Potterverse

Nikhil Chalakkal, Chengyao Ma, Ryan Chi, and Pratik Keni

ISOM 673 - Social Network Analytics

Final Project Report

Premise

The Harry Potter series has been a resounding success both as books and movies, and one of the essential parts of this magical world is how J.K Rowling vividly depicts each character with various background settings and personality designs. In her writings, different beasts can communicate in fantastic ways, wizards and muggles are bonded by blood, and a vast magical network is growing and even integrated into the real world.

As a part of this project, we explore the interpersonal relationships between the major characters. By doing so, we are hoping to achieve two goals:

1. Explain why specific plots and relationships develop that way.

2. Explore what groups or clusters the characters truly stand for.

Aside from the information from the books and movies, we will try and understand this complex network through social network analysis. At the end of this project, we are hoping to gain some insights from an analytical point of view, to see how our results differ from traditional story plots. If you haven't read the books or watched the movies, please note that there are spoilers ahead!

Data Details

Collection

We tried to see if data was available online for this, and we got a dataset which had details of the network in the Harry Potter Universe. This was obtained from GitHub, from

user Efe Karakus. He and a few others had built and cleaned this data for a personal project to see if the network exhibited the same properties as the real world. The data was downloaded from here and consisted of a list of 515 edges between 65 nodes. The data was undirected, but the edges were classified as friendly or unfriendly. Snapshots of the data set are given below:

Node (Character) List:

ID	Name	Type	For Edges Bio I
0	Regulus Arcturus Black	Pre-Book	Brother of Sirius. Used to be a Death Eater but defected.
1	Sirius Black	Wizard	Best friend of James Potter and godfather of Harry.
2	Lavender Brown	Wizard	Killed by a werewolf. She was a gryffindor student who dated Ron.
3	Cho Chang	Wizard	Ravenclaw student who dated Cedric Diggory and Harry Potter.
4	Vincent Crabbe, Sr.	Wizard	Father of Crabbe and death-eater who escaped Azkaban.
5	Vincent Crabbe	Wizard	Slytherin student who was best friends with Goyle and followed Draco.
6	Bartemius (Barty) Crouch, Sr.	Wizard	Head of the department of Internation Magical Cooperation. Killed by his son.
7	Bartemius (Barty) Crouch, Jr.	Wizard	Death Eater who impersonated Alastor Moody.
8	Fleur Delacour	Wizard	Participated in the Triwizard tournament and married Bill Weasley.
9	Cedric Diggory	Wizard	Participated in the Triwizard tournament and got killed by Voldemort.
10	Aberforth Dumbledore	Wizard	Albus' brother and owner of Hog's Head.

Edge List:

Original				Without Duplicates				Final Edge List				Directed Edge List		
Source	Target	Type		Source	Target	Type		Source	Target	Type		Source	Target	Type
0	1	-		9	3	+		9	3	+		0	1	-
0	25	-		9	39	+		9	39	+		0	25	-
0	45	-		9	45	-		9	45	-		0	45	-
1	0	-		0	1	-		0	1	-		0	4	+
1	11	+		10	11	+		10	11	+		1	0	-
1	21	+		10	21	+		10	21	+		1	11	+
1	25	-		10	39	+		10	39	+		1	21	+
1	31	+		10	58	+		10	58	+		1	25	-
1	33	-		10	59	+		10	59	+		1	31	+
1	34	-		11	1	+		11	1	+		1	33	-

Affiliation Matrix (Characters and Potterverse Groups):

ID	Dumbledore's Army	Death Eater	Ministry of Magic	Staff	Student	Quidditch Player	Order of the Phoenix I	Order of the Phoenix II
0	0	1	0	0	0	0	0	0
1	0	0	0	0	0	0	1	1
2	1	0	0	0	1	0	0	0
3	1	0	0	0	1	1	0	0
4	0	1	0	0	0	0	0	0
5	0	1	0	0	1	1	0	0
6	0	0	1	0	0	0	0	0
7	0	1	0	0	0	0	0	0
8	0	0	0	0	1	0	0	1
9	0	0	0	0	1	1	0	0
10	0	0	0	0	0	0	1	1

Cleaning

This data was cleaned up to get the dataset that we used - there were duplicates in the data, as well as edges that had two different attributes in different directions, which was

incorrect given that the data was undirected. After all these details were removed, we got a final dataset having 456 edges between 65 nodes. This was the dataset we initially used.

Further Data Collection

In the process of working on the data and getting different results, we realized that the data was not sufficient - there were a lot of characters and edges missing, which was resulting in very counterintuitive and skewed results. For example, Harry and Voldemort had only 9 and 8 connections respectively, whereas Dumbledore had 36, and Ron Weasley had 34.

