

Research Summary:

Communication is central to the evolution of social systems. The advent of the Web 2.0 technology through a plethora of social media websites (e.g. Facebook, MySpace, Blogger, LiveJournal, Flickr, YouTube) has let users of the Web participate in a wide variety of online communication. Users can express their opinions on personal blogs as well as can share media objects and engage themselves in discussion. Right from shopping a new car to getting suggestions on investment or to search for the next holiday destination, people have started to rely heavily on other opinionated people in the blogosphere. Moreover, personal experiences as well as thoughts and opinions on external events also manifest themselves through variable degrees of online chatter in several people's blogs and social profiles.

Under all these affordances provided by the online social media, it is interesting to observe rich communication dynamics, interesting social roles of people as well as static or volatile community emergence and evolution. Such dynamics can lend interesting insights to understand user behavior, mine useful knowledge applicable in business domains as well as study and reflect on the growth and characteristics of online communities at micro and macro scales.

I have worked on three coupled directions regarding analysis of human communication in online social media – communication dynamics and individuals, that deals with emergent local properties; communication dynamics and online communities, that deals with emergent global properties; and third, communication dynamics and rich media, that deals with the emergent properties of the shared media elements centered around local and global properties.

Communication and Individuals:

Communication characterization within a social network is useful in suggesting alternative ranking strategies for 'impactful' individuals, groups and communities in social networks. For example, a news media organization or a corporation might be interested to know which are the sources of reliable information and who are the people capable of suggesting answers to specific questions. It could also help user behavior prediction underlying several sociological phenomena as well as support tools for marketing analysis and operations management that are required to cater to user needs in response to temporally changing external events and impulses.

The work in [9] deals with characterizing individuals in social networks via several information roles – roles emerging out of their communication activity, such as early responders and late trailers.

Papers [7,8,13] deal with representation of communication in social media with different types of contextual attributes – network topology based context, social context (e.g. strength

of ties), topical context (e.g. content of communication) and user identity context (e.g. who are the people involved in communication). Such contextual representation is then used to predict the flow of communication on a large dataset crawled from MySpace. We compute two attributes – probability of communication between a pair of users and the associated delay in communication. Our model yields excellent prediction results with error rates lower than ~15%. We observe that while the probability of communication is highly dependent on contextual factors, delay appears to be more or less habitual for the pair of users concerned.

Communication and Online Communities:

Analyzing communication in large scale online networks is of great value in opinion tracking in business intelligence – for example, corporations are interested in knowing how they and their products are being received at large to customers. They also want to understand shareholder sentiment, as well as to maximize exposure on a certain product. Hence they are interested in knowing which communities are discussing them and the dynamics of such interaction.

We attempt to address some of the challenges that come along in the evolution of communication-triggered social processes in [1,5,6,10,11,12]. Our goal is to focus on developing methods that can help us understand human-communication better on online social media. In [11], we study the phenomenon of emergence of 'social synchrony' in social actions on online social media, specifically Digg. User representations of communication, actions and hidden user states relating to emergent synchrony are developed based on the idea of dynamic Bayesian networks. States of the users are predicted and a cumulative aggregate of 'active' users in a set of evolved social networks are used to determine the presence or absence of a synchrony in the future.

In [5] we characterize communication dynamics in social networks at multiple scales – individual, group and community scales. We focus on technology communities in Engadget for the purpose. The dynamics of communication at different scales are then used to predict stock market movements of corresponding technology companies, with significantly low error rates of prediction. We analyze blog communication dynamics in [6] through individuals and their communication properties and use them to determine correlation with stock market movements of companies based on a Support Vector Regression prediction framework. Our experiments on a large dataset from the blog Engadget indicate satisfactory correlation.

In [1,10,12] we study different global dynamics of online communities. Our goal is to understand different significance measures of communities over time. Such measures can enable end users to choose between a set of thematically similar groups, as well as can help represent the overall network dynamics based on only a handful of prototypical

groups that captures its context and content of communication over time.

Communication and Rich Media:

Nevertheless, apart from impact of communication on the dynamics of the individuals' actions, roles and the community in general, social media sites are observed to generate rich dialogue of communication centered round shared media objects, e.g. YouTube, Flickr etc. The accompanying challenges that come along include the manner in which the evolving properties (e.g. popularity, themes) of media artifacts impact users, also the dynamics that affect the popularity of rich media and connection of media objects to diverse online communities.

The papers [2,4,14] deal with some of these challenges. In [4, 14] we propose a computational framework to predict the "interestingness" property of conversations associated with media objects; specifically over videos in the popular site YouTube. Interestingness is defined to depend on two properties – the conversational themes and the co-participation of individuals in communication. We have excellent results in favor of our algorithm on a dataset from YouTube.

Rich media communities often have several temporally changing themes and capturing these dynamics can help understand the on-going communication and interests of a community. In [3] we provide an algorithm to extract themes from image social media groups in Flickr. It uses visual, tag, user, and time features to determine themes. We have satisfactory qualitative and quantitative results in understanding the dynamically changing patterns in different Flickr groups.

