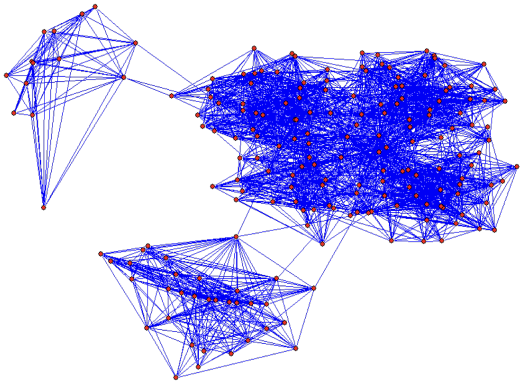
SNA
cse533

Unmasking the Avengers: A Narrative Network Analysis of the MCU Heroes

CSE 533, Social Network Analysis

Wired

Kathan Bhavsae	AU2040170
Manav Shah	AU2040088
Devyash Shah	AU2040152
Kush Patel	AU2040137



Ahmedabad University

Contents

1 | Abstract

This project analyzes the social network of Marvel's movies and comics by creating graphs of the interactions between characters. The nodes in the graph represent heroes, while the edges represent their conversations or appearances in the same comic book. The sentiment of positive and negative interactions was extracted from the movie scripts, which allowed for the creation of graphs of positive and negative relations to overall evaluate the character in terms of sentiment through social networks. Additionally, strong and weak ties graphs were generated for the movies to get insight into how strongly related the characters are in particular. For the comics, the edges were based on the appearance of characters in the same comic. The findings provide insight into the social dynamics of the Marvel universe, including character evaluation through degree centrality and community evaluation in the comics.

2 | Introduction

The Marvel universe has captured the hearts and imaginations of millions of fans worldwide. With its vast array of characters and interwoven storylines, it presents a rich source of material for social network analysis. This project aims to explore the social network of the Marvel universe by analyzing the interactions between its characters, with a focus on character evaluation and community evaluation.

For the movies, sentiment analysis was performed on the scripts to extract positive and negative interactions between characters, allowing for the creation of graphs of positive and negative relations. These graphs provide insight into the emotional dynamics between characters and their impact on the overall story arc. Additionally, character evaluation was performed using degree centrality, which measures the importance of a character within the network. By analyzing these graphs, we can gain a better understanding of the relationships between characters in the movies and their significance to the story.

For the comics, the edges in the graph were based on the appearance of characters in the same comic book. This data allowed for the creation of community graphs, which group characters based on their interactions and relationships. Community evaluation was performed to analyze the structure of these groups and identify key characters within them. Additionally, character evaluation was performed using degree centrality, which provides insight into the importance of a character within their community and the overall network.

Overall, this project provides a unique perspective on the social dynamics of the Marvel Universe. By analyzing the interactions between characters in the movies and the comics, we can gain valuable insights into the relationships that drive the story forward. The findings of this project have implications for fans, writers, and researchers alike and offer a new way of looking at the Marvel universe.

3 | Related Work

Marvel Universe Character Network" by Isabel Meirelles: This project created a character network of the Marvel universe and analyzed its evolution over time. The research found that the network became more complex over time as new characters were introduced and storylines became more interwoven.

"The Network of Thrones" by Andrew Beveridge and Jie Shan: This study analyzed the social network of characters in the Game of Thrones universe using network science. The research found that the social network of the characters was highly centralized, with a small group of characters playing a crucial role in the story.

"The Structure and Dynamics of Fictional Universes" by Shawn Graham: This study explored the use of network science in the analysis of fictional universes. The research argued that the study of fictional universes can provide insights into the dynamics of real-world social networks.

"The Social Network of DC Comics: Detective Comics" by Zdenko Kovacic: This study analyzed the social network of characters in DC Comics using social network analysis and centrality measures. The research found that Batman was the most important character in the network, while Superman was the most central in terms of social ties.

4 | Model, Approach and Methods

Movies

For the data extraction of Nodes and relationships from the script we initially used the Natural Language Processing models provided by the NLTK library of Python.
The algorithm for that was simple first the word tokenization of the script will take place thus, each word will get part of the speech tags such as verb, pronoun, adjective, and many more. After part of the speech tagging based on the tags, characters, and relations are defined, such as proper nouns for the names and the import events titles, and action words to describe any interaction or action between characters for edges. Different rules of grammar were employed by us to define the relationships between characters, but since our rules were unable to capture all the relations and actions from the script we relied on the AI tools to extract the nodes and edges of relation from the scripts.

