# Final project

[redacted]

## Introduction

This semester, we've spent a lot of time analyzing networks that exist in the real world, using them as examples to learn about various theories of social interaction. For this project, I thought it would be interesting to take a step away from these 'real' networks, and instead relate some of those same social network theories to a piece of popular media. In doing this, I want to find the differences and similarities between real and fictional networks, and analyze them in a way that may allow for a better understanding of why those differences exist. More specifically, I want to see if my chosen media relies on any network theories to further the plot and/or characterization, or conversely, if it avoids portraying certain aspects of those theories in order to maintain a believable narrative.

For this project, I decided to use the network of Star Wars character interactions from the 'networkdata' package we've used in our labs. I chose to analyze this network specifically for the primary reason that Star Wars is an extremely popular franchise, and if there are any deviations from what would be expected of a real-world network, there's a chance the films' popularity could be partially attributed to these fictional aspects of social interaction. In other words, I think it would be interesting to discover some differences in the social networks of the movies that may be perceived by audiences as 'better' than real life, and therefore make the world of Star Wars more desirable. For the purposes of this project, I decided to focus on the networks of the prequels: Episode I – The Phantom Menace (1999), Episode II – Attack of the Clones (2002), and Episode III – Revenge of the Sith (2005). This is mostly due to the fact that these networks have more characters; they're a bit more reliable in terms of actually applying network theories because of this.

In addition to the Star Wars network data, I wanted to include visualizations of one 'real' network to use in comparison. For this purpose, I chose 'ffe_elite' a network of interactions between French financial elite that we've already worked with a few times in class. I chose to use this network as a comparison because the data contains similar node attributes, allowing for some more precise comparisons.

## Network Theories

The two theories I want to focus on are centrality and the idea of structural holes.

Centrality measures "the locations of individuals in terms of how close they are to the 'center' of the action in a network" (Hanneman & Riddle, 2005). There are several different measures for centrality; the ones covered in this course are degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality. For this project, I'm focusing on betweenness centrality. This is a measure of centrality that calculates the number shortest paths that travel through each nose. In other words, "what's the number of shortest paths that [k] lies on between other nodes?" (Jackson 2016). The theoretical idea behind calculating betweenness centrality is that if a node is placed along a higher number of these 'shortest paths,' it may have access to more unique information, resources, or power than another node. This relates to the second idea I want to look at with this project: the idea of a structural hole.

The term 'structural hole' refers to "the situation in which [a node] connects individuals who are themselves connected but who, without the presence of [that node] would have no connection with one another" (Kadushin, 2018). To analyze this within my chosen networks, I made basic visualizations of each network, and noted instances where nodes or groups of nodes would not be connected in any way without the presence of a 'bridge' over the structural hole.

## Methods

To analyze these networks and the differences between them, I first did some visualizations of node attributes. I chose the 'ffe_elite' network as a comparison to the Star Wars network because of the number of node attributes available within its data, so I wanted to play around with both of them to find some similarities.

After some basic visualizations, I focused specifically on structural holes: I wanted to observe both graphs to find nodes that may act as 'bridges' for information, connecting groups that wouldn't otherwise share any edges. To answer my research questions, I compared any instances of structural holes I found in the networks, noting any significant similarities and differences in the nodes that acted as bridges over these gaps in information.

Another aspect of social network theory I wanted to focus on with this project was centralization. Using R, I calculated betweenness centrality with each network to find any similarities. I also identified the most central nodes of each network, and compared the node attributes of the nodes with highest betweenness centrality.

## Analysis

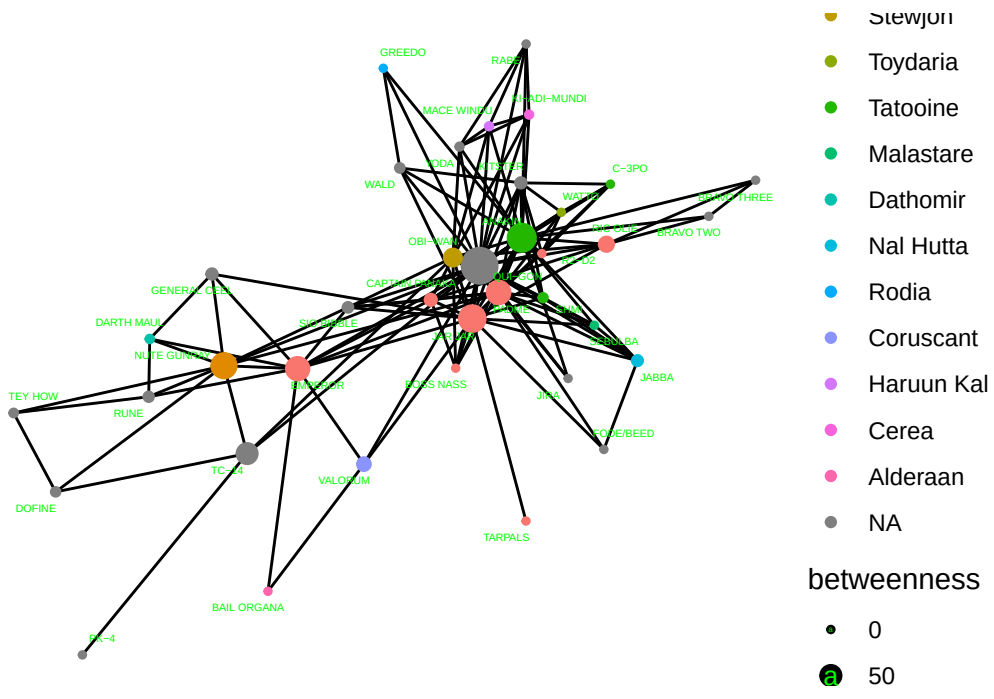
Below are graphs of each of the Star Wars prequels in order (1-3). The edges represent character interaction, and each node corresponds to a character.

