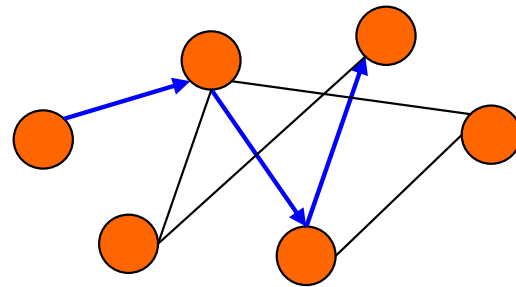


Dynamic Prediction of Communication Flow using Social Context



Munmun De Choudhury

Hari Sundaram

Ajita John

Dorée Duncan Seligmann

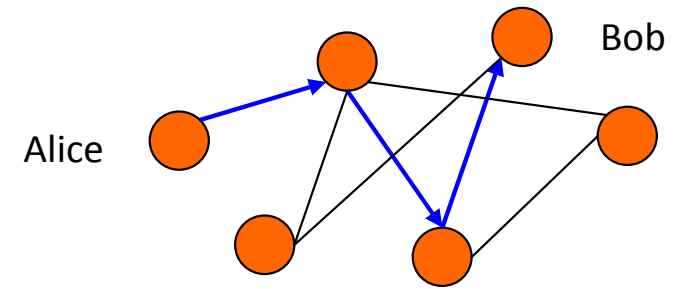


*Arts, Media & Engineering
Arizona State University, Tempe*

*Collaborative Applications Research
Avaya Labs, New Jersey*

Introduction

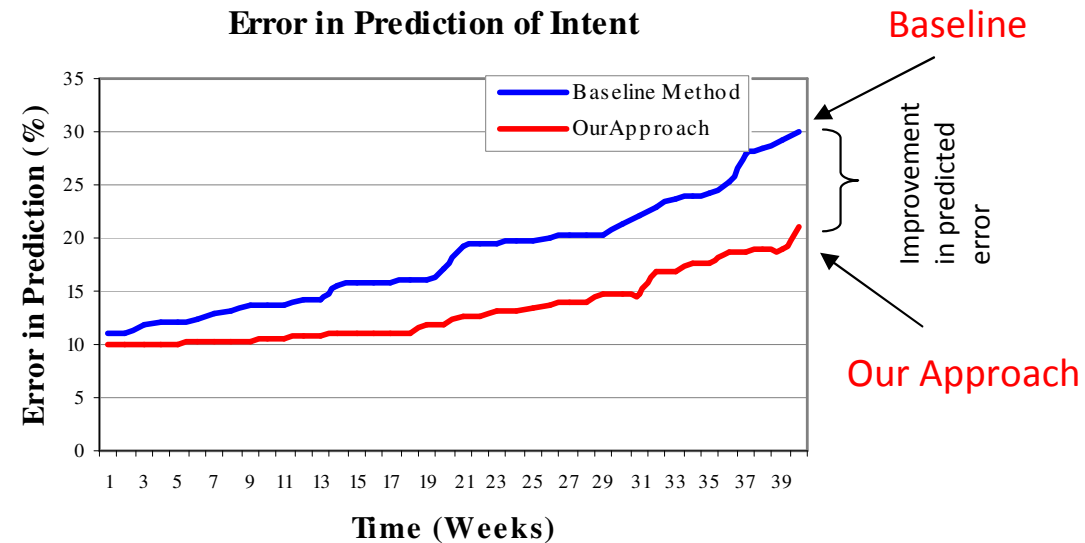
A temporal representation framework using communication and **social context** to efficiently predict communication flow in social networks.



- Why is the problem important?
 - Determine information propagation and the roles of people in the process.
 - Targeted advertising, spread of fashions and fads, innovations, consumer interests etc.
 - Determine community evolution.

Our Approach

- Estimate two properties of communication flow: the **intent to communicate** and **associated delay** [De Choudhury '07].
 - Design features to characterize communication and *social context*.
 - Select a subset of optimal contextual features using a voting strategy from five different feature selection techniques.
 - Use features in a Support Vector Regression framework to predict the intent and the delay.

