DASC – 804: Social Network Analysis
MPHIL (DS) – Semester 1 of 2023
Course Instructor: Dr. Abdul Waheed Mahesar
Faculty of Engineering and Technology
University of Sind, Jamshoro

Class Assignment
Department: Data Science
Submitted by: Zulfiqar Ali Shaikh
[redacted]

Exploring the Wizarding World: A Social Network Analysis of the Harry Potter Movie Series

ABSTRACT

This research presents a comprehensive social network analysis of the Harry Potter movie series, offering valuable insights into the intricate web of relationships among the beloved wizarding characters. Leveraging a dataset of 65 characters and their interactions, we explore into the social structure of the Wizarding World, unveiling the presence of distinct communities and the pivotal roles of specific characters within these communities. Key findings include a moderate modularity score, illustrating the network's modular nature, and visual representations that highlight the interconnectedness of characters. Centrality measures reveal characters of significant influence, while clustering coefficients emphasize the tightly-knit groups within the Wizarding World. Our analysis provides a deeper understanding of the complex social fabric of the Harry Potter universe, offering a foundation for further exploration into the dynamics of this magical realm.

INTRODUCTION

The Harry Potter movie series, based on the beloved novels by J.K. Rowling, has captivated audiences worldwide since its cinematic debut in 2001. The magical world of Harry Potter, complete with its unique characters, intricate plotlines, and enchanting settings, has not only provided an unparalleled source of entertainment but has also sparked academic curiosity and engagement. This interest has led to various research avenues exploring different aspects of the Harry Potter series.

Social Network Analysis (SNA) is a powerful methodology used to study the interrelationships among characters, organizations, or entities within a narrative context. In the case of Harry Potter, this technique allows us to investigate deeper into the intricate web of relationships among the characters and their roles within the storyline. By analyzing the social network embedded within the narrative, we can uncover hidden patterns, dynamics, and insights that may not be apparent through traditional narrative analysis.

This research embarks on a journey through the wizarding world of Harry Potter to investigate the social network that binds the characters in the movies. By applying SNA to the cinematic adaptations, we aim to identify key characters, explore their interactions, and gain a better understanding of the social structure that underlies the magical universe. The analysis will consider the importance of characters in the narrative, the strength of their relationships, and the impact of their interactions on the unfolding of events.

Moreover, this study seeks to contribute to the growing body of research that applies SNA to works of fiction and popular culture, highlighting the potential of this methodology in gaining insights into complex narrative structures. The Harry Potter series, with its intricate character relationships and richly woven narrative, provides a compelling context for such analysis.

In this exploration of the social network within the Harry Potter movie series, we anticipate uncovering valuable insights into the power of social connections, character dynamics, and narrative complexity. Through this research, we hope to enrich our understanding of this beloved series and, by extension, the broader application of Social Network Analysis in the field of popular culture and literary studies.

Social Network Analysis is a powerful tool used to study social structures and relationships by representing them as a network of nodes (characters) and edges (connections between characters). By examining these relationships, we can better understand the underlying structure of the Harry Potter series and the social dynamics at play in the wizarding community.

DATA COLLECTION

To conduct this analysis, we utilized a publicly available dataset from GitHub, specifically the repository maintained by efekarakus [1]. This dataset is comprised of two main files:

1. **characters.csv**: This file contains a comprehensive list of 65 Harry Potter characters. Each entry includes essential information such as character id, name, label, and a brief character biography. This data forms the foundation for the nodes in our network.

2. **relations.csv**: In this file, we find a list of undirected relations between the characters. These relations are labeled as either alliances or enemies, denoted by the symbols '+' and '-', respectively. Each entry in this file includes the source character, the target character, and the type of relationship. This data will serve as the basis for defining the edges in our network.

Software used for analysis is Python, a versatile and powerful programming language, was employed to conduct the Social Network Analysis. By leveraging popular Python libraries such as NetworkX, Matplotlib, Numpy, and Pandas, we were able to construct the network and perform various analytical tasks on it.

In the following section of Analysis, we will examine into the steps of our analysis, including the construction of the social network, measures of centrality, community detection, and the interpretation of results. Ultimately, this analysis will provide a unique perspective on the intricate web of relationships within the Harry Potter series.

ANALYSIS OF SOCIAL NETWORK

The first step involves importing the nodes (characters) from the 'characters.csv' file and creating a mapping of character ids to labels. This allows us to identify all the characters included in the analysis. Next, the edges (relationships) are imported from the 'relations.csv' file. The '+' relationships are represented as 1, and the '-' relationships as 0. These edges form the connections between characters.

Assigning relationships between positive (allies) and negative (enemies) relationships by filtering the edges accordingly. This step identifies the nature of each connection in the network. A graph is constructed using NetworkX library, with nodes representing characters and edges representing their relationships. This graph in figure.1 provides a visual representation of the Wizarding World's social network.