After getting the nodes and edges we still had to work on the textual output since the output provided by the tools was also text and thus an appropriate logic was designed by us to construct the graph from that text data. We go the data in the format **Character1 and Character2: Positive/ Negative or Weak/ Strong**. We had about 250-300 lines in this format and thus we had to use loops to generalize the logic for all the lines. The logic was to first declare the empty bucket characters the loop for individual scripts of any movie was iterating over the characters of the string. The characters are added to that bucket as soon as the bucket is matched to any name of the characters list that we had from the Kaggle file data set. We also defined a flag called ”f” and whenever ”and” is detected the flag will be updated to 1 and after that, if the character is detected then we know that it’s the second character since we already had it. Thus, the edge will be formed between characters, and flag ”f” will be updated to 0 to detect new relations, Note that after detecting characters and ”and” the bucket will be emptied to detect new objects.

For the graphs of the sentiment analysis and strong and weak ties, we had the data related to positive/negative and the strong/ weak tires thus using the algorithm of the buckets discussed above the edges were labeled. For sentiment analysis, the edges were represented by different colors that are red for negative and green for positive edges, For strong and weak ties the edges were labeled as s (for strong) and w (for weak).

Comics

For the comics, we had the data in Excel regarding the characters of the comics. We had two pieces of data, one was regarding the characters and the comics, and the other was the data on the edges between the characters. We had a total of 12000 comics out of which we constructed the graphs of 11249. We were unable to construct the graphs of all the comics because not all the characters from both characters were correlating and that’s why we were only able to make the graphs for 11249. The algorithm for creating the graphs for all the comics is as follows. A list is created containing the dictionary for all the characters which will contain the names of comics in which the characters have appeared. After that a dictionary of empty graphs will be created for comics in other words the key of that dictionary is the comics name and the value of the graph for that comic. After that, the set of all the characters from the comic and hero network data is formed which will be used further, After that the data of the edges network which contains the character names side by side will be iterated over, and the character which has any comic in common, an edge will be formed between them and that network will be added to the graph of that comic. Now we faced some the problems in this construction such as the name of the character were not completely the same even though they represented the same characters. We solved that problem by using Jaccard similarity rule to compare the names for both the data sets and the names with highest similarity were considered same.

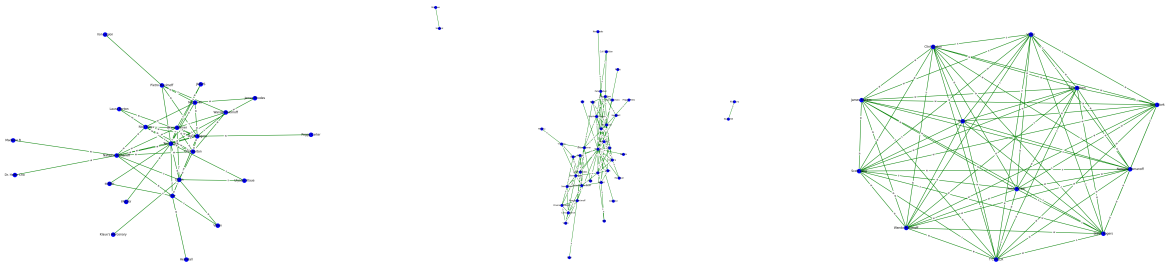
Community detection and evaluation algorithm

We merge multiple graphs into a single graph iteratively based on a set of rules to study the evolution of communities. Community louvain algorithm is used by us to find the communities at each step and a random color function to generate random colors, which are assigned to nodes based on their community. Each community will have different colour.We plot the graph using the Fruchterman-Reingold algorithm to lay out the nodes of the graph and tried to keep the visualisation such that the nodes and edges dont overlap each other.

A list of keys in the dictionary containing all the graphs is used to iterate through all the graphs, and merged keys is used to keep track of the keys of the merged graphs. The first graph is used as the starting point for the new merged graph. We then check whether any nodes in the new graph are present in the other graphs. If so, it merges the other graphs with the new graph. This process continues until all the graphs have been merged.

