Why Do We Converse on Social Media? An Analysis of Intrinsic and Extrinsic Network Factors

Munmun De Choudhury
School of Communication and Information
Rutgers, The State University of New Jersey
m.dechoudhury@rutgers.edu

Hari Sundaram
School of Arts, Media and Engineering
Arizona State University, Tempe, AZ 85281, USA
hari.sundaram@asu.edu

ABSTRACT

We are motivated in our work by the following question: *what factors influence individual participation in social media conversations?* Conversations around user posted content, is central to the user experience in social media sites, including Facebook, YouTube and Flickr. Therefore, understanding *why* people participate, can have significant bearing on fundamental research questions in social network and media analysis, such as, network evolution, and information diffusion.

Our approach is as follows. We first identify several key aspects of social media conversations, distinct from both online forum discussions and other social networks. These aspects include intrinsic and extrinsic network factors. There are three factors *intrinsic* to the network: social awareness, community characteristics and creator reputation. The factors *extrinsic* to the network include: media context and conversational interestingness. Thereafter we test the effectiveness of each factor type in accounting for the observed participation of individuals using a Support Vector Regression based prediction framework. Our findings indicate that factors that influence participation depend on the *media type*: YouTube participation is different from a weblog such as Engadget. We further show that an optimal factor combination improves prediction accuracy of observed participation, by ∼9–13% and ∼8–11% over using just the best hypothesis and all hypotheses respectively. Implications of this work in understanding individual contributions in social media conversations, and the design of social sites in turn, are discussed.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social And Behavioral Sciences

General Terms

Algorithms, Experimentation.

Keywords

Conversations, Participation, Social media.

1. INTRODUCTION

Today, rich media sites including Flickr and YouTube as well as weblogs including Engadget and Huffington Post have emerged as popular channels for the expression of individual interests, ideas and opinions. These rich media sites allow users to share content, including uploading images, text and videos. Importantly, the shared content allows users to communicate with other users, through comments on the shared media object. We define a sequence of temporally-ordered comments on the shared media object, as a "conversation."

Conversations are important to understand the nature of the underlying social network [2]. In particular, conversations can be used to study the following: user behavior [11] and information roles, including content dissipators, impact on information cascades [4], and influence propagation. Hence, it is important to understand user participation in the context of social media conversations. For example, why do certain conversations exhibit continued and increasing participation from individuals? In this light, our work in this paper is motivated by the following question: *what are the factors that influence individual participation in social media conversations?* Notice that by "participation," we mean that a user has posted comments on a conversation.

Understanding the motivations behind participation of individuals in social media conversations involves several challenges. These challenges are related to key aspects of the social network: the inherent culture of interaction within the greater community, the affinity of the community to invite new individuals, the standard practices of social actions and the goal and purpose of the community-wide interactions. Contemporary online communities support different types of social interaction, and cater to different kinds of audiences. Rich media sites, for example, including YouTube and Flickr, primarily cater to sharing of media objects. On the other hand, blog forums such as Engadget or Huffington Post are directed towards technology-savvy or liberal political audiences who intend to remain engaged in interactions around news events. Therefore, it is likely that different social media sites will have different factors driving conversational participation within their sites. Furthermore, it is likely that there are differences between the motivations of newcomers to participate, compared to the existing members.

In this light, identifying factors influencing participation in each of these sites, and how they vary across the types of sites and participants, is critical. In particular, a careful analysis of participation can help contextualize network phenomena (e.g. distribution of information roles, or network dynamics including changes to the structure and information flow) within these sites. An application of our work includes better design of social media websites — in particular, sites where individuals interact with a shared media object (videos, photos, blogs).

