

The background of the slide is a cinematic image from Star Wars. It shows the Millennium Falcon flying low over a vast, brown, desert-like landscape. In the distance, two suns are visible in a hazy sky, creating a warm glow. The ship's iconic blue lightsaber-like engines are glowing at the bottom. The title text is overlaid on the left side of the image.

# STAR WARS UNIVERSE USING FORCE DIRECTED GRAPH IN D3

Rohit Srinivas, Vaibhav Beohar, Vaishali Khandelwal

# PROJECT MOTIVATION

Demonstrate the use of Force Directed graph using D3.js combined with the eagerness to learn network skills and our love for Star Wars. The primary motivation is to showcase graph technology using the Star Wars universe example, using characters as embodiments of a network.

## MARKET POTENTIAL FROM GARTNER'22:



Innovations  
fueled by Graph  
Technology\*



Market by 2026\*



Compound  
Annual Growth\*



\*Source(s):

<https://www.dbta.com/DBTA-Downloads/WhitePapers/The-Graph-Revolution-Accelerating-Analytics-and-AI-in-the-Enterprise-11692.aspx>

<https://towardsdatascience.com/graph-ml-in-2022-where-are-we-now-f7f8242599e0>

# TARGETED USERS

- Star Wars fans
- Graph technology experts
- Data and Visualization enthusiasts

## PRIMARY TASKS

- Visualizing the Star Wars character network for every episode individually
- Interpreting and analyzing the character importance from the network
- Character lifespan through the episodes
- Interpreting primary graph metrics like betweenness and degree of centrality from the network

# DATA OVERVIEW

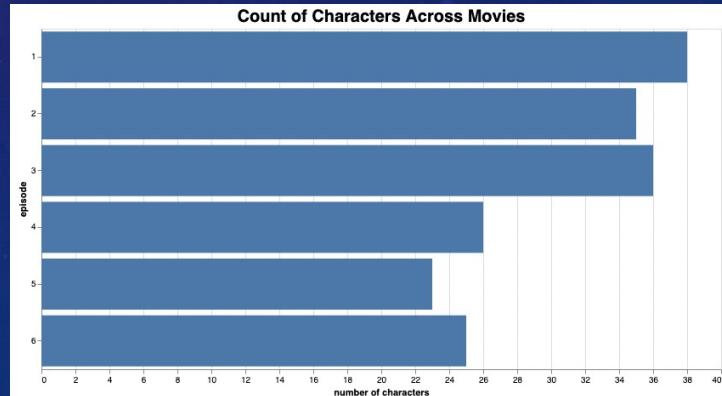
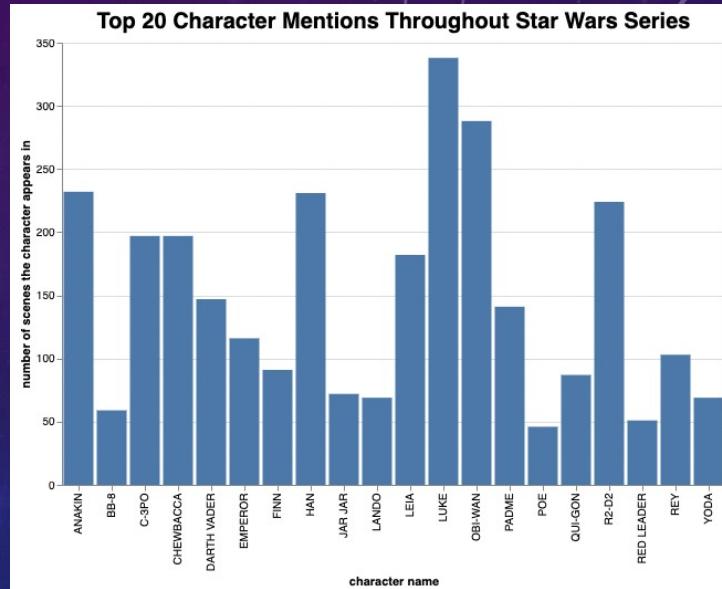
Data representing characters and their interactions for all seven movies has been extracted from the JSON files and mined here.