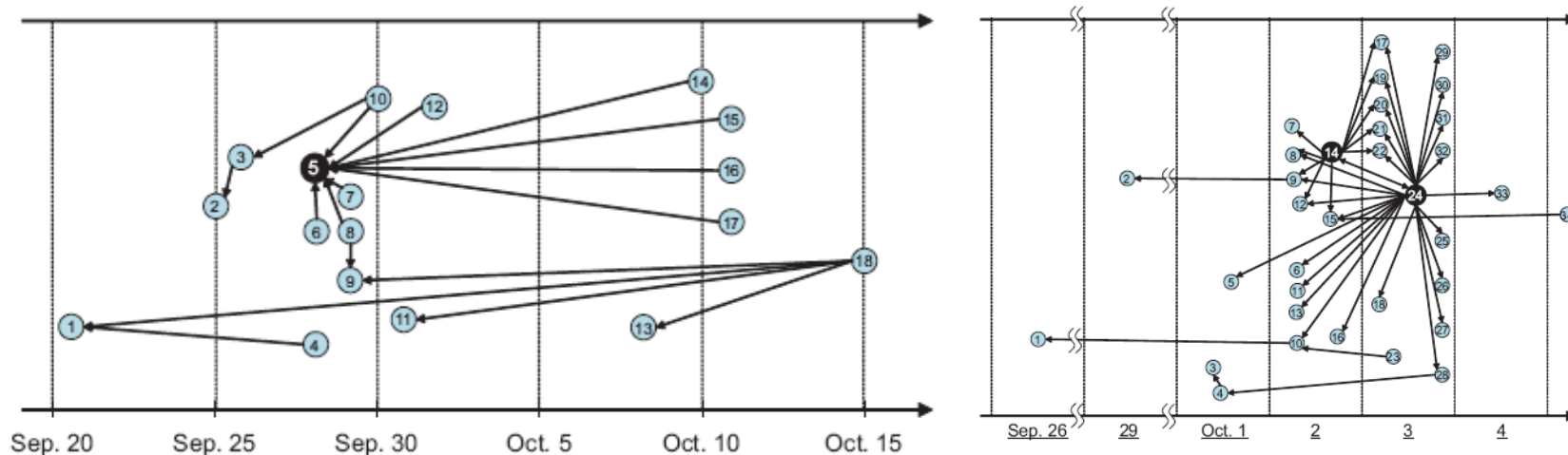


Error in Prediction of Intent to communicate

- Experimental results on MySpace dataset with effective prediction (error ~12-15%).

Related Work

- Work on information diffusion [Gruhl, Tomkins '04].
- Early adoption based flow model for recommendation systems [Song et al. '06].
- Work on information roles of people [Nakajima et al. '05, Song et al. '06].
- Limitations:
 - information flow estimated from indirect evidence.
 - context not been considered.
 - Roles of people with respect to long term habitual behavior not analyzed.



Agitators and Summarizers [Nakajima et al. '05]

Outline

Introduction / Related work

Problem Statement

Communication Context

Social Context

Prediction

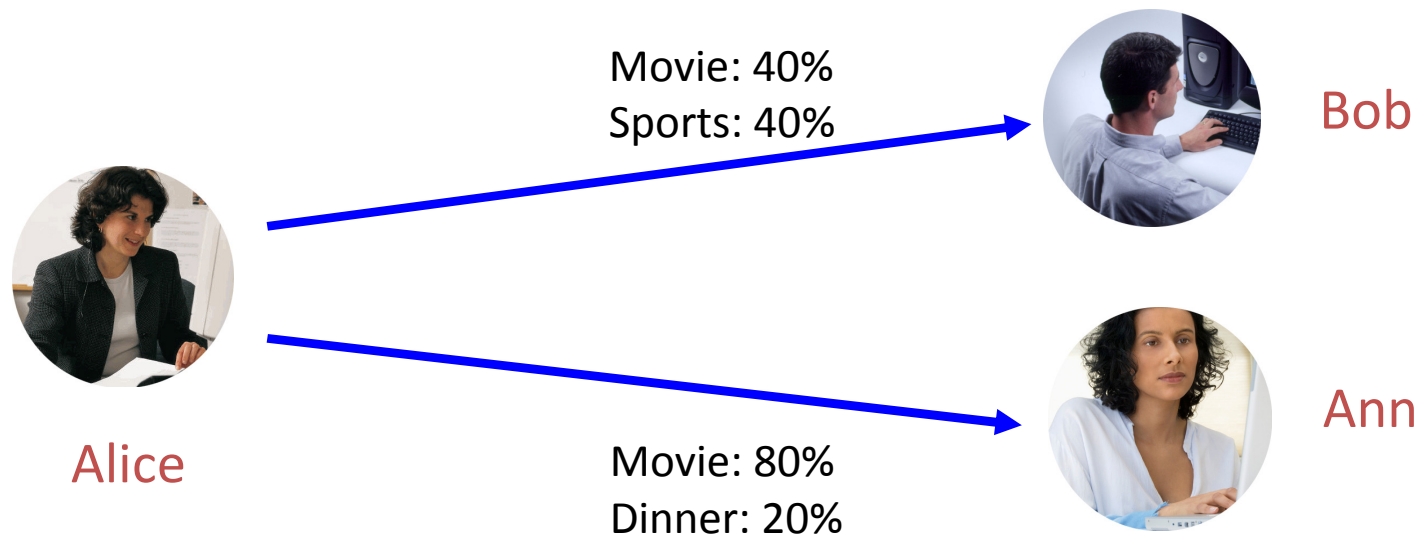
Experimental Results

Conclusions

- Two sub-problems:
 - Intent to communicate*
 - Communication Delay*
- Role of context

