

SNA U1 - Project

Priscila Guadalupe Tzuc Alonzo
Data Engineering
Universidad Politécnica de Yucatán
Ucú, Yucatán, México
2109148@upy.edu.mx

Abstract—This study presents a comprehensive network analysis of Victor Hugo's classic novel "Les Misérables", revealing the intricate social dynamics and character interactions within the narrative. The network consists of 77 characters and 254 interactions, and shows a tightly connected structure with an average path length of 2.64 and a clustering coefficient of 0.57. The centrality measures identify key characters, likewise identifying key figures such as Valjean, Gavroche, Javert and Marius as pivotal characters due to their high degree, interrelatedness, proximity and eigenvector centralities.

On the other hand, the degree distribution indicates that while most characters have limited interactions, a few, such as Valjean, are highly connected, reflecting a typical social network pattern.

Overall this analysis provides a detailed understanding of the social structure and relationships in "Les Misérables," enhancing our appreciation of the complexity of the novel and the pivotal role played by key characters in driving the plot. The results underscore the utility of network analysis in literary studies by offering nuanced insights into the interactions between characters and narrative dynamics.

INTRODUCTION

This analysis focuses on a network derived from the classic novel "Les Misérables" by Victor Hugo. The network consists of 77 nodes and 254 edges, representing the interactions between characters in the story. Each node represents a character, and each edge indicates co-occurrences between characters within the same chapter.

The "Les Misérables" network is an undirected graph, where the interactions between characters are mutual and bidirectional. This type of network is crucial for understanding social structures within the narrative, as it highlights the relationships and interactions between different characters. Analyzing such a network helps in identifying key figures, understanding community dynamics, and exploring the social fabric of the story.

These are the points that will be covered in the document:

- **Literary Analysis:** Understanding character dynamics and their significance within the narrative.
- **Community Detection:** Identifying clusters or groups of characters who frequently interact.
- **Centrality Measures:** Highlighting the most influential characters based on various centrality metrics.
- **Educational Tool:** Providing a visual and analytical tool for students studying literature or social networks.

- **Network Science Research:** Using the dataset as a benchmark for testing new algorithms and methods in network analysis.

The dataset has been extensively used in multiple research papers focusing on social network analysis, community detection, and network visualization. General results from these analyses typically highlight key characters, their interactions, and the overall structure of the network, providing a deeper understanding of the narrative's social fabric.

The "Les Misérables" network can be classified as follows:

- **Undirected:** The edges have no direction, representing mutual co-occurrences between characters.
- **Unweighted:** All edges are treated equally, without any weight indicating the strength of interactions.
- **Non-Planar:** Given the number of nodes and edges, the network cannot be drawn on a plane without edge crossings. [1]

Visualization of the Network

Below is a visualization of the "Les Misérables" character network, providing a clear view of the structure and relationships within the story.

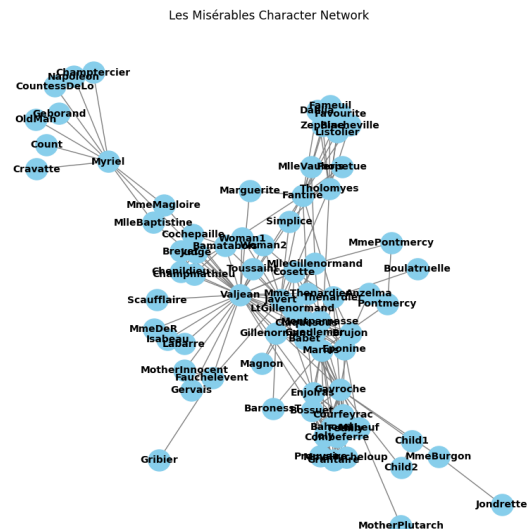


Fig. 1.

I. NETWORK CHARACTERISTICS

Here are the detailed results of the network analysis for the "Les Misérables" character network:

- 1) **Size of the Network:** 77
 - The network consists of 77 nodes, each representing a character in the novel.
- 2) **Number of Links:** 254
 - There are 254 edges, indicating the interactions between the characters.
- 3) **Average Path Length:** 2.64
 - The average shortest path length between any two nodes is approximately 2.64. This suggests that on average, any character is separated by about 2.64 steps from another.
- 4) **Clustering Coefficient:** 0.57
 - The network has a clustering coefficient of 0.57, indicating a relatively high tendency of characters to form tightly knit groups.
- 5) **Average Distance:** 2.64
 - The average distance, being the same as the average path length, is 2.64.
- 6) **Diameter:** 5
 - The longest shortest path between any two nodes in the network is 5, meaning the greatest distance between any pair of characters is five steps.
- 7) **Eccentricity:**
 - Eccentricity represents the maximum distance from a node to all other nodes. Some examples include:
 - Myriel: 3
 - Valjean: 3
 - Fantine: 3
 - Cosette: 3
 - Javert: 3
 - Gavroche: 3
- 8) **Radius:** 3
 - The radius, representing the smallest eccentricity in the network, is 3. This means the minimum distance from the center to any node is three steps.
- 9) **Periphery:** ['Napoleon', 'Geborand', 'Cravatte', 'OldMan', 'Valjean']
 - Nodes such as 'Napoleon', 'Geborand', 'Cravatte', 'OldMan', and 'Valjean' are on the periphery, indicating they are the farthest from the center of the network.
- 10) **Center:** ['Valjean', 'Marius', 'Cosette', 'Javert', 'Gavroche']
 - Nodes such as 'Valjean', 'Marius', 'Cosette', 'Javert', and 'Gavroche' are central, indicating they are closest to all other nodes. [2]

II. CENTRALITY MEASURE ANALYSIS

For analyzing the "Les Misérables" character network, the following centrality measures have been chosen: Degree Centrality, Betweenness Centrality, Closeness Centrality, and Eigenvector Centrality. These measures provide a comprehensive understanding of the importance and influence of each character within the network. [3]

1. Degree Centrality

Degree Centrality measures the number of direct connections a node has. Characters with the highest degree centrality are the most connected within the network.

2. Betweenness Centrality

Betweenness Centrality measures the extent to which a node lies on the shortest paths between other nodes. Characters with high betweenness centrality act as bridges, facilitating interactions between different parts of the network.

3. Closeness Centrality

Closeness Centrality indicates how close a node is to all other nodes in the network. Characters with high closeness centrality can quickly reach all other characters.

4. Eigenvector Centrality

Eigenvector Centrality measures a node's influence based on the number and quality of its connections. Characters with high eigenvector centrality are influential not just because of their connections but also because they are connected to other well-connected characters.

RESULTS AND ANALYSIS

Degree Centrality

Top Characters: Valjean, Gavroche, Javert, Marius, Thenardier

Interpretation: Valjean, the main character, has the highest number of connections, followed by other central figures such as Gavroche and Javert. These characters are pivotal to the storyline due to their numerous interactions.

Betweenness Centrality

Top Characters: Valjean, Gavroche, Javert, Marius, Enjolras

Interpretation: Valjean acts as a key bridge within the network, facilitating connections between different character groups. Gavroche and Javert also play crucial roles in connecting various parts of the narrative.

Closeness Centrality

Top Characters: Valjean, Gavroche, Javert, Marius, Cosette

Interpretation: Characters with high closeness centrality, such as Valjean and Gavroche, can efficiently reach all other characters, indicating their central role in the network's communication flow.

Eigenvector Centrality

Top Characters: Valjean, Gavroche, Javert, Marius, Enjolras

Interpretation: Valjean and Gavroche are highly influential not just due to their direct connections but also because they are connected to other important characters, amplifying their overall influence.

VISUALIZATION OF THE CENTRALITY MEASURES

To visualize the centrality measures, the following plots highlight the top characters based on each measure:

	Degree Centrality	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality
Valjean	0.473684	0.569989	0.644068	0.267818
Gavroche	0.289474	0.165113	0.513514	0.317839
Marius	0.250000	0.132032	0.531469	0.259111
Javert	0.223684	0.054332	0.517007	0.184225
Thenardier	0.210526	0.074901	0.517007	0.187808
Fantine	0.197368	0.129645	0.480606	0.089759
Enjolras	0.197368	0.042553	0.481013	0.267179
Bossuet	0.171053	0.030754	0.475000	0.242131
Courfeyrac	0.171053	0.005267	0.400000	0.232467
Joly	0.157895	0.002185	0.393782	0.221554

Fig. 2.