To get better results, we took inspiration from Efe Kakarus and decided to build our own dataset. We did this by getting data off the Harry Potter Characters Wiki Page. This webpage gave us an exhaustive list of characters in the series, which came up to a total of 199 characters, along with their brief bios as well.

Further Data Cleaning

This data was cleaned up further to build a second data set that augmented the first one. During the cleaning process, animals, ghosts, magical creatures, painting names, historical figures, and characters not featured in the duration of the books were removed - there were some essential people who were retained (Frank and Alice Longbottom, Lily and James Potter) though. Once this was done, we had a total of 152 characters to work with.

Developing an edge list for 152 characters would have taken weeks (152 * 152 = ~ 23k edges) if it was done manually. So, we did something different – a shortcut, if you will – we created a matrix for all 152 characters with affiliations to 10 different groups. We used this affiliation matrix to determine relationships between different actors - how many different groups they were co-membered in as an indicator of a relationship between them.

So finally, we used the following different datasets:

1. Initial (found) dataset of 456 edges between 65 nodes with edge attributes of friendly and unfriendly relationships.

2. Final (created) dataset of 2249 edges between 124 nodes, created from a 124 x 124 element square co-membership matrix of 124 characters.

3. (Created) Affiliation matrix of 152 characters as members of 12 different groups.

Analysis Results

Data Processing

As mentioned earlier, we created a dataset with affiliations to ten different groups. We also classified the actors as dead or alive (by the end of the book), so in total, we had an affiliation matrix for 152 characters and 12 groups. These groups (along with the number of members in them) were:

Sl No	Group	Members
1	Dumbledore's Army	25
2	Death Eater	23
3	Ministry of Magic	18
4	Hogwarts Staff	26
5	Hogwarts Students	46
6	Quidditch Player	20
7	Order of the Phoenix I	17
8	Order of the Phoenix II	29
9	Inquisitorial Squad	5
10	Auror	9
	Dead	31
	Alive	121

However, given that the Inquisitorial Squad and Auror groups did not have many members, we excluded them from the analysis, and focused on the 8 larger groups. Also, for easier visualization, all the networks were built on the three largest and most distinct groups - Students, the Orders of the Phoenix, Death Eaters - and all other groups classified as Others.

Network Structure

We visualized the smaller network (456 edges x 65 nodes) that we had, classified by the groups that the nodes were a part of, and the edge attributes - friendly or unfriendly. The friendship network has green edges (as shown above), and the enemy network contains red edges (as shown below). Blue nodes represent the Hogwarts Students group, orange nodes

represent members of the Orders of the Phoenix, green nodes represent the Death Eaters, and gray nodes represent other groups, such as the Ministry of Magic, Hogwarts staff, etc.
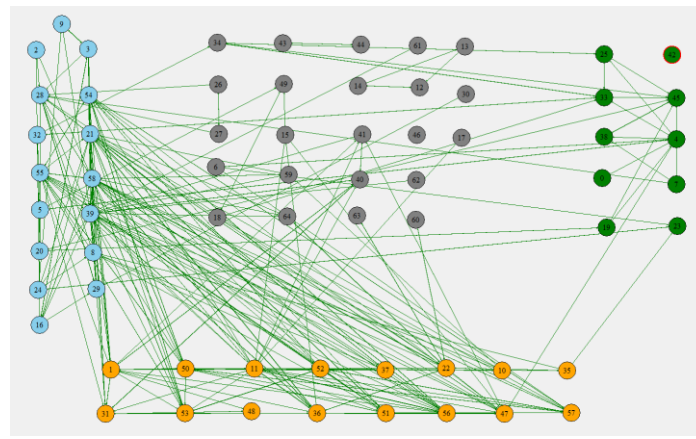

At first glance of the friends' network, students are mostly interconnected, as well as linked with Order of the Phoenix members, which makes sense since they were aligned groups in the books. Also, as expected, neither of these groups have a lot of friendly ties with the Death Eaters, who were the enemies in the Potterverse.

