Analysis of Networks in the Star Wars Universe

By: Jared, Nicky, and Pierson

Introduction:

Our group approached this project with the intention of better understanding networks in action. In order to accomplish this goal, our group decided that we needed to select a data set that all group members could understand and relate with. After a quick discussion we realized that we were all fans of the Star Wars universe, so we decided to conduct our research using Star Wars networks data found on GitHub. Specifically, we chose to focus on a file containing information about the interactions between all characters across the films[1]. This data set contains a list of 110 characters who appear across various movies within the Starwars universe and provides context about the number of times characters speak in the same scene as another character. This data was last updated in September of 2018, meaning that it only contains information on characters who appeared in Episodes I-VII and excludes those who first appeared in Episodes VIII of IX.

Topics of Focus and Research Question:

We intended to use this Star Wars network data to better understand the way in which characters in the movie franchise are connected to each other. One issue we ran into, however, is the fact that the network data we pulled from GitHub was very sparse. The data contained information about the number of scenes each character spoke in across the films and similarly provided a pair of characters with the number of scenes in which the two characters both spoke, but the data failed to provide any more meaningful information. This network data seemed to be a good place to start, but we realized that we would need more information to truly bring this project to life and to better understand the networks within the Star Wars universe. We were able to accomplish this goal by extracting information from the Star Wars official data bank which contains detailed descriptions of almost all characters across the universe[2]. We chose to pull in

[1] https://github.com/evelinag/star-wars-network-data/blob/master/starwars-full-interactions.json
[2] https://www.starwars.com/databank

information about Affiliation, Location, and Species of the characters which we then used to help build our research question. We then came up with the following research questions: are we able to use the Affiliations (which group of faction a character belongs to), Location (where the character spends a majority of their time), and Species (species and whether they are a droid) variables to better understand why specific characters would appear and speak in the same scene as one another, and does this additional information provide further context that can help us analyze the films? These questions are important because they would help us understand which variables are the best predictors of interactions between characters. We could apply these findings the next time we watch Star Wars films, and focus on those factors which appear to be important to better understand the network between characters. We would consider this project to be a success if we can use our conclusions to shape what factors to focus on when watching Star Wars films in the future.

Research Question and Hypothesis:

To reiterate, our research question asks whether we are able to use the Affiliations, Location, and Species variables to better understand why specific characters interact with one another? We hypothesize that the Affiliations and Location variables will be important in understanding this relationship, and that Affiliations will have the most significant impact on these interactions out of all three variables.

We combined two metrics in order to conduct this analysis. First, we calculated the weighted degree for each node (character). This statistic essentially sums the weight or strength of all connections for a node and provides insights about the relative importance of each character within the data set; the most influential individuals within a network are those with the highest weighted degree score. Appendix 1 shows the weighted degree for each node where the

darker color signifies a higher weighted degree. In our example, the weighted degree for a character is calculated by finding all their direct interactions and then adding the total number of scenes between these pairs. We can see that Obi-Wan, C-3PO, Anakin, and Padme have the highest weighted degree, meaning they are the most influential characters across the Star Wars storyline. Appendix 2 shows the scatterplot with the distribution of weighted degree values for this data set. Appendix 3 shows a similar graph, but this time uses color to show the total number of scenes that a character appears in across the Star Wars films. Han Solo and Luke stand out in this visualization because they appear in the highest number of scenes total, which is interesting because they aren't among the characters with the highest weighted degree score. This finding is important because it gives the best indicator of which characters are important in the storyline, and we can focus on these characters when stratifying the data based on Affiliations, Location, and Species in order to see how these variables impact the main characters.

