BANA 277: Customer and Social Analytics

Project Report

on

Marvel Hero Social Network Analysis

Submitted by: Group 2B

Abhinandan Verma
Amanda Wei
CK Anand Prakash
Deyn Li
Kingsley Liu
Urvi Vaidya

MARVEL
25 .com
KIRKMAN
KUHN
MARVEL
TEAM-UP
DIRECT EDITION
RATED A
$2.99 US $3.75 CAN

Table of Contents

INTRODUCTION

The Marvel Comic Universe is a vast and complex world with a large number of characters, each with their own unique backstory and relationships with other characters. Network analysis is a powerful tool that can be used to understand the relationships between these characters and uncover patterns and insights that may not be immediately apparent. In this report, we will analyze the Marvel dataset, which contains information about characters, comics, and relationships between characters in the Marvel Universe.

Project Background:

The Marvel Comic Universe is a vast, interconnected fictional world that has captivated audiences for decades. Created by Stan Lee and Jack Kirby in the 1960s, the Marvel Universe is home to some of the most iconic and beloved characters in popular culture, including Spider-Man, Iron Man, Captain America, and the X-Men.

At its core, the Marvel Universe is a world of superheroes, where individuals with extraordinary abilities use their powers to fight for justice and defend the innocent. However, it is also a world of complex relationships and alliances, where heroes and villains alike must navigate a web of interconnected relationships in order to achieve their goals.

One of the defining features of the Marvel Universe is its intricate continuity. The stories and characters of the Marvel Universe have evolved over time, with new characters and plotlines being introduced and old ones being retired or reimagined. This continuity has allowed for a richly textured and layered world, with characters and events interweaving to create a complex and dynamic narrative tapestry.

The Marvel Cinematic Universe (MCU), which began with the release of Iron Man in 2008, has further expanded and popularized this fictional world, bringing the characters and stories of the Marvel Universe to a wider audience than ever before.

Given the complexity and richness of the Marvel Universe, there is a great deal of interest in understanding the social networks that exist within it. By analysing the relationships between characters, we can gain insights into the structure and dynamics of this universe, as well as explore the themes and motifs that underpin the stories it tells. This project aims to do just that, using graph theory and network analysis techniques to uncover the intricate social network of the Marvel Universe.

Project Objective:

The objective of this project is to analyze the social network of the Marvel Universe, using graph theory and network analysis techniques to uncover patterns and relationships between characters. Specifically, we aim to:

a. Identify the key players in the Marvel Universe, based on their centrality and influence within the network.

b. Analyze the relationships between characters, in order to identify sub-groups or communities within the Marvel Universe.

c. Explore the implications of our findings for our understanding of the Marvel Universe as a whole, as well as the broader themes and cultural significance of this fictional world.

By undertaking this analysis, we hope to shed new light on the intricate web of relationships that exists within the Marvel Universe, and deepen our understanding of the complex social dynamics that drive this fascinating fictional world.

Data Collection and Pre-processing

About Dataset:

The Marvel dataset, an extensive collection of data, is available on Kaggle and contains information on over 20,000 characters and 57,000 comics. The dataset comprises diverse character attributes, such as name, affiliation, and appearance in comics. Additionally, it provides valuable insights into different relationships among characters, such as co-appearances, alliances, and family ties. This remarkable dataset presents a unique opportunity to explore and analyze the intricate relationships between characters in the Marvel Universe using network analysis techniques.

1. *nodes.csv:* Contains two columns indicating the name and the type (comic, hero) of the nodes.
2. *edges.csv:* Contains two columns (hero, comic), indicating in which comics the heroes appear.
3. *hero-edge.csv:* Contains the network of heroes which appear together in the comics.

For our network analysis, we sourced our dataset from Kaggle's "The Marvel Universe Social Network" (*https://www.kaggle.com/datasets/csanhueza/the-marvel-universe-social-network?select=hero-network.csv*), utilizing the hero-edge.csv file. The file contains 559,666 rows of data organized into two columns, "hero1" and "hero2". Each row represents a pairing of two characters, capturing every occurrence of their appearance together in the comics. We leveraged this dataset to conduct our network analysis of the Marvel Universe, providing unique insights into the relationships among characters in this complex fictional universe.

Data Interpretation:

Our analysis involved using Gephi, a powerful software tool for network analysis, to create visualizations and calculate statistical metrics. We began by inputting our nodes into Gephi as an adjacency matrix, enabling us to visualize all the nodes (heroes) and the links between them (edges) as a network graph. As characters appear together in comics, an edge is created, and the frequency of these co-appearances determines the thickness or strength of the connection between the heroes.

To create the network graph, we conducted statistical calculations to determine the diameter of the network, the average path, degree, centrality, and other relevant metrics. These calculations allowed us to generate visualizations of our network. We were able to change the colour, size, and labels of the nodes and edges based on different metrics. Furthermore, we could group or cluster nodes to visualize communities within the network, making it easier to highlight different relationships and patterns.

Overall, using Gephi provided us with an interactive and dynamic approach to visualizing the complex relationships among characters in the Marvel Universe. The statistical calculations we performed enabled us to obtain valuable insights into the structure and nature of the network, ultimately leading to a more comprehensive understanding of the relationships between characters.

Network Construction and Visualizations:

To enhance our comprehension of the network structure and operations of the project, we have employed several methodological approaches for conducting a thorough network analysis. The aim of this analysis is to gain a deeper understanding of the relationships, interactions, and communication patterns among the different nodes and actors within the network.

Network analysis is a methodology used to study the relationships and interactions among nodes in a network. It involves the use of mathematical and statistical techniques to analyze the network structure and identify important features and patterns. Two key concepts in network analysis are centralities and modularity.

Key Concepts and Definitions

Centralities refer to measures that assess the relative importance of nodes in a network. There are several types of centralities, including degree centrality, betweenness centrality, and eigenvector centrality.

