What Makes Conversations Interesting? Themes, Participants and Consequences of Conversations in Online Social Media

Munmun De Choudhury Hari Sundaram Ajita John Dorée Duncan Seligmann
Arts Media & Engineering, Arizona State University Collaborative Applications Research, Avaya Labs
Email: {munmun.dechoudhury,hari.sundaram}@asu.edu, {ajita,doree}@avaya.com

ABSTRACT
Rich media social networks promote not only creation and consumption of media, but also communication about the posted media item. What causes a conversation to be interesting, that prompts a user to participate in the discussion on a posted video? We conjecture that people participate in conversations when they find the conversation theme interesting, see comments by people whom they are familiar with, or observe an engaging dialogue between two or more people (absorbing back and forth exchange of comments). Importantly, a conversation that is interesting must be consequential – i.e. it must impact the social network itself.

Our framework has three parts. First, we detect conversational themes using a mixture model approach. Second, we determine interestingness of participants and interestingness of conversations based on a random walk model. Third, we measure the consequence of a conversation by measuring how interestingness affects the following three variables – participation in related themes, participant cohesiveness and theme diffusion. We have conducted extensive experiments using a dataset from the popular video sharing site, YouTube. Our results show that our method of interestingness maximizes the mutual information, and is significantly better (twice as large) than three other baseline methods (number of comments, number of new participants and PageRank based assessment).

Categories and Subject Descriptors
H.1.2 [**Models and Principles**]: User/Machine Systems, I.2.6 [**Artificial Intelligence**]: Learning, J.4 **[Social and Behavioral Sciences]**: Sociology.

General Terms
Algorithms, Experimentation, Human Factors, Verification.

Keywords
Conversations, Interestingness, Social media, Themes, YouTube.

1. INTRODUCTION
Today, there is significant user participation on rich media social networking websites such as YouTube and Flickr. Users can create (e.g. upload photo on Flickr), and consume media (e.g. watch a video on YouTube). These websites also allow for significant communication between the users – such as comments by one user on a media uploaded by another. These comments reveal a rich dialogue structure (user A comments on the upload, user B comments on the upload, A comments in response to B's comment, B responds to A's comment etc.) between users, where the discussion is often about themes unrelated to the original video. Example of a conversation from YouTube [1] is shown in Figure 1. In this paper, the sequence of comments on a media object is referred to as a conversation. Note the theme of the conversation is latent and depends on the content of the conversation.

The fundamental idea explored in this paper is that analysis of communication activity is crucial to understanding repeated visits to a rich media social networking site. People return to a video post that they have already seen and post further comments (say in YouTube) in response to the communication activity, rather than to watch the video again. Thus it is the content of the communication activity itself that the people want to read (or see, if the response to a video post is another video, as is possible in the case of YouTube). Furthermore, these rich media sites have notification mechanisms that alert users of new comments on a video post / image upload promoting this communication activity.

Figure 1: Example of a conversation from YouTube. A conversation is associated with a unique media object comprising several temporally ordered comments on different latent themes from different authors (participants of the conversation).

We denote the communication property that causes people to further participate in a conversation as its "interestingness." While the meaning of the term "interestingness" is subjective, we decided to use it to express an intuitive property of the communication phenomena that we frequently observe on rich media networks. Our goal is to determine a real scalar value

WWW 2009, April 20–24, 2009, Madrid, Spain.
ACM 978-1-60558-487-4/09/04.

corresponding to each conversation in an objective manner that serves as a measure of interestingness. Modeling the user subjectivity is beyond the scope of our paper.

What causes a conversation to be interesting to prompt a user to participate? We conjecture that people will participate in conversations when (a) they find the conversation theme interesting (what the previous users are talking about) (b) see comments by people that are well known in the community, or people that they know directly comment (these people are interesting to the user) or (c) observe an engaging dialogue between two or more people (an absorbing back and forth between two people). Intuitively, interesting conversations have an engaging theme, with interesting people.

A conversation that is deemed interesting must be consequential – i.e. it must impact the social network itself. Intuitively, there should be three consequences (a) the people who find themselves in an interesting conversation, should tend to co-participate in future conversations (i.e. they will seek out other interesting people that they've engaged with) (b) people who participated in the current interesting conversation are likely to seek out other conversations with themes similar to the current conversation and finally (c) the conversation theme, if engaging, should slowly proliferate to other conversations.

There are several reasons why measuring interestingness of a conversation is of value. First, it can be used to rank and filter both blog posts and rich media, particularly when there are multiple sites on which the same media content is posted, guiding users to the most interesting conversation. For example, the same news story may be posted on several blogs, our measures can be used to identify those sites where the postings and commentary is of greatest interest. It can also be used to increase efficiency. Rich media sites, can manage resources based on changing interestingness measures (e.g. and cache those videos that are becoming more interesting), and optimize retrieval for the dominant themes of the conversations. Besides, differentiated advertising prices for ads placed alongside videos can be based on their associated conversational interestingness.

It is important to note that frequency based measures of a video (e.g. number of views, number of comments and number of times it has been marked by a user as a favorite) do not adequately capture interestingness because these measures are properties of the video (content, video quality), not the communication. Furthermore, the textual analyses of comments alone are not adequate to capture conversational interestingness because it does not consider the dialogue structure between users in the conversation.

1.1 Our Approach

There are two key contributions in this paper. We characterize conversational *themes* and communication properties of participants for determining the "interestingness" of online conversations (sections 3, 4). Second, we measure the consequence of conversational interestingness via a set of communication *consequences*, including activity, cohesiveness in communication and thematic interestingness (section 5).

There are three steps to our approach. First we detect conversational themes using a sophisticated mixture model approach. Second we determine interestingness of participants and interestingness of conversations based on a random walk model. We also propose a novel joint optimization framework of interestingness that incorporates temporal smoothness constraints to effectively compute interestingness. Third, we compute the consequence of a conversation deemed interesting by a mutual information based metric. We compute the mutual information between the interestingness with consequence-based measures: activity, cohesiveness and thematic interestingness.

To test our model, we have conducted extensive experiments using a dataset from the highly popular media sharing site, YouTube [1]. We observe from the dynamics of conversational themes, interestingness of participants and of conversations that (a) conversational themes associated with significant external happenings become "hot", (b) participants become interesting irrespective of the number of their comments during times of significant external events, and (c) the mean interestingness of conversations increase due to the chatter about important external events. During evaluation, we observe that our method of interestingness maximizes the mutual information by explaining the consequences significantly better than three other baseline methods (our method 0.83, baselines 0.41).

1.2 Related Work

Now we discuss prior work from the following three facets useful in solving this problem.

Analysis of Media Properties: There has been considerable work conducted in analyzing dynamic media properties, (e.g. associated tags on a media object). In [4] Dubinko et al. visualized the evolution of tags within Flickr and presented a novel approach based on a characterization of the most salient tags associated with a sliding interval of time. Kennedy et al. in [9] leveraged the community-contributed collections of rich media (Flickr) to automatically generate representative views of landmarks. Nether work captures the dynamics of the associated conversations on a media object.

Theme Extraction: There has been considerable work in detecting themes or topics from dynamic web collections [10,13,12,16]. In [12] the authors study the problem of discovering and summarizing evolutionary theme patterns in a dynamic text stream. The authors modify temporal theme extraction in [13] by regularizing their theme model with timestamp and location information. In other work, authors in [16] propose a dynamic probability model, which can predict the tendency of topic discussions on online social networks. In prior work, the relationship of theme extraction with the co-participation behavior of the authors of comments (participants) has not been analyzed.