We have attached visuals of the graphs (Appendix 4-6) which are partitioned using the three variables mentioned above. We are using the size of the node to represent the total number of scenes for a character, and the line color to represent the number of scenes in which a set of characters interacts. We are judging the impact of each variable on our ability to better understand the storyline in the Star Wars films, and are primarily focused on how these variables influence the network of the most important characters. We are ultimately looking for variables which associate people who interact into the same grouping and provide some insights into the story. Appendix 4, which looks into Associations, provides some insightful observations about the networks within Star Wars. We can see that characters who share Affiliations seem to be clustered around and interact with each other. One thing that's interesting, however, is that the most important characters including Anakin and Obi-Wan interact with characters who identify

with multiple Affiliations besides their own. This is likely due to the fact that these characters are the protagonists of the story and have more scenes than most other characters, meaning they are more likely to interact with a wider scope of characters. This graph ultimately suggests that Affiliations play an important role in whether characters interact with each other. Appendix 5 showcases the network through the scope of location. We can see some clustering based on location, including characters who spent time on Naboo and Tatooine, but these groupings are somewhat hard to identify. Additionally, Star Wars has a massive universe and many of the main characters travel to various planets; Luke is listed under Tatooine but travels throughout the universe across the movies. The location variable can be somewhat useful, but we don't believe this is the best way to analyze the networks because a lot of overlap between groups exists and that a lot of characters aren't necessarily isolated to just one location. Finally, Appendix 6 shows networks broken out by Species. We can see that humans dominate the Star Wars universe as they make up nearly ⅔ of the whole dataset. There are a couple observations that are grouped together who aren't humans (Neimoidians on the left side of the graph) but there are not enough non-human observations to really say whether or not species is an important indicator of interactions between characters.

Based on these findings, it is appropriate to say that we have supported our hypothesis that Affiliations is the best variable through which to understand networks in the Star Wars universe. Characters who share Affiliations are clustered together meaning they primarily interact with one another. This makes sense because people within the same affiliation share a political goal, and are likely to work together in some capacity to accomplish this goal. Individuals in the Rebel Alliance, for example, are extremely unlikely to share scenes with

characters in the Galactic Republic because these are warring factions who would be fighting if they ended up face to face.

Hypothesis Dependencies:

Our hypothesis depends on whether characters interact closely in relation to certain shared characteristics. For example, we hypothesize that characters within the Jedi Order would tend to interact more often with one another (in more scenes together). This rationale exists from previous understanding of the Star Wars film series where certain characters interact more often than others. From our analysis, our hypothesis is more likely to be true for larger affiliation groups. This can be seen in Appendix 4, in factions such as Rebel Alliance, the Resistance, First Order and Jedi Order. They have the highest percentage of affiliates within our dataset and are closely affiliated within their networks. The Galactic Republic is the only large affiliation that is not seen to be closely related within its own network. However, this group is built of minor characters within the Star Wars universe, whose purpose runs through the main characters in the series. They do not carry much screen-time in the movie so do not have a close connection amongst the group. From this analysis, the hypothesis truly depends on the screen time throughout the film series. Lesser screen time will hold little to no effect on the strength of relationships within affiliations. One could argue that this time-dependency will hold no relation because characters with lesser screen time. In rebuttal, the smallest size bubbles in the affiliation network demonstrates relationships within their affiliation. Overall, screen time holds a power over the closeness of the relationship between characters.

Refined Testable Hypothesis 1: Star Wars characters, with a high proportion of screen-time, with certain affiliations, gender/species, etc. will interact more amongst their network than

others. Specifically, the relationship between qualities and network strength relies on the screen-time of the characters.

The hypothesis also depends specifically on the locations of characters to build a strong network within. This hypothesis depends on a weak location variable that relates where characters are most associated. One will argue that characters tend to move to different planets every few scenes and don't build strong relationships on each world. But, the one strong location relationship in hypothesis is on the planet Tatooine. Important Star Wars storyline pieces are centered around this planet, so a stronger relationship is found. To improve this hypothesis, the variable could be adjusted to the "Location of Interaction" of the characters. Characters tend to move from planet to planet with a high frequency. This new metric will evaluate a novel network within the Star Wars universe. The color of the network bubbles can be adjusted to represent the planet on which they had their highest frequency of interactions. This method will develop a better understanding of which characters tend to spend more time together in addition to affiliation.