Unlike the friends' network, which is denser on one side between two groups, the enemies are more concentrated from the Death Eaters. As seen (and as was observed in the book), they were universal enemies, which is shown in the graph as well. On the contrary, only a few Students and members of Orders of Phoenix establish the enemy relationship, which is again as per expectation, given their characters were aligned in the book.
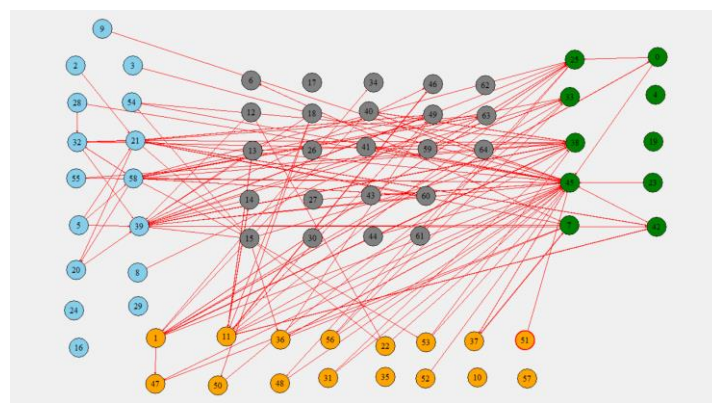
We also visualized the larger network (2249 edges x 124 nodes) that we had, classified by the groups that the nodes were a part of. As expected, this is a much denser

network, given there are a lot more edges involved. However, the groups all remain the same as before, but they're shown as tighter clusters in this network. This is expected, given that this network was built from an affiliation matrix. So, the different groups that people were affiliated to are shown as tight clusters. This network is shown below:
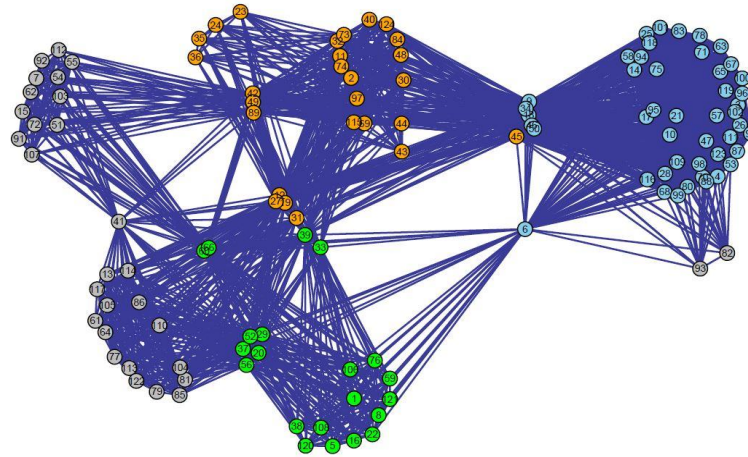
Central Figures in the Network

We used the network combining both friend and enemy relationships to figure out the top 10 central figures in the Harry Potter network. Three methods were used to show the centrality in different perspectives, based on three iGraph built-in functions:

Highest Closeness	Highest Betweenness	Highest Eigen Centrality
Harry Potter	Ron Weasley	Ron Weasley
Ron Weasley	Albus Dumbledore	Hermione Granger
Hermione Granger	Hermione Granger	Harry Potter
Lord Voldemort	Lord Voldemort	Albus Dumbledore
Albus Dumbledore	Harry Potter	George Weasley
Sirius Black	Sirius Black	Lord Voldemort
Ginny Weasley	Rubeus Hagrid	Ginny Weasley
Fred Weasley	Minerva McGonagall	Fred Weasley
George Weasley	Alastor Moody	Molly Weasley
Molly Weasley	Seamus Finnigan	Sirius Black

Interestingly, the titular characters (Harry Potter and Lord Voldemort) are not the central characters in the universe. The closeness centrality results are the most intuitive. Harry has the highest closeness – which makes sense, given that he can reach almost everyone in the Potterverse – followed by other titular characters – Harry's best friends, Ron and Hermione, the main antagonist, Lord Voldemort, and Harry's mentor and guide, Albus Dumbledore.

However, the betweenness centrality shows Harry quite low down – this could be because given Ron's huge family, and Dumbledore's age and huge number of connections, they are more central when it comes to bridging different groups. Eigen Centrality is also intuitive – given Ron and Hermione are connected to the most important characters in the book – Harry, Lord Voldemort, and Dumbledore, they are the most important by association.

However, if you look at all the measures of centrality, it seems that Ron and Hermione are easily more central to the Potterverse than Harry Potter himself. This result, while slightly counterintuitive, seems possible given that the two of them are Harry's best friends and connected very closely to all the titular characters in the book.

