Network Analysis Project Report

Vincenzo Collura, Artificial Intelligence

0.0.1 Abstract

Network theory is often used to analyze works of fiction and film. This article aims to compare two of the most famous and successful sitcoms ever filmed: Friends and The Big Bang Theory. The two television series have much in common, both revolve around the events of a group of friends struggling with their daily lives, their jobs, their love affairs, and so on. Moreover, several Internet sites and fan pages of the two claim how similar these two works are to each other, comparing the characters and dynamics in the two sitcoms. Therefore, this article uses network theory in order to objectively and numerically verify these common intuitions. These sitcoms have been compared and the commonalities in terms of the general structure of the networks, characters, and dynamics between them have been listed.

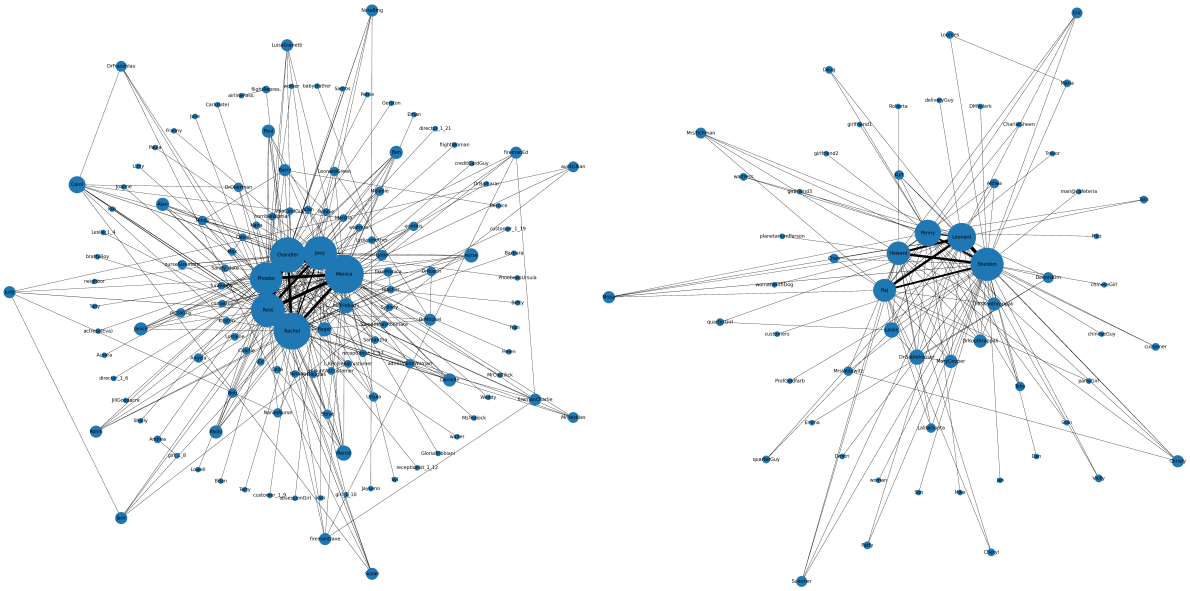

Figure 1: Friends and The Big Bang Theory nets respectively. The nets were shown by changing the thickness of the edges according to weight, the size of the nodes according to degree, and the layout used uses the cores of the net, chosen because it is very clear.

0.0.2 Context

The general field covered in this article is cinematography and specifically the sub-field of TV series. With today's streaming technology, TV series are becoming more and more popular. As this article says [1] now TV Series must be taken seriously. Due to their format (usable anywhere and anytime) and the participatory qualities of using the Internet (tweeting, chat forums), series allow a new form of education, entertainment, expressing complex issues through storytelling

and characters. What makes series so popular? In addition, of course, to classic elements of classic cinematography, so plot, cast, cinematography, etc.. the structure of the social network of characters underlying the plot can also offer some insights as stated in this article [2], the use of network theory is great way to shed light on issues related to the social network underlying a TV show.
In this article, the focus will be on comparing two sitcoms whose popular intuition is that they are very similar: Friends and The Big Bang Theory (TBBT). For those who have not seen the two series find a brief and immediate description on Wikipedia [3, 4]. Comparisons are mostly made on the Internet by fans of the two series, as can be seen, for example, in the forum [5] and website [6] and other endless resources around the web on the subject.