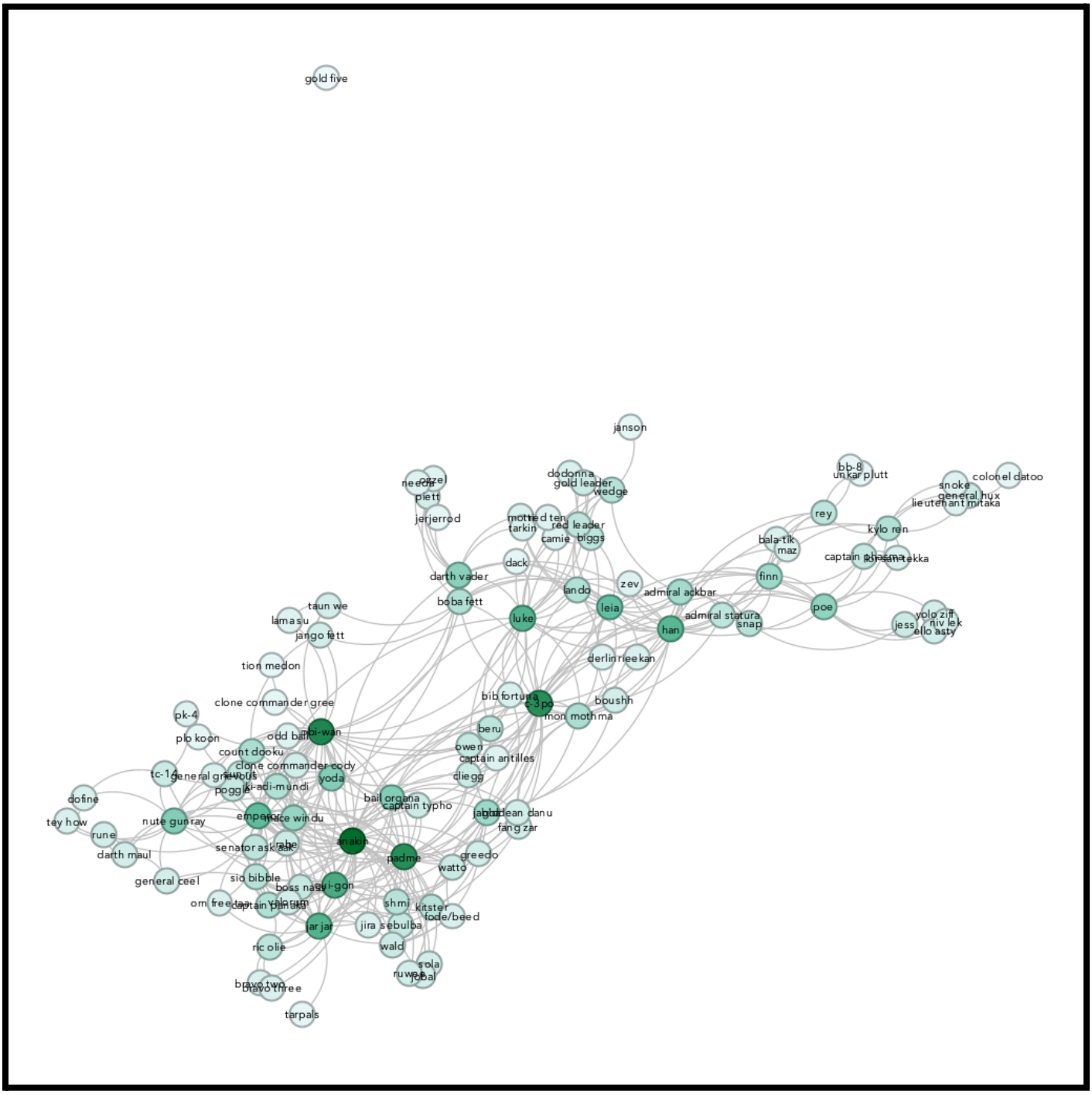
Refined Testable Hypothesis 2: Star Wars characters, with certain affiliations, gender/species, etc. will interact more amongst their network than others. Specifically, the relationship between character qualities and network strength relies on the screen-time of the characters and the locations at which they interact.