Social Media Communication Analysis: There has been considerable work on analyzing discussions or comments in blogs [5,8,15] as well as utilizing such communication for prediction of its consequences like user behavior, sales, stock market activity etc [2,3,6,11]. In [3], we analyzed the communication dynamics (of conversations) in a technology blog and used it to predict stock market movement. However, in prior work, the relationship or impact of a certain conversation property with respect to other attributes of the media object has not been considered. In this work, we characterize the consequences of conversations based on the impact of the themes and the communication properties of the participants.

The rest of the paper is organized as follows. We present our problem formulation in section 2. In sections 3, 4 and 5 we describe our computational framework involving detection of conversation themes, determining interestingness of participants and of conversations. Section 6 discusses the experiments using the YouTube dataset. Section 7 discusses the conclusions.

2. PROBLEM FORMULATION

In this section, we discuss our problem formulation – definitions, data model and problem statement and the key challenges.

2.1 Definitions

We now define the major concepts involved in this paper.

Conversation: We define a conversation in online social media (e.g., an image, a video or a blog post) as a temporally ordered sequence of comments posted by individuals whom we call "participants". In this paper, the content of the conversations are represented as a stemmed and stop-word eliminated bag-of-words.

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

Interestingness of Participants: Interestingness of a participant is a property of her communication activity over different conversations. We propose that an interesting participant can often be characterized by (a) several other participants writing comments after her, (b) participation in a conversation involving other interesting participants, and (c) active participation in "hot" conversational themes.

Interestingness of Conversations: We now define "interestingness" as a dynamic communication property of conversations which is represented as a real non-negative scalar dependent on (a) the evolutionary conversational themes at a particular point of time, and (b) the communication properties of its participants. It is important to note here that "interestingness" of a conversation is necessarily subjective and often depends upon context of the participant. We acknowledge that alternate definitions of interestingness are also possible.

Conversations used in this paper are the temporal sequence of comments associated with media elements (videos) in the highly popular media sharing site YouTube. However our model can be generalized to any domain with observable threaded communication. Now we formalize our problem based on the following data model.

2.2 Data Model

Our data model comprises the tuple $\{C, P\}$ having the following two inter-related entities: a set of conversations, C on shared media elements; and a set of participants P in these conversations. Each conversation is represented with a set of comments, such that each comment that belongs to a conversation is associated with a unique participant, a timestamp and some textual content (bag-of-words).

We now discuss the notations. We assume that there are N participants, M conversations, K conversation themes and Q time slices. Using the relationship between the entities in the tuple $\{C,P\}$ from the above data model, we construct the following matrices for every time slice q, $1 \le q \le Q$:

a. $\mathbf{P_F}^{(q)} \in \Re^{N \times N}$: Participant-follower matrix, where $\mathbf{P_F}^{(q)}(i,j)$ is the probability that at time slice q, participant j comments following participant i on the conversations in which i had commented at any time slice from 1 to (q-1).

b. $\mathbf{P_L}^{(q)} \in \Re^{N \times N}$: Participant-leader matrix, where $\mathbf{P_L}^{(q)}(i,j)$ is the probability that in time slice q, participant i comments following participant j on the conversations in which j had commented in any time slice from 1 to (q-1). Note, both $\mathbf{P_F}^{(q)}$ and $\mathbf{P_L}^{(q)}$ are asymmetric, since communication between participants is directional.

c. $\mathbf{P_C}^{(q)} \in \Re^{N \times M}$: Participant-conversation matrix, where $\mathbf{P_C}^{(q)}(i,j)$ is the probability that participant i comments on conversation j in time slice q.

d. $\mathbf{C_T}^{(q)} \in \Re^{M \times K}$: Conversation-theme matrix, where $\mathbf{C_T}^{(q)}(i,j)$ is the probability that conversation i belongs to theme j in time slice q.

e. $\mathbf{T_S}^{(q)} \in \Re^{K \times 1}$: Theme-strength vector, where $\mathbf{T_S}^{(q)}(i)$ is the strength of theme i in time slice q. Note, $\mathbf{T_S}^{(q)}$ is simply the normalized column sum of $\mathbf{C_T}^{(q)}$.

f. $\mathbf{P_T}^{(q)} \in \Re^{N \times K}$: Participant-theme matrix, where $\mathbf{P_T}^{(q)}(i,j)$ is the probability that participant i communicates on theme j in time slice q. Note, $\mathbf{P_T}^{(q)} = \mathbf{P_C}^{(q)}.\mathbf{C_T}^{(q)}$.

g. $\mathbf{I_P}^{(q)} \in \Re^{N \times 1}$: Interestingness of participants vector, where $\mathbf{I_P}^{(q)}(i)$ is the interestingness of participant i in time slice q.

h. $\mathbf{I_C}^{(q)} \in \Re^{M \times 1}$: Interestingness of conversations vector, where $\mathbf{I_C}^{(q)}(i)$ is the interestingness of conversation i in time slice q.

For simplicity of notation, we denote the i^{th} row of the above 2-dimensional matrices as $\mathbf{X}(i,:)$.

2.3 Problem Statement

Now we formally present our problem statement: given a dataset $\{C,P\}$ and associated meta-data, we intend to determine the interestingness of the conversations in C, defined as $\mathbf{I_C}^{(q)}$ (a non-negative scalar measure for a conversation) for every time slice q, $1 \le q \le Q$. Determining interestingness of conversations involves two key challenges:

a. How to extract the evolutionary conversational themes?

b. How to model the communication properties of the participants through their interestingness?

Further in order to justify interestingness of conversations, we need to address the following challenge: what are the consequences of an interesting conversation?

In the following three sections 3, 4 and 5, we discuss how we address these three challenges through: (a) detecting conversational themes based on a mixture model that incorporates regularization with time indicator, regularization for temporal smoothness and for co-participation; (b) modeling interestingness of participants; and of interestingness of conversations; and using a novel joint optimization framework of interestingness that incorporates temporal smoothness constraints and (c) justifying interestingness by capturing its future consequences.

3. CONVERSATIONAL THEMES

In this section, we discuss the method of detecting conversational themes. We elaborate on our theme model in the following two sub-sections – first a sophisticated mixture model for theme detection incorporating time indicator based, temporal and co-participation based regularization is presented. Second, we discuss parameter estimation of this theme model.

3.1 Chunk-based Mixture Model of Themes

Conversations are dynamically growing collections of comments from different participants. Hence, static keyword or tag based assignment of themes to conversations independent of time is not useful. Our model of detecting themes is therefore based on segmentation of conversations into 'chunks' per time slice. A chunk is a representation of a conversation at a particular time slice and it comprises a (stemmed and stop-word eliminated) set of comments (bag-of-words) whose posting timestamps lie within the same time slice. Our goal is to associate each chunk (and hence the conversation at that time slice) with a theme distribution. We develop a sophisticated multinomial mixture model representation of chunks over different themes (a modified pLSA [7]) where the theme distributions are (a) regularized with time indicator, (b) smoothed across consecutive time slices, and (c) take into account the prior knowledge of co-participation of individuals in the associated conversations.

