CS 8803 Social Computing: Networks (Time)

Munmun De Choudhury
munmund@gatech.edu
Week 8 | October 8, 2014
Assignment I: Some numbers

- Max: 20.4
- Mean: 18.28
- Median: 18.4

- For those of who hadn’t completed Assignment I, if you give me/Joe a legitimate reason, you can take a make-up Assignment IV, similar in vein to Assignment I.
  - Will be given on Nov 19, the week before Thanksgiving.
  - Will be due on Dec 3.
Term Project Presentations: Midterm

- 5% of your grade
- Nine groups in all
- Each group will get 8 minutes; ~6 minutes of presentation, ~2 minutes of Q&A
- No more than ~8 slides
- Everyone doesn’t need to present
- Main structure of the presentation:
  - What is the project?
  - Why is it important?
  - What have others (e.g., prior work) done on this or similar topic?
  - What are the specific things you plan to accomplish by the semester end?
  - What have you done so far?
  - What are the remaining steps from midterm to final, including action items per group member?
# Term Project Presentations: Midterm

<table>
<thead>
<tr>
<th>Group name/topic</th>
<th>Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>les redditorsians</td>
<td>Ashwini Khare, Revant Kumar, Suren Nihalani, Prajwal Prasad</td>
</tr>
<tr>
<td>Help Yelp!</td>
<td>Thomas Loalbo, Florian Foerster, Perron Jones, Christina Masden, Jitesh Jagadish</td>
</tr>
<tr>
<td>Triple C + P/Pro-ED and Instagram</td>
<td>Stevie Chancellor, Trustin Clear, James Crouch, Jessica Pater</td>
</tr>
<tr>
<td>Loneliness, emotion, and imagery</td>
<td>Unaiza Ahsan, Jose Delgado, John Dugan, Omer Semerci</td>
</tr>
<tr>
<td>Collaboration and GitHub</td>
<td>Sneha Iyengar, Netra Kenkarey, Srinivas Eswar, Shankar Vishwanath</td>
</tr>
<tr>
<td>Two Sides to a Story/Topical Polarization and Social Media</td>
<td>Alex Godwin, Anand Sainath, Sanjay Obla Jayakumar, Vinodh Krishnan</td>
</tr>
<tr>
<td>User Interest Modeling on Social Media</td>
<td>Alvin Khong, Saajan Shridhar, Mrinal Kumar</td>
</tr>
<tr>
<td>Twitter - Entertainment Data Analysis</td>
<td>Harikumar Venkatesan, Karthik Krishna Subramanian, Divya Vijayaraghavan</td>
</tr>
<tr>
<td>Social Media (Twitter) and amusement parks</td>
<td>Arjun Srinivasan, Suraksha Suresh Pai</td>
</tr>
</tbody>
</table>
Midterm Project Report

• 20% of grade
• Due: October 20
• Structure of Report:
  • If building a tool: design process, mockup, and an early prototype if possible
  • If analyzing data: data collection method/key properties of the data, plan for analysis
  • Report length: 4-5 pages, single column, single spaced format submitted through T-Square
• Clearly articulate in an extra page individual contribution
• Typically you will not need to submit the code, unless some exception arises
Why is studying network evolution important?

- Anomaly detection and computer network management
- Graph extrapolation and prediction
- Graph sampling
- Moderation and group management
Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations
Summary

• First quantitative study of evolution of social graphs.

• Research questions:
  • How do real graphs evolve over time?
  • What are “normal” growth patterns in social, technological, and information networks?

• Most earlier work was on studying structure of static graphs
  • Preferential attachment model (“rich gets richer”) gives strong bounds to diameters of graphs, and that they grow slowly as number of nodes grow

• Findings:
  • As graphs get more edges, they tend to become more dense
  • Average distance between nodes shrinks over time, instead of increasing as a logarithmic function of the number of nodes

• Contribution: a model called the “forest fire model” that mimicked this kind of graph evolution
Forest Fire Model

- Properties of the model:
  - some type of “rich get richer” attachment process, to lead to heavy-tailed in-degrees
  - some flavor of the “copying” model, to lead to communities
  - some flavor of Community Guided Attachment, to produce a version of the Densification Power Law
  - and an ingredient that leads to shrinking diameters

- A new node “burns” links outwards, with a certain probability followed in-links and out-links of nodes at the end of the newly burnt links, and continues to expand recursively
Group formation in large social networks: membership, growth, and evolution
Summary

