Assignment 1

Hanniu Sun & Thomas Haase

February 7, 2025

Table of contents

Task 1 **2**

Task 2 **3**

Task 3 **4**

Task 4 **5**

Task 5 **6**

Task 6 **8**

Task 7 **9**

Task 8 **10**

Bibliography **11**

Dataset

The chosen data contains information of support between 64 characters in the Harry Potter books, made available by Goele Bossaert and Nadine Meidert. It was downloaded on the RSiena webpage.

Task 1

> *Write an R-script that reads your data. Transform your data into a format that you will uselater. Argue about the format you have chosen.*

The used data is accessable in the format of adjacency matrices. For each book the directed network of the support of one character for another is represented in the form of an adjacency matrix. Besides the 6 matrices 2 tables with the characternames and some characterattributes (schoolyear, gender, house) are made available.

For this assignment the available data of all books is going to be combined into one dataset. To achieve this the adjacencymatrices are added to create a weighted adjacency matrix of the support that the characters give each other throughout the whole story from book one until book six.

In the article of the dataset support between the characters is defined as emotional support, instrumental help, giving information that will help others and praising each other (Goele Bossaert 2013). Based on this self-connections are nonsensical as they do not represent how characters help each other. Therefore the diagonal is going to be removed.

Task 2

Justify what you do with isolates and multiple components if there are any.

An isolate is a node that is not connected to other components. A component is a connected subgraph that is not part of any other connected subgraph. A component is weakly connected if nodes are reachable without taking the direction of edges into account. In contrast a strongly connected component is reachable even when taking the direction of edges into account.

```
The graph contains 27 isolated vertices.
```

Component Summarystatistics
Without Isolates

Statistic	Weakly Connected	Strongly Connected
N components	2	7
Min. size	5	2
Mean size	18.50	11.00
Max. size	32	26

The high number of isolated vertices distorts the summarystatistics as most of them seem to consists of only a singular node. Therefore they have been filtered out for the calculation of the summarystatistics. We do not delete them from the graph because it is valuable for interpreting to keep the high number of isolated characters in mind.

The weakly connected portion of the network contains 2 components. Within these components, the size varies considerably, with the smallest component having 5 characters and the largest having 32 characters. The strongly connected portion contains 7 components. These components tend to be smaller, with sizes ranging from 2 to 26 characters, and an average size of 11 characters per component. The support structure of Harry Potter therefore consists of several tightly-knit groups (the strongly connected components) and a couple of larger, more loosely connected groups (the weakly connected components). The strongly connected components likely represent close-knit groups of characters who consistently support each other, while the weakly connected components might represent broader social circles with less consistent mutual support.

Task 3

Report and interpret the density of your network. Create a table that contains further descriptive network statistics. Please include average degree (in-degree and out-degree if both are relevant), standard deviation of degree (in-degrees and out-degrees), reciprocity, and transitivity.

Density Summarystatistics

Parameter	Value (AVG ± SD)
Density	0.05
In-Degree	3.1 ± 4.6
Out-Degree	3.1 ± 4.2
Reciprocity	0.85
Transitivity	0.53

Network density is defined as the proportion of actually observed ties among the potentially observable ones. The observed density of ~5% indicates a sparsely connected network. Support therefore seems to be really selecive in the Harry Potter universe. An explanation is the high amount of isolate nodes which are probably the main reason for the low observed density.

The average in- and out-degree are the same because for every character who gives support another character needs to be the receiving support, otherwise no support could have been given in the first place. The higher in-degree standard deviation shows a higher spread in the distribution of support characters are receiving. There are some characters receiving no support and some who receive support from many characters. The standard deviation of out-degree is a bit lower than the standard deviation of in-degrees, indicating underlining that more characters give similar amount of support while some characters receives more support than others.

