INVITED PAPER

Multimedia Semantics: Interactions Between Content and Community

Research on pattern analysis and monitoring of web-based social communities is reviewed in this paper; the authors study how media-rich social networks provide insight into multimedia research problems.

By Hari Sundaram, *Member IEEE*, Lexing Xie, *Member IEEE*, Munmun De Choudhury, Yu-Ru Lin, and Apostol Natsev

ABSTRACT | This paper reviews the state of the art and some emerging issues in research areas related to pattern analysis and monitoring of web-based social communities. This research area is important for several reasons. First, the presence of near-ubiquitous low-cost computing and communication technologies has enabled people to access and share information at an unprecedented scale. The scale of the data necessitates new research for making sense of such content. Furthermore, popular websites with sophisticated media sharing and notification features allow users to stay in touch with friends and loved ones; these sites also help to form explicit and implicit social groups. These social groups are an important source of information to organize and to manage multimedia data. In this article, we study how media-rich social networks provide additional insight into familiar multimedia research problems, including tagging and video ranking. In particular, we advance the idea that the contextual and social aspects of media are as important for successful multimedia applications as is the media content. We examine the interrelationship between content and social context through the prism of three key questions. First, how do we extract the context in which social interactions occur? Second, does social interaction provide *value* to the media object? Finally, how do social media facilitate the repurposing of shared content and engender cultural memes? We present three case studies to examine these questions in detail. In the first case study, we show how to discover structure *latent* in the social media data, and use the discovered structure to organize Flickr photo streams. In the second case study, we discuss how to determine the *interestingness* of conversations—and of participants—around videos uploaded to YouTube. Finally, we show how the analysis of visual content, in particular tracing of content remixes, can help us understand the relationship among YouTube participants. For each case, we present an overview of recent work and review the state of the art. We also discuss two emerging issues related to the analysis of social networks—robust data sampling and scalable data analysis.

KEYWORDS | Emergence; large scale; semantics

I. INTRODUCTION

Social media platforms including Facebook, Twitter, and Digg have made it easy to upload, tag, share, and interact with content as well as to communicate with other users. As a specific outcome, the ease of sharing and communication has led to rapid emergence and dissemination of cultural memes. The information from these social media platforms—about individuals, their interactions on the social network, and the social structures to which they belong—is an invaluable resource for understanding complex online social phenomena.

Over the past 40 years, traditional methods of studying social processes including information diffusion, expert identification, or community detection have focused on longitudinal studies of relatively small groups. The

Manuscript received June 17, 2011; revised December 20, 2011; accepted March 4, 2012. Date of publication June 1, 2012; date of current version August 16, 2012.

H. Sundaram is with the School of Arts Media and Engineering, Arizona State University, Tempe, AZ 85281 USA (e-mail: hari.sundaram@asu.edu).
L. Xie was with IBM Research, Hawthorne, NY 10532 USA. He is now with the Australian National University, Canberra, ACT 0200, Australia.
M. De Choudhury is with Microsoft Research, Redmond, WA 98052 USA.
Y.-R. Lin is with Northeastern University, Boston, MA 02115 USA.
A. Natsev was with IBM Research, Hawthorne, NY 10532 USA. He is now with Google Research, Mountain View, CA 94043 USA.

Digital Object Identifier: 10.1109/JPROC.2012.2191529

widespread use of social websites like Facebook, Twitter, Digg, Flickr, and YouTube has provided new avenues for researchers to study social processes at *very large scales*.[1] Social networks have made application programming interfaces (APIs) available to researchers to access user-generated content and user interactions. We can now acquire and store electronic data for very large populations over extended intervals. The result is that studies of social processes on a scale of millions of people—inconceivable just a decade ago—are becoming routine. As a specific example, consider the problem of diffusion of information. Now, we can conduct large-scale empirical studies on diffusion on many different kinds of activities, including blog posts [1], Internet chain-letter data [2], social tagging [3], Facebook news feed [4], and online games [5].

The goal of this paper is to analyze human activity on or around multimedia objects, including images and video, in online social networks. The analysis will provide us with contextual cues to better understand the meaning of multimedia objects.

A. Multimedia Semantics

Multimedia, including images, audio, and video, has become an increasingly popular medium to archive and to communicate about our social and professional lives. Multimedia data are forecast to occupy more than 60% of global consumer IP traffic by 2015 [6]: Multimedia is quickly becoming an important medium of communication in the online world.

Effective organization and search of the multimedia content remains a major challenge, despite significant work by the multimedia and computer vision communities (see [7] for a review). There are several reasons why effective search and organization of multimedia content is hard. *First*, the multimedia semantics space is unlimited.[2] By "semantic space," we mean the space of meanings associated with online multimedia content. The familiar adage "a picture is worth a thousand words" captures this sentiment. We cannot effectively describe rich multimedia objects including video with a few words—a video requires a rich text description to capture its meaning. *Second*, an examination of videos in online repositories reveals a scarcity of tags—many have no tags, while a small fraction have a few tags. This is important: Text-based search is the dominant mechanism of multimedia information search. The relative scarcity of tags precludes effective video search based on textual descriptors. *Third*, the semantic space is not only large, but it is also constantly evolving. Consider, for example, the popular object "iPhone"—this word and its associated meaning began to form only when it was introduced in 2007.

Multimedia and computer vision researchers have long explored words and their relationships to visual content. Existing studies on multimedia semantics related to visual content have typically focused on specific visual categories: objects and scenes. Among the most researched visual categories of interest include faces [9], generic objects [10], scenes [11], and landmarks [12]. Relationships among visual categories are deemed important to help recognition, including region-level hierarchies [13], small tree taxonomies for video retrieval [14], and special-purpose medium-scale visual ontologies (in hundreds) for the broadcast news domain [15].

In principle, the scale of the multimedia data—available from social networks—should ease the task of learning object and scene classifiers. Even uncommon concepts would appear in enough photos and videos in online networks to develop robust classifiers. If we could develop robust classifiers for enough concepts, then text-based search of multimedia objects would be significantly enhanced.