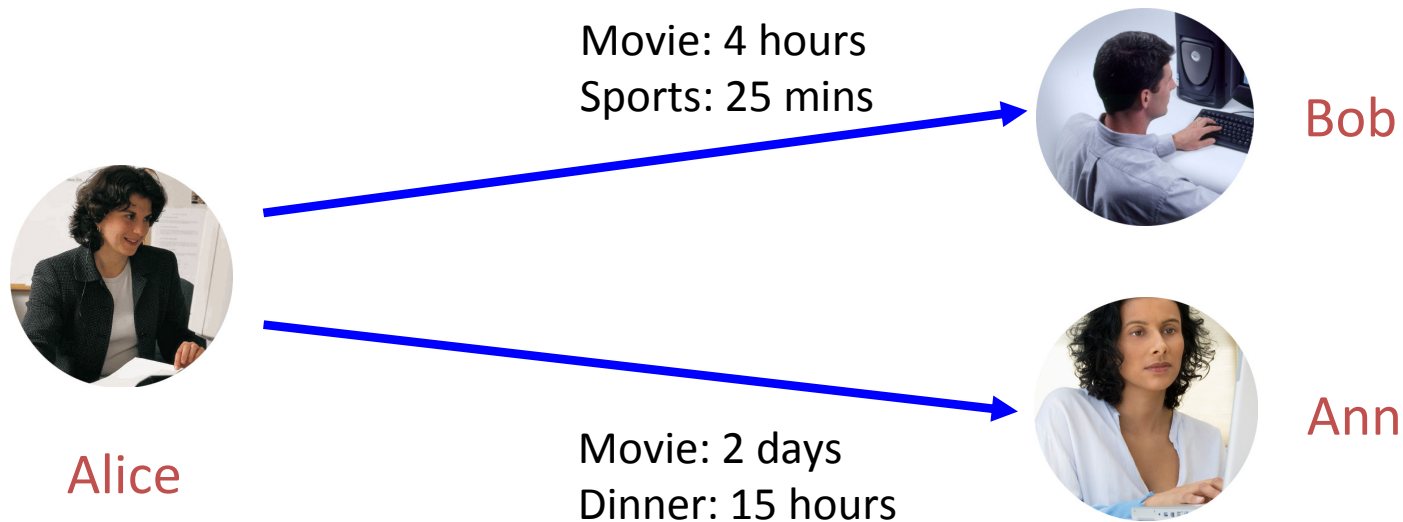
What is Intent to Communicate?

- The probability that a person will engage herself in some communication (given a particular topic and at a certain point of time) with another person.
 - It is contingent upon several factors or features defined by the communication context [De Choudhury et al. 07] and **social context**.



What is Delay in Propagation?

- The amount of time passed between the reception of a message (on a certain topic) and the corresponding response by a person.

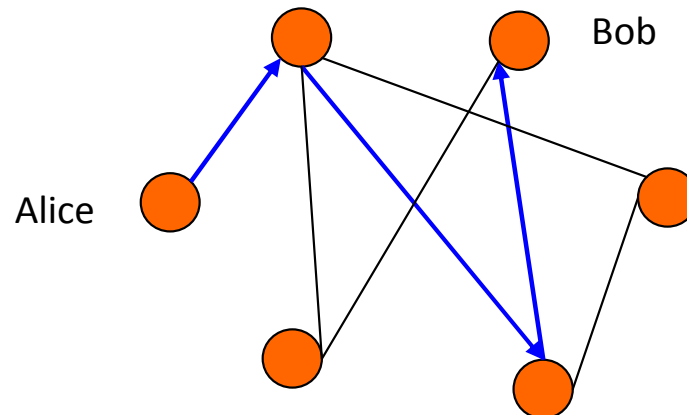


What is the role of context?