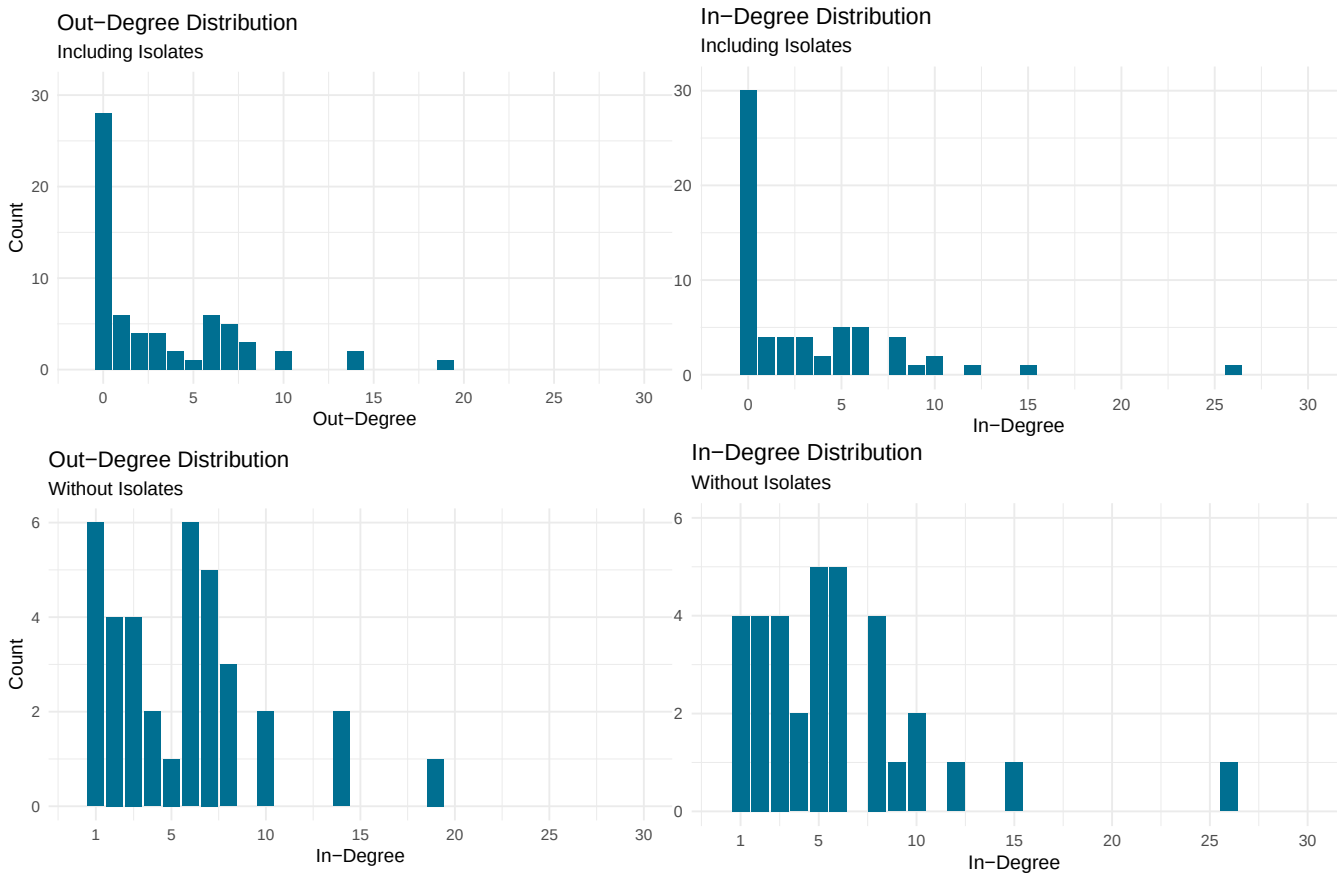
Reciprocity is defined as the proportion of actually reciprocated ties among the potentially reciprocable ones. The reciprocity value of 0.85 indicates that support is usually mutual in the Harry Potter universe, suggesting strong bonds between connected characters. This makes sense because the focus of the books lies on a view main characters, on whose interaction all 6 books are focusing.

The transitivity is the proportion of actually observed transitively closed triples of nodes among the count of all potentially closed triples in the network. It shows how likely a character B who gets support by character A is willing to support a supporter of character A. The chance of this happening is ~53%.