Recommendations and Conclusions:

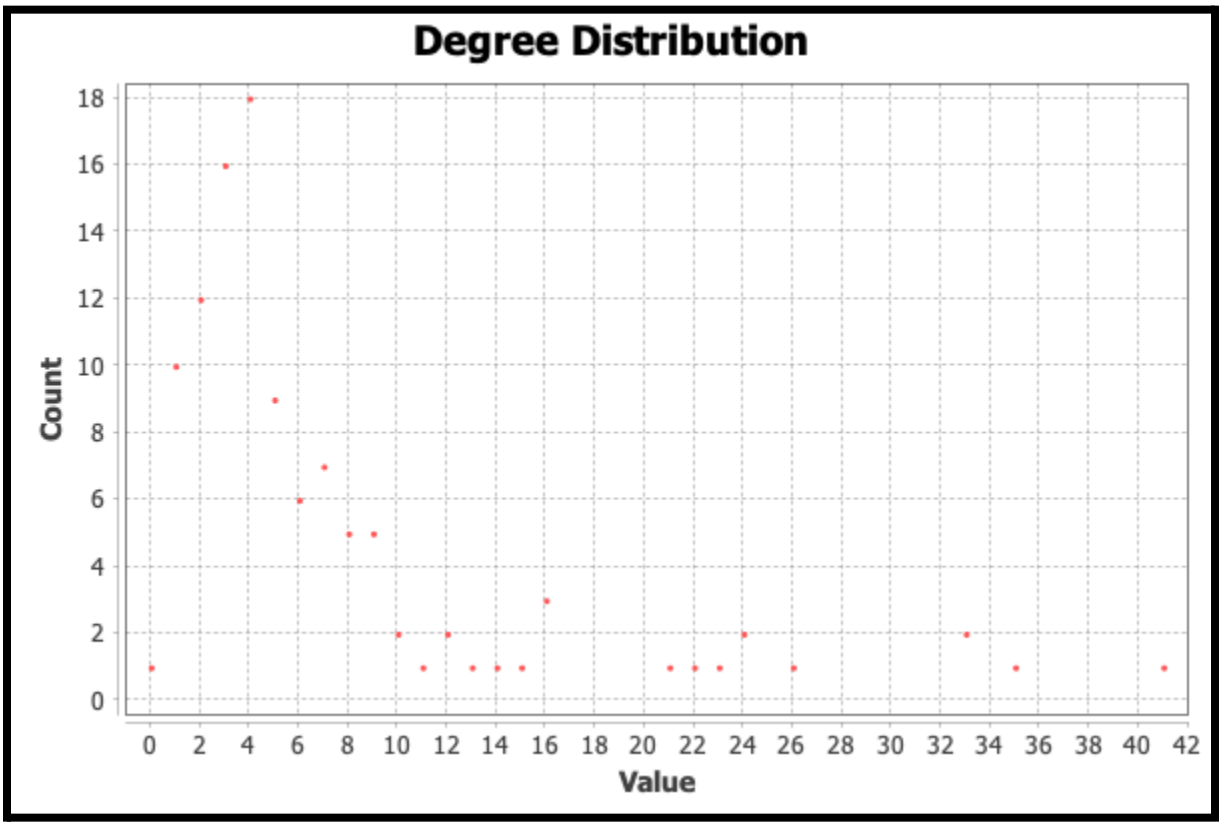
This analysis of the Star Wars universe does not just hold implications for the characters within the series. If the hypotheses were supported, the general knowledge could be utilized by even managers at consulting firms. The dataset and character behaviors are remarkably similar to those of travel consultants. The interpersonal strength of travel consultants diminishes as they

spend more time away from their peers. This is similar to characters within the Star Wars as they build relationships with characters they do not interact with too often. While also developing incredibly strong connections with those in their same affiliation (essentially same business). A manager could find this important when deciding to send employees on a work trip or keep them in the office to develop stronger inter-office relationships. They could sacrifice personal relationships in the form of developing a new connection by sending an employee on a work trip. Salesmen could also use the ideology that more interactions mean stronger relationships. If they send employees to continually follow up with current connections they will develop a bond. This tactic may not work at earlier stages of a network but can be used to further a current relationship. Overall, this analysis holds benefits for Star Wars enthusiasts as well as members of the business community. The structure of networks is an important concept for successful teams and fictional relationships.