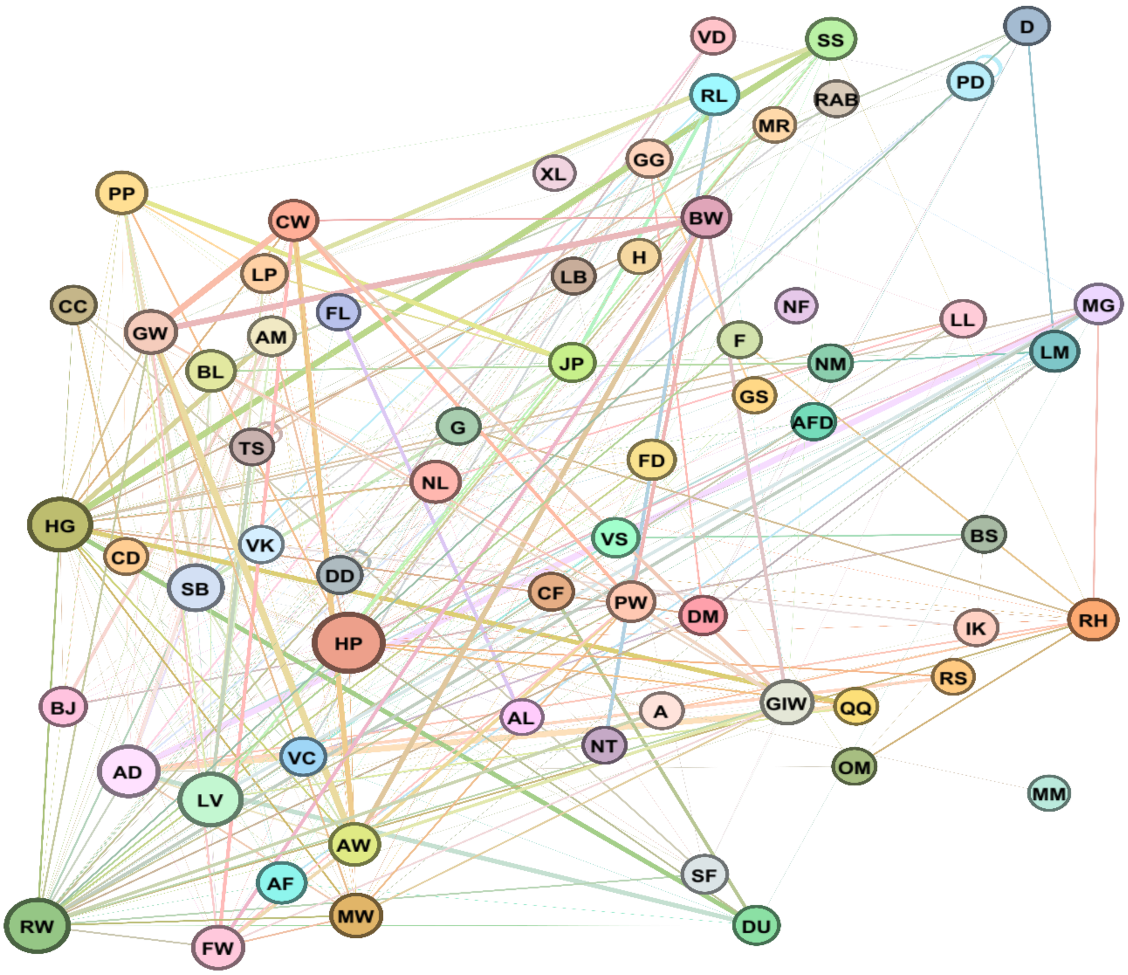

Figure. 1 (Network Graph)

General Network Information, including the number of Nodes (65 characters), the number of Edges (333 relationships), Graph Density (0.160, indicating moderate connectivity), Degree Centrality (0.75, suggesting that most characters are moderately connected), and Network Diameter (4, the maximum shortest path between any two characters).

The network graph is visualized with a circular layout in figure.2. Node size corresponds to character importance, with positive relationships (allies) shown in blue and negative relationships (enemies) in red. Character labels are added for clarity.

