

A complex network graph with numerous nodes and edges, rendered in a light gray color, serves as a background for the slide. The nodes are small squares, and the edges are thin lines connecting them, forming a dense, interconnected web.

Dynamics of Communication in Online Social Media

Munmun De Choudhury

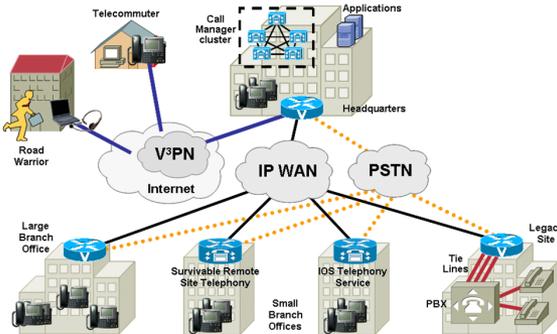
PhD Student, Dept of Computer Sc & Engg, ASU

Advisor: *Prof Hari Sundaram*

Current State-of-the-Art



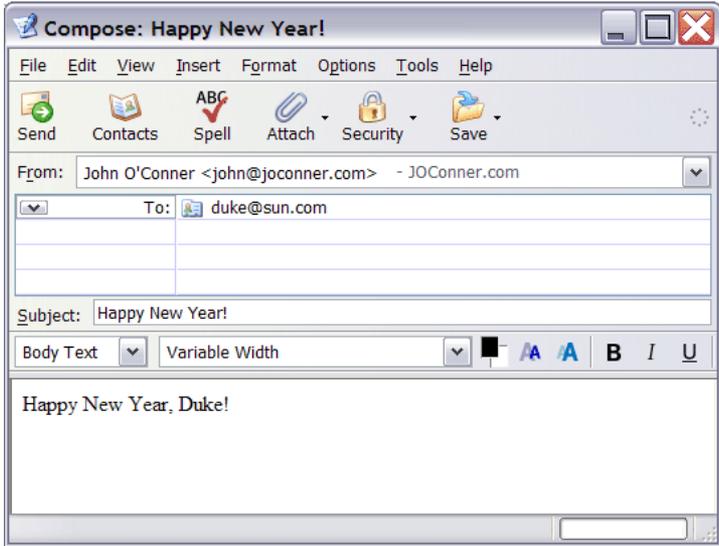
Face-to-face



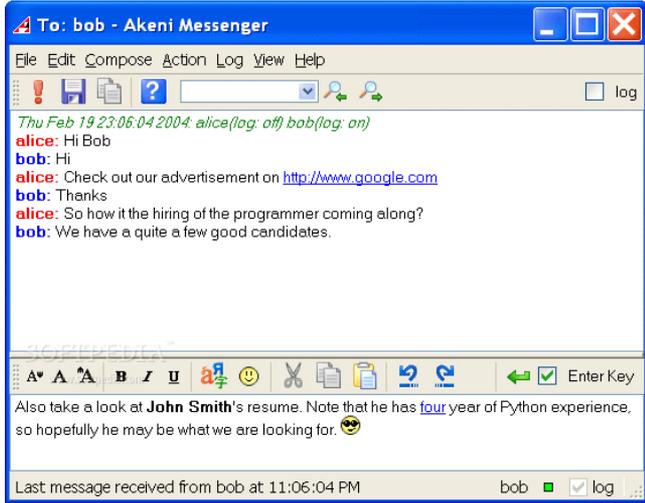
IP Telephony



Text Messaging



Emails



Instant Messaging

Modern communication modes

facebook Home Profile Friends Inbox 45 Mumun De Choudhury Settings Logout Search

Now you can have a username for your Facebook profile
Easily direct friends, family, and coworkers to your profile with a Facebook username. Set your username now.

What's on your mind?

Shanta Pratyusha is wondering when "jimiki-jeans" will be discarded as a dressing faux pas. can see anymore of those lovely long heavy gold stone studs, additional ear rings mis paired with jeans lol ... i guess it is meant to be fun but I know what you are saying. =)) how about the embroidered jeans ?? :-)

Facebook

Slashdot News for Nerds. Stuff that Matters.

Slashdot CSS Redesign Contest
Posted by CmdrTaco on Wednesday April 26, @12:59PM

Why Login? Why Subscribe?

Sections: Main, Apple, AskSlashdot, Books, Developers, Games, Hardware, Interviews, IT, Linux, Politics, Science, YRO

Advertisement: AMD Get Open Source AMD64 Platform Tools at the AMD Resource Center

Slashdot

digg Join Digg About Login Search Digg

News, Videos, Images

- Kid Playing Video Games Gets Attacked by a Dog
- Sen. Webb: Bush using "fear tactics" for more war funding
- How to Use Bleach to Create Your Own T-Shirt Design (PICS)
- Facebook Founder Finds He Wants More Privacy
- Open source hardware gift guide
- Strain Closes
- Chavez Loses Constitution

Digg

LIVEJOURNAL Express Yourself, Share Your Life, Connect with Friends Online

Top 10 in All

Featured in LiveJournal

About LiveJournal: 15.4 million members and counting

LiveJournal

engadget

Navman's S70 navigation system surfaces

Listening to your music

The only iPhone review you need

engadget:mobile

Engadget

Twitter

Reddit

Blogger

MySpace

MetaFilter

Orkut

YouTube



Flickr

Some Social Media Statistics

YouTube	139M users; US\$200M [Forbes].
Flickr	3.6B images; 50M users.
Facebook	200M active users; 1B pieces of content (web links, news stories, blog posts, notes, photos, etc) shared each week.
MySpace	110M monthly active users; 14 B comments on the site.
Digg	3M unique users; \$40M.
Engadget	1,887,887 monthly visitors.
Huffington Post	8.9M visitors.
Live Journal	19,128,882 accounts.

What are the impacts of such large-scale online social communication?

- *Microscopic* level

evolution of shared media characteristics (WWW 09)

information diffusion (WI 07, HT 08, SocialCom 09)

- *Macroscopic* level

group evolution (HT 08, CIKM 08, ICME 09)

network representations (Y! Research summer work)

Motivating Applications ...



What are the interesting conversations on the Blogosphere post Yahoo!'s Bing deal?

Motivating Applications ...



What has been the public buzz on the new Nikon D3000 SLR?

Motivating Applications ...



A collage of various news and social media websites. At the top left is the ReadWriteWeb logo. To its right is the Slashdot website with a headline about a CSS redesign contest. Below that is the Digg website with a headline about a kid playing video games getting attacked by a dog. In the center is the Mashable website with a headline about social media comments. To the right is the Intel Centrino Duo advertisement. At the top right is the New York Times website. At the bottom left is a snippet of a PopURLs article about Mashable, Disqus, and UberVU.

Which is the best news source to read about the recent Twitter crash?

Problem 1

Collaborators: Ajita John, Doree Duncan Seligmann (Avaya Labs)

Iowa Caucus Victory Speech



Why do people *repeatedly* come back to the same YouTube video?

Iowa Caucus Victory Speech



Iowa Caucus Victory Speech



Iowa Caucus Victory Speech



Iowa Caucus Victory Speech



They are returning to do *more than* watch the same video

Iowa Caucus Victory Speech



mushroomagicman (2 days ago) [Reply](#) | [Spam](#) +1

ShitePride61. Your prisons are full of people of colour because they are run by racist Morons like you.

ariasdfghjkl (2 days ago) [Reply](#) | [Spam](#) 0

Обама мусульманский террорист.

rockysman6 (2 days ago) [Reply](#) | [Spam](#) +1

your right about the guns!

WhitePride61 (2 days ago) [Reply](#) | [Spam](#) 0

You got to love the Lies these Democrats put out!
You can't believe a word they say!
How many more troops did Insane Hussein just send to the WAR????
But he was going to bring all the Troops home!
And then Insane Hussein, said he would get rid of all the Worlds Nukes...LOL!!!!
What kind of a Fairyland does this moron live in???

rockysman6 (2 days ago) [Reply](#) | [Spam](#) 0

you know they have to fight that war in afghanistan wp61 he should have never said he was going to bring a singl troop home,campaign promises to get elected and we can't get rid of nukes he has knows that.i voted for the guy.but didn't mccain say we could be in iraq for a 100 years?

WhitePride61 (2 days ago) [Reply](#) | [Spam](#) 0

See now there is the Problem!
You are just thinking of the 2 Party's that have Killed this Country!
There were many more running for Prez!
I think it was some where around 80 people running for Prez!
Anyone is better than a Republican or a Dumocrate!!!!

rockysman6 (2 days ago) [Reply](#) | [Spam](#) 0

Comment removed by author

WhitePride61 (2 days ago) [Reply](#) | [Spam](#) 0

Baby Bush was scared of the N Korean President.
Does anyone see the pattern ?
Baby Bush was just dog-gone scared.

rockysman6 (2 days ago) [Reply](#) | [Spam](#) 0

you are killing me!

tocklove7 (2 days ago) [Reply](#) | [Spam](#) 0

I am not a liberal because labels are weak and tend to promote close-minded un-intelligent people.
People that choose teams over right and wrong LABELS ARE WEAK !
My motto is:Your either right or wrong.
CHOOSE ONE !

axeman196822 (2 days ago) [Reply](#) | [Spam](#) 0

North Korea never launched anything when Bush was in office, they waited for Obama the pathetic coward. I'm sure he will apoligize to N.Korea for Bush not giving them the go-ahead on their nukes like Obama is, because he is such a swell guy.

We think it is the *conversations* around the video they find interesting

An example of an interesting conversation ...