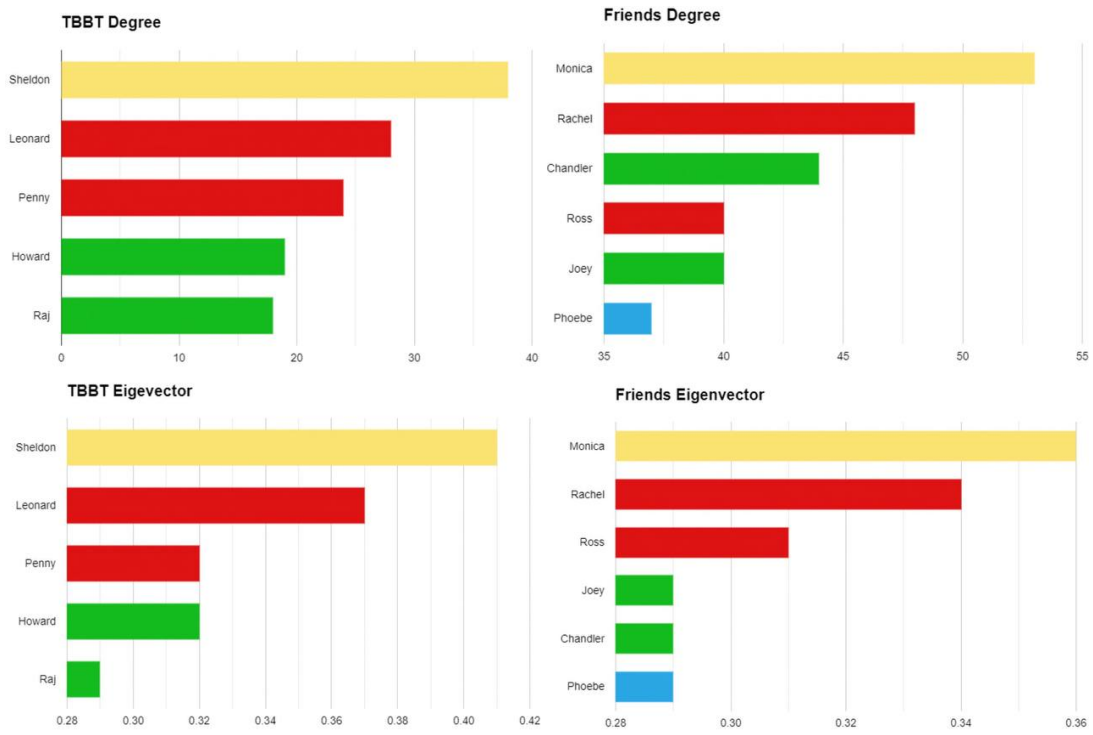

Figure 2: Centrality measures applied to the two sitcoms: degree, eigenvector. With comparisons highlighted: red for the love line, green for the comic duo, yellow for the protagonist among the protagonists, and blue for the most secondary among the protagonists.

0.0.3 Problem and Motivation

As is well known, the two sitcoms under consideration are among the most popular sitcoms ever filmed. Popular opinion is that the two are very similar to each other, in that, both are about a group of friends struggling with their everyday lives, work, awkward situations, love lives, etc.. The use of network theory is increasing on issues concerning the social networks underlying film narratives and narrative works in general. As the following paper [7] says, problems due to narratives can be solved through character network analysis. Character network analysis can help us draw conclusions about the similarities between our two works under consideration by comparing them. Obviously plot, cast, and direction are key aspects and, but the structure of the play's social network and the characters underlying the plot can also offer some insights. The main questions then are: to what extent can the structure of the social networks behind these shows contribute to discussions on the similarity between the two series? Given the great success

of these two series can one infer what is the basic network for a successful sitcom? As done in this article [8] where they analyzed the character networks of Stargate and Star Trek and found that their structures are quite similar. What are the types of characters in them? For example [9], the most central characters in Game of Thrones are studied. Finding these characteristics that unite the two networks could be important in writing future sitcoms, evaluating new ones, evaluating scripts, identifying a sort of basic network to follow for a successful sitcom.