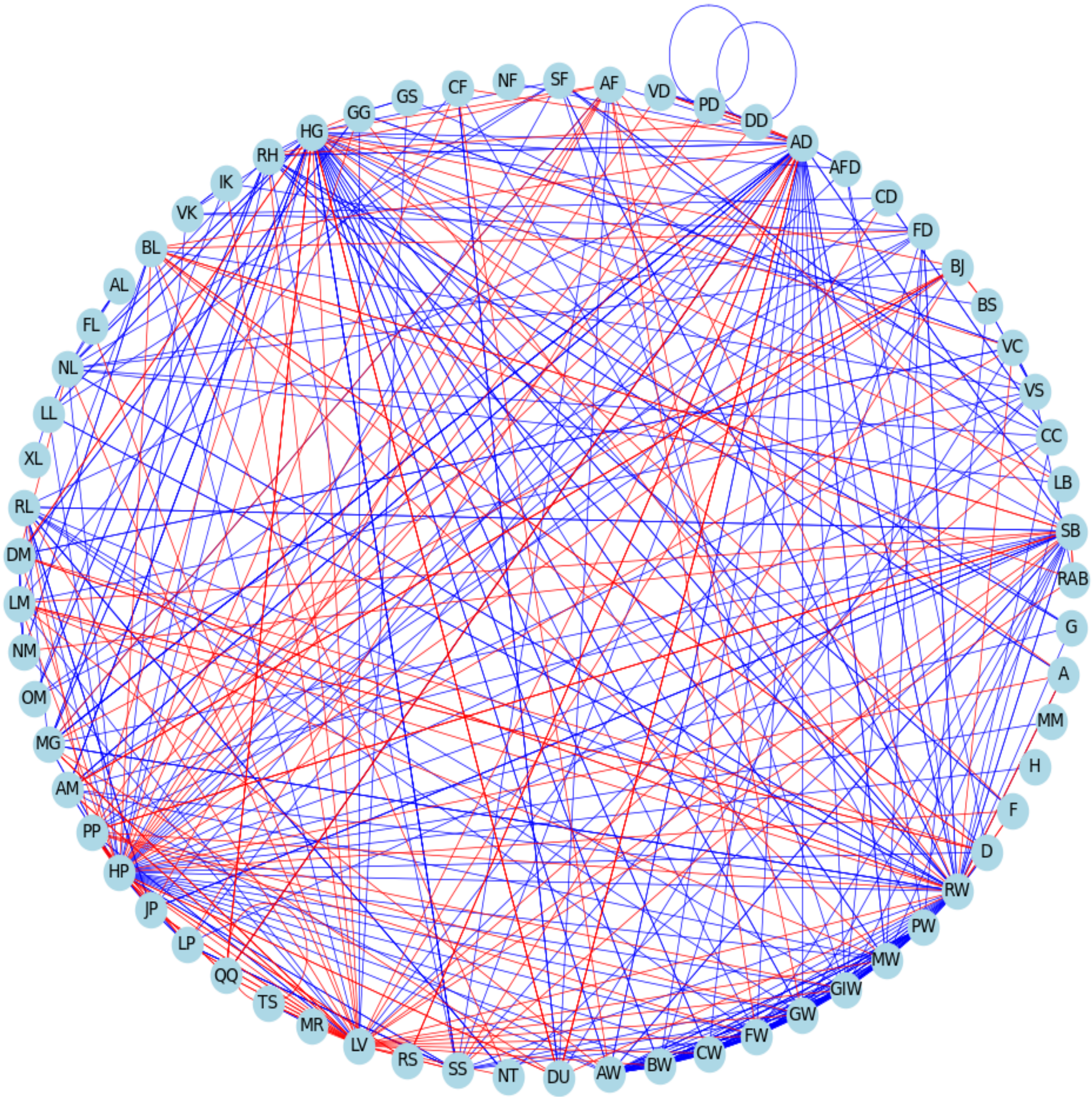

Figure. 2 (Circular Graph showing both Positive and Negative relationship)

A Dual Circle Graph is created, emphasizing the top 5 characters (nodes) with the highest degrees (most connections) inside the circle. Node sizes reflect the degree of connectivity. Positive relationships are shown in blue, while negative relationships are shown in red in figure.3.

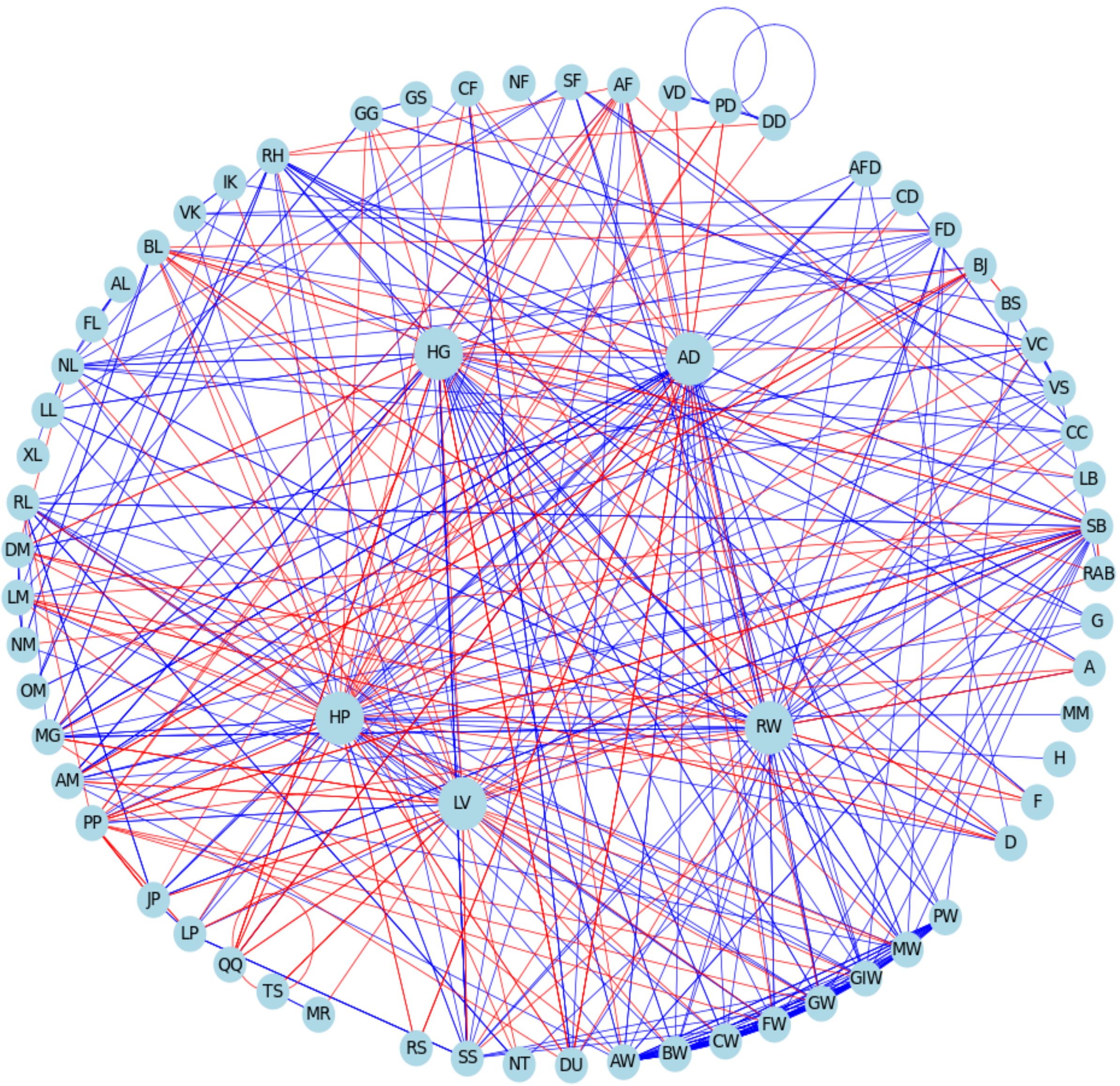

Figure. 3 (Dual Circular Graph showing both Positive and Negative relationship)

A similar Dual Circle Graph is created, emphasizing the top 5 nodes with the highest degrees inside the circle, but this time, only positive relationships (allies) are shown in blue. This visualization highlights the cooperation and alliances within the Wizarding World in figure.4.