Our Contributions. We define the participation of individuals on a social media conversation as "collective participation." There are two aspects to it: newcomers and existing participants[1]. The for-

[1]Our fundamental unit of analysis is a conversation around a media object. We are investigating why a person, already within a social media site, and commenting on posts on a topic, joins a new conversation on the same topic. Hence a newcomer in our context is limited to the scope of a conversation, and different from the perceived notion of a newcomer in prior literature on why people join social media websites.

mer, includes individuals who have not posted a comment or reply on the particular conversation thus far. The latter includes participants who have posted at least one comment or reply at an earlier point in time. We identify *intrinsic* as well as *extrinsic* network factors influencing collective participation from both newcomers and participants who have posted comments on a conversation earlier. Following these two categories of factors, we develop one hypothesis for each factor to test the influence of the factor on participation. There can be several ways to qualitatively validate the proposed hypotheses including via ethnographic studies. We adopt a quantitative prediction approach, based on Support Vector Regression.

We tested our hypotheses on two dataset classes — two rich media datasets, Flickr and YouTube, and two blog forum datasets, Engadget and Huffington Post. Our results indicate that different factors influence conversations from the two data classes differently. On one hand, extrinsic network factors explain participation on rich media conversations. On the other hand, intrinsic network factors seem to explain participation on blog forums.

Since a combination of factors may better explain collective participation, we propose a Bayesian Information Criterion (BIC) metric to test hypothesis combinations. Interestingly, we find that including *all* of the intrinsic and extrinsic network factors does not yield the best prediction accuracy. This reveals that there is likely to be a complex set of factors responsible for the nature of participation observed on different social media conversations today.

The rest of this paper is organized as follows. In section 2 we discuss prior work, followed by a discussion of the nature of social media conversations. Sections 4, 5 and 6 present the datasets, the factors behind participation and the prediction framework. In section 6 we present our experiments. We test the impact of combining multiple hypotheses in section 7. We conclude with a discussion of the implications of this work and its contributions in sections 8, 9.

2. RELATED WORK

Over several years, sociologists have been interested in understanding individual participation that underpins social movements. Dixon et al. [7] considered aggregate network processes that may condition the costs and benefits of participation in social movements. Recent work on understanding participation over the Internet has focused on factors associated with continued contribution of individuals on newsgroups, discussion forums, and online communities and networks and social media [1, 10, 12].

Lampe et al. [9] examined the participation of users on the technical community Slashdot and substantiated three explanations for participation. Joyce and Kraut [8] studied the factors behind participation in newsgroups. In the context of social networks, Burke et al. [3] studied content contribution on Facebook. On the rich media site Flickr, Nov et al. [10] studied how the tenure in a community affects individuals' participation.

Limitations of Prior Work: The state-of-the-art has made significant contributions to understanding factors behind voluntary participation in physical and online communities. A key property of these online communities is the following: there are clear incentives behind an individual's participation in the discussion forum, in editing a Wikipedia article or posting/tagging a photo on Facebook. Therefore, from the prior literature we gain the insight that participation can in these contexts be explained by considering intrinsic factors within the social network. Such factors include, the awareness of a participant to feedback/responses from her peers or her familiarity with the peers in the past.

However, prior research has not investigated participation in the context of the conversations in rich media around which a social network evolves. It is natural to conjecture that *a combination of factors* is likely to impact participation. Addressing these concerns is a major focus in this work.

Table 1: Details of the four datasets crawled in 2009.

Dataset	#Participants	#Conversations	#Comments
Rich Media Datasets			
YouTube	17,736,361	272,810	145,682,273
Flickr	4,304,525	305,258	26,557,446
Blog Forum Datasets			
Engadget	78,740	45,073	6,580,256
Huff Post	59,282	24,479	4,748,837

3. WHY ARE SOCIAL MEDIA CONVERSATIONS DIFFERENT?

Social media conversations possess unique characteristics. These features of social media conversations are different from online forum discussions, where user participation has been typically investigated. The key features of conversations include: community, presence of shared media and conversational interestingness.

Community. Shared media conversations can promote cohesive interaction amongst community members. Members of the community can interact in a specific conversation due to several reasons. First, individuals can come together because they share a common interest in the topic. Second, individuals may be interested in expressing their opinion on a media object related to a recent event. Finally, they may be interested in exchanging ideas with familiar community members, whom they observe participating in the conversation. Thus, an individual's observations of the larger community is likely to influence her participation in a conversation.