- mushroomagicman** (2 days ago) [Reply](#) | [Spam](#) +1
- ShitePride61. Your prisons are full of people of colour because they are run by racist Morons like you.
- ariasdfghjkl** (2 days ago) [Reply](#) | [Spam](#) 0
- Obama мусульманский террорист.
- rockysman6** (2 days ago) [Reply](#) | [Spam](#) +1
- your right about the guns!
- WhitePride61** (2 days ago) [Reply](#) | [Spam](#) 0
- You got to love the Lies these Democrats put out!
You can't believe a word they say!
How many more troops did Insane Hussein just send to the WAR????
But he was going to bring all the Troops home!
And then Insane Hussein, said he would get rid of all the Worlds Nukes... LOL!!!!
What kind of a Fairyland does this moron live in???
- rockysman6** (2 days ago) [Reply](#) | [Spam](#) 0
- you know they have to fight that war in afghanistan wp61, he should have never said he was going to bring a singl troop home, campaign promises to get elected, and we can't get rid of nukes he has knows that i voted for the guy, but didn't mccain say we could be in iraq for a 100 years?
- WhitePride61** (2 days ago) [Reply](#) | [Spam](#) 0
- See now there is the Problem!
You are just thinking of the 2 Party's that have Killed this Country!
There were many more running for Prez!
I think it was some where around 80 people running for Prez!
Anyone is better than a Republican or a Dumocrate!!!!
- rockysman6** (2 days ago) [Reply](#) | [Spam](#) 0
- Comment removed by author*
- WhitePride61** (2 days ago) [Reply](#) | [Spam](#) 0
- Baby Bush was scared of the N Korean President.
Does anyone see the pattern ?
Baby Bush was just dog-gone scared.
- rockysman6** (2 days ago) [Reply](#) | [Spam](#) 0
- you are killing me!
- tocklove7** (2 days ago) [Reply](#) | [Spam](#) 0
- I am not a liberal because labels are weak and tend to promote close-minded un-intelligent people.
People that choose teams over right and wrong LABELS ARE WEAK!
My motto is, Your either right or wrong.
CHOOSE ONE !
- axeman196822** (2 days ago) [Reply](#) | [Spam](#) 0
- North Korea never launched anything when Bush was in office, they waited for Obama the pathetic coward. I'm sure he will apologize to N.Korea for Bush not giving them the go-ahead on their nukes like Obama is, because he is such a swell guy.

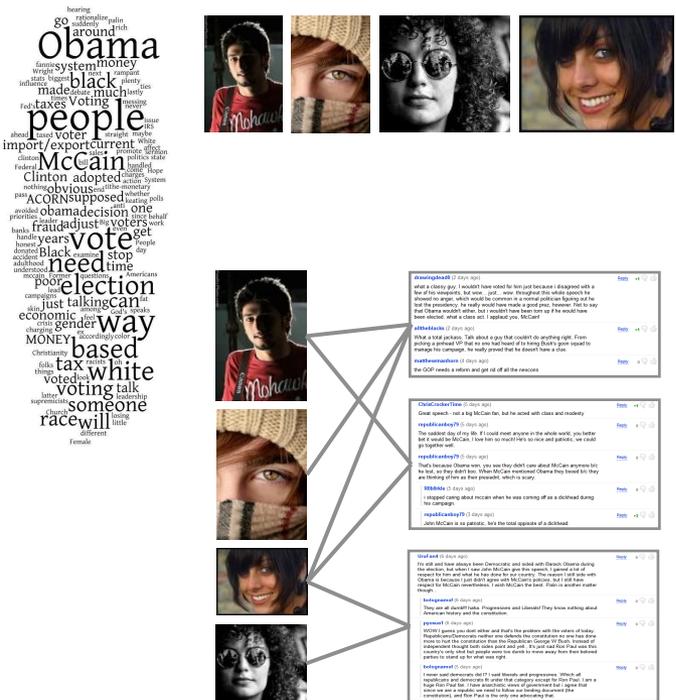


Our Contributions

- Goal:
 - What causes a conversation to be interesting, that prompts a user to participate in the discussion on a posted video?

- Approach:

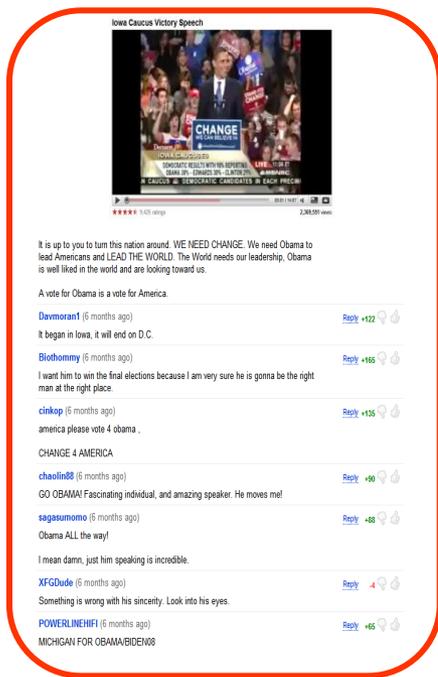
- Detect conversational themes.
- Determine interestingness of participants and interestingness of conversations based on a random walk model.
- Measure the consequence of a conversations.
- Excellent results on a dataset from YouTube.



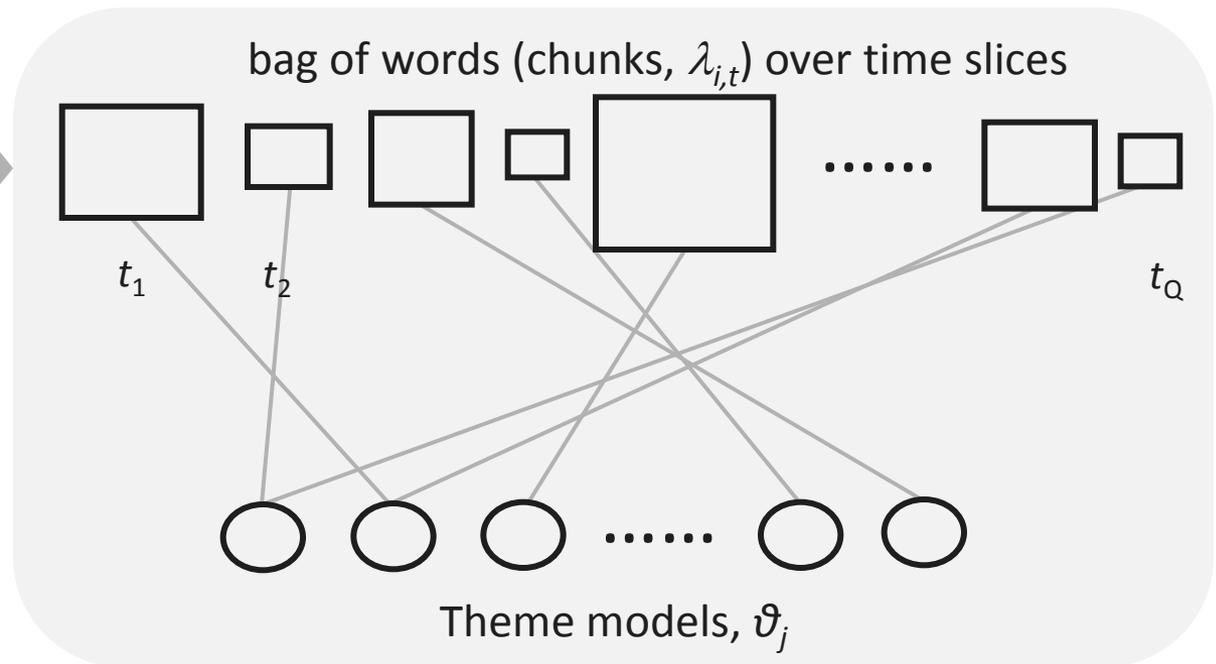
How do we determine interestingness?

Conversational Themes

- Conversational themes are sets of salient topics associated with conversations at different points in time.



conversation i



$$p(\theta_j | \lambda_{i,t}, t) = ?$$

Theme Model

- Temporal Regularization

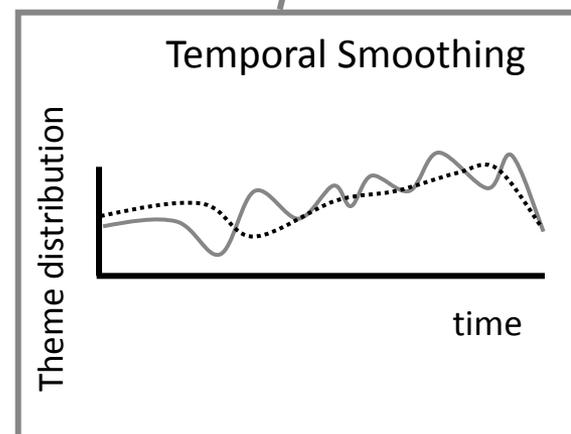
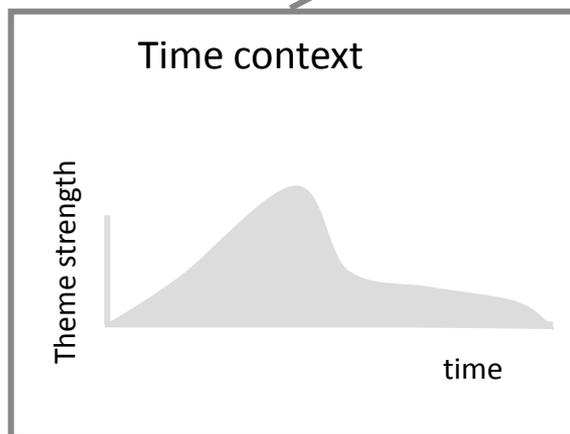
- A word w in the chunk can be attributed either to the textual context $\lambda_{i,t}$, or the time slice t
- Smoothness of theme models over time

$$L(C) = \sum_{\lambda_{i,t} \in C} \sum_{w \in \lambda_{i,t}} n(w, \lambda_{i,t}) \cdot \log \sum_{j=1}^K \left(p(w, \theta_j | \lambda_{i,t}, t) + \exp(-d_T(j)) \right)$$

Chunk context



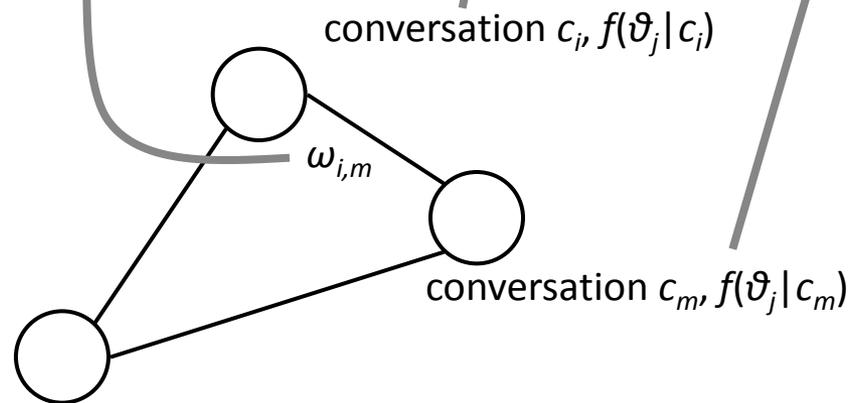
A screenshot of a movie page with a red box highlighting the text content.



