

CSCE 489 Cover Page

Homework Assignment #05 (100 points)

First Name: Sami

Last Name: Chowdhury

UIN: 423006267

Any assignment turned in without a fully completed cover page will receive ZERO POINTS.

Please list all below all sources (people, books, webpages, etc.) consulted for this assignment:

Web Material (URL)

1. <http://cneurocv.s.rmki.kfki.hu/igraph/doc/R/iterators.html>
2. <http://www.sthda.com/english/wiki/ggplot2-barplots-quick-start-guide-r-software-and-data-visualization>
3. <https://www.sharpsightlabs.com/blog/highlight-data-in-ggplot2/>

Recall that University Regulations, Section 42, define scholastic dishonesty to include acquiring answers from any unauthorized source, working with another person when not specifically permitted, observing the work of other students during any exam, providing answers when not specifically authorized to do so, informing any person of the contents of an exam prior to the exam, and failing to credit sources used. Disciplinary actions range from grade penalties to expulsion. Please consult the Aggie Honor System Office for additional information regarding academic misconduct – it is your responsibility to understand what constitutes academic misconduct and to ensure that you do not commit it.

I certify that I have listed above all the sources that I consulted regarding this assignment, and that I have not received nor given any assistance that is contrary to the letter or the spirit of the collaboration guidelines for this assignment.

Today's Date: 7/27/19

Printed Name (in lieu of a signature): Sami Chowdhury

Q1. [18 points]

Explain the following concepts and provide concrete (real-world applications) examples of usage (provide sharp and objective explanations and examples):

(A) Random network

- A. **Random Networks** are networks where the vertices are connected by links in a completely random manner. This network mimics a binomial distribution. They have fewer connections and are spread out in the network. Not as many connections are needed to fully connect the network, so it is very sparse with a moderate diameter. Concrete examples of random networks include yeast protein interactions, science collaboration, and power grids. These systems all randomly connect together in very sparse networks to become fully connected.

(B) Regular network

- A. **Regular networks** contain deterministic rules to connect to nodes and are highly ordered. These are very rare in nature, but do appear. LAN connections are built in a bus regular network framework, and diamonds are another example of regular networks. These seem to be the only few concrete examples of regular networks. They have few connections and are clustered. They require many connections to be fully connected and have a large diameter.

(C) Small-world network

- A. **Small-world networks** are an intermediate between regular and random networks. A regular network that is fully connected has a few edges randomly rewired to create a small-world network. With these random rewiring, short cuts are made with significant impacts such as reducing path lengths by half. This type of network represents many real-life networks that humans interact with. For example, the internet, world wide web, email and actor networks (also known as the Six Degrees of Bacon) are all examples of small-world networks. They are estimated by dividing the natural logarithm of the number of nodes by the natural logarithm of the average degree of each node. The number of clusters for these nodes decreases as the number of node degrees increase.

(D) Scale-free network

- A. **Scale-free networks** are also known as complex networks. One of the most concrete and important examples of a scale-free network is the world wide web. Scale-free networks are networks whose degree distribution follows a Power law. Scale-free networks have natural hubs of information and grow fast with N. Another example would be a network of air traffic in the United States. There are many nodes with few links and a few hubs with a large number of links in scale-free networks. This is one of the most significant networks that humans have created to build many frameworks and solve problems such as the size of craters on the moon. They are used for very large networks.

(E) Centrality measures

- A. **Centrality measures** identify the most important vertices within a graph. They help us identify the most influential people in a network. They are key infrastructure nodes in the Internet networks and describe super spreaders, or virus hubs, of disease. They are also used to

analyze social networks. Centrality measures help us understand how connected and influential each node is in the larger picture of the network. They also help us identify nodes that can propagate the most amount of information. Centrality measures are used to classify the most important characters in a book series, such as Game of Thrones. Centrality measures are limited to only being able to rank the vertices by importance, not quantify the differences in importance. Still, they can be applied in many applications like social network analysis and defining the most important characters in a novel series to help us comprehend the most important information from the nodes of the networks

(F) Community detection

- A. Community detection involves finding group of nodes that have a higher likelihood of connecting to each other than to nodes from other communities. Community detection helps us further classify groups of data that can provide more properties about the network. They are most widely seen in social network and friend circles. Another example is how Vincent Bondel was able to figure out how the country of Switzerland is split in terms of language being spoken. Community detection can help us find hubs, such as how Bondel found that Brussels was a hub between the two communities where both Dutch and French were spoken. Other real-world examples include scientific citation networks and protein networks. By identifying communities within a network, we can understand which communities influence link weights, how fast information travels within communities, and the presence of degree correlations. Nodes that act as intermediaries between two communities help identify information brokers in networks.