In practice, things are complicated. One cannot effectively use all photos tagged with the same keyword as the training set for a concept classifier. Learning a concept classifier on a set of images tagged with the same keyword implicitly assumes that the photographs share the same context in which the keyword is meaningful. While this may be true in a carefully designed dataset (e.g., the Corel dataset), this is a strong assumption on Flickr, a popular photograph-sharing social network. In general, the context in which the keyword makes sense to the photo is only known to the photo author or annotator. On Flickr, for example, there are thousands of photographs labeled as "Yamagata"—some are of the town, some refer to a visual artist (Hiro Yamagata), while still others refer to a singer (Rachel Yamagata). Unsurprisingly, concept classifiers that train concept classifiers by using images from online repositories tagged with the same keyword perform poorly on unseen photos. That is, we cannot successfully classify a photo whose context is unknown.

To support effective search and organization of multimedia content, we need to identify the context in which a multimedia object exists. Media available on social networks, including Flickr and YouTube, are associated with rich context. A shared media item is associated with a variety of information, including the identity of the person who uploaded it, associated tags, identities of people who commented on the item, and the number of times it is viewed or marked as a "favorite." These user actions, including "upload," "tag," and "comment," are timestamped.

A media item such as a photograph on Flickr therefore exists as part of a meaningful interrelationship among several attributes including time, visual content, users, and actions. The semantics of media objects as well as human activity on social media platforms needs to be understood as a relationship between people, actions, artifacts, and

[1] Facebook, for example, has about 800 million users as of November 2011.

[2] See [8] for an article on an attempt to bound the number of semantic concept detectors needed for video retrieval.

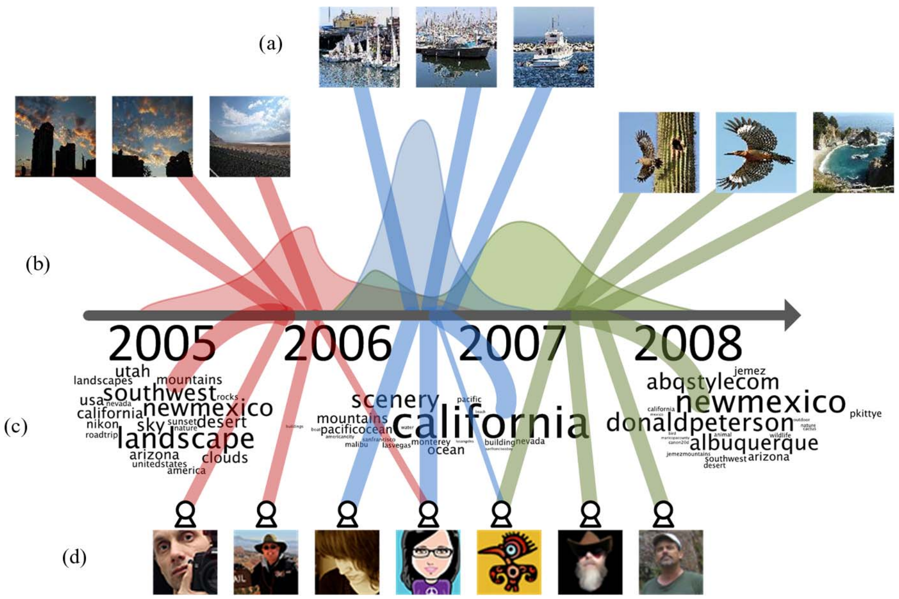


Fig. 1. ***Relational structure in social media streams, which reveals the strong relationship among multiple facets: (a) photos (visual content), (b) time, (c) tags, and (d) users. This figure presents partial multirelational structure extracted from the data of Flickr group "The Southwest United States." It illustrates three major themes in the group photo stream: "landscape," "California," and "New Mexico." Users principally contributed to these themes during three time frames 2005–2006 (landscape), 2006–2007 (California), and 2007–2008 (New Mexico).***

supportive contextual metadata. Fig. 1 illustrates the relationships among visual content, time, tags and users.

B. Semantics Arising From Social Interaction

The semantics that arise from social interaction, including commenting, sharing, and tagging, around media objects—denoted in this paper as *interaction semantics*—are distinct from the semantics of the media object. Rather than asking "what is the meaning of this photo or video?" *we seek the semantics of the relationship* between people, actions, and media.

An example is useful to illustrate the difference: A Flickr group on "Arizona Travel" may have a lot of posts on Sedona, a popular destination, in July from people who live in Phoenix but travel there to escape the heat. There are fewer posts in December, when it is cold in Sedona. Now, even if the meaning of each individual photo is known, the meaning of the relationship between location (Sedona), time (summer), specific users, and photo colors is not explicit in the data. This relationship may exist because the active members of the group are friends who live in Phoenix and plan an annual summer retreat together in Sedona. In other words, the relationship—among photo visual features, photo capture time, tagging, and commenting on the photo—arises due to human activity, both online and in the physical world. In this case, the *interaction semantics*—the meaning of the relationship—while not explicit, are known only to the group members. These semantics cannot be easily discovered by accessing the photo stream via a single object or attribute (e.g., photo tags) or through a simple aggregation of attributes. The discovery of latent structure in such social media platforms can point to emergent cultural behaviors. Interestingly, these behaviors may not even be explicitly identifiable by members of the network.

Characteristics of social network data preclude simple representations of the social context. First, social media data typically involves *multiple social relations*. In Flickr for example, there are several relations including user-to-user relation (friendship or commenting), user-to-photo relation (tagging or "like"), photo-to-location relationship, and photo-to-time relationship. Second, the activity patterns in social media often reflect not just a single user's routine behavior and interests, but the community structure. In Facebook and Twitter, for example, communities emerge due to users' topical interests and collaboration on projects. In Flickr, media and ideas are shared within communities of friends. Understanding social media content requires knowledge about communities which carry relevant context. The specific knowledge is important because the process of content creation and sharing in social media is driven by community activities and interests.

