Discussion Topics and Ego Networks on Twitter

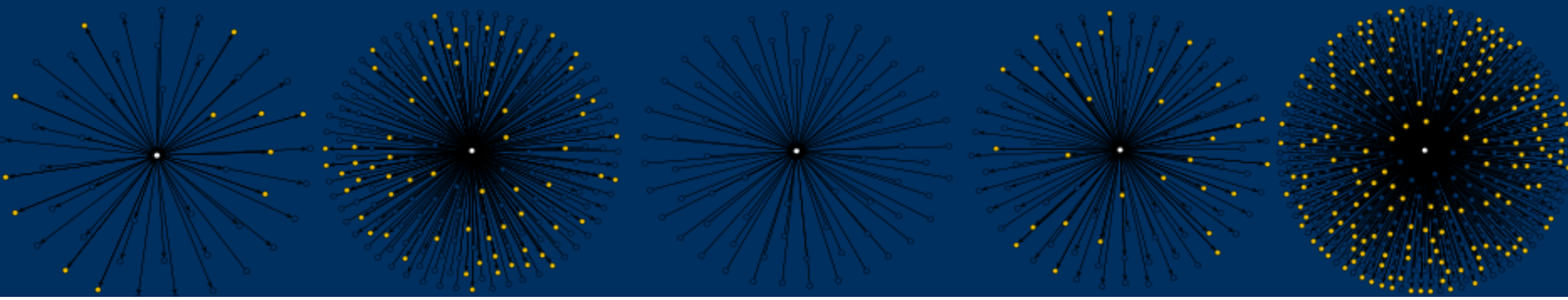
Emma S. Spiro

University of California, Irvine
Department of Sociology

Presented at MURI AHM May 25, 2010

This material is based on research supported by the Office of Naval Research

under award N00014-08-1-1015.



UCIrvine

NCASD
NETWORKS, COMPUTATION, and SOCIAL DYNAMICS

Calit2

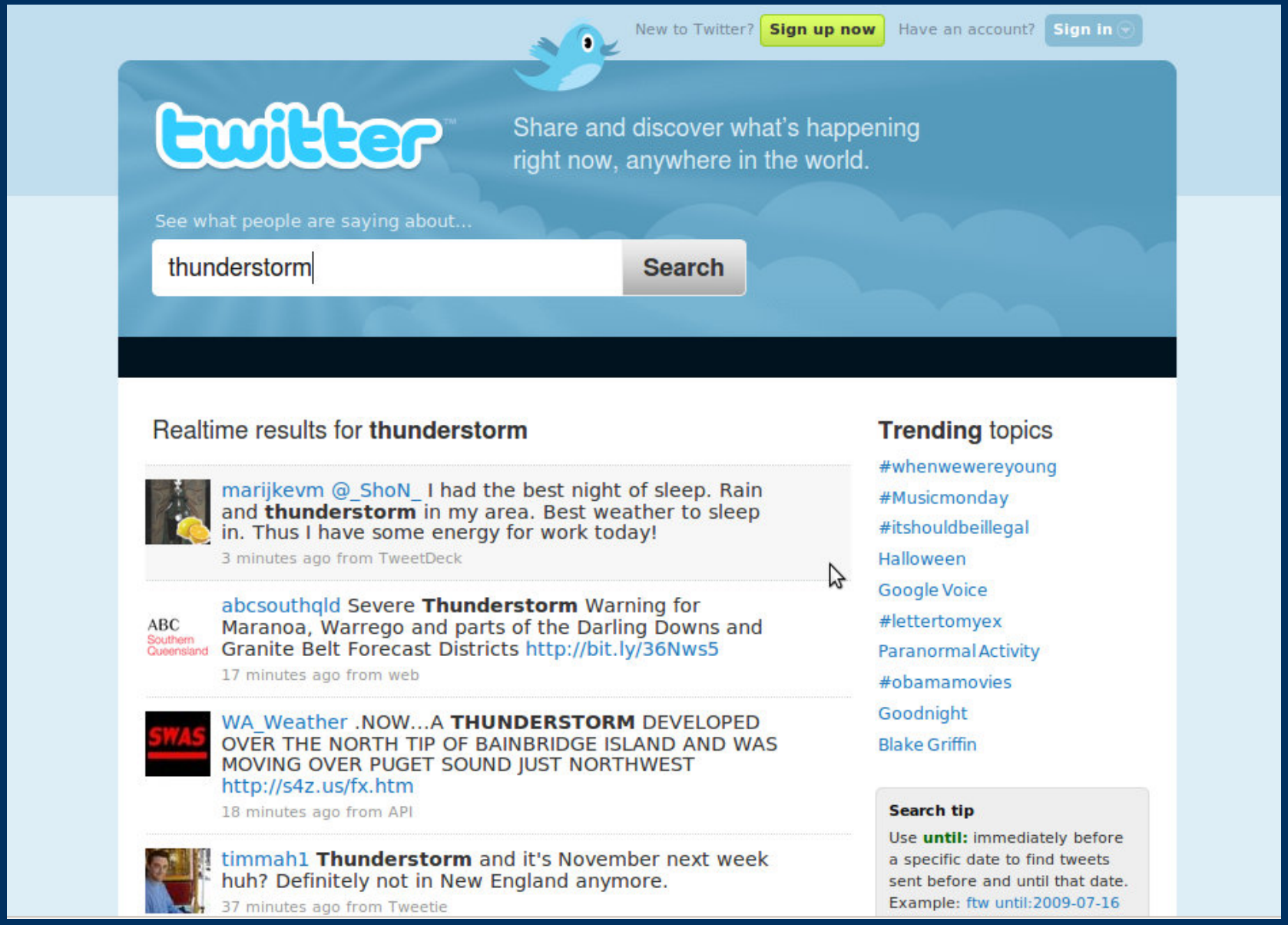
Scalable Methods for the
Analysis of Network-Based Data