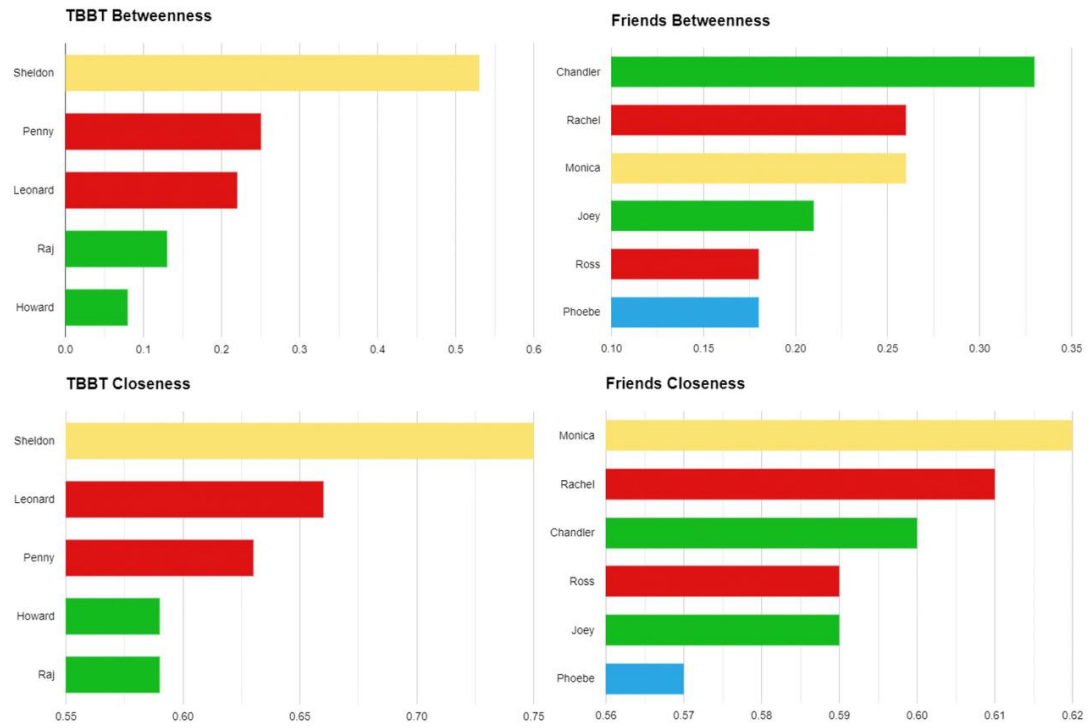

Figure 3: Centrality measures applied to the two sitcoms: betweenness, closeness. With comparisons highlighted: red for the love line, green for the comic duo, yellow for the protagonist among the protagonists, and blue for the most secondary among the protagonists.

0.0.4 Datasets

The datasets are online and publicly available at these two links: Friends, The Big Bang Theory. The data was manually collected. This was done by the author of this paper [10], Each episode was watched and notes were taken, regarding how many times who interacts with whom, i.e., based on the actual interactions of characters in each scene of each episode. An interaction happens when two characters talk (even if one talks and the other just listens) or touch or have eye contact. This means that, since not necessarily every character does interact with all others in a scene, each scene is not a complete graph. The datasets contain essentially all interactions or dialogues present between characters in all episodes of all seasons. But the data made available to the public by the author for The Big Bang Theory covers only the first 22 episodes, so it was decided to analyze only the first season of Friends. As a result the data analyzed covers the first season of the sitcom Friends and the first season plus 5 episodes of the second season of the sitcom The Big Bang Theory. Inasmuch as the first season of Friends consists of 24 episodes, while that of TBBT consists of 17 episodes. The Python programming language was used to calculate all measurements thanks to the NetworkX library [11]. For ease of viewing the data was also pre-processed and transformed from an edge list to an edge list with weight of that

relative edge, where the weight is equal to the number of times the same edge appears in the list. All the work done can be found at this link.

