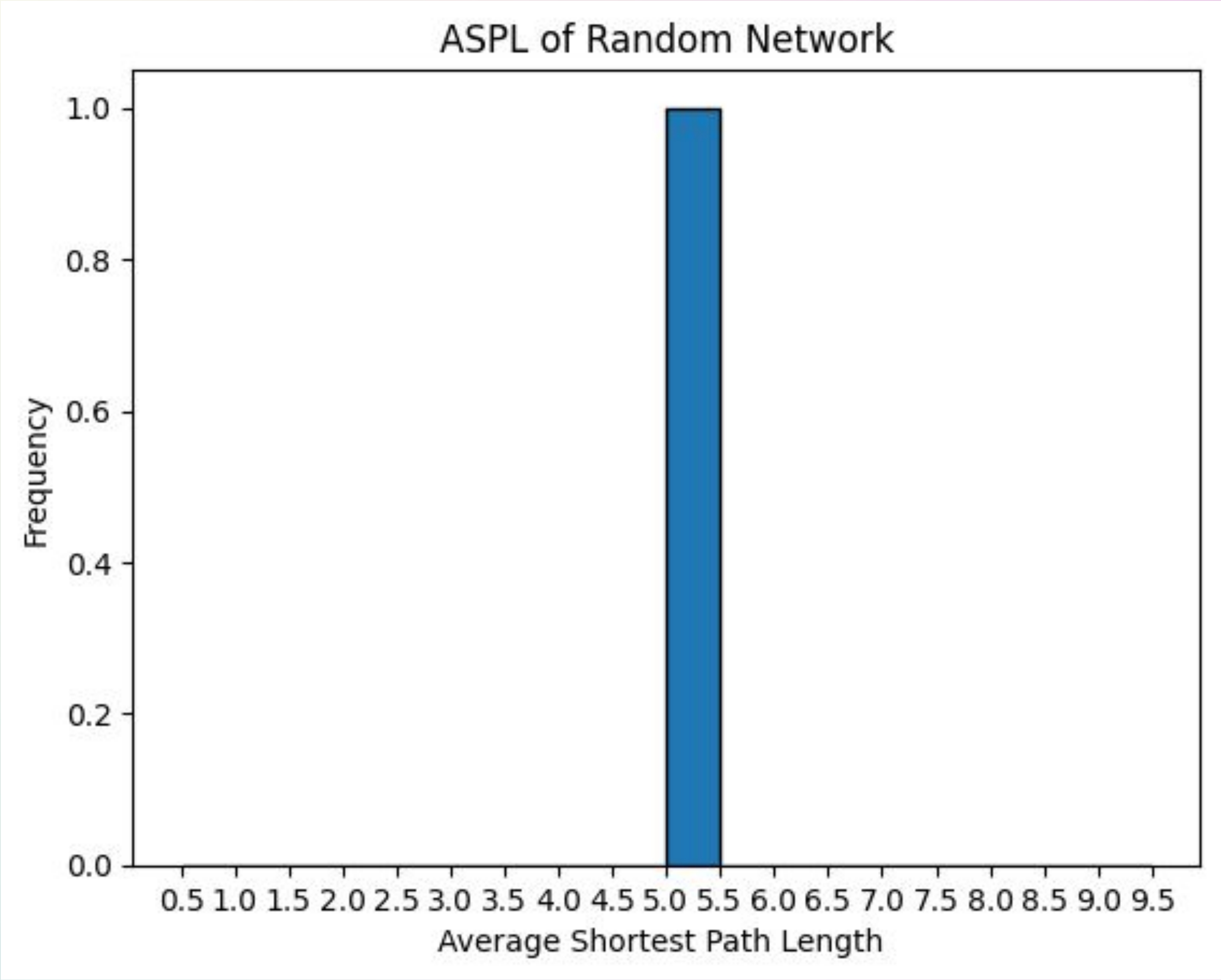




Average Shortest Path

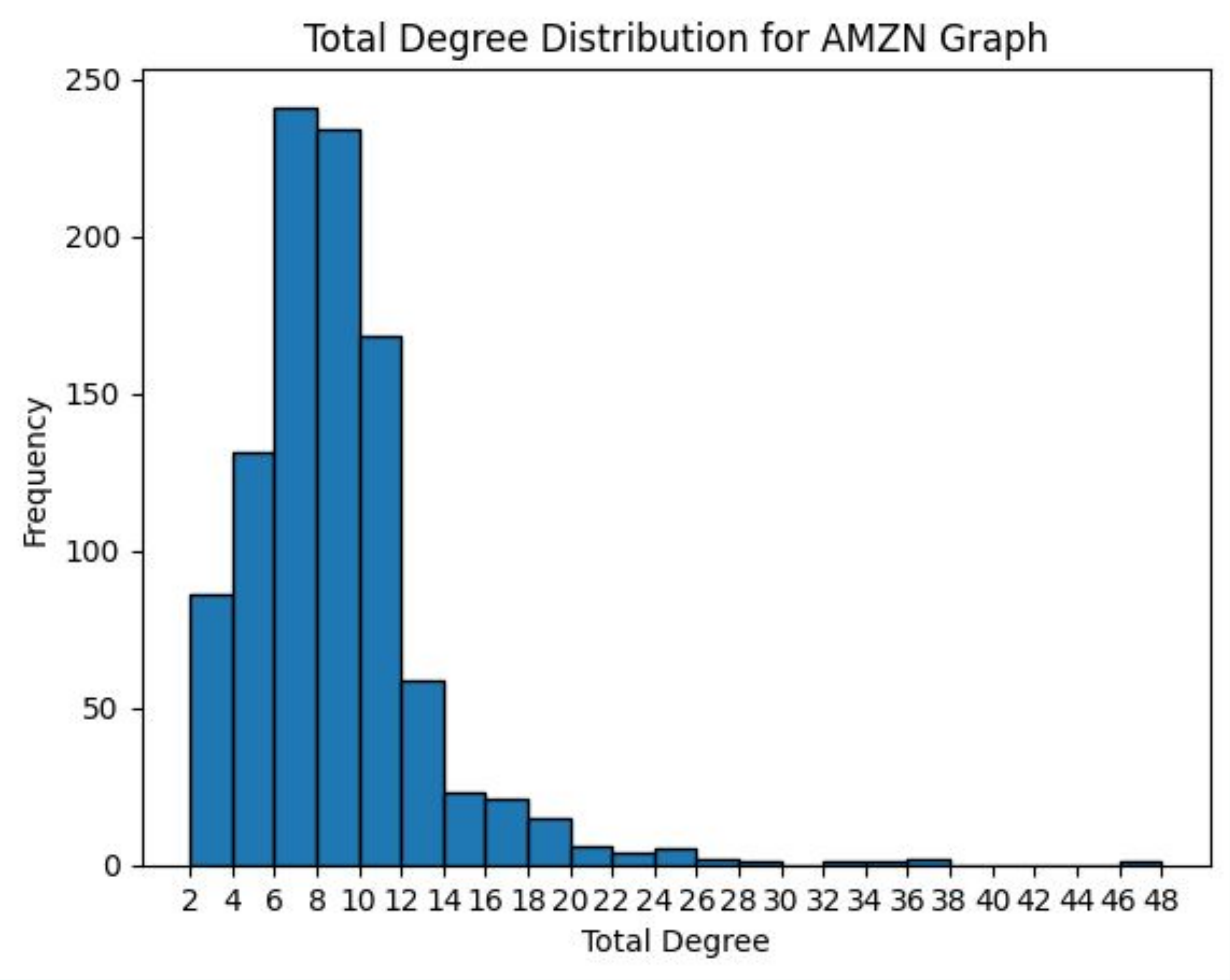
We took all the connected components and found their shortest path length. It is interesting to note that while our network had around 22 SCC, the random network had only one SCC which makes sense as the number of SCC is bounded by log(N) in a directed random graph where p<=c/n