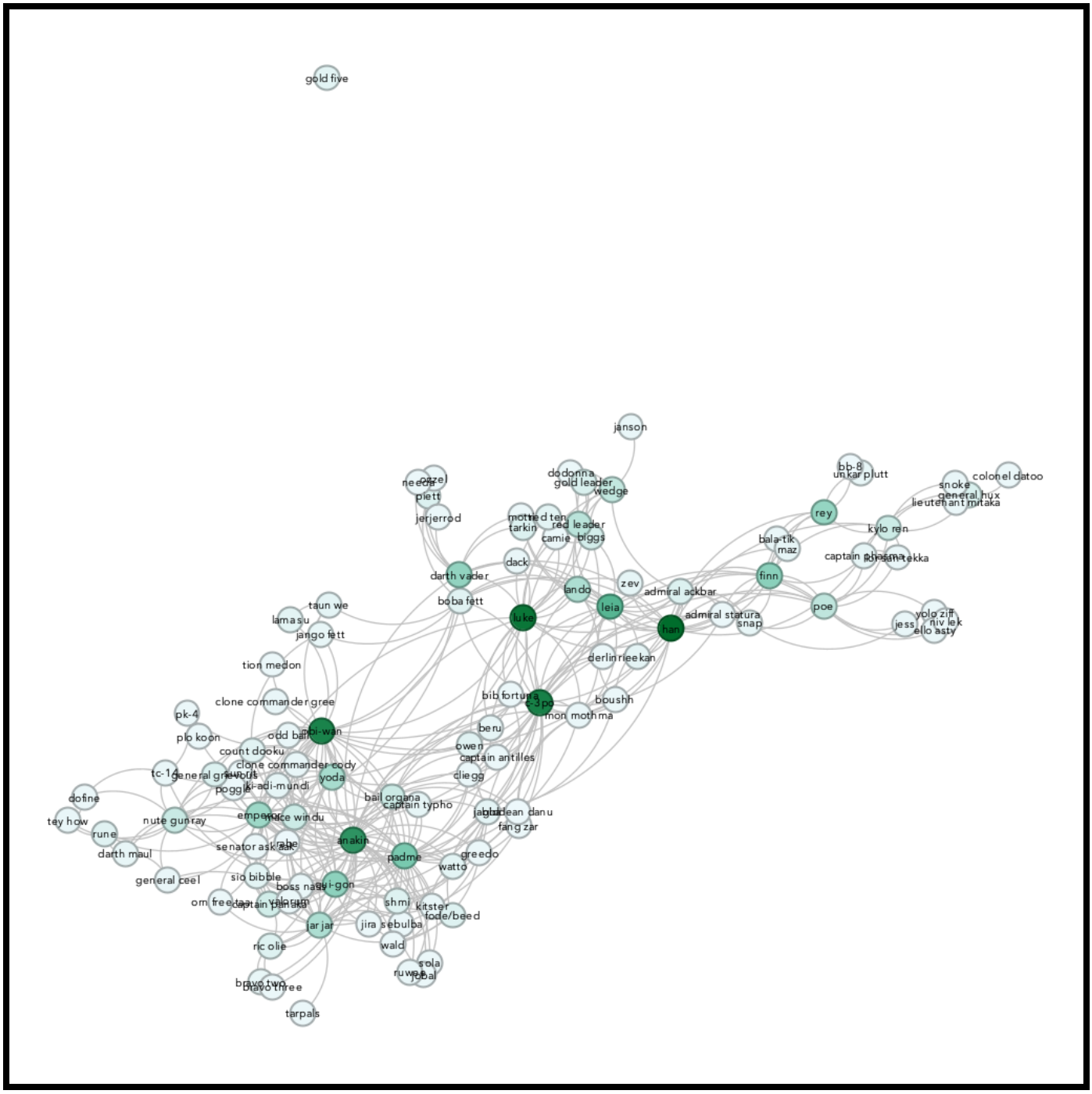
Appendix