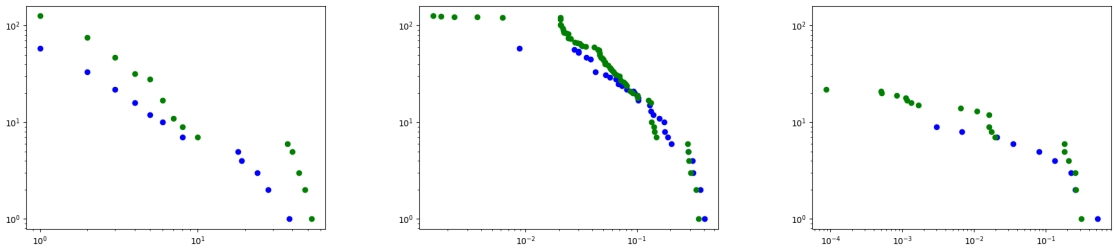

Figure 4: Respectively degree, eigenvector, betweenness distributions. Blue TBBT, green Friends.

0.0.5 Validity and Reliability

As anticipated, our datasets consist of manually extracted data from Bazzan [10]. Bazzan manually annotated all the data, defining an interaction as two characters talking, touching or making eye contact. Although there may be human interpretation errors in this dataset, this is the most reliable method of extracting the social network. Even in this article [12] Edward compared different extraction methods for Friends using manually extracted and automated datasets, providing evidence that automated data extraction methods are reliable for most but not all analyses. Indeed, he calls the manual extraction method used by Bazzan the "gold standard" for character social networks. So we can say that however abstract our model is still very valid, even one of the most valid representations for these types of studies. Moreover in our networks which are very small it is easy to see that there are no major errors of omission, commission at least among the few main characters. Also since what we are prentending to consider are simple interactions between characters there can be no retrospective and attribution errors. The least reliable part of the whole project, however, is the second part of the results. The part about the comparison between the protagonists, as the analysis of centrality and similarity measures are not the only aspects to be analyzed for such an analysis, so some interpretations on the measures are based on my perceptions and those of common opinion found around the web. So these results are less reliable because other researchers may perceive things differently than we do.

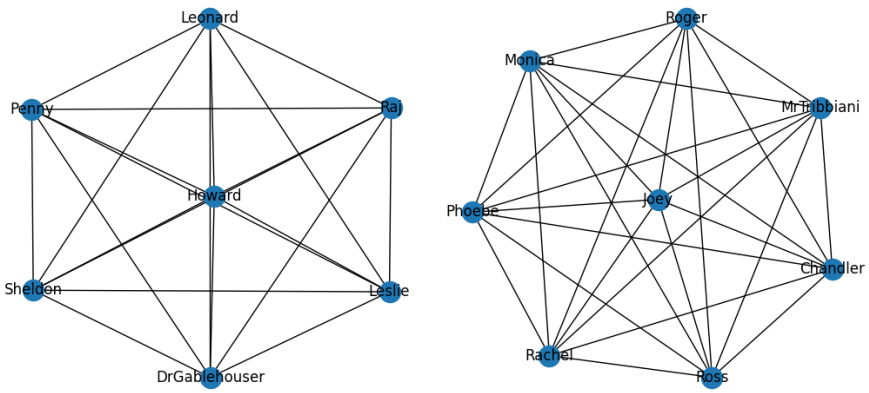

Figure 5: The Big Bang Theory and Friends maximal cliques, respectively.