Because I'm comparing these networks with the ffe_elite network, I wanted to visualize an aspect shared by nodes in both networks in order to make some comparisons. A node attribute that is found in both sets of data is 'birthplace,' or 'homeworld' in the Star Wars data (I'm treating the two as equivalent for the purpose of analysis). In these graphs, I've colored the nodes according to this attribute.

The second measure found in these graphs is betweenness centrality, as discussed earlier. each node's size corresponds to its betweenness centrality, with a larger size denoting a higher centrality level.

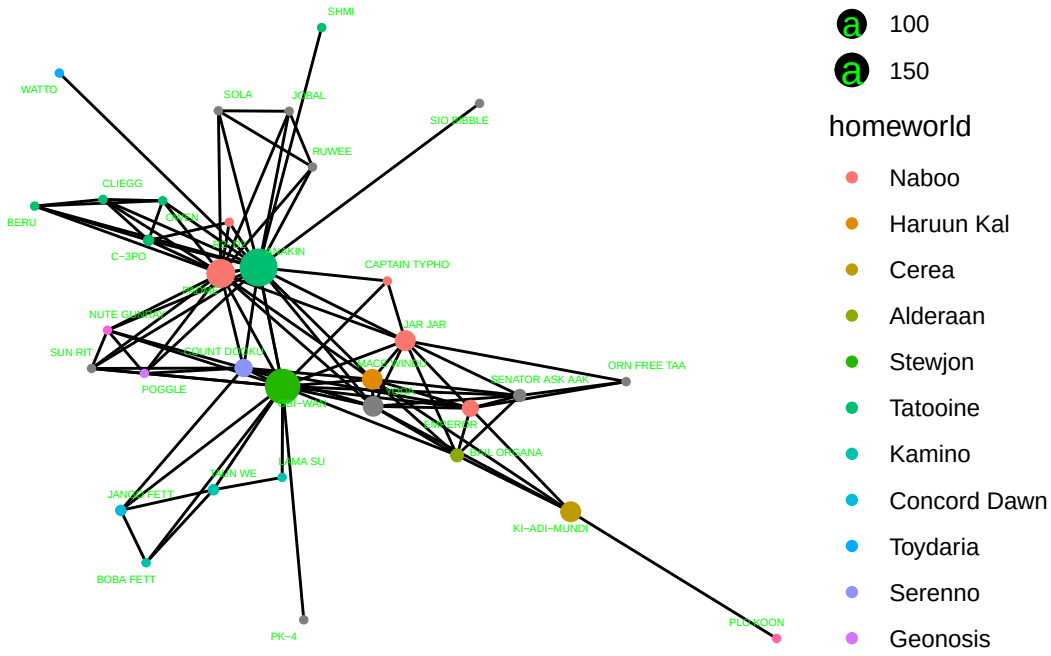
Star Wars EP 1 (The Phantom Menace):

```
G <- as_tbl_graph(starwars[[1]])


G |>
  mutate(homeworld = as_factor(homeworld), betweenness = centrality_betweenness()) |>
  ggraph(layout = "fr") +
  geom_edge_link(aes(edge_alpha = weight), show.legend = FALSE) +
  geom_edge_fan() +
  geom_node_point(aes(color = homeworld, size = betweenness)) +
  geom_node_text(aes(label = name, size = .5), color = "green", repel = TRUE) +
  theme_void()
```



Star Wars EP 2 (Attack of the Clones):