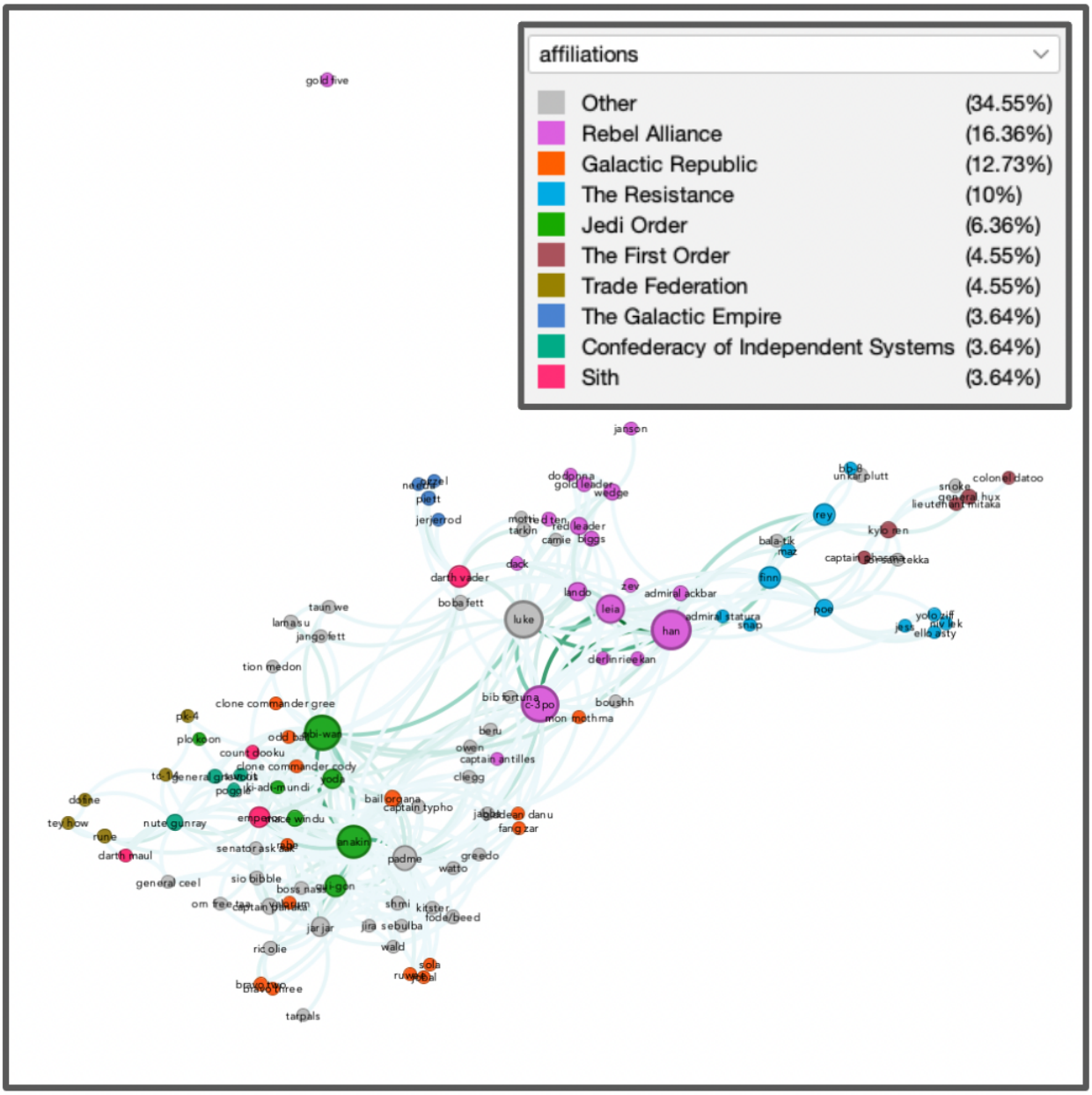
Appendix 1- Weighted Degree Graph