Q2. [82 points]

In this exercise, you are suppose to use the `igraph` for R.

The network

The network describes the the social network of the The Fellowship of The Ring, the first of the three volumes of the novel The Lord of The Rings. In this network, we have 36 **vertices**, one for each character. An **edge** between characters means that they interacted, spoke of one another, or that another character spoke of them together. There is an **edge** between two characters if their names appeared within 15 words of one another. The **weight** of the edge is the **number of interactions** between two character within the book.

The network is divided in two files:

- The `fellowship.vertices.csv` has two columns ID (unique integer identifying the character) and label (name of the character).
- The `fellowship.edges.csv` has three columns Source (source/from node) and Target (target/to node) and Weight (number of interactions between Source and Target). Note that all **edges are undirected**, simplifying the network.

Using the `igraph` for R, you should build the network of the Fellowship of the Ring and do an analysis similar to the Network of Thrones, presented in class.

The **centralities** you should evaluate are: Betweenness, Closeness, Degree, Weighted Degree, Eigenvector, PageRank, Contribution Centrality, and Hub score. All centralities were exemplified in the **igraph** lab sessions on eCampus, except the Contribution centrality, which can be computed using the source code available on eCampus. It is as simple as

```
con = contrib_centrality(your_network_object)
```

The community detection algorithms you should use are: Newman-Girvan and Louvain both also exemplified on the **igraph** lab sessions on eCampus.

Compare the results of each centrality and comment on their meaning for this specific network (not general networks). Show the ranking of the first 10 characters for each centrality. Conclude what are the three most important characters in the book and why. You may find useful to watch again the video about the introduction to Network Science with the analysis of the Network of Thrones. In case, two centralities provide the same ranking, explain why you think they are the same.

In case the two community detection algorithms find different communities, explain what are the similarities and differences among the communities they detected. Try to explain the meaning of each community. Explain which algorithm seems to provide the best community division based on your knowledge of the Fellowship of The Ring.

Your report structure and content is of your choice. Your ability to properly collect, analyze, and present the results is part of your evaluation. Do your best in choosing how to present your results.

Grading distribution:

- 22 points for your R files (correctness, reproducibility, and quality of code/comments).
- 60 points for your report with your analysis (quality of the provided information and quality of presentation).

Provide all the useful information to allow you TA to reproduce your results.

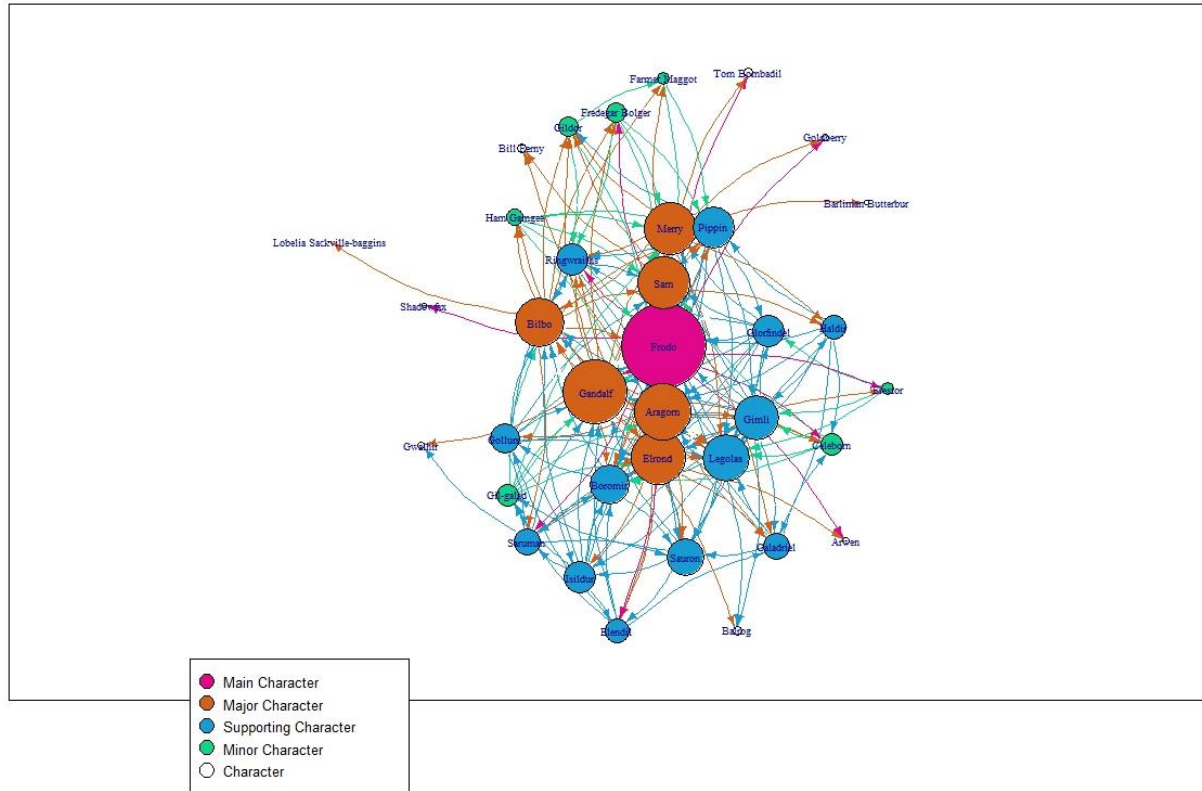
You must turn in all of your files (R, report, and any other file you judge important for reproducing your results)) in one single zip file.