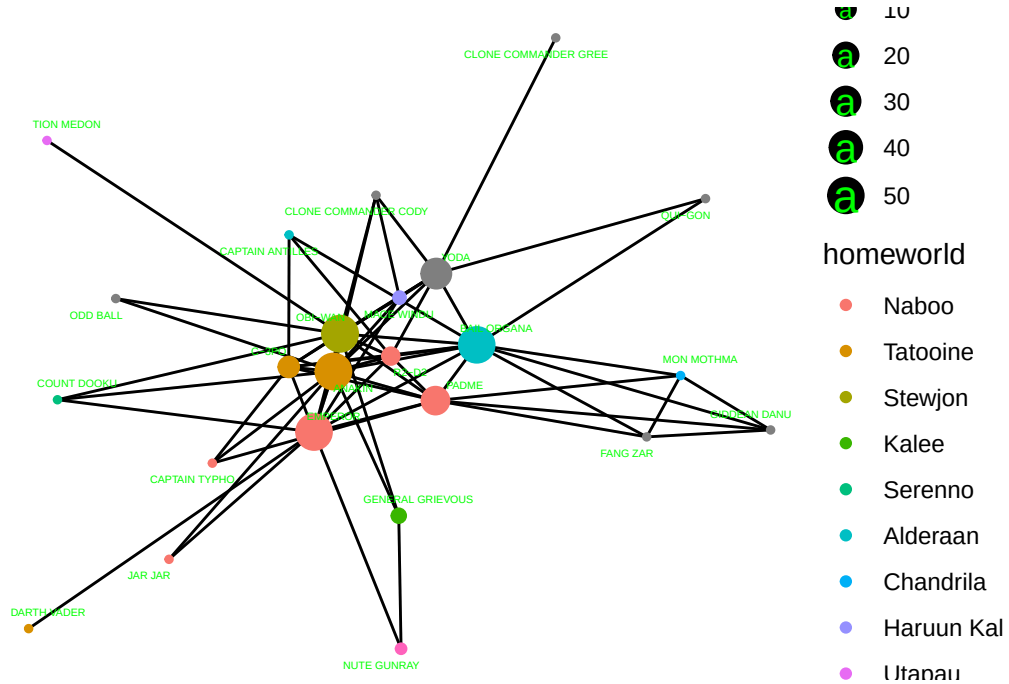
```
G <- as_tbl_graph(starwars[[2]])


G |>
  mutate(homeworld = as_factor(homeworld), betweenness = centrality_betweenness()) |>
  ggraph(layout = "fr") +
  geom_edge_link(aes(edge_alpha = weight), show.legend = FALSE) +
  geom_edge_fan() +
  geom_node_point(aes(color = homeworld, size = betweenness)) +
  geom_node_text(aes(label = name, size = .5), color = "green", repel = TRUE) +
  theme_void()
```




Star Wars EP 3 (Revenge of the Sith):

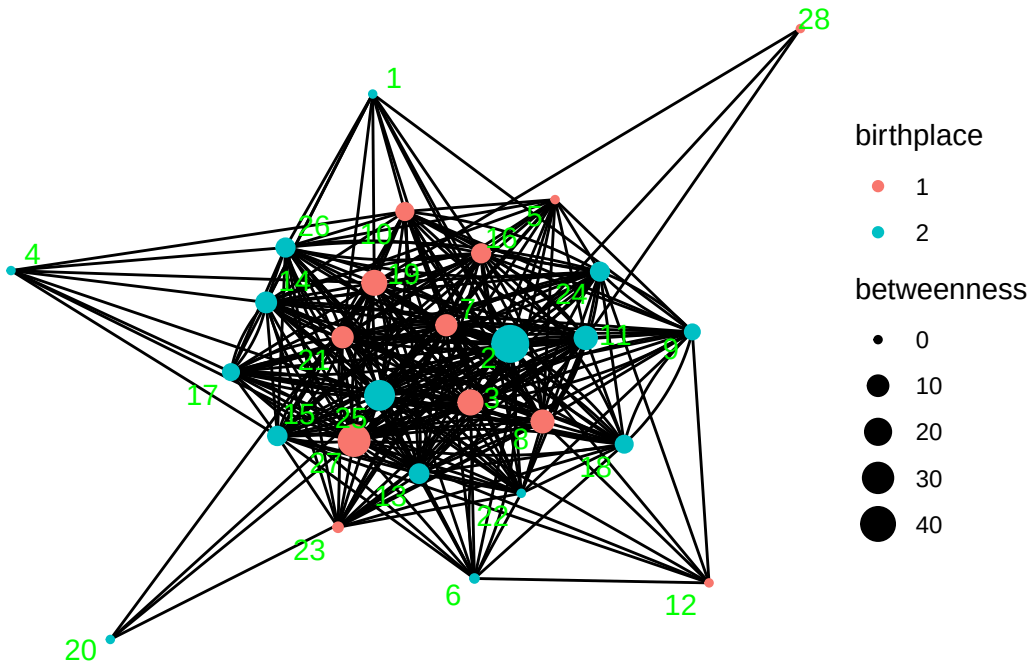
```
G <- as_tbl_graph(starwars[[3]])


G |>
  mutate(homeworld = as_factor(homeworld), betweenness = centrality_betweenness()) |>
  ggraph(layout = "fr") +
  geom_edge_link(aes(edge_alpha = weight), show.legend = FALSE) +
```

```
  geom_edge_fan() +
  geom_node_point(aes(color = homeworld, size = betweenness)) +
  geom_node_text(aes(label = name, size = .5), color = "green", repel = TRUE) +
  theme_void()
```