0.0.6 Measures

The measures will focus on two key points: a comparison between the two networks, and a comparison between individual characters. So as to identify similarities, strengths, main features, and differences between the two networks in general, but also carping on a possible similarity between the main characters.
To compare the two networks we will use their properties: density that is the percentage of possible relationships in the network that are actually present, conceptually, it provides an idea of how dense a graph is in terms of edge connectivity. Diameter that is equal to the percentage of possible relationships in the network that are actually present, this measure can suggests how close all the characters are to each other as well as average shorter path length.
Also to compare the two networks we will analyze groups of nodes: cliques subsets of vertices of an undirected graph such that every two distinct vertices in the clique are adjacent, the study of cliques can tell us, by observing which and how many other characters are part of the larger cliques, if in addition to main characters there are other secondary characters or if these only stop at interactions with the individual main characters, i.e. if the story is centered on the common life of the principals or on their parallel lives and acquaintances. K-cores maximal subgraphs that contains nodes of degree k or more, by studying the main cores we can figure out which characters are part of it from the main to the secondary. Clustering coefficient that quantifies the density of triads i.e., strongly connected triangles of nodes in a network, can be used to study the clustering coefficient at the network level but also at the local level, so at the node level. Communities that are subsets of nodes within the graph such that connections between the nodes are denser than connections with the rest of the network, we can thus observe whether the main characters have communities of their own or in common with each other. But also the distributions of centrality measures to understand, for example, whether the networks are scale-free.

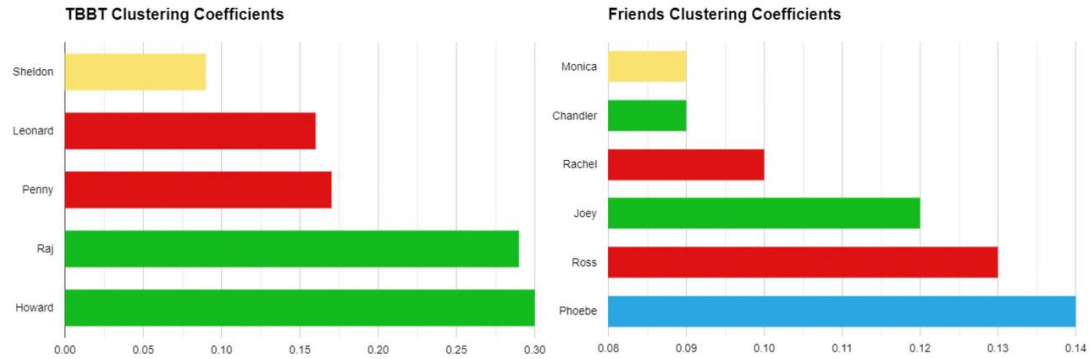

Figure 6: Clustering coefficients. With comparisons highlighted: red for the love line, green for the comic duo, yellow for the protagonist among the protagonists, and blue for the most secondary among the protagonists.

In terms of comparing individual characters, we will mainly use centrality measures such as: degree, the simplest centrality measure for a node, the number of connections that a node has to other nodes in the network. Eigenvector centrality that is an extended form of degree, which takes into account not only how many neighbours a node has, but also how central those neighbours themselves are. Betweenness centrality, based on shortest paths, that measures the extent to which a node lies on paths between other nodes. The assumption here is that paths lying on “trafficked” shortest paths have a more central role in the network, as gateways favoured

by their closeness to (reach) the other nodes. Closeness centrality uses the shortest paths in networks, measuring the mean distance from a node to other nodes. These will allow us to better compare characters, understand which ones are the most important and what role they play. Obviously the analysis of centrality measures alone is not enough but they will be supported by regular equivalence which is a similarity measure, two nodes are said to be regularly equivalent if they have the same profile of ties with members of other sets of actors that are also regularly equivalent. This measure was implemented by hand using the following video as a basis [13].

0.0.7 Results

Commonalities between the two shows identified through network analysis, applied measures show that:

1. Low diameter and average shortest path length values, this tells us that the two series are all about the interactions between the protagonists or of the protagonists with secondary characters, with a few (negligible) exceptions.

2. Few other characters, apart from the main ones, in the cliques with the highest number of nodes. So few secondary characters have interactions with all the main ones. The story also and above all develops on the parallel plots of these characters and their sub-groups and not on events involving all of them.

3. In the k-cores with higher k we find all secondary characters who have the same kind of relationship with one of the main characters, i.e.: parents and siblings, love relationships. With very few exceptions.

4. A community for each main character, each containing friendships and kinships of each specific character.

5. High clustering coefficients, so, an approximate 10% to 60% probability that two neighbors of a node are themselves neighbors.

6. Both are scale-free networks, so, highly robust networks that can survive the failure of a sensible number of their nodes.