Shared Media. Social media conversations take place in the context of a shared media object, including a video on YouTube, or a post on the technology blog, Engadget. Naturally, the content (and context) of the media object—e.g., visual features of an image/video, textual content of a blog post, their associated metadata etc. are likely to impact an individual's desire to participate in the associated conversation. Hence, analysis of factors behind voluntary participation in these conversations needs to consider the properties of the shared media.

Conversational Interestingness. Temporal theme evolution is a key characteristic of social media conversations. New themes slowly emerge due to new user comments, and over time, the conversation topic can bear little resemblance to the original conversation topic [5]. Theme popularity affects the participants who comment in such themes. In [5], the authors operationalize temporal evolution of a conversation by the "interestingness" measure of the conversation. We conjecture that the degree of interestingness of a conversation, influences individual participation.

4. DATASETS

We now provide an overview of the datasets used in the paper. A key goal in this work is to understand the factors affecting collective participation in different types of social media conversations, i.e. a temporal sequence of comments and replies. We identify two different conversational contexts: conversations centered around a shared rich media object (image, video) and conversations centered around shared textual content, including blogs. We utilize two datasets from each of the two categories—two rich media websites, Flickr (http://www.flickr.com/) and YouTube (http://www.youtube.com/), and two blog forums, Engadget (http://www.engadget.com/) and Huffington Post (http://www.huffingtonpost.com/).

We describe the details of each dataset in Table 1. Note that for

Table 2: Media context on multiple rich media and blog datasets.

DATASET	MEDIA CONTENT FEATURES	MEDIA META-DATA
	Rich Media Datasets	
YouTube	Visual features of the video—color (color histogram, color moments), texture (GLCM, phase symmetry), shape (radial symmetry, phase congruency) and keypoint location features (SIFT) [6]	Number of views[a], number of 'favorites'[a], ratings, number of linked sites, time elapsed since video upload (recency), video duration
Flickr	Visual features of the photo: same as YouTube	Number of tags[a], number of notes, number of views[a], number of 'favorites'[a], number of associated groups, time elapsed since photo upload
	Blog Forum Datasets	
Engadget	tf-idf (term frequency-inverse document frequency) based features of the blog content; where the content is represented as a stemmed and stop-word eliminated bag-of-words	Number of tags[a], time elapsed since blog was posted (recency), number of Facebook "likes"[a], length of the post
Huffington Post	tf-idf based features of the blog content: same as Engadget	Same as Engadget

[a] Variable is log-transformed to correct for skew.

the purpose of comparison, we consider approximately the same length (~ 147 days) for all the four datasets.

5. FACTORS IN SOCIAL PARTICIPATION

5.1 Intrinsic Network Factors

The nature of the social network in which the conversation is embedded influences *intrinsic* network factors. Intrinsic network factors include: an individual's 'social awareness,' 'community characteristics' and 'reputation' of the media creator.

Social Awareness. Participation of individuals in social media conversations is dependent upon factors that induce social awareness in an individual. We utilize three measures of social awareness:

Familiarity: We quantify the degree of *familiarity* of an individual associated with a conversation to be the fraction of the mean number of times they co-participated with all individuals in any prior conversation on the same topical category.

Feedback: Next we quantify the degree of *feedback* of an individual associated with a conversation to be the fraction of the mean number of replies she received from other participants on the same conversation in the past.

Dialogue: Dialogue is a measure of the overall back and forth communication (comment/reply) that has happened between the participants in the past. Presence of dialogue among the participants in a conversation is given by the ratio of the frequency of all the replies to all the comments in it.

HYPOTHESIS 1. *Collective participation on a social media conversation is affected by the degree of social awareness of the participating individuals, including their familiarity with other participants in the conversation, feedback from others and dialogue among others.*

Community Characteristics. Properties of the overall community also influence collective participation in conversations. We consider a community to be a set of individuals who engage in commentary centered around a broad topic. A typical community in our dataset, for example, on YouTube is a set of individuals who write comments or replies around shared videos on "News & Politics", which belongs to the YouTube-defined topical taxonomy. Engadget and Huff Post also have such taxonomies featuring topics such as "Tech", "Lifestyle" and "Media", "Comedy" etc. respectively. For Flickr, a community is the set of individuals who associate their photos to different photo pools, defined topically (e.g. "Black and White Photography", "Nature lovers").