Remember that your work will be checked for plagiarism. You should talk and collaborate with your friends, but your coding and writing should be done by your own.

Have fun!

Report

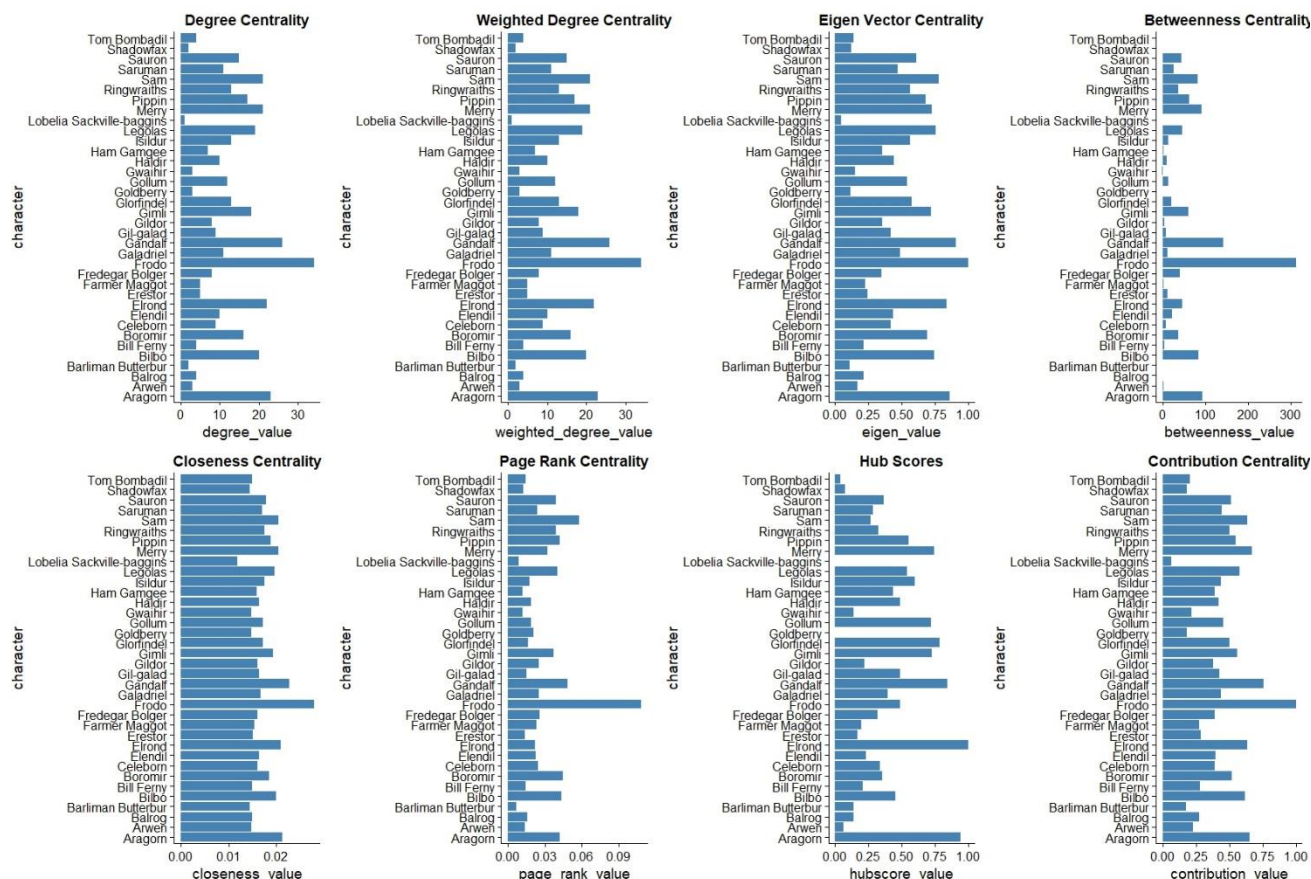
Below is my network that I created of the Fellowship of the Ring