First point: the diameter of the networks, with value 4 for TBBT and 5 for Friends, suggests how close all characters are. In particular, it is easy to see, also looking at the nets in figure 1, how all secondary characters are immediately connected to at least one of the main characters, with the exception of some characters who are still only two edges away from the main characters. Confirming this statement is the average shortest path length of 2.24 in TBBT and 2.5 in Friends, i.e. all characters are reachable from the others with the help of only two edges on average.
Second point: analysing the cliques, we can see, as expected, that obviously all the main characters of the two series are part of the maximal clique in their respective networks, plus some secondary characters who have interacted with everyone in some way. Of these cliques, in TBBT there are 4 with sizes between 6 and 7, while in Friends there are 9 with sizes between 7 and 8. So in both series besides the main characters there are only a few others. This means that in each case the story also and above all develops on the parallel plots of these characters and their subgroups and not on events involving all of them. Two examples of maximal cliques are in the figure 5.
Third point: now let us focus on the cores, which you can find in the figure 7. As expected,

within the main cores we find the main characters of both sitcoms. What can be noticed, however, is that within the main core and in the immediately following cores all secondary characters have the same type of relationship with one of the secondary characters, i.e.: parents and siblings, love relationships. For example (listing from the core with the highest k to those with the lowest k): Dr Koothrappali and Mrs Koothrappali are Raj's parents, Mary Cooper is Sheldon's mum, Leslie has a love relationship with Leonard, Missy Sheldon's sister, Eric Penny's boyfriend etc. On the other hand, in Friends we find: Roger who is Phoebe's boyfriend, Mr. Tribbiani who is Joey's father, Paul Ross father-in-law, Ben Ross son etc.
Moreover, in all the calculated measures of centrality, immediately after the main protagonists we always notice the same secondary characters, those in the cores with the highest k.
Fourth point: the study of communities, as we can see in figure 9, produced 4 communities for TBBT and 6 for Friends. In particular for Friends it has produced one community per character, which makes a lot of sense, because in each of these communities there are the friendships and kinships of each specific character, and few of these are in common with other main characters, or they are but only with some of them. We are therefore bound to the fact with the cores. In particular, we can see that in general all the communities are well separated, as mentioned, except for some characters; the two closest communities (between which there are more edges) are Chandler's and Joey's, which I expected to find even in the same community, but it is clear that there is a very strong bond between the two housemates. For TBBT we find the exact same results, one community per character containing friendships, people and relatives of that group. With the exception of Raj and Howard's community, which is the same, confirming the similarity between the Raj-Howard and Chandler-Joey comic duo.
Fifth point: the two networks have a high clustering coefficient, 0.48 for TBBT and 0.51 for Friends, i.e. a probability of between 10% and 60% that two neighbours of a node are themselves neighbours.
Sixth point: we can see how, by plotting the distributions of degree, eigenvector and betweenness centrality, figure 4, the curves of the two shows are very similar, almost equal, in each case. In particular, the curve in front follows a power-law behaviour. Networks whose degree distribution follows a power-law behaviour are usually called scale-free networks. The name is derived from the fact that power-laws are scale invariant. Scale-free networks are highly robust networks, capable of surviving the failure of an appreciable number of nodes.