Theme Model (Contd.)

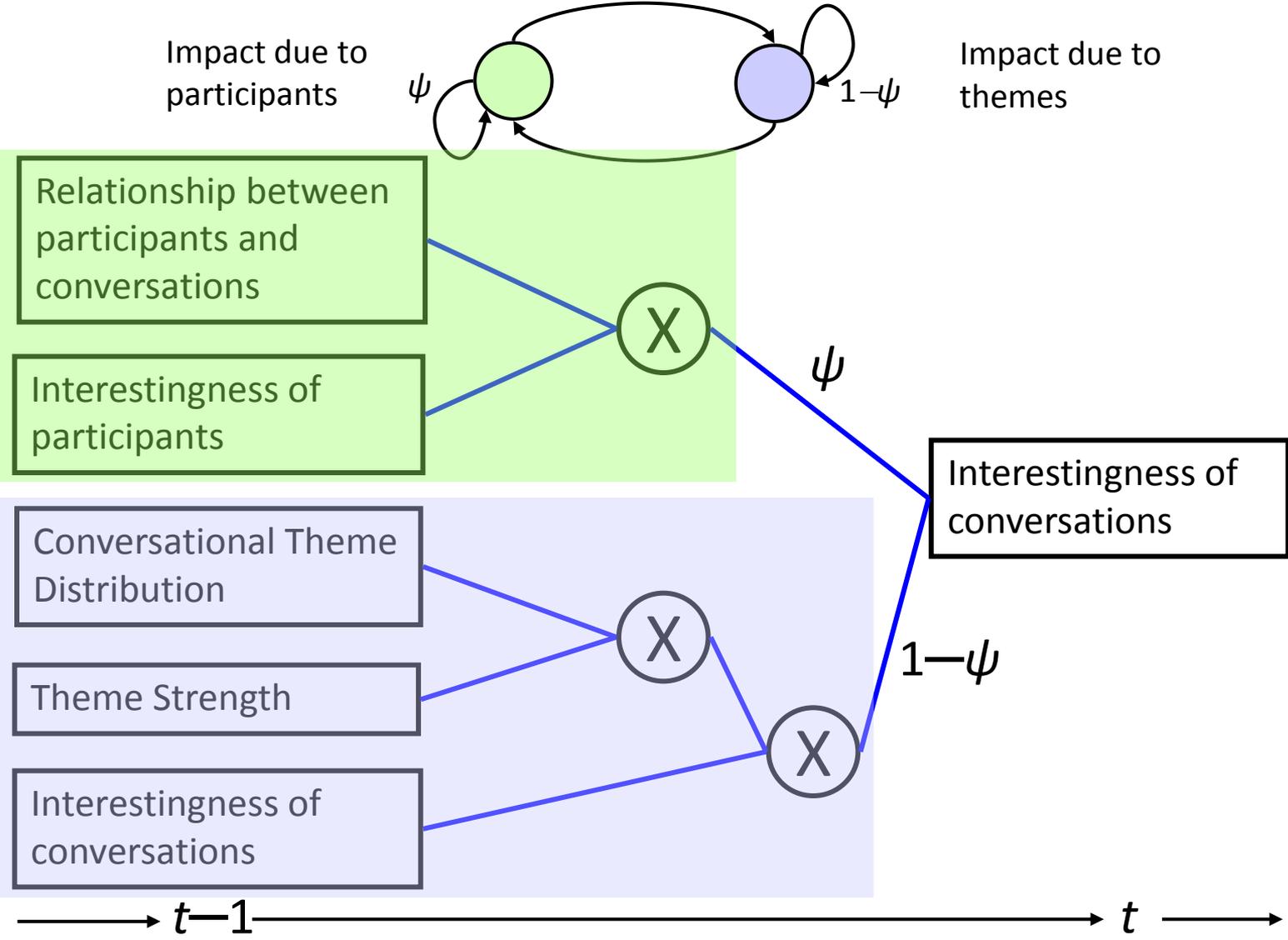
- Co-participation based Regularization
 - If several participants comment on a pair of chunks, their theme distributions are likely to be close to each other.

$$R(C) = \sum_{c_i, c_m \in C} \sum_{j=1}^K \left(\omega_{i,m} - \left(1 - \left(f(\theta_j | c_i) - f(\theta_j | c_m) \right)^2 \right) \right)^2.$$

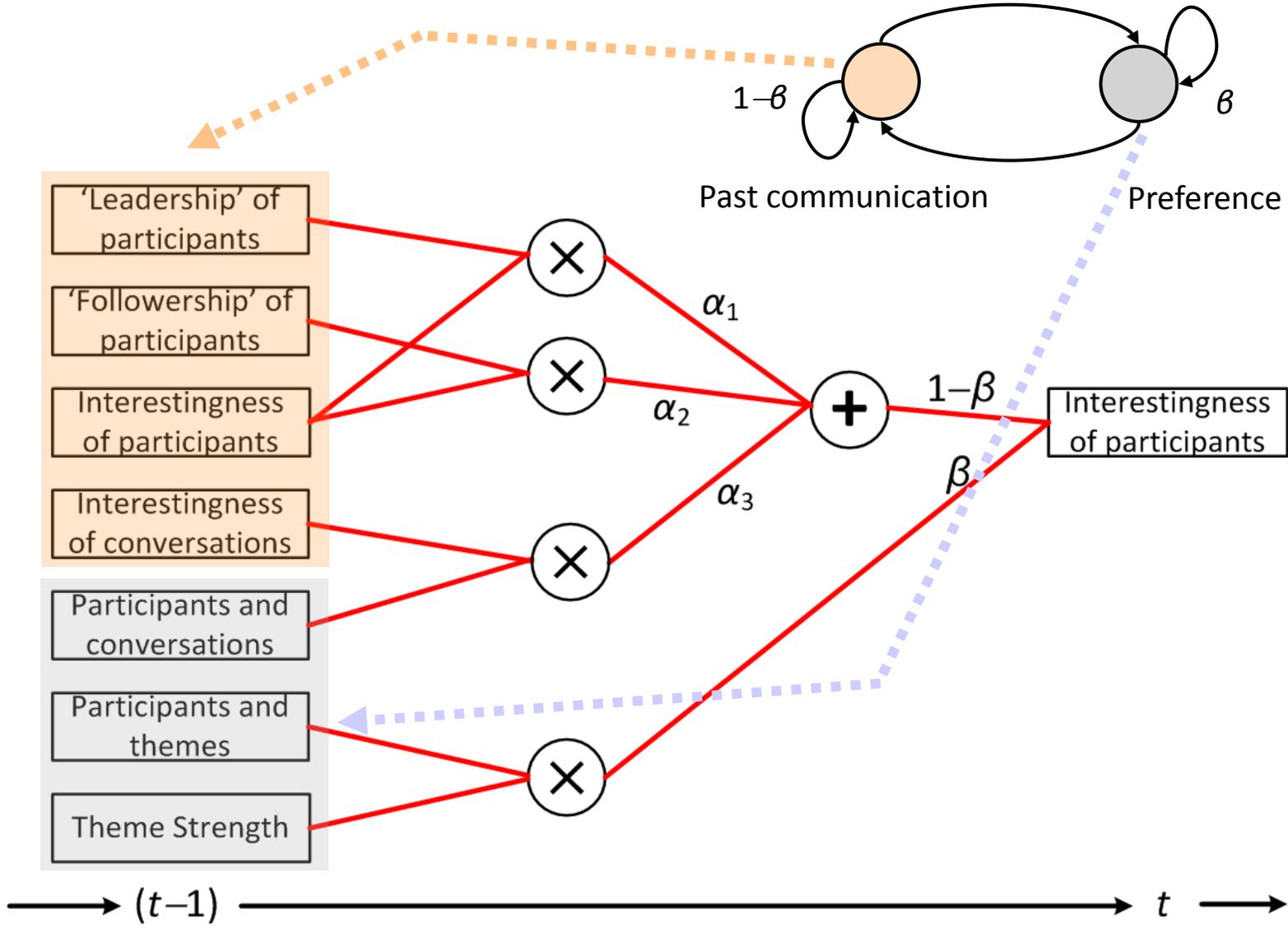


$$O(C) = -(1 - \zeta) \cdot L(C) + \zeta \cdot R(C)$$

Interestingness of Conversations



Interestingness of Participants



Joint Optimization of Interestingness

- A joint optimization framework, which maximizes the two interestingness measures for optimal $\mathbf{X}=(\alpha_1, \alpha_2, \alpha_3, \psi)$ and also incorporates temporal smoothness:

$$g_1(\mathbf{X}) = \rho \cdot \|\mathbf{I}_p(\mathbf{X})\|^2 + (1 - \rho) \cdot \|\mathbf{I}_c(\mathbf{X})\|^2 + \exp(-d_p) + \exp(-d_c)$$

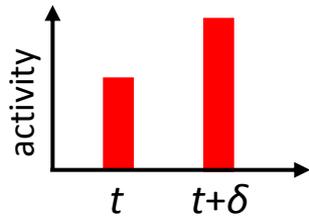
Interestingness of
participants

Interestingness of
conversations

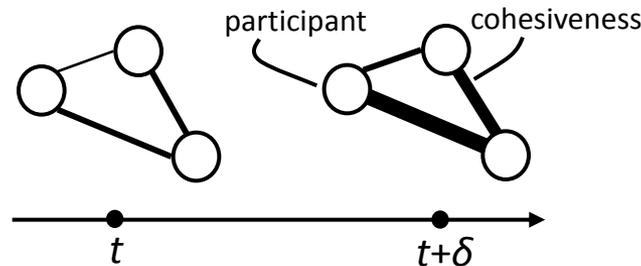
Regularization of
participants'
interestingness

Regularization of
conversations'
interestingness

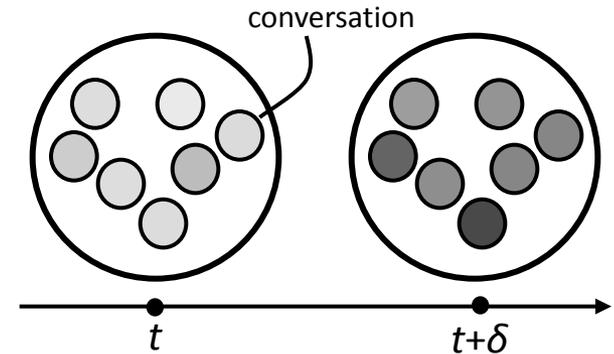
What happens *after* a conversation becomes interesting?