1. *Degree Centrality:*

   - Measures the number of connections (or edges) a node has in a network.
   - Nodes with a high degree centrality are considered to be more important or influential than those with a low degree centrality.
   - Useful for identifying hubs or connectors within a network.

2. *Betweenness Centrality:*

   - Measures the extent to which a node lies on the shortest path between other nodes in a network.
   - Nodes with a high betweenness centrality are considered to be important for maintaining communication between different parts of the network.
   - Useful for identifying nodes that act as brokers or bridges between different groups within a network.

3. *Eigenvector Centrality:*

   - Measures the importance of a node based on the importance of its neighbours.

- Nodes that are connected to other nodes with high centrality will have a higher eigenvector centrality than nodes that are connected to nodes with low centrality.
- Useful for identifying nodes that are influential because they are connected to other influential nodes.

4. *Number of Triangles:*

- Measures the number of triangles that a node is part of in a network.
- A triangle is a set of three nodes that are connected to each other.
- Nodes that are part of many triangles are considered to be important for maintaining local cohesion and information flow within a network.

5. *Eccentricity Centrality:*

- Measures the distance between a node and all other nodes in a network.
- Nodes with a low eccentricity centrality are considered to be important for maintaining global cohesion and information flow within a network. These nodes are often referred to as "central nodes" or "hubs" in the network.

Modularity, on the other hand, refers to the degree to which a network can be divided into distinct subgroups, or modules.

A module is a group of nodes that are more densely connected to each other than they are to nodes in other modules. Modularity is a measure that quantifies the extent to which a network is divided into modules.

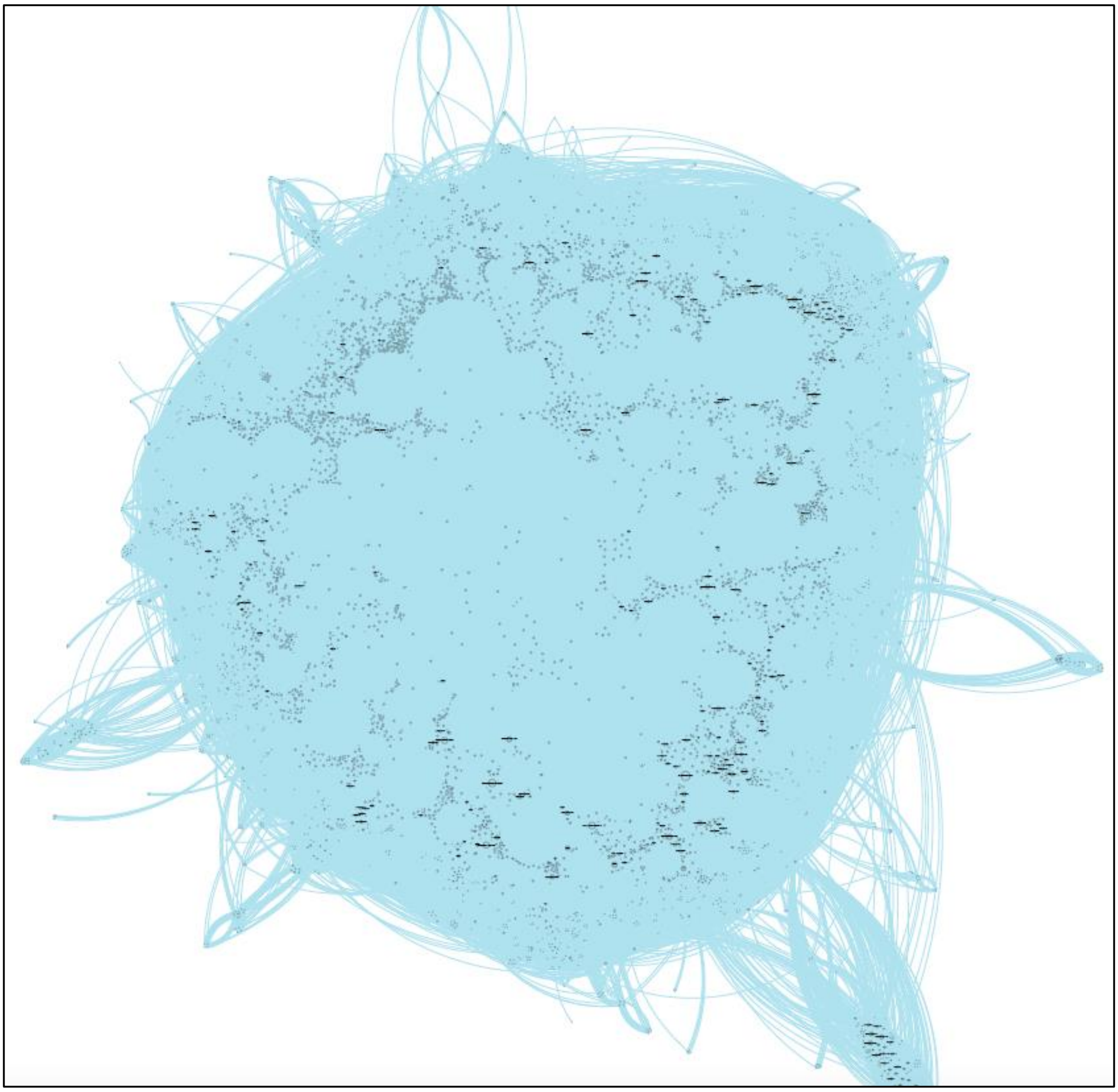
Networks with high modularity are characterized by a clear division of nodes into distinct groups, while networks with low modularity are more homogenous and less modular.