For each episode, the dataset includes:

- Interaction dataset - Social network with links between characters defined by times the characters speak in the same scene
- Mentions dataset - Social network with links between characters defined by the number of times they are mentioned in the same scene
- Social network with character links and mentions for all 7 episodes

As a first step, we performed EDA on the dataset to ensure -

- It fit graph structure (i.e., presence of nodes and links)
- Aligned with our intuition of Star Wars universe



# WHY GRAPHS?

Through the EDA, it was clear that a graph network would fit this dataset. The primary reasons for using a graph-based approach for this project:

- Characters would serve as nodes
- Interactions between the characters would serve as edges
- Using graph metrics like betweenness and degree of centrality would help understand the importance of characters and their role in the movie
- Explore a new graphing technology to build on D3 skillset

...And last but not the least, Graphs are the future of viz!

# GRAPH METRICS

Using metrics from the networkX python library:

**Degree Centrality** - Number of connections the node has in the network.

- Corresponds to the total number of scenes where each character speaks.
- To find extremely connected individuals or popular individuals

**Betweenness** - How many shortest paths in the network lead through the node

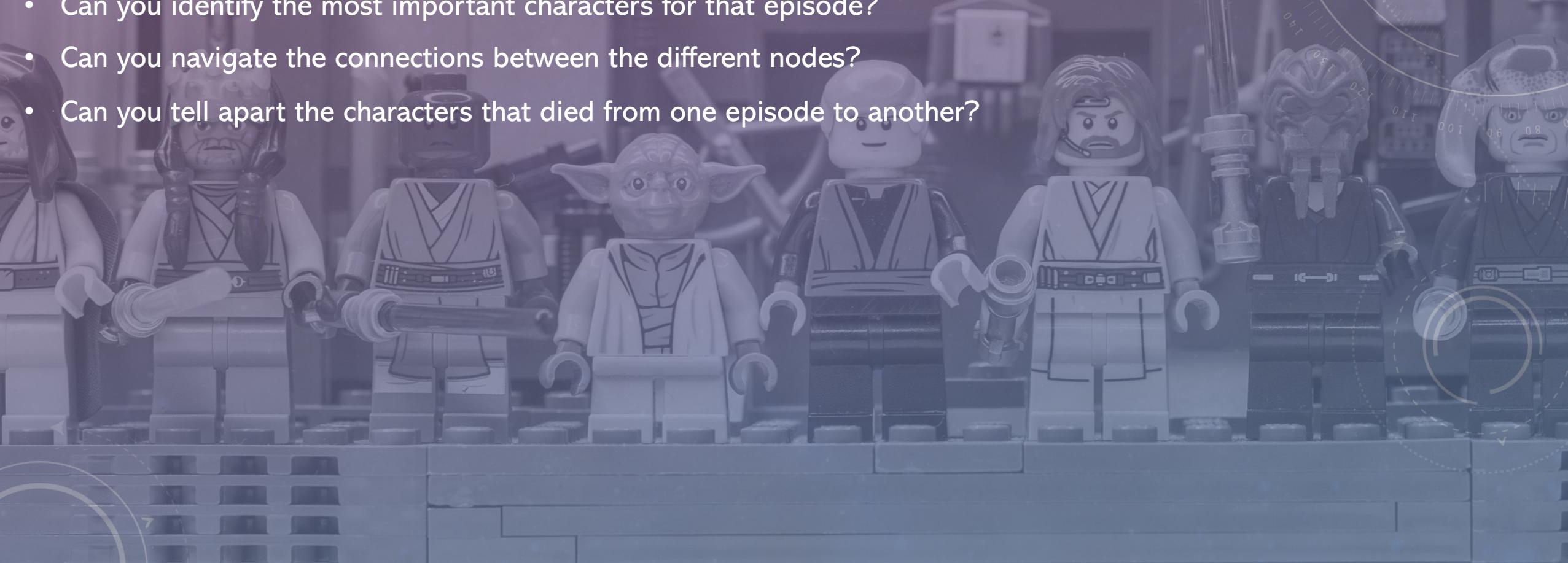
- Shows which nodes are “bridges” in the network (how many times each node falls on all shortest paths in the network)
- Could indicate authority over disparate clusters. Obi Wan or R2D2 seem to be more “centrally” connected to the storyline vs Anakin the sequels!