We consider different properties, structural and temporal, to characterize online communities:

Community size is defined as the number of unique individuals who have posted a comment or a reply at least once on all conversations associated with media objects belonging to a certain topic.

Community activity is given by the mean number of postings of comments and replies across all the individuals in the community.

Community cohesiveness is defined as the mean clustering coefficient of the communication graph. The graph is induced by the co-participation of individuals commenting (or replying) to all conversations associated with media objects belonging to a certain topic. In this graph, the nodes are the individuals participating on conversations on the topic, while an edge between two nodes indicates that they have commented/replied together at least once.

Community sustenance is defined as the mean degree of retention of communicating individuals over time. Sustenance is given by the fraction of the number of individuals who repeatedly return to the community over time to post comments / replies on conversations belonging to the particular topic.

HYPOTHESIS 2. *Collective participation on a social media conversation is affected by the characteristics of the larger community, including its size, how active and cohesive its members are, and to what degree it is able to sustain its members over time.*

Creator Reputation. A creator is an individual who uploads a video on YouTube, shares a photo on Flickr or write blog posts on Engadget or Huffington Post. Since conversations are typically centered around a media object, the identity or characteristics of the creator is likely to play an important role in the communication. We consider the following attributes to quantify creator reputation: number of media objects uploaded by the individual, his or her number of (social) contacts in the community, and the duration of his or her 'tenure' i.e. the time elapsed since s/he joined the site.

HYPOTHESIS 3. *Collective participation on a social media conversation is affected by the reputation of the creator of the associated media artifact, including his or her activity in media creation, his network authority score and tenure in the larger community.*

5.2 Extrinsic Network Factors

We note that participants also receive external 'information signals' through *extrinsic* network factors, that may be due to an image/video posted in response to an external event, or associated with emergent themes due to conflicting opinions. These include the 'media context' and 'conversational interestingness'.

Media Context. As mentioned earlier, a distinct feature of participation on social media conversations is that it takes place around a shared media object. Hence the media context is also useful in analyzing the degree of collective participation over time. We consider two kinds of collective participation media contexts: the visual/textual content (features) of the media object, and media metadata. A detailed description of the two different aspects of the media context is described in Table 2.

HYPOTHESIS 4. *Collective participation on a social media conversation is affected by the context associated with the media artifact, including its visual or textual content as well as media metadata, including its ratings, views, tags and recency of upload.*

Conversational Interestingness. A typical aspect of social media conversations is that they engender communication around the shared media spanning a variety of external events. As a result, *we conjecture that collective participation will be significantly affected by the evolving nature of the conversation itself.* We consider a subjective temporal property of the conversations: known as "interestingness". We utilize the interestingness model proposed in [5] to compute this measure as a real scalar value in the range [0,1]. Interestingness of a conversation at any given time depends on its themes (popular themes featured in a conversation are likely to make it interesting to individuals and facilitate participation); and also the prior communication activity of its participants.

HYPOTHESIS 5. *Collective participation on a social media conversation is affected by the characteristics of the conversation itself, such as its interestingness over time, where interestingness is characterized by the popularity of the conversational themes and the communication properties of the participants around those themes.*

Given the various factors behind participation, we now discuss how we can evaluate the impact of each factor type or hypothesis on collective participation in social media conversations.

6. A PREDICTION FRAMEWORK

We propose a prediction approach to evaluate each hypothesis in explaining observed participation. We utilize an incremental Support Vector Regression model to predict the degree of observed participation, that can be attributed to each of the five different types of factors.