Conventional view of communication flow in the community

Gladwell's approach: information flows through specific "hot" points

Watt's approach: "hot" points are just coincidental; information flow depends on the vulnerability of people to communication



Communication flow depends upon: how Alice and Bob communicate (communication context), and what intrinsic roles Alice and Bob play in their communication (social context)

Outline

Introduction / Related work

Problem Statement

Communication Context

Social Context

Prediction

Experimental Results

Conclusions

- What is communication context?
- Overview of Neighborhood, Topic and Recipient context

Neighborhood



Topic



Recipient

Communication Context

- Communication context [De Choudhury, Sundaram et al. '07] is the set of attributes that affect communication between two individuals.
- Contextual attributes are dynamic [Dourish '02].
 - relationship between messages
 - past communication behavior of a person
 - response patterns from others

Features of Communication Context

- Neighborhood context
 - It refers to the effect of the user's social network on her communication.
 - Susceptibility
 - Backscatter
- Topic context
 - It refers to the effect of the semantics of a user's *past* communication on a topic on her future communication.
 - Message Coherence
 - Temporal Coherence
 - Topic Relevance
 - Topic Quantity
- Recipient context

Outline

Introduction / Related work

Problem Statement

Communication Context

Social Context

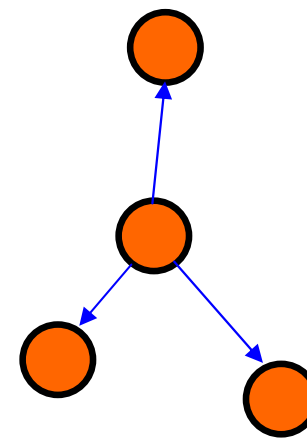
Prediction

Experimental Results

Conclusions



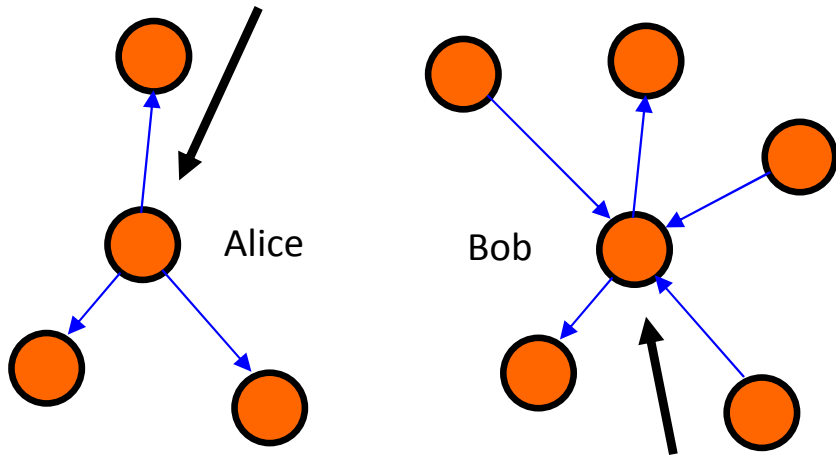
- What is social context?
- Features of social context



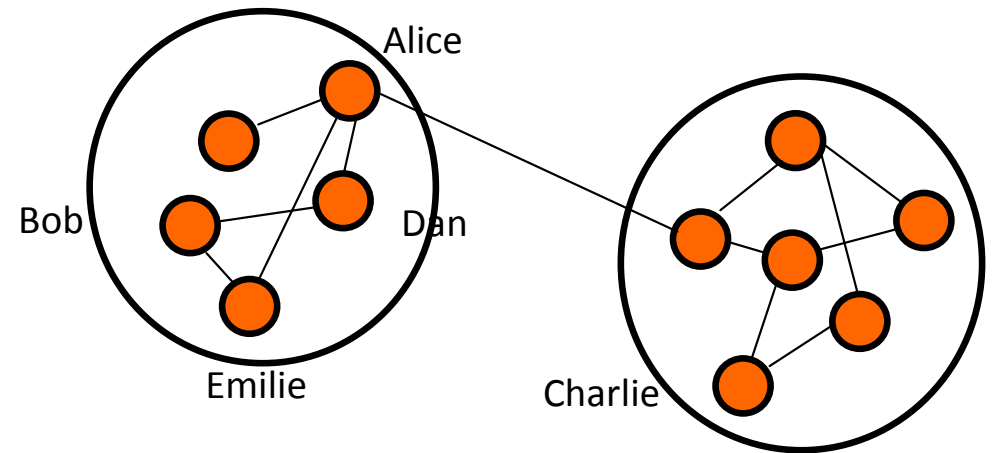
Social Context

- Social context:
 - set of contextual attributes that refer to who is communicating with whom and what is the strength of relationship shared between them.
- Two features: **information roles** and **strength of ties**.

Star network with only out-going communication links: high tendency to send messages.



Several in-coming links: receptor of messages.



Same clique – strong tie – higher chance of communication. Different cliques – weak tie – lower chance of communication.

Information Roles

- Information role is a contextual attribute acquired over time which impacts a person's communication behavior.
- Three different categories of roles of people:
 - *generators*, people who generate information by themselves or from other sources,
 - *mediators*, people who act as transmitters of information between people, and
 - *receptors*, people who mostly receive messages.