Let us assume that a conversation c_i is segmented into Q non-overlapping chunks (or bag-of-words) corresponding to the Q different time slices. Let us represent the chunk corresponding to the i^{th} conversation at time slice q ($1 \le q \le Q$) as $\lambda_{i,q}$. We further assume that the words in $\lambda_{i,q}$ are generated from K multinomial theme models $\theta_1, \theta_2, \ldots, \theta_K$ whose distributions are hidden to us. Our goal is to determine the log likelihood that can represent our data, incorporating the three regularization techniques mentioned above. Thereafter we can maximize the log likelihood to compute the parameters of the K theme models.

However, before we estimate the parameter of the theme models, we refine our framework by regularizing the themes temporally as well as due to co-participation of participants. This is discussed in the following two sub-sections.

3.1.1 Temporal Regularization

We incorporate temporal characterization of themes in our theme model [13]. We conjecture that a word in the chunk can be attributed either to the textual context of the chunk $\lambda_{i,q}$, or the time slice q – for example, certain words can be highly popular on certain time slices due to related external events. Hence the theme associated with words in a chunk $\lambda_{i,q}$ needs to be regularized with respect to the time slice q. We represent the chunk $\lambda_{i,q}$ at time slice q with the probabilistic mixture model:

$$p(w:\lambda_{i,q},q)=\sum_{j=1}^{K} p\left(w,\theta_j \mid \lambda_{i,q},q\right), \tag{1}$$

where w is a word in the chunk $\lambda_{i,q}$ and θ_j is the j^{th} theme. The joint probability on the right hand side can be decomposed as:

$$\begin{aligned} p\left(w,\theta_j \mid \lambda_{i,q},q\right) &= p\left(w \mid \theta_j\right).p\left(\theta_j \mid \lambda_{i,q},q\right) \\ &= p\left(w \mid \theta_j\right).\left(\left(1-\gamma_q\right).p\left(\theta_j \mid \lambda_{i,q}\right)+\gamma_q.p\left(\theta_j \mid q\right)\right), \end{aligned} \tag{2}$$

where γ_q is a parameter that regulates the probability of a theme θ_j given the chunk $\lambda_{i,q}$ and the probability of a theme θ_j given the time slice q. Note that since a conversation can alternatively be represented as a set of chunks, the collection of all chunks over all conversations is simply the set of conversations C. Hence the log likelihood of the entire collection of chunks is equivalent to the likelihood of the M conversations in C, given the theme model. Weighting the log likelihood of the model parameters with the occurrence of different words in a chunk, we get the following equation:

$$\begin{aligned} L(C) &= \log p(C) \\ &= \sum_{\lambda_{i,q}\in C}\sum_{w\in\lambda_{i,q}} n\left(w,\lambda_{i,q}\right).\log\sum_{j=1}^{K} p\left(w,\theta_j \mid \lambda_{i,q},q\right), \end{aligned} \tag{3}$$

where $n(w, \lambda_{i,q})$ is the count of the word w in the chunk $\lambda_{i,q}$ and $p(w, \theta_j | \lambda_{i,q}, q)$ is given by equation (2).

However, the theme distributions of two chunks of a conversation across two consecutive time slices should not too divergent from each other. That is, they need to be temporally smooth. For a particular topic θ_j this smoothness is thus based on minimization of the following L^2 distance between its probabilities across every two consecutive time slices:

$$d_T(j)=\sum_{q=2}^{Q}\left(p\left(\theta_j \mid q\right)-p\left(\theta_j \mid q-1\right)\right)^2. \tag{4}$$

Incorporating this distance in equation (3) we get a new log likelihood function which smoothes all the K theme distributions across consecutive time slices:

$$\begin{aligned} &L_1(C) \\ &= \sum_{\lambda_{i,q}\in C}\sum_{w\in\lambda_{i,q}} n\left(w,\lambda_{i,q}\right).\log\sum_{j=1}^{K}\left(p\left(w,\theta_j \mid \lambda_{i,q},q\right)+\exp\left(-d_T(j)\right)\right). \end{aligned} \tag{5}$$

Now we discuss how this theme model is further regularized to incorporate prior knowledge about co-participation of individuals in the conversations.

3.1.2 Co-participation based Regularization

Our intuition behind this regularization is based on the idea that if several participants comment on a pair of chunks, then their theme distributions are likely to be closer to each other.

To recall, chunks being representations of conversations at a particular time slice, we therefore define a participant co-occurrence graph $G(C,E)$ where each vertex in C is a conversation c_i and an undirected edge $e_{i,m}$ exists between two conversations c_i and c_m if they share at least one common participant. The edges are also associated with weights $\omega_{i,m}$ which define the fraction of common participants between two conversations. We incorporate participant-based regularization based on this graph by minimizing the distance between the edge weights of two adjacent conversations with respect to their corresponding theme distributions.

The following regularization function ensures that the theme distribution functions of conversations are very close to each other if the edge between them in the participant co-occurrence graph G has a high weight:

$$R(C)=\sum_{c_i,c_m\in C}\sum_{j=1}^{K}\left(\omega_{i,m}-\left(1-\left(f\left(\theta_j \mid c_i\right)-f\left(\theta_j \mid c_m\right)\right)^2\right)\right)^2, \tag{6}$$

where $f(\theta_j|c_i)$ is defined as a function of the theme θ_j given the conversation c_i and the L^2 distance between $f(\theta_j|c_i)$ and $f(\theta_j|c_m)$ ensures that the theme distributions of adjacent conversations are similar. Since a conversation is associated with multiple chunks, thus $f(\theta_j|c_i)$ is given as in [14]:

$$f\left(\theta_j \mid c_i\right)=p\left(\theta_j \mid c_i\right)=\sum_{\lambda_{i,q}\in c_i} p\left(\theta_j \mid \lambda_{i,q}\right).p\left(\lambda_{i,q} \mid c_i\right). \tag{7}$$

Now, using equations (5) and (6), we define the final combined optimization function which minimizes the negative of the log likelihood and also minimizes the distance between theme

distributions with respect to the edge weights in the participant co-occurrence graph:

$$O(C) = -(1-\varsigma).L_1(C) + \varsigma.R(C), \quad (8)$$

where the parameter ς controls the balance between the likelihood using the multinomial theme model and the smoothness of theme distributions over the participant graph. It is easy to note that when ς=0, then the objective function is the temporally regularized log likelihood as in equation (5). When ς=1, then the objective function yields themes which are smoothed over the participant co-occurrence graph. Minimizing $O(C)$ for $0 \leq \varsigma \leq 1$ would give us the theme models that best fit the collection.

3.2 Parameter Estimation

Now we discuss how we can learn the hidden parameters of the theme model in equation (8). Note, the use of the more common technique of parameter estimation with the EM algorithm in our case involves multiple computationally intensive iterations due to the existence of the regularization function in equation (8). Hence we use a different technique of parameter estimation based on the Generalized Expectation Maximization algorithm (GEM [14]). The update equations for the E and M steps in estimation of the theme model parameters are illustrated in the Appendix (section 9). With the learnt parameters of the theme models, we can now compute the probability that a chunk $\lambda_{i,q}$ belongs to a theme θ_j:

$$\begin{aligned} & p(\theta_j \mid \lambda_{i,q}, q) \\ &= \sum_w p(\theta_j \mid w)\Big((1-\gamma_q).p(w \mid \lambda_{i,q}) + \gamma_q.p(w \mid q)\Big) \\ &= \sum_w \Big(p(w \mid \theta_j).p(\theta_j)/\,p(w)\Big).\Big((1-\gamma_q).p(w \mid \lambda_{i,q}) + \gamma_q.p(w \mid q)\Big). \end{aligned} \quad (9)$$

All the parameters on the right hand side are known from parameter estimation. A chunk $\lambda_{i,q}$ being the representation of a conversation c_i at a time slice q, the above equation would give us the conversation-theme matrix $\mathbf{C_T}$ at every time slice q, $1 \leq q \leq Q$. Now, we discuss how the evolutionary conversational themes can be used to determine interestingness measures of participants and conversations.

