

# Identifying Important Characters and Communities in the Marvel Comics Network

Kelvin Tiongson  
kelvinjtiongson@lewisu.edu  
DATA-51000-001, Summer 2020  
Data Mining and Analytics  
Lewis University

## I. INTRODUCTION

The goal of this analysis was to identify important Marvel comics characters and communities. The analysis takes a look at a network of Marvel comics characters. The dataset that was used in the analysis was downloaded from an open network data repository [1]. We will be able to discover which characters are closely connected and who is important within the groups.

The future sections of this report describe the dataset, methodology, results along with a discussion, and a conclusion. Section II contains a description of the dataset. Section II contains information about the nodes, edges, and general statistics about the dataset. The methodology for analysis is presented in section III. For the network analysis, I implemented different measures and methods such as closeness centrality, harmonic closeness centrality, betweenness centrality, eigenvector centrality, PageRank, and HITS in order to understand which nodes are important and identify existing communities. In section IV, the reports and results of the analysis are discussed. Finally, section V provides conclusions about the important characters and the communities found within the network.

## II. DATA DESCRIPTION

Table I, Table II, and Table III describe the dataset used in the analysis. Table I describes the nodes and Table II describes the edges. Table III describes the count of nodes, count of edges, the average clustering coefficient, and the average path length of the network. The goal of the analysis was to identify important, well connected superheroes and superhero. From Table I, the nodes had an ID and a label. The label was the name of the Marvel character. None of the attributes used were normalized when creating the network analysis. From Table II, the edges had a source, target, and type. The source was the starting location of the edge and the target was the ending location of the edge. The type was the type of connection of the edge. Table III shows that the original dataset had over ten thousand nodes and over one hundred and seventy-eight thousand edges. From Table III, the original network had a low tendency for clustering with an average clustering coefficient of 0.53. The average path length is not very close to the  $\ln(\text{nodes})$  which is approximately 9.256. Therefore, the original dataset would not have been considered of being a small world. Figure 1 represents the original network.

TABLE I. NODES

Attribute	Type	Example Value	Description
ID	Nominal (string)	"10470"	Id of the Marvel character
Label	Nominal (string)	"BLACK PANTHER/T'CHAL"	Name of the Marvel character

TABLE II. EDGES

Attribute	Type	Example Value	Description
Source	Nominal (string)	"10470"	Start location of the edge
Target	Nominal (string)	"10613"	End location of the edge
Type	Nominal (string)	"Undirected"	Type of connection of the edge

TABLE III. GENERAL STATISTICS

Network Property	Value
Nodes	10,469
Edges	178,115
Average Clustering Coefficient	0.53
Average Path Length	2.889

