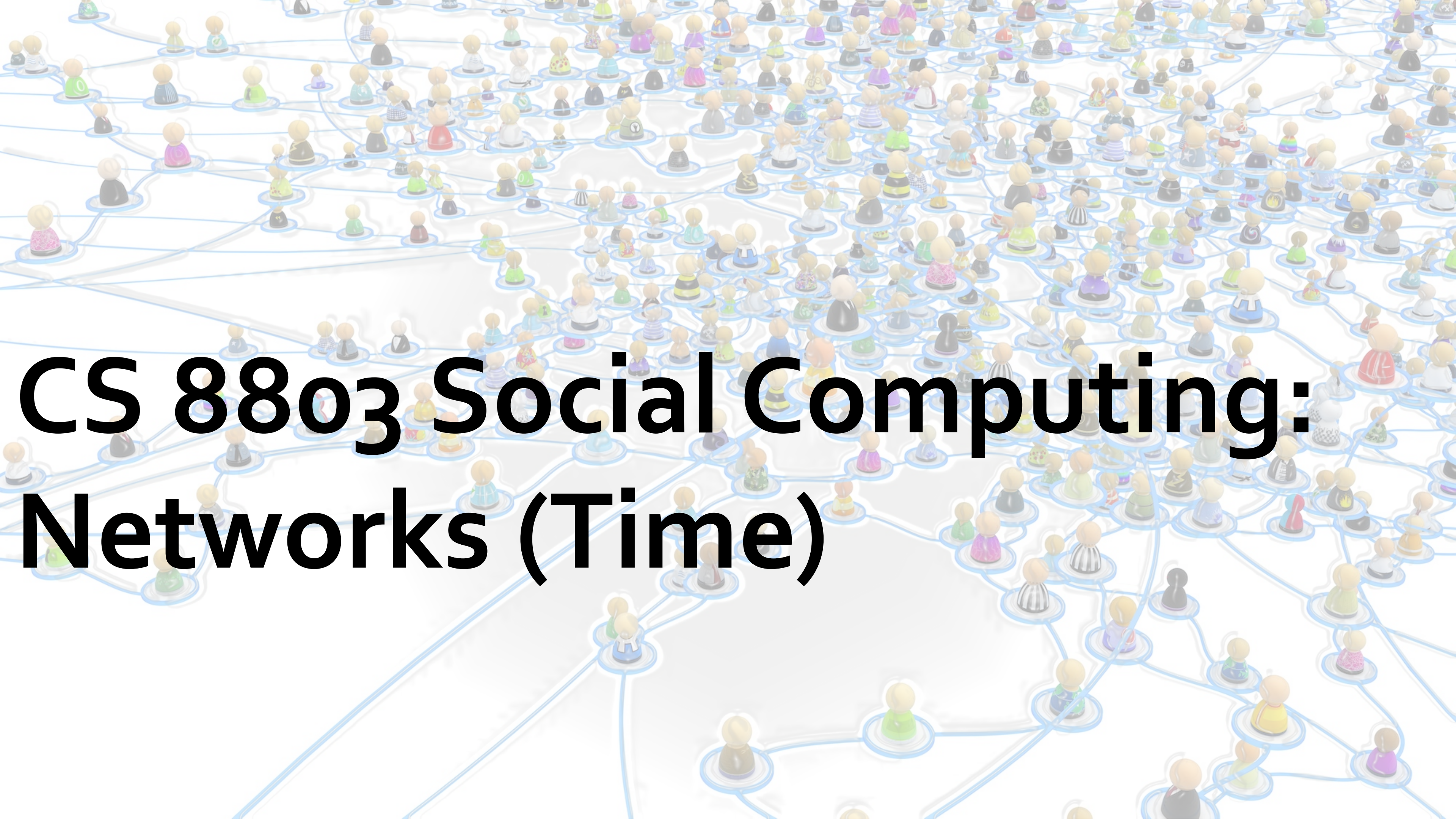


CS 8803 Social Computing: Networks (Time)

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

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

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 com- munity 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

Summary

Table 1: Features.

Feature Set	Feature
Features related to the community, C. (Edges between only members of the community are $E_C \subseteq E$.)	Number of members ($\|C\|$). Number of individuals with a friend in C (the *fringe* of C) . Number of edges with one end in the community and the other in the fringe. Number of edges with both ends in the community, $\|E_C\|$. The number of open triads: $\|\{(u,v,w)\|(u,v) \in E_C \wedge (v,w) \in E_C \wedge (u,w) \notin E_C \wedge u \neq w\}\|$. The number of closed triads: $\|\{(u,v,w)\|(u,v) \in E_C \wedge (v,w) \in E_C \wedge (u,w) \in E_C\}\|$. The ratio of closed to open triads. The fraction of individuals in the fringe with at least k friends in the community for $2 \leq k \leq 19$. The number of posts and responses made by members of the community. The number of members of the community with at least one post or response. The number of responses per post.
Features related to an individual u and her set S of friends in community C.	Number of friends in community ($\|S\|$). Number of adjacent pairs in S ($\|\{(u,v)\|u,v \in S \wedge (u,v) \in E_C\}\|$). Number of pairs in S connected via a path in E_C. Average distance between friends connected via a path in E_C. Number of community members reachable from S using edges in E_C. Average distance from S to reachable community members using edges in E_C. The number of posts and response made by individuals in S. The number of individuals in S with at least 1 post or response.