4. INTERESTINGNESS

In this section we describe our interestingness models and then discuss a method that jointly optimizes the two types of interestingness incorporating temporal smoothness.

4.1 Interestingness of Participants

We pose the problem of determining the interestingness of a participant at a certain time slice as a simple one-dimensional random walk model where she communicates either based on her past history of communication behavior in the previous time slice, or relies on her independent desire of preference over different themes (random jump). We describe these two states of the random walk through a set of variables as follows.

We conjecture that the state signifying the past history of communication behavior of a participant i at a certain time slice q, denoted as $\mathbf{A}^{(q-1)}$ comprises the variables: (a) whether she was interesting in the previous time slice, $\mathbf{I_P}^{(q-1)}(i)$, (b) whether her comments in the past impacted other participants to communicate and their interestingness measures, $\mathbf{P_F}^{(q-1)}(i,:).\mathbf{I_P}^{(q-1)}$[1], (c) whether she followed several interesting people in conversations at the previous time slice $q-1$, $\mathbf{P_L}^{(q-1)}(i,:).\mathbf{I_P}^{(q-1)}$, and (d) whether the conversations in which she participated became interesting in the previous time slice $q-1$, $\mathbf{P_C}^{(q-1)}(i;:).\mathbf{I_C}^{(q-1)}$. The independent desire of a participant i to communicate is dependent on her theme distribution and the strength of the themes at the previous time slice $q-1$: $\mathbf{P_T}^{(q-1)}(i,:).\mathbf{T_S}^{(q-1)}$. The relationships between all these different variables involving the two states are shown in Figure 2.

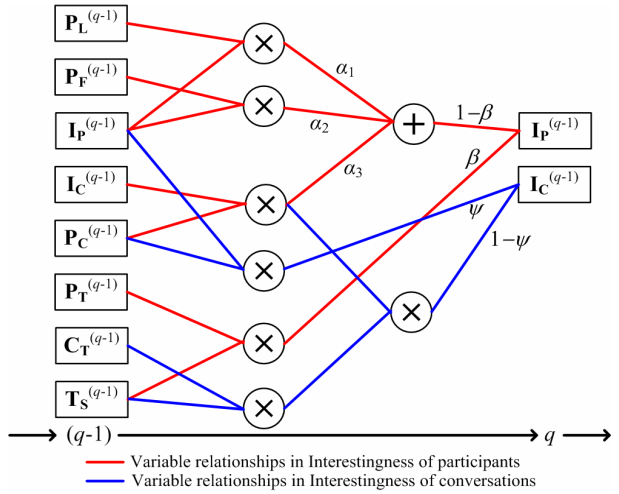

Figure 2: Timing diagrams of the random walk models for computing interestingness of participants ($\mathbf{I_P}^{(q-1)}$) and of conversations ($\mathbf{I_C}^{(q-1)}$). The relationships between different variables affecting the two kinds of interestingness are shown.

Thus the recurrence relation for the random walk model to determine the interestingness of all participants at time slice q is given as:

$$\mathbf{I_p}^{(q)} = (1-\beta).\mathbf{A}^{(q-1)} + \beta.\left(\mathbf{P_T}^{(q-1)}.\mathbf{T_S}^{(q-1)}\right),$$

$$\text{where } \mathbf{A}^{(q-1)} = \alpha_1.\mathbf{P_L}^{(q-1)}.\mathbf{I_P}^{(q-1)} + \alpha_2.\mathbf{P_F}^{(q-1)}.\mathbf{I_P}^{(q-1)} + \alpha_3.\mathbf{P_C}^{(q-1)}.\mathbf{I_C}^{(q-1)}. \quad (10)$$

Here α_1, α_2 and α_3 are weights that determine mutual relationship between the variables of the past history of communication state $\mathbf{A}^{(q-1)}$, and β the transition parameter of the random walk that balances the impact of past history and the random jump state involving participant's independent desire to communicate. In this paper, β is empirically set to be 0.5.

4.2 Interestingness of Conversations

Similar to interestingness of participants, we pose the problem of determining the interestingness of a conversation as a random walk (Figure 2) where a conversation can become interesting based on two states: the first state is when participants make the conversation interesting, and the second state is when themes make a conversation interesting (random jump). Hence to determine the interestingness of a conversation i at time slice q, we conjecture that it depends on whether the participants in conversation i became interesting at $q-1$, given as, $\mathbf{P_C}^{(q-1)}(i,:)^t.\mathbf{I_P}^{(q-1)}$, or whether the conversations belonging to the strong themes in $q-1$ became interesting, which is given as,

[1] To recall, $\mathbf{X}(i,:)$ is the i^{th} row of the 2-dimensional matrix $\mathbf{X}$.

$diag(\mathbf{C_T}^{(q-1)}(i,:).\mathbf{T_S}^{(q-1)}).\mathbf{I_C}^{(q-1)}$. Thus the recurrence relation of interestingness of all conversations at time slice q is:

$$\mathbf{I_C}^{(q)} = \psi.\left(\mathbf{P_C}^{(q-1)}\right)^t.\mathbf{I_P}^{(q-1)} + (1-\psi).diag\left(\mathbf{C_T}^{(q-1)}.\mathbf{T_S}^{(q-1)}\right).\mathbf{I_C}^{(q-1)}, \quad (11)$$

where ψ is the transition parameter of the random walk that balances the impact of interestingness due to participants and due to themes. Clearly, when ψ=1, the interestingness of conversation depends solely on the interestingness of the participants at $q-1$; and when ψ=0, the interestingness depends on the theme strengths in the previous time slice $q-1$.

4.3 Joint Optimization of Interestingness

We observe that the measures of interestingness of participants and of conversations described in sections 4.1 and 4.2 involve several free (unknown) parameters. In order to determine optimal values of interestingness, we need to learn the weights α_1, α_2 and α_3 in equation (10) and the transition probability ψ for the conversations in equation (11). Moreover, the optimal measures of interestingness should ensure that the variations in their values are smooth over time. Hence we present a novel joint optimization framework, which maximizes the two interestingness measures for optimal (α_1, α_2, α_3, ψ) and also incorporates temporal smoothness.