# USABILITY TESTS

We conducted user tests on a wide array of user personas to get holistic feedback on our project. The user persona included data enthusiasts, Star Wars fans and users with neither of these backgrounds.

The primary tasks included:

- Can you identify the most important characters for that episode?
- Can you navigate the connections between the different nodes?
- Can you tell apart the characters that died from one episode to another?



The graph transition from episode to episode is too jarring... can you make it more subtle!

As a user, it's easy for me to tell who's connected to whom..

What do the metrics mean? Can you place a clear explanation somewhere?

Would be nice to have clear labeling and meta-data.

What does each circle represent? How do I make sense circle represents?

Can we place the slider in the center? Also, the overall layout of the page needs to be more user friendly?

Metrics and labeling of nodes could use some standardization..

As a future enhancement, you could perhaps provide a "search character" functionality to easily find and highlight characters!

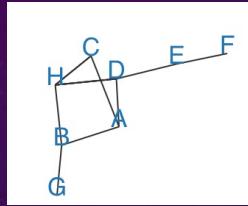
Great viz!!

It's fun to play with!

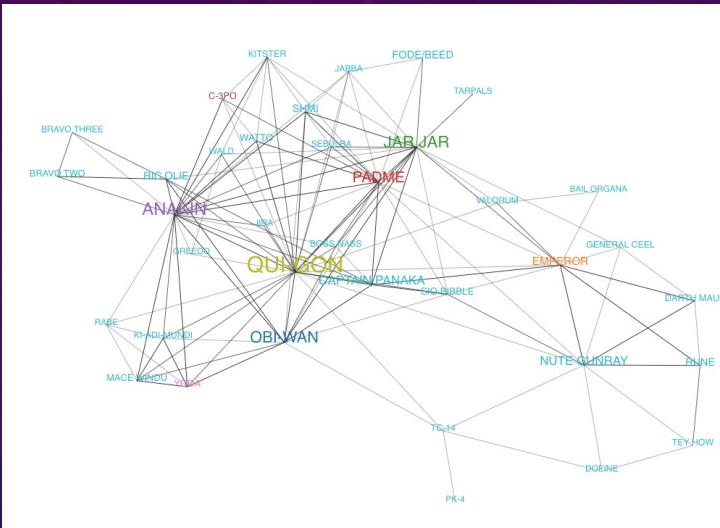


## USER FEEDBACK

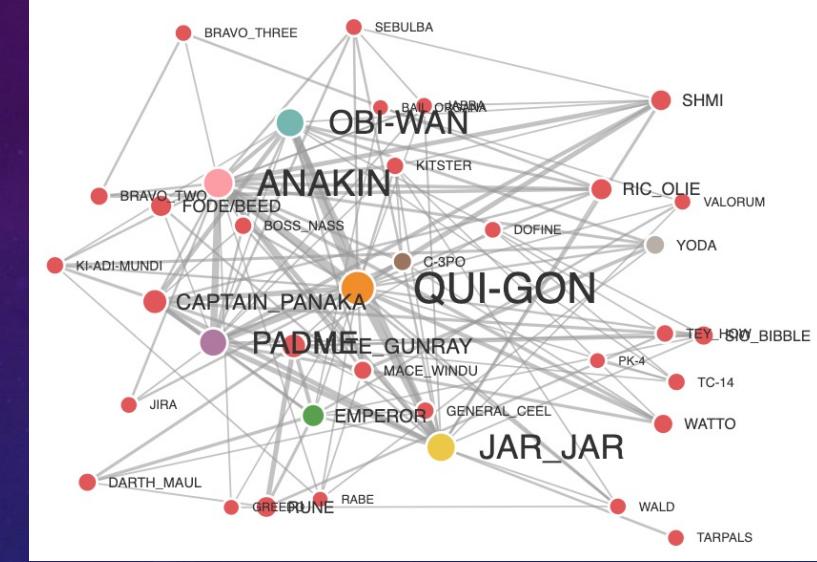
# ITERATIONS, FIXES, AND USER FEEDBACK UPDATES