First, we discuss the construction of our "ground truth" for quantifying the influence of each type of factor towards newcomer and existing user participation. Since we are using a temporal regression model, we model our framework, and compute the ground truth data over a set of time slices[2] (say, K). The ground truth participation on a conversation from newcomers is given by the number of comments in it at a given time slice, by individuals who had not posted any comments (or replies) on the same conversation at any earlier time slice. Similarly, the ground truth for existing participants is the number of comments on the conversation by individuals who had posted at least one comment (or reply) on the same conversation earlier.

Second, we compute the various factors defined in the previous section over a time slice, such that *each hypothesis* can be defined as a feature set over *all its associated factors*. We are interested in observing how each feature set individually quantifies the observed participation, from both newcomers and existing participants at a future time slice.

Thereafter we utilize the data, corresponding to each feature set, over the first p time slices ($p < K$) to train a SV Regression model

[2] In this work, each time slice is equal to 1 day.

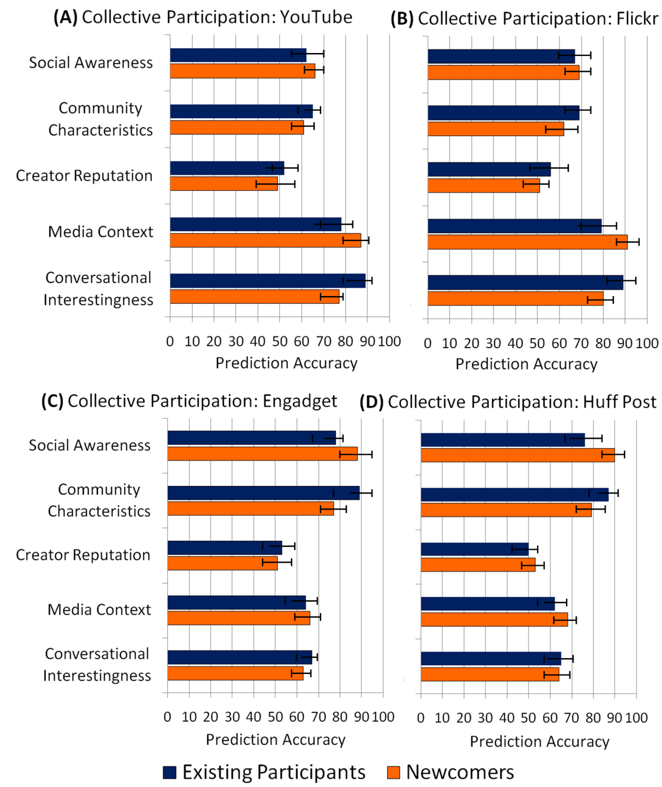

Figure 1: Prediction accuracies (higher numbers are better) of collective participation in social media conversations over four different datasets; corresponding error bars are also shown to illustrate the deviations.

(based on a Gaussian RBF kernel). We then use the learnt parameters to obtain the predicted measures of participation (i.e. the number of comments from newcomer and existing participants) over time slices $p + 1$ through K. The effectiveness of the chosen feature set (or hypothesis) is therefore given by the mean percentage accuracy in predicting the values of the number of comments from newcomers and existing participants respectively.

7. VALIDATING HYPOTHESES

We conduct elaborate experimental studies on all the four datasets introduced in section 4, in order to find empirical grounding on the five different hypotheses behind collective participation. For the datasets, we choose the first 97 days ($\sim 65\%$) as the training set and the next 50 days ($\sim 35\%$) as test set in each case.

In Figure 1, we present the prediction performance of using different feature set categories in accordance with the different hypotheses framed in section 4. The performance is evaluated based on the corresponding percent accuracy metric.

Rich media vs. Blog Forums. We observe differences in the feature sets that yield the best prediction performance across the two dataset types. For rich media data, extrinsic network factors (media context and conversational interestingness; mean accuracy $\sim$80%) seem to be better predictors of participation compared to social awareness and community characteristics. This is because the nature of the shared media is central to triggering users to participate in conversations. For blog forums data, intrinsic network factors (social awareness and community characteristics; mean accuracy $\sim$78%) seem to better predict participation compared to the others. This is because participation on these websites are often driven by personal opinions on technology or political happenings.