• Early study of the evolution of communities
• Research questions:
  • what are the structural features that influence whether individuals will join communities?
  • Which communities will grow rapidly?
  • How do the overlaps among pairs of communities change over time?
• Focus on two datasets: LiveJournal and DBLP
• Study
  • how the evolution of communities relates to properties such as the structure of the underlying social networks
  • How to measure movement of individuals between communities, and how such movements are closely aligned with changes in the topics of interest within the communities
• Findings:
  • the tendency of an individual to join a community is influenced not just by the number of friends he or she has within the community, but also crucially by how those friends are connected to one another
An underlying premise in diffusion studies is that an individual's decision to adopt a new behavior is influenced by the number of friends already in the community. The probability of adopting a new behavior increases with the number of friends already engaging in the behavior — in this case, the number of users that have "joined" a conference from one year to the next.

The plots for LJ and DBLP exhibit qualitatively similar shapes to the plots in Figure 1, although the curves in their case become noisier at smaller values of $k$. While these curves represent a good start towards membership in the context of diffusion studies, a very large sample may be required to begin seeing the shape of the curve clearly.

We find it striking that the curves for LJ and DBLP have such similar shapes (including the deviations for $k = 1$). This suggests that for computing the probability of joining a community, it is sufficient to consider only members of the community and the other in the fringe. We now consider a range of other features related to the community, $C$. (Edges between only members of the community are $E_C \subseteq E$.)

### Table 1: Features.

<table>
<thead>
<tr>
<th>Feature Set</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Features related to the community, $C$.</strong> (Edges between only members of the community are $E_C \subseteq E$.)</td>
<td>Number of members ($</td>
</tr>
<tr>
<td><strong>Features related to an individual $u$ and her set $S$ of friends in community $C$.</strong></td>
<td>Number of friends in community ($</td>
</tr>
</tbody>
</table>
The Life and Death of Online Groups: Predicting Group Growth and Longevity
Summary

- What factors predict whether a community will grow and survive in the long term?
- Main idea: investigate the role that two types of growth (growth through diffusion and growth by other means) play during a group’s formative stages
- Results: 79% accuracy in predicting growth of groups in the short-term, while 78% for those in a longer term spanning two years
- Findings:
  - Group clustering does increase the diffusion growth of a group, but that groups which grow primarily through diffusion reach smaller sizes eventually.
  - Past growth rates predict short term growth; incorporating network structures e.g. size of GCC improves prediction of longer term group growth
In each task, we generate models for the four growth metrics. In evaluating our models, we consider two different logistic regression. The first is a simple logistic regression aimed at predicting whether a group will grow or not. The second, more complex model, includes features that indicate a natural threshold for group size around 150 members.

For small and large groups, Table 3 shows the group size at prediction time to the group size at the time of the growth. The variance for small and large groups is shown in Table 3 below.

### Table 3: Number of groups, short-term (2-month) and long-term (2 year) growth rates.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Small Groups (10–100 Members)</th>
<th>Large Groups (150–1000 Members)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly Growth Rate</td>
<td>G (0.7) 0.9</td>
<td>G (0.7) 0.9</td>
</tr>
<tr>
<td>Fringe Growth Rate</td>
<td>C (0.5) 0.9</td>
<td>C (0.5) 0.9</td>
</tr>
<tr>
<td>Group Transitivity</td>
<td>S (0.7) 0.9</td>
<td>S (0.7) 0.9</td>
</tr>
<tr>
<td>Transitivity Ratio</td>
<td>ALL (0.5) 0.9</td>
<td>ALL (0.5) 0.9</td>
</tr>
<tr>
<td>Group Density</td>
<td>60 Days (0.5) 0.9</td>
<td>60 Days (0.5) 0.9</td>
</tr>
<tr>
<td>Density Ratio</td>
<td>180 Days (0.5) 0.9</td>
<td>180 Days (0.5) 0.9</td>
</tr>
</tbody>
</table>