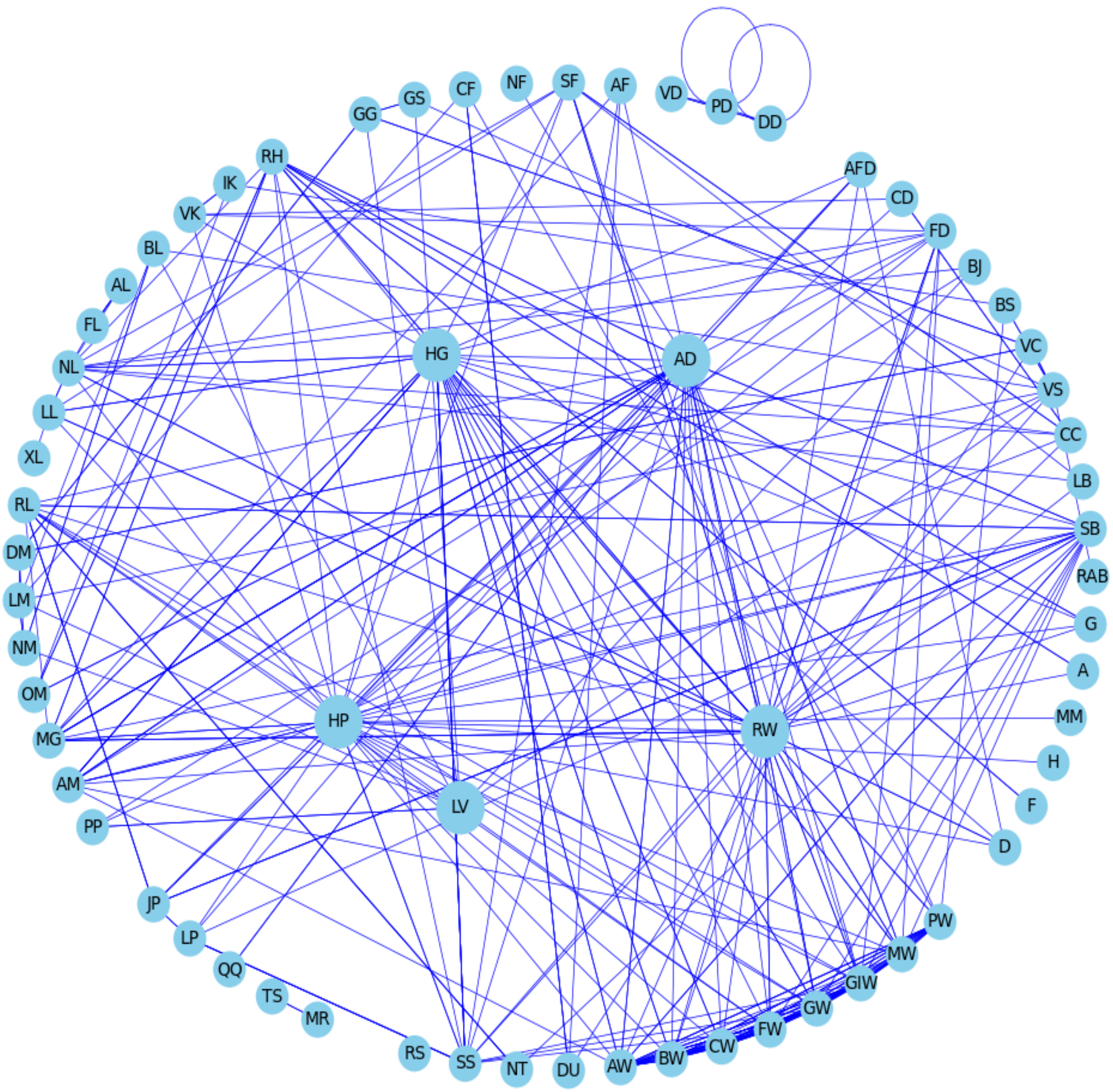

Figure. 4 (Dual Circle Graph Showing Positive Relationship)

This Dual Circle Graph showcases the negative relationships (enmities) within the network. Emphasizing the top 5 nodes with the highest degrees inside the circle. Red edges illustrate the conflicts and rivalries among characters in figure.5.