Participant activity



Participant Cohesiveness



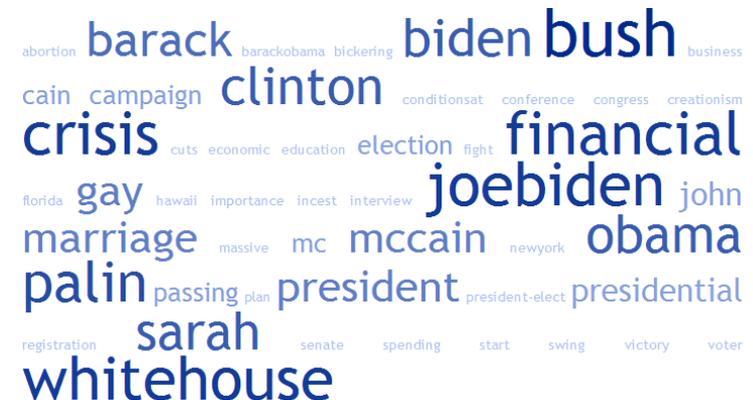
Thematic interestingness

Three consequence metrics of interestingness

YouTube Dataset

- ‘News & Politics’ category on YouTube – rich communication on highly dynamic events.
 - 132,348 videos
 - ~ 9M unique participants
 - ~ 89M comments
 - 15 weeks from June 20, 2008 to September 26, 2008

The screenshot shows the YouTube homepage interface. At the top, there are navigation tabs for Home, Videos, Channels, and Community. Below this is a search bar and an 'Upload' button. The main content area is divided into several sections: 'Videos being watched right now...', 'Featured Videos' (with thumbnails for 'The Jonas Brothers & Miley Cyrus...', 'Mumbies Gets A Job...Part 1', 'FND Films - Parkour Kids!', and 'Eli's Dirty Jokes - Episode 28...'), and 'Spotlight Videos'. The 'Spotlight Videos' section has a red circle around the 'Most Viewed' tab, which is selected. Below the tabs are three video thumbnails with titles: 'How To File Your Taxes on the Last Day', 'How to file your taxes online using onions and electrolytes', and 'Your Money: Avoiding Tax Audits'. On the right side, there are promotional banners for Google Chrome, 'Get the Latest News Every Day', and 'What's New'.

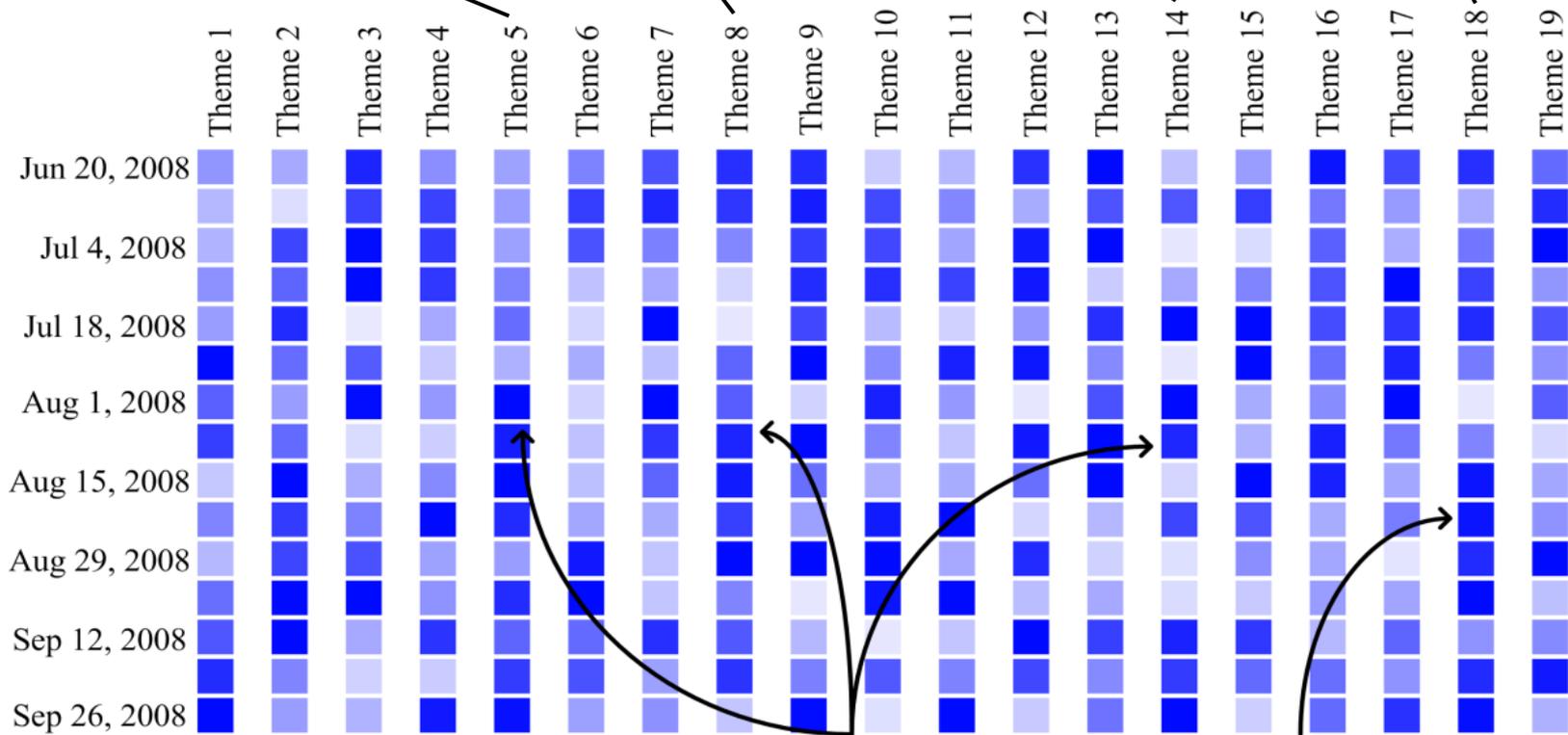


want, agree, put, just, think, respect, ones, towards, outlaws, make, army, people, force, freedom, point, train, opinion, first, soldiers, military, really, one, identity, get, draft, propaganda, like

world, overpopulated, abortion, wrong, choice, technology, population, Pro-choice, say, fetus, around, child, life, well, raped, like, just, people, pro, just, cases, choose, condom, sperm, fact, destroying, murder, every, social, much, an, right

private, problems, insurance, like, offered, worse, tooth, care, health, want, see, get, service, need, together, let, government, national, people, healthcare, pay, system, America, possible, live, business, can

Palin, real, answering, one, campaign, economy, question, Biden, plan, Joe, sarah, system, try, Article, less, try, businesses, Branch, high, see, energy, said, polarization, just, tax, small, expert, even, like, coverage, make, fact, truth, time, made, test, McCain, Executive, job, get