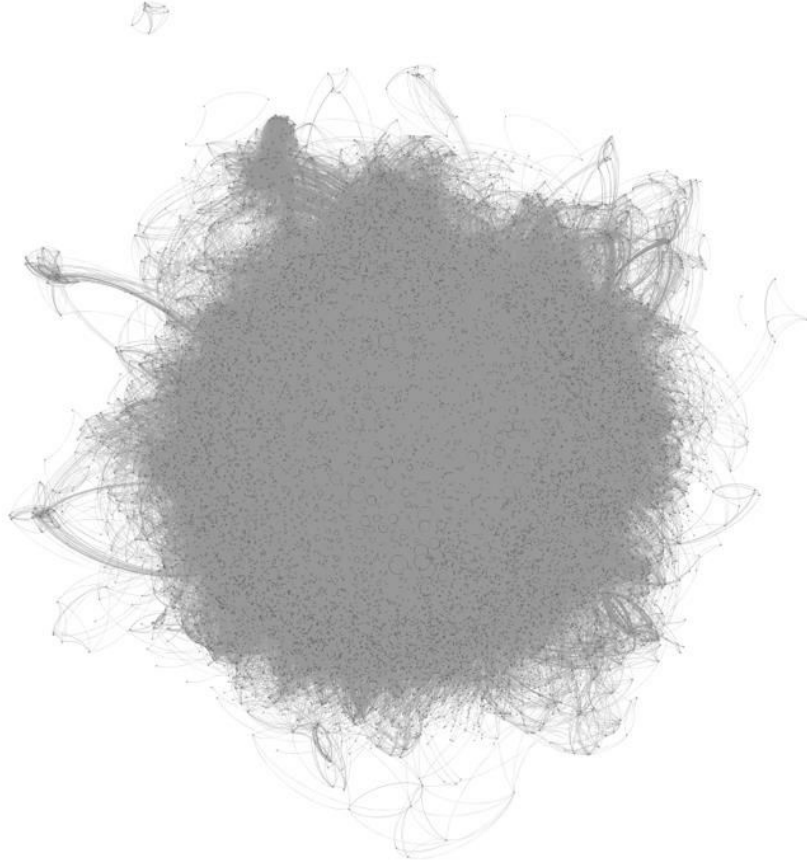


Fig. 1. Original network

### III. METHODOLOGY

The original network that was created showed every node and edge from the dataset. Figure 1 represents the original network. As shown in Figure 1, it is hard to identify certain features of the network. In order to get a better look, I filtered the network with a degree range. I set the minimum degree to be 848 so each node in the graph has at least 848 edges interacting with it. The new filtered network has 41 nodes and 663 edges. Figure 2 shows the filtered network. Table IV shows the general statistics of the new filtered network.

TABLE IV. GENERAL STATISTICS OF FILTERED NETWORK

Network Property	Value
Nodes	41
Edges	663
Average Clustering Coefficient	0.889
Average Path Length	1.191

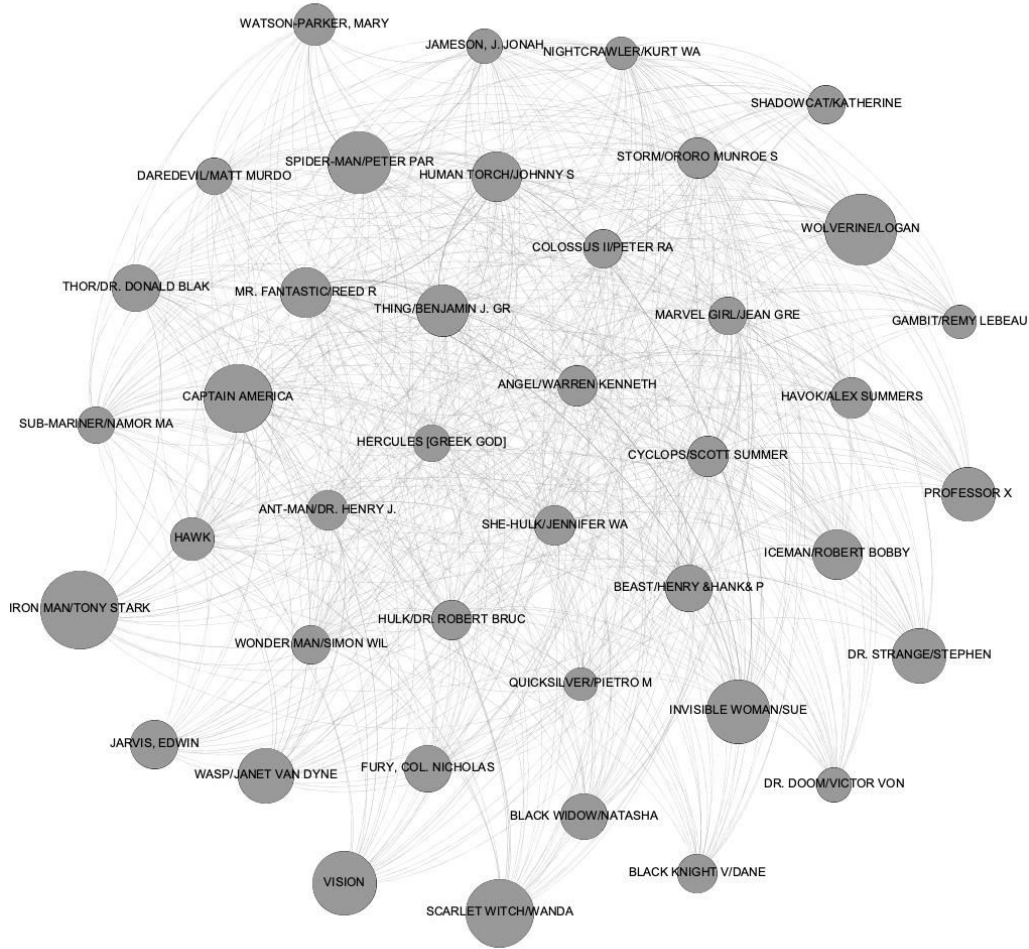


Fig. 2. Filtered Network

After filtering the original network, I was able to identify important nodes and communities. I used different measures and methods to find important nodes. The measures I looked at were the node's closeness centrality, harmonic closeness centrality, betweenness centrality, and eigenvector centrality. The methods used to supplement the determination of the important nodes were PageRank and HITS. In order find important communities I partitioned the network using modularity. After running the modularity method, I was able to discover four communities.