5 | Results

Follow are the resulting graphs representing the Social Network of Marvel Cinematic Universe's Movies:



(a) Age of Ultron **(b)** Infinity War **(c)** Civil War

Figure 5.1: Strong and Weak Ties Graphs

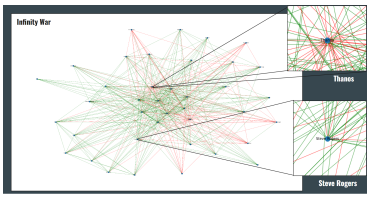
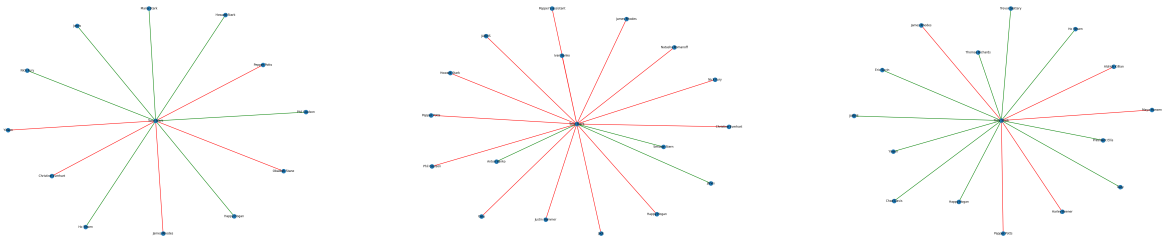


Figure 5.2: Sentiment Analysis of Infinity War

By using the Graphs Strong and weak ties we have also evaluated individual character's evaluation by isolating that character from all the movies in which he/ she has appeared. Similar kind of studied can be carried out for case of the positive and negative relations.

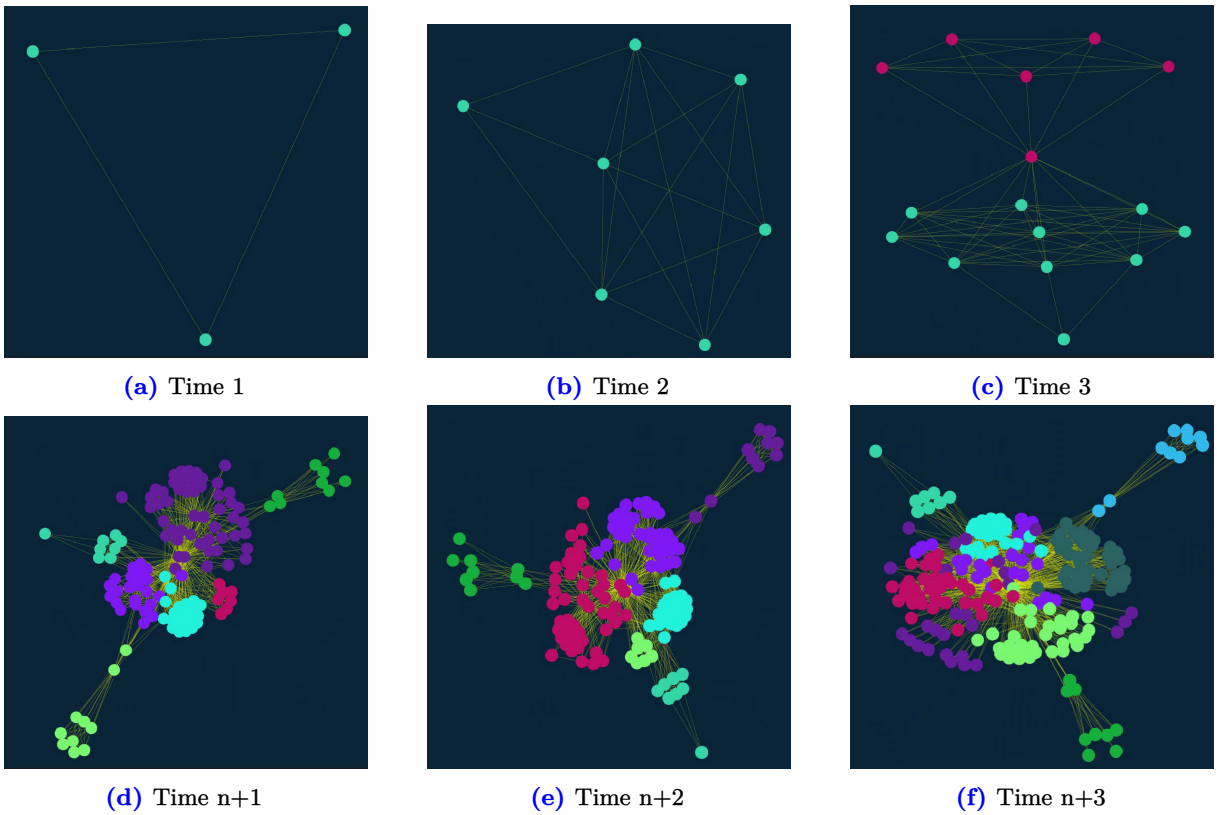
Following are the graphs of Tony Stark's evaluation.



(a) Iron Man-1 **(b)** Iron Man-2 **(c)** Iron Man-3

Figure 5.3: Tony's evolution through out Iron Man franchise.

