Project Proposal - CS 605.404 Social Media Analytics Class

Max Robinson max.robinson@jhu.edu
Johns Hopkins University,
Baltimore, MD 21218 USA

1. Research Question

A foundational problem in social network analysis is the ability to identify community structure just from the provided network. A problem that arises from attempting to find these communities is the amount of compute time required to find these communities.

Several algorithms currently exist for finding communities among social networks. The Klaus-Newman-Moore and Louvain algorithms are currently quite popular. These algorithms are used widely and have been shown to perform reasonably quickly when tackling large scale networks.

This proposal suggests investigating the performance of a newer algorithm called Truss Community Detection, or just Truss for short, developed by Jon Cohen. The research would compare the Truss algorithm to the performance of Klaus-Newman-Moor and Louvain, based on relative runtime to complete a task.

2. Target Venue

The planned target venue for this paper would be the IEEE/ACM ASONAM conference. This conference looks at a wide range of different areas and a paper showing a possibly more performance community detection algorithm would be a hit there.

3. Data Collection

The data used for the experiment will be gathered from the Stanford SNAP datasets. These datasets provide large networks that are available to everyone for experiment reproduction. These datasets also constitute a modern standard for large social networks, and as a result provide a good baseline for modern comparisons.

Result data will be collected in the form of the amount of time needed to identify communities of nodes in the network. These run-times measures will be taken several times and averaged out over the runs to provide an averaged run-time with standard deviations to attempt to account for variance in uncontrolled operating system (OS) properties on the machine running the experiments. The data collected will attempt to be similar, if not exatly the same, as the data that was collected as part of Louvain's paper, detailing the Louvain algorithm.