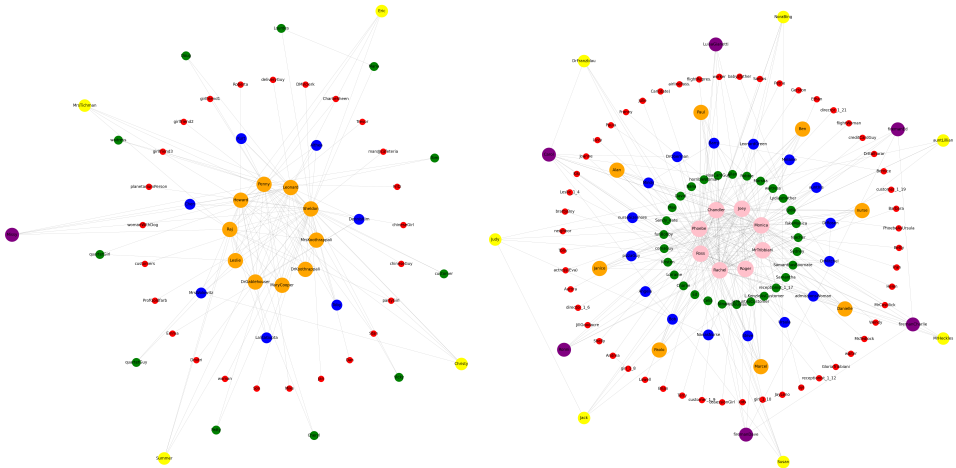

Figure 7: TBBT and Friends k-cores respectively. The larger the node size the larger k and nodes of the same color belong to the same core.

Comparisons between characters, applied measures show that:

1. The most relevant character among the protagonists is Sheldon for TBBT and Monica for Friends.

2. Both series have a main love line: Leonard-Penny, Ross-Rachel. So Ross' correspondent in TBBT is Leonard, and Rachel's correspondent is Penny.

3. Both have a comic duo: Raj-Howard, Chandler-Joey. However, it appears from the applied measures that Howard has a very similar role to Chandler. While Raj is not so reviewable in Joey but more in Phoebe.

All the results for the centrality measures are given in the figure 2 and figure 3. Looking at the measures: degree, eigenvector and closeness centrality, the results are the same, i.e., in TBBT we have a clear distinction between Sheldon and the rest of the main characters, whose values gradually go down from Leonard, his roommate, to Penny; the latter two have two very close central positions and as we know there is a love line between the two, and then we move on to the Raj, Howard pair whose values are more or less identical. While there is no prominent character in Friends, as we see, however, the highest rank is Monica's by any measure, the host in which many scenes in the sitcom take place, while Phoebe's is the lowest. These two values encapsulate the remaining 4 main characters who have very similar values to each other, among these characters are Ross and Rachel who make up the love line of the series and Chandler and Joey who make up the comedy duo line.
The same results were also obtained with clustering coefficients, figure 6, ordering them in descending order of value. Nodes tend to aggregate and connect internally to their "groups", therefore, in networks exhibiting this behavior, nodes belonging to small groups are forced to have a low degree, but at the same time their local clustering coefficient tends to be larger because each group, being mostly detached from the rest of the network, increases its internal clustering coefficient.
With the centrality of Betweenness, figure 3, the remarks made about TBBT remain the same and are even more confirmed, it changes a little bit on the other hand as far as Friends is concerned where things are somewhat subverted in support of the fact that unlike TBBT in Friends everyone is a protagonist more or less equally, Monica stands out but very little.

	TBBT character	Friends character
1	Sheldon	Monica
2	Leonard	Rachel
3	Penny	Ross
4	Howard	Chandler
5	Raj	Phoebe

Figure 8: Regular equivalence results. With comparisons highlighted: red for the love line, green for the comic duo, yellow for the protagonist among the protagonists, and blue for the most secondary among the protagonists.

Studying the regular equivalence the results found, figure 8, confirm the considerations made above. The regular equivalence found that the most similar characters are Sheldon and Monica, as expected. As for Leonard and Penny, the respective characters are Rachel and Ross, the

two love lines as expected but in reverse. As for the comedy duo, the assumptions made with the measures of centrality were only partially met, namely the similarity between Howard and Chandler, while Raj is compared to Phoebe and not Joey. This last result was to be expected since the values are always the lowest as precisely we see with Phoebe. While Joey tends to have average values with the other characters.
So are the two sitcoms similar or not? The conclusion is that the two basic networks are certainly very similar to each other on a network level, the six points listed above are very strong commonalities and offer insights into the basic network properties of a sitcom. The finding of two dynamics so fundamental to the series' plot as the love story between two of the main characters and the comedy duo are also symptoms of strong similarity.
While the three points concerning the comparison between the characters can offer good pointers as to the models to look at when wanting to write, for example for a new series, the most relevant character among the main characters using Sheldon and Monica as examples, a comedy duo using the two aforementioned comedy duos and the characters that are part of it or the characters that are part of a love story.