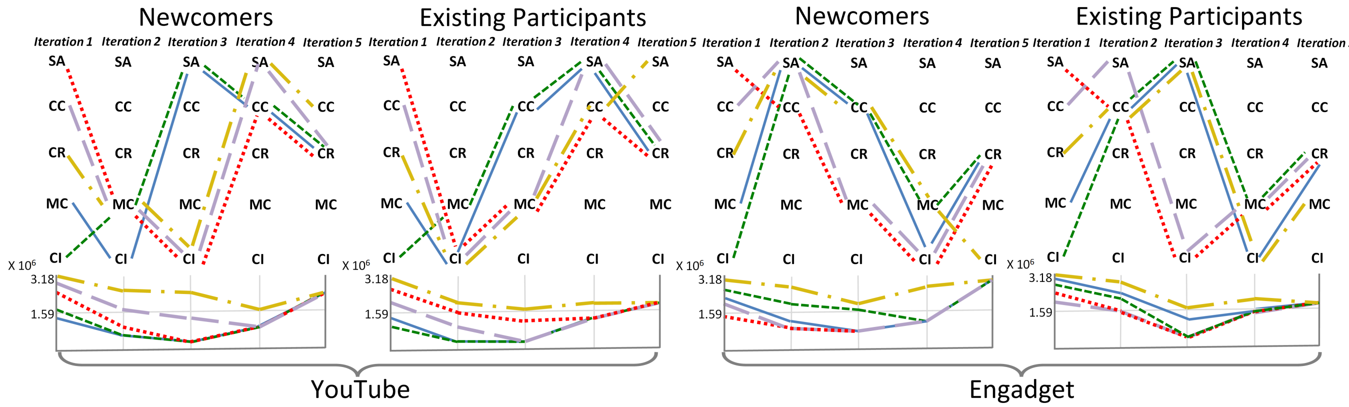

Figure 2: Performance of combining different feature categories (or hypotheses) in predicting collective participation. For each starting feature set, we show which feature sets were selected at each iteration, that minimizes the BIC. The plot at the bottom shows the actual BIC measures of the combinations at each step (lower BIC values are better). Here SA: Social Awareness, CC: Community Characteristics, CR: Creator Reputation, MC: Media Context and CI: Conversational Interestingness.

Hence the overall community's response and behavior to a certain event are likely to be important factors behind participation.

Newcomers vs. Existing Participants. There are also significant differences across the factors that affect participation from newcomers and existing participants. Conversational interestingness and community characteristics perform relatively better for all datasets in the case of existing participants. This is because over time they are able to 'learn' a community's dynamics: its nature of activity as well as can judge better (via comparison) the interestingness of the on-going conversations. Newcomers seem to rely more on media context and social awareness. This is because their participation is more likely to be triggered by the popularity of the media shared, or by how the rest of the participants are reacting to their comments.

Creator Reputation. Creator reputation does not explain collective participation well for any of the datasets (accuracy ∼49%). We believe that there are two explanations: large number of authors in rich media and restrictive media authorship in blogs. Anyone can upload a media object in rich media websites. Since there is no restriction on *who* can upload—the number of creators on rich-media sites is very large. This overabundance of creator choice, in rich media sites, makes the creator a less likely candidate as the *sole* attribute on which to filter media. On the two blog forums analyzed in this work, only a fixed number of Editors can create content. Since all the content on these two blogs are created by editors, the reputation of the editor makes little difference to participation.

We conclude that *not* all the stated hypotheses are able to quantify the observed participation equally well. In essence, there are differences across the two classes of datasets, as well as between the participation from newcomers and existing participants.

8. COMBINING MULTIPLE HYPOTHESES

Collective participation in online media will typically be manifested due to a collection of factors, rather than a single factor type. In this section, we therefore use a Bayesian Information Criterion (BIC) based measure to determine such an optimal set of factors.

BIC Measure. Our goal is to determine an "optimal" number of factors, typically smaller than the set of all factors, that can best explain the observed participation. This problem can be reduced to a model selection problem: hence we utilize a measure frequently used in model selection, known as Bayesian Information Criterion (BIC) to find the optimal factor set combination.