Network Visualizations and Analysis

1. Baseline Network:

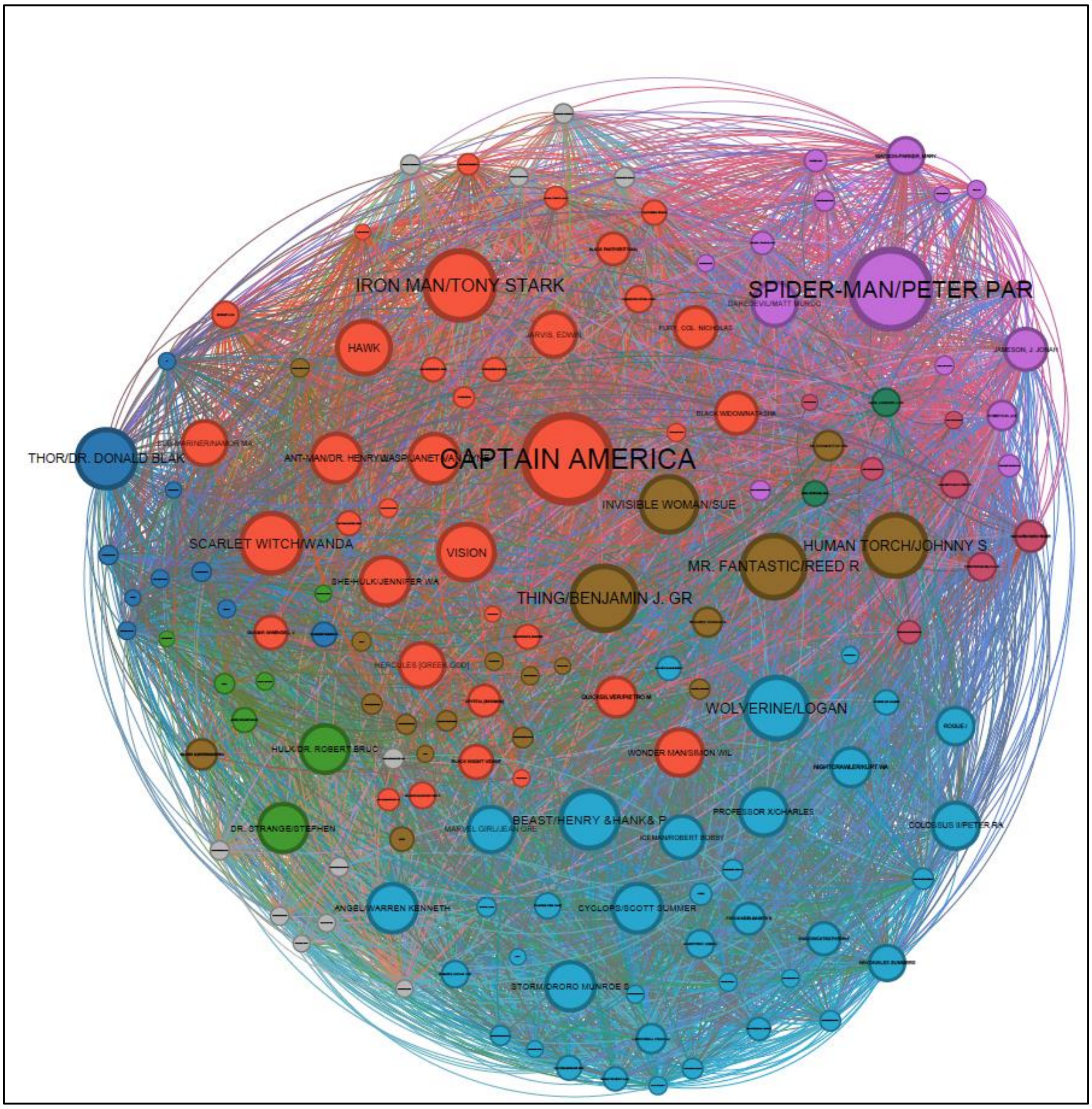
When visualized in Gephi, the initial network generated from the MCU dataset reveals a dense undirected network that includes all hero-hero node connections. Each node in this network represents a superhero, and each edge signifies when two superheroes appeared together in a comic book.

Despite its comprehensive nature, this network alone does not provide much information about the dependencies of heroes on one another, nor does it reveal the specific characteristics of individual heroes. Nonetheless, this initial network serves as a crucial foundation for further exploration and analysis, enabling us to uncover the intricate relationships and connections between characters in the Marvel Universe.



2. Degree Centrality with Modularity Network:

Our exploration of the Hero-Hero relationships in the Marvel Cinematic Universe (MCU) involves the use of an undirected network graph. We employ Degree centrality in conjunction with a modularity class filter set to 340 degrees to analyze the complex web of relationships among the characters. Upon careful examination, we observe the emergence of multiple clusters within the graph, with each representing small communities of interconnected heroes. These clusters serve to highlight the intricate and nuanced relationships among the characters within the MCU. Our analysis of these clusters, along with other statistical measures, provides valuable insights into the underlying patterns and dynamics that govern the interactions among the characters in this vast and dynamic universe. By utilizing this methodology, we are able to gain a comprehensive understanding of the complex interplay between heroes and their various affiliations within the MCU.



Key clusters include

Orange Cluster	Avengers
Blue Cluster	X-Men
Brown Cluster	Fantastic 4
Purple Cluster	Spiderman
Green Cluster	Dr. Strange/Hulk

Upon close examination of the network graph, some notable observations have emerged. Of particular interest are the distinct clusters formed around Spider-Man and Doctor Strange. Spider-Man's cluster prominently features Daredevil and Iron Fist, indicating a strong connection between Spider-Man and the Defenders group, which includes Daredevil, Luke Cage, Iron Fist, Jessica Jones, and Punisher.