Talk Outline

- Project Motivation
- Introduction to Twitter
- Data Collection
- Communication Dynamics
- Structural Characteristics of Personal Networks

Project Motivation

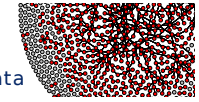
- Informal communication channels are often the primary means by which time-sensitive hazard information first reaches members of the public.
- Social media technologies, e.g. micro-blogging, provide a means for gathering, sorting and disseminating information — a venue for collective problem solving.
- Relatively little is known about the dynamics of informal information communication in emergencies or hazards.


New to Twitter? Sign up now
Have an account? Sign in
twitter
Share and discover what's happening right now, anywhere in the world.
See what people are saying about...
thunderstorm
Search
Realtime results for thunderstorm
marijkevm @_ShoN_ I had the best night of sleep. Rain and thunderstorm in my area. Best weather to sleep in. Thus I have some energy for work today!
3 minutes ago from TweetDeck
abcsouthqld Severe Thunderstorm Warning for Maranoa, Warrego and parts of the Darling Downs and Granite Belt Forecast Districts http://bit.ly/36Nws5
17 minutes ago from web
WA_Weather .NOW...A THUNDERSTORM DEVELOPED OVER THE NORTH TIP OF BAINBRIDGE ISLAND AND WAS MOVING OVER PUGET SOUND JUST NORTHWEST http://s4z.us/fx.htm
18 minutes ago from API
timmah1 Thunderstorm and it's November next week huh? Definitely not in New England anymore.
37 minutes ago from Tweetie
Trending topics
#whenwewereyoung
#Musicmonday
#itshouldbeillegal
Halloween
Google Voice
#lettertomyex
Paranormal Activity
#obamamovies
Goodnight
Blake Griffin
Search tip
Use until: immediately before a specific date to find tweets sent before and until that date.
Example: ftw until:2009-07-16

