CSC 555+455 SoCDAI
Hypotheses to validate for P1 (Part 2)
Due: 10:00 pm ET Sept 6, 2020

This document contains the two hypotheses (crowdsourced from you all in part 1) that you need to validate for Part 2 of your assignment P1 along with all the other analysis of the social network.

Validate both the hypotheses on all episodes and then on the overall (combined) social network of the entire star-wars universe (all episodes combined). Report anything interesting you discover. Also, report if your findings from each episode hold up to be true in the overall universe or does it change.

Hypothesis 1: Characters can be characterized as on the light side or dark side by comparing their homophily to the subnetwork they belong to.

Argument to support the hypothesis: Since the Dark and the Light side are opposing forces we can expect to find more interactions within each group compared to across the groups. And as our graph is constructed based on the interactions between characters it will capture this information.

Instructions on how you can analyze this:
You have to analyze if it is possible to identify if a given character belongs to the Dark side or the Light side using some metrics of the social network constructed for the Star Wars universe?

- You can assume you have information on one important character in the social graph. You can choose any character you wish to use (explain why you chose that character), let's say, you know who Luke Skywalker is in the social network (but have no idea who the other nodes represent), can you identify the two communities (Dark side and the Light side) starting with this one representative node?
- Report your results in terms of accuracy for the top 10 predicted nodes, predicting if a given node belongs to the dark side or light side. Include the metric you are using to determine this and how do you determine which are your top 10 predictions (it can be based on some threshold or some confidence score derived from some metric based on the social network).
- Here is a blog that will help you with characters in Star Wars: https://www.deseret.com/2015/12/17/20579103/a-beginner-s-guide-to-star-wars-and-its-characters

Hypothesis 2: The character of a series, developed to be the "main" character will have the highest degree centrality, however the most "integral" character of a series will have the highest betweenness.

Argument to support the hypothesis: The main character is many times not the most integral character, thus a character with the highest degree centrality is unlikely to also have the highest degree centrality.

Instructions on how you can analyze this:

- Test this for 5 central characters for each episode and see how this changes in the overall universe (all episodes put together)
- Essentially you have to analyze based on the graph who are the central characters (state the metric you use). Does it contradict from what you expected?
- Mention the metric you use, is there any other metric which is better or worse to measure the importance of a node and does it reveal anything else about these graph.