Using word graphs to find users in a given location

Utkarsh Bali, Rochester Institute of Technology, utkarshbali.51@gmail.com
Advisor: Dr. Christopher M. Homan, Rochester Institute of Technology, cmh@cs.rit.edu

Hypothesis
Words whose subgraphs have strong community properties are likely to be a part of the jargon of a community.

Experiment

- Tweets from ~86,000 Rochester users.
- Followers graph
- For each word: one subgraph for all who used it.
  - 1000 Metropolis-Hastings samples from followers graph of similar size & edge count of word graph.
  - Compare samples with the word graph:
    - Degree
    - Triangles
    - Clustering coefficient
    - Largest connected component

Results

![Figure 2: Graph and statistics for users who tweeted the word #rochesterny](image)

<table>
<thead>
<tr>
<th>Regression Coefficients</th>
<th>Average Degree</th>
<th>-0.0018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustering Coefficient</td>
<td>0.00035</td>
<td></td>
</tr>
<tr>
<td>Avg. Triangles</td>
<td>0.0076</td>
<td></td>
</tr>
<tr>
<td>Largest Conn Comp</td>
<td>0.00058</td>
<td></td>
</tr>
<tr>
<td>Overlap</td>
<td>-0.00012</td>
<td></td>
</tr>
<tr>
<td>Variance Score</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>0.79</td>
<td></td>
</tr>
</tbody>
</table>

Conclusion
Model selects location-based words with 37% accuracy
Analyze more terms & community properties for a greater number of terms.
Use this model to find more users in a particular city or a region.