The joint optimization framework is based on the idea that the optimal parameters in the two interestingness equations are those which maximize the interestingness of participants and of conversations jointly. Let us denote the set of the parameters to be optimized as the vector, $\mathbf{X} = [\alpha_1, \alpha_2, \alpha_3, \psi]$. We can therefore represent $\mathbf{I_P}$ and $\mathbf{I_C}$ as functions of $\mathbf{X}$. We define the following objective function $g(\mathbf{X})$ to estimate $\mathbf{X}$ by maximizing $g(\mathbf{X})$:

$$g(\mathbf{X}) = \rho.\left\|\mathbf{I_P}(\mathbf{X})\right\|^2 + (1-\rho).\left\|\mathbf{I_C}(\mathbf{X})\right\|^2,$$
$$\text{s.t. } 0 \le \psi \le 1, \alpha_1, \alpha_2, \alpha_3 \ge 0, I_P \ge 0,\ I_C \ge 0, \alpha_1 + \alpha_2 + \alpha_3 = 1. \quad (12)$$

In the above function, ρ is an empirically set parameter to balance the impact of each interestingness measure in the joint optimization. Now to incorporate temporal smoothness of interestingness in the above objective function, we define a L^2 norm distance between the two interestingness measures across all consecutive time slices q and $q-1$:

$$d_P = \sum_{q=2}^{Q}\left(\left\|\mathbf{I_P}^{(q-1)}(\mathbf{X})\right\|^2 - \left\|\mathbf{I_P}^{(q-1)}(\mathbf{X})\right\|^2\right),$$
$$d_C = \sum_{q=2}^{Q}\left(\left\|\mathbf{I_C}^{(q-1)}(\mathbf{X})\right\|^2 - \left\|\mathbf{I_C}^{(q-1)}(\mathbf{X})\right\|^2\right). \quad (13)$$

We need to minimize these two distance functions to incorporate temporal smoothness. Hence we modify our objective function,

$$g_1(\mathbf{X}) = \rho.\left\|\mathbf{I_P}(\mathbf{X})\right\|^2 + (1-\rho).\left\|\mathbf{I_C}(\mathbf{X})\right\|^2 + \exp(-d_P) + \exp(-d_C)$$
$$\text{s.t. } 0 \le \psi \le 1,\ \alpha_1, \alpha_2, \alpha_3 \ge 0,\ I_P \ge 0,\ I_C \ge 0$$
$$\text{and } \alpha_1 + \alpha_2 + \alpha_3 = 1. \quad (14)$$

Maximizing the above function $g_1(\mathbf{X})$ for optimal $\mathbf{X}$ is equivalent to minimizing $-g_1(\mathbf{X})$. Thus this minimization problem can be reduced to a convex optimization form because (a) the inequality constraint functions are also convex, and (b) the equality constraint is affine. The convergence of this optimization function is skipped due to space limit.

Now, the minimum value of $-g_1(\mathbf{X})$ corresponds to an optimal $\mathbf{X}^*$ and hence we can easily compute the optimal interestingness measures $\mathbf{I_P}^*$ and $\mathbf{I_C}^*$ for the optimal $\mathbf{X}^*$. Given our framework for determining interestingness of conversations, we now discuss the measures of consequence of interestingness followed by extensive experimental results.

5. INTERESTINGNESS CONSEQUENCES

An interesting conversation is likely to have consequences. These include the (commenting) activity of the participants, their cohesiveness in communication and an effect on the interestingness of the themes. It is important to note here that the consequence is generally felt at a future point of time; that is, it is associated with a certain time lag (say, δ days) with respect to the time slice a conversation becomes interesting (say, q). Hence we ask the following three questions related to the future consequences of an interesting conversation:

Activity: Do the participants in an interesting conversation i at time q take part in other conversations relating to similar themes at a future time, $q+\delta$? We define this as follows,

$$Act^{(q+\delta)}(i) = \frac{1}{\left|\varphi_{i,q+\delta}\right|}\sum_{k=1}^{\left|\varphi_{i,q+\delta}\right|}\sum_{j=1}^{\left|P_{i,q}\right|}\mathbf{P_C}^{(q+\delta)}(j,k), \quad (15)$$

where $P_{i,q}$ is the set of participants on conversation i at time slice q, and $\varphi_{i,q+\delta}$ is the set of conversations m such that, $m \in \varphi_{i,q+\delta}$ if the KL-divergence of the theme distribution of m at time $q+\delta$ from that of i at q is less than an empirically set threshold: $D(\mathbf{C_T}^{(q)}(i,:) \parallel \mathbf{C_T}^{(q+\delta)}(m,:)) \le \varepsilon$.

Cohesiveness: Do the participants in an interesting conversation i at time q exhibit cohesiveness in communication (co-participate) in other conversations at a future time slice, $q+\delta$? In order to define cohesiveness, we first define co-participation of two participants, j and k as,

$$O^{(q+\delta)}(j;k) = \frac{\mathbf{P_P}^{(q+\delta)}(j,k)}{\mathbf{P_C}^{(q+\delta)}(j,:)}, \quad (16)$$

where $\mathbf{P_P}^{(q+\delta)}$ is defined as the participant-participant matrix of co-participation constructed as, $\mathbf{P_C}^{(q+\delta)}.(\mathbf{P_C}^{(q+\delta)})^t$. Hence the *cohesiveness* in communication at time $q+\delta$ between participants in a conversation i is defined as,

$$Co^{(q+\delta)}(i) = \frac{1}{\left|P_{i,q}\right|}\sum_{j=1}^{\left|P_{i,q}\right|}\sum_{k=1}^{\left|P_{i,q}\right|}O^{(q+\delta)}(j;k). \quad (17)$$

Thematic Interestingness: Do other conversations having similar theme distribution as the interesting conversation c_i (at time q), also become interesting at a future time slice $q+\delta$? We define this consequence as thematic interestingness and it is given by,

$$TInt^{(q+\delta)}(i) = \frac{1}{\left|\varphi_{i,q+\delta}\right|}\sum_{j=1}^{\left|\varphi_{i,q+\delta}\right|}\mathbf{I_C}^{(q+\delta)}(j). \quad (18)$$

To summarize, we have developed a method to characterize interestingness of conversations based on the themes, and the interestingness property of the participants. We have jointly optimized the two types of interestingness to get optimal interestingness of conversations. And finally we have discussed three metrics which account for the consequential impact of interesting conversations. Now we would discuss the experimental results on this model.

6. EXPERIMENTAL RESULTS

The experiments performed to test our model are based on a dataset from the largest video-sharing site, YouTube, which serves as a rich source of online conversations associated with shared media elements. We first present the baseline methods.

6.1 Baseline Methods

We discuss three baseline methods for comparison of our computed interestingness. We define the first baseline interestingness measure of a conversation based on the number of comments in a particular time slice so that it satisfies the following two constraints as in [4]: (a) a conversation is interesting at a time slice when it has several comments in that time slice, and (b) a conversation should not be considered interesting if all its comments are in a particular time slice and no comments occur in other time slices. The second baseline is based on the idea of novelty in participation: if several new participants join in a conversation at time q who did not appear at any time slice before q, then it implies the conversation is interesting. The third baseline is based on ranking conversations using the PageRank algorithm on the participant-co-occurrence graph $G(C,E)$ discussed in section 3.1. This is based on the motivation that if the participants of several conversations co-communicate on another conversation, it makes the latter interesting as it appeals to a large number of individuals.

6.2 Experiments

Here we present the experiments conducted on YouTube dataset.

6.2.1 Dataset

We executed a web crawler to collect conversations (set of comments) associated with videos in the "Politics" category from the YouTube website. For each video, we collected its timestamp, tags, its associated set of comments, their timestamps, authors and content. We crawled a total set of 132,348 videos involving 8,867,284 unique participants and 89,026,652 comments over a period of 15 weeks from June 20, 2008 to September 26, 2008. In the crawled data, there are a mean number of ~67 participants and ~673 comments per conversation. The reason behind choice of the Politics category is due to the rich dynamics related to the US Presidential elections over the said time period.

6.2.2 Results

Now we discuss the results of experiments conducted to test our framework.

Conversational Themes: In order to analyze the interestingness of conversations, we have extracted theme distributions of YouTube conversations at different time slices based on our theme model discussed in section 3. The number of themes K for the theme model is computed to be 19 for the dataset, which is given by the number of positive singular values of the word-chunk matrix, a popular technique used in text mining.