C. Key Questions

We shall focus on three key questions, answers to which will help us understand and use social context in multimedia research.

1) How do we extract context in which social interactions occur? We would like to discover an

interpretable structure in social media streams. Specifically, how do communities emerge in online social networks? When do new associations between tags and photos emerge? We believe that latent interaction structure can facilitate an effective exploration and summarization of social media.

2) Does social interaction provide *value* to the media object? The value must arise solely due to the social interaction around the object, outside of any intrinsic media semantics.

3) How does social media facilitate the repurposing of shared content and engender cultural memes? What does the presence of memes tell us about the different roles—content creation, repurposing, and reposting—played by the members of the network?

We shall address these three questions through three case studies. In the first case study, presented in Section II-A, we show how to discover structure *latent* in the data. The structure relates people, actions, and content on Flickr. Then, we show how we can use the discovered structure to organize Flickr photo streams. In the second case study, Section II-B, we discuss how to determine the *interestingness* of conversations—and of participants—around videos uploaded to YouTube. The conversational interestingness is a very different way of ranking video content, not based on visual content, but instead driven by social interaction around the content. Finally, in Section II-C, we show how analysis of visual content—tracing content remixes, in particular—can help us understand the relationship among YouTube participants.

D. Example Applications

The discovery of relational semantics can significantly influence applications in multimedia and other domains, including content organization, recommendation algorithms, and social network analysis.

The rapid growth of content on social media sites creates several interesting challenges. First, the content in a photo stream (either for a user or a community) is typically organized using a temporal order, making the exploration and browsing of content cumbersome. Second, sites including Flickr provide frequency-based aggregate statistics including popular tags and top contributors. Users can access a subset of the content by clicking on these tags/contributors. However, these aggregates do not reveal the rich relational structure inherent in the community sharing and interaction. As discussed by Shamma *et al.* [16], harnessing the contextual information for media understanding is one of the most difficult challenges in media pragmatics/applications.

Social media metadata can be used to improve media retrieval. Existing research on tagging includes improving tag recommendation [17], [18] and analyzing usage patterns of tagging systems [19]. Negoescu and Gatica-Perez [19] present a large-scale analysis of Flickr groups and propose a topic modeling approach for representing a group based on the co-occurrence of groups and tags. Zunjarward *et al.* [20] propose a framework for annotating events in images by exploiting the social networks of annotators. Kennedy *et al.* [21] propose a framework for generating knowledge—representative tags—for a location, and for extracting place and event semantics for a tag. Their work suggests that community-generated media and tags can improve access to multimedia resources.

The multirelational structure can be used to provide effective recommendations along any attribute [22]. When the user is looking at a particular photo, we could use the set of relations in which the photo exists, and then recommend other photos, tags, and related peers. The multirelational data can provide additional context over the (photo, tag) pairs that have been used to recommend tags in automated annotation algorithms. It can help identify peers and context (including feature distributions, activities, time) in which they are related to the current user.

In [23], for example, we describe methods for finding users the right communities of interests in order to gain useful feedback on their uploaded content. A recommendation framework based on learning a latent space representation of the groups is developed to recommend the most likely groups for a given image. Based on the same crawl of Flickr dataset, we observe that: 1) the tagging and communication-based features of images help improve recommendation performance significantly against image content alone; and 2) groups (the photo-sharing communities on Flickr) can be effectively represented by their features (image content, tags, and communication activity) in a latent space which is useful for recommending more interesting groups. The potential of community-contribution information in multimedia collection has been discussed in [22], including areas important for multimedia access: annotation, distribution, and retrieval. In [24], the authors propose a similar joint Nonnegative Matrix Factorization framework to leverage a secondary source to improve retrieval performance from a primary dataset. The effectiveness of the proposed method is demonstrated through a social media retrieval application.

We can exploit the communication characteristics in social media sites to make predictions on user behavior, sales, and stock market activity [25]–[28]. Antweiler *et al.* [29] determine correlations between communication activity in Internet message boards and stock volatility and trading volume. Gruhl *et al.* in [27] correlate postings in blogs, media, and webpages with the sales ranks of books on Amazon.com. They devise carefully handcrafted queries to find matching posts which can be indicators of future sales ranks of books. Similarly, in [26], we analyzed the communication dynamics (of conversations) in a technology blog and used it to predict stock market movement.

E. Emerging Research Areas

We need several computational elements to understand the nuanced patterns of linking among individuals and communities that occur through social media, at different structural levels of interaction. There are several nontrivial research challenges associated with analyzing social network datasets, and we highlight two of them: data bias and data diversity.

The first challenge is a basic methodological question: how to sample social network datasets to ensure validity of the analysis. The datasets relevant to the analysis are enormous in size, diverse in form and content, and are growing rapidly. Importantly, researchers are limited in the ability to acquire the data due to an information acquisition bottleneck: The social network APIs restrict the amount of data that can be retrieved per minute. This acquisition rate is several orders of magnitude smaller than the rate of production of information in the network.

The second challenge is associated with scalable storage and analysis of data—the data exhibits temporal evolution and significant diversity. In social networks, user interactions and community interests are constantly evolving, often tracking real-world events. Social media data are also multifaceted: typically involving multiple types of relationships including friendship, and co-commenting on a news story. Entities in social networks may also have different attributes, e.g., location, age, profession. The multidimensional and multirelational nature of these interactions increases the computational complexity of the algorithms. Consequently, a framework for managing user data in a form amenable for large-scale data analysis is a key step in supporting significant technical advances in social media. The computational ceiling arising due to finite resources requires us to pursue innovative strategies to address the data scalability challenges.

In this paper, we would like to understand how media-rich social networks can provide additional insight into familiar multimedia research problems. In particular, we advance the idea that the contextual and social aspects of media semantics are as important for successful multimedia applications as is the media content. An important idea is that a framework for extracting useful information from social media data needs to therefore scale not only with the data scale, but also against diversity of data facets. We shall explore these ideas in detail with several case studies.