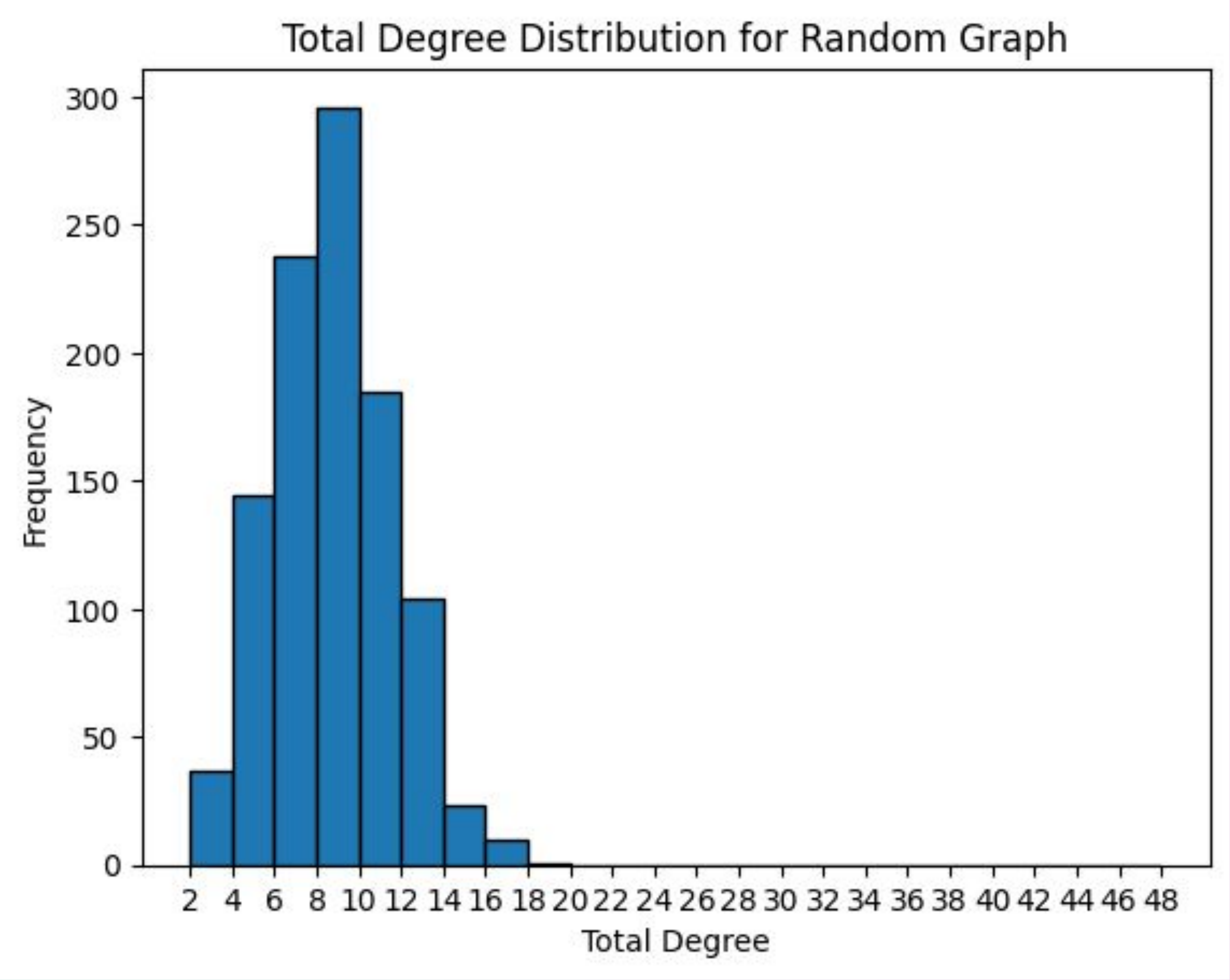




Total Degree Distribution

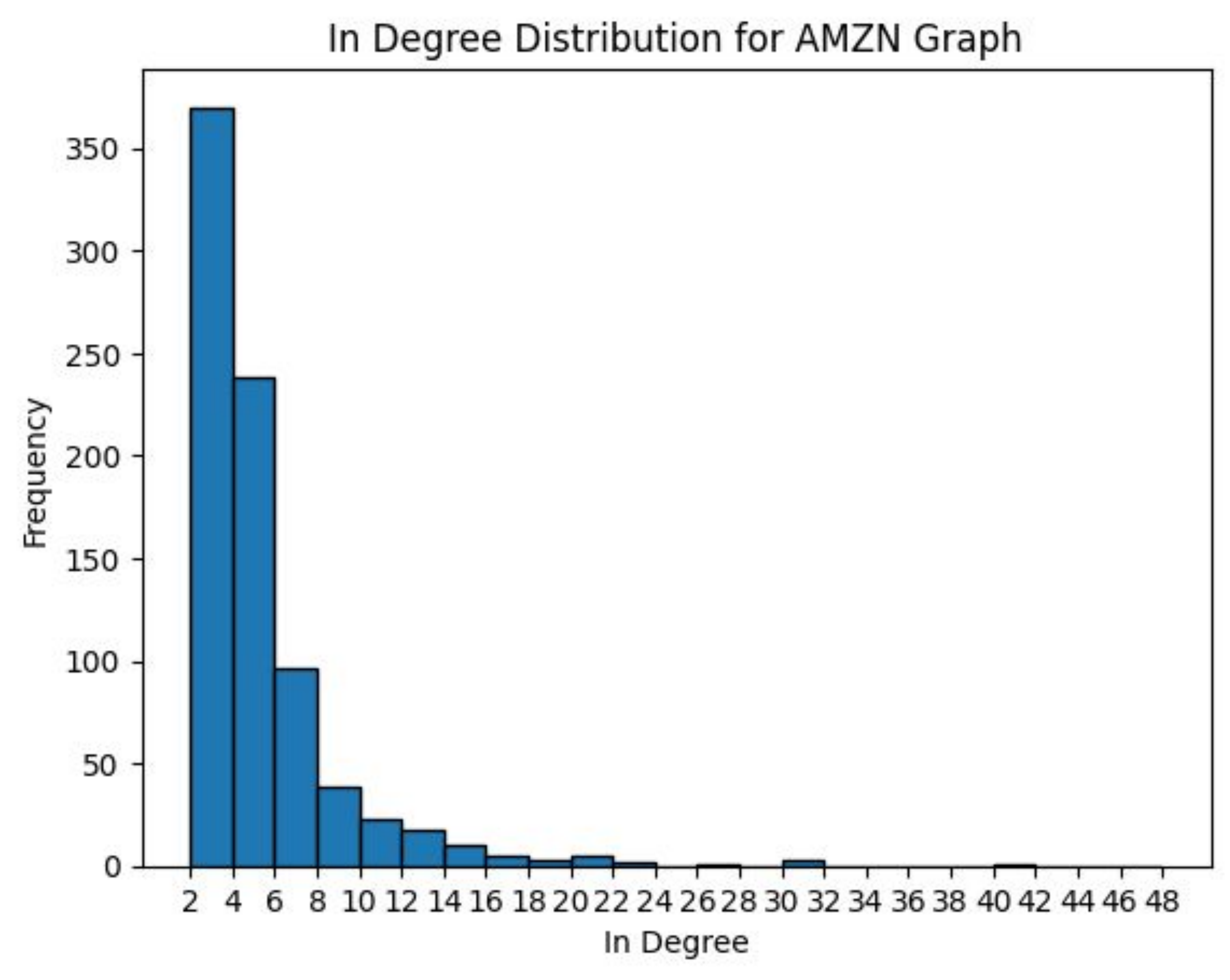
The degree distribution of our network is very similar to the binomial distribution of the random graph.