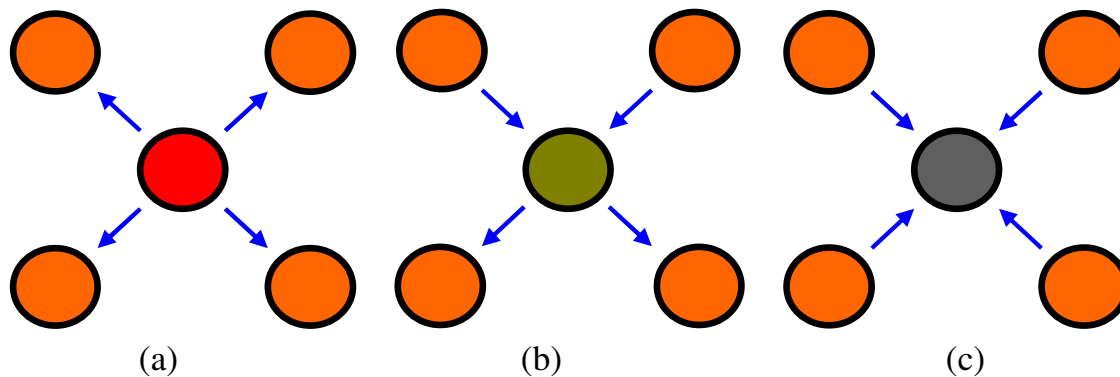
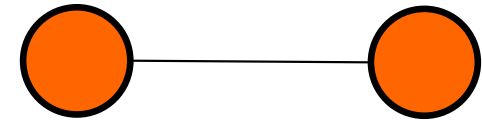


Figure : (a) Generators (**red**) (b) Mediators (**green**) (c) Receptors (**gray**).

Strength of ties

- The nature of relationship between two people affects communication [Granovetter '73].
- Strength of a tie between two people:
 - overlap of their friends' circles.
- The pattern of relationships between actors reveals the likelihood that individuals will be exposed to particular kinds of information.
 - Weak tie and flow of new information.
 - Strong tie and flow of existing information.

Weak tie



Topic: iPhone 3G [new information]

Strong tie



Topic: Joe's birthday party last week [existing information]

Outline

Introduction / Related work

Problem Statement

Communication Context

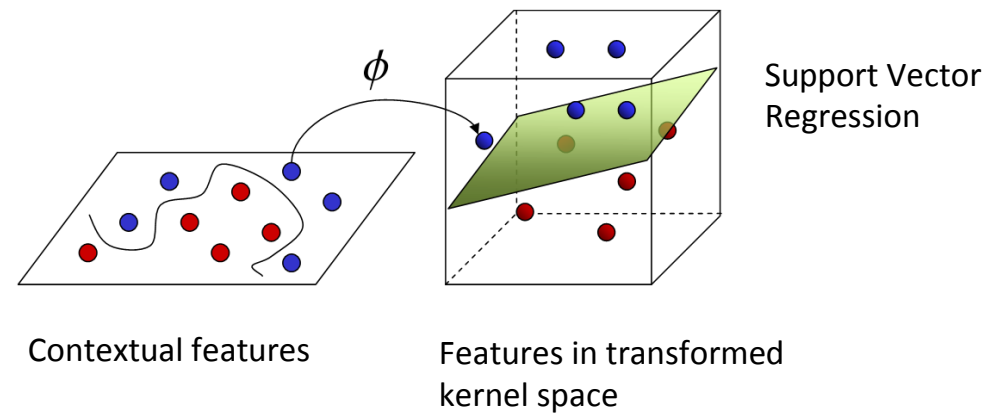
Social Context

Prediction

Experimental Results

Conclusions

- SVR approach
- Feature Selection



Prediction of Intent and Delay

- Support Vector Regression framework
 - Use features of communication and social context and actual frequency of communication over N weeks for training a Support Vector Machine regressor.
 - SVR further used to predict intent and delay at the $(N+1)^{\text{th}}$ week.
- Feature Selection
 - Need to eliminate redundant features dynamically.
 - Several static and dynamic feature selection techniques:
 - Correlation coefficient
 - Mutual information
 - Decision trees
 - Principal Component Analysis (PCA)
 - k NN based estimation
 - Use of a 'voting strategy' to determine the best k features at the beginning of each time slice.