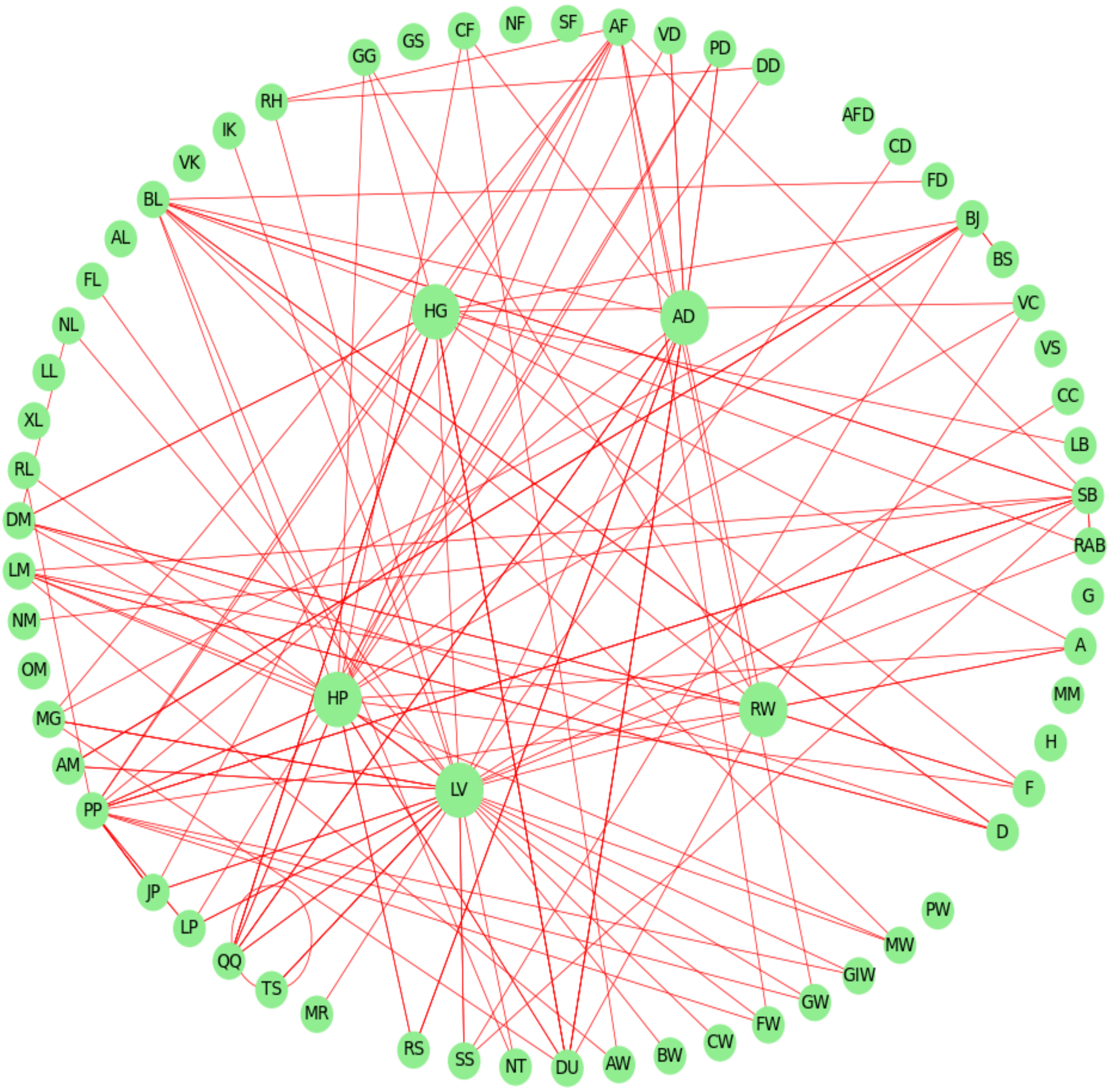

Figure. 5 (Dual Circle Graph Showing Negative Relationship)

Degree Centrality:

The analysis identifies the top 10 characters (nodes) with the highest degree centrality, which signifies their level of influence within the network. The top character (node) 39 has a degree of 48, indicating extensive connections shown in table. 1.

TOP 10 NODES WITH HIGHEST DEGREE		
Serial. No	**Node**	**Degree**
1	39	48
2	58	37
3	21	36
4	45	36
5	11	32
6	1	23
7	55	18
8	53	17
9	54	17
10	56	17

(Table. 1)

The top 10 characters (nodes) with the highest degrees include key figures such as Node 39, Node 58, Node 21, and Node 45, showcasing their centrality and importance in the Wizarding World.

Betweenness Centrality:

The betweenness centrality for all nodes in the network, measuring how central each character is in terms of connecting other characters. The top 10 characters (nodes) with the highest betweenness centrality are identified. These characters serve as important bridges or intermediaries within the Wizarding World's social network shown in table. 2.

TOP 10 NODES WITH HIGHEST BETWEENNESS CENTRALITY		
Serial. No	**Node**	**Betweenness Centrality**
1	39	0.282
2	45	0.203
3	11	0.1
4	58	0.099
5	21	0.094
6	28	0.045
7	29	0.031
8	1	0.028
9	22	0.019
10	25	0.017

(Table. 2)

Node 39, with a betweenness centrality of 0.282, holds a crucial position in connecting other characters, indicating significant influence. Characters like Node 45, Node 11, Node 58, and Node 21 also play pivotal roles in connecting various parts of the network.

Closeness Centrality:

The closeness centrality for all nodes, measuring how close each character is to all others in the network. The top 10 characters (nodes) with the highest closeness centrality are identified. These characters are best positioned to access information quickly and have strong connections to others shown in table. 3.

TOP 10 NODES WITH HIGHEST CLOSENESS CENTRALITY		
Serial. No	**Node**	**Closeness Centrality**
1	39	0.8
2	58	0.703
3	21	0.696
4	45	0.688
5	11	0.653
6	1	0.593
7	55	0.571
8	53	0.566
9	54	0.566
10	56	0.566

(Table. 3)

Node 39 has the highest closeness centrality of 0.8, indicating it is most centrally located within the network. Other characters in the top 10, such as Node 58, Node 21 and Node 45, also have high accessibility to the network.

Global Clustering Coefficient:

The global clustering coefficient, which provides insight into the network's overall tendency for characters to form tightly-knit groups or communities. The global clustering coefficient is 0.612, suggesting a relatively high degree of clustering in the network. This indicates that characters within the Wizarding World often form cohesive groups or communities.

Binary Global Clustering Coefficient:

The binary global clustering coefficient, which quantifies the likelihood that the connections of one character also connect to each other (transitivity). The binary global clustering coefficient is 0.413, indicating that there is a moderate level of transitivity in the network. Characters' relationships are somewhat interconnected, but not overwhelmingly so.

Average Binary Local Clustering Coefficients:

The average binary local clustering coefficient for all nodes in the network, offering a measure of how clustered the network is at the individual character level. The average binary local clustering coefficient is 0.612, which aligns with the global clustering coefficient. This suggests that on average, characters in the Wizarding World tend to form tightly-knit groups.

Binary Local Clustering Coefficients:

The binary local clustering coefficient for each individual character in the network, revealing how connected they are to their immediate neighbors shown in table. 4.