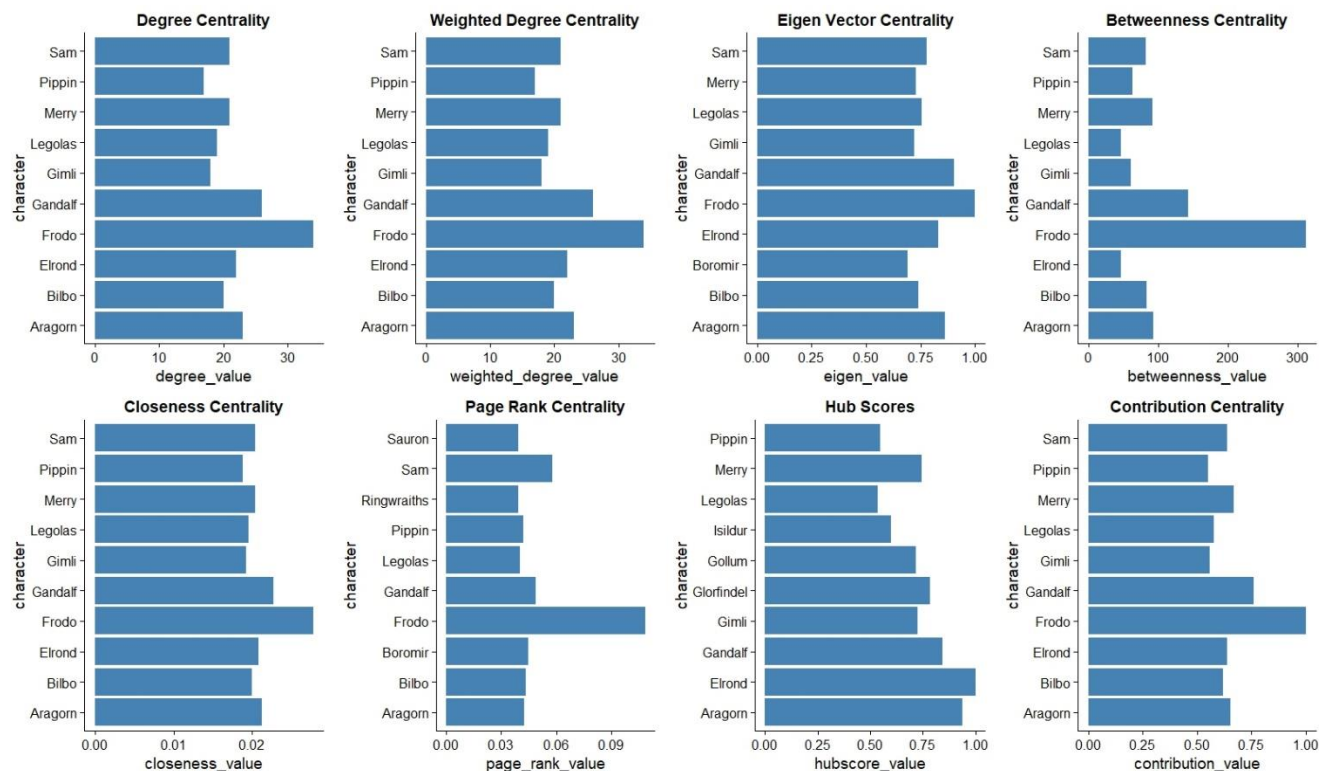
This image is also saved in the directory as “Network.jpeg”

This network was built with the components layout.