Outline

Introduction / Related work

Problem Statement

Communication Context

Social Context

Prediction

Experimental Results

Conclusions

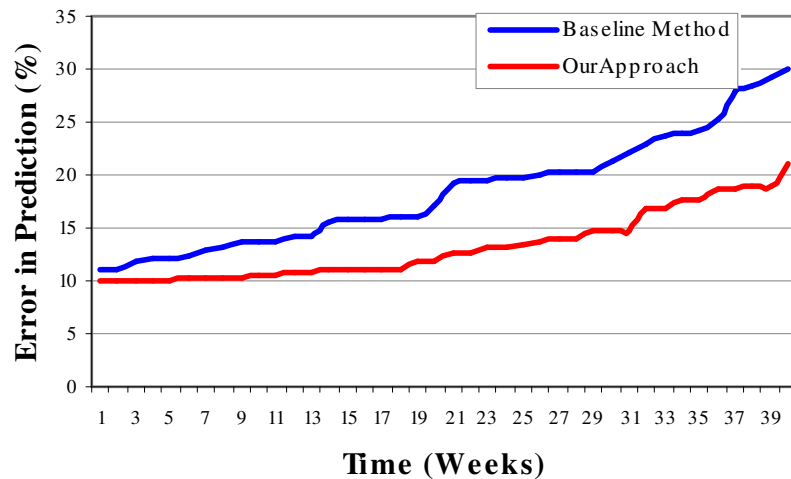


MySpace.com

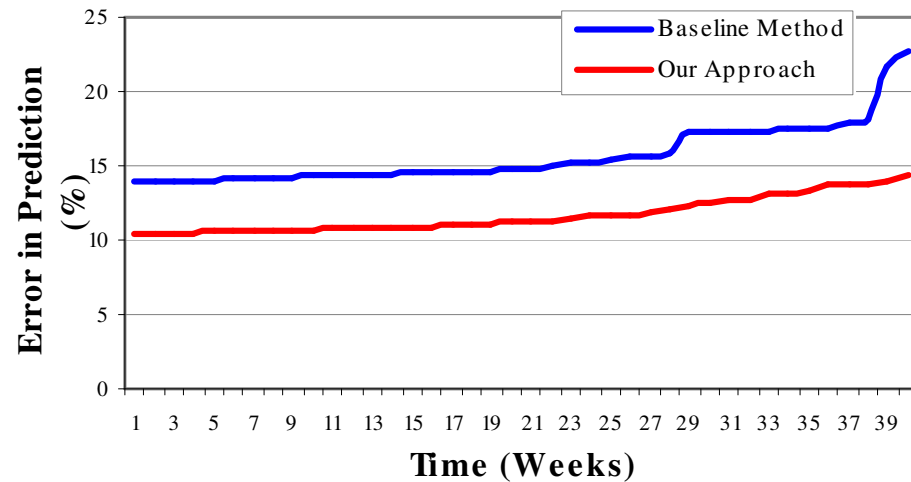
- Prediction of intent and delay
- Qualitative Evaluation

Prediction Results

Error in Prediction of Intent



Error in Prediction of Delay



- Data description:
 - 20,000 users from MySpace, 1,425,010 messages from September 2005 to April 2007.
- Baseline Method:
 - Consider all features of communication context without feature selection [De Choudhury et al. 2007].
- The training set of 70 weeks.
- Experiments over a period of 40 weeks, averaged over eight different contacts of a person Charlie and four topics.
 - Prediction error of 12-15 % compared to 25-30 % using the baseline method.

