Project Proposal

Harald Ng
Chuan Su

September 28, 2019

1 Problem Description

Today social network sites and apps, such as Twitter, LinkedIn, have been becoming part of our dairly lives. Recommendation systems of friend, connection and group are becoming the fundamental part of social network systems. We believe a great recommendation system is based on user's networking, friendship graph analysis.

In this project we are going to implement a social network analysis tool providing two major functionalities:

- **Friend recommendation** Find pairs of user who are not connected but have at least X number connections in common.

- **Community detection** Find cliques where the number of outgoing edges is at most one more than the number of nodes in the cliques. For example, if a clique has 3 nodes and there are no more than 4 edges from nodes in this clique to nodes outside, then this clique is defined as a community.

2 Tools

The data will be processed **Spark GraphX** and the results will be stored to **HDFS**. The **Scala** programming language will be used for development. If there will be sufficient time, visualizing the results will be implemented as well.

3 Data

We have not decided on the dataset to be used. But we have narrowed down our choices to a graph dataset from `http://konect.uni-koblenz.de/networks`. The nodes of the graph represents users and the edges represents friendships in the social network.