The results of the experiments on theme evolution are shown in a visualization in Figure 3. The visualization gives a representation of the set of 19 themes (columns) over the period of 15 weeks (rows from June 20, 2008 to September 26, 2008) of analysis. The themes are associated with representative "word clouds" which describe the content of the conversations associated with the themes. The strength of a theme ($\mathbf{T_S}$) at a particular time slice is shown as a blue block, whose higher intensity indicates that several conversations are associated with that theme. Since our dataset is focused on the Politics category, we observe that the word clouds are representative of the political dynamics about the 2008 US Presidential elections in the said period. For example, themes 5, 8 and 14 are consistently discussed over time in different conversations since they are about the major issues of the elections – 'abortion', 'war', 'soldiers' and 'healthcare'. Moreover, themes become strong about the same time when there is an external event related to its word cloud – theme 18 becomes strong when Palin and Biden are appointed as the VP nominees. This is intuitive because external events often manifest themselves on popular online discussions.

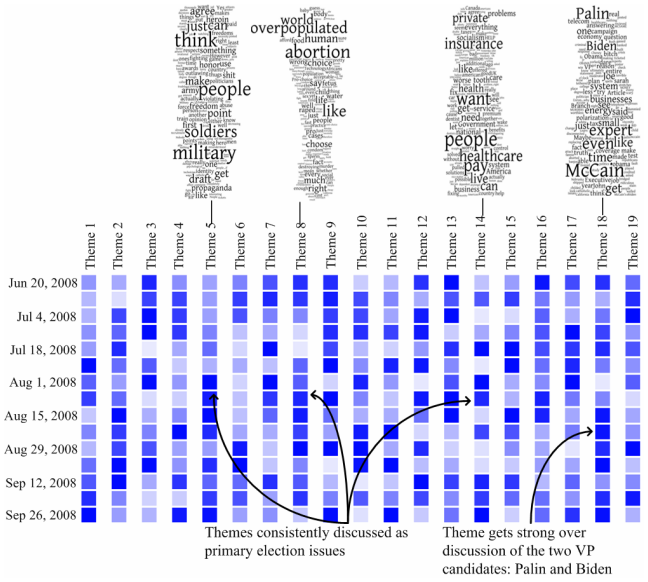

Figure 3: Evolution of conversational themes on the YouTube dataset: rows are weeks and columns are themes. The strength of a theme (number of conversations associated with it) at a particular week is shown as a blue block: strength is proportional to intensity of block. The themes are associated with their word-clouds; only a few themes are shown for clarity. We observe the dynamics of theme strengths with respect to external events.

Interestingness of participants and conversations: The results of interestingness of the participants are shown in Figure 4. We have shown a set of 45 participants over the period of 15 weeks (June 20, 2008 to September 26, 2008) by pooling the top three most interesting participants over all conversations from each week. From left to right, the participants are shown with respect to decreasing mean number of comments over all 15 weeks. The figure shows plots of the comment distribution and the interestingness distribution for the participants at each time slice along with the Pearson correlation coefficient between the two distributions. From the results, we observe that on the last three weeks (13, 14, 15) with several political happenings, the interestingness distribution of participants does not seem to follow the comment distribution well (we observe low correlation). Hence we conclude that during periods of significant external events, participants can become interesting despite writing fewer comments – high interestingness can instead be explained due to their preference for the conversational theme which reflects the external event.

The results of the dynamics of interestingness of conversations are shown in Figure 5. We show a temporal plot of the mean and maximum interestingness per week in order to understand the

relationship of interestingness to external happenings. From Figure 5, we observe that the mean interestingness of conversations increased significantly during weeks 11-15. This is explained when we observe the association with large number of political happening in the said period.

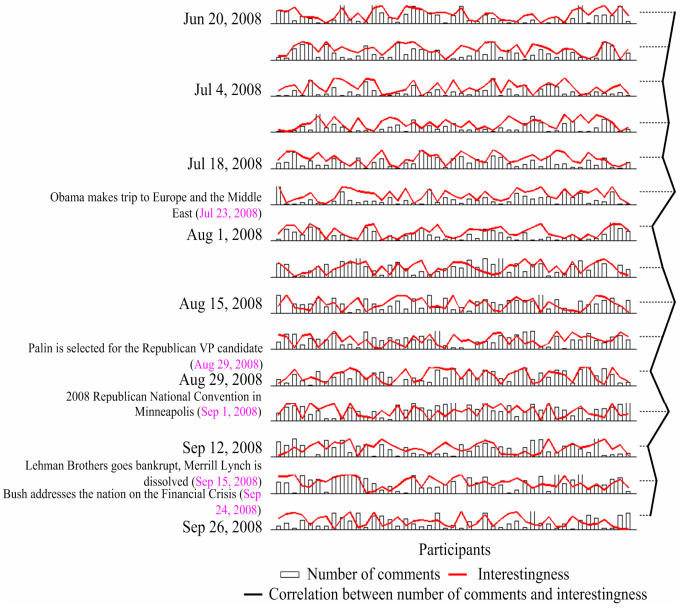

Figure 4: Interestingness of 45 participants from YouTube, ordered by decreasing mean number of comments from left to right, is shown along with the corresponding number of comments over 15 weeks (rows). The Pearson correlation coefficient between the number of comments and interestingness is also shown; which implies that interestingness of participants is less affected by number of comments during periods of significant external events.

Hence it seems that more conversations in general become highly interesting when there are significant events in the external world – an artifact that online conversations are reflective of chatter about external events. However, certain highly interesting conversations always occur at different weeks irrespective of events. This implies that conversations could become interesting even if the themes they discuss are not very popular at that point of time – rather, the interestingness in such cases could be attributed to the communication activity of the participants.

Relationship with media attributes: Now we explore the relationships between our computed interestingness of conversations and the attributes of their associated media objects. We consider correlation (using the Pearson correlation coefficient) between interestingness (averaged over 15 weeks) and number of views, number of favorites, ratings, number of linked sites, time elapsed since video upload and video duration which are media attributes associated with YouTube videos. From Table 1, we observe that there is low correlation of each of these attributes to conversations with high interestingness. We further observe that time elapsed since video upload and video duration have negative correlation with high interestingness – this is intuitive because videos which are recently uploaded and generate lot of attention quickly are likely to be highly interesting; also, most interesting conversations have been observed to be those which are short in duration. This justifies that media attributes cannot always be indicators of the interestingness of the conversations.

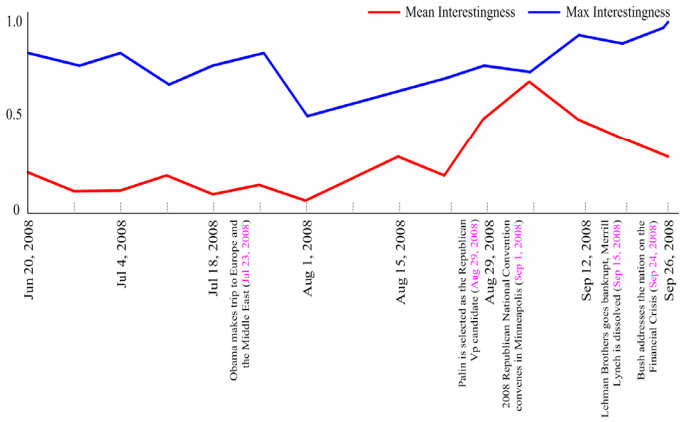

Figure 5: Mean and Max Interestingness of all conversations from the YouTube dataset are shown over 15 weeks (X axis). Mean interestingness of conversations increases during periods of several external events; however, certain highly interesting conversations always occur at different weeks irrespective of events.

