Analysis of Social Networks

An Overview of my Project: Looking at the Networks in BoJack Horseman

Lauren Fowler

I. Introduction

For this project I was interested in looking at the social networks in the show, BoJack Horseman, as it is a commentary on what it is like to be a washed-up actor in Hollywood while struggling with mental illnesses, like substance abuse. This assignment provides visual analyses of social networks in the show, as well as analysis of actor prominence, and analysis through ERGMs (exponential random graph models).

There are two parts to social networks, nodes, and edges. Nodes are the characters/people that make up the network. For this assignment, I classified a node as any character that was a main or reoccurring character as per the show's Wikipedia. An edge is the relationship between two nodes. I deemed if there was an edge between two nodes based on my opinion and knowledge of the show.

In this assignment, I looked specifically at perceived friendship and economic exchange relationships. I defined perceived friendship as if a character perceived themselves to have a friendship relationship with another character, based on my interpretation of the show. I chose to look at this relationship because the main character in the show, BoJack, suffers from borderline personality disorder, which is a mental illness that is characterized by unstable relationships, mood swings, and uncertainty in how the person affected views themselves and their relationships with others. Therefore, in viewing this relationship, I was curious in seeing if BoJack perception of his friendships differed from the perception of the characters. When looking at economic exchange, I again made the judgement about if the relationship was present or not based on if the characters interacted (at least once) for jobs, money, or an exchange of goods. I chose to look at this relationship because being an actor and living in Hollywood is a large part of the show's premise, and I was interested in seeing if the economic exchange network may have influenced the perceived friendship network, meaning that who the characters work with tend to be who they are friends with.

When analyzing the networks, I also wanted to specifically look at how character type and occupation of the characters may have affected the networks. This show is a cartoon show with

some characters depicted as animals and others as human beings. When researching the show, I found that the creators chose to do this because they wanted to make the animal characters more relatable to the viewers and felt as though it would be easier for viewers to connect to fictional animal character rather than a depiction of a fictional person. Therefore, I am speculating that this is why four of the five main characters are animals. Thus, with this, I was interested in seeing if the prominent characters tended to be the ones that were animals compared to humans. I also wanted to look at occupation, because I was interested in seeing if there was economic exchange across different occupations and/or friendships that formed across differing occupations. For this, I chose five categories of occupations and then grouped the characters according to their occupation in the show.

Overall, this assignment served as a guide to analyzing social networks as we were able to gather our own data, visualize it, and then analyze it. The main purpose of the assignment was to learn the tools for analyzing social networks, and while conclusions were still drawn about the show, this assignment mainly served as a tool so that, in the future, we have the skills and knowledge to analyze real networks.

II. Exploratory Visualizations of the Networks: Perceived Friendship

Figure 1 is a depiction of what I deemed the total perceived friendship network in the entire show BoJack Horseman. This was created using R. The nodes with the edges are the ones that are a part of the network, while the nodes on the outside are characters in the show that are not a part of the friendship network. Letters are the nodes stand for the characters initial. Single arrow edges depict that the relationship was one sided – the node that arrow is coming from has a perceived friendship with the node that arrow is pointing to, but this relationship is not reciprocated. The double arrows reflect reciprocity. This figure just gives us an overall visualization of the network where we can begin to make some judgements.

Figure 2 depicts the same network, but now nodes are colored according to the occupation of the characters. Figure 3 also illustrates the same network, but with nodes colored according to character type. Again, this is just simply a visualization where we can begin to make some assumptions, but everything is pure speculation.

III. Actor Prominence: Perceived Friendship

In this section of the project, I looked at actor prominence and centrality of the total perceived friendship network. When looking at the centrality of a network, we are looking at who is the most central, meaning who may have the most power, influence, or prestige in the group. There are three different ways to study centrality when looking at a network: by degree, closeness, and/or betweenness. Degree centrality is the measure of the edges each node has in the network, and it comes from the idea that those who are more central to a network will be more connected, and thus, as a result have more edges, a higher score of degree centrality. Closeness centrality is the measure of closeness (distance) a node is to others in the network, which is from the idea that nodes that are more central can communicate more quickly and easily with other nodes in the network compared to less centralized nodes. It is calculated as the average shortest path length from the node to every other node in the network. Therefore, having a lower score in this case is indicative of a more central node. Lastly, betweenness centrality is the measure of how often a node is in the shortest path between two other nodes. This idea stemming from that those who are central to a network are often gatekeepers of information and serve as bridges connecting nodes in the network. It measures the percentage of shortest paths that must go through the specific node, so a higher score means a more central node. For this portion of the project, I decided to include the degree centrality and betweenness centrality scores and I did not include closeness centrality as it showed similar results to the betweenness centrality score and network, and I had a limit on pages for this assignment.

Again, degree centrality is a measure of the number of edges a node has in the network. Since I have a directed network, meaning that not all relationships are reciprocal (arrows) I had a choice of looking at in-degree centrality, which is just analyzing the number of edges that one receives from, or out-degree centrality, which analyzes the number of edges going out from a node. Since I was looking at perception of friendship, I chose to look at in-degree, as if a node was not reciprocated with an edge, I wanted that to be reflected in the centrality score. To start, I removed all the isolates from the network, and they are not influential in the network, but would impact the centrality scores (figure 4).

In Figure 5, the degree() function in the igraph package of R was used to plot degree centrality. Larger nodes depict a higher degree centrality score.

Commented [LF1]: Delete the score in R Lauren

Figure 6 is another way to look at the degree centrality of the network, by looking at the distribution of the of the edges in the network. From the graph, we see that most of the nodes in

the network have a low degree of centrality score, while only a few have high degrees. Therefore, the network is regarded as highly centralized around just a few characters, in this case the main characters.