Qualitative Evaluation of Predicted Intent

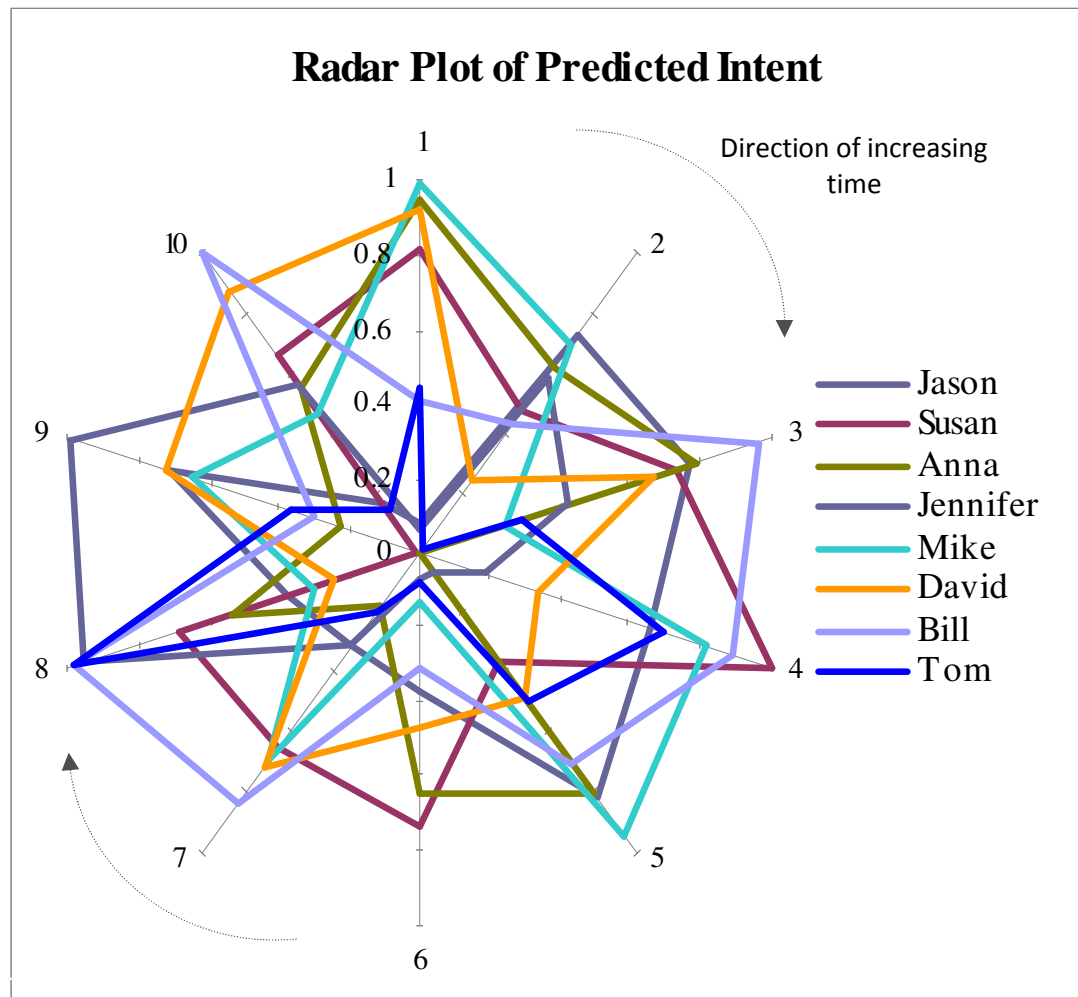


Figure : Dynamics of predicted intent across time and contacts for a user.

Qualitative Evaluation of Predicted Delay

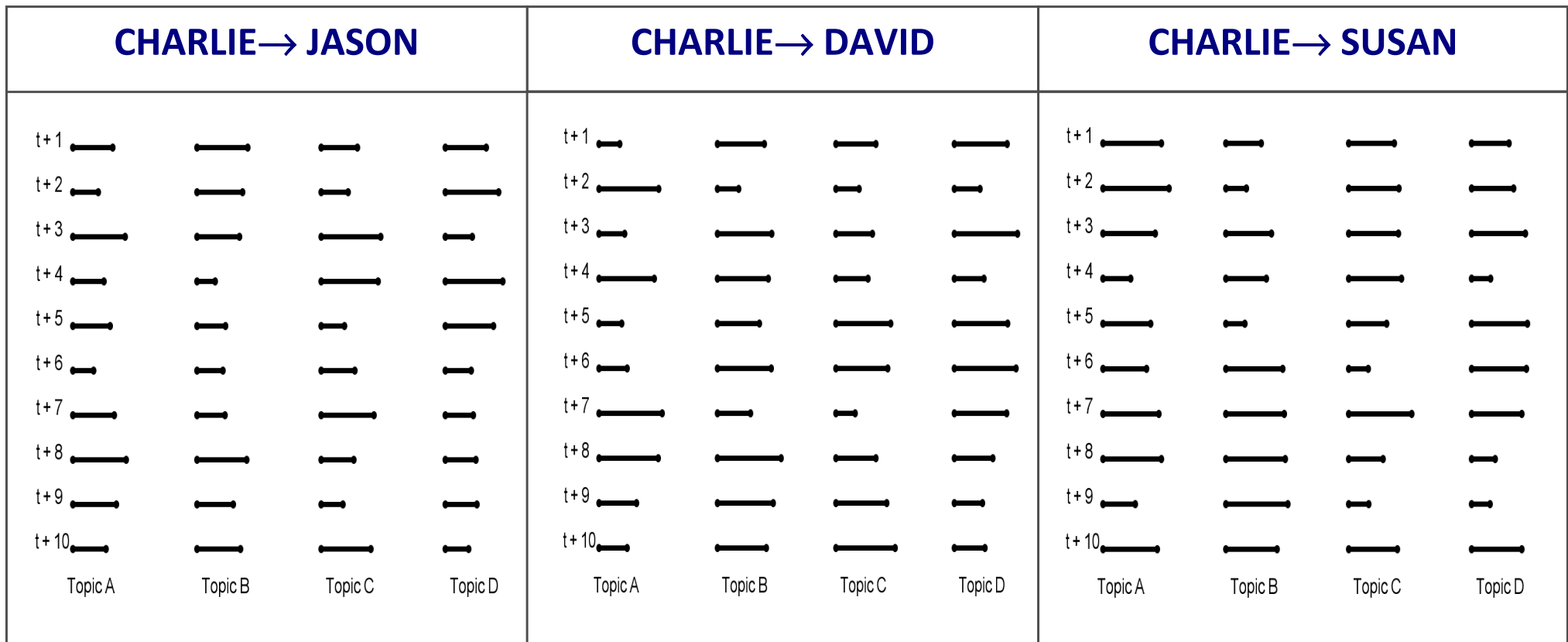


Figure : Dynamics in delay between Charlie and three of her contacts. The visualizations show dynamics for ten consecutive weeks (Y axis) and for four topics (X axis). The length of each horizontal line is proportional to the measure of delay.