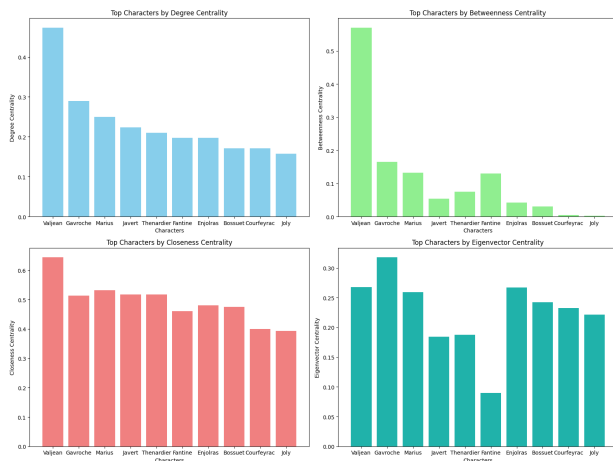


Fig. 3.

III. DEGREE DISTRIBUTION

DEGREE DISTRIBUTION ANALYSIS OF THE "LES MISÉRABLES" NETWORK

The degree distribution of the "Les Misérables" network provides insights into how interactions are spread among the characters. Here's a detailed analysis based on the degree distribution plot:

Key Observations:

1) Most Common Degrees:

- The majority of characters have a relatively low degree, indicating that they interact with only a few other characters.

- This is reflected in the histogram where the bins corresponding to lower degrees have higher frequencies.

2) Central Characters:

- Characters with high degrees are central figures in the story. They interact with many other characters and are key to the narrative.
- For instance, Valjean, the protagonist, has one of the highest degrees, indicating his numerous interactions throughout the novel.

3) Peripheral Characters:

- Many characters have a degree of 1 or 2, suggesting they are minor characters with limited interactions.
- These characters might appear briefly or in specific contexts within the story.

4) Degree Distribution Shape:

- The distribution follows a pattern where a few characters have very high degrees while the majority have low degrees.
- This is typical of social networks and can be indicative of a scale-free network where a few nodes (characters) are highly connected.

5) Interpretation of the Tail:

- The tail of the distribution, representing high-degree characters, shows a steep drop-off, indicating that very few characters have a high number of interactions.
- These high-degree characters are likely central to multiple subplots and are involved in various critical interactions. [4]

Implications for "Les Misérables":

1) Character Importance:

- High-degree characters are likely to be the main drivers of the plot. Their actions and interactions significantly impact the story's progression.
- For example, Valjean, Javert, and Gavroche are central figures whose interactions drive much of the narrative.

2) Narrative Structure:

- The presence of many low-degree characters suggests a rich and diverse cast with numerous minor roles contributing to the depth of the story.
- This diversity can add complexity and realism to the narrative, depicting a broad social landscape.

3) Community Detection:

- The degree distribution can help identify potential clusters or communities within the network. High-degree characters might serve as hubs connecting different groups.
- Understanding these clusters can provide insights into the social structure and relationships within the novel.

4) Literary Analysis:

- Analyzing the degree distribution helps in understanding the prominence and significance of characters based on their connectivity.
- It also aids in identifying central themes and interactions that are crucial to the story.

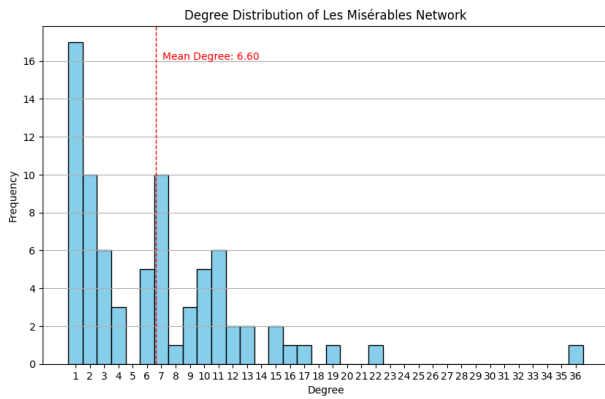


Fig. 4.

IV. COMMUNITY DETECTION

Community Structure:

- The network is divided into several communities, each represented by a different color.
- These communities indicate groups of characters that are more closely connected with each other than with those in other groups.

Central Figures:

- Central characters like Valjean, Javert, and Cosette are likely to be part of larger or more interconnected communities.
- These central figures often act as bridges between different communities.

Narrative Clusters:

- Communities can represent different subplots or interactions within the novel.
- Analyzing these clusters helps in understanding the structure of the story and the relationships between different groups of characters.

Visualization Insights:

- The visualization helps identify which characters belong to which community.
- It also shows how tightly knit certain groups are compared to others.

