

Chris Ayala
CSE 8331 - Data Mining

Social Network Mining - Abstract

Social networks provide a treasure trove of human information. Even when only looking at relationships (for example, a list of friends or followers), we are able to gather insight into what connects us. What groups connect certain people? Can we provide recommendations based on existing friendships? Can we predict who will become friends in the future? These questions, and more, can be answered by analyzing relationships between people.

If we visualize a social network as a directed graph, then we can apply some well-known graph theory algorithms. Nodes of the graph represent users, and edges represent relationships. We can look at different cuts of the network, and analyze the resulting subsets for quality. By optimizing where we cut the graph, we can accurately find sub-groups of people within the network, and through human or automated analysis determine what it is about these people that connect them. Tools like Gephi and Cytoscape can help visualize the data before and after this clustering takes place, and in this tutorial I will be demonstrating how this is done using Gephi. A large challenge of social network mining is handling the dataset size: millions of users may have hundreds of connections apiece, making sampling and subsetting an integral part of the process.