The rest of this paper is structured as follows. In Section II, we present three case studies. In Section III, we discuss related work. In Section IV, we discuss open issues related to sampling and scalable analysis. Finally, in Section V, we present our conclusions.

II. CASE STUDIES: SOCIAL ACTIVITY AROUND MEDIA OBJECTS

In this section, we present three case studies that allow us to shed light on the three key questions outlined in Section I-C. In particular, we examine the extraction of latent social network structure, deriving value for media objects accruing from social interaction, and finally develop a deeper understanding of the social media participants. The first two studies analyze the social dynamics on a social network *around* a rich media object, whereas the last study analyzes media content to understand the relationship among participants.

A. Case Study: Understanding Collective Human Activity Through Multirelational Structure Discovery

We present a method for discovering multirelational structures from social media streams on Flickr, a popular social media site. Example user–photo relationships in this structure include photo tags, photo sharing, commenting, and so on. These relations encode the interaction semantics that are only interpretable to the participants of the social network. That is, the *explanation* for the existence of a stable relationship between a specific set of people, location, time, photos, and tags, while known to the users, may not be explicitly encoded in the data. [30], [31]. Our goal is to extract relational structures within a group of online users and their shared content, through *soft clustering* that reflects these relations. As will be shown in the rest of this section, such structure will lead to richer interpretation of the data, as well as better tagging algorithms for photos.

1) *Dataset:* We extract relational structure from Flickr group photo pools.[3] We define a *group photo stream* (or group, for short) to be a collection that includes: photos posted in a *shared space*, all users who posted the photos, and tags associated with the photos. In this paper, the shared space specifically refers to Flickr group pools. Our framework extends to other shared spaces including Facebook groups and event repositories such as Eventful or Upcoming.

We use the Flickr API[4] to collect data from a sample of 120 Flickr groups. The distribution of group size in our sample follows the overall group size distribution on Flickr. We download all publicly available photos for each group. Our dataset consists of 111 108 photos, 8117 unique users, and 102 607 unique tags in total. The photo post times range from January 1, 2004 to January 8, 2009, which enables us to analyze long-term temporal patterns in this collection.

2) *Methodology:* There are two key ideas in our framework: extraction of relations from social media streams and extraction of relational clusters.

Extraction of relations in social media streams: Relations connect different aspects of the photo stream data. Specifically, this work models visual content,

[3] http://www.flickr.com/groups/
[4] http://www.flickr.com/services/api/

associated tags, photo owners, and post times. We analyze the relationship between groups of objects of the same type. We call a set of entities of the same type a *facet*. A photo facet, for example, is a set of photos; a user facet is a set of users. We call the interactions among items in different facets a *relation*. A relation can involve two (i.e., binary relation) or more facets. In this work, we investigate pairwise relations; our method, however, can be extended via tensor-based representations, to handle higher-order relations [31].

We use matrices to capture the relationship between any two facets. We represent the visual facet with a number of visual descriptors that have been found effective in image content analysis, including color (color histogram and moments), texture [32], and shape [33], [34], as well as interest points [35]. We concatenate the visual features to form a $L = 1064$-dimensional vector for each photo. We normalize the features to lie within [0, 1] with a logistic function. Let $\mathcal{P}$ be the set of photos in a group. We obtain a photo-feature matrix $\mathbf{W}^{(F)} \in \mathbb{R}^{|\mathcal{P}|\times L}$, where the ith row is the feature vector of the ith photo. Here, the photo-feature matrix $\mathbf{W}^{(F)}$ represents the relation between photos and the their visual features. We use three facets to encode contextual features: ownership, tags, and photo posting time. Let $\mathcal{U}$ be the set of users who post photos to the group, i.e., photo owners. In a similar manner, we can construct a photo-user matrix $\mathbf{W}^{(\mathcal{U})} \in \mathbb{R}^{|\mathcal{P}|\times|\mathcal{U}|}$, where each entry $\mathbf{W}_{ij}^{(\mathcal{U})} = 1$ if the ith photo is posted by the jth user, and 0 otherwise. Matrices $\mathbf{W}^{(Q)}$ and $\mathbf{W}^{(T)}$, representing photo-tag and photo-time relationships, are defined in a similar manner. The matrix-based relational data model can easily incorporate additional context. It is possible, for example, to add the EXIF metadata from photos and include location, camera model, and settings. The EXIF metadata can be represented in a manner similar to the basic contextual information discussed in this section. Fig. 2 shows the four relation matrices mentioned above.

Extraction of relational clusters: We formulate the extraction of relational clusters as an optimization problem. The optimization objective is to find a set of *soft clusters* that best represents various simultaneous relations between the photos and other facets such as visual features, associated tags, photo owners, and post times. In this setup, we assume that an entity, including a photo, a tag, or a user, can belong to multiple clusters, with weights that indicate the likelihood of membership. The cluster assignment ensures that the observed relationship between entities is well approximated by the soft-assignment to the set of clusters.

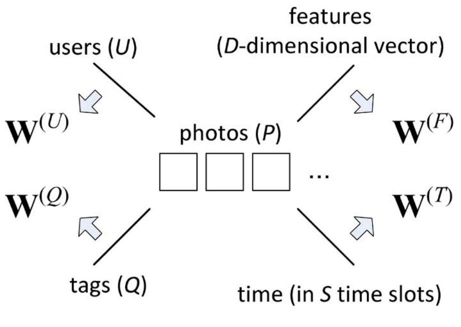


Fig. 2. ***Data of the group photo stream over time can be represented as four matrices: photo-feature matrix*** $\mathrm{W}^{(F)}$***, photo-user matrix*** $\mathrm{W}^{(\mathcal{U})}$***, photo-tag matrix*** $\mathrm{W}^{(Q)}$***, and photo-time matrix*** $\mathrm{W}^{(T)}$***.***