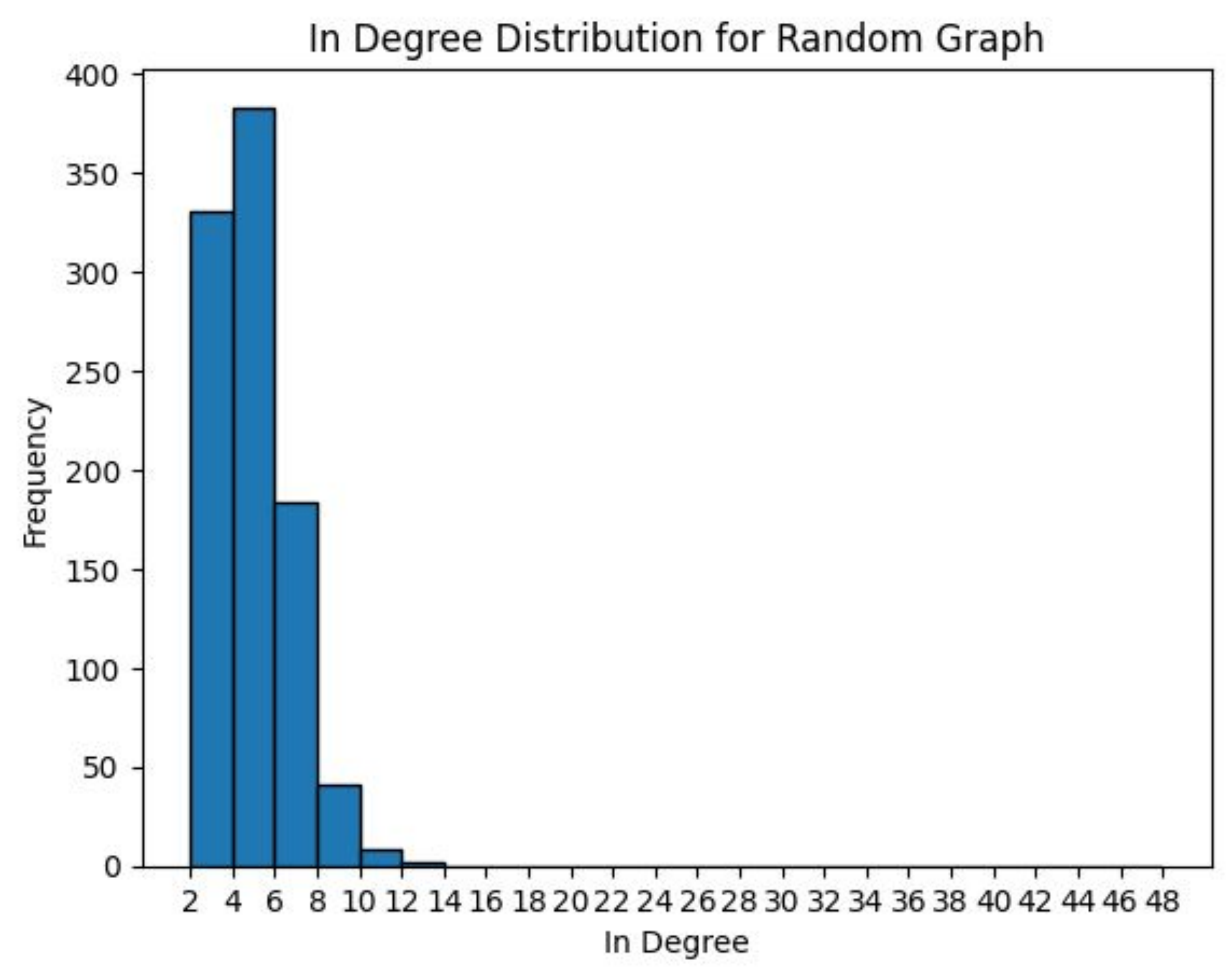




In-Degree Distribution

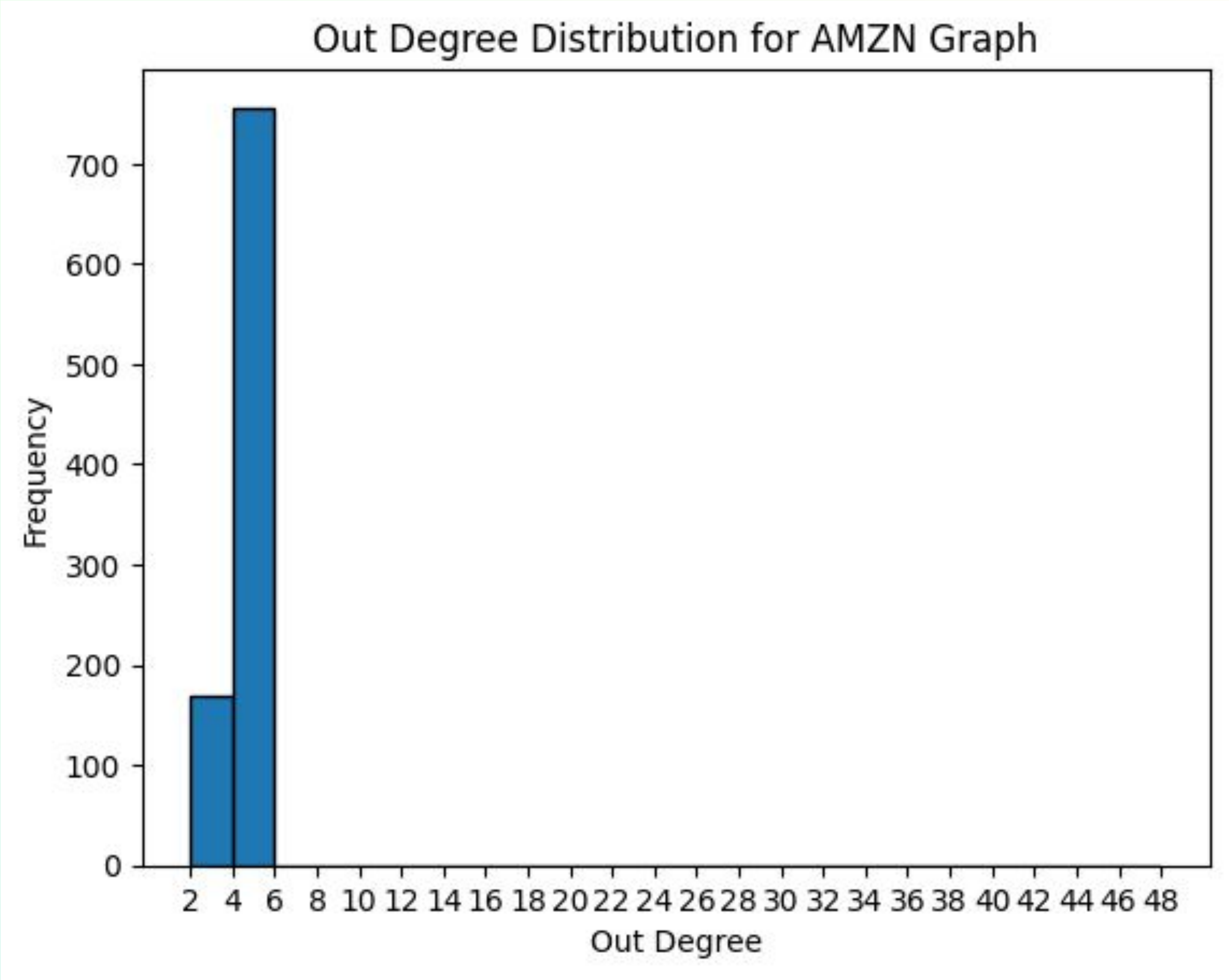
As you can see even the In-Degree distribution of the graph is almost same binomial-ish except for the tails, therefore does it make sense to assume that the out degree distribution will also be the same? No! We shall find the reason why in the next slide