Consequences of Interestingness: Now we present the results of measuring consequence of interestingness on the YouTube dataset captured by the three metrics discussed in section 5 – activity, cohesiveness and thematic interestingness. In order to compare the performance of our method, we use the three baseline methods – interestingness based on comment frequency (B_1), interestingness based on novelty of participation (B_2) and interestingness based on PageRank (B_3).

Table 1: Correlation coefficient between interestingness and media attributes. For convenience of interpretation, we segment conversations to have three types of interestingness, low ($0 \le I_C \le 0.33$), mid ($0.34 \le I_C \le 0.66$) and high ($0.67 \le I_C \le 1$).

Media Attribute	Corr. for Low Interestingness ($0 \le I_C \le 0.33$)	Corr. for Mid Interestingness ($0.34 \le I_C \le 0.66$)	Corr. for High Interestingness ($0.67 \le I_C \le 1$)
Number of views	0.24	0.78	0.53
Number of favorites	0.17	0.69	0.48
Ratings	0.10	0.38	0.51
Number of linked sites	0.18	0.62	0.61
Time elapsed since video upload	0.38	0.01	-0.29
Video duration	0.44	0.13	-0.14

To observe the consequential impact of interestingness, we determine its correlation to activity, cohesiveness and thematic interestingness using five methods – our interestingness measure with temporal smoothing (I_1), our interestingness measure without temporal smoothing (I_2), and the three baseline methods B_1-B_3. As discussed in section 5, the three consequence metrics would be felt after a certain time lag with respect to the point at which a conversation became interesting. Hence for each metric and method pair, we need to determine by what time lag the metric trails the interestingness with maximum correlation. Since interestingness of a conversation and its associated activity, cohesiveness or thematic interestingness computed over different time slices (weeks) can be considered to be time-series, we determine the cross-correlation between interestingness and each

of the consequence-based metrics for various values of lags (-40 to 40 days for leading and trailing consequences). The lag corresponding to which the correlation is maximum, is taken as the 'best lag'.

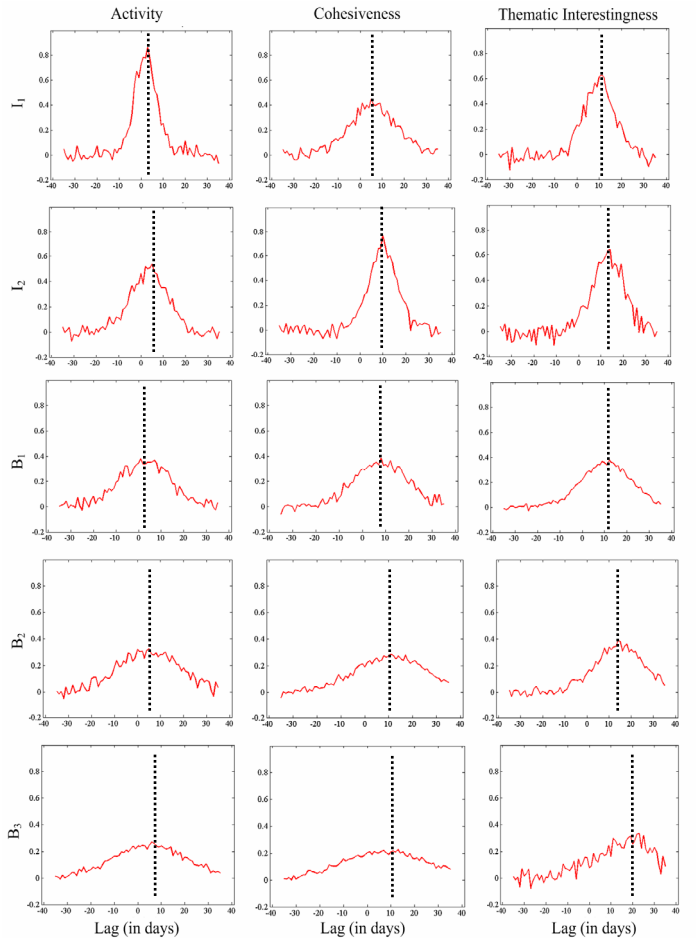

Figure 6: Best lag for correlation of interestingness measures to the three consequence-based metrics: activity, cohesiveness and thematic interestingness. Our method with temporal smoothing (I_1) is seen to be sharply correlated with the three metrics of consequences having the following lags – 3 days for activity, 6 days for cohesiveness and 11 days for thematic interestingness.

Figure 6 shows the correlation between the consequence-based metrics and interestingness of conversations computed using various methods for various lags, averaged over the entire period of 15 weeks. We observe that incorporating temporal smoothing significantly improves correlation (I_1 over I_2) for our method and this is explained by the fact that interestingness of conversations exhibits considerable relationship across time slices. We finally conclude from these results that our computed interestingness appears to have significant consequential impact on the three metrics due to high correlation compared to all baseline methods (mean correlation of 0.71 over all three metrics) – all the three baseline methods appear to have more or less flat correlation plots (mean correlation of 0.35 over all three metrics). Hence interestingness of conversations determined through our method could be predictors of communication dynamics in social media.

6.2.3 Evaluation against Baseline Methods

In this section we compare the efficiency of our algorithm in computing interestingness against the previously introduced baseline methods (B_1-B_3). We evaluate to what extent the consequence-based metrics (activity, cohesiveness and thematic interestingness) can be explained by each method using its best lag (from Figure 6). The measure chosen to demonstrate this evaluation is mutual information between interestingness and each metric: activity, cohesiveness and thematic interestingness.

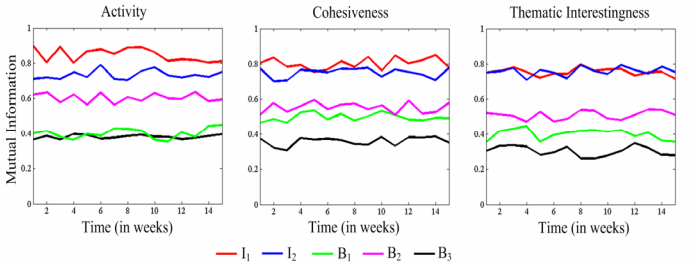

Figure 7: Evaluation of our computed interestingness I_1 and I_2 against baseline methods, B_1 (comment frequency), B_2 (novelty of participation), B_3 (co-participation based PageRank). Our method incorporating temporal smoothness (I_1) uses its best lags, 3 days for activity, 6 days for cohesiveness and 11 days for thematic interestingness and maximizes the mutual information for the three consequence-based metrics (activity, cohesiveness and thematic interestingness).

The results of evaluation are shown in Figure 7. We observe that our method I_1 maximizes mutual information for all three metrics (mean 0.83) – implying that our computed interestingness can successfully explain the three consequences compared to the baseline methods (mean 0.41). The baseline methods perform poorly because they have relatively flat correlation with the three consequences. This implies that our methods are effective in explaining the consequences reasonably.