twitter
Have an account? Sign in
Get short, timely messages from Tori Gabor.
Twitter is a rich source of instantly updated information. It's easy to stay updated on an incredibly wide variety of topics. Join today and follow @gabortori.
Give it a try ›
Get updates via SMS by texting follow gabortori to 40404 in the United States
Codes for other countries
gabortori
wow. tornado alerts, totally messed up my glee schedule.
2 minutes ago via web
@thereadyset my day was pretty decent. i got a tennis ball size bruise at soccer practice ;) how was your day?!?!
about 4 hours ago via web in reply to thereadyset
love is were it's at.
4:14 PM May 2nd via web
@thereadyset omg i wish you were gonna be at warped tour!!! i loovve you <3
8:45 AM May 2nd via web in reply to thereadyset
Name Tori Gabor
Location Appleton, Wisconsin
45 following
14 followers
0 listed
Tweets 53
Favorites
Following
View all...

Why Twitter?

- Twitter represents an extremely large social network — 100 million users.
- Tie formation and destruction are rapid and widespread.
- Combination of text and interpersonal networks.
- Extreme heterogeneity in terms of network properties as well as communication behavior.
- Scalable methods and models.



Modeling Discussion Topics on Twitter

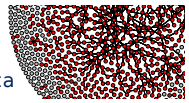
Consider the population of individuals talking about a given topic. Can we make predictions about

- the dynamics of this communication?
- the network properties of this discussion group?
- For now, sampling-based approaches.

Project Activities

Using automated data collection methods we collect information

- on the dynamics of communication content.
- on the properties of communicants' online interpersonal networks.



Twitter Data Collection, Part I - Topic Dynamics

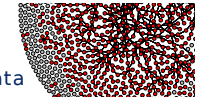
- Public, global content is searchable by keyword.
- Begin with a list of topics each characterized by a set of keywords.
- We include a control topic in which words are chosen from Ogden's word list.
- Automated data collection designed to capture all public tweets containing the given keyword.
- Potential missing data.

Twitter Data Collection, Part II - Personal Networks

- Each user on Twitter has a personal network consisting of friends (out-ties) and followers (in-ties).
- For each keyword we sample 20 recently active users each day and keep them in the sample for 7 days.
- For each user we obtain a list of alters, as well as various covariates if available.
- Potential covariates: location, privacy settings, timezone, account creation date, activity level, language.

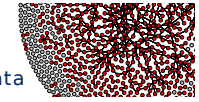
Research Questions

- What seasonality exists within a discussion?



Research Questions

- What seasonality exists within a discussion?
- How do exogenous events affect communication dynamics?