The Life and Death of Online Groups: Predicting Group Growth and Longevity

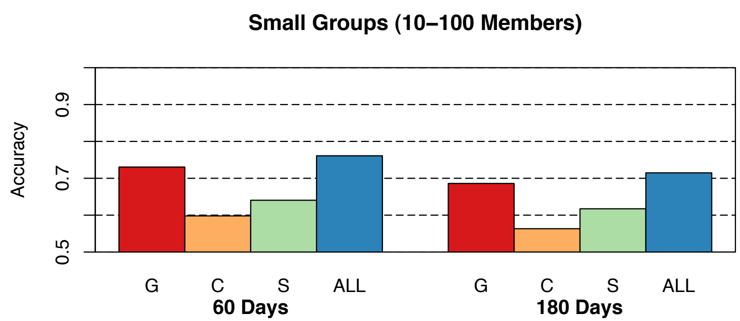
Summary

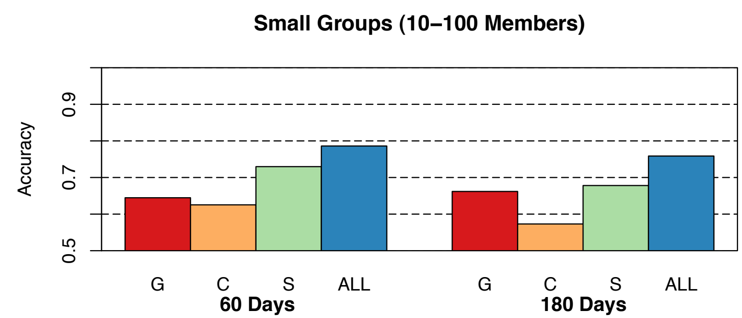
- What factors predict whether a community will grow and survive in the long term?
- Main idea: investigate the role that two types of growth (growth through diffusion and growth by other means) play during a group's formative stages
- Results: 79% accuracy in predicting growth of groups in the short-term, 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

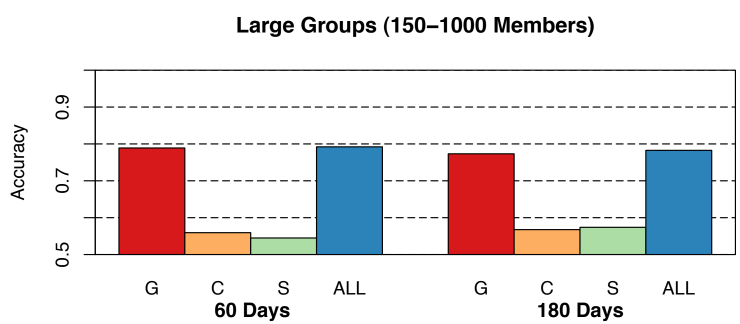
Summary

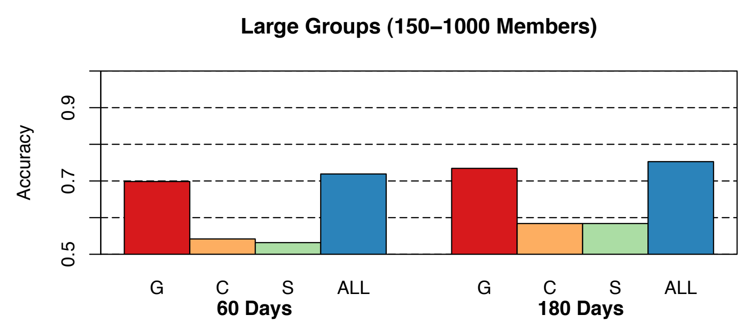
Category	Feature	Description
Growth	Monthly Growth Rate	Fraction of users who joined in the prior month
	Fringe Growth Rate	Fraction of users who joined in the prior month who joined from the fringe
Connectivity	Group Transitivity	Transitivity of network formed by group members
	Transitivity Ratio	Ratio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

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









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.

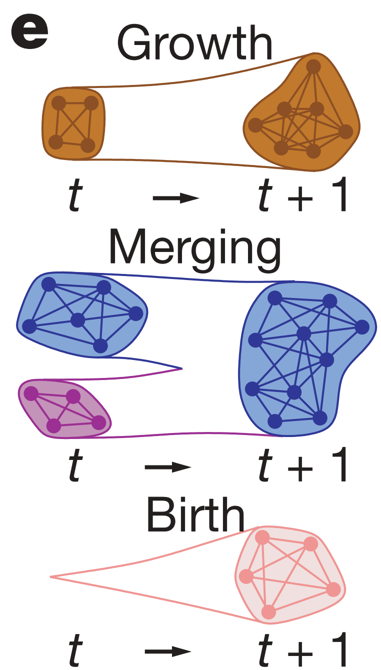
Relate the densification law of social graphs given in Leskovec et al. to two theories we have studied: (1) structural balance and triangle closure, (2) 4-6 degrees of separation

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?



Palla, Barabasi, & Vicsek, Nature Letters 2007

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