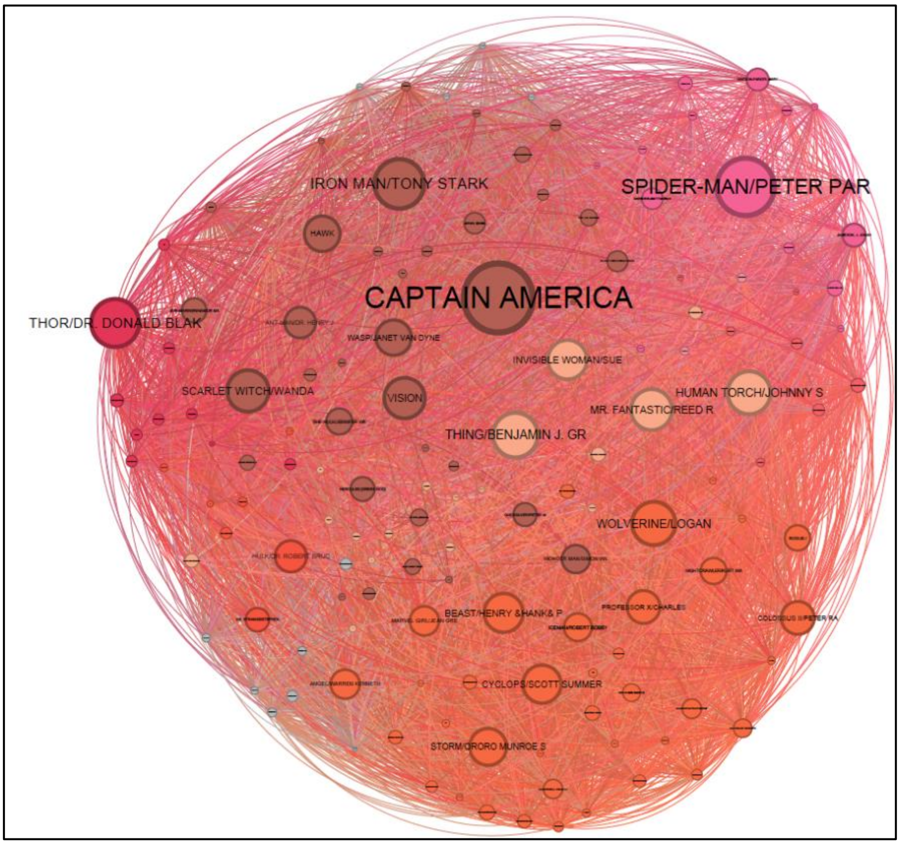
Similarly, Doctor Stranger's cluster highlights his association with Hulk and Wong, indicating that the Mystical Realm and Hulk share more commonalities than Hulk does with the Avengers. This finding is surprising and emphasizes the complexity of relationships between characters in the Marvel Comic Universe and Marvel Cinematic Universe.

Furthermore, the analysis reveals that Hulk and She-Hulk belong to different cluster classes, with She-Hulk being part of the Avengers cluster. This further emphasizes the intricate relationships between the characters in the Marvel Universe and provides valuable insights into the structure of the Hero-Hero network.

3. Weighted Degree Centrality with Modularity Network:

Through the use of Weighted Degree Centrality, we were able to identify the Heroes in the graph who have appeared most frequently in the comics. Additionally, by utilizing modularity, we gained insight into the types of comics in which these heroes were featured.

For instance, it is challenging to imagine any Avengers comic without Captain America or Iron Man, as they have a high degree of centrality in the network. However, there are comics where notable characters like Hawkeye or Black Widow do not appear as frequently. By analysing the weighted degree centrality, we were able to identify the variance in character appearances and the critical players in the Marvel Cinematic Universe.