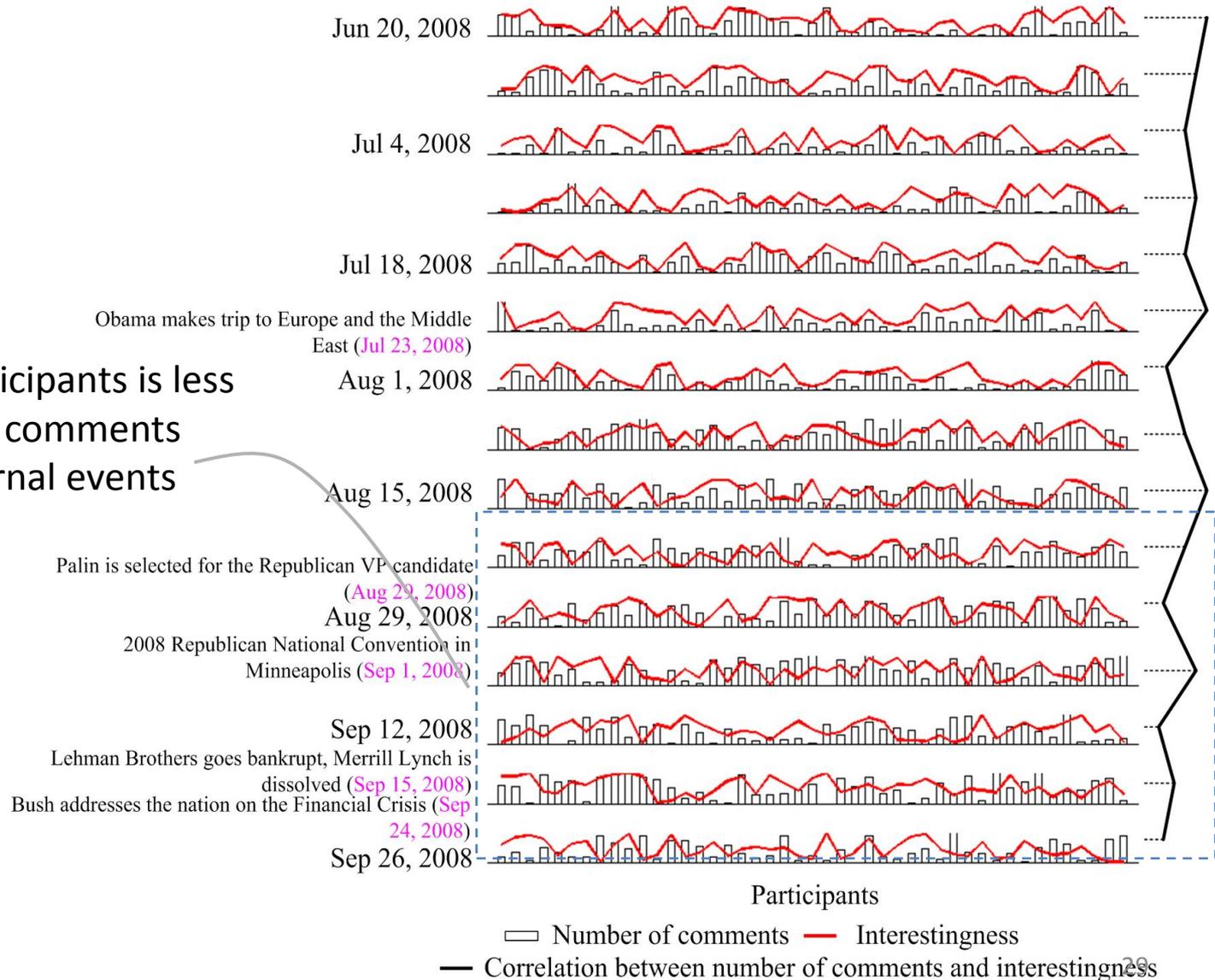


Themes consistently discussed as primary election issues

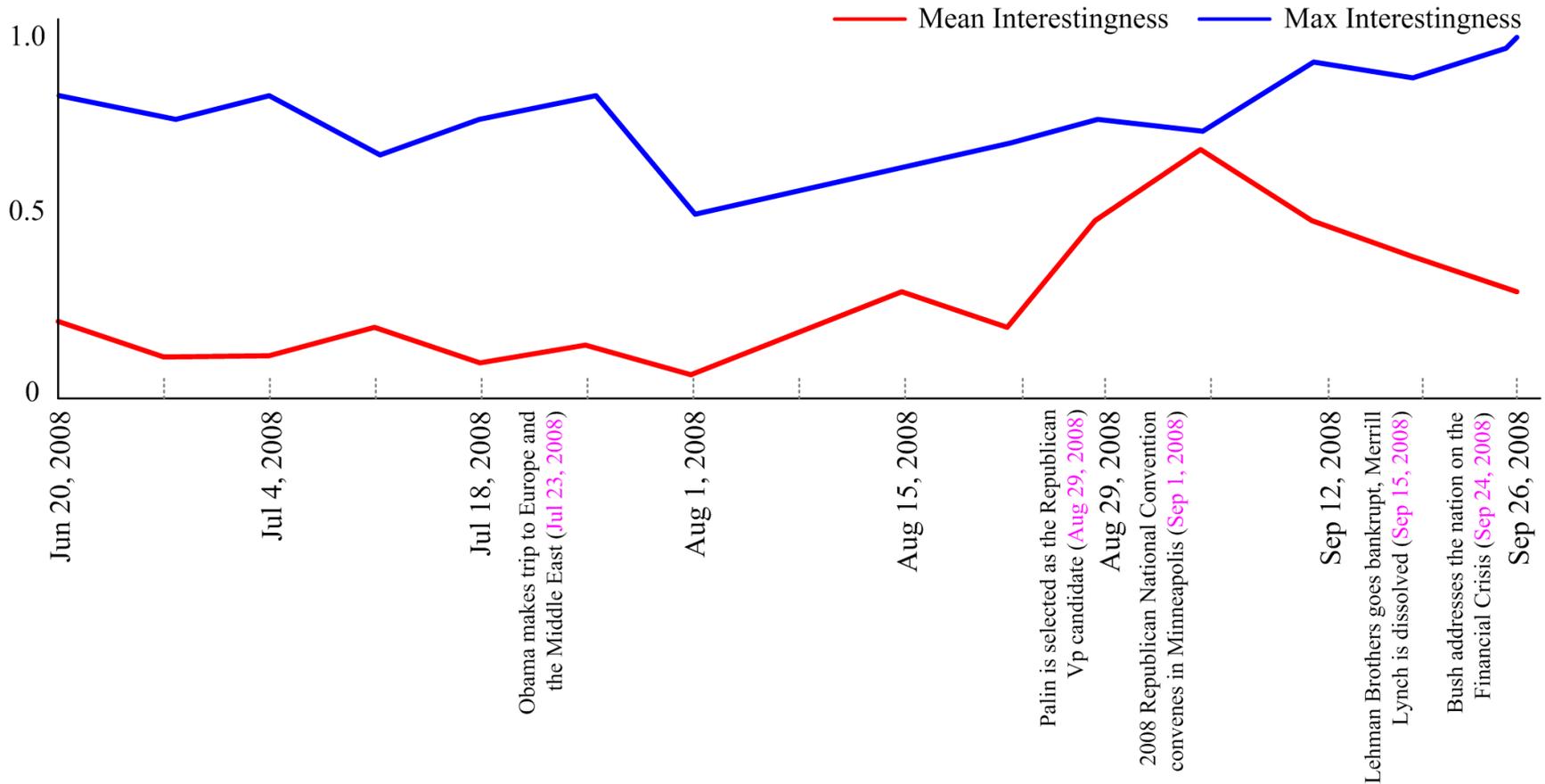
Theme gets strong over discussion of the two VP

Analysis of Interestingness of Participants

Interestingness of participants is less affected by number of comments during significant external events



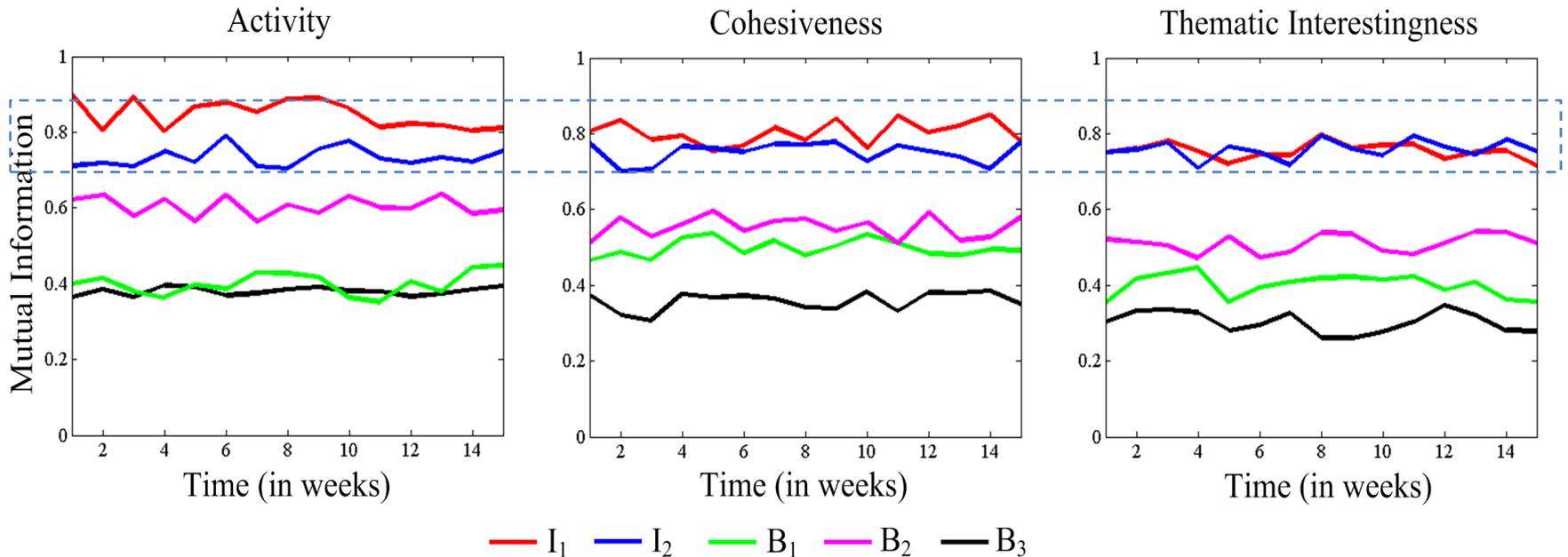
Analysis of Interestingness of Conversations



- Mean interestingness of conversations increases during periods of several external events; however, certain highly interesting conversations always occur at different weeks irrespective of events.

Evaluation using Consequences

- Interestingness is computed using five techniques –
 - our method with temporal smoothing (I_1),
 - our method without temporal smoothing (I_2) and
 - the three baseline methods,
 - B_1 (comment frequency),
 - B_2 (novelty of participation),
 - B_3 (co-participation based PageRank).