In Figure 7, the betweeness() function in the igraph package of R was used to plot degree centrality. Larger nodes depict a higher betweenness centrality score.

Figure 8 looks at the betweenness centrality of the network, by looking at the distribution of this in the network. From the graph, we see that most of the nodes in the network have a low betweenness centrality score, while only a few have high degrees. Therefore, the network is regarded as highly centralized around just a few characters, in this case the main characters.

The point of these visualizations is to get a deeper understanding of the social network in the show. From this we see more clearly who the central characters in the friendship network are and how important and influential they are in the show.

IV. Exploratory Visualizations of the Networks: Economic Exchange

In this section we look at the same visualizations as prior, but with our economic exchange network.

Figure 9 is a depiction of what I deemed the total economic exchange network in the entire show BoJack Horseman. This was created using R. I decided to treat this network as an undirected network, meaning there are no arrows, and every edge is reciprocated because economic exchange relationships always involve two parties (both characters need to be involved for some exchange to happen).

Figure 10 depicts the same network, but now nodes are colored according to the occupation of the characters. Figure 11 also illustrates the same network, but with nodes colored according to character type. Again, this is just simply a visualization where we can begin to make some assumptions, but everything is pure speculation.

V. Actor Prominence: Economic Exchange

For this section we are looking at the same measures of centrality just for the economic exchange network. This network is not directed so we do not have to worry and in-degree or out-degree centrality. All the isolates were also removed from the network as to not impact centrality scores (figure 12).

In Figure 13, the degree() function in the igraph package of R was used to plot degree centrality. Larger nodes depict a higher degree centrality score.

Commented [LF2]: Delete the score in R Lauren

Figure 14 shows the distribution of the edges in the network. From the graph, we see that a majority of the nodes in the network have a low degree of centrality score, while only a few have high degrees, and one has significantly more than others. Therefore, the network is regarded as highly centralized around just a few characters, in this case the main characters, specifically BoJack.

In Figure 15, the betweeness() function in the igraph package of R was used to plot degree centrality. Larger nodes depict a higher betweenness centrality score.

Figure 16 looks at the betweenness centrality of the network, by looking at the distribution of this in the network. From the graph, we see that most of the nodes in the network have a low betweenness centrality score, while very few have high degrees with one having a much higher degree than all the others. Therefore, the network is regarded as highly centralized around just a few characters, in this case the main characters.

With these visualizations, we begin to understand our network deeper. Both the friendship and economic exchange networks are highly centralized around the main characters in the show, as would be expected. It is interesting to note that our main character is very centralized in the economic exchange network, while in the perceived friendship network, the network is still centralized, but this centralization is more evenly distributed amongst the main characters.

VI. ERGMs: Perceived Friendship

ERGMs (exponential random graph models) are way of statistically analyzing social networks. ERGMs consider the presence and absence of edges in networks and then provide models for the network structures. They allow us to look at theories for why networks look the way they do – why did edges form in the network in the way they have. For this project, I was interested in looking at sociality and selective mixing as possible reasons to why edges may have formed in the ways that they did. We can think of sociality as “do members of group A tend to have more edges than members of group B?”. In the case of this project, I was interested in looking at if members of certain occupations tended to have more edges than members of other occupations to understand how occupation influenced the relationships in the show. I also looked

at character type to see if animal characters had more edges than human characters, as they were the more dynamic of the character types. When thinking about selective mixing, we are asking the question, how much more likely is it for two individuals in the same group to have a relationship (edge) compared to two individuals in different groups. Again, I looked at occupation and character type for this. Lastly, triadic closure was also looked at. Triadic closure looks at reciprocity; how likely is it for a "triangle" to form between three nodes when two of them already have an existing relationship.

Again, this analysis was done in R using an ERGM package. Figure 17 is the summary of the models that were run. The model is a summary of the probability. The ERGM code considers what theories you are looking at (ex: sociality, selective mixing) and then generate all possible networks that could form with the nodes you have. It then compares your network to the averages of all the ones created to tell you if the network is forming at random or if there is evidence of the factor/theory you are looking at.

Going through the summary table. Model 1 does not take any theories into account; it just looks at the edges and tell us how likely it for an edge to form in our network is. The significance does not really matter here, as this is just a baseline probability for how likely it is to form an edge in our network which is used for comparison to the other models.

In Model 2, sociality is considered, which is what the term *nodefactor* means. The term, *nodefactor.Animal.Human* is looking if those who are character type human are more likely to form an edge compared to those of animal type animal. Sociality was also looked at for occupation, comparing all other categories of occupation to the actor/actress category.

Model 3 then also considers selective mixing by occupation, which is what the term, *nodematch* represents. Here we are asking, how much more likely is it form two characters of the same occupation to form a tie compared to two of differing occupations.

Model 4 finally takes into all the above and triadic closure.

To see, which model represented our network the best, goodness of fit statistics was run. For this, we look to see if there are high p-values because a low p-value would mean our graph is

improbable. I also looked at MCMC stats. What we want to see here is all the graphs converging, meaning that our model is a good fit.

VII. ERGMs: Economic Exchange

The same thing was done as above, but this time for the economic exchange network.

VIII. Conclusion

Overall, this project was a way of learning about social networks and how to analyze them. Using a T.V show allowed the opportunity to have ample data for the analysis and draw conclusions about the show. In the future, the tools used for this assignment will be useful when analyzing real networks that we see in the world today. Social network analysis is field that is emergently becoming more popular and important in understanding human social life, so have the skills to analyze networks effectively will be important and valued in the coming future.