#### IV. RESULTS AND DISCUSSION

From Table IV, the network's average path length is 1.191 and the average clustering coefficient of the network is 0.889. From Table III, I declared that the original network did not have the small world property. However, Table IV shows that the filtered network has a high tendency for clustering. The filtered network's average path length is much closer to the  $\ln(\text{nodes})$  which is approximately 3.714. This is much closer than the original network. Since there is a high tendency for clustering and the average path length is close to the  $\ln(\text{nodes})$ , I would say that this filtered network satisfies being a small world.

Table V shows the different measures used in identifying important nodes. Table VI shows the methods used. From both tables, you can see that the top nineteen characters all have the same level of importance. For people who do not follow the comics, most of the characters in Table V and Table VI could have been seen in movies or tv series. For example, from the Avengers movie series, one would be able to identify Spider-Man, Hawkeye, Captain America, Hulk, Hawkeye, Thor, Ant-Man, and Iron Man [2]. Jean Grey, Storm, Cyclops, Beast, and Nightcrawler could have been seen in the X-Men movies [3]. Mr. Fantastic, Thing, Human Torch are in the Fantastic Four movies [4]. Daredevil has been in a movie and is a popular tv series within Netflix [5]. So, it is certain that these Marvel comic characters are the most well-known, important ones. Although, it was interesting to see that some characters such as Iron Man and Wolverine are not among the top important nodes. Table V and Table VI shows that there are characters that were surprising to find important. Some of the ones that stood out the most in terms of their importance were Sub-Mariner, Hercules, and Jameson. After researching the surprisingly, important characters, I found that most of them appeared alongside some of the important characters. Sub-Mariner appeared alongside characters such as Captain America, Hulk, Human Torch, and Iron Man [6]. Hercules appeared alongside the Avengers [7]. Jameson was not a fighting character, but he is important because he was the owner of a newspaper company whom always seemed to report on Spider-Man within the comics [8].

TABLE V. TOP 25 NODES AND THEIR CENTRALITY MEASURES

Name	Closeness Centrality	Harmonic Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
Sub-Mariner	1	1	0.008965	1
Hulk	1	1	0.008965	1
Spider-Man	1	1	0.008965	1
Hawkeye	1	1	0.008965	1
Captain America	1	1	0.008965	1
She-Hulk	1	1	0.008965	1
Jean Grey	1	1	0.008965	1
Storm	1	1	0.008965	1
Mr. Fantastic	1	1	0.008965	1
Thing	1	1	0.008965	1
Beast	1	1	0.008965	1
Cyclops	1	1	0.008965	1
Thor	1	1	0.008965	1
Human Torch	1	1	0.008965	1
Colossus	1	1	0.008965	1
Ant-Man	1	1	0.008965	1
Hercules	1	1	0.008965	1
Daredevil	1	1	0.008965	1
Nightcrawler	1	1	0.008965	1
Quicksilver	0.97561	0.9875	0.00768	0.983136
Wonder Man	0.97561	0.9875	0.00768	0.983136
Angel	0.97561	0.9875	0.00768	0.983136
Jameson, J. Jonah	0.97561	0.9875	0.007907	0.983136
Iron Man	0.701754	0.7875	0	0.674908
Wasp	0.701754	0.7875	0	0.674908