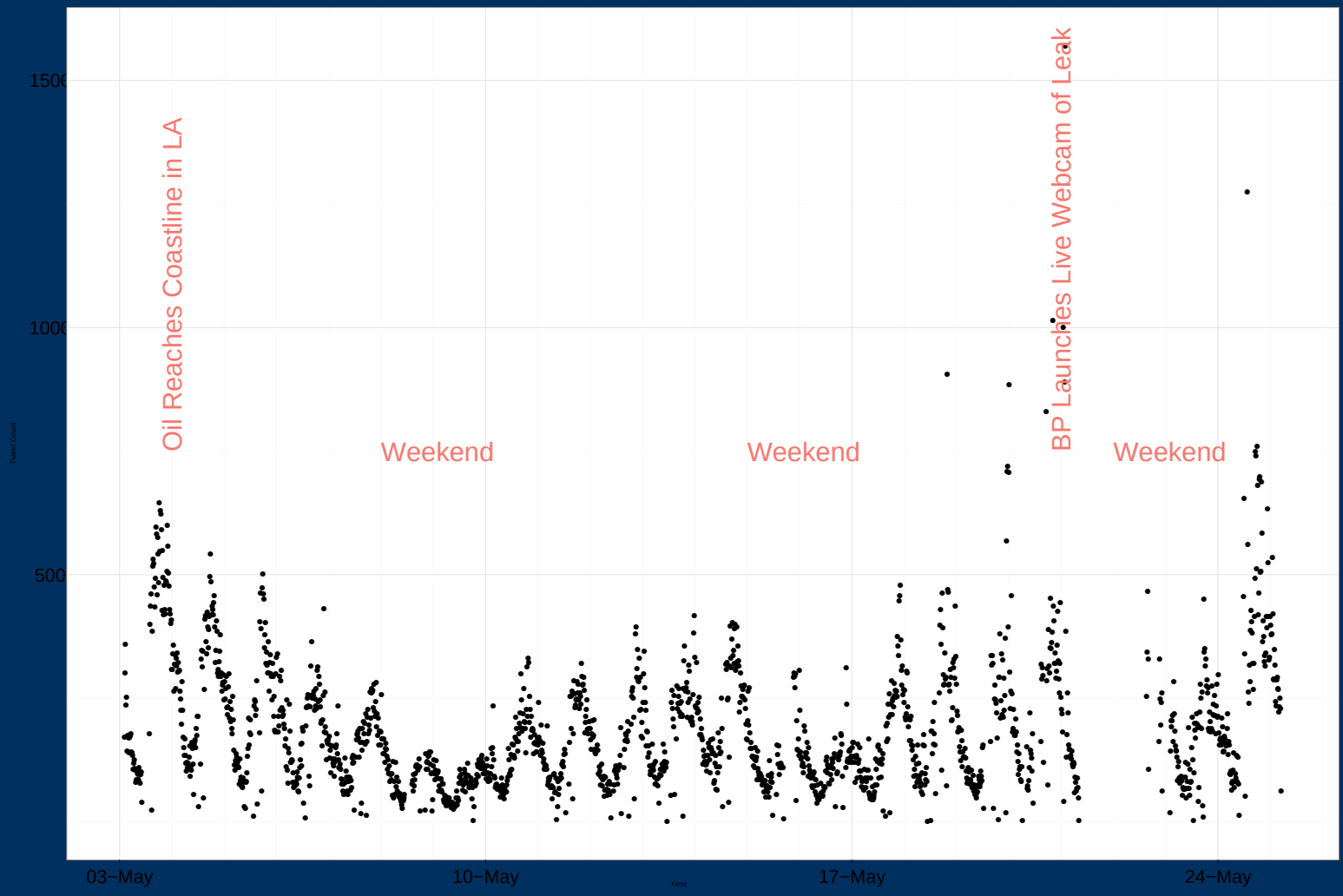
Research Questions

- What seasonality exists within a discussion?
- How do exogenous events affect communication dynamics?
- What are the structural characteristics of the interpersonal networks of the discussant group?

Research Questions

- What seasonality exists within a discussion?
- How do exogenous events affect communication dynamics?
- What are the structural characteristics of the interpersonal networks of the discussant group?
- Individual level prediction?

Oil spill: Seasonality and Exogenous Events

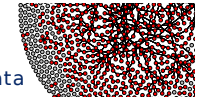


Structural Characteristics of Personal Networks

- Mean degree of topic participants.
- Is an increase in overall mean degree due to those already present in the discussion gaining alters or is it due to high degree individuals entering the discussion?

