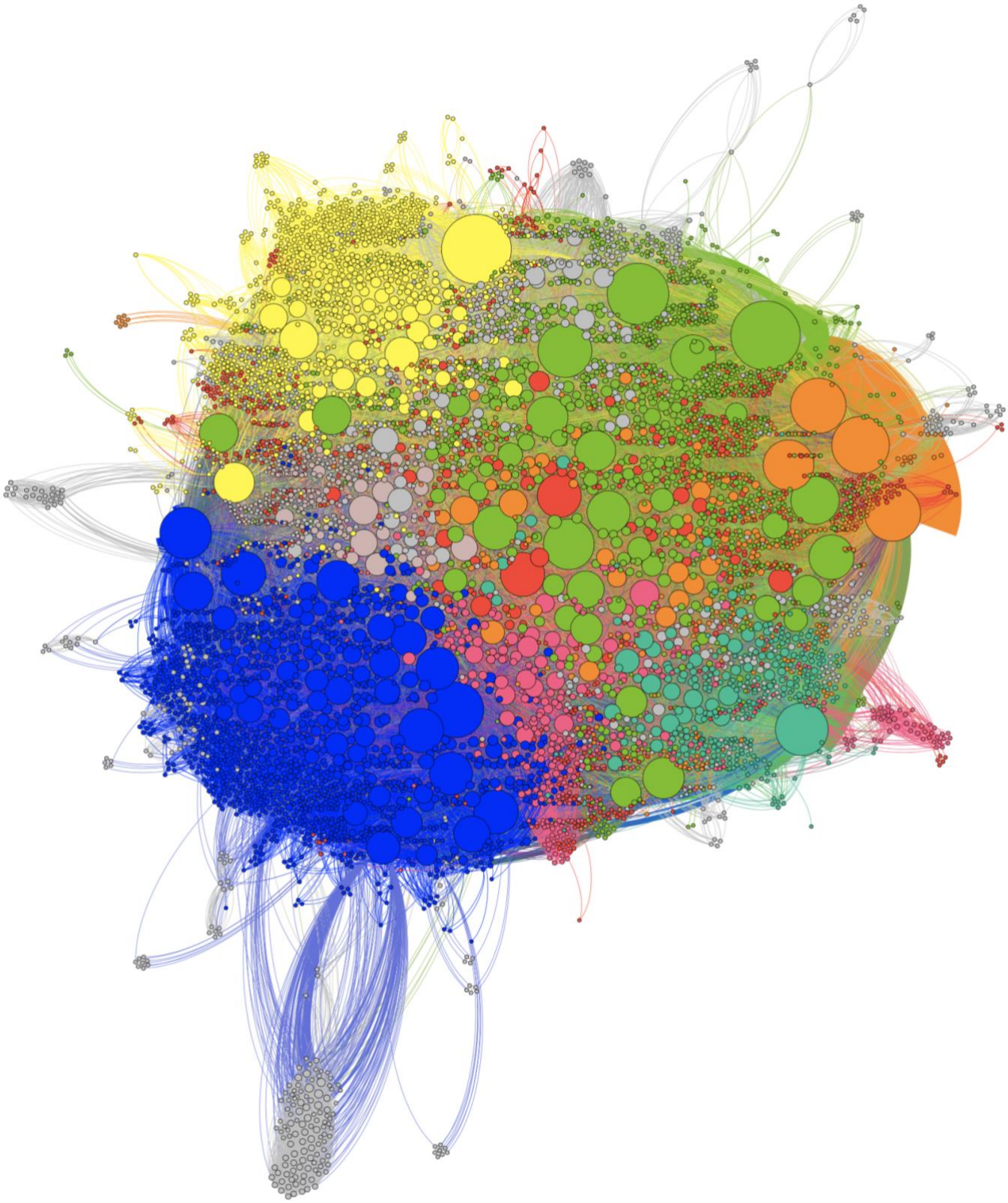
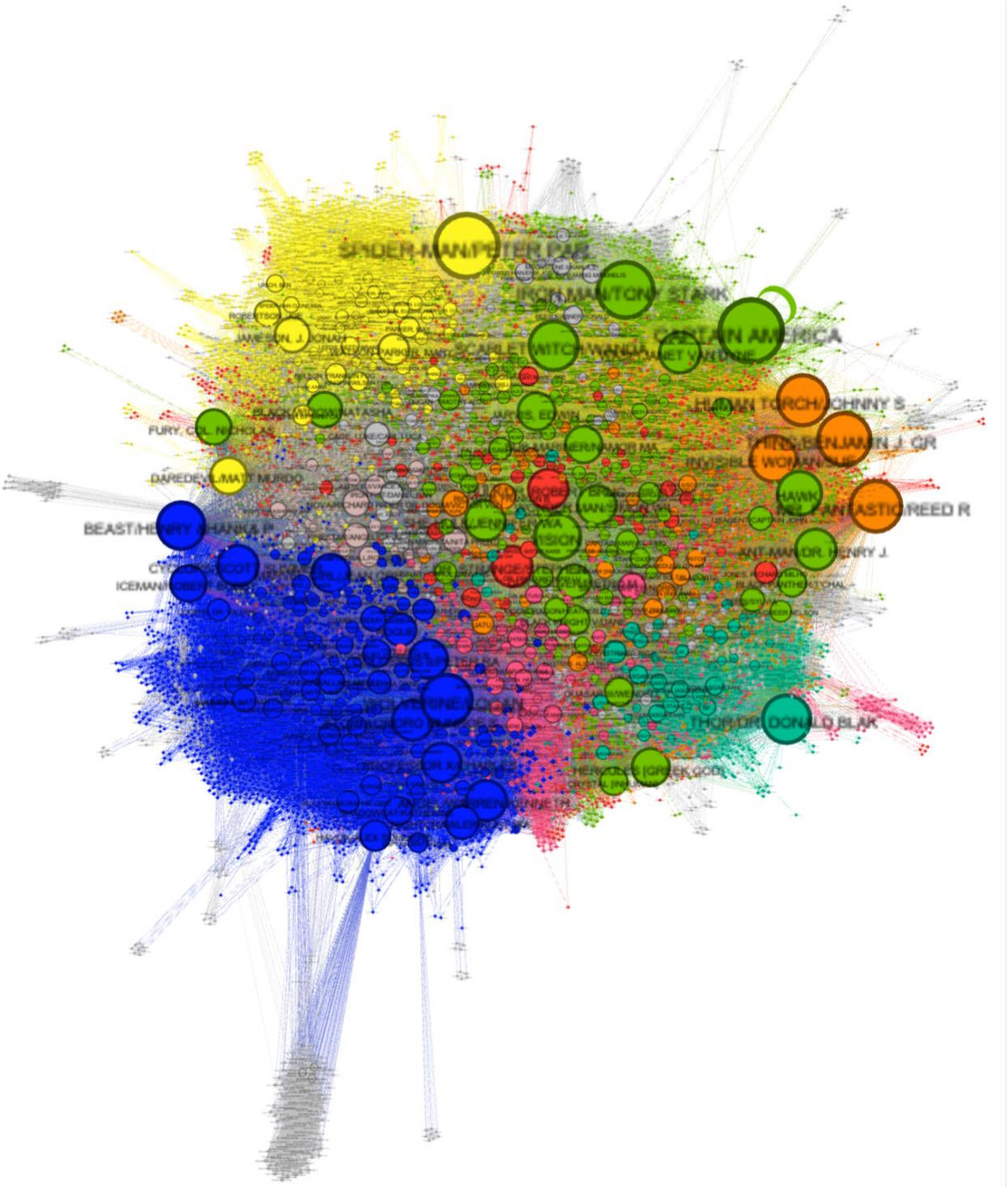
ECS637U
Digital Media and Social Networks
Coursework 2 – Network Analysis

