

# CSC 555/495 Social Computing Project 1: Social Analytics

Due date: Aug 31 23:59 EST

August 18, 2017

## 1 Introduction to Social Analytics

Social Analytics involves analyzing, interpreting and measuring interactions and associations between people, topics and ideas. This assignment will introduce you to the basics of formal approaches in social network analysis. The assignment has been divided into two parts:

### 1.0.1 Part 1:

**Work interactively with the class and identify interesting characteristics of a sample dataset.**

In this section, you will work interactively to hypothesize potential trends across a given sample dataset. Give reasons and make arguments to support or reject a hypothesis. Use the message board for all your discussion. The instructors will monitor and guide the discussion and finally summarize and vote up the best trends that emerge from the discussion. Based on the summary, you will be asked to further discuss and develop evaluation measures to test your hypothesis. For our discussion, we will use the [Kaggle Influencers in Social Networks dataset](#). You can access the dataset [here](#).

To begin with, explore the dataset to understand the attributes that have been collected. The above URL also links to a paper for your reference. Once you have an intuition for the nature of data collected, try and reason about how you would characterize the graph and evaluate it using some properties of the social network. As an example of some standard evaluation measures, consider the following sociometric network properties for the sample dataset density, centrality, reciprocity and number of components. Think about other measures that you would like to calculate for the graph. Try and come up with some interesting properties that you would like to evaluate on the dataset.

As part of the assignment, you will also be introduced to concepts of argumentation theory in practice. To this end, we have come up with a text based argumentation scheme for use on our message board. For simplicity of maintenance, we request you kindly adhere to the rules given below while posting to the message board.

#### **Argumentation Rules:**

- Every new hypothesis <unityID><serial number>, will be <unityid>'s <serial number>th post.
- All hypotheses must have a justification. To mention your justification simply prepend your statement with the text <Justification>.
- A new post must be created in the Project1 Hypothesis forum for a new hypothesis. All discussions for the hypothesis must take place as replies within that post. The only exception is when <Replace> is being used.

1) To propose an original hypothesis not previously considered, prepend the unique ID with the text <Propose> e.g., <Propose> uniqueID

2) To suggest an alternative hypothesis (uniqueID2) to an existing hypothesis (uniqueID1), prepend uniqueID1 with the text <Replace> and append the new uniqueID (uniqueID2). e.g., <Replace> uniqueID1 uniqueID2

3) To oppose a hypothesis (uniqueID1) in support of another student's hypothesis (uniqueID2), prepend uniqueID1 with <Oppose> and uniqueID2 with <Support> e.g., <Oppose> uniqueID1 <Support> uniqueID2

4) If you would like to question the validity or the claim made by a hypothesis without proposing a new hypothesis, prepend the uniqueID of the hypothesis with <Raise> e.g., <Raise> uniqueID

5) To increase the weight of a question/suggestion raised with <Raise> uniqueID, use a simple <+1> followed by <Raise> <uniqueID> e.g., <+1> <Raise> uniqueID

6) Finally, use a new line to start a statement that includes any of the above tags.

Once the deadline (TBD) for Part 1a has passed, no further discussions will take place. Everyone will have one day to vote for their top hypotheses. Voting will be done through a Survey (to be posted after Part 1a has finished).

The top two hypotheses from this round will be used for Part 2.

### 1.0.2 Part 2:

**Work individually on the programming task to test your hypotheses.**

In this section, you will programatically evaluate some of the sociometric properties decided upon in Part 1.

You may use any programming language of your choice, as long as you write your own code, and make sure that your instructions for compilation and execution are clear and complete.

Use the above dataset to calculate the following properties of the graph:

- 1) The number of nodes
- 2) The number of edges
- 3) Clustering coefficient
- 4) The graph density
- 5) The average degree

### 1.0.3 Part 3: Bonus Credit

Find *the most influential node in the graph*. You may need to look up techniques online to do this.