0.0.8 Critique

Certainly this analysis provides many interesting insights to consider, although it does not provide a detailed 360-degree base network. This is because our data was limited to only the interactions between the characters while we could, for example, also evaluate to the writing of the characters, thus providing a more detailed comparison between them. Another very interesting thing that could have been done between characters is role detection, as explained by this article [14]. In addition, although the first season is of crucial importance to the development of a series it was very limiting not to include the rest of the seasons that we unfortunately did not have available. Thanks to the seasons after the first one the study could be performed a study on the temporal aspects of the series as well, such as done with Friends in this article [2]. Another fact is that there are two series analyzed, but there are many other equally successful series that can be analyzed in order to prove our conclusions.

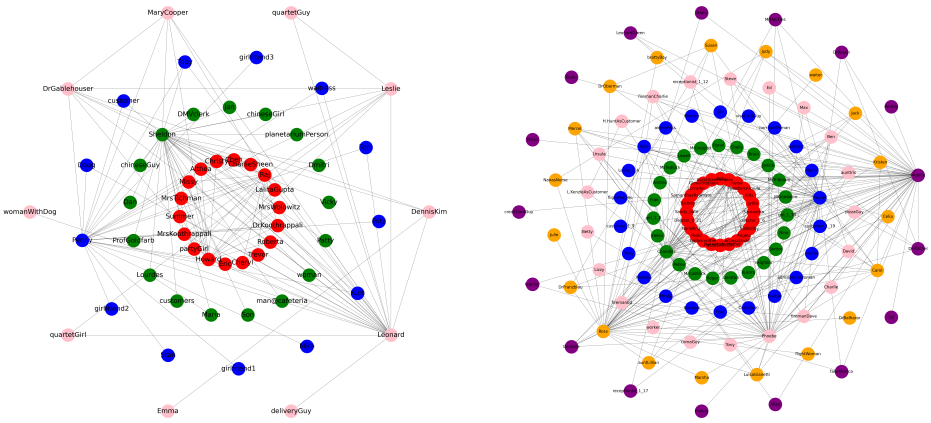

Figure 9: TBBT and Friends communities respectively. Nodes of the same color belong to the same community.

Bibliography

References

[1] Sandra Laugier. Taking tv series seriously. *Open Philosophy*, 5(1):250–253, 2022.

[2] Ana L. C. Bazzan. I will be there for you: clique, character centrality, and community detection in friends. *Computational and Applied Mathematics*, 39(3), jun 2020.

[3] Wikipedia. The big bang theory — wikipedia, l'enciclopedia libera, 2022. [Online; in data 23-novembre-2022].

[4] Wikipedia. Friends — wikipedia, l'enciclopedia libera, 2022. [Online; in data 23-novembre-2022].

[5] Am i the only one who has noticed a lot of similarities between ... - quora.

[6] Daniel Rugunya. Fans think 'big bang theory' stole some of 'friends' storylines, Dec 2021.

[7] Vincent Labatut and Xavier Bost. Extraction and analysis of fictional character networks: A survey. *ACM Comput. Surv.*, 52(5), sep 2019.

[8] M. S. A. Tan, E. A. Ujum, and K. Ratnavelu. A character network study of two Sci-Fi TV series. In *Frontiers in Physics: 4th International Meeting*, volume 1588 of *American Institute of Physics Conference Series*, pages 246–251, March 2014.

[9] Andrew Beveridge and Jie Shan. Network of thrones. *Math Horizons*, 23(4):18–22, 2016.

[10] Ana Lúcia Cetertich Bazzan. Similar yet different: the structure of social networks of characters in seinfeld, friends, how i met your mother, and the big bang theory. *Revista de Informática Teórica e Aplicada*, 27(2):66–80, Apr. 2020.

[11] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.

[12] Michelle Edwards, Lewis Mitchell, Jonathan Tuke, and Matthew Roughan. The one comparing narrative social network extraction techniques. 2018.

[13] Kanchan Bhale. *Regular Equivalence*. YouTube, Apr 2020.

[14] Vincent Labatut and Xavier Bost. Extraction and analysis of fictional character networks: A survey. *CoRR*, abs/1907.02704, 2019.