Index:

Graph1: Marvel Comic-book Universe Social Network (Nodes = 6421, Edges = 167112)



Graph1.1 (Graph1 with Labelled nodes):

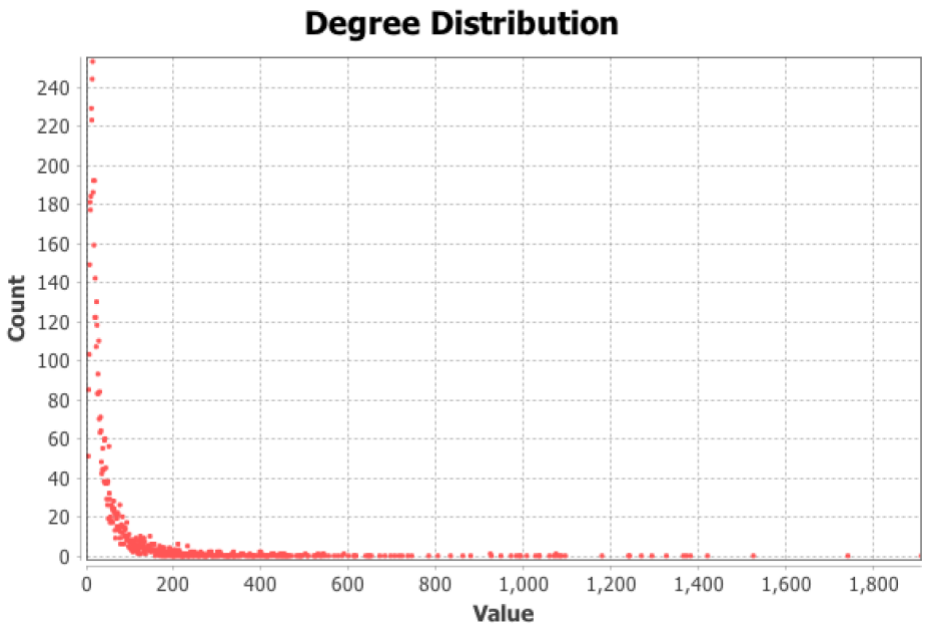


Graph1 Information:

Graph1 is based on a marvel comic book universe dataset where each node represents a comic book character and each edge represents whether these characters have met. It is an undirected graph.

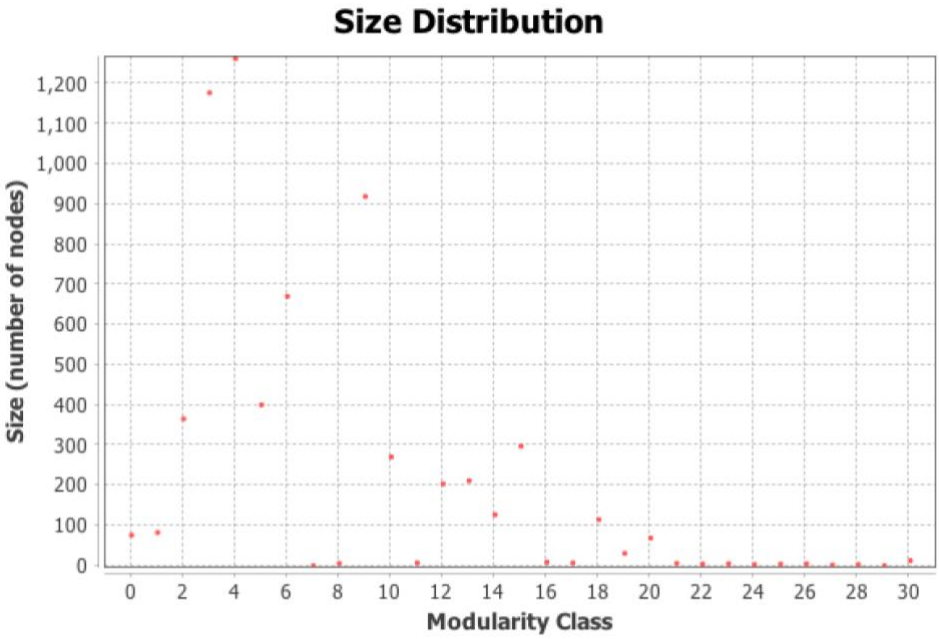
Graph1 measurements:

- Average Degree Distribution = 52.052



- Modularity = 0.499

  Number of Communities = 31



- Average Clustering Coefficient = 0.781

- Centrality

- Eigenvector Centrality

Parameters:

Network Interpretation: undirected

Number of iterations: 100

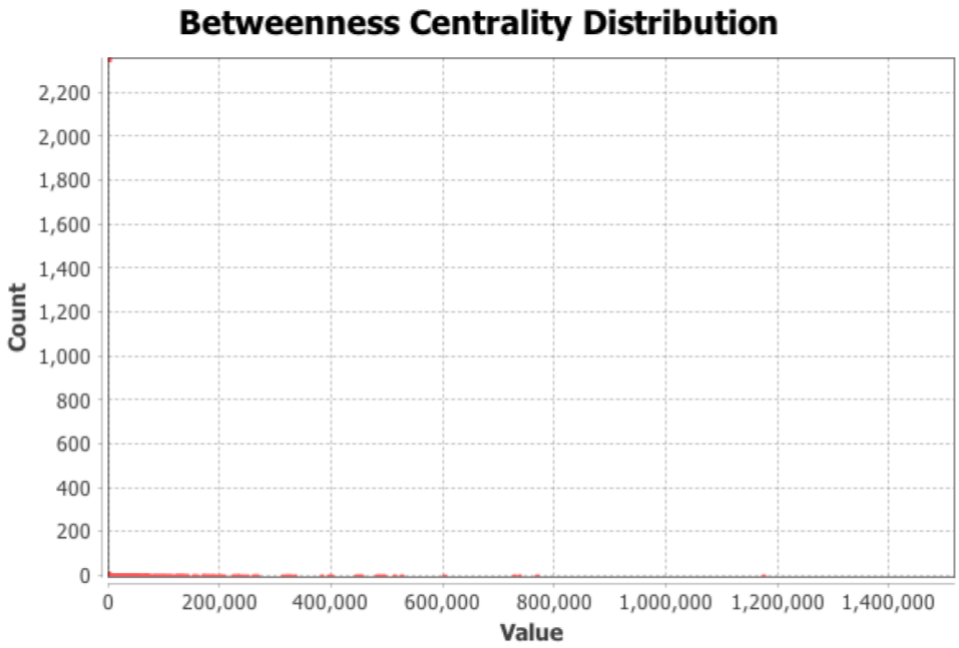
Sum change: 0.08939555928421349

-Betweeness Centrality