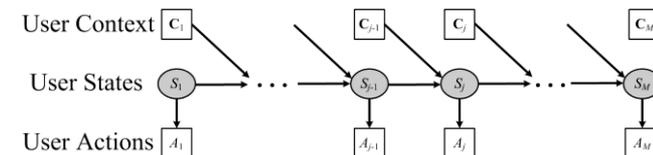
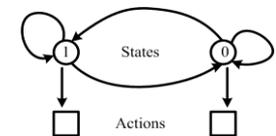
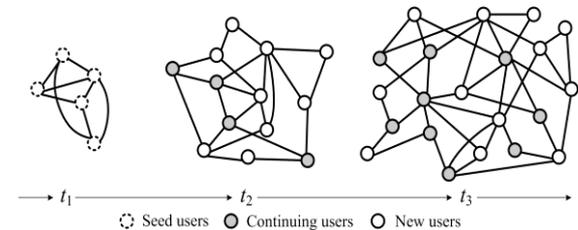


Problem 2

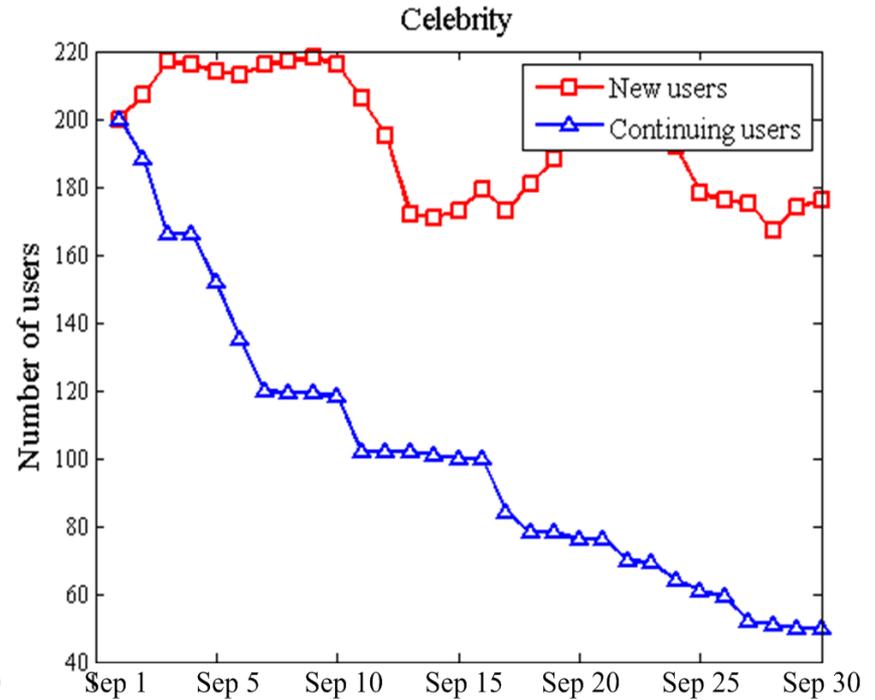
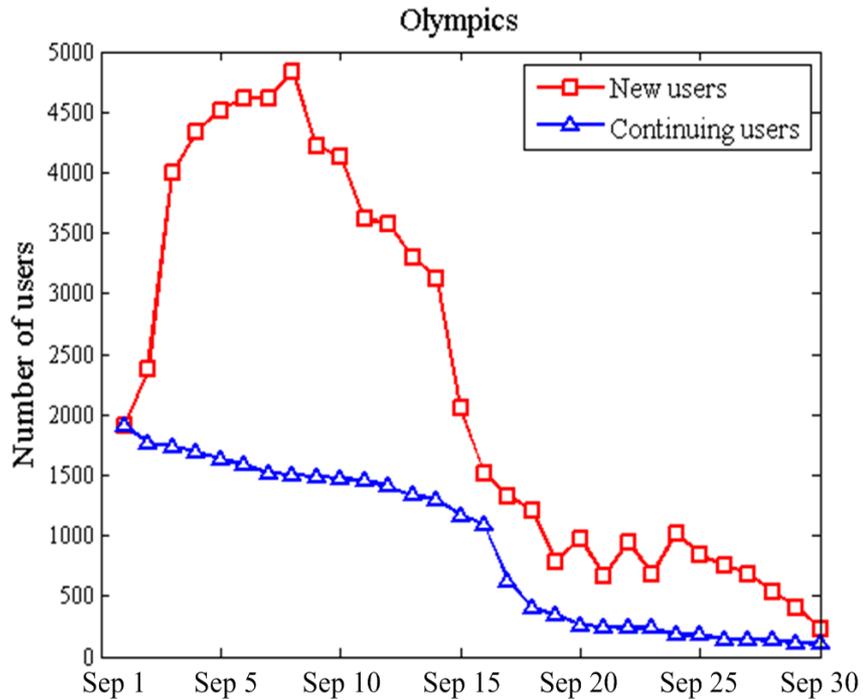
Collaborators: Ajita John, Doree Duncan Seligmann (Avaya Labs)

Social Synchrony

- Goal:
 - a framework for predicting social *synchrony* in online social media over a period of time into the future.
- Approach:
 - Operational definition of social synchrony.
 - *Learning* – a dynamic Bayesian representation of user actions based on latent states and contextual variables.
 - *Evolution* – evolve the social network size and the user models over a set of future time slices to predict social synchrony.
- Excellent results on a large dataset from the popular news-sharing social media Digg.



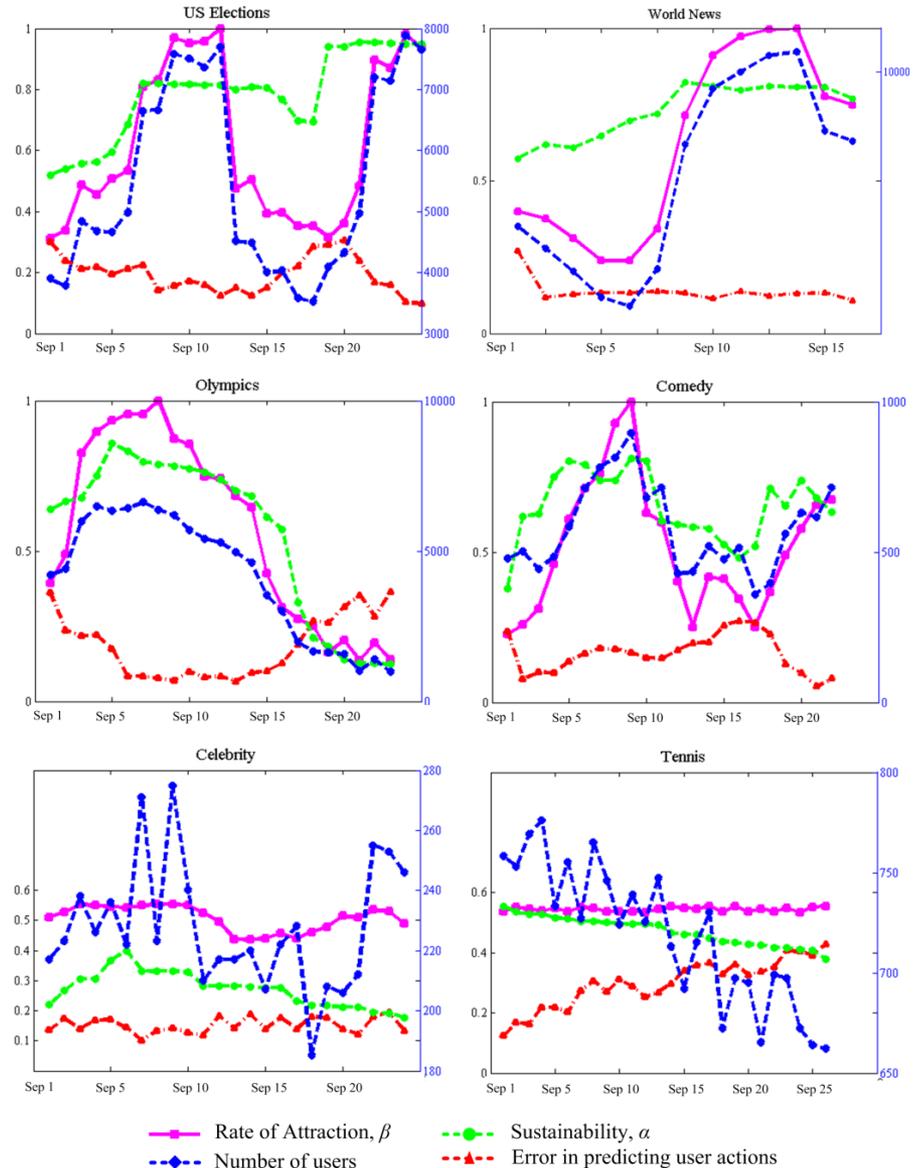
Observations from Digg



Topic 'Olympics' is observed to exhibit synchrony where old users continue to be involved in the action of digging stories, as well as large number of new users join in the course of time (Sept 3-Sept 13).

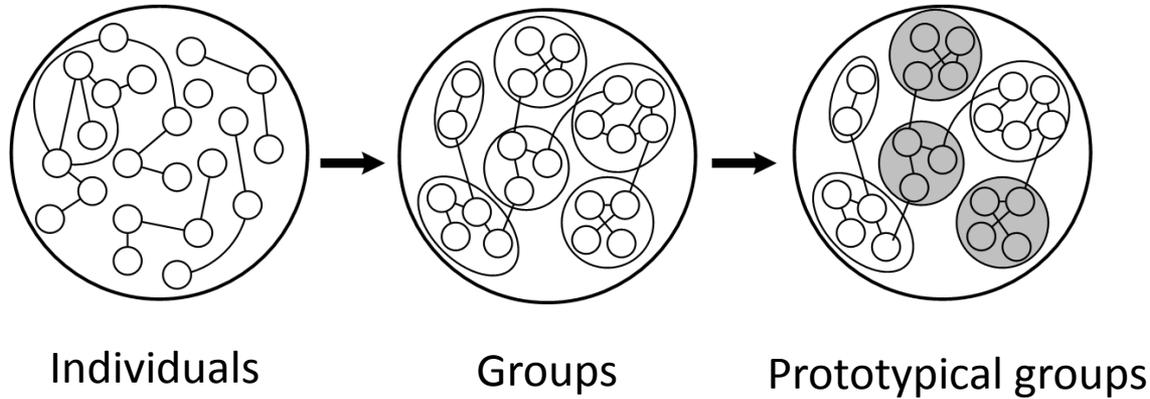
Experiments on Digg Data

- Digg dataset
 - August, September 2008
 - 21,919 users, 187,277 stories, 7,622,678 diggs, 687,616 comments and 477,320 replies.
 - Six sample topics – four inherently observed to have synchrony.