These metrics paint a picture of a network where support relationships, while selective, tend to be strong and mutual when they do exist, with a moderate clustering among connected characters. This aligns well with the narrative structure of Harry Potter, where distinct groups of a few main and secondary characters forming strong, reciprocal support systems.

Task 4

Create a graph of the degree (out-degree and in-degree if they are relevant) distributions in your network. What do you observe? What do you imply from the observed degree distributions?



Degree distribution is defined as the probability distribution of degrees across all nodes in the network. The high count of nodes with degree 1 in the upper plot are an artifact of the high amount of isolates in the network. Therefore the lower plots are being created. There are 3 different groups visible in the Out-degree distribution. Group one consists of characters supporting less than 5 persons, Group two consists of characters supporting between 5 and 10 characters while three characters support more than 10 characters. The In-degree distribution looks more even. There are almost all the time 4 charaacters who get support from 1-10 characters. Outliers are also visible here, especially one extreme one that gets support from over 25 characters.

In conclusion the network seems to have a core group of mutually supporting characters where a few characters (likely the main protagonists) serve as major support hubs. Many peripheral characters exist who have minimal involvement in the support network (filtered isolates and high amount of vaguely connected characters).

Task 5

Create a formatted table that contains descriptive information for node level variables.

Table 1: Descriptive Statistics of metric Harry Potter Character Variables

	Mean	Median	SD
Indegree	5.43	5.00	4.95
Outdegree	5.43	6.00	4.18
Eigenvector	0.20	0.05	0.27
Betweenness	0.01	0.00	0.02

Centralitymeasures are normalized
Isolates removed

Table 2: Descriptive Statistics of nominal Harry Potter Character Variables

		N	Percent
Gender	Male	22	59
	Female	15	41
House	Gryffindor	21	57
	Hufflepuff	6	16
	Ravenclaw	5	14
	Slytherin	5	14

Isolates removed

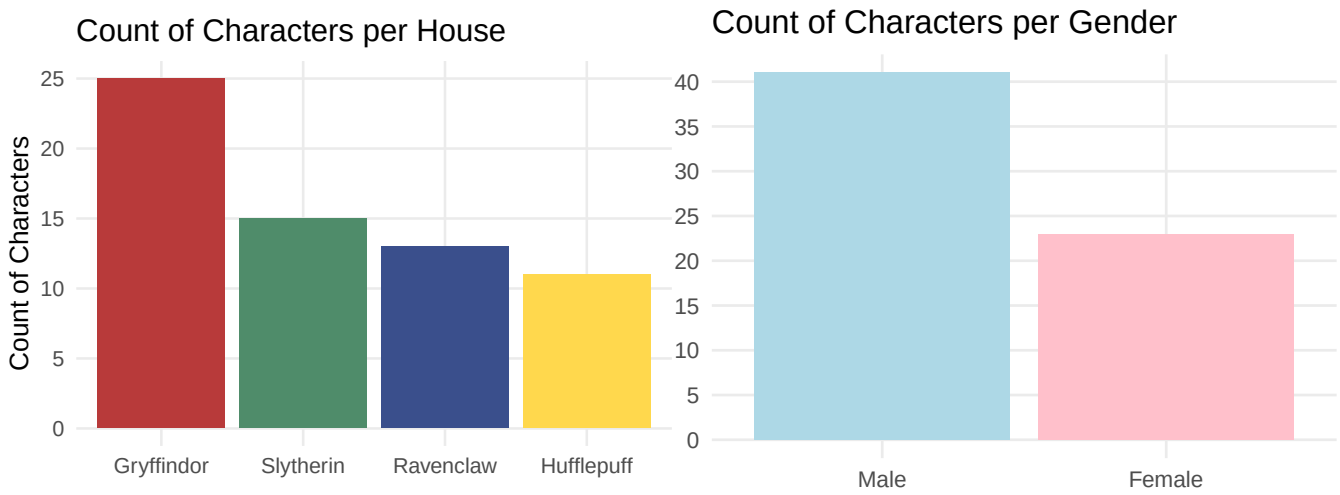
Table 1 presents key network metrics that describe how characters support one another. The indegree and outdegree measures both average 5.43, but show substantial variation with standard deviations of 4.95 and 4.18 respectively. This indicates that while characters typically have around 5-6 supportive connections, there is considerable spread - some characters may have very few connections while others maintain extensive support networks.

