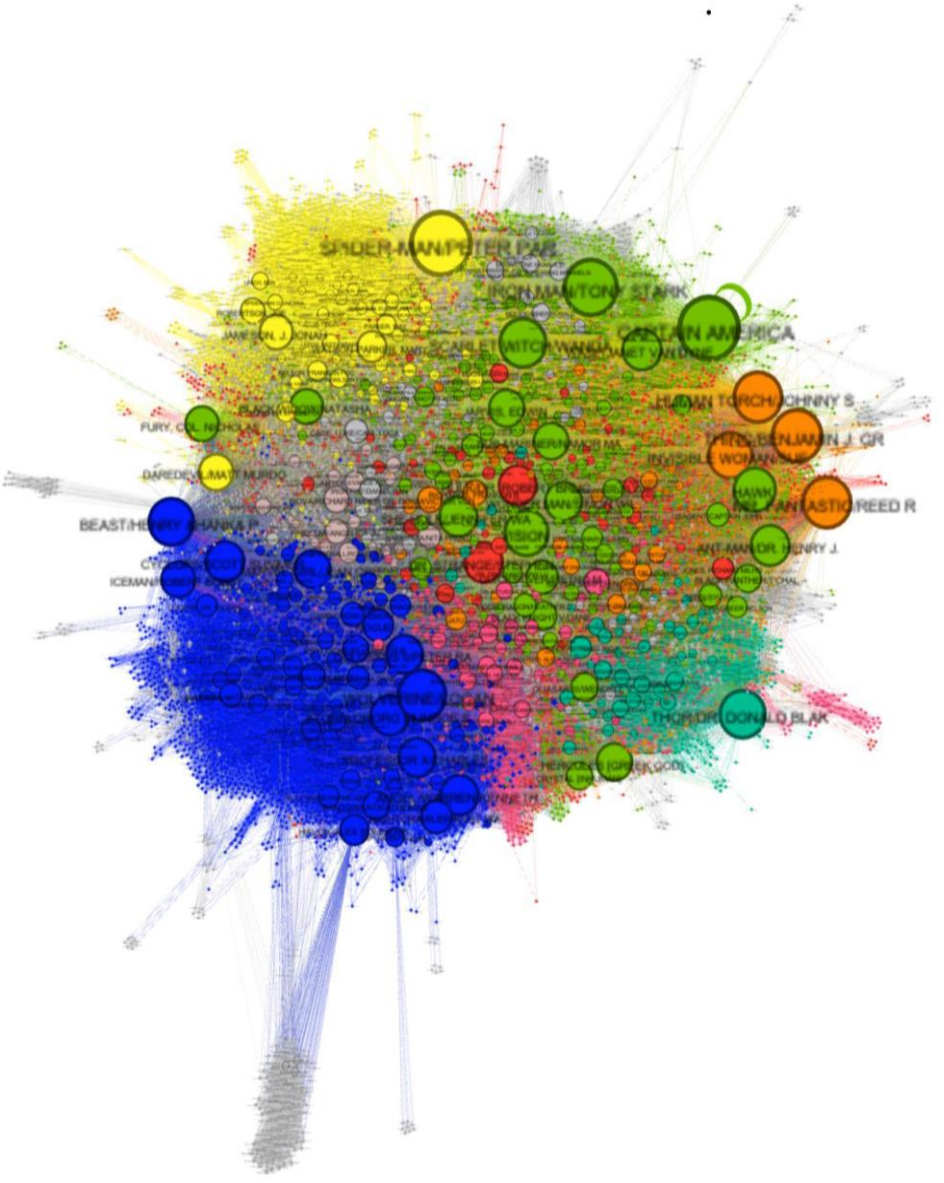
Marvel Comic Book Universe Dataset and Analysis:

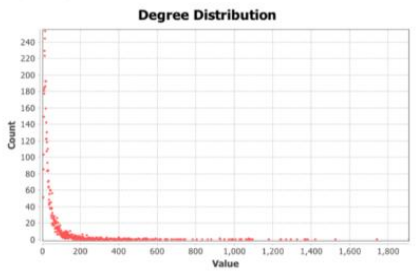
Overall Proposal and Dataset: The Marvel comic book universe dataset we obtained contains 6421 nodes where each node represents a character in Marvel's entire comic book universe which is made up of multiple comic book franchises. The graph consists of 167112 edges where each edge represents whether 2 characters have appeared in the same comic book together. The largest yellow node represents Spiderman and the yellow cluster represents Spiderman's comic book franchise. Similarly the blue cluster represents the X-men franchise and the 4 orange nodes on the right represents the Fantastic Four franchise and so on.

Algorithm Information and The Problem We Want to Solve: Given Marvel wants to create a new comic book franchise including the top 5 popular characters in their entire comic book universe we used the K-Means clustering algorithm to divide the data into clusters to detect communities and then used different colours to represent each community. We then calculated the betweeness centralities of each node and made the nodes with higher betweeness centralities larger to easily identify the most popular characters/nodes. From the graph we found the following characters amongst the most popular: Spiderman, Captain America, Ironman, Scarlet Witch, Thor and the members of the Fantastic Four franchise.

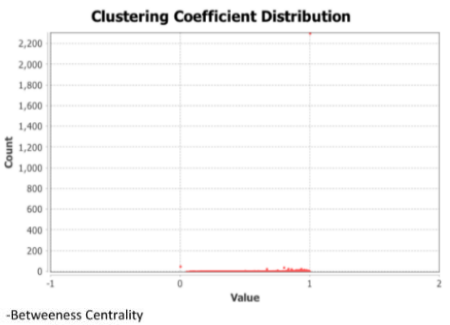


Network Statistics:

- Average Degree Distribution = 52.052



- Average Clustering Coefficient = 0.781



-Betweeness Centrality
Diameter: 5
Radius: 1
Average Path Length: 2.6383862245530985

- Modularity = 0.499
Number of Communities = 31