Degree Distribution Dynamics

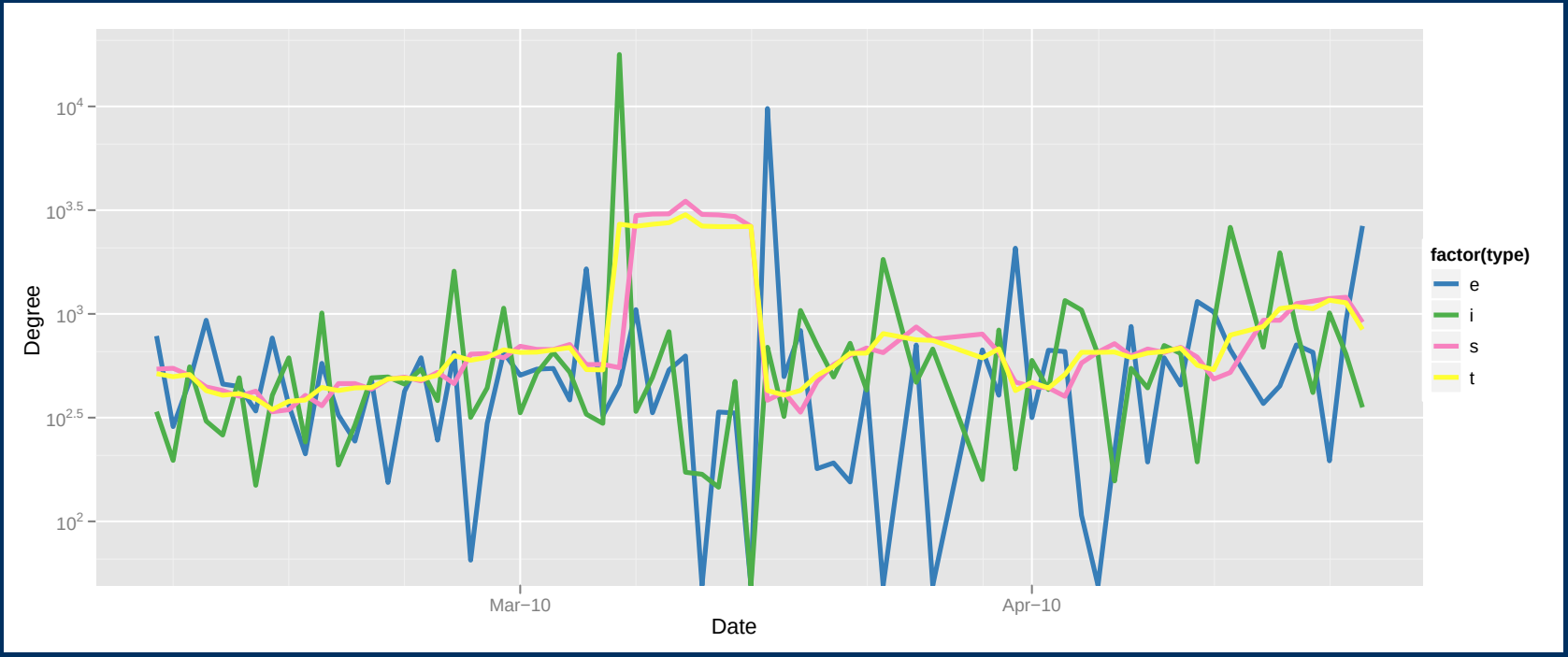
- Consider the mean degree in the population, i.e. topic discussion sample, over time.



