COMPUTING AND INFORMATION SYSTEMS DOCTORAL COLLOQUIUM

JULY 8th

2013 PROGRAM AND PROCEEDINGS

gsa graduate student association

NICTA

Google

vlsci LIFE SCIENCES COMPUTATION

THE UNIVERSITY OF MELBOURNE

How tightly connected are communities?

[Extended abstract]

Minh Van Nguyen
Dept. Computing and Information Systems
The University of Melbourne
Melbourne, Australia
minguyen@student.unimelb.edu.au

Michael Kirley
Dept. Computing and Information Systems
The University of Melbourne
Melbourne, Australia
mkirley@unimelb.edu.au

Rodolfo García-Flores
CSIRO Mathematics, Informatics and Statistics
Clayton, Australia
Rodolfo.Garcia-Flores@csiro.au

Categories and Subject Descriptors: H.2.8 Database Management: Database applications – Data mining

General Terms: Measurement; Experimentation

Keywords: Community structure; Graph partitioning

1. INTRODUCTION

Networks are ubiquitous in modern society. From the Internet to social networks, a network can be divided into clusters where the nodes in each cluster are tightly connected among themselves, with sparse connection between a cluster and the rest of the network. Clusters that satisfy this property are known as communities [1]. In terms of the Internet, communities represent clusters of Autonomous Systems that, once extracted, allow us to identify a minimum set of links whose removal would fragment the Internet. While there are efficient algorithms to extract communities, a fundamental issue remains: How well connected are the nodes in a community?

2. METHODS

We propose a technique to measure how tightly connected is a community. Denote by d_{ij} the distance, or minimum number of links, separating two nodes i and j in a community C. A community whose nodes are tightly connected among themselves must be such that the sum of the distance between all pairs of nodes is as small as possible. That is, the sum $W(C) = \sum_{i<j} d_{ij}$ should be minimal. If T is a minimum spanning tree of C, the ratio $W(C)/W(T)$ quantifies the probability that C has a topology similar to T. We define the compactness ratio $W^* = 1 - \frac{W(C)}{W(T)}$ to measure how tightly connected is the community. The closer that W^* is to 1, the more tightly connected is the community. A related measure is the clustering coefficient [2], which in terms of social networks quantifies the probability that two friends have a friend in common. The compactness ratio measures the global connectedness across a community, in contrast to the clustering coefficient which measures the average local cliquishness of a node.

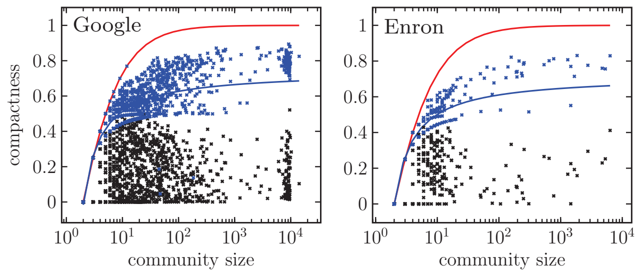

Figure 1: (Color online) The compactness of communities in two networks. In each plot, black dots represent the compactness W^* of communities and a blue dot represents the ideal compactness $W^*_{K_n}$ when a community has all possible edges. Each blue curve models the ideal compactness and the red curve provides an upper bound on both W^* and $W^*_{K_n}$.

3. RESULTS

We have computed the compactness of communities in various social, information, technological, and biological networks. See Figure 1 for results for two real-world networks. For large communities on n nodes, we find that the compactness changes at a rate that is at most proportional to $\frac{1}{n(\log n)^2}$. The largest communities in some networks have low clustering coefficient (< 0.02), yet high compactness ($W^* > 0.3$). We have verified that the high compactness of a community can be attributed to edges that act as shortcuts in the community. The shortcuts connect nodes having high numbers of links to nodes with low numbers of links, thereby decreasing the minimum number of links that separate distant nodes. The presence of shortcut edges means that the overall structure of a community can be highly compact, despite a low clustering coefficient.

4. REFERENCES

[1] S. Fortunato. Community detection in graphs. *Phys. Rep.*, 486:75–174, 2010.

[2] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393:440–442, 1998.