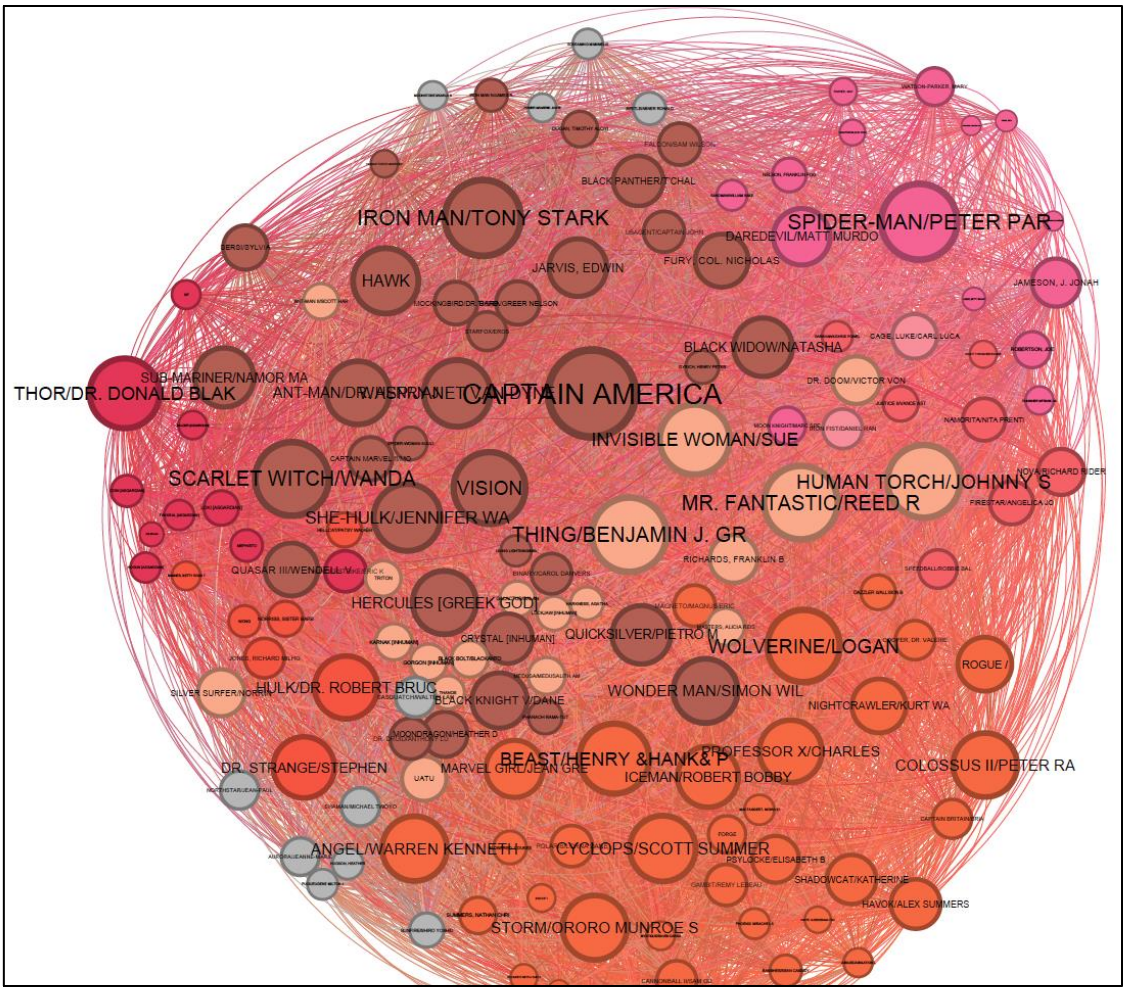


4. Eigen-Vector Centrality with Modularity Network

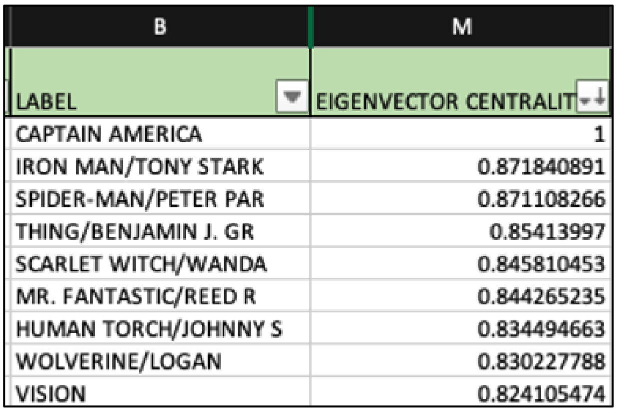
Eigenvector centrality is a crucial metric in network analysis, as it helps us understand which superhero is connected to other highly connected superheroes. Our analysis reveals that Captain America has the highest eigenvector centrality score, indicating that he is connected to a large number of highly connected superheroes in the Marvel Cinematic Universe (MCU) network.

Captain America's score can be attributed to his longevity and extensive history in the Marvel Universe. As a founding member of the Avengers and a prominent figure in many key storylines, Captain America has had ample opportunities to form strong connections with other heroes. These connections may have been forged through shared experiences, common enemies, or simply the fact that Captain America is seen as a respected leader among superheroes.

Overall, the eigenvector centrality analysis provides valuable insights into the structure of the MCU network and the key players within it. It highlights the importance of Captain America and his extensive connections, as well as the interconnectivity of other highly connected superheroes.



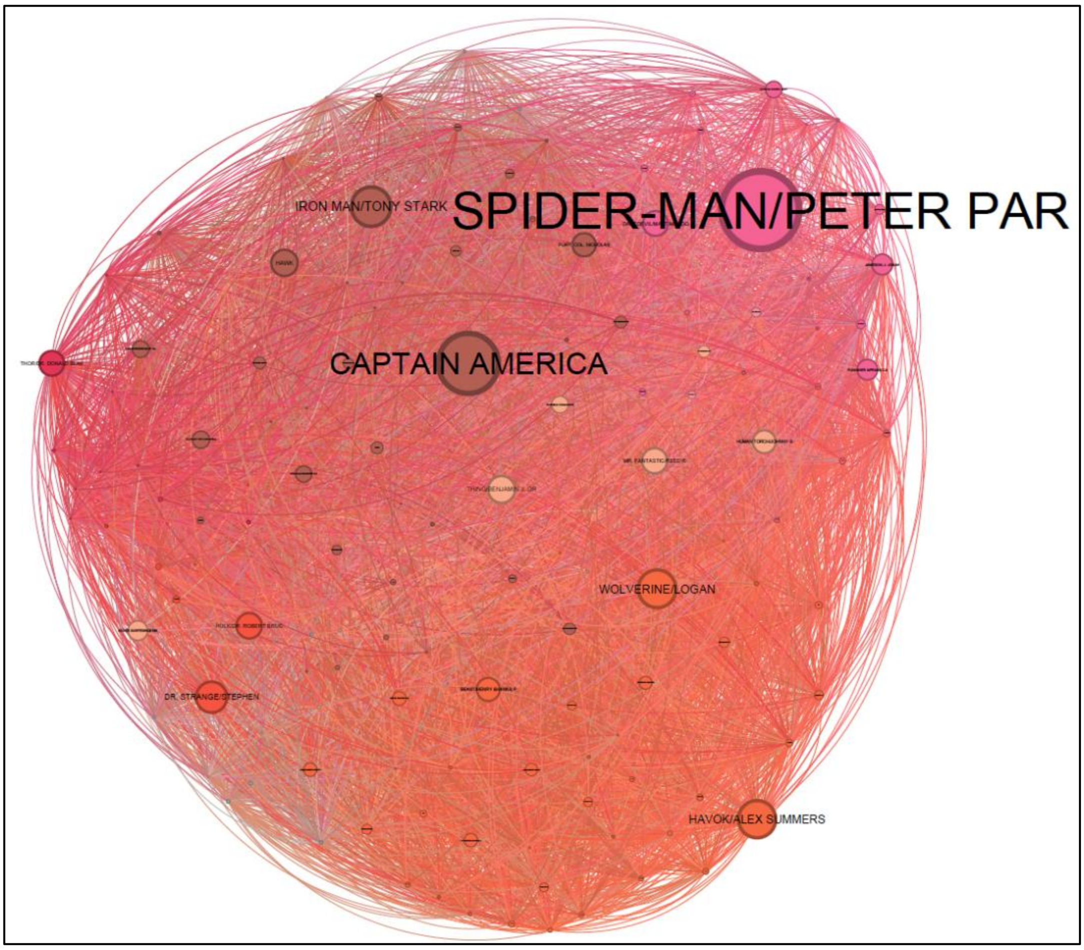
An illustrative example from the comics reveals that in the third issue of Spider-Man, the most prominent heroes that assisted Spidey are Ironman, Captain America, Wanda, Logan, and Daredevil. The top-10 superheroes with highest eigen-vector centralities are:



B	M
LABEL	EIGENVECTOR CENTRALIT
CAPTAIN AMERICA	1
IRON MAN/TONY STARK	0.871840891
SPIDER-MAN/PETER PAR	0.871108266
THING/BENJAMIN J. GR	0.85413997
SCARLET WITCH/WANDA	0.845810453
MR. FANTASTIC/REED R	0.844265235
HUMAN TORCH/JOHNNY S	0.834494663
WOLVERINE/LOGAN	0.830227788
VISION	0.824105474

5. Betweenness Centrality with Modularity Network:

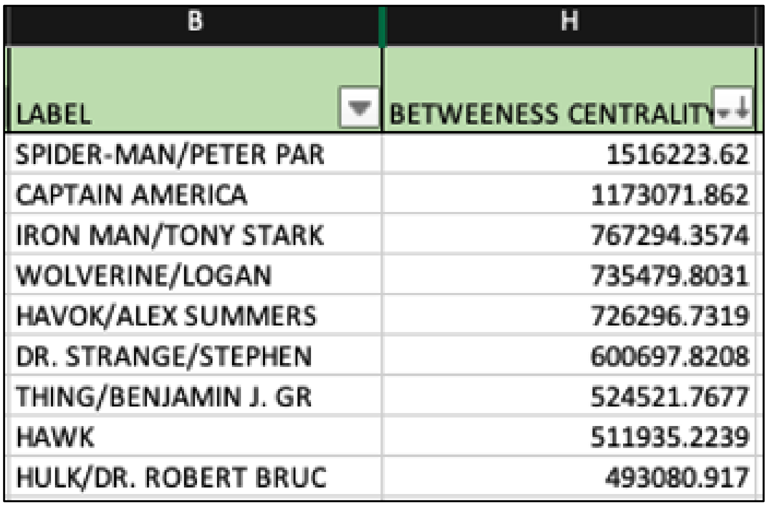
Betweenness centrality helps us understand which superheroes can act as a bridge in connecting superheroes with each other. For instance, a superhero with high betweenness centrality may be critical in preventing the network from fragmenting into disconnected subgroups. They may be the ones who bring together different superheroes from disparate groups, thereby enhancing collaboration and facilitating the exchange of information and resources. On the other hand, a superhero with low betweenness centrality may be relatively isolated from the rest of the network, limiting their ability to influence the flow of information or connect different groups of superheroes.



Based on our analysis of the network graph, we can infer that Spider-Man is the most central superhero in the Marvel Cinematic Universe (MCU) with connections to almost all other superheroes. This observation is supported by the high Betweenness Centrality score of Spider-Man, indicating his ability to act as a bridge and facilitate connections between other superheroes.

A demonstrative example that can be examined is the potential introduction or connection between Deadpool and Ironman in the context of the Marvel Cinematic Universe. He throws out some names like Doctor X, Wolverine, and Storm, but it turns out the right answer is Doctor Strange, as they were in a comic book together (#4, baby!). Deadpool chats up Doctor Strange, who then goes and talks to Tony Stark, and just like that, it's done! Doctor Strange is the man! Deadpool → Dr. Strange (bridge) → Iron Man

The top-10 most connected superheroes in Marvel Comic Universe are:



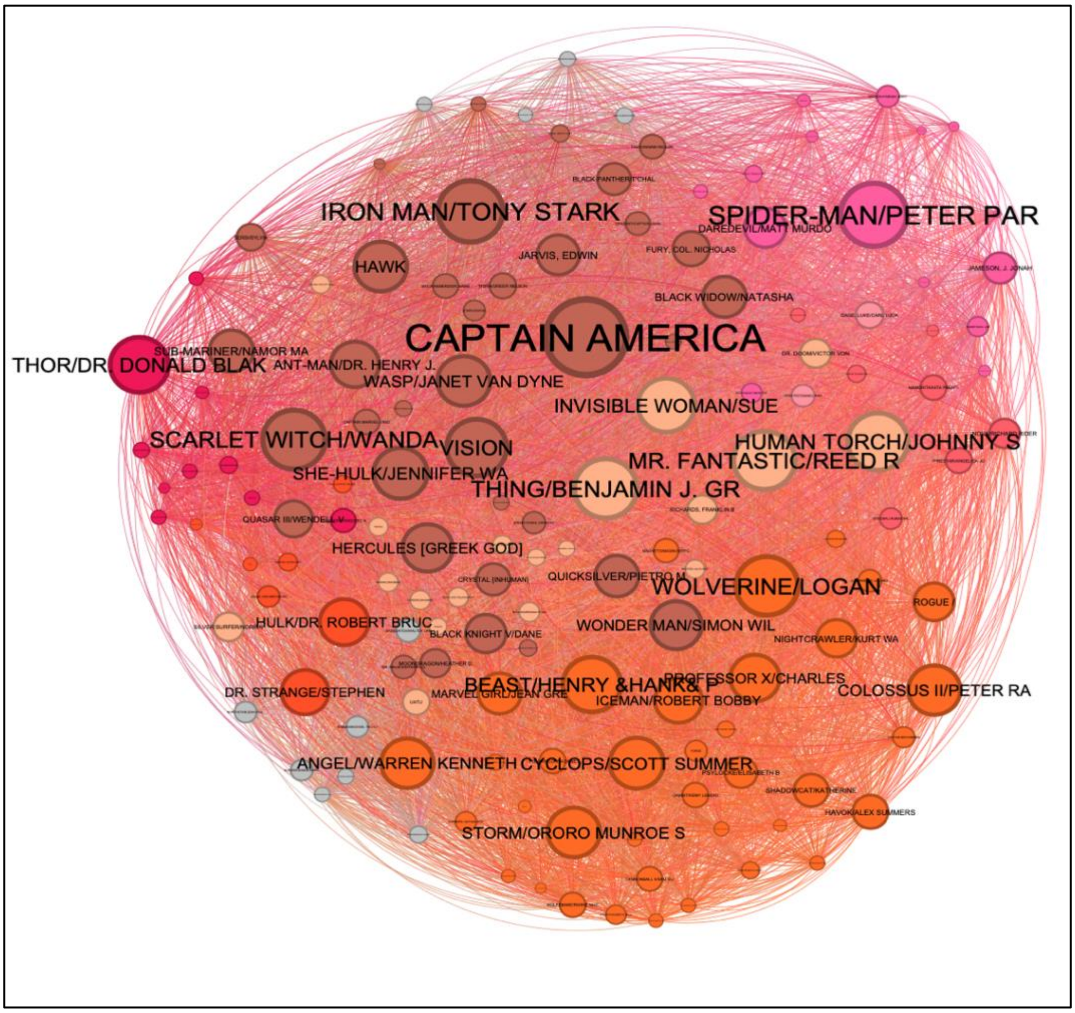
B	H
LABEL	BETWEENESS CENTRALITY
SPIDER-MAN/PETER PAR	1516223.62
CAPTAIN AMERICA	1173071.862
IRON MAN/TONY STARK	767294.3574
WOLVERINE/LOGAN	735479.8031
HAVOK/ALEX SUMMERS	726296.7319
DR. STRANGE/STEPHEN	600697.8208
THING/BENJAMIN J. GR	524521.7677
HAWK	511935.2239
HULK/DR. ROBERT BRUC	493080.917

6. Number of Triangle Centrality with Modularity

The analysis conducted on the number of triangles with modularity class is considered to elaborate further on understanding the communities and the cohesiveness of those communities in the Marvel heroes' network. Below are the top five heroes with many triangles in the Marvel heroes' network.

Name	No. of triangle
Captain America	88849
Spiderman	73690
Ironman	71764
Thing	69252
Scarlet Witch	69028

According to the graph and statistics, these five heroes have the top five numbers of triangles indicating their strong presence and tight connections to groups and communities. Furthermore, the higher number of triangles can identify as essential connectors or hubs in the network for different groups and communities. The number of triangles is also an excellent indicator of transitivity. The higher the number of triangles one node has, the higher the likelihood of nodes connecting to other nodes if they have standard connections.

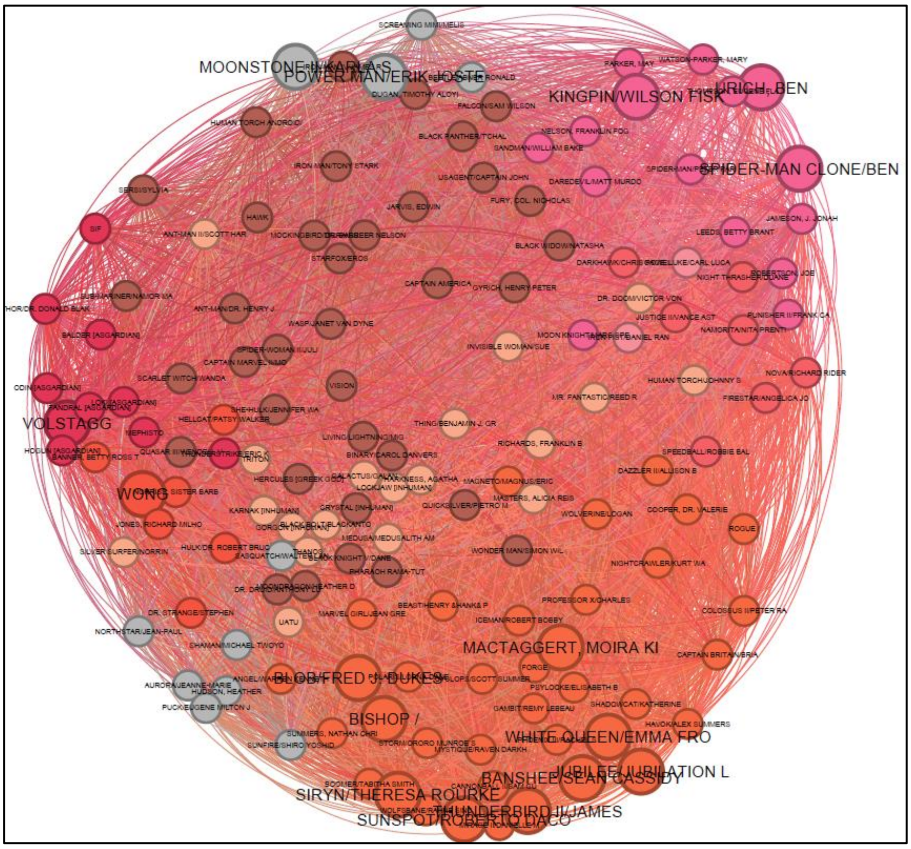


The highest number of triangles that Captain America has suggests that he is often the central role and the most crucial hero in the subgroups and the Marvel heroes' network. It is trivial that the number of triangles of Captain America is consistent with his role in the movies and comics. He is the leader of the Avengers and shows a prominent presence in other groups or communities, such as The Illuminati and Secret Avengers.