We find soft relation clusters with nonnegative joint matrix factorization techniques. We now discuss how to factor a single relation, photo-visual feature matrix $\mathbf{W}^{(F)}$—the others follow analogously. We represent each cluster k, $k = 1, \ldots, K$, with a feature vector $\mathbf{z}_k$ of length L. We define p_{ik} to be the probability that a particular photo i belongs to the kth cluster and define λ_k to be the cluster probability. Our goal is to determine the coefficient z_k based on the likelihood that a photo i belongs to the kth cluster. Here $\vec{Z}^{(F)} = \{z_k\}$ is a $K \times L$ matrix, $\vec{P} = \{p_{ik}\}$ a $|\mathcal{P}| \times K$ matrix, and $\boldsymbol{\Lambda} = \{\lambda_k\}$ a $K \times K$ diagonal matrix. $\boldsymbol{\Lambda}$ is shared across all considered relations and indicates the size of each obtained cluster. The soft clustering is a good one if z_k, weighted by $\vec{P}$ and $\boldsymbol{\Lambda}$, approximates the ith row in $\mathbf{W}^{(F)}$

$$\mathbf{W}_i^{(F)} \approx \sum_k \lambda_k p_{ik} z_k = \left(\mathbf{P}\boldsymbol{\Lambda}\mathbf{Z}^{(F)}\right)_i. \qquad (1)$$

We can approximate $\mathbf{W}_i^{(F)}$ by minimizing a cost measure $D(\mathbf{W}^{(F)}\|\vec{P}\Lambda\vec{Z}^{(F)})$, where $D(\cdot\|\cdot)$ is a measure of approximation cost between two matrices. We use Kullback–Leibler (KL) divergence between two matrices.[5] The KL divergence is a natural measure of the dissimilarity between two distributions. Hence, to obtain $\vec{P}$, $\boldsymbol{\Lambda}$, and $\vec{Z}^{(F)}$, we minimize the following objective function:

$$J\left(\mathbf{P}, \boldsymbol{\Lambda}, \mathbf{Z}^{(F)}\right) = D\left(\mathbf{W}^{(F)}\|\mathbf{P}\boldsymbol{\Lambda}\mathbf{Z}^{(F)}\right). \qquad (2)$$

All the objective functions—with respect to different facets—can be written out similarly and combined to minimize the overall cost

$$J\left(\mathbf{P}, \boldsymbol{\Lambda}, \left\{\mathbf{Z}^{(r)}\right\}\right) = \sum_{r\in\{F,U,\ldots\}} D\left(\mathbf{W}^{(r)}\|\mathbf{P}\boldsymbol{\Lambda}\mathbf{Z}^{(r)}\right),$$
$$\text{s.t.} \quad \mathbf{P} \in \mathbb{R}_+^{|\mathcal{P}|\times K}, \boldsymbol{\Lambda} \in \mathbb{R}_+^{K\times K}, \mathbf{Z} \in \mathbb{R}_+^{K\times I_r},$$
$$\sum_i \mathbf{P}_{ik} = 1 \quad \forall k, \sum_k \boldsymbol{\Lambda}_k = 1 \qquad (3)$$

[5] Using matrices to represent distributions, the KL divergence between matrices $\vec{A}$ and $\vec{B}$ is defined by $D(\vec{A}\|\vec{B}) = \sum_{ij}(\vec{A}_{ij}\log\vec{A}_{ij}/\vec{B}_{ij} - \vec{A}_{ij} + \vec{B}_{ij})$, where $\sum_{ij}\vec{A}_{ij} = \sum_{ij}\vec{B}_{ij} = 1$.

where $Z^{(r)}$ is the matrix representing the rth relation and I_r denotes the dimensionality of the second dimension of the coefficient matrices. The joint objective function can be easily extended to incorporate additional relations or to incorporate weights on the relations or facets. We develop an iterative algorithm for minimizing the objective function in (3). We also automatically determines the number of clusters by introducing structure cost penalty. Algorithm details can be found in the full paper [31].

3) Experimental Results: We now discuss our experimental results. Fig. 3 shows the results from two Flickr groups: "sky's the limit" (denote as group A, with 2278 photos), and "Full Frame Sensor group" (denote as group B, with 3961 photos). We extract the clusters using the joint optimization in (3), obtaining five thematic clusters in group A and six in group B. For each thematic cluster, we show the top three photos based on the data likelihood p_{ik} for a photo i to belong to theme k. We can determine through other coefficient matrices the most likely users who post photos belonging to the theme and the most likely tags associated with the theme photos. The middle part of Fig. 3 shows aggregated cluster strength over time for group A and group B. We can see that the theme strengths vary over time; some themes, such as A2 and B2, only appear at certain time periods and then diminish. Some others—A3 and A5 are examples—appear, then fall and then reappear. We have observed that these themes emerge due to dedicated users (e.g., the "bird" images in A4 are taken by the same user), tag co-occurrences (e.g., "sunset" in A2, "water" in B6, etc.), as well as similar visual content (e.g., A2, A4, A5, B2, B5, B6, etc.). These empirical results suggest that our analysis captures the dynamics of group patterns and gives meaningful summary of group photo streams.

How can we quantify the meaningfulness of an extracted structure? We designed a prediction task to examine this question. The intuition is that if our algorithm captures the relational structure, it should be able to predict missing data tuples from the same group photo stream. Specifically, we use the photo-tag relation in our prediction task: Our prediction task is to predict tags for a new photo. Our prediction results outperform baseline methods such as feature and tag-frequency-based methods by 44% $\sim$ 390% on precision at 10 (P@10) [31].

The success of our prediction framework may be attributed to the "event locality" in Flickr photos: Many photos are well correlated to either global events or to events that are directly observed by the users. This implies that the use of tags is highly correlated to the event context. The event context includes information related to a particular user, time, and visual appearance. The relational data model helps to capture the event context.