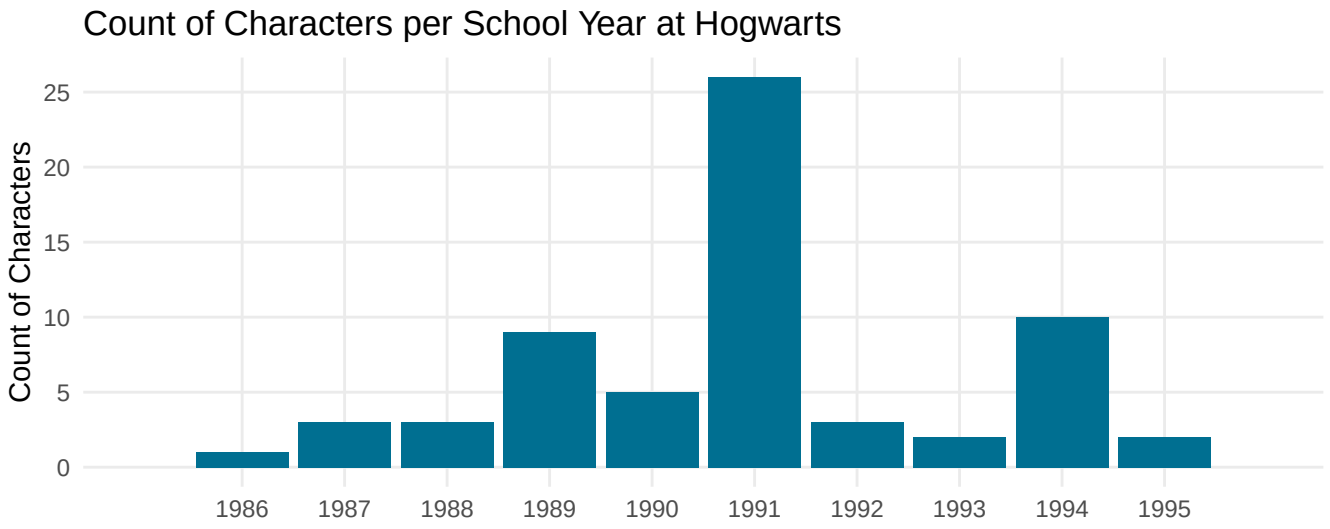
The eigenvectorcentrality weights the degree measures with the degree measures of the directly connected nodes. The eigenvector centrality measures show a mean of 0.20 with a standard deviation of 0.27. The standard deviation being larger than the mean indicates significant variation in characters' influence within the network, suggesting the presence of both highly central figures and more peripheral characters.

Betweennesscentrality measures brokerage by the sum of fractions of shortest paths between any 2 nodes that pass through a given node. The betweenness centrality in the network displays a very low mean of 0.01 with a small standard deviation of 0.02, indicating that across the network, very few characters serve as crucial bridges between different groups - most support relationships appear to be direct rather than requiring intermediaries.

Table 2 provides demographic information about the character network. The gender distribution is relatively balanced, with 59% male and 41% female characters (keep in mind that the isolates are removed for these statistics!). Regarding house affiliations, Gryffindor represents the largest group at 57% of characters, which aligns with the books' narrative focus. The remaining houses - Hufflepuff (16%), Ravenclaw (14%), and Slytherin (14%) - have roughly equal representation in the support network.

The following plots shed light on demographic information including isolates.