### Table 2: Features used in all growth and longevity models.

- **Summary**

<table>
<thead>
<tr>
<th>Category</th>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth</td>
<td>Monthly Growth Rate</td>
<td>Fraction of users who joined in the prior month</td>
</tr>
<tr>
<td></td>
<td>Fringe Growth Rate</td>
<td>Fraction of users who joined in the prior month who joined from the fringe</td>
</tr>
<tr>
<td>Connectivity</td>
<td>Group Transitivity</td>
<td>Transitivity of network formed by group members</td>
</tr>
<tr>
<td></td>
<td>Transitivity Ratio</td>
<td>Ratio of group transitivity to transitivity of entire community</td>
</tr>
<tr>
<td></td>
<td>Group Density</td>
<td>Density of network formed by group members</td>
</tr>
<tr>
<td></td>
<td>Density Ratio</td>
<td>Ratio of group density to density of entire community</td>
</tr>
<tr>
<td>Structural</td>
<td>Clique Ratio</td>
<td>Largest fraction of group members whose edges form a clique</td>
</tr>
<tr>
<td></td>
<td>Disconnected Ratio</td>
<td>Fraction of group members who are not a part of the group’s largest connected component</td>
</tr>
</tbody>
</table>

As expected, groups which were established in Ning communities that are eventually small or large were able to attract many members through wider appeal. That is, individuals join fringe communities like sports teams and close friends to exchange private messages. Regardless of where the online fan club group is located, its members’ external friendship ties play a role in the group’s growth. The growth of fringe communities can be predicted by the transitivity of network formed by group members. Fringe growth was found to be higher for small groups, hinting that these structural features may be relevant for small groups. For groups that reach eventually smaller sizes, it was found that individuals who were in the fringe of a community were more likely to join. Only 16% of those in Ning communities that are eventually large had come from the fringe if 90% of the group joined from the fringe. For example, the same pattern was observed when examining growth rates over periods of 30 or 60 days. Again, this suggests that fringe growth of groups is due to diffusion processes in subsequent growth.

In our second modeling task, we examined the role of diffusion processes in subsequent growth. We observed the same pattern when examining growth rates over periods of 30 or 60 days. Again, this suggests that fringe growth is due to diffusion processes in subsequent growth. The diffusion process is indicated by its members’ external friendship ties, which positively predict growth. However, transitivity and growth have similar effects on the probability of edges among members of the group. A decrease in transitivity indicates a decrease in the likelihood that a group will grow rapidly, matching earlier observations about diffusion processes. In Table 5, we show coefficients for the combined models (as well as coefficients for the single feature models). The results reveal that adding features pertaining to network connectivity and the role of diffusion processes in subsequent growth. The table shows that adding features pertaining to network connectivity and the role of diffusion processes in subsequent growth. The table shows that adding features pertaining to network connectivity and the role of diffusion processes in subsequent growth.

In summary, our findings provide insight into the relationship between the amount of growth a group experiences and the factors that influence it. We have shown that group dynamics, such as the transitivity of network formed by group members, have a significant impact on group growth. Additionally, we have observed that fringe growth is due to diffusion processes in subsequent growth. The diffusion process is indicated by its members’ external friendship ties, which positively predict growth. However, transitivity and growth have similar effects on the probability of edges among members of the group. A decrease in transitivity indicates a decrease in the likelihood that a group will grow rapidly, matching earlier observations about diffusion processes. In conclusion, our models still achieve relatively high accuracy in predicting whether a group will grow, hinting that these structural features may be relevant for small groups.
Summary

- Groups with one or more cores of tightly connected members and a periphery of members loosely connected or entirely disconnected from this core should experience increased and prolonged growth (low transitivity, small connected components, and large cliques).
- The densely connected core allows for the swift transmission of resources and the loose periphery allows for the presence of structural holes, or ties which bridge clusters, allowing members on the periphery to bring new information or members to the core.
Relate the densification law of social graphs given in Leskovec et al. to two theories we have studied: (1) structural balance and triangle closure, (2) 4-6 degrees of separation
Backstrom et al found that topical changes (or movement bursts) were associated with movement of individuals between communities. What factors could be similar or distinct in the case of social network communities?
The design of today’s social media sites may allow for lesser community movement, simply because one could lurk on one community while being active on the other. How do you envision these communities to evolve over time?
Many communities on social media form due to external (and uncontrolled) events, e.g., the #ebola outbreak. Can the models of community evolution examined in Backstrom et al or Kairam et al explain these instantaneous group formations?
How would the findings of Backstrom et al and Kairam et al generalize to social media sites which are more content focused than friendship focused?
How do we characterize group splitting or group merging over time?
Next class

- NO CLASS on Monday – Fall break.
- Wednesday 10/15
- Term Project Presentations I (Midterm)
- No assigned readings