We conducted a pilot user study with 12 participants to understand the utility of using Flickr-group relational semantics to organize Flickr photos. To conduct the study, we developed an interactive interface to present the results of thematic cluster extraction [31]. The user study used both 5-scale rated survey questions as well as semi-structured interviews. The user study results indicate that users found the extracted clustering results to well represent the major group themes. Furthermore, user study participants pointed out that the clustering results not only reveal how users describe the group data, but also lead users to discover evolution of the group activity.

Our work improves annotation accuracy by introducing local or temporal context in data analysis. Furthermore, the algorithm proposed in our work can be easily extended with additional facets to transfer knowledge from readily available auxiliary data.

B. Case Study: Characterizing Interestingness of Conversations on Social Media Sites

In this case study, we examine a simple question: Does the presence of social interaction—facilitated by a social media site—around a media object add to the value of the media object? The emergence and growth of social media websites such as YouTube and Flickr has created new opportunities for individuals to share rich multimedia objects online. A striking feature of these sites is the extensive interaction among community participants: community members return[6] repeatedly—not to watch the video again—but to *participate in the discussion around the video*.

A video uploaded on YouTube, for example, may generate many comments around the video. These comments often include conversational structure—back-and-forth comments between several community members—akin to a *conversational thread*. The theme of the conversation is latent; it is not only affected by the video, but also by the content of the comments.

There are two key characteristics of these conversations: conversation theme and how they can affect the meaning of the media object that resulted in the conversation. First, over time, the conversation will drift to topics unrelated to the video content. The change in topic affects participation: A flame war may dissuade participation, whereas a thoughtful, respectful dialog may encourage people to voice their thoughts. Second, the comments provide additional context to the video. Therefore, the semantics associated with the video—by the participants in the conversation—will evolve to accommodate the new context. Comments on a music video, for example, may provide context about the band and the circumstances under which the video was recorded, which is not obtainable from watching the video. Keith Jarrett's well-known recording—the Köln concert, for example—is known not only for the music,

[6] Social media sites, including YouTube, encourage participation though notification. There are several cases when users are notified: Users are notified of activity on their uploaded content; users who comment are notified of replies on the video.

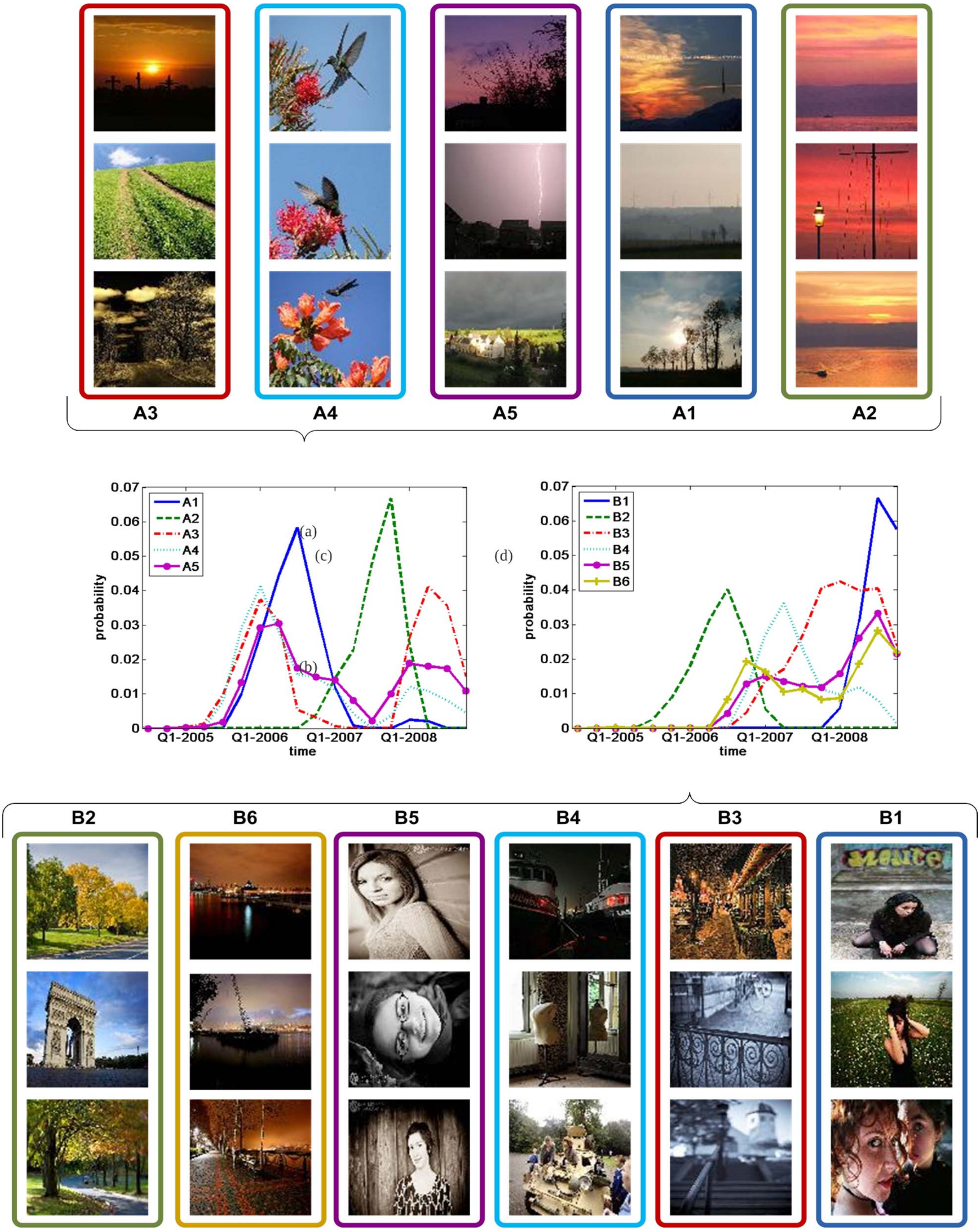


Fig. 3. *Theme discovery for Flickr group A "sky's the limit" (top) and group B, "Full Frame Sensor group" (bottom). Theme evolutions are shown in two middle plots, with the mid-left plot from "sky's the limit" and the mid-right plot from "Full Frame Sensor group." Themes were obtained using joint analysis in (3). The results show that group patterns emerge due to dedicated users, tag co-occurrences, as well as similar visual content.*

but also for the recording context: The performer was experiencing severe pain while performing.