Problem 3

Collaborators: Ajita John, Doree Duncan Seligmann (Avaya Labs)



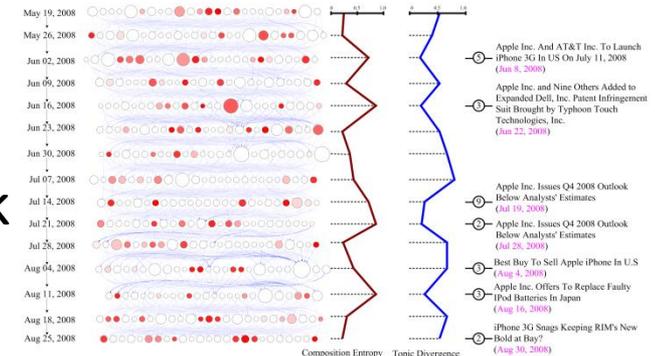
How do groups evolve centered around communication in the Blogosphere?

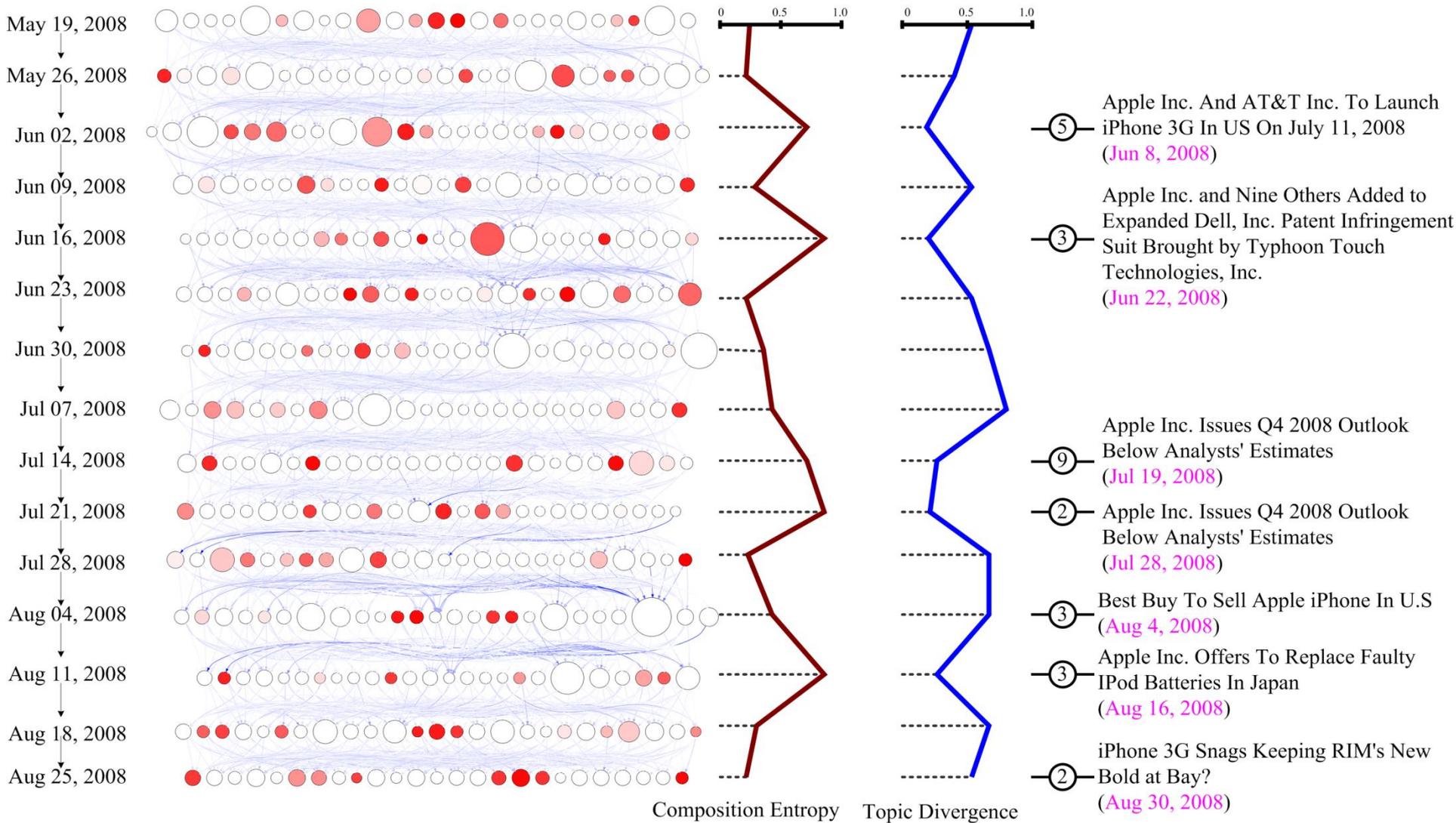
Mining Prototypical Social Groups

- Goal:
 - How do we characterize communication at the level of individuals and groups?
 - How do we extract groups?
 - Which are the prototypical groups given a topic?

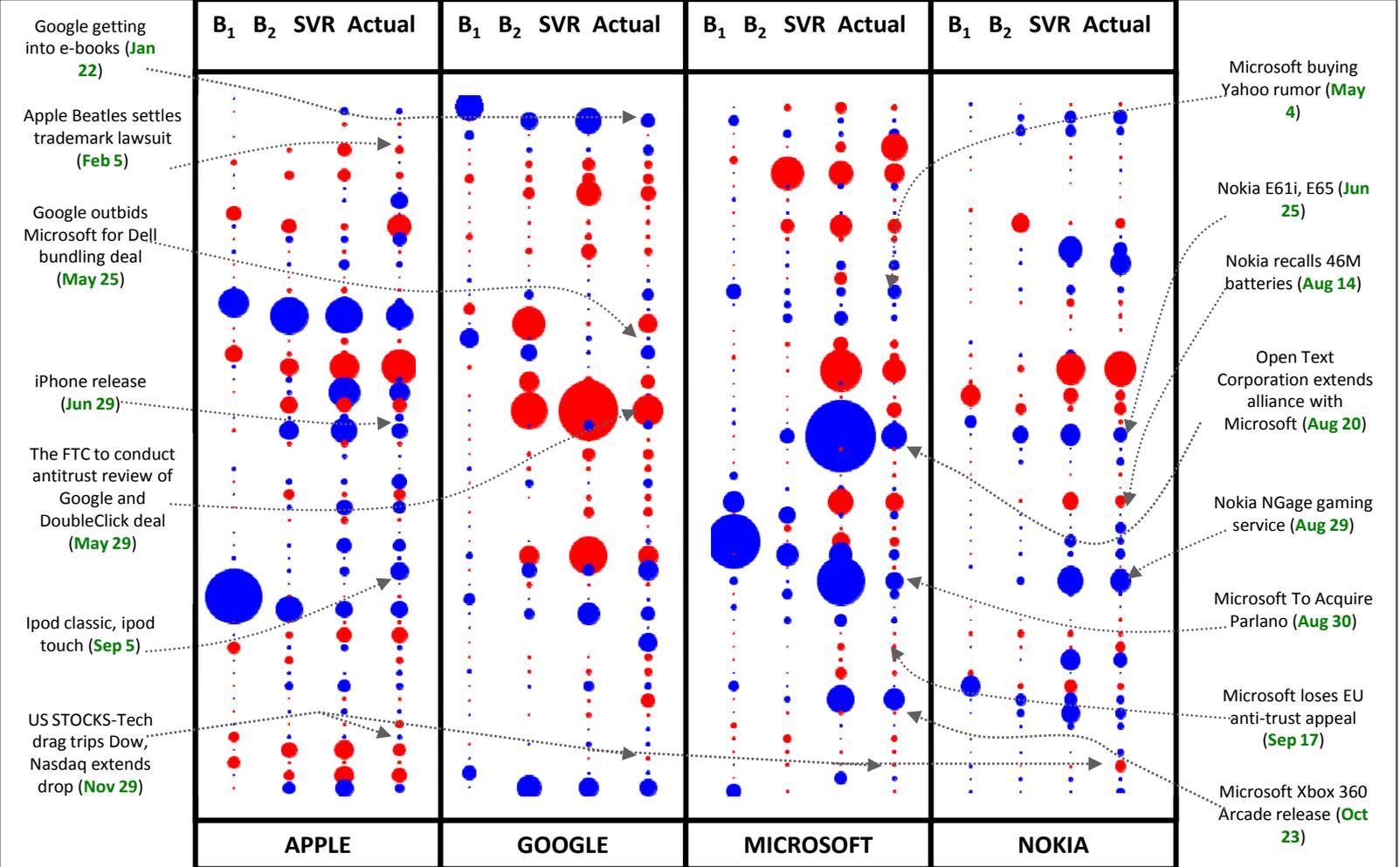
■ Approach:

- Extract individual characteristics based on communication.
- Extract groups based on a random walk model based clustering algorithm.
- Determine prototypical groups – composition entropy, topic divergence.