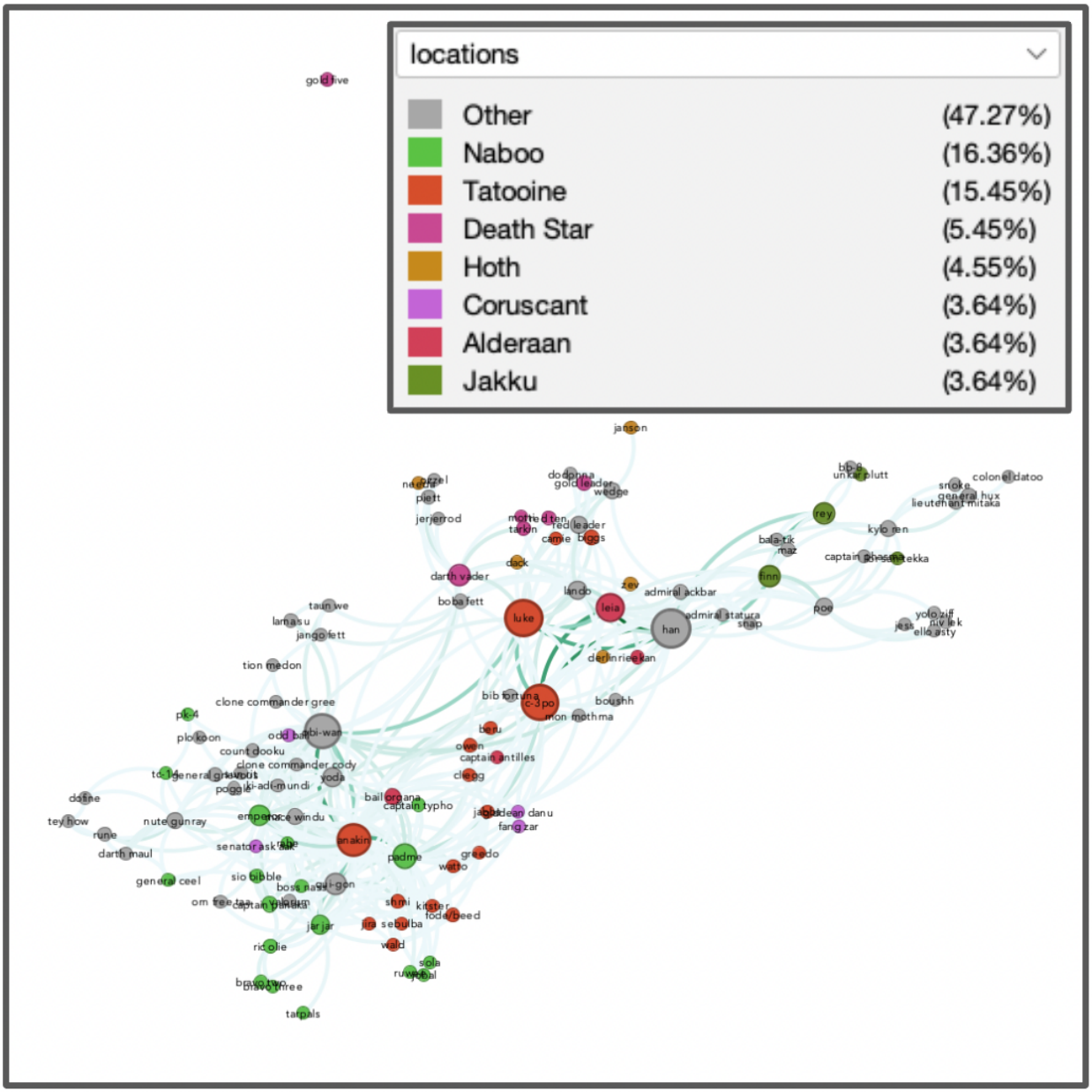
Appendix 2- Degree Distribution Scatter Plot