What properties, including properties of the conversation and contextual metadata, prompt community members to participate in a conversation? A measure that captures properties correlated with participation would be of value in different applications. First, we can use the measure to rank and filter both blog posts and rich media. The same news story, for example, may be posted on several blogs. A conversational measure can be used to identify sites where the postings and commentary have the highest likelihood of participation. Second, the conversational measure can also be used to increase efficiency. We can cache media objects associated with conversations with high measure.

We shall call our measure of a conversation, its "interestingness."[7] We recognize the pitfalls of using a word that connotes subjective engagement with an objective measure. However, the word "interesting" appears to best describe a measure for a conversation that correlates with increased participation—the phenomena of users returning to participate in the conversation associated with a YouTube video for example.

Our insight is that interesting conversations have an engaging theme with interesting people [36]. We propose a computational framework to determine an objective measure of conversational interestingness. One by-product of our work is that we can also compute a measure of "interestingness" of participants.

The framework has three innovations: extraction of conversational themes, a measure of interestingness, and a method to cross-validate the interestingness measure. We extract conversational themes for determining the "interestingness" of online conversations. A theme is either visual (content features of the associated media object) or textual (topical assignment based on the comment contents). Fig. 4 shows example conversational themes from a YouTube dataset. We detect visual themes via content features of the associated media object. We detect textual themes using a sophisticated mixture model approach. We use a random walk model for characterizing the communication properties of participants of conversations. Then, we use a novel joint optimization framework—that utilizes the themes and the communication properties of participants—with temporal smoothness constraints to compute the interestingness measure. A framework to compute the interestingness of a conversation is not enough—we need to show that the measure is meaningful. Therefore, we examine the *consequence* of a conversation with high interestingness measure.

1) YouTube Dataset: In our experiments, we crawled YouTube, a popular video-sharing site, to create a dataset comprising 272 810 videos, involving 17 736 361 unique participants and 145 682 273 comments. We collected the dataset during 2008–2009.[8] We looked at several broad categories on the YouTube website: "Comedy," "Education," "Entertainment," "News & Politics," "Sports," "Music." For each video, apart from downloading the video, we collected contextual metadata: timestamp, tags, associated set of comments and their timestamps, and authors.

2) Methodology: Let us assume that we have a set $\mathcal{C}$ of conversations and a set $\mathcal{P}$ of participants who have posted comments on any conversation $c_i \in \mathcal{C}$, and metadata associated with comments and the media object. Our goal is to determine a measure of interestingness for each conversation $c_i \in \mathcal{C}$. The interestingness measure of a conversation—at each point in time—is a nonnegative scalar. Determining interestingness of conversations involves three key challenges: extracting conversational themes, deriving the interestingness measure, and cross-validating the measure of interestingness.

Conversational themes: Participation in a conversation is influenced by the video object content and the comments. To represent visual content, we use a number of well-known features. These features include color (color histogram and color moments), texture (GLCM and phase symmetry), shape (radial symmetry and phase congruency), and interest points (SIFT).

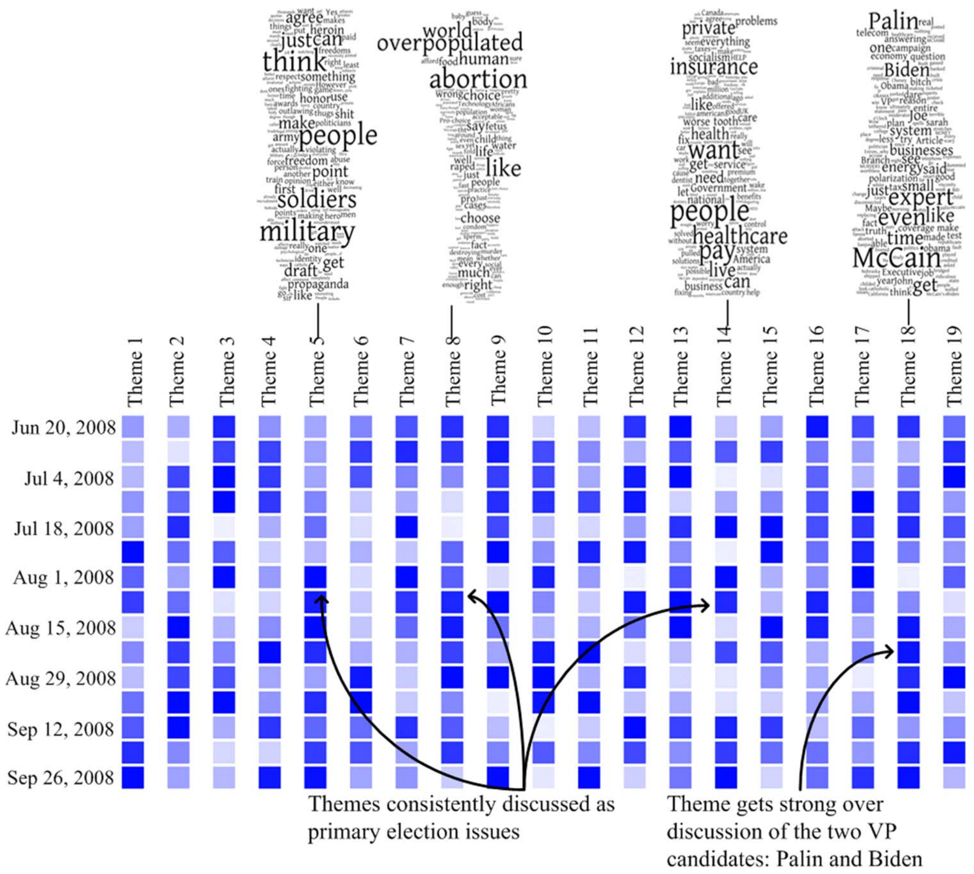
Conversations are growing collections of comments from different participants. Interestingness of a conversation—at any point in time—depends on the content of the comments associated with it. Hence, we propose a mixture model for the conversational themes. The mixture model is regularized over time and over the participants.