Diameter: 5

Radius: 1

Average Path Length: 2.6383862245530985



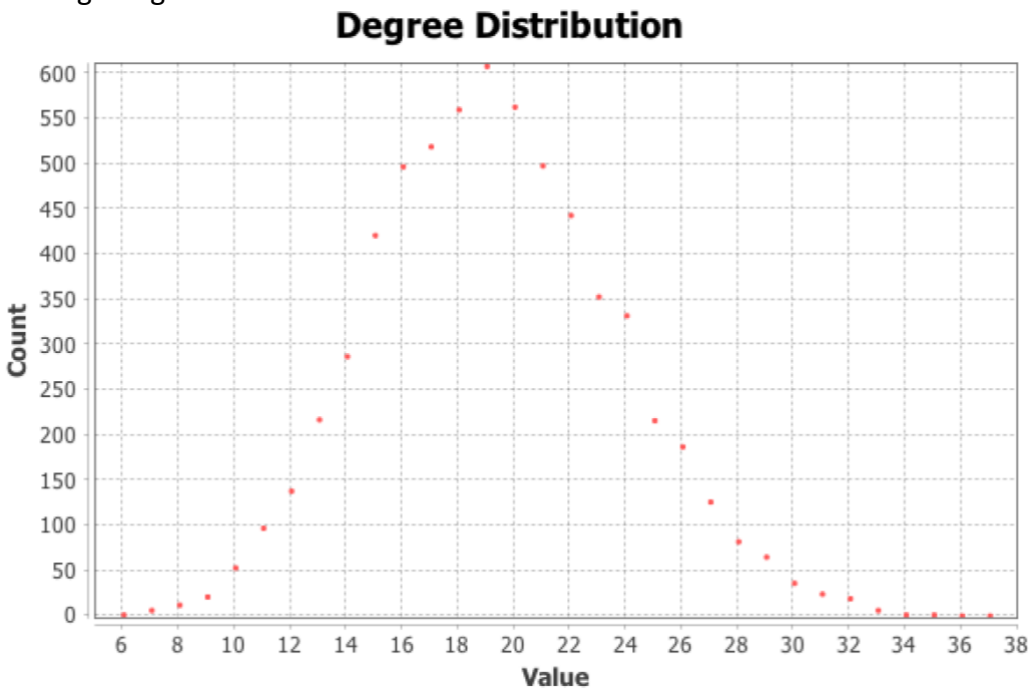
Graph2: Random Graph (Nodes = 6421, Edges = 61965)

Graph2 Information:

Graph2 is an undirected random graph generated from the same number of nodes as Graph1.

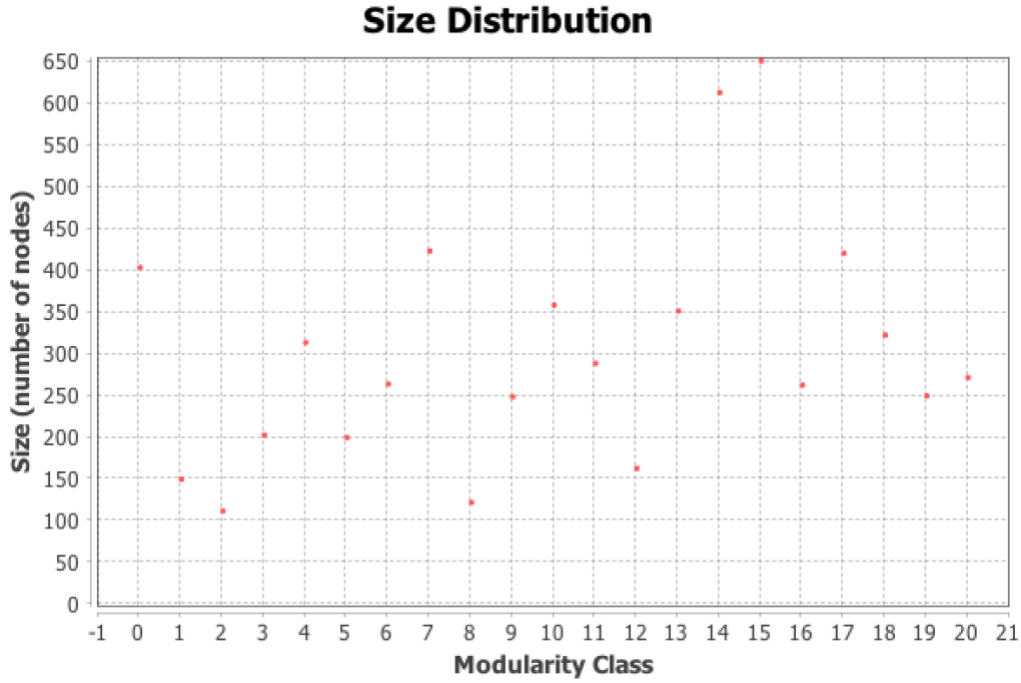
Graph2 Measurements:

