Identifying Sociological Trends in Facebook Networks

Amanda L. Traud ▪ Department of Mathematics ▪ Carolina Population Center Trainee ▪ The University of North Carolina at Chapel Hill

Introduction

- Networks are a set of objects, or nodes, that are connected in some way. They can consist of objects of any kind from chemicals to people. The networks studied consisted of users on Facebook and their friends.
- Facebook is a social networking website that plays a prominent role in modern society. Data on Facebook users can be gathered in a way that resembles real-world social networks. Using this data, a network for each college or university could be set up for each college or university. Five of these networks were examined closely.
- From those five networks, algorithms were used to detect communities, or tightly connected sets of users. These communities were correlated with characteristics given by those users.

Methods

- Community Detection
  - Why is important?
  - Breaking networks into communities can help us understand the structure of the network.
  - Community Detection can be used on many types of networks, including social networks, disease networks, and neural networks, just to name a few.

- Newman’s Leading Eigenvector Method
  - In this method, one tries to maximize a quantity called modularity of the network. The formulas for this method are described above and the size of our networks.

- Similarity Coefficients
  - To compare communities to given characteristics, one calculates similarity coefficients.
  - These coefficients are based on the idea that the characteristics given are a new set of communities.
  - For the five similarity coefficients we calculated, we had to pair every node with every other node and count pairs. All five of the similarity coefficients calculated are based on different combinations of the quantities a, b, c, and d.

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td># number of pairs of nodes in the same community in the first set, but not in the same community in the second set.</td>
<td># number of pairs of nodes in the same community in the second set, but not in the same community in the first set.</td>
<td># number of pairs of nodes in the same community in both sets.</td>
<td># number of pairs of nodes in different communities in both sets.</td>
</tr>
</tbody>
</table>

Graph of Z-scores versus Average Z-score for each category. Note: the slope is approximately 1.

Z-Score Calculation

- After calculating similarity coefficients, a Z-score was calculated. The Z-score was approximately the same for each distribution, specifically the Z-scores.
- Permutation tests were performed on the similarity coefficients to create a distribution of coefficients. Many other tests were performed, including calculating the frequency and ranking interquartile ranges of the distributions to make sure the distribution was Gaussian. (See below for a distribution example.)

Results

- While a number of different similarity coefficients appear in the literature, the statistics of those as obtained through permutation tests were approximately the same, specifically the Z-scores.
- Most schools break up into communities due to graduation year; however, some, like Caltech, break up into communities based on different life stages. It is not clear why the networks at similarity coefficients used, the Z-score was approximately the same for each community and characteristics. Comparing to the distribution.
- When the similarities in a similarity coefficient were found, the Z-score was approximately the same for each category, as shown by the plot below.

Conclusion/Contact info

- These analyses were performed both ignoring and respecting the fact that there are missing data. In almost no case did the differences change the qualitative conclusion.

- Contact: Amanda L. Traud (altraud@email.unc.edu)

Acknowledgements

- My Collaborators: Eric Kelsic, Peter J. Mucha, Mason A. Porter
- UNC-Chapel Hill KAGE Program, Director: Dr. Valerie Ashby
- National Science Foundation HRD-0450099, DMS-0645369
- UNC Graduate Education & the Profession