TABLE VI. TOP 25 NODES AND THEIR PAGE RANK, AUTHORITY, AND HUB SCORES

Name	Page Rank	Authority	Hub
Sub-Mariner	0.029547	0.179841	0.17984
Hulk	0.029547	0.179841	0.17984
Spider-Man	0.029547	0.179841	0.17984
Hawk	0.029547	0.179841	0.17984
Captain America	0.029547	0.179841	0.17984
She-Hulk	0.029547	0.179841	0.17984
Marvel Girl	0.029547	0.179841	0.17984
Storm	0.029547	0.179841	0.17984
Mr. Fantastic	0.029547	0.179841	0.17984
Thing	0.029547	0.179841	0.17984
Beast	0.029547	0.179841	0.17984
Cyclops	0.029547	0.179841	0.17984
Thor	0.029547	0.179841	0.17984
Human Torch	0.029547	0.179841	0.17984
Colossus II	0.029547	0.179841	0.17984
Ant-Man	0.029547	0.179841	0.17984
Hercules	0.029547	0.179841	0.17984
Daredevil	0.029547	0.179841	0.17984
Nightcrawler	0.029547	0.179841	0.17984
Quicksilver	0.028856	0.176815	0.176815
Wonder Man	0.028856	0.176815	0.176815
Angel	0.028856	0.176815	0.176815
Jameson, J. Jonah	0.02887	0.176517	0.007907
Iron Man	0.018094	0.12141	0.12141
Wasp	0.018094	0.12141	0.12141

In order to find important communities, I used the Modularity method. As a result, there were four discovered communities. This is shown in Figure 3. From Figure 3, you can see that the communities are separated by color. The purple community represents characters that are all in the Spider-Man series. The orange community represents the characters that are connected with the Avengers. One unique finding in this group is Daredevil. From what is displayed in his movie and television series, you would not think about associating Daredevil within the Avengers community. However, he is connected to the Avengers because he is a crimefighter in New York alongside Captain America and Daredevil considers Captain America and Black Panther as one of his closest confidantes [9]. The pink community are all characters connected to the Fantastic Four. Invisible Woman, Thing, Human Torch, and Mr. Fantastic are the main four superheroes within the group with Dr. Doom being one of their main nemesis. The green community are all characters that are associated with X-Men group. From Table VII, we can see that the biggest community is the Avengers. Second, would be the X-Men and third being the Fantastic Four. The Spider-Man community is the smallest with only three Marvel comic characters. Figure 4 shows the frequencies of each community. So, from Figure 3 and Figure 4 it is evident that the most connected community within the Marvel social network are the Avengers.

