Effect of Gender on Centrality measures

Project Description

The objective of this project to check does Gender or Education influence how central the person is in her or his Facebook network. For this analysis a real sample data from Facebook will be used. Degree centrality and eigenvector centrality will be used as measure of centrality for this project.

Data

The sample data for this project is obtained from Stanford university's snap (Stanford Network Analysis Project) website. The dataset contains 10 networks and for this project one of the networks (network 3980). If time permitting the analysis will be done for one more network.

http://snap.stanford.edu/data/egonets-Facebook.html

The 3980 network has 59 nodes and 292 edges. There are 42 properties available, but for this project the gender and education type will be used. All the properties of this network are anonymized. So, the analysis will only reveal if the centrality is correlated to a factor without providing information on which factor it is. That is, we could analyze if one gender has higher centrality over other gender, but will not be able to find which gender (i.e. male or female) has higher centrality.

Approach

Host the dataset to github page and load data using urllib. Parse the data using networkX package and create a network. Use networkx inbuilt functions to compute degree centrality and eigenvector centrality for groups by gender and then or groups by education type. Conduct t-test to check if the centrality of nodes in two different gender group are statistically different from each other.

Statistical Measure

Network Analysis – Degree Centrality and Eigenvector Centrality

Hypothesis testing – T-test and Chi-Square test

Technology

Python (IPython notebook) and NetworkX package