**Week 4 Assignment**

**Centrality measures can be used to predict (positive or negative) outcomes for a node.**

**Your task in this week’s assignment is to identify an interesting set of network data that is available on the web (either through web scraping or web APIs) that could be used for analyzing and comparing centrality measures across nodes. As an additional constraint, there should be at least one categorical variable available for each node (such as “Male” or “Female”; “Republican”, “Democrat,” or “Undecided”, etc.)**

**For this week’s assignment, you are not required to actually load or analyze the data.**

**DATA SET**

Character interactions from the book series “A song of Ice and Fire” (a.k.a Game of Thrones) by George R. R. Martin, with the categorical variable of which of the major religion they belong to.

This data set may not exist as such, but could be obtained from IBM’s Watson, since Watson has already been used for various text analysis on the series, it seems plausible that Watson could also identify the interactions between characters and which religion they are associated with via text analysis.

**BACKGROUND**

<https://en.wikipedia.org/wiki/A_Song_of_Ice_and_Fire>

A Song of Ice and Fire is a series of epic fantasy novels by the American novelist and screenwriter George R. R. Martin. The first volume of the series, A Game of Thrones, was begun in 1991 and first published in 1996. Martin, who initially envisioned the series as a trilogy, has published five out of a planned seven volumes.

A Song of Ice and Fire takes place on the fictional continents Westeros and Essos. The point of view of each chapter in the story is a limited perspective of a range of characters growing from nine, in the first novel, to thirty-one by the fifth. Three main stories interweave a dynastic war among several families for control of Westeros, the rising threat of the supernatural Others in the northernmost reaches of Westeros, and the ambition of Daenerys Targaryen, the deposed king's exiled daughter, to assume the Iron Throne.

<http://www.techinsider.io/number-of-characters-in-game-of-thrones-outweighs-those-in-shows-2016-4>

While there are only 31 point of view characters, there are in fact 2,103 characters mentioned at some point in the series. This was determined by IBM’s Watson via text analysis.

<http://mashable.com/2016/06/06/game-of-thrones-religion/#sFUdhk9VxZqz>

There are many different religions in the world created in the series. There are five main religions and several minor religions and sects. Other religions will be classified as “Other” for the categorical variable.

The main religions are:

The Old Gods,

Faith of the Seven,

R'hllor (Lord of Light),

The Drowned God,

The Many-Faced God

**In addition to identifying your data source, you should create a high level plan that describes how you would load the data for analysis, and describe a hypothetical outcome that could be predicted from comparing degree centrality across categorical groups.**

**LOADING THE DATA FOR ANALYSIS**

Nodes: Any characters mentioned in the book (2,103 characters)

Edges: Any interaction between characters

Categorical Variable: Religion of the character

The data would be loaded with Python networkx as a series of edges to create a graph.

Religion would be added as an attribute of each node

**HYPOTHETICAL OUTCOME**

Measuring power and influence:

Degree Centrality: popularity

Closeness Centrality: ability to move information from one side of the network to another

Betweenness Centrality: communication bottlenecks/ community bridges

Eigenvector Centrality: finding the “grey cardinal” – someone who doesn’t show up in the first three metrics but is actually very powerful

A hypothetical outcome would be to verify via network analysis what the author has already told us about the characters. Will the characters who we know to be popular, information movers, or communication bottlenecks/bridges, or grey cardinals rank at the top of the list for the appropriate centrality measures?

Using the categorical variable, religion, we can further determine if certain religions are more common among characters who rank highly in particular centrality measures.