By performing community detection and analyzing the resulting communities, we gain a deeper understanding of the social structure and dynamics within the "Les Misérables" network. This analysis is useful for literary studies, network analysis, and understanding complex social interactions. [5]

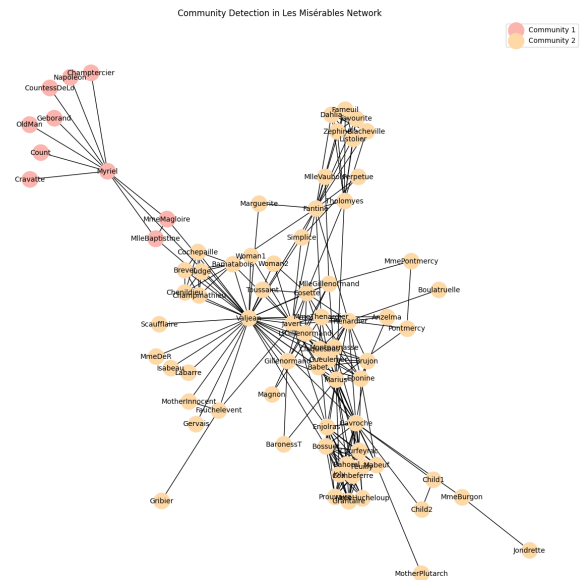


Fig. 5.

CONCLUSION

The comprehensive network analysis of "Les Misérables" reveals intricate social dynamics and interactions between characters within Victor Hugo's novel. With a network consisting of 77 characters and 254 interactions, the analysis demonstrates that characters are closely connected, as indicated by an average path length of 2.64, on the other hand the clustering coefficient of 0.57 suggests a tendency for characters to relate to each other, and the grouping coefficient of 0.57 suggests a tendency for characters to form tightly knit groups, reflecting the complex and intertwined nature of the narrative.

Central characters such as Valjean, Gavroche, Javert, and Marius emerge as pivotal figures in the network. These characters exhibit heightened measures of centrality, such as centralities of degree, betweenness, proximity, and eigenvector, underscoring their importance and influence in the story. Valjean, in particular, stands out as the most central character, with extensive interactions throughout the network, reinforcing his role as protagonist.

Analysis of the degree distribution reveals that while most characters have a low degree, indicating limited interactions, a few characters possess a high degree, underlining their centrality and importance. This pattern is typical of social networks, where a small number of nodes (characters) are highly connected, contributing to the overall cohesion of the network.

The detection of communities further enriches the analysis by identifying distinct groups within the network. These communities represent subplots and groups of characters that interact with each other more frequently. The visualization of these communities, each represented by a unique color, provides a clear depiction of the social structure and rela-

tionships within the novel. The central characters often act as bridges between the various communities, highlighting their crucial role in maintaining the fluidity and connectivity of the narrative.

The combination of network characteristics, centrality measures, degree distribution, and community detection provides a comprehensive understanding of the social fabric of "Les Misérables." The analysis reveals how closely connected characters drive the plot and how central figures facilitate interactions between different parts of the network. By examining these aspects, network analysis underscores the complexity and depth of Victor Hugo's literary masterpiece, providing a nuanced perspective on the relationships and interactions that shape the story. This detailed examination allows us to better appreciate the novel's rich social dynamics and the pivotal role played by key characters in the narrative.

REFERENCES

- [1] gephi, "Datasets," GitHub, Jul. 23, 2021. <https://github.com/gephi/gephi/wiki/Datasets> (accessed Jun. 21, 2024).
- [2] M. Gamper, "Social Network Theories: An Overview," Springer eBooks, pp. 35–48, Jan. 2022, doi: https://doi.org/10.1007/978-3-030-97722-1_3.
- [3] S. Gómez, "Centrality in Networks: Finding the Most Important Nodes," Business and Consumer Analytics: New Ideas, pp. 401–433, 2019, doi: https://doi.org/10.1007/978-3-030-06222-4_8.
- [4] N. Meghanathan, "Theoretical Network Models." Accessed: Jun. 21, 2024. [Online]. Available: <https://www.jsu.edu/nmeghanathan/files/2016/01/CSC641-Sp2016-Module-5-TheoreticalNetworkModels.pdf>
- [5] R. Cazabet, G. Rossetti, and F. Amblard, "Dynamic Community Detection," Encyclopedia of Social Network Analysis and Mining, pp. 1–10, 2017, doi: https://doi.org/10.1007/978-1-4614-7163-9_383-1.