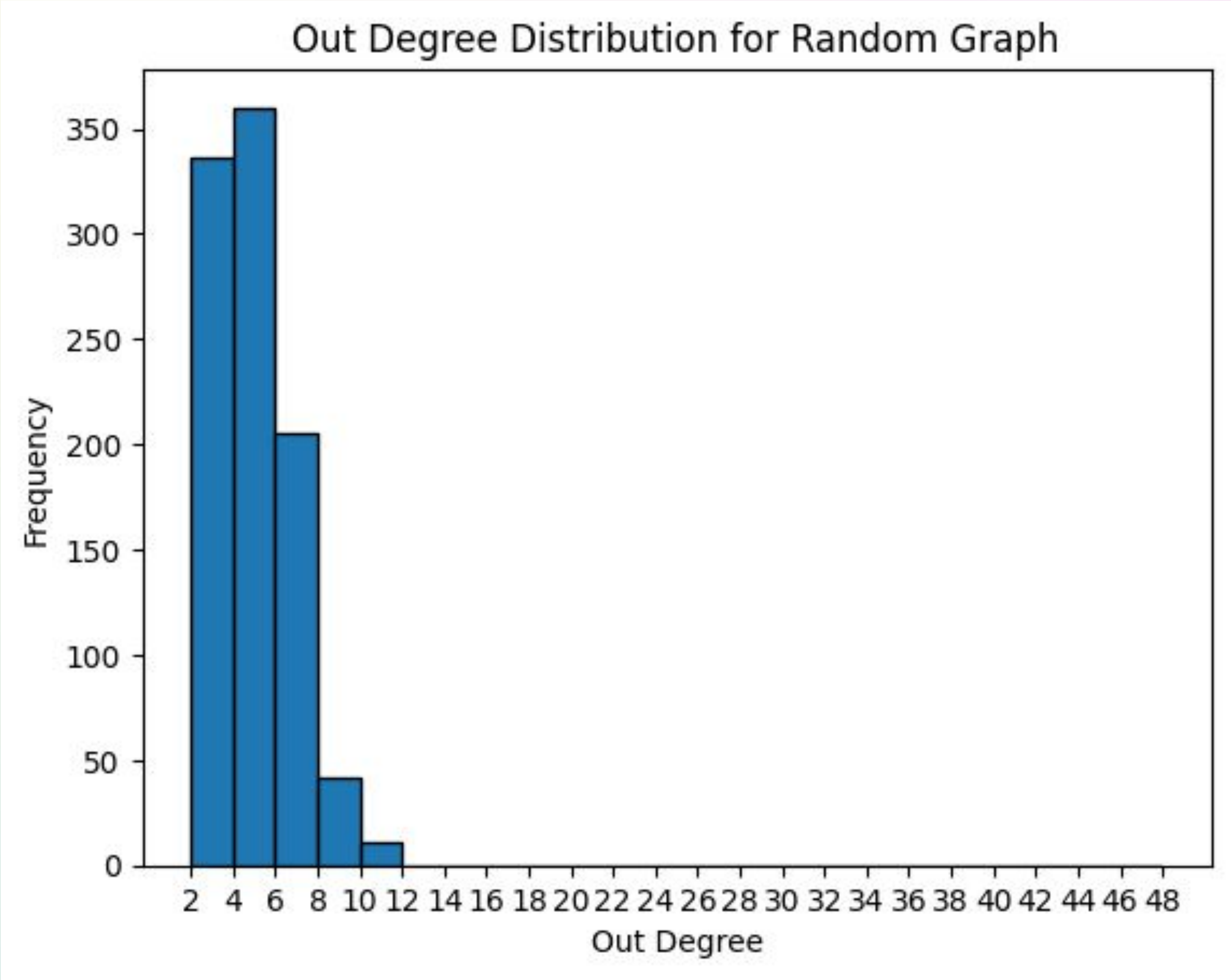




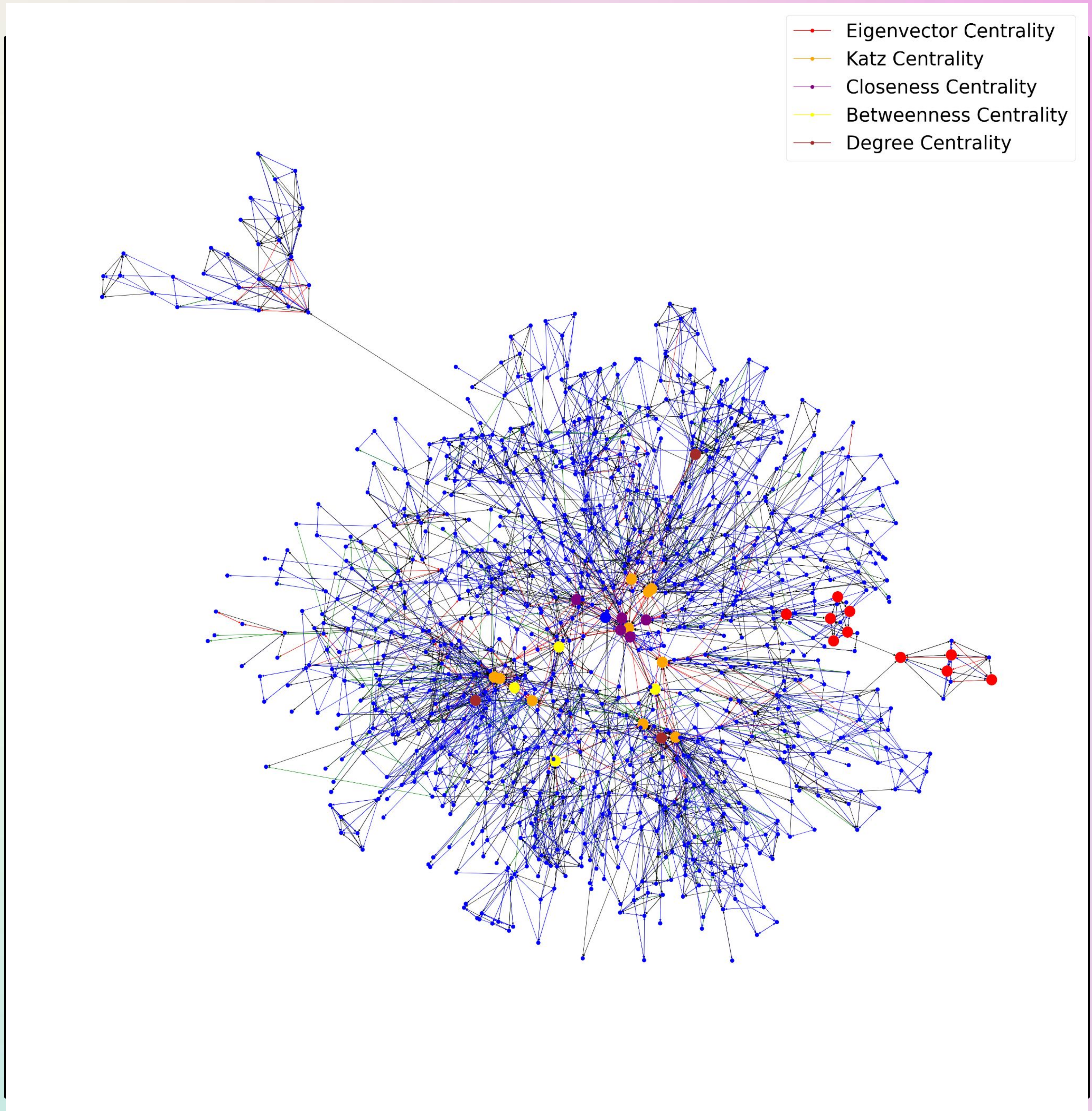
Out-Degree Distribution

As you can see they are different, this is due to the fact that a person is not gonna buy 12 different kind of products at the same time.