- Average Degree Distribution = 9.65

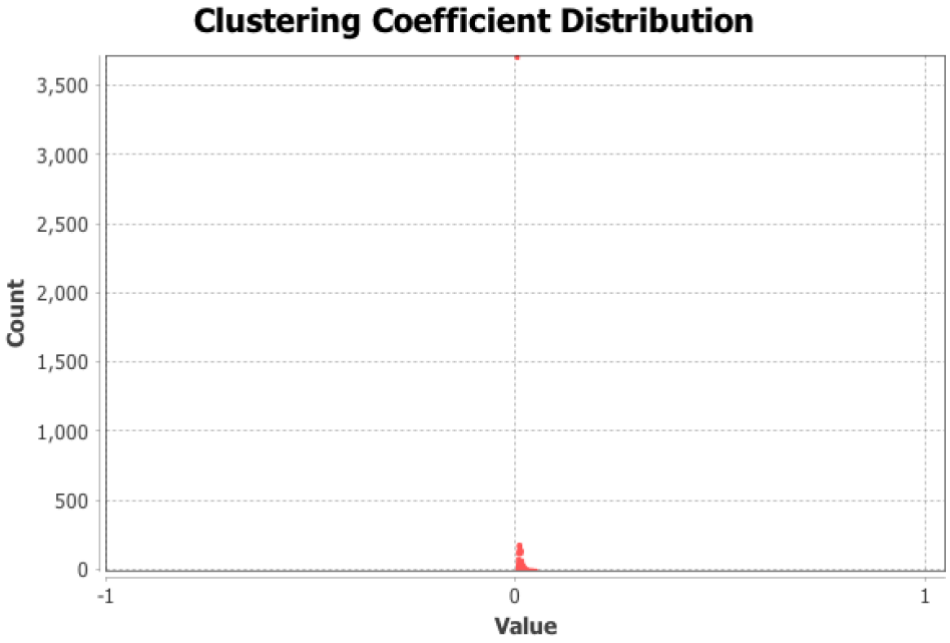


- Modularity = 0.147

  Number of Communities = 21



- Average Clustering Coefficient = 0.003



- Centrality

  - Eigenvector Centrality

    Parameters:

    Network Interpretation: undirected

    Number of iterations: 100

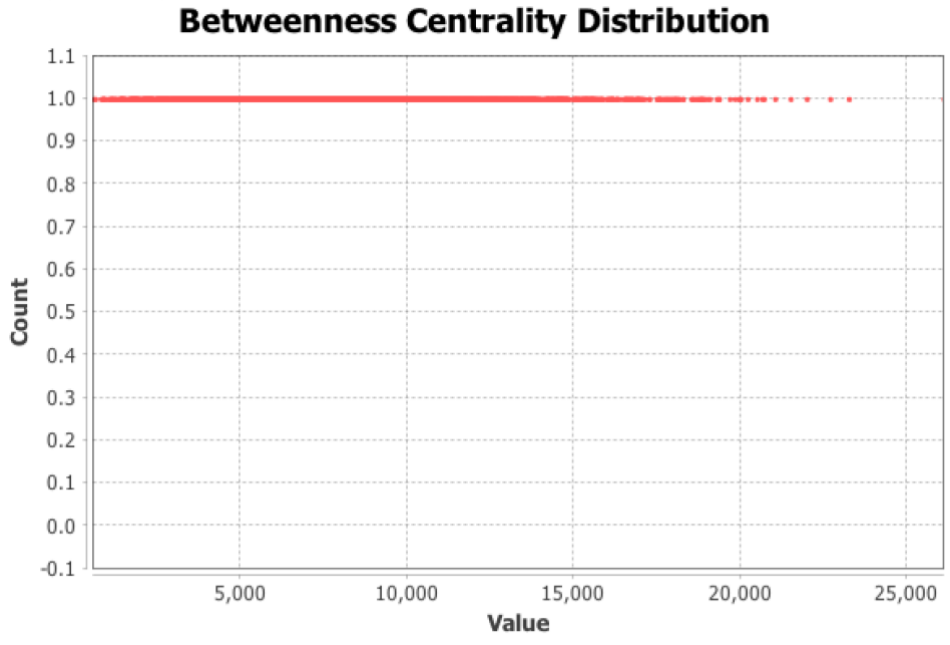
    Sum change: 0.0917076847283477

  -Betweeness Centrality

    Diameter: 5

    Radius: 4

    Average Path Length: 3.2657539683117265



Network Analysis

Degree Distribution:

To classify networks into different categories, the average degree distribution must be known. The degree of a node is equal to the number of edges that are attached to it. Graph1 has an average degree distribution of 52.052 whereas Graph2 has an average degree distribution of 9.65 this is because Graph2 has a significantly smaller amount of edges at 61965 compared to Graph1 which has 167112. This means on average there will be less edges per node in Graph2.

Clustering Coefficient:

The average clustering coefficient of Graph1 is 0.781. This value defines the proportion of a node's (character's) neighbours which are connected by an edge. Graph2 has a significantly lower value at 0.003 since it has less edges than Graph1 and therefore will have a smaller proportion of neighbours per node connected by an edge.

Modularity:

This is a measure of the structure of networks. Its designed to measure the strength of division of a network into groups/communities (tightly coupled nodes). The modularity of Graph2 is much weaker compared to Graph1 as due to Graph2 being a random graph it does not allow for visible division of communities. Instead of divided colours as shown in Graph1, Graph2 shows colours scattered all over the graph without a visible structure of grouped nodes of the same colour.

Centrality:

Centrality is about finding the most important nodes in a network. This can be used to identify the most influential people in a social network.
Graph1 shows the distinct division of communities. Community detection can be measured based on betweeness centrality (measure of centrality in a graph based on the shortest paths). Graph1 has an average path length of about 2.64, which is shorter than that of Graph2 with value of about 3.27, therefore it would be faster to send information through Graph1 compared to Graph2 on average. The betweeness centrality per central node in Graph2 will never be as high as that of Graph1 since Graph2 does not have highly influential nodes like Graph1.

The distinct separation of communities is shown through different colours in Graph1/ Graph1.1. Graph1.1 briefly shows the labels of the Marvel Comic book characters and allows us to see multiple comic book franchises owned by Marvel as communities, for example; the blue community represents the Xmen Franchise, yellow represents the Spiderman franchise,

green represents the Avengers franchise and orange represents the Fantastic Four franchise etc.

We can see from Graph1 that the green community is more spread out. This is due to the fact that most characters in this community have their own comic book franchises. They are also introduced in multiple comic book franchises other than their own. This is the reason for overlapping communities (when a node is a part of more than 1 community) and the scattering of certain members in these communities across Graph1. However, the blue community is less scattered and more clustered towards the bottom left of the graph, this is due to the fact that Marvel tends to keep this franchise (the x-men) separate to their other franchises except a few exceptional characters.

The size of the nodes on both graphs represents the degree of the node. Therefore, the bigger the node size, the higher the degree and from this we can determine the central nodes in the network. From Graph1.1 we can see characters Captain America and Spiderman have the largest node sizes followed by; Iron Man, members of the Fantastic Four and Wolverine etc., meaning they are the most influential characters in the network. Captain America (the biggest green node in the top right of Graph1/1.1) is the most central node and therefore is the most important node in the network. This node could help disseminate information in the network faster and could help protect the network from breaking. The removal of these large central nodes would have a large impact on the network and their communities.

However, the same amount of information cannot be analysed from Graph2 as most nodes are of similar size, thereby increasing difficulty to determine central nodes. Removal of certain larger nodes in Graph2 will have a less off an impact than in Graph1. There are also no visible communities that can be analysed due to the randomness of the graph. Any communities that were detected are scattered throughout the graph in random patterns making the graph harder to read.