NETWORK SCIENCE
Final Project Proposal

Guoxi Liu

1 Dataset

The dataset we are going to use in the final project is the Flickr dataset. Flickr is an image hosting and video hosting website, web services suites, and online community. The dataset contains two parts: users and groups. If there is an edge between two users, it means that these two bloggers are friends; if there is an edge between a user and a group, it means the user is a member of the group.

The whole dataset contains 80,513 users and 195 groups. There are 107,741 membership pairs and 5,899,882 friendship pairs. The dataset is downloaded from ASU Social Computing Data Repository.

2 Problems

The connections in the social network are not uniformly distributed, which means that the network can be divided into several communities and there are dense connections in the same community but sparse connections between different communities. We are going to do community detection with Flickr dataset in the final project.

Apart from clustering, new friend and group suggestion are also important in social networks. We are going to recommend some new groups for a specific user according to his/her friendship and some possible friends to a particular blogger based on the groups he/she has joined.

3 Methods

Modularity is one measure of the structure of networks. It was designed to measure the strength of division of a network into modules. So modularity can be used in optimization methods for detecting community structure in networks. Apart from modularity maximization, there are many other algorithms for community detection, e.g., Label Propagation Algorithm, infomap. We are going to implement these different algorithms, test them on Flickr dataset and compare their results.

As for friend and group suggestion, we are going to compute random-walk based similarities between vertices and use collaborative filtering method, then compare their results.