Below is the graph derived from the ffe_elite data, also modified to show centrality and birthplace in the same ways as the Star Wars networks:

```
G <- as_tbl_graph(ffe_elite) |>
  activate(nodes) |>
  mutate(node_id = row_number(), birthplace = as_factor(birthplace), betweenness = centrality

ggraph(G, layout = "fr") +
  geom_edge_fan() +
  geom_node_point(aes(color = birthplace, size = betweenness)) +
  geom_node_text(aes(label = node_id), color = "green", repel = TRUE) +
  theme_void()
```




When observing these graphs, a few things become apparent. I first want to analyze the Star Wars graphs. I don't think the centrality shown in the graphs is very important by itself - because these networks are fictional, the centrality of any node is determined solely by plot relevance. Despite this, I do think it's interesting to observe how certain characters (e.g., Emperor Palpatine) become more relevant throughout the series through the lens of centrality.

My actual research questions lie in the comparison between these graphs and ffe_elite, so I'll move on to that graph now. Before creating the graph, I was hoping that there would be some relationship between birthplace and betweenness centrality. Looking at the graph, I don't think any significant relationship exists. I think that nodes from birthplace '2' do tend to have a slightly higher betweenness centrality, but it's not enough of a difference to be a really interesting result.

Comparing the two graphs, I think there's a few things worth noting. Following the 'birthplace' and 'homeworld' attributes, it's clear in both the Star Wars and ffe_elite datasets that two nodes being from the same region makes them more likely to share a connection. In addition to this, there are instances of 'structural holes' found in both datasets. Among the more notable 'bridges' in the Star Wars graphs are Emporer Palpatine, Anakin, and Obi-Wan (Obi-Wan's centrality is much higher in the second and third installments than in Episode I). The most notable in the ffe_elite data are nodes 11, 27, and 25.

## Discussion

I think the results of my analysis have a few interesting implications. I stated above that sharing the 'birthplace' attribute is strongly related to the likelihood that nodes are connected in both the fiction and nonfiction networks. This observation obviously makes complete sense for real-world interaction, but an interesting deviation from this pattern occurs in the Star Wars networks. While it remains true that region of origin is still related to node connection, this is only true of minor and supporting characters. The main protagonists and antagonists, in contrast, do not share this attribute. It's not just unlikely, it's completely absent from the networks. This is really interesting to me, and definitely answers my search for a notable difference between the networks. I think this discovery can be attributed to the nature of the fictional network: because it takes place in an expansive galaxy, Star Wars uses its characters' geographical backgrounds for worldbuilding. By having a character be from a location in the world of the story, writers can provide information about that location through the character. Additionally, Star Wars is a sci-fi franchise, and some of its novelty stems from the inclusion of unique planets and alien species. Because of this, it makes sense that there's diversity in the homeworlds of the main characters.

One of the aspects that remained relatively similar across Star Wars and ffe_elite, and one of the aspects I initially set out to compare between the two, was structural holes. While there wasn't any similarity in the nodes that acted as bridges over these structural holes, there was a similar percentage of structural holes in each network. I think that if I had access to more node attributes, it may be easier to find a pattern, and maybe there would be a more significant finding. However, because the data was limited in this way, I'm content with the basic affirmation that structural holes do exist in both datasets, and occur at a similar frequency in both the fiction and nonfiction networks.

Overall, I'm pretty happy with my findings. I discovered one difference and one similarity between the two sets of data, and I think both have some interesting implications regarding social connection, storytelling, and the aspects of networks that audiences want to see portrayed onscreen. In future research, I think it would be interesting to study some other fictional networks, and compare a variety of plotlines and genres to a variety of real-world networks. I also think the lack of node attributes was a limitation when answering my research questions; it would be interesting to replicate these in another project with similar networks whose nodes had more documented similarities. I think my basic findings are really interesting, and I'd love to see what the results would look like if the research was ever expanded.

## Sources

(2016). *Social and Economic Networks 2.3 Week 2: Centrality Measures.* Retrieved from https://www.youtube.com/watch?v=0unzqsPaPk8&ab_channel=SocialandEconomicNetworks.

Hanneman, R. A., & Riddle, M. (2005). Chapter 10: Centrality and Power. In *Introduction to social network methods*.

Kadushin, C. (2018). *Understanding social networks: Theories, concepts and findings.* Oxford University Press.