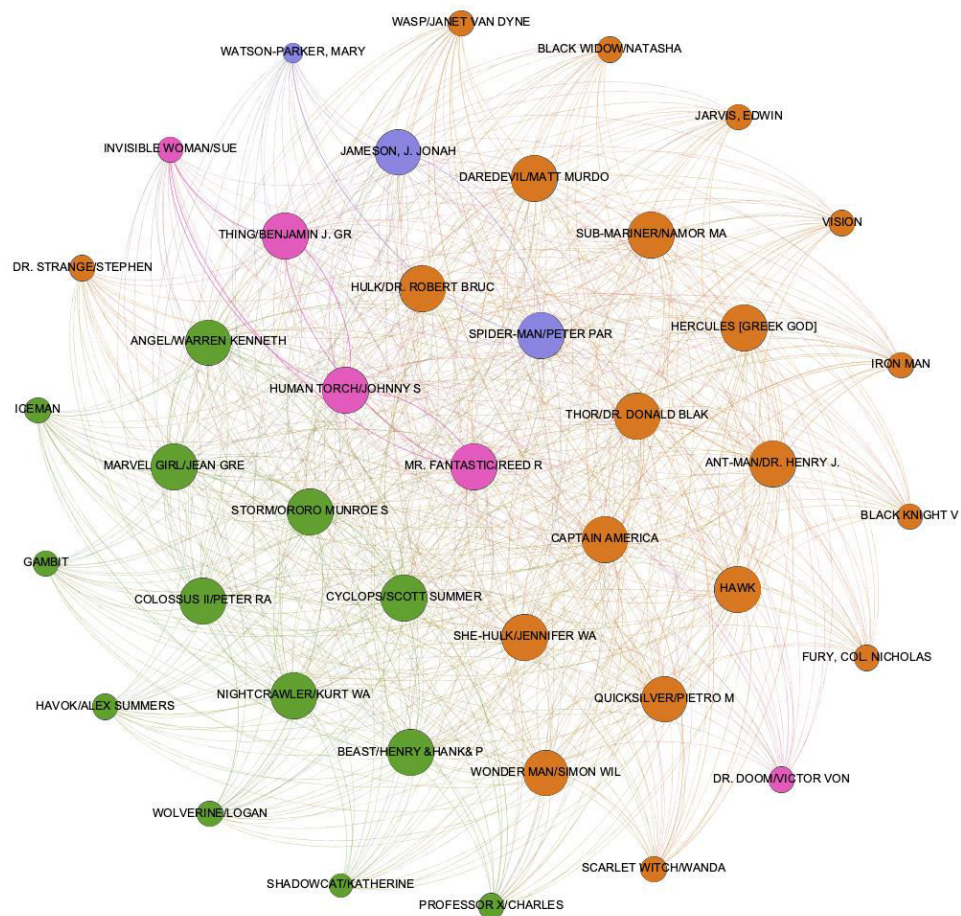


Fig. 3. Filtered Network and Their Respective Communities

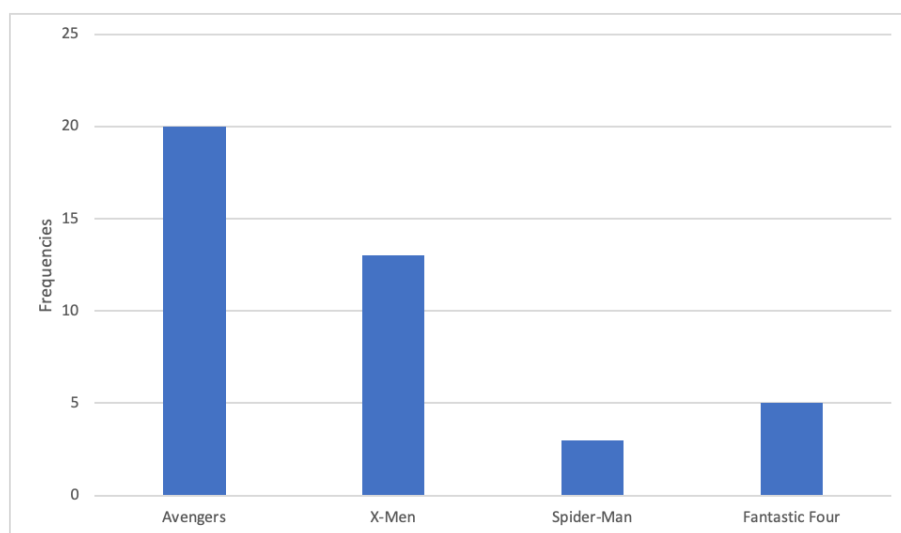


Fig. 4. Frequency chart of the communities

## V. CONCLUSIONS

I graphed the network of the Marvel comics characters. In order to get a better look at more important nodes I filtered the network to the nodes that have a degree greater than or equal to 848. This gave me a filtered network of 41 nodes and 663 edges. I implemented different measures and methods to identify important nodes and communities within the superheroes. In order to find important nodes, the measures used the node's closeness centrality, harmonic closeness centrality, betweenness centrality, and eigenvector centrality. The methods used were PageRank and HITS. In order to find communities within the network, I implemented the modularity method. There were four communities that were discovered. The four communities were all a part of a certain group or team within the Marvel universe. Among the important nodes, there were nineteen nodes that had the same measures, rank, and scores. An interesting finding of the important nodes were the characters Sub-Mariner, Hercules, and Jonah Jameson. Also, Iron Man and Wolverine were not as important as some of their comrades in their respective communities. The four communities discovered were the X-Men, Fantastic Four, Avengers, and Spider-Man. Then I graphed the frequencies of members within each group. The Avengers had the most characters in their group, followed by the X-Men, then the Fantastic Four, and last being the Spider-Man group. So, there are nineteen Marvel comics characters that are the most important within the Marvel universe and there are four major communities within the network with the Avengers having the most members among those four communities.

## REFERENCES

- [1] <https://github.com/gephi/gephi/wiki/datasets>
- [2] <https://www.marvel.com/teams-and-groups/avengers/on-screen>
- [3] [https://www.imdb.com/title/tt0290334/?ref\\_=ttls\\_li\\_tt](https://www.imdb.com/title/tt0290334/?ref_=ttls_li_tt)
- [4] [https://www.imdb.com/title/tt0486576/?ref\\_=tt\\_sims\\_tt](https://www.imdb.com/title/tt0486576/?ref_=tt_sims_tt)
- [5] <https://www.netflix.com/title/80018294>
- [6] <https://www.marvel.com/characters/sub-mariner>
- [7] [https://marvel.fandom.com/wiki/Hercules\\_\(Earth-616\)](https://marvel.fandom.com/wiki/Hercules_(Earth-616))
- [8] <https://www.marvel.com/characters/j-jonah-jameson/in-comics>
- [9] <https://www.marvel.com/characters/daredevil-matthew-murdock/in-comics>