Node	Binary Local Clustering Coefficient	Node	Binary Local Clustering Coefficient
0	0.833	**33**	0.5
1	0.478	**34**	0.7
2	1	**35**	0.167
3	0.762	**36**	0.709
4	0.222	**37**	0.756
5	0.619	**38**	0.629
6	0	**39**	0.201
7	0.5	**40**	0.857
8	0.491	**41**	0.75
9	0.667	**42**	1
10	0.9	**43**	1
11	0.327	**44**	1
12	0.667	**45**	0.248
13	0.7	**46**	1
14	0.833	**47**	0.626
15	0.705	**48**	0.833
16	0.714	**49**	0.714
17	0	**50**	0.714
18	0.6	**51**	0.859
19	0.667	**52**	0.939
20	0.667	**53**	0.757
21	0.284	**54**	0.757
22	0.451	**55**	0.673
23	0	**56**	0.706
24	0.4	**57**	0.691
25	0.462	**58**	0.299
26	1	**59**	0.714
27	0.667	**60**	1
28	0.396	**61**	0
29	0.667	**62**	0
30	0	**63**	1
31	0.727	**64**	1
32	0.556		

(Table. 4)

Different characters have varying binary local clustering coefficients. For instance, Node 2 has a perfect coefficient of 1.000, indicating that all of its connections are connected to each other. Node 21 has a lower coefficient of 0.284, indicating a lower level of clustering in its immediate network.

Communities in the Network:

The Louvain method is used to identify communities or groups of characters within the network. The partitions characters into different communities based on their connections shown in table. 5.

Community	Nodes
2	[0,4,5,6,7,18,19,20,23,25,32,33,34,35,38,45,49]
1	[1,11,15,17,31,36,40,41,42,47,48]
0	[2,3,8,9,10,16,21,22,24,26,27,28,29,30,37,39,58,59,60,61,62,63,64]
6	[12,13,14]
2	[43,44]
4	[46]
5	[50,51,52,53,54,55,56,57]

(Table. 5)

The network is divided into several communities, each represented by a group of nodes. The number of communities indicates the existence of distinct social groups or factions within the Wizarding World. Characters are grouped together based on their interactions and relationships. Community 0 is the largest and most interconnected, while smaller communities like Community 6, Community 2 and Community 4 represent more isolated or specialized groups shown in figure.6.

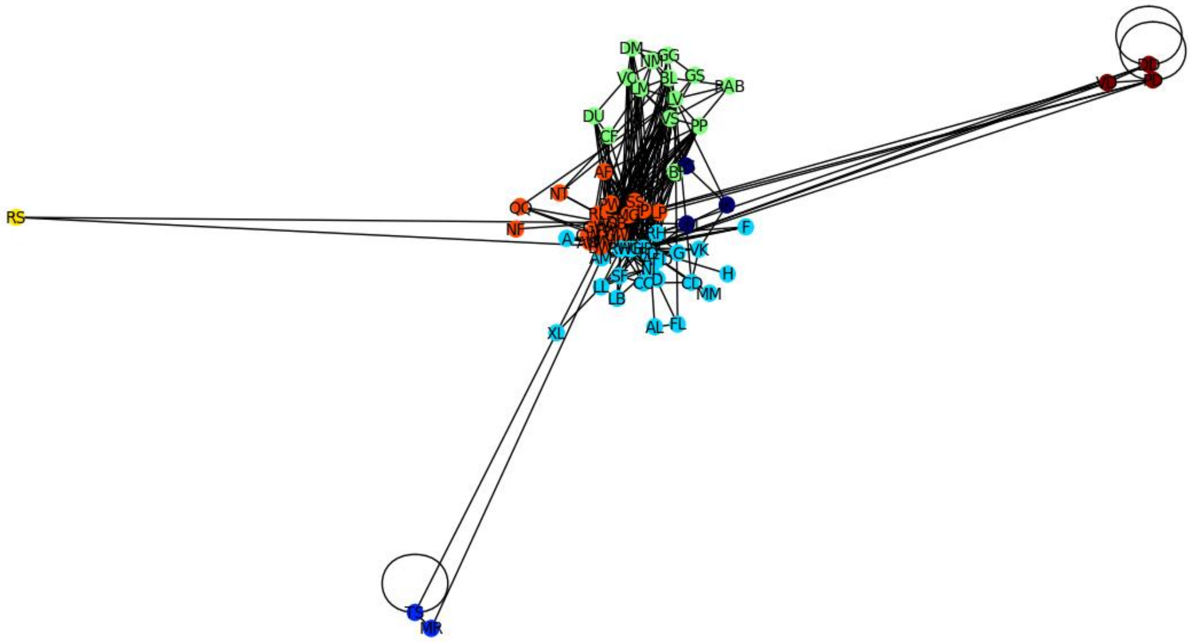

Figure. 6 (Network Communities)

Modularity of Network:

The modularity of the network based on the partition of communities within the network. The modularity score is 0.381, indicating that the network's structure exhibits a level of separation into distinct communities. A higher modularity score suggests that the network's structure is more modular, meaning characters are more connected within their respective communities.

Modularity Classes of Communities:

A Dictionary is created that assigns nodes to their corresponding modularity classes, providing insight into which characters belong to each community shown in table. 6.

Modularity Classes	Nodes
2	[0,4,5,6,7,18,19,20,23,25,32,33,34,35,38,45,49]
1	[1,11,15,17,31,36,40,41,42,47,48]
0	[2,3,8,9,10,16,21,22,24,26,27,28,29,30,37,39,58,59,60,61,62,63,64]
6	[12,13,14]
2	[43,44]
4	[46]
5	[50,51,52,53,54,55,56,57]

(Table. 6)

Characters are divided into different modularity classes, highlighting the existence of distinct social groups or communities within the Wizarding World. Each class represents a group of characters who are closely interconnected within their community.

The subgraphs for each modularity class and show the calculation of the number of nodes and edges within these subgraphs shown in table. 7.

Modularity Class	Nodes	Edges
0	17	34
1	11	24
2	23	67
3	3	5
4	2	2
5	1	0
6	8	28

(Table. 7)

Subgraphs represent the different communities or modularity classes identified within the network. Some communities are more substantial, with a larger number of nodes and edges, while others are smaller and more isolated.

The visual representations of each subgraph, offering a graphical view of the individual communities within the network shown in figure. 7.