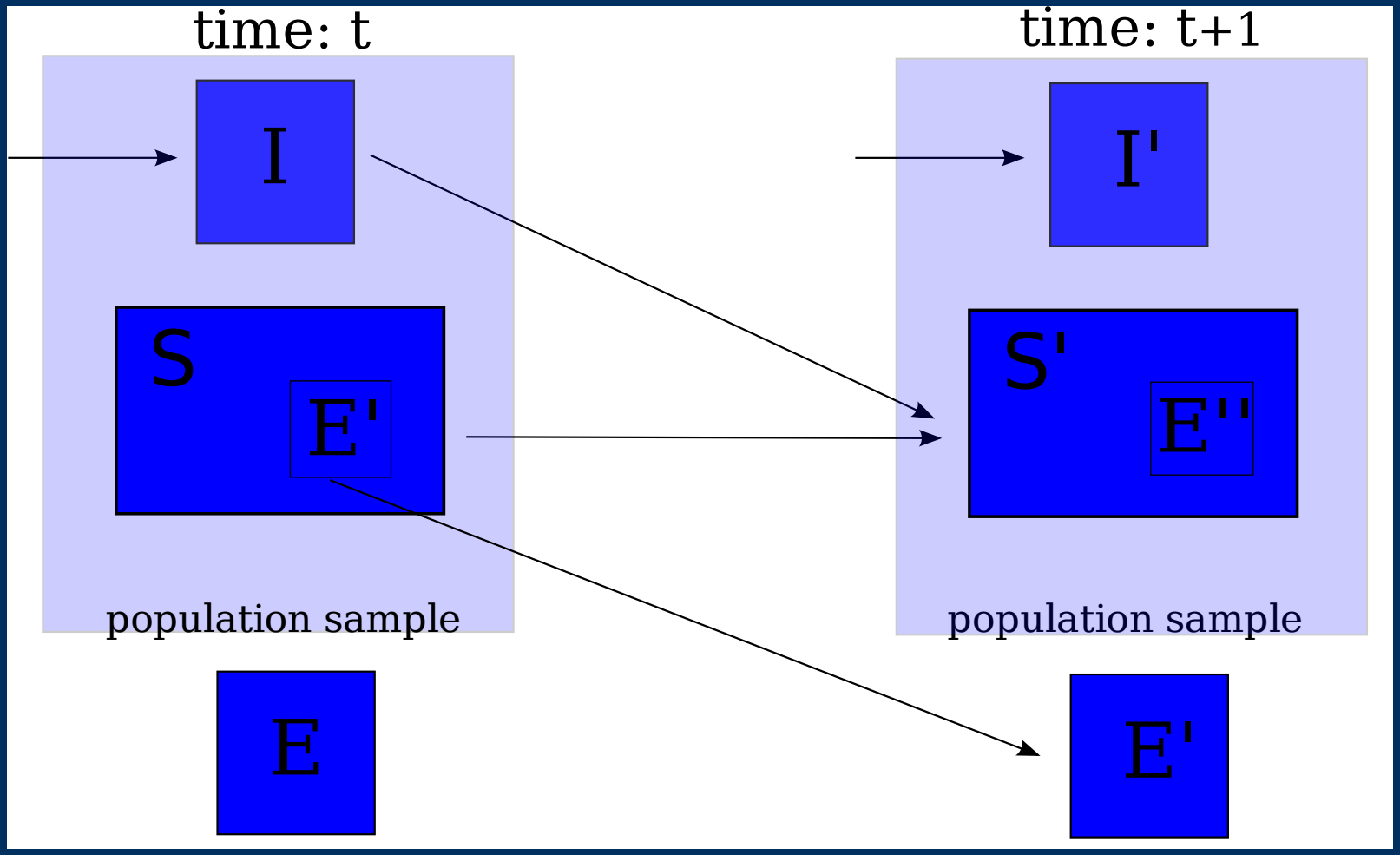
Degree Distribution Dynamics

- Consider the mean degree in the population, i.e. topic discussion sample, over time.
- The mean degree is affected by different population processes: those entering the sample (immigrants), those who stay in the sample (non-migrants), and those who leave the sample (emigrants).