6.2.4 Discussion

From the experimental results we have gained several insights. First, interestingness of participants is observed to be less correlated with the number of comments written by them during periods involving several significant events. High interestingness during such periods can be explained by other communication properties of participants, like preference for themes reflective of the events or co-participation with other interesting participants. Second, mean interestingness of conversations increases during periods of significant external events – implying that conversations often involve active discussion about evolutionary themes reflective of external events. Third, evaluation shows that our method can successfully explain the consequences on participants and themes. To summarize, interestingness of conversations is an important property associated with online social media because it captures the dynamics of the participants and the themes, in contrast with static analysis of media content.

7. CONCLUSIONS

We have developed a computational framework to characterize the conversations in online social networks through their "interestingness". Our model comprised the following parts. First we detected conversational themes using a mixture model approach. Second we determined interestingness of participants

and interestingness of conversations based on a random walk model. Third, we established the consequential impact of interestingness via metrics: activity, cohesiveness and thematic interestingness. We conducted extensive experiments using dataset from YouTube. During evaluation, we observed that our method maximizes the mutual information by explaining the consequences (activity, cohesiveness and thematic interestingness) significantly better than three other baseline methods (our method 0.83, baselines 0.41).

Our framework can serve as a starting point to several interesting directions to future work. We believe that incorporating visual features of the media objects associated with the conversations can boost the performance of our algorithm. It would also of use in resource allocation to determine if there are particular time-periods during which conversations become interesting. Moreover, because alternative definitions of a subjective property like interestingness are always possible, in the future we are interested in observing how such a property is connected to the structural and temporal dynamics of an online community.

8. REFERENCES

[1] *YouTube* http://www.youtube.com/.

[2] E. ADAR, D. S. WELD, B. N. BERSHAD, et al. (2007). *Why we search: visualizing and predicting user behavior*. Proceedings of the 16th international conference on World Wide Web. Banff, Alberta, Canada, ACM: 161-170.

[3] M. D. CHOUDHURY, H. SUNDARAM, A. JOHN, et al. (2008). *Can blog communication dynamics be correlated with stock market activity?* Proceedings of the nineteenth ACM conference on Hypertext and hypermedia. Pittsburgh, PA, USA, ACM: 55-60.

[4] M. DUBINKO, R. KUMAR, J. MAGNANI, et al. (2006). *Visualizing tags over time*. Proceedings of the 15th international conference on World Wide Web. Edinburgh, Scotland, ACM: 193-202.

[5] V. GÓMEZ, A. KALTENBRUNNER and V. LÓPEZ (2008). *Statistical analysis of the social network and discussion threads in slashdot*. Proceedings of the 17th international conference on World Wide Web. Beijing, China, ACM: 645-654.

[6] D. GRUHL, R. GUHA, R. KUMAR, et al. (2005). *The predictive power of online chatter* Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining Chicago, Illinois, USA 78-87

[7] T. HOFMANN (1999). *Probabilistic latent semantic indexing*. Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. Berkeley, California, United States, ACM: 50-57.

[8] A. KALTENBRUNNER, V. GOMEZ and V. LOPEZ (2007). *Description and Prediction of Slashdot Activity*. Proceedings of the 2007 Latin American Web Conference, IEEE Computer Society: 57-66.

[9] L. S. KENNEDY and M. NAAMAN (2008). *Generating diverse and representative image search results for landmarks*. Proceeding of the 17th international conference on World Wide Web. Beijing, China, ACM: 297-306.

[10] X. LING, Q. MEI, C. ZHAI, et al. (2008). *Mining multi-faceted overviews of arbitrary topics in a text collection*. Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. Las Vegas, Nevada, USA, ACM: 497-505.

[11] Y. LIU, X. HUANG, A. AN, et al. (2007). *ARSA: a sentiment-aware model for predicting sales performance using blogs*. Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. Amsterdam, The Netherlands, ACM: 607-614.

[12] Q. MEI and C. ZHAI (2005). *Discovering evolutionary theme patterns from text: an exploration of temporal text mining* Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining Chicago, Illinois, USA ACM Press: 198-207

[13] Q. MEI, C. LIU, H. SU, et al. (2006). *A probabilistic approach to spatiotemporal theme pattern mining on weblogs*. Proceedings of the 15th international conference on World Wide Web. Edinburgh, Scotland, ACM: 533-542.

[14] Q. MEI, D. CAI, D. ZHANG, et al. (2008). *Topic modeling with network regularization*. Proceeding of the 17th international conference on World Wide Web. Beijing, China, ACM: 101-110.

[15] G. MISHNE (2006). *Leave a Reply: An Analysis of Weblog Comments*, Third annual workshop on the Weblogging ecosystem (WWE 2006), Edinburgh, UK,

[16] Y. ZHOU, X. GUAN, Z. ZHANG, et al. (2008). *Predicting the tendency of topic discussion on the online social networks using a dynamic probability model*. Proceedings of the hypertext 2008 workshop on Collaboration and collective intelligence. Pittsburgh, PA, USA, ACM: 7-11.

9. APPENDIX

We discuss the parameter estimation of the conversational theme model in section 3 using the Generalized Expectation Maximization algorithm (GEM). Specifically, in the E-step, we first compute the expectation of the complete likelihood $\Theta(\Psi; \Psi^{(m)})$, where Ψ denotes all the unknown parameters and $\Psi^{(m)}$ denotes the value of Ψ estimated in the m^{th} EM iteration. In the M-step, the algorithm finds a better value of Ψ to ensure that $\Theta(\Psi^{(m+1)}; \Psi^{(m)}) \geq \Theta(\Psi^{(m)}; \Psi^{(m)})$. First we empirically fix the free transition parameters involved in the log likelihood in equation (8): γ_q to be 0.5 for all q and ς as well to be 0.5. For the E-step, we define a hidden variable $z(w,\lambda_{i,q},j)$. Formally we have the **E-step**:

$$z\left(w,\lambda_{i,q},j\right) = \frac{p^{(m)}\left(w\mid\theta_j\right)\left(\left(1-\gamma_q\right)p^{(m)}\left(\theta_j\mid\lambda_{i,q}\right)+\gamma_q p^{(m)}\left(\theta_j\mid q\right)\right)}{\sum_{j'=1}^{K} p^{(m)}\left(w\mid\theta_{j'}\right)\left(\left(1-\gamma_q\right)p^{(m)}\left(\theta_{j'}\mid\lambda_{i,q}\right)+\gamma_q p^{(m)}\left(\theta_{j'}\mid q\right)\right)}. \quad (19)$$

Now we discuss the **M-step**:

$$\Theta(\Psi;\Psi^{(m)}) = \left(1-\varsigma\right)\left(\sum_{\lambda_{i,q}\in C}\sum_{w\in\lambda_{i,q}} n\left(w,\lambda_{i,q}\right).\sum_{j=1}^{K} z\left(w,\lambda_{i,q},j\right).a_j + L_\lambda + L_q + L_j\right) \quad (20)$$

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

where $L_\lambda=\alpha_\lambda(\sum_j p(\theta_j|\lambda_{i,q})-1)$, $L_q=\alpha_q(\sum_j p(\theta_j|q)-1)$ and $L_j=\alpha_j(\sum_w p(w|\theta_j)-1)$ are the Lagrange multipliers corresponding to the constraints that $\sum_j p(\theta_j|\lambda_{i,q}) = 1$, $\sum_j p(\theta_j|q) = 1$ and $\sum_w p(w|\theta_j) = 1$. Based on several iterations of E and M-steps, GEM estimates locally optimum parameters of the K theme models. Details of convergence of this algorithm can be referred in [14].