Count of Characters per House
Count of Characters
25
20
15
10
5
0
Gryffindor
Slytherin
Ravenclaw
Hufflepuff
Count of Characters per Gender
40
35
30
25
20
15
10
5
0
Male
Female

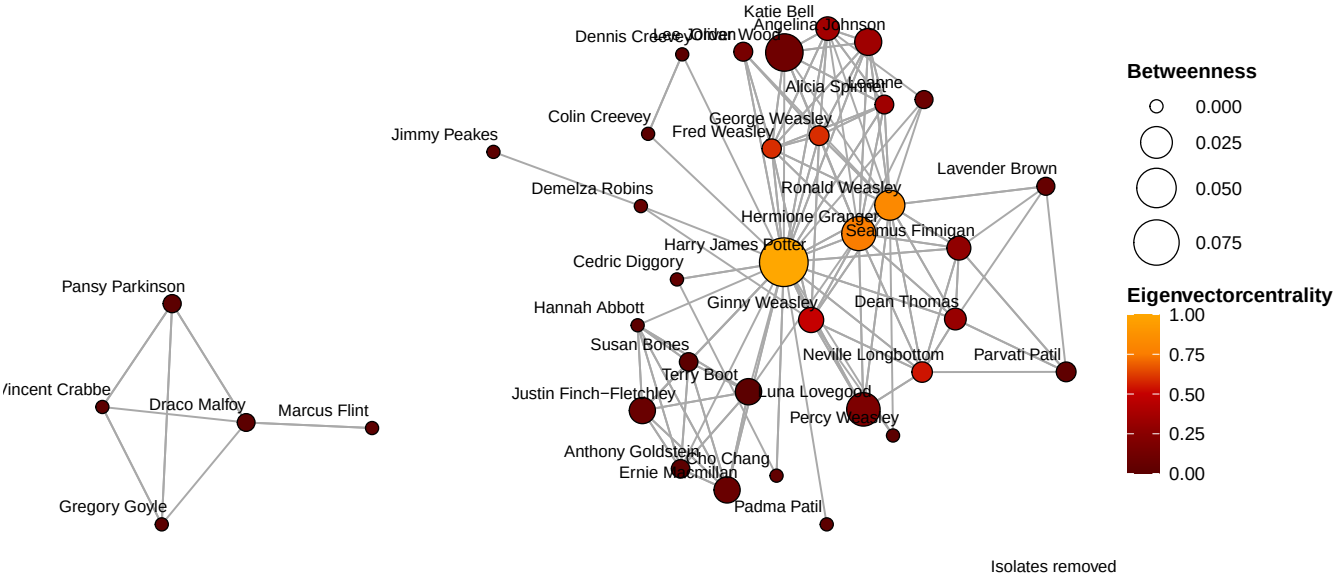

Count of Characters per School Year at Hogwarts
Count of Characters
25
20
15
10
5
0
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995

Task 6

Taking into consideration network size, select a form of visualization that could be meaningful. Visualize your network and color the nodes according to a selected actorvariable or according to a selected measure of centrality.

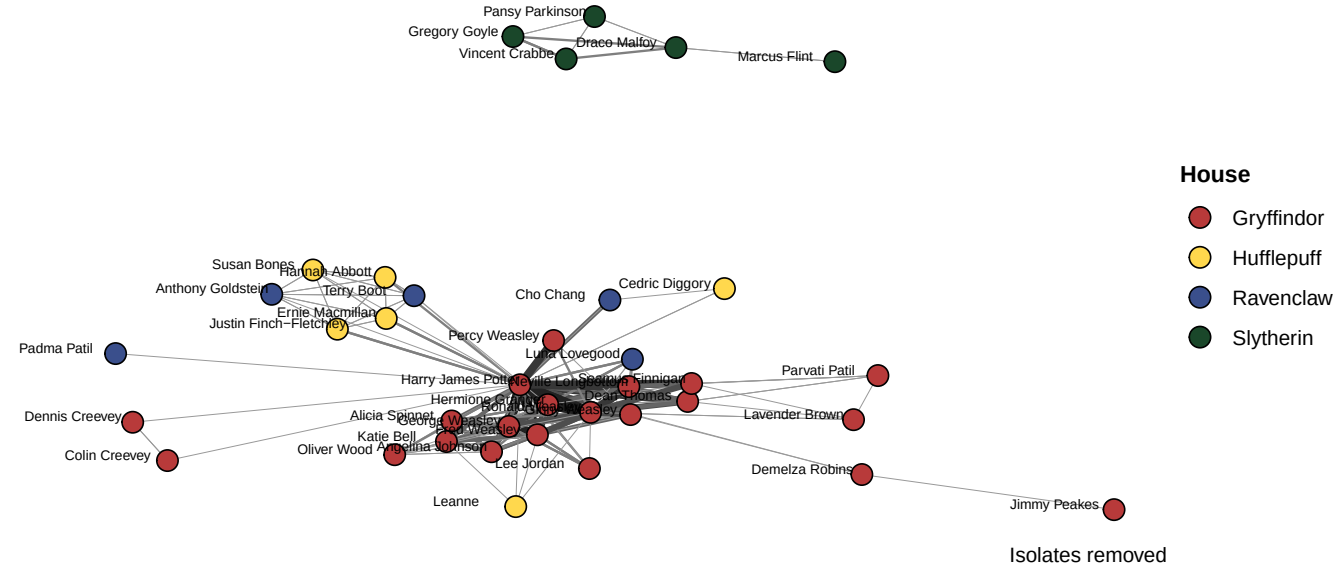
Support of Harry Potter Characters for each other