We develop an iterative approach to determine the optimal hypothesis combination using the BIC measure. We start with a random hypothesis (called *seed*), and sequentially add hypotheses to it. The feature vector corresponding to the chosen starting hypothesis is used to predict the collective participation using the Support Vector Regression technique discussed in Section 6. Using the prediction model, we then compute the BIC measure of the combination at the current iteration. The same exercise is repeated for various choice of seed hypotheses. Finally, the optimal hypothesis combination is the one that *minimizes the Bayesian Information Criterion (BIC)* measure, for predicting participation.

Results. We present the results of combining hypotheses to predict collective participation in Figure 2. The figure has two parts. In the top part, we show a visual representation of which hypotheses were chosen at each iteration for each starting seed hypothesis. This is shown using linear paths between the hypotheses (each path has a different starting seed hypothesis). That is, in the figure, in the prediction of newcomer participation for YouTube, at iteration 3, for the starting seed hypothesis CI, we have the optimal combination as {CI, MC, SA}, shown with the green dotted path. Next in the bottom part of the figure, we show the BIC value of each hypothesis combination at each iteration (shown in a line plot with the same color and style as the corresponding path in the top part).

The results indicate that combining hypotheses does *indeed* appear to improve the prediction of collective participation for both newcomers and existing participants. It appears that the combinations that perform the best are the ones which have the starting seed as the best performing hypothesis in Figure 1. However, surprisingly enough, using all information in terms of all five hypotheses *does not* yield the best prediction. In fact the best performance, as seen in the BIC curves in Figure 2 are given by hypotheses combinations in the middle of the curve—i.e., a selective few hypotheses quantify collective participation in the best manner.

Table 3: Summary of results of combining hypotheses in prediction of collective participation.

Dataset	Newcomers	Existing Participants
YouTube	{MC, CI, SA}	{MC, CI, CC}
Flickr	{MC, CI, SA}	{MC, CI, CC}
Engadget	{SA, CC, MC}	{SA, CC, CI}
Huff Post	{SA, CC, MC}	{SA, CC, CI}

A summary of the best performing combinations is shown in Table 3. We also present in Table 4 the prediction accuracies for these best performing combinations and compare them to those of using

Table 4: Prediction accuracies using (I) just the best performing hypotheses (Figure 2), (II) optimal hypotheses combination (Figure 3), and (III) all five hypotheses.

Dataset	Newcomers			Existing participants		
	I	**II**	III	I	**II**	III
YouTube	79%	**88%**	80%	80%	**92%**	81%
Flickr	82%	**92%**	82%	80%	**91%**	82%
Engadget	76%	**89%**	78%	76%	**87%**	76%
Huff Post	83%	**93%**	82%	81%	**90%**	80%

just the best performing hypothesis (ref. Figure 1) and the combination of all five hypotheses. The best combination improves prediction accuracy significantly by ∼9–13% and ∼8–11% respectively over using just the best hypothesis and all hypotheses.

The combinations that work best (see Table 3), for example, for rich media are ones which utilize extrinsic network factors: MC and CI. For blogs, intrinsic network factors SA and CC play a key role. For newcomers MC and SA are important across both blogs and rich media, while for existing participants, CC and CI are key across all datasets.

It is reasonable to conclude from these experiments that collective participation on social media conversations are guided by *a complex set of factors*. However, different factors dominate depending on the type of the site: rich media site participation depends more on the properties of the conversation itself, while the blog forums particpation are guided by the social attributes including awareness and community behavior.

9. DISCUSSION

We now discuss some of the implications of our work. This is an early paper on the differences amongst factors behind social media conversational participation, including media features, intrinsic and extrinsic network features. Prior work has not explored, via quantitative analysis, the roles of each category towards participation. We view comments, as social features. Prior work did not establish quantitatively, for example, if comments were more important to blogs or to rich media sites. It is clear from our work that interestingness of conversations matters more, on Flickr, than on Engadget, to support conversational participation. We hope that future research can use our results as a basis for design exploration. Preliminary design considerations for a photo sharing social media can therefore, as an example provide extensive commentary platforms around different themes of media engagement that would encourage users to participate on conversations extensively. Developing concrete design principles, however, requires additional controlled user studies based on outcomes of this paper, focused exclusively on social media site design. Social media design principles is an important problem that will be explored in future research.