Lightning: Mean Degree Dynamics



Degree Dynamics Decomposition



Degree Dynamics Decomposition

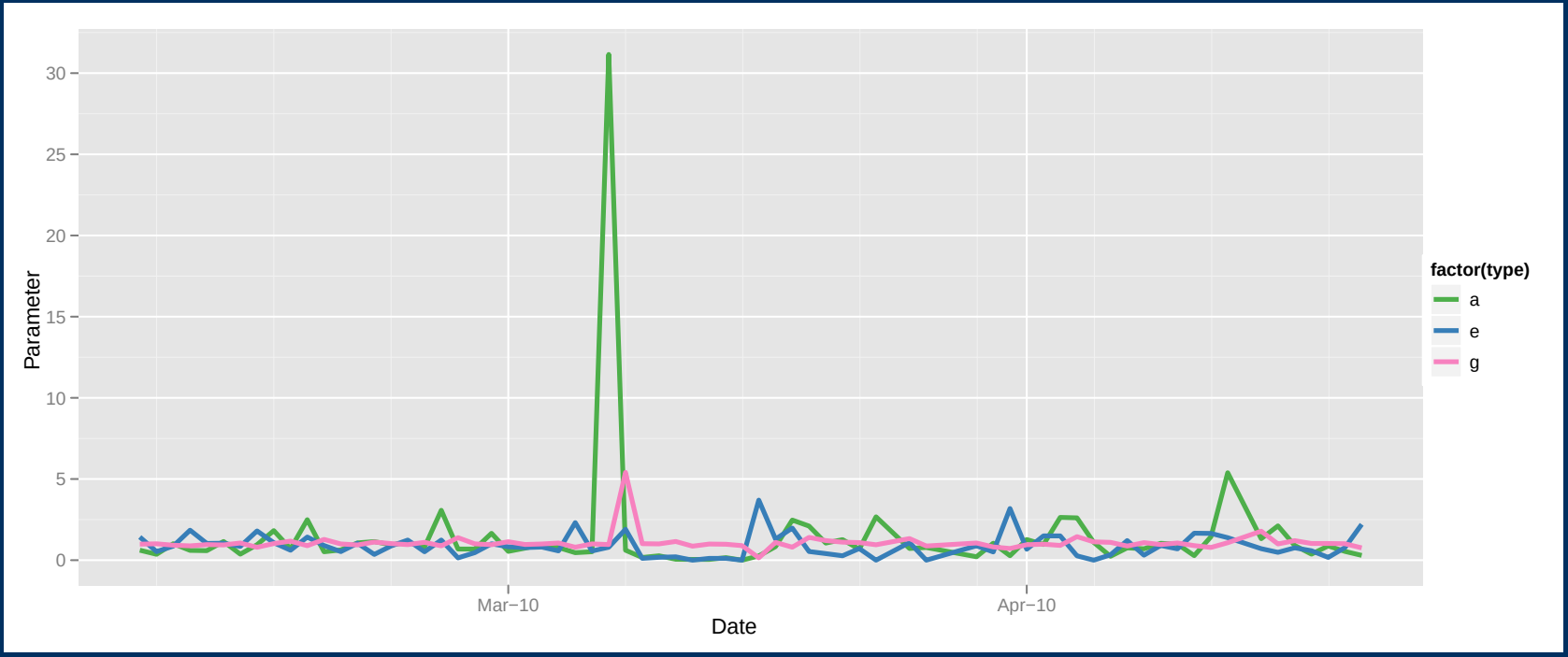
$$\begin{aligned}
\bar{d}_{t+1} - \bar{d}_t &= \frac{\bar{d}_{t+1}(I')N^{I'} + \bar{d}_{t+1}(S')N^{S'}}{N_{t+1}} - \frac{\bar{d}_t(S')N^{S'} + \bar{d}_t(E')N^{E'}}{N_t} \\
&= \frac{\bar{d}_{t+1}(I')N^{I'}}{N_{t+1}} + \frac{\bar{d}_{t+1}(S')N^{S'}}{N_{t+1}} - \frac{\bar{d}_t(S')N^{S'}}{N_t} - \frac{\bar{d}_t(E')N^{E'}}{N_t}
\end{aligned}$$

Degree Dynamics Decomposition

- Let $\bar{d}_{t+1}(I') = \alpha \bar{d}_t(S')$, $\bar{d}_t(E') = \epsilon \bar{d}_t(S')$, $\bar{d}_{t+1}(S') = \gamma \bar{d}_t(S')$.
- Intuitively, we are expressing the respective degrees of the immigrants, emigrants, and (t+1) stayers in terms of what the stayers' degrees were at time t.
- Likewise, let $w_{t+1}^{I'} = N^{I'}/N_{t+1}$, $w_t^{E'} = N^{E'}/N_t$, and $w_{t+1}^{S'} = N^{S'}/N_{t+1}$ be the relative population weights for the three groups.

$$\bar{d}_{t+1} - \bar{d}_t \quad = \quad \alpha \bar{d}_t(S') w_{t+1}^{I'} + \gamma \bar{d}_t(S') w_{t+1}^{S'} \left[1 - \frac{N_{t+1}}{\gamma N_t} \right] - \epsilon \bar{d}_t(S') w_t^{E'}$$

Lightning: Mean Degree Dynamics



Summary

- Modeling large scale dynamic networks with text component.
- Scalability.
- Activity sampling and egocentric properties.

Future Work

- Time-series analysis of the topic data.
- Complete sampling of discussant groups.
- Decomposition of the change in average number of shared partners or other statistics.
- Statistical models of topics on dynamics networks.
- Questions?