Co-membership Matrix & Attributes

We used only 7 different groups out of 10 (Dumbledore's Army, Death Eaters, Ministry of Magic, Hogwarts Staff, Hogwarts Students, Quidditch Players, and the Orders of the Phoenix) to determine these relationships, as the other groups either did not make sense (dead or alive) or did not have enough actors affiliated to them. This gave us a total of 124 characters in the affiliation matrix, which was then converted into a co-membership matrix. This matrix, so created, had the following properties:

   a. Highest Degree Centrality: Severus Snape, Ron Weasley and George Weasley with 4 affiliations each.

   b. Participation Rate: 1.63 groups / character – this is a measure of how active the characters are.

   c. Largest value of n for an n-clique: 46 members.

   d. Member of maximum cliques: Severus Snape, a member of 9 cliques.

Similarity Measures Between Different Groups

For the 12 groups identified above, we computed 2 similarity measures – the Cosine Similarity and the Jaccard Similarity – between all of them, to numerically verify how the

characters overlapped among the groups. The results are in the 2 matrices below:

Cosine Similarity:

	Dumbledore's Army	Death Eater	Ministry of Magic	Staff	Student	Quidditch Player	Order of the Phoenix I	Order of the Phoenix II	Dead	Alive	Inquisitorial Squad	Auror
Dumbledore's Army	0	0	0	0	0.70772139	0.31304952	0	0.14855627	0.10776318	0.4	0	0
Death Eater	0	0	0.14744196	0.24535825	0.03074377	0.04662524	0.1011443	0.07744031	0.26215206	0.3032937	0.09325048	0
Ministry of Magic	0	0.14744196	0	0.04622502	0	0	0	0.13130643	0.1693335	0.2999847	0	0.1571348
Staff	0	0.24535825	0.04622502	0	0	0	0.1902606	0.18208926	0.14089399	0.3922323	0	0
Student	0.7077214	0.03074377	0	0	0	0.5604734	0	0.13689642	0.10592543	0.5629602	0.32969024	0
Quidditch Player	0.3130495	0.04662524	0	0	0.5604734	0	0	0.12456822	0.1204829	0.3455741	0.3	0
Order of the Phoenix I	0	0.10114435	0	0.1902606	0	0	0	0.54045282	0.30492479	0.2204869	0	0.404226
Order of the Phoenix II	0.1485563	0.07744031	0.13130643	0.18208926	0.13689642	0.12456822	0.5404528	0	0.16675934	0.4051535	0	0.1856953
Dead	0.1077632	0.26215206	0.1693335	0.14089399	0.10592543	0.1204829	0.3049248	0.16675934	0	0	0.08032193	0.2993422
Alive	0.4	0.30329369	0.2999847	0.39223227	0.5629602	0.34557414	0.2204869	0.40515347	0	0	0.16262313	0.1212121
Inquisitorial Squad	0	0.09325048	0	0	0.32969024	0.3	0	0	0.08032193	0.1626231	0	0
Auror	0	0	0.15713484	0	0	0	0.404226	0.18569534	0.29934217	0.1212121	0	0

Jaccard Similarity:

	Dumbledore's Army	Death Eater	Ministry of Magic	Staff	Student	Quidditch Player	Order of the Phoenix I	Order of the Phoenix II	Dead	Alive	Inquisitorial Squad	Auror
Dumbledore's Army	0	0	0	0	0.5106383	0.18421053	0	0.08	0.05660377	0.177419	0	0
Death Eater	0	0	0.07894737	0.13953488	0.01470588	0.02380952	0.05263158	0.04	0.14893617	0.125	0.03703704	0
Ministry of Magic	0	0.07894737	0	0.02325581	0	0	0	0.06818182	0.08888889	0.112	0	0.08
Staff	0	0.13953488	0.02325581	0	0	0	0.1025641	0.1	0.0754717	0.176	0	0
Student	0.5106383	0.01470588	0	0	0	0.34693878	0	0.07142857	0.05479452	0.336	0.10869565	0
Quidditch Player	0.18421053	0.02380952	0	0	0.34693878	0	0	0.06521739	0.0625	0.137097	0.13636364	0
Order of the Phoenix I	0	0.05263158	0	0.1025641	0	0	0	0.35294118	0.17073171	0.078125	0	0.23809524
Order of the Phoenix II	0.08	0.04	0.06818182	0.1	0.07142857	0.06521739	0.35294118	0	0.09090909	0.190476	0	0.08571429
Dead	0.05660377	0.14893617	0.08888889	0.0754717	0.05479452	0.0625	0.17073171	0.09090909	0	0	0.02857143	0.14285714
Alive	0.17741935	0.125	0.112	0.176	0.336	0.13709677	0.078125	0.19047619	0	0	0.03278689	0.03174603
Inquisitorial Squad	0	0.03703704	0	0	0.10869565	0.13636364	0	0	0.02857143	0.032787	0	0
Auror	0	0	0.08	0	0	0	0.23809524	0.08571429	0.14285714	0.031746	0	0

Based on our knowledge of the characters and groups within the Potterverse, both the similarity coefficients seem to be accurate. The values highlighted in orange in both matrices above illustrate their validity within the Potterverse:

1) Student – Dumbledore's Army (Cosine Similarity 0.70, Jaccard Similarity 0.51): Dumbledore's Army only consisted of Hogwarts students, but not all Hogwarts students were members of Dumbledore's Army.