It is possible that we have not exhaustively examined the set of factors influencing participation. Unobserved variables, including participant demographics, cultural norms, sentiment and linguistic style may also affect participant behavior. Fleshing out other factors driving participation remains a ripe area for future research.

Additionally, we have considered only one kind of participation on social media sites: posting comments on conversations. An individual may also participate in other ways: individuals can participate by rating comments, sharing posts and comments of interest. We would be interested to see in future work, if our intrinsic and extrinsic factors can explain other forms of participation.

Finally, we acknowledge that beyond cross-category differences, understanding the role of each feature within each category can be useful, e.g. to understand positive and negative influences at the level of an individual feature. The number of features being large, a systematic analysis at the feature level has not been presented in the current work; however can be an interesting future investigation.

10. CONCLUSION

In this paper, we investigated several factors to explain participation in social media conversations. Investigating the factors allows us to understand the nature of the underlying social network, including network structure and evolution, and information roles, and influence propagation. Efficient design of social media sites is one potential application of our work.

Our contributions addressed limitations of prior work. Prior work typically looked at intrinsic network factors affecting the awareness of the participant, and paid little attention to extrinsic network factors, including conversational dynamics and content. We incorporated both intrinsic and extrinsic factors in our work. A second difference is that that we investigated how a combination of factors influence participation on social media conversations. We plan to investigate several research directions in the future, including a careful analysis of creator reputation and various other factors that may influence participation.

11. REFERENCES

[1] Gerard Beenen, Kimberly Ling, Xiaoqing Wang, Klarissa Chang, Dan Frankowski, Paul Resnick, and Robert E. Kraut. Using social psychology to motivate contributions to online communities. In *CSCW '04*, pages 212–221.

[2] Fabricio Benevenuto, Fernando Duarte, Tiago Rodrigues, Virgilio A.F. Almeida, Jussara M. Almeida, and Keith W. Ross. Understanding video interactions in youtube. In *MM '08*, pages 761–764.

[3] Moira Burke, Cameron Marlow, and Thomas Lento. Feed me: motivating newcomer contribution in social network sites. In *CHI '09*, pages 945–954.

[4] Meeyoung Cha, Alan Mislove, and Krishna P. Gummadi. A measurement-driven analysis of information propagation in the flickr social network. In *WWW '09*, pages 721–730.

[5] Munmun De Choudhury, Hari Sundaram, Ajita John, and Dorée Duncan Seligmann. What makes conversations interesting?: themes, participants and consequences of conversations in online social media. In *WWW '09*, 331-340.

[6] Munmun De Choudhury, Hari Sundaram, Yu-Ru Lin, Ajita John, and Dorée Duncan Seligmann. Connecting content to community in social media via image content, user tags and user communication. In *ICME 2009*, pages 1238–1241.

[7] Marc Dixon and Vincent J. Roscigno. Status, networks, and social movement participation: The case of striking workers. *American Journal of Sociology*, 108(6):1292–1327, 2003.

[8] Elisabeth Joyce and Robert E. Kraut. Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication*, 11:2006, 2006.

[9] Cliff Lampe and Erik Johnston. Follow the (slash) dot: effects of feedback on new members in an online community. In *GROUP '05*, pages 11–20. ACM.

[10] Oded Nov, David Anderson, and Ofer Arazy. Volunteer computing: a model of the factors determining contribution to community-based scientific research. In *WWW '10*, pages 741-750.

[11] Fan Qiu and Yi Cui. An analysis of user behavior in online video streaming. In *VLS-MCMR '10*, pages 49–54.

[12] Vivek K. Singh, Ramesh Jain, and Mohan S. Kankanhalli. Motivating contributors in social media networks. In *WSM '09*, pages 11–18.