

A Threshold-Based Girvan-Newman Implementation

Samanvitha Basole Ketki Kulkarni
BS, Computer Science MS, Computer Science

San Jose State University

November 18, 2016

Contents

1	Introduction and Motivation	2
2	Problem Description	2
3	Tools and Technologies	2
4	Strategy	2
5	Datasets to be used	3
6	Evaluation Method	3
7	References	3

1 Introduction and Motivation

A community in large networks is a group of nodes in which the proportion of edges is greater than the betweenness of edges. The study of community detection is a standard problem in analyzing large and complex social networks. Many researchers have concentrated in solving this problem using graph partitioning, hierarchical clustering, partitional clustering and spectral clustering.

2 Problem Description

In this project, we plan to analyze large social networks using different hierarchical clustering algorithms. Every algorithm uses different approach to detect communities, and thus underlying structure of those communities differs. We plan to apply these algorithms and compare the resulting communities based on modularity measure, centralities, local bridges, and community structure.

Modularity is ratio of intra-community links to the inter-community links. For accurate communities, this measure is high. Some of the algorithms we are planning to use are Girvan-Newman algorithm based on edge betweenness, Louvain Method, Label Propagation, Random Walks, and Fast Greedy approach to detect communities. We will also compare these algorithms based on their accuracy and efficiency.

3 Tools and Technologies

1. Python
2. iGraph package
3. PyCharm/Canopy

4 Strategy

To achieve our goal of comparing different algorithms, we start by implementing a method in python to read datasets and generate a graph object. Then,

we will apply different algorithms, already implemented in iGraph package of Python to each of the datasets mentioned below. Finally, we will visualize the result using `plot()` method of iGraph and evaluate the result.

5 Datasets to be used

1. Youtube network
2. Barabasi-Albert model
3. California road network

6 Evaluation Method

To gauge the effectiveness of each algorithm, we will evaluate the results produced by these algorithms on the basis of community structure, Modularity, and Centralities.

7 References

- <http://snap.stanford.edu/data/index.html>
- <http://netsg.cs.sfu.ca/youtubedata/>
- <https://snap.stanford.edu/data/com-Youtube.html>
- <http://snap.stanford.edu/data/roadNet-CA.html>
- <http://dm.kaist.ac.kr/jaegil/papers/bigcomp14.pdf>
- <http://www.pnas.org/content/99/12/7821.full.pdf>