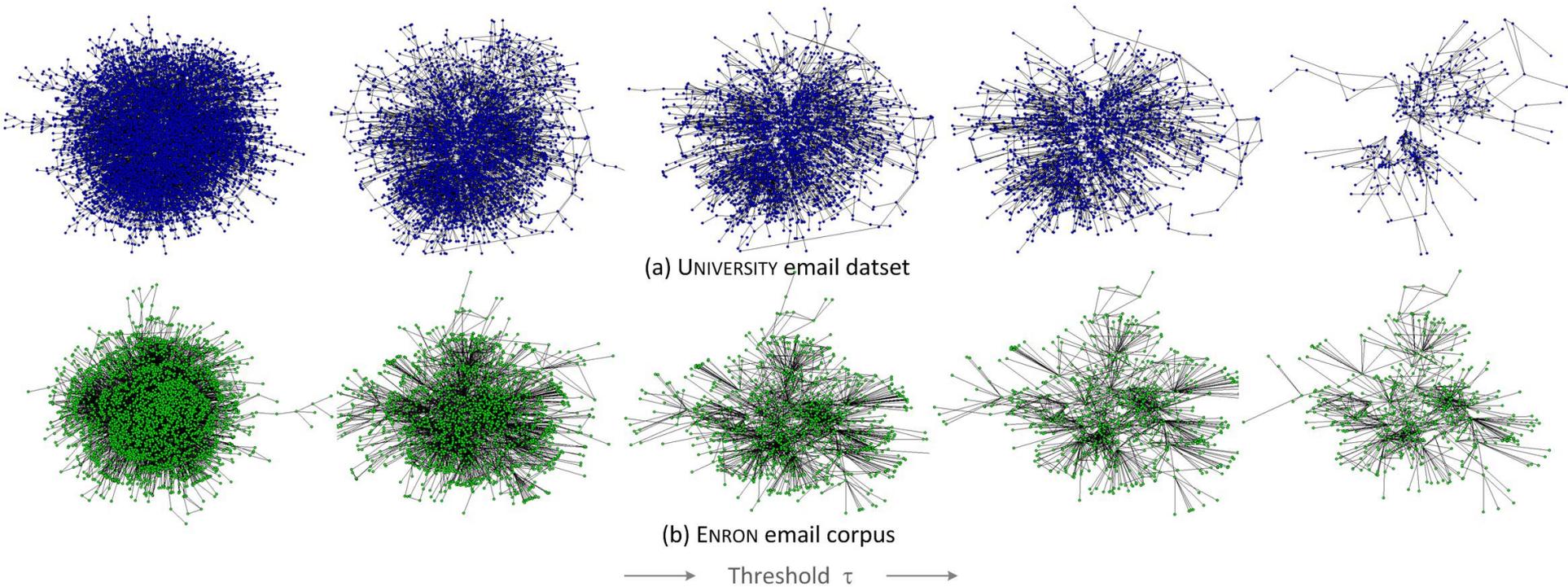
Evaluation



Visualization of stock movements with time on vertical scale. B₁ and B₂ are two baseline techniques. Blue bubbles indicate positive movement and red bubbles negative movements. Sizes of the bubbles represent magnitude of movement. The SVR prediction is found to follow the *movement trend* very closely with an error of 26.41 %.

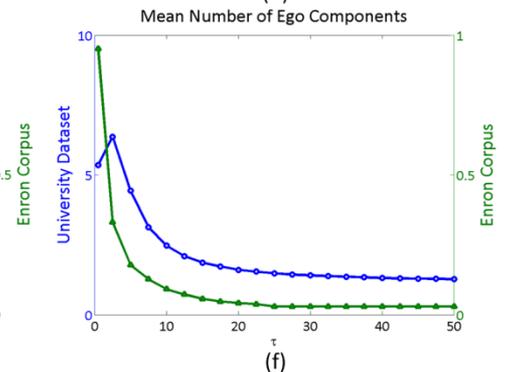
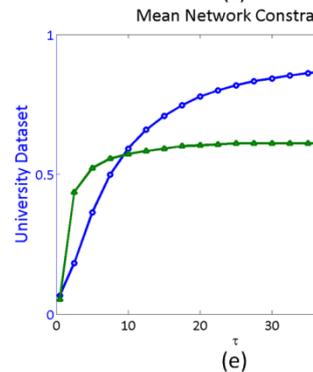
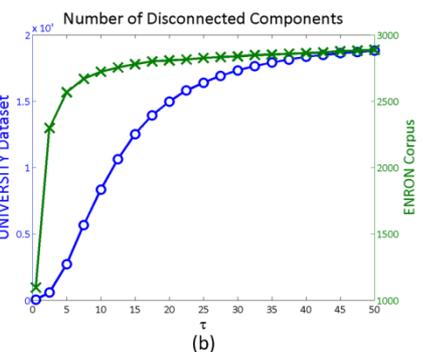
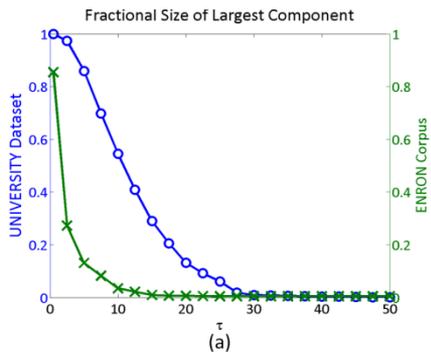
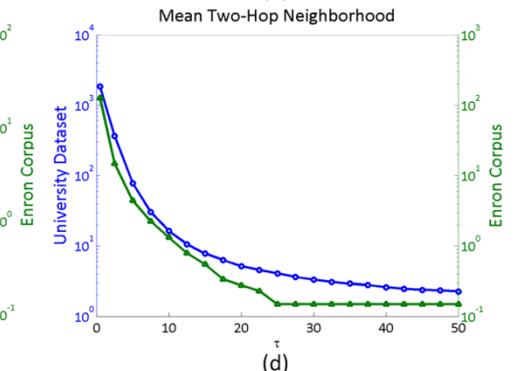
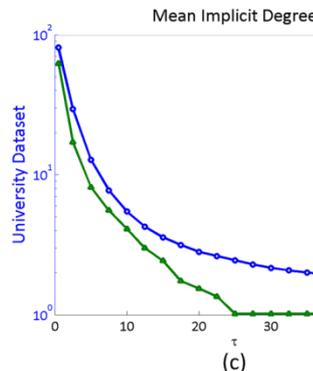
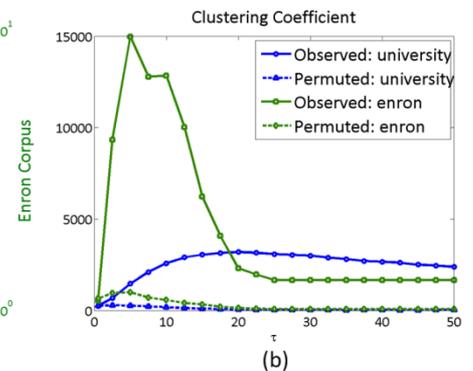
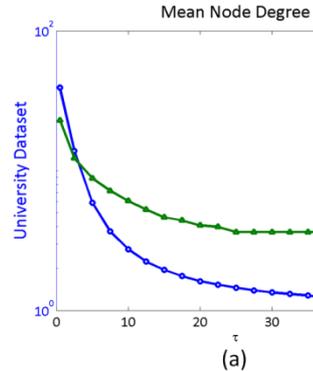
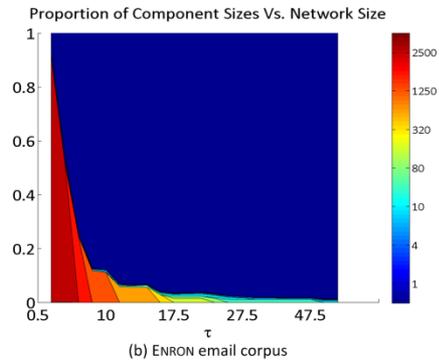
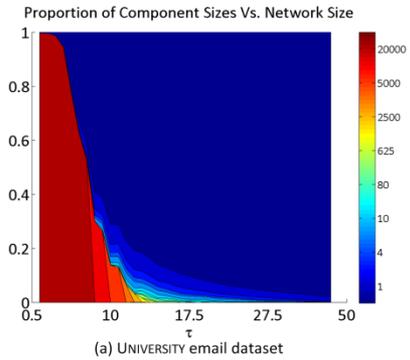
Problem 4

Collaborators: Winter Mason, Jake Hofman, Duncan Watts (Yahoo! Research)



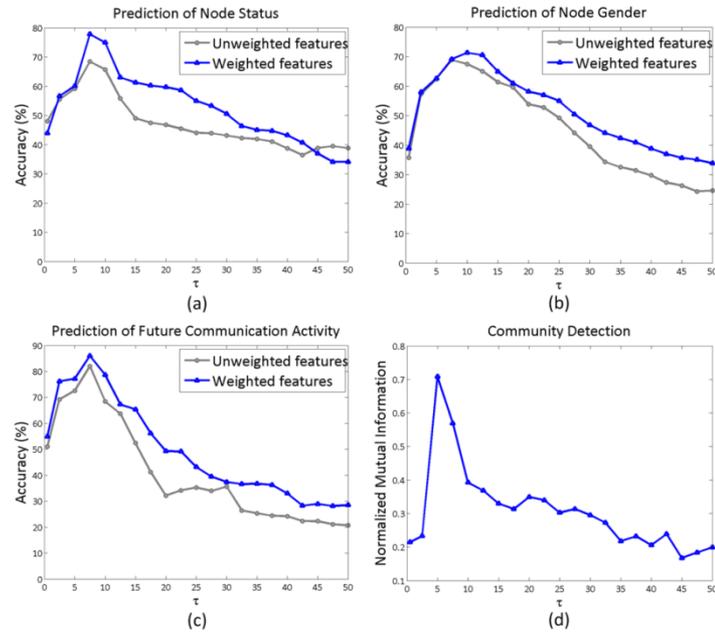
How do we infer optimal network structures based on social communication?

Empirical Observations

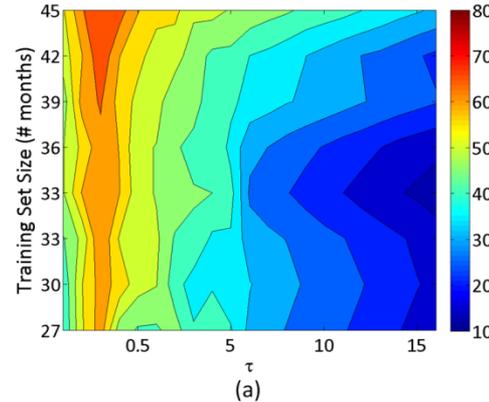


—○— UNIVERSITY email dataset —×— ENRON email corpus

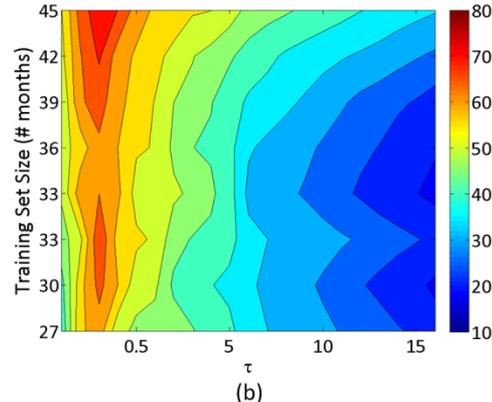
Prediction Tasks



Prediction of Future Communication Activity (Unweighted)



Prediction of Future Communication Activity (Weighted)





Summary...