MATLAB FS12 – Research Plan

Group Name: Serbian Brate

Group participants names: Campara Milan, Roveri Riccardo

Project Title: Investigation of community structure using online social network data (might be changed)

General Introduction

In general, the study of friendship networks and social networks can be very useful for many reasons, such as the prediction of the transmission of information or diseases.

The analysis of the organization of social networks has been recently gaining a lot of attention with the growth of social network websites, such as Facebook or MySpace, which provide researchers with huge amounts of relational data and users information data.

Thus, the investigation of correspondences between social network communities and real word organizational structures can be considered a relevant and up-to-date topic.

The Model

Our goal is to compare quantitatively a set of algorithmically identified communities to partitions determined by demographic labels integrated in the data (such as attended university, class or year).

We will model the communities using groups of nodes with more internal connections than external ones. Communities are going to be identified by spectral optimization, while the pair counting method will be used for the quantitative comparison with the existing demographical groups.

Fundamental Questions

At the end of the project we want to be able to answer to the following questions:

- How useful is the study of communities structure to find and measure a correlation between online and offline social networks?
- Which pair-counting indices provide a better interpretation about the similarity score?
  - What similarity score values indicate a good correlation?
- Taking a universities context as an example, which are the factors that influence the structure of communities in the university at most?
  - Are these factors the same in every university?

Expected Results

It is expected to get an understanding about the importance of the communities structure's analysis in order to compare online social networks with real world clusters.

We also expect to notice some differences in the factors which influence university communities, depending on the type of the university (for example, in big universities the hometown may assume more weight than in a small university).

References

Comparing Community Structure to Characteristics in Online Collegiate Social Networks,
Amanda L. Traud, Eric D. Kelsic, Peter J. Mucha, Mason A. Porter

Tastes, ties, and time: A new social network dataset using Facebook.com,
Kevin Lewis, Jason Kaufman, Marco Gonzalez, Andreas Wimmer, Nicholas Christakis

A dynamic model of friendly association networks,
Joseph M. Whitmeyer , Cynthia N. Yeingst

Research Methods

As previously explained, we will use the pair counting method to compare quantitatively offline communities with online networks.

Dataset

We are likely going to use a Facebook dataset accessible at the webpage:
http://thedata.org/.