By Measures of Centrality



Support of Harry Potter Characters for each other

By Houses, Edges colored by Weight



Task 7

Check assortativity in the network by a chosen measure of degree centrality. What do you observe? Provide an interpretation of your findings. What kind of theoretical arguments could possibly explain your results?

"Biggest man on Campus" based on different criterias

Top Closeness Centrality	Top Betweenness Centrality	Top Pagerank
Harry James Potter	Harry James Potter	Harry James Potter
Ronald Weasley	Oliver Wood	Ronald Weasley
Hermione Granger	Hermione Granger	Hermione Granger

Assortativity

Degree	Pagerank	Gender	House	Schoolyear
−0.121	−0.139	−0.098	0.709	0.329

Closeness Centrality is a metric in network analysis used to measure the average distance between a node and all other nodes, it reflects the efficiency with which the node can transmit information or influence others across the entire network. In this case, we can see from the table that only less than 5 nodes have high closeness centrality, a major part of the nodes barely have the closeness centrality, indicating that these characters are far from the main stream of the Harry Potter novel, they do not even have connections to Harry Potter.

Betweenness Centrality is a metric in network analysis that measures the importance of a node as a bridge or intermediary, reflecting how many pairs of nodes have shortest paths passing through it. If a node lies on the shortest paths between many pairs of nodes, its betweenness centrality significantly increases. And from my perspectives, betweenness centrality describes nodes as a metro stations, highly valued stations can be regarded as the central stations or popular transfer stations. As the histogram shows, approximately 90 percent nodes fall on the value 0, which means absolutely disconnected with other. There are several observations fall on the right side, they might act like hubs in this social network.

The core idea of PageRank is: A node's importance depends not only on how many other nodes link to it but also on the importance of those linking nodes themselves. It assumes weak connectivity between disconnected components, thereby addressing inconsistent ranking metrics in non-fully connected networks.the histogram shows that a few core nodes dominate the network.