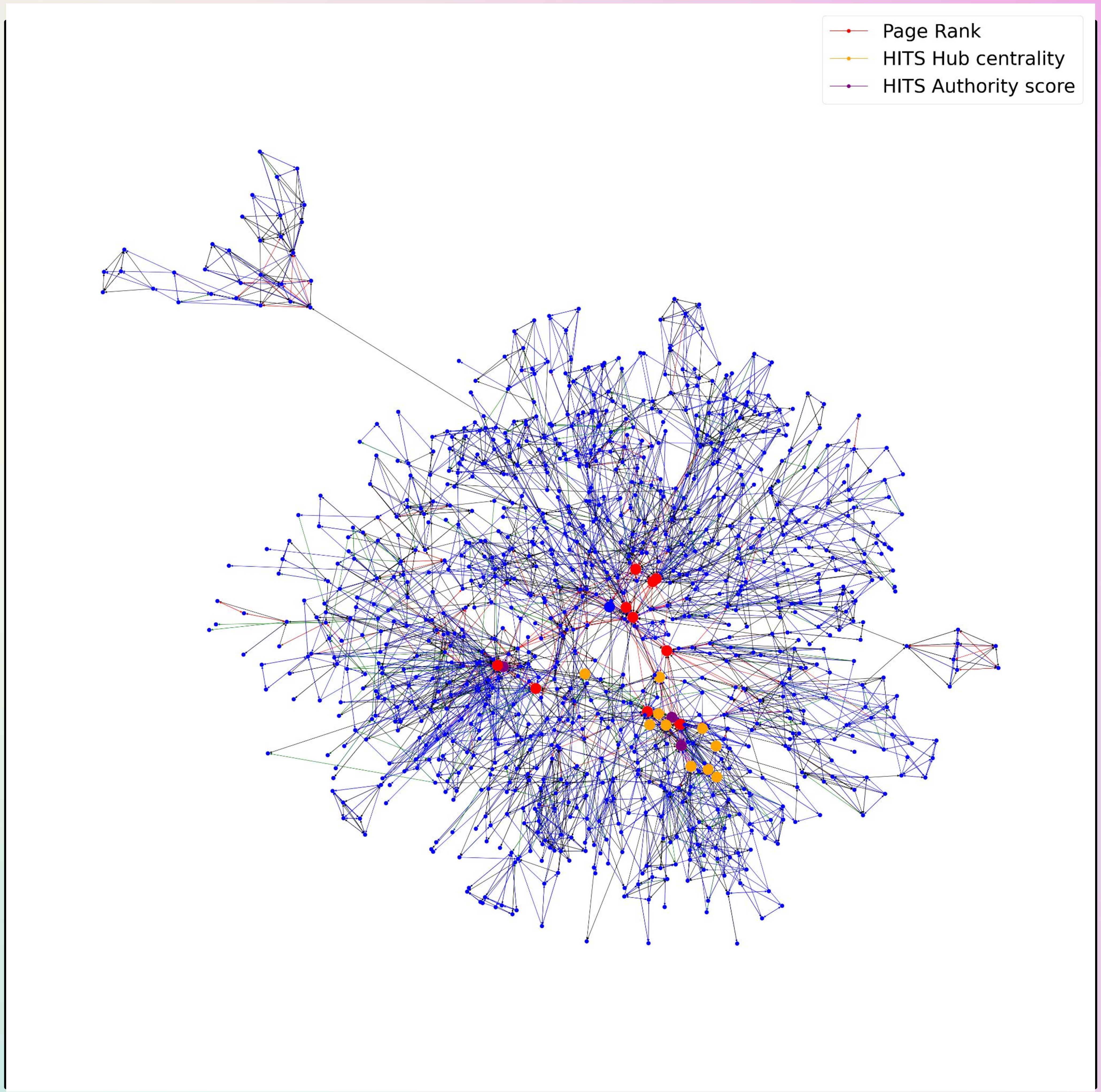




Network Visualized with the top 10 nodes by all types of centrality colorized