Below are the Centrality Measures listed below in Bar Graph Representation, saved as “Overall Plots.jpeg”



The Top 10 Values were extracted into bar graphs and tables below, saved as “Top10Centralities.jpeg”



These tables are saved as ResultsPT1 and ResultsPT2 jpeg files

	character	degree_value	character.1	weighted_degree_value	character.2	eigen_value	character.3	betweenness_value
1	Frodo	34	Frodo	34	Frodo	1.0000000	Frodo	312.04034
2	Gandalf	26	Gandalf	26	Gandalf	0.9025936	Gandalf	142.66540
3	Aragorn	23	Aragorn	23	Aragorn	0.8605167	Aragorn	93.61783
4	Elrond	22	Elrond	22	Elrond	0.8327810	Merry	92.01129
5	Sam	21	Sam	21	Sam	0.7789196	Bilbo	84.03944
6	Merry	21	Merry	21	Legolas	0.7539940	Sam	82.24403
7	Bilbo	20	Bilbo	20	Bilbo	0.7415829	Pippin	62.92689
8	Legolas	19	Legolas	19	Merry	0.7268037	Gimli	60.95428
9	Gimli	18	Gimli	18	Gimli	0.7222067	Legolas	46.77915
10	Pippin	17	Pippin	17	Boromir	0.6917149	Elrond	46.76724

	character	closeness_value	character.1	page_rank_value	character.2	hubscore_value	character.3	contribution_value
1	Frodo	0.02777778	Frodo	0.10805706	Elrond	1.0000000	Frodo	1.0000000
2	Gandalf	0.02272727	Sam	0.05764297	Aragorn	0.9376447	Gandalf	0.7577287
3	Aragorn	0.02127660	Gandalf	0.04865455	Gandalf	0.8440874	Merry	0.6681403
4	Elrond	0.02083333	Boromir	0.04485620	Glorfindel	0.7841911	Aragorn	0.6543154
5	Sam	0.02040816	Bilbo	0.04330196	Merry	0.7440379	Elrond	0.6371825
6	Merry	0.02040816	Aragorn	0.04239137	Gimli	0.7262274	Sam	0.6366899
7	Bilbo	0.02000000	Pippin	0.04181657	Gollum	0.7188259	Bilbo	0.6184580
8	Legolas	0.01960784	Legolas	0.04020306	Isildur	0.5977792	Legolas	0.5781468
9	Gimli	0.01923077	Sauron	0.03928987	Pippin	0.5482899	Gimli	0.5593927
10	Pippin	0.01886792	Ringwraiths	0.03907636	Legolas	0.5372667	Pippin	0.5483827

Lastly, the communities detected by the NG and the Louvian algorithms are displayed below. The NG community is displayed first, followed by the louvian algorithm graph. Saved as “NG_communities.jpeg” and “LOU_communities.jpeg”

Analysis:

For this specific network, we can deduce many characteristics based on the centrality measures. The Top 10 measure bar plots are listed again below.

	character	degree_value	character.1	weighted_degree_value	character.2	eigen_value	character.3	betweenness_value
1	Frodo	34	Frodo	34	Frodo	1.0000000	Frodo	312.04034
2	Gandalf	26	Gandalf	26	Gandalf	0.9025936	Gandalf	142.66540
3	Aragorn	23	Aragorn	23	Aragorn	0.8605167	Aragorn	93.61783
4	Elrond	22	Elrond	22	Elrond	0.8327810	Merry	92.01129
5	Sam	21	Sam	21	Sam	0.7789196	Bilbo	84.03944
6	Merry	21	Merry	21	Legolas	0.7539940	Sam	82.24403
7	Bilbo	20	Bilbo	20	Bilbo	0.7415829	Pippin	62.92689
8	Legolas	19	Legolas	19	Merry	0.7268037	Gimli	60.95428
9	Gimli	18	Gimli	18	Gimli	0.7222067	Legolas	46.77915
10	Pippin	17	Pippin	17	Boromir	0.6917149	Elrond	46.76724

	character	closeness_value	character.1	page_rank_value	character.2	hubscore_value	character.3	contribution_value
1	Frodo	0.02777778	Frodo	0.10805706	Elrond	1.0000000	Frodo	1.0000000
2	Gandalf	0.02272727	Sam	0.05764297	Aragorn	0.9376447	Gandalf	0.7577287
3	Aragorn	0.02127660	Gandalf	0.04865455	Gandalf	0.8440874	Merry	0.6681403
4	Elrond	0.02083333	Boromir	0.04485620	Glorfindel	0.7841911	Aragorn	0.6543154
5	Sam	0.02040816	Bilbo	0.04330196	Merry	0.7440379	Elrond	0.6371825
6	Merry	0.02040816	Aragorn	0.04239137	Gimli	0.7262274	Sam	0.6366899
7	Bilbo	0.02000000	Pippin	0.04181657	Gollum	0.7188259	Bilbo	0.6184580
8	Legolas	0.01960784	Legolas	0.04020306	Isildur	0.5977792	Legolas	0.5781468
9	Gimli	0.01923077	Sauron	0.03928987	Pippin	0.5482899	Gimli	0.5593927
10	Pippin	0.01886792	Ringwraiths	0.03907636	Legolas	0.5372667	Pippin	0.5483827