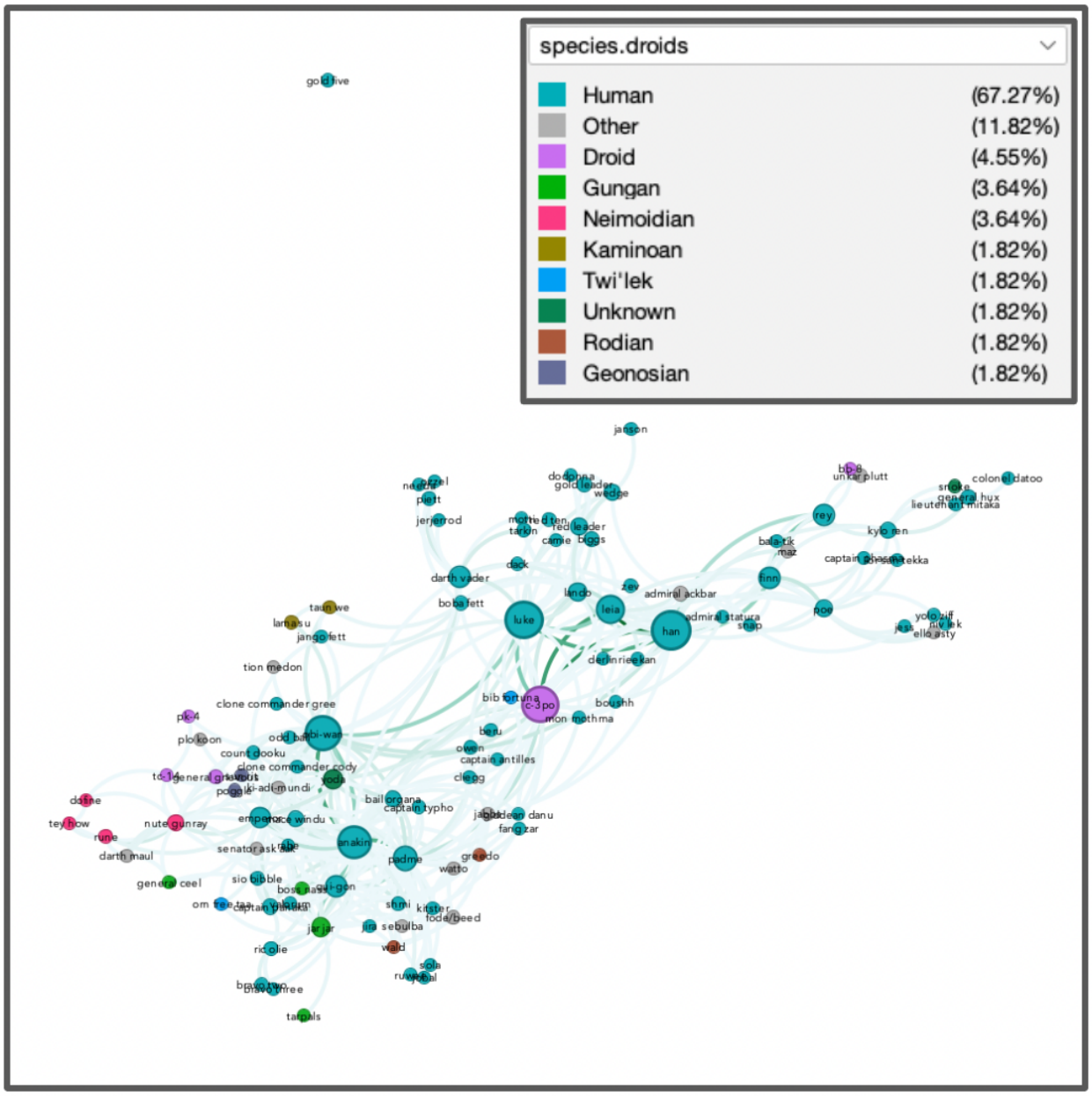
Appendix 3- Number of Total Scenes Graph



Appendix 4- Affiliations Graph



Appendix 5- Location Graph



Appendix 6- Species Graph