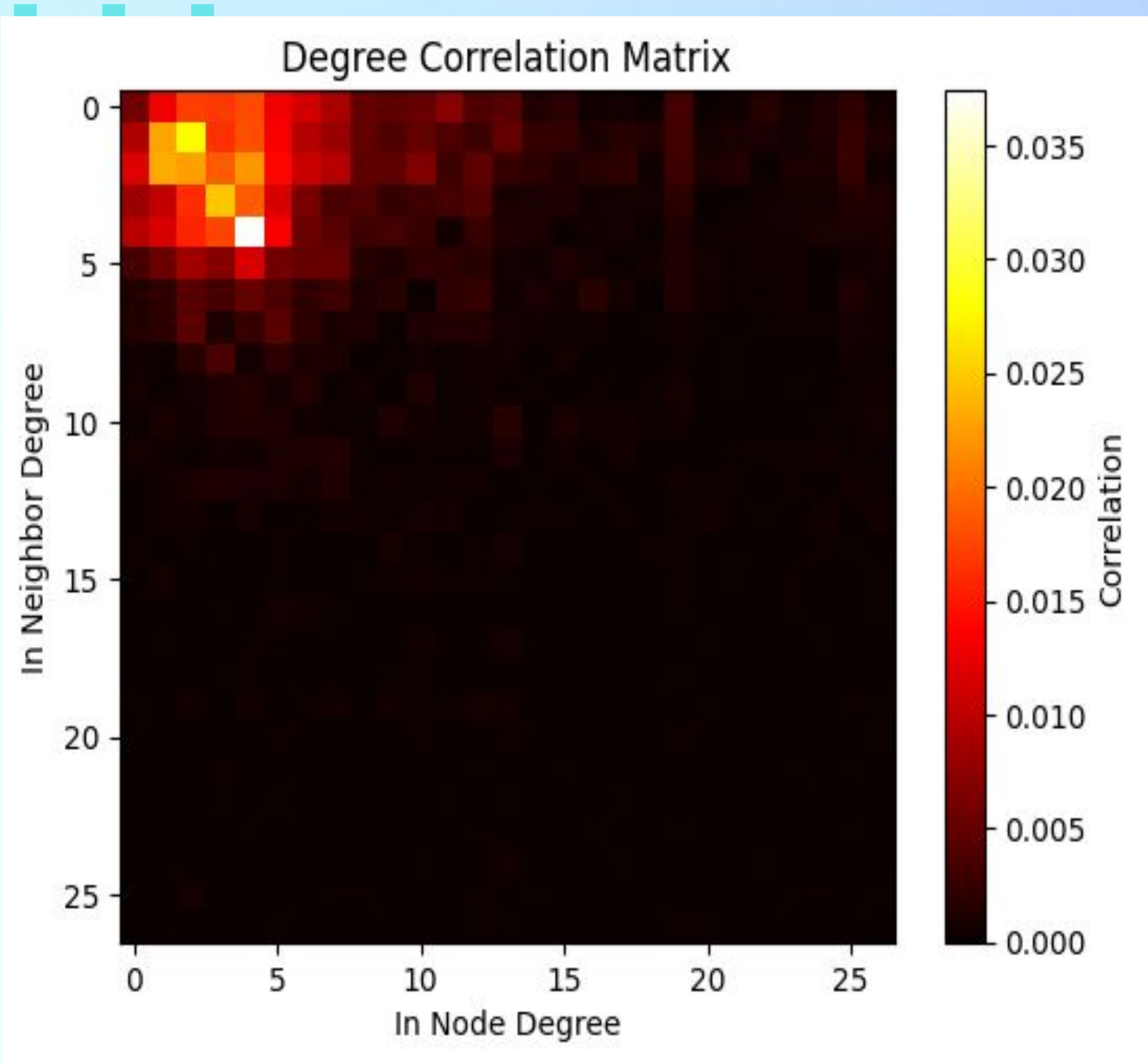
Network Visualized with the top 10 nodes by Page-Rank/Hits

