DATA620 – HW4
Honey Berk

Proposal for Project 1

Dataset

Twitter user data via Twitter REST API
https://dev.twitter.com/rest/public

Goal

Use the REST API to collect a sample of tweets over the course of a manageable period of time (e.g., one day). Choose a popular hashtag and analyze the network of users tweeting with that hashtag to identify central users and calculate their metrics. Note that the hashtag will be one that is used long-term (e.g., #photography), not one that is momentarily popular.

Nodes

Twitter users

Edges

Connections

Categorical Variables

Gender, Location, Verified Status

Procedure

Load data into Jupyter Notebook via the Twitter REST API, Data may require filtering/pruning, depending on number of nodes and edges.

Analysis

The NetworkX package will be used to analyze the dataset, including a search for "celebrities" via degree, closeness and betweenness.

Hypothetical Outcomes

- Do users with similar interests – assumed due to use of established hashtag – tend to follow each other?
- Do long-term hashtags essentially function as online communities, much like Facebook groups?
- Do users with verified status score higher on degree or other metrics?
- Do users tend to follow users of the same gender?
- Do hashtag "communities" tend to be bounded by geographic location?