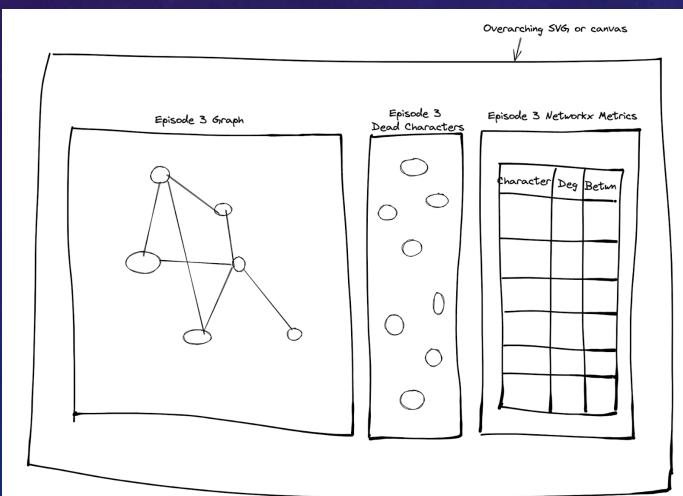
From this..



To this..



...ideation..



...and problem solving along the way...

- Using NetworkX to create graph metrics (degree, betweenness)
- Reading D3 documentation to fix JSON files
- JavaScript issues with event-handling (mouseover, tooltip)
- Hyperparameter tuning with D3 Force library
- Bootstrap and integration with Flask
- SVG issues!!!

# DEMO..

[Link to the demo](#)



# KEY INSIGHTS..

Social networks attain peak density in episodes 1 & 2 and subsequently fade out as the series progresses (not necessarily bad)

Episode 1 has the highest number of characters in the series and the most "connectedness" (as seen in the graph visualization)

Some characters have a recurring appearing in subsequent episodes even when they die earlier in the series (e.g., Qui-Gon in episode 3)

## Original Trilogy (Episodes 4, 5 and 6):

The original trilogy series are sparser with lesser characters than the prequels. This results in more social interactions, leading to more linear story lines and ease of narration. One can also notice that the smaller number of central characters also leads to tighter cohesion of degree and betweenness centrality measures

## Prequel Trilogy (Episodes 1, 2 and 3):

Qui-Gon is the most important character in episode 1; but then dies in episode 2 (as can be seen by an orphan node on the right). Anakin then becomes the most connected character in episode 2, while Obi-Wan Kanobi is not too far behind. Thus, it's interesting to see the interplay of character importance across different episodes in the prequels.

The prequel trilogy has more nodes overall, with many more connections (as can be seen by larger social networks with denser connections). It is also interesting to note that though Anakin is one of the most important characters, his network metrics are not always the highest! This means that though overall he's the most connected character in the films (high degree centrality), not all relations necessarily flow "through" him (low betweenness scores)

*TL;DR - Prequel trilogy has more overall characters, but no centrality. The stories are not bound by a central character (even though Anakin should be binding them all together). On the other hand, the sequels have sparser networks but more linear and tightly-knit storylines.*



# TECHNIQUES USED..

- Python libraries like Matplotlib, Pandas, Plotly for EDA
- D3.js for force directed graph
- NetworkX package in python for deriving network metrics
- Website setup in HTML (Bootstrap Framework)



# PROJECT NEXT STEPS..

## Make the graph viz more interactive

- Character search functionality
- Adding directionality to links (e.g., Antagonist nodes repel, friendly nodes attract)
- Creating timelines for better visualization of stories
- Incorporating more user feedback
- Expand to more movie series (e.g., Game of Thrones, Harry Potter)
- Make this a living project!

## Continuing our graph learning journey

- Creating “real world” knowledge graphs using D3-Force (e.g., Corporate hierarchy graphs, LinkedIn social networks)
- Visualization D3 and enhanced JavaScript functionalities to visualize data science use-cases

A black and white photograph of two lightsabers crossed in a dark space. The sabers are held diagonally, their blades forming an 'X' shape. The light from the blades creates bright, glowing lines against the dark background.

# MAY THE (D3) FORCE BE WITH US!