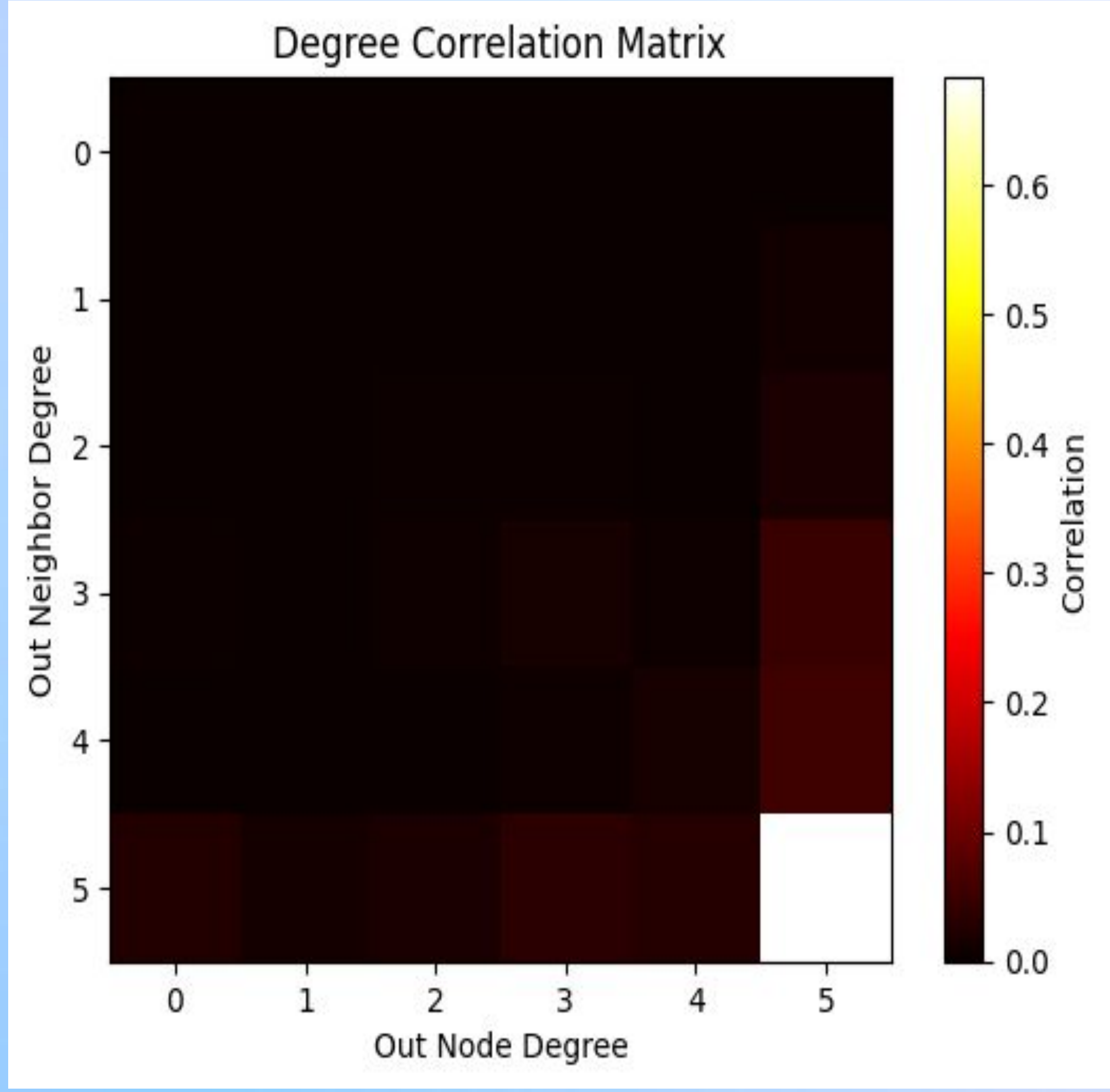


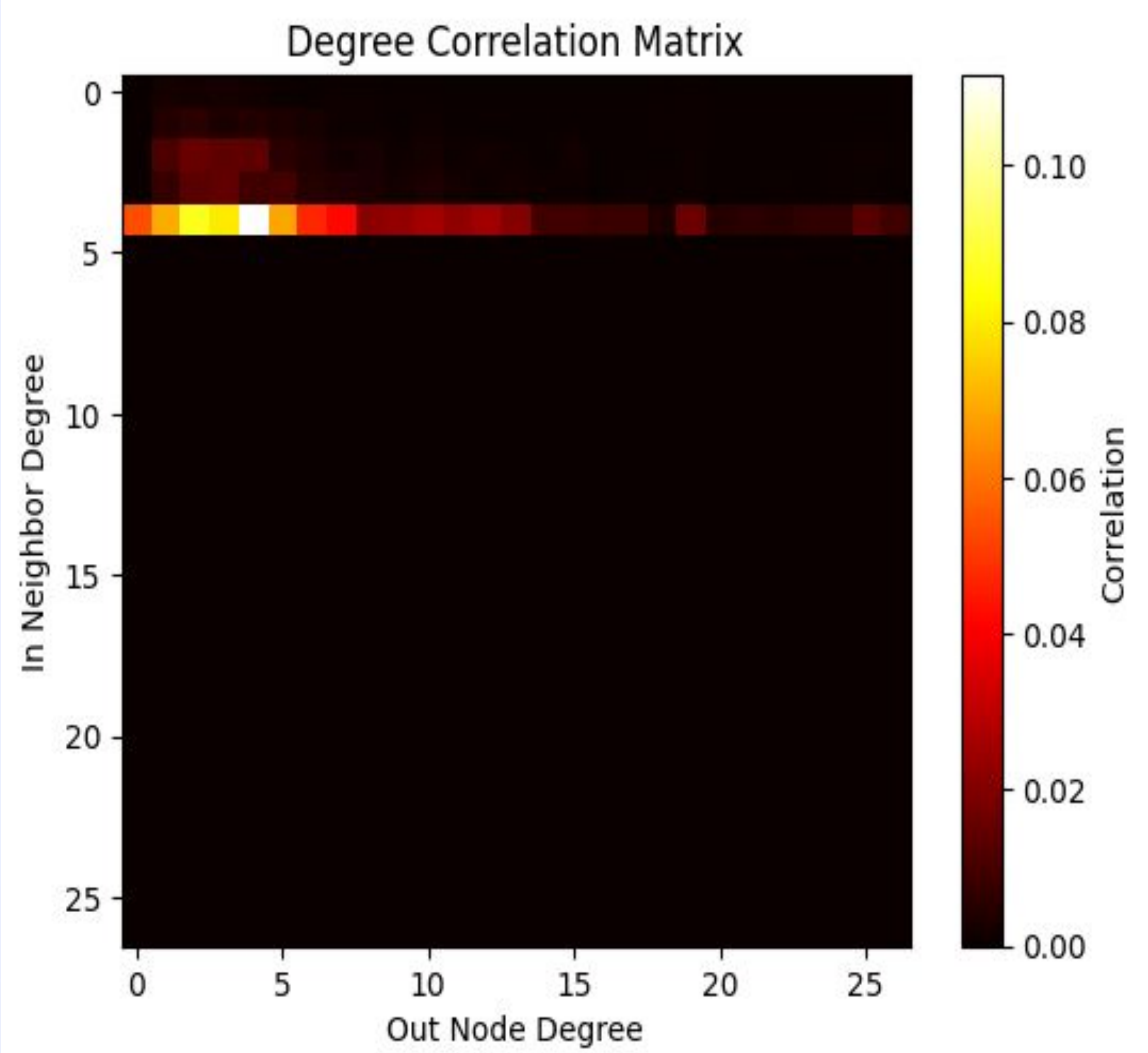
Assortative Mixing Matrix

We now visualise the Assortative Mixing Matrix for all three cases:- (x='in',y='in'), (x='out',y=out), (x = 'out',y='in').

If you are confused please check the slides for In-Degree and Out-Degree histogram.







THANK YOU!

Please don’t hesitate to ask any follow up questions!