7. Eccentric centrality with Modularity

Eccentricity is a vital measurement in network analysis for us to analyze the distance from a given hero to the farthest from a given hero in the Marvel hero network. Regarding the Marvel hero network analysis, the highest number of eccentricities is 5, and the lowest is 1. The peripheral nodes(heroes) are nodes with high eccentricity far away from other nodes in the network. The central nodes are typically nodes with low eccentricity.

Eccentricity:5	Eccentricity:1
211	15



After examining the common centrality, such as degree centrality, betweenness centrality, and closeness centrality, we concluded that these 15 heroes with low eccentricities are considered peripheral nodes in the network and are lesser important or known heroes like Master of Vengeance, Oswald and Steel Spiders.

Result and Discussion

Network analysis plays a vital role in understanding the structure and complexity of the Marvel hero network. Based on our research, we can draw three key conclusions. Firstly, the network exhibits a small-world phenomenon, with a high clustering coefficient of 0.781, indicating solid local connections among Marvel heroes. The network diameter of 5 suggests that it is relatively easy to navigate the network, while the average path length of 2.638 highlights the efficiency of communication among heroes. (However, they might come across multiple universes to get to each other)

Secondly, the analysis reveals that famous heroes such as Spiderman, Captain America, and Ironman have the most links and are central to Marvel's profits. However, the network also features a long tail of nodes with few links, including lesser-known heroes. This insight highlights the potential for Marvel to explore new storylines and expand its character roster to reach new audiences.

Lastly, network analysis offers valuable insights into identifying essential characters, informing plot development, and shaping marketing strategies in the MCU. Furthermore, it can be applied to analyze audience engagement and identify key influencers, enabling targeted marketing and business strategies for the franchise.

Key Takeaways:

1. *Identification of Clusters*: The study identified clusters of superheroes that had strong connections with each other. For example, Spider-Man had a cluster with Daredevil and Iron Fist, while Doctor Strange had a cluster with Hulk and Wong.

2. *Weighted Degree Centrality*: This metric helped identify the superheroes that appeared in the most comics. For example, Captain America and Iron Man appeared in most Avengers comics, while Hawkeye and Black Widow were not present in some Avengers comics.

3. *Eigenvector Centrality*: This metric helped identify the superheroes who had strong connections with other highly connected superheroes. For example, Captain America had the largest connections in the MCU, indicating his extensive history and longevity in the comics.

4. *Betweenness Centrality*: This metric helped identify the superheroes who acted as a bridge in connecting other superheroes. For example, Spider-Man had a high betweenness centrality, indicating his ability to connect with many other superheroes.

5. *Visualization of Network*: The network visualization helped provide a visual representation of the connections between superheroes, allowing for a better understanding of the intricate relationships within the MCU.

Conclusion

In conclusion, the network analysis of the Marvel Cinematic Universe (MCU) dataset has provided valuable insights into the intricate relationships among the superheroes in the comic universe. By using different network metrics such as Degree centrality, Eigenvector centrality, and Betweenness centrality, we were able to identify the key superheroes and their connections to each other.

We found that some superheroes, such as Spider-Man and Captain America, have strong and numerous connections with other heroes, while others like Hulk have relatively fewer connections. Additionally, the analysis revealed that some superheroes belong to distinct clusters, suggesting their unique relationships with other heroes.

Overall, this study has demonstrated the power of network analysis in understanding complex systems such as the MCU comic universe. The insights gained from this analysis can be useful for understanding the characteristics of different superheroes and for predicting their interactions in future comics.

Future Scope

Based on the insights gained from this study, there are several potential areas for future research in the analysis of the Marvel Comics Universe network.

Firstly, a more in-depth examination of the clusters formed by the different superhero groups could provide additional insights into the dynamics of the Marvel Universe. By exploring the relationships between these groups, it may be possible to identify new patterns of interaction or uncover previously unknown connections between characters.

Secondly, it would be interesting to conduct a longitudinal analysis of the network, tracking changes in the relationships between characters over time. This could shed light on the evolution of the Marvel Universe and help identify key turning points or milestones in its development.

Lastly, the use of additional network metrics and visualization techniques could enhance the analysis further. For instance, the use of dynamic network visualizations could enable us to track changes in the network over time more effectively. Additionally, incorporating sentiment analysis techniques could provide insights into the nature of the relationships between characters and help to identify patterns of cooperation, conflict, or rivalry.

References and Citations

Citations:

1. Hanneman, R. A., & Riddle, M. (2005). Introduction to social network methods. University of California, Riverside, 1(1), 1-13.
2. Marvel Comics. (2021). Marvel Comics - Home of Spider-Man, The Avengers, and the X-Men. *https://www.marvel.com/comics*
3. Kaggle. (n.d.). Marvel Universe Social Network. Retrieved from *https://www.kaggle.com/csanhueza/the-marvel-universe-social-network*
4. Marvel GitHub source: *https://syntagmatic.github.io/marvel/*

References:

1. Jackson, M. O. (2008). Social and economic networks. Princeton University Press.
2. Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications (Vol. 8). Cambridge University Press.
3. Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). Analysing social networks. Sage Publications.
4. Freeman, L. C. (1978). Centrality in social networks conceptual clarification. Social networks, 1(3), 215-239.
5. Brandes, U. (2001). A faster algorithm for betweenness centrality. Journal of mathematical sociology, 25(2), 163-177.