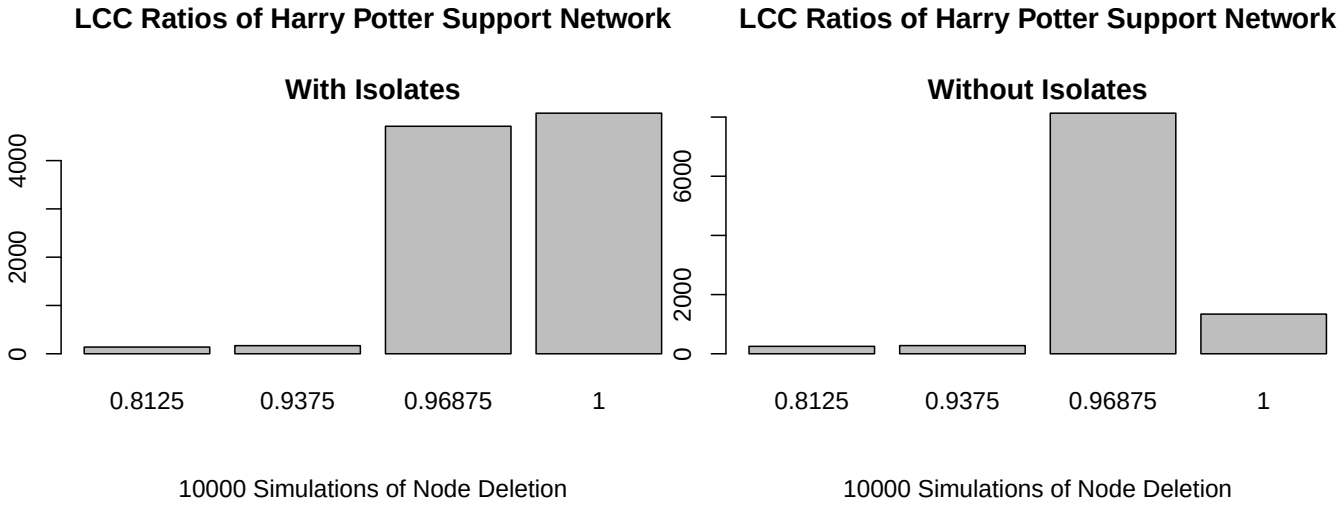
We list top 3 nodes of each centrality calculation, it seems that Harry James Potter and Hermione Granger are the most important nodes in this network.

Assortativity is a measure of how similar nodes in a network are to their neighbors based on some attribute. It tells us whether similar nodes tend to connect with each other or not. Here we calculated 5 kinds of assortativity: 1.degree 2.pagerank 3.gender 4.house 5.schoolyear. The result illustrates that: 1.characters are not likely to team up according to the degree, pagerank and gender attribute,we can say that there is no gender segregation within the network. 2.assortativity based on house is the highest among 5 assortativities, 0.709, which is approching 1, indicating that characters are prone to group up due to the

house factors, people intend to have connection with people from same house. 3.assortativity based on schoolyear factor is 0.329, denoting that people have a tendency of teaming up with people who are in the same schoolyear, but the tendency is not as significant as the house one, however, stronger that degree and pagerank and gender.

Task 8

And finally, a thought exercise. Assume that the robustness or vulnerability of the network is examined. Try to come up with a measurement of robustness / vulnerability and argue for the usefulness of your measure. Speculate about some implications for the concrete network.



In this part we evaluate the robustness of Harry Potter network with the method of simulating the random removal of nodes. After a random node is removed we calculate the sizte of the largest connected component. To get a ratio of change we divide this size by the largest connected component of the graph with the removed node by the size of the largest connected component of the original graph. This procedure is repeated 10000 times. The calculated ratio is a measure of grpah stability. In terms of the Harry Potter support network: How much does the size of the biggest component change after removing one character from the network?

The shown diagrams present the count of ratios calculated by the simulation. In both diagrams (with and without Isolates), the most frequent ratio is around 0.97, showing that when most characters are removed, the network maintains about 97% of its original largest connected component size. This suggests the network is highly robust. In both simulations very few node removals resulted in ratios of 0.81 or 0.94, indicating that it's rare for character removals to significantly fragment the network. We theorize that in these cases one of the few main-characters was removed from the network. After removing the isolates a big drop in the ratio 1 is visible because the removal of a isolate does not change the size of the biggest component. A small amount of ratio 1 remains, because of the second component of opponents of the main characters also does not change the size of the main component.

In conclusion the supportstructure of Characters in Harry Potter is very stable and only would take significant damage in case of the removal of a main character.

Bibliography

Goele Bossaert, Nadine Meidert. 2013. "'We Are Only as Strong as We Are United, as Weak as We Are Divided' a Dynamic Analysis of the Peer Support Networks in the Harry Potter Books." *Open Journal of Applied Sciences*, no. 3 (June): 174–85.