The degree centrality shows how many other characters the specific character has interacted with. From its bar plot, we can see that Frodo is the character who has interacted the most amongst all the characters. It would be logical than to conclude that they are the main character of the Fellowship of the Ring, which does check out when examining the series. The other 9 top characters would then follow suit in being other impactful major characters, while Frodo would be the main character. Other characters not listed here would be listed as minor or supporting characters, with the next most popular after the top 10 being the supporting characters. The weighted degree centrality supports the same conclusions as the degree centrality, this time displaying the number of interactions the character listed has participated in. It again follows suit that the characters with the most character interactions would also have the most interactions participated in.

The following centrality measure is the eigen vector centrality. This measure displays how much credit each character has with the importance of its neighbors. In other words, this measure shows which characters are connected to other important characters. In this bar plot, we can see that most of the characters who ad leading interactions prevail again, but this time, Bilbo has a higher value because they are close with more important characters than Boromir. Legolas also becomes the new placeholder for the 6th most important character with this measure as they know more important characters compared to Bilbo and Merry.

The next centrality to analyze is the betweenness measure. Betweenness depicts how frequently a vertex lies on a short path between pairs of other vertices. The people with the highest values would be brokers of information.

Similar trends seem to with the exception that Bilbo and Merry become better brokers of information. Their networks are closer than the networks of Sam and Elrond, who fall behind in this measure. Next, closeness measures the number of links a character has to traverse to get to all other vertices. In other words, this centrality measures how extensive the connections are for each character. This measure trends similarly to the degree centrality measures, showing that the people who have interacted with the most people have the longest distances to traverse.

Page rank is another centrality that measures how important the character is based on how many other characters interact with the specific node. Sam and Boromir shine higher in this regard as they have more characters interacting with them. Surprisingly, Ringwraiths also shows up in the top 10 list, meaning that multiple characters have linked with this character, even though they may not have interacted with other characters. Even though they may be in an isolated section of the network, their importance is noted with this centrality. Lastly, we have the hub score and authority scores of the characters. Hubs are classified as characters with a large number of outgoing links while authorities are classified as characters with a large number of incoming links. Hubs would basically indicate the most forward and social characters while authorities would display the most charming or beloved characters. We can see here that Elrond, Aragorn, and Gandalf are the characters who establish the most connections towards other people initially. For the first time, Frodo does not appear in this centrality measure's top 10 list, showing instead with the authority measure that he is one of the most interacted with characters. These measures break down the degree centrality measures to figure out how the connections are established and by who initially establishes them.

Moving on to the community detection results, both algorithms found different sized communities for this network. The NG method has communities that have only one or two characters contained within them, while the Louvain method have 3 large communities grouping similar characters in a much more general sense. The NG method seems to just distinguish between who has more than 5 lines in the Fellowship of the Ring. On the other hand, the Louvain algorithm does a better job at detecting factions within the Fellowship of the Ring. Based on my knowledge, the Louvain method better divides the community of characters in this scenario with meaningful and rich communities. These communities are filled with more people and have more distinction compared to the NG community algorithm's communities. It has concrete communities that are large enough to make similar comparisons to. The results of both algorithms have been taken as a table below

	NG Algorithm				Louvain Algorithm		
Number of Communities	4				3		
Size of each	32	1	1	2	8	13	15
Modularity	0.008613913				0.170223		
Membership	Every one Else	Barliman	Lobelia	Tom, Gold berry	Legolas, Gimli, Glorfindel, Celeborn, Haldir, Galadriel, Erestor, Balrog,	Gandalf, Elrond, Aragorn, Boromir, Gollum, Sauron, Galad, Isildur, Saruman, Elendil, Gwaihir, Shadowfax, Arwen	Frodo, Sam, Merry, Bilbo, Ringwraiths, Pippin, Fredegear, Gildor, Ham, Farmer, Bill, Barliman, Lobelia, Tom, Goldberry