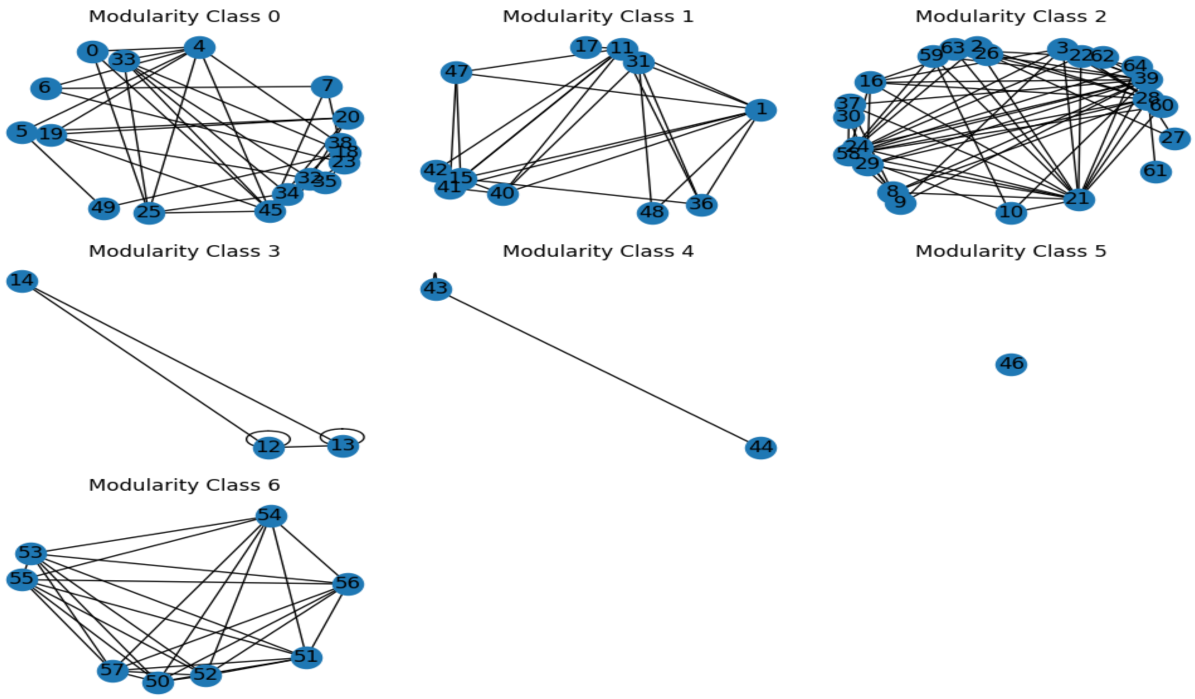

Figure. 7 (Subgraph of Communities)

The visual representations provide an intuitive way to see how characters within each community are connected. Characters within the same community have stronger connections with each other, creating more densely connected subgraphs.

The visual representation of the entire network graph, with different colors for modularity classes. Nodes are labeled according to their character names shown in figure. 8.

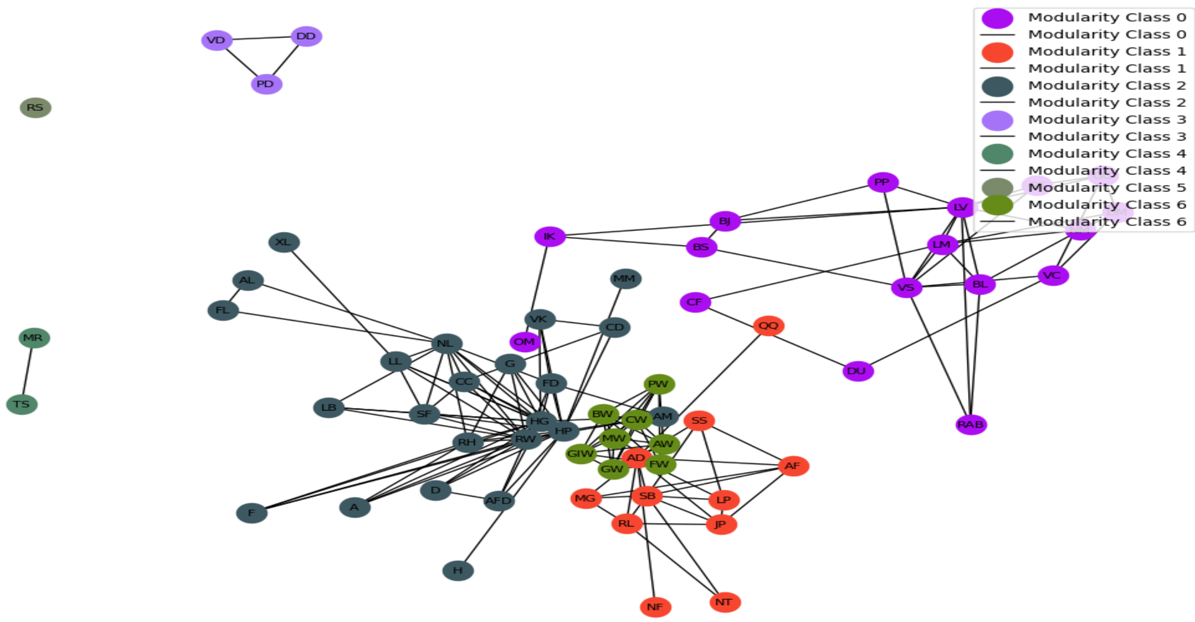

Figure. 8

This visual representation provides a comprehensive view of the entire network, highlighting the different modularity classes with distinct colors. The use of character names as labels enhances the readability and understanding of the network.

The visualization of the network as a connected graph, with nodes belonging to the same communities sharing the same colors. The graph demonstrates how characters are connected within the entire network, with nodes of the same modularity class sharing a common color. It offers a clear visualization of the social structure within the Wizarding World, showcasing the communities and connections between characters in figure. 9.