Following figures represents the evaluation of the community in starting in the chronological order. Different colors are used for representing different communities.



(a) Time 1 **(b)** Time 2 **(c)** Time 3

(d) Time n+1 **(e)** Time n+2 **(f)** Time n+3

Figure 5.4: Community evaluation with time

Narrative network analysis and visualization techniques were used to gain a deeper understanding of character dynamics, relationships, and community evolution in the analysis of various MCU movies and Marvel comics.

The study investigated character connections in terms of strong and weak ties, providing insights into the significance of various types of relationships in films. Using network graphs to visualize these ties revealed the nature of character interactions and highlighted the dynamics that drive the narrative.

Furthermore, the analysis of positive and negative relationships between characters in MCU films provided a clearer picture of the films' alliances and conflicts. The network graphs emphasized the complexities of character relationships by displaying the distribution and intensity of these interactions.

Finally, analyzing character appearances and edge information from Marvel comic books revealed patterns of community formation and growth. The analysis iterative approach identified instances where communities merged based on the presence of at least one common character in both comics. This dynamic visualization depicted the evolution and transformation of Marvel comics' character communities.
Overall, the project was a success in that it used network analysis techniques to investigate character relationships, interactions, and community evolution in both MCU films and Marvel comic books. This analysis' findings contribute to a better understanding of the narrative structure, character development, and the forces that shape these interconnected universes.

6 | Conclusion

Overall, This is an extensive project that used narrative network analysis and network visualization techniques to investigate the connections, relationships, and community evolution among characters in MCU films and Marvel comics, emphasizing the importance of understanding the intricate character dynamics that shape the narratives.

The study looked at the role of strong and weak ties in shaping character interactions and dynamics in MCU films, providing valuable insights into the series' narrative structure and character development. The investigation of positive and negative relationships between characters in the MCU films revealed more about the changing nature of relationships and the factors that contributed to developing the overarching storyline.

The project provided insights into community formation and growth by analyzing the appearance of characters in Marvel comics and the connections between them, demonstrating the utility of an iterative approach to exploring community evolution. This method can be applied to other film franchises, large-scale comic universes, or character ensembles to better understand how relationships, shared appearances, and character interactions influence the narrative structure and character development over time.

Finally, the use of network analysis techniques in the project has resulted in a better understanding of character dynamics, relationships, and community evolution in both MCU films and Marvel comics. This analysis' findings contribute to a better understanding of the narrative structure, character development, and the forces that shape these interconnected universes.

7 | References

[1] Duton, P. (2021). Marvel Cinematic Universe Dialogue: Dialogue from 18 of the MCU movies, and character / movie information. Retrieved from `https://www.kaggle.com/datasets/pdunton/marvel-cinematic-universe-dialogue`

[2] Bearman, P. S., & Stovel, K. (2000). Becoming a Nazi: A Model for Narrative Networks. *Poetics*, 27(2-3), 69-90.

[3] Sandberg, S. (2022). Narrative Analysis in Criminology. *Journal of Criminal Justice Education*, 33(2), 212-229. doi:10.1080/10511253.2022.2027479

[4] Sudhahar, S., De Fazio, G., Franzosi, R., & Cristianini, N. (2015). Network Analysis of Narrative Content in Large Corpora. *Natural Language Engineering*, 21(1), 81-112. doi:10.1017/S1351324913000247

[5] Marvel Comics. (n.d.). Series. Retrieved October 20, 2023, from https://www.marvel.com/comics/series

8 | References

A | Appendix A title

A. Data Collection and Processing

- MCU Movies
  - Movie scripts
  - Character interactions and relationships
- Marvel Comics
  - 12,000 comic dataset
  - Character appearances and edge information

B. Network Analysis

1. Strong and weak ties
2. Positive and negative relationships
3. Community evolution

C. Graph Visualization

- MCU movies
  - Strong and weak ties
  - Positive and negative relationships
- Marvel Comics
  - Community evolution
  - Merging communities based on common characters

D. Tools and Libraries

1. Python
2. Networkx
3. Matplotlib
4. Jupyter Notebook

E. Results

-
  - Strong and weak ties
  - Positive and negative relationships
- Community evolution in Marvel Comics
  - Patterns of community formation and growth
  - Merging communities based on common characters

F. Conclusions

Significance of network analysis in understanding narrative structure and character development Application to other movie franchises or character ensembles Insights gained from the analysis of MCU movies and Marvel comics

G. Future Work

Expanding the analysis to include more movie franchises or comic universes Investigating the impact of specific events or storylines on character relationships and community evolution Developing predictive models for character interactions and community formation