2) Death Eater – Dumbledore's Army (Cosine Similarity 0.0, Jaccard Similarity 0.0): Death Eaters and Dumbledore's Army are two key opposing groups within the plot, and there's no overlap between the members of both groups.

3) Student – Staff (Cosine Similarity 0.0, Jaccard Similarity 0.0): None of the Hogwarts students are also staff members, and vice versa.

4) Quidditch Player – Student (Cosine similarity 0.56, Jaccard Similarity 0.34): All Quidditch players at Hogwarts are students, but not all students play Quidditch for their respective House teams.

5) Order of the Phoenix II – Order of the Phoenix I (Cosine Similarity 0.54, Jaccard Similarity 0.35): There is some overlap between the Orders, but not 100%, because some members of the first order were killed by Voldemort.

6) Alive – Dead (Cosine Similarity 0.0, Jaccard Similarity 0.0): Obvious reasons – all characters that are alive by the end of the story have clearly not died, and all characters who are dead won't be coming back to life!

Given this, it can be inferred that the Cosine and Jaccard similarities provide valid measures of similarities between different groups in the Potterverse. Interestingly, all Jaccard coefficients are generally smaller (in value) than all corresponding Cosine coefficients.

Clustering
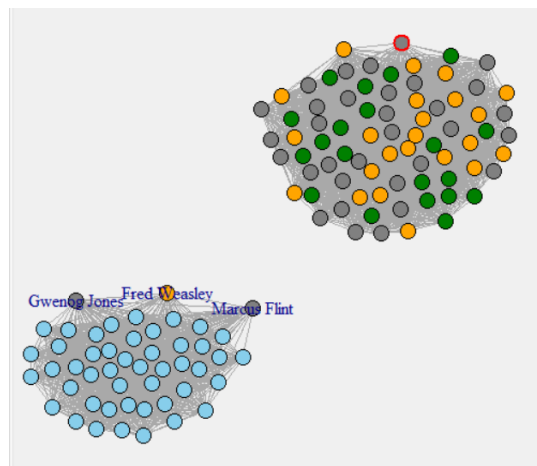
We conducted a correlational clustering based on the affiliation information of 124 characters (as in the co-membership matrix). We build the adjacency matrix from the co-membership matrix, and by doing 8 iterations, have a list of characters that end up in two clusters. For our clustering algorithm, we considered the 8 different groups from the co-membership matrix. Once done, we did a visualization – and as before, color of vertices here stands for four main groups: Death Eaters, Students, Orders of the Phoenix, and Others.

From the visualization we can see that the group of Students is quite cohesive – they're all in the same cluster. Most of other three groups of characters mingle with each other in the other cluster. This can be explained as follows: Hogwarts Students clearly have a

high similarity with Dumbledore's Army and Quidditch Players, which we know are both students-only groups. In addition, the Students have extremely low similarity (between 0 and 0.13) with all other groups, which we know are primarily composed of non-students (as seen in the Cosine and Jaccard similarity matrices). Given that all the groups which consist of students do not at all mingle with the other groups, it pre-empts that the students will be in a completely different cluster as compared to the other groups, and this is validated by the visualization. The second cluster, expectedly, proves to have groups that don't mingle with students or with groups having students – Death Eaters, Orders of the Phoenix, and Others.

There are just three exceptions – Marcus Flint and Gwenog Jones, who do not belong to any other affiliations, are senior members of the Quidditch team. Another exception, Fred Weasley is a member of Dumbledore's Army and a Quidditch player as well. According to the dataset, he is also a member of the Order of the Phoenix, which is not the case: he should be classified as a student. Hence, the clustering algorithm here even helps us to identify possible flaws in the dataset.

Conclusion

In conclusion, using analysis of the networks of the Potterverse, we have been able to explain, with some success, why characters were aligned the way they were in the books. We have also got some counterintuitive results about the central characters, and seen how different groups interact, and how similar they are to each other.

We have been a little limited by availability of data and feel that with more robust data (which will need to be created!) we can do a more in-depth analysis and come up with some deeper insights on how people in the Potterverse interact with each other.