Outline

Introduction / Related work

Problem Statement

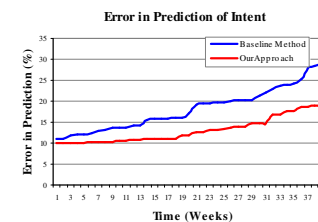
Communication Context

SVM Based prediction

MySpace dataset

Experimental Results

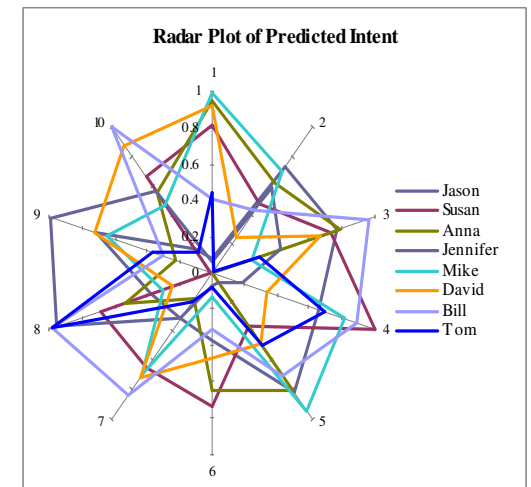
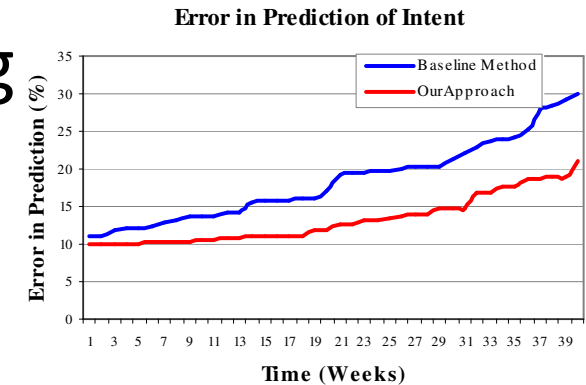
Conclusions



- Summary and Contributions
- Future Work

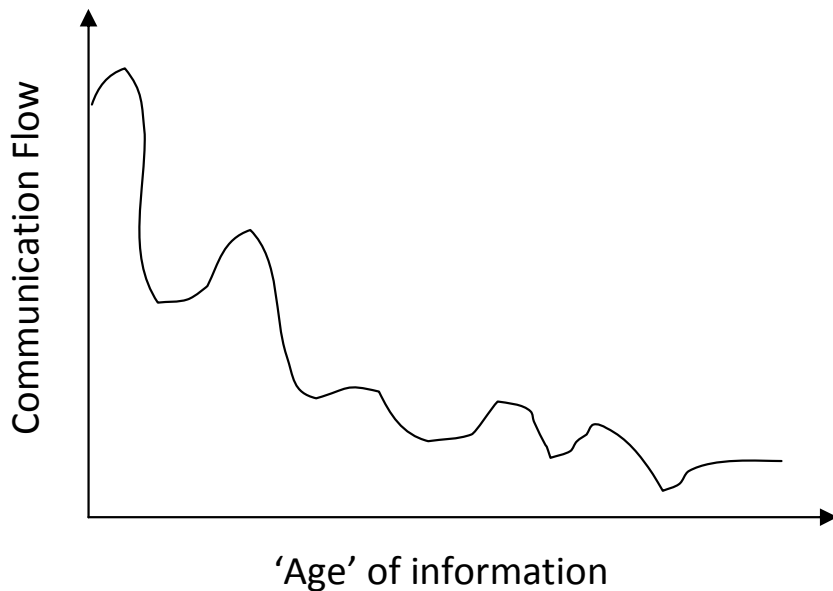
Summary and Conclusions

- Predict communication flow in large scale social networks by incorporating **social context**.
- Excellent results on a real world dataset MySpace.com
 - error ~12-15 % compared to 25-30 % using only communication context.
- **Conclusions**
 - Social context is key to predicting communication flow.
 - Intent to communicate appears to be a property of context; delay seems to be a property of habitual characteristics of people.



Future Work

- How does the 'age' and nature of information impact communication flow?
- Model of collective user behavior.
- Inherent property of a communication media that affects communication flow.



Emergent collective user behavior

Thanks!