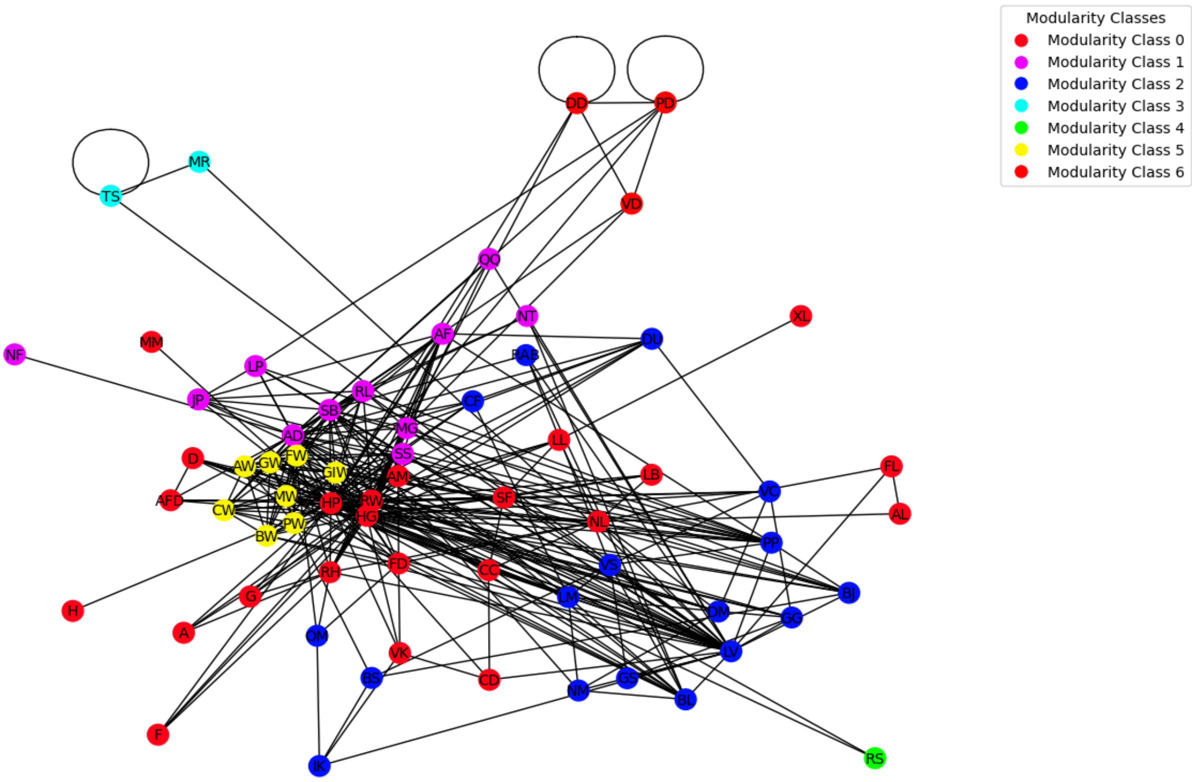

Figure. 9 (Complete Interconnected Community Graph)

CONCLUSION

The social network analysis of the Harry Potter movie series, as portrayed in the provided dataset, has unveiled a multitude of insights into the intricate web of relationships among the wizarding characters. Through various network analysis metrics and visualizations, we have gained a deeper understanding of the social structure within the Wizarding World.

The analysis explored centrality measures, including degree centrality, betweenness centrality, and closeness centrality. These metrics provided valuable insights into the characters' prominence and influence within the network. The identification of characters with high degree centrality, betweenness centrality, and closeness centrality underscores their pivotal roles in connecting and bridging different parts of the Wizarding World.

The network's clustering coefficients were examined, both at the global and local levels. The global clustering coefficient indicated a high level of interconnectedness within the network, suggesting that characters tend to form tightly knit groups. The local clustering coefficients for individual nodes demonstrated the extent to which characters tend to form connections with each other. These coefficients highlighted characters who are central to forming connections within their communities.

The analysis revealed a network with a modularity score of 0.381. This indicates that the network is moderately modular, with characters forming distinct communities or social groups. The modularity classes identified through the analysis offer a fascinating glimpse into the complex interactions between characters. The characters are not randomly connected; instead, they form distinct communities, each with its own set of interactions and relationships.

Subsequent analysis involved the creation of subgraphs for each modularity class. This allowed us to investigate deeper into the communities. The results showed variations in the size and complexity of these communities. Some communities were tightly knit, with a larger number of nodes and edges, while others were smaller and more isolated. This suggests that the Wizarding World is not a huge society, but a collection of smaller, interconnected groups, each with its own dynamics.

Visualizations played a crucial role in comprehending the network's structure. By creating visual representations of subgraphs and the entire network, we gained a clearer picture of how characters are connected. These visualizations illustrated that characters within the same community share more connections, resulting in denser subgraphs. The use of character names as labels enhanced the readability of the network, making it easier to identify and understand the relationships.

Finally, the analysis visualized communities as connected networks, with nodes sharing the same modularity class assigned the same color. This visual representation effectively conveyed the intricate social structure of the Wizarding World, showcasing the interconnected communities.

In conclusion, the social network analysis of the Harry Potter movie series has illuminated the presence of distinct communities and the characters' roles within them. The network is not a huge entity but a collection of interconnected social groups, much like the complex and multifaceted world depicted in J.K. Rowling's novels. The analysis not only provides valuable insights into the relationships among characters but also serves as a testament to the depth and richness of the Wizarding World's social fabric.

This analysis may serve as a foundation for further investigations into the dynamics of the Harry Potter universe, shedding light on the complexity of interpersonal relationships and the underlying structure of this fictional world.

REFERENCES:

1. GitHub Repository: Potter Network
2. Harry Potter - Wikipedia
3. List of Harry Potter cast members | Harry Potter Wiki | Fandom