However, analysis of the dynamic properties of media objects does not reveal their connection to a community in general. This is because social media sites are popular not just for the content or popularity, but also due to the accompanying social interaction. In image sharing sites such as Flickr, enthusiastic photographers are interested in receiving critical comments on their photos. Note that simply uploading an image onto Flickr does not ensure rich social interaction or *reachability* to other users for critical feedback. To address this issue, the paper [2] focuses on how media objects can be associated with communities by leveraging user communication. Specifically it develops a recommendation framework that can connect images on Flickr to appropriate groups. Groups are represented in a latent space and images characterized through their visual content, and historical data on user tagging and communication activity. Experiments yield satisfactory precision and recall measures, suggesting the recommendation framework is useful.

In the future I intend to investigate more on the connection between the visual properties of the media objects and the social aspects of online communities. I am also interested to

observe how traditional sociological models of social network growth and evolution based on a top-down approach apply to today's online media networks, which are primarily bottom-up in nature, governed by highly localized communication.

Publications:

- [1] Choudhury, M. D. (2009). *Modeling and Predicting Group Activity over Time in Online Social Media*. Accepted to appear in Proceedings of the Twentieth ACM Conference on Hypertext and Hypermedia (Torino, Italy, June 29 - July 1, 2009). HT '09.
- [2] Choudhury, M. D., Sundaram, H., Lin, Y., John, A., and Seligmann, D. D. *Connecting Content to Community in Social Media via Image Content, User Tags and User Communication*. Accepted to appear in Proceedings of the 2009 IEEE International Conference on Multimedia & Expo (Cancun, Mexico, June 28- July 3, 2009). ICME '09. (Acceptance rate: 23%)
- [3] Lin, Y-R., Sundaram, H., Choudhury, M. D., Kelliher, A. (2009). *Temporal Patterns in Social Media Streams: Theme Discovery and Evolution using Joint Analysis of Content and Context*. Accepted to appear in Proceedings of the 2009 IEEE International Conference on Multimedia & Expo (Cancun, Mexico, June 28- July 3, 2009). ICME '09 (*Invited Paper*).
- [4] Choudhury, M. D., Sundaram, H., John, A., and Seligmann, D. D. (2009). *What Makes Conversations Interesting? Themes, Participants and Consequences of Conversations in Online Social Media*. Accepted to appear in Proceedings of the Eighteenth International World Wide Web Conference (Madrid, Spain, April 20-24, 2009). WWW '09. (Acceptance rate: 11%)
- [5] Choudhury, M. D., Sundaram, H., John, A., and Seligmann, D. D. (2008). *Multi-scale Characterization of Social Network Dynamics in the Blogosphere*. In Proceedings of the Seventeenth ACM Conference on Information and Knowledge Management (Napa Valley, California, USA, October 26-30, 2008). CIKM '08. ACM, New York, NY, 1515-1516. (Acceptance rate: 16%)
- [6] Choudhury, M. D., Sundaram, H., John, A., and Seligmann, D. D. (2008). *Can Blog Communication Dynamics be correlated with Stock Market Activity?* In Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia (Pittsburgh, Pennsylvania, USA, June 19 - 21, 2008). HT '08. ACM, New York, NY, 55-60.
- [7] Choudhury, M. D., Sundaram, H., John, A., and Seligmann, D. D. (2008). *Dynamic Prediction of Communication Flow Using Social Context*. In Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia (Pittsburgh, Pennsylvania, USA, June 19 - 21, 2008). HT '08. ACM, New York, NY, 49-54.

- [8] Choudhury, M. D., Sundaram, H., John, A., and Seligmann, D. D. (2007). *Contextual Prediction of Communication Flow in Social Networks*. In Proceedings of the IEEE/WIC/ACM international Conference on Web intelligence (Silicon Valley, California, USA, November 02 - 05, 2007). Web Intelligence. IEEE Computer Society, Washington, DC, 57-65. (Acceptance rate: 17%)
- [9] Choudhury, M. D., Sundaram, H., John, A., and Seligmann, D. D. (2008). *Modeling Information Roles for Understanding Stock Market Dynamics*. Eighth Annual Grace Hopper Celebration of Women in Computing Conference - Poster Session (Keystone, Colorado, USA, October 1-4, 2008). GHC '08.
- [10] Choudhury, M. D., Sundaram, H., John, A., and Seligmann, D. D. (2009). *Which are the Representative Groups in a Community? Extracting and Characterizing Key Groups in Blogs*. Submitted.
- [11] Choudhury, M. D., Sundaram, H., John, A., and Seligmann, D. D. *Social Synchrony: Predicting User Actions over Time in Online Social Media*. In Preparation.
- [12] Choudhury, M. D., Sundaram, H., John, A., and Seligmann, D. D. *Extraction, Characterization and Utility of Prototypical Groups of Communication Sets in the Blogosphere*. Journal in Preparation.
- [13] Choudhury, M. D., Sundaram, H., John, A., and Seligmann, D. D. *Predicting Efficient Communication Paths in Online Social Networks*. Journal in Preparation.
- [14] Choudhury, M. D., Sundaram, H., John, A., and Seligmann, D. D. *Interestingness: A Dynamic Property of Conversations on Online Social Media*. Journal in Preparation.