We temporally segment each conversation into nonoverlapping chunks (or bag-of-words). Each chunk corresponds to one time slice. We assume that words in a chunk are generated from K multinomial theme models with latent distributions. The theme model parameters are computed by maximizing the data likelihood.

We need to regularize themes associated with words in a chunk with respect to time [37] and coparticipation. Temporal regularization is important: A word, for example, can become highly popular at a specific time due to related external events. The intuition to regularize conversational theme distributions by coparticipation is as follows: If two chunks share participants, the chunk theme distributions will be more similar than the case when the two chunks share no participants. We define a participant co-occurrence graph $G(V, E)$ where each vertex is a conversation $c_i \in \mathcal{C}$ and an undirected edge $e_{i,m}$ exists between two conversations c_i and c_m if they share at least one common participant. Each edge $e_{i,m}$ is associated with

[7] We are motivated by Flickr's proprietary measure of "interestingness" for photographs uploaded to the site. Flickr's algorithm to compute the interestingness of a photo is unpublished.

[8] A large portion of the dataset is available for download at: http://www.public.asu.edu/mdechoud/datasets.html.



Themes consistently discussed as primary election issues

Theme gets strong over discussion of the two VP candidates: Palin and Biden

Fig. 4. ***Evolution of conversational themes on the YouTube dataset: The rows represent weeks, and the columns represent 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.***

weight $\omega_{i,m}$, which is computed as the fraction of participants common to a pair of conversations. We incorporate participant-based regularization by constraining the *similarity* between two theme distributions to be proportional to the edge weight $\omega_{i,m}$.

Determining interestingness: Our main insight is that interestingness of conversations and participants *mutually reinforce* each other: Interesting people are to be found at interesting conversations. Therefore, our framework determines the interestingness measures for both participants and conversations through a joint optimization. We compute both measures through temporal recurrence relations. The interested reader can find the detailed analysis in [36].

In our framework, a participant who communicates can be in two states. In one state, she is influenced by her past history of communication, and in the other, her communication is independent of her past communication history. A two-state model helps account for participants introducing new themes in their comments, independent of their past communication history.

The participant measure of interestingness is influenced by the state of the participant. The measure—when in the state affected by the participant's prior communication history—depends on the following aspects. First is the value of the participant interestingness measure at the previous time slice. Second, the measure depends on the measure of interestingness of other participants, who were influenced to comment by the participant whose interestingness we are trying to measure. Third, the measure increases when she comments on posts by persons with high interestingness measure. Finally, the participant measure depends on measure of interestingness of the conversation—participating in a conversation with a high value of interestingness increases interestingness of the participant. The measure, when the participant is *not affected* by prior history, depends on the following aspects. The measure is influenced by the themes associated with the participants' comments. Second, the measure is influenced by conversational theme strength at the previous time slice. The theme strength includes measures for both the visual and textual themes.

The interestingness measure for a conversation can be defined in a similar manner. In the case of conversations, however, a conversation becomes interesting when it attracts participants with high measure of interestingness or when the conversational themes are engaging. The conversational measure depends on the following aspects. First, it depends on the measure of interestingness of the participants in the conversation at the previous time slice. Second, the measure depends on the strength of the themes in the conversation at the previous time slice, including the visual and the textual theme strength. We use theme strength as a proxy for theme interestingness.

We use a joint optimization framework to learn the optimal values of the model parameters to jointly maximize the interestingness measures for both participants and conversations.

Evaluating iInterestingness: We need to cross-validate the measure for interestingness—after all, there is no *a priori* reason to believe that our measure, while intuitive, is meaningful. We are guided by a simple observation: Any interesting conversation is likely to have consequences. The consequences include participant activity, increased coparticipation, and theme sustenance. Given a conversation at time t_0, we ask the following three questions, at a later time $t_0 + \delta$.

1) *Activity*: Do the participants in an interesting conversation c_i at time t_0 take part in other conversations relating to similar themes at $t_0 + \delta$?
2) *Cohesiveness*: Do the participants in an interesting conversation c_i at time t_0 exhibit cohesiveness in communication, that is, tend to coparticipate in other conversations at $t_0 + \delta$?
3) *Thematic similarity*: Do other conversations with a theme distribution similar to conversation c_i at time t_0 also become interesting at $t_0 + \delta$?

Given appropriate measures for each of the three consequences—activity, cohesiveness, and thematic similarity—a good measure of conversational interestingness ought to predict each consequence well. These measures are described in more detail in an earlier publication [36].

3) Empirical Studies: There are five measures of interestingness: two measures proposed by us and three baseline measures. Our measures are as follows: interestingness measure with temporal smoothing I_1 and interestingness measure without temporal smoothing I_2. We use three baseline measures for conversational interestingness. The first baseline interestingness measure (B_1) of a conversation is the number of comments per time slice. This measure satisfies the following two constraints [38]. First, a conversation is interesting at a time slice when it has several comments in that time slice, and second, 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 measure (B_2) is based on the idea of novelty in participation: New participants join a conversation at time t_0 and who did not appear in the same conversation at any time slice prior to t_0. The intuition behind the measure is that interesting conversations attracts new participants. The third baseline measure (B_3) ranks conversations using the PageRank algorithm on the participant co-occurrence graph. The intuition is that if participants who coparticipate on several interesting conversations also coparticipate on another conversation, then this new conversation must be appealing.

We present the results of measuring consequence of interestingness on the YouTube dataset captured by the three measures of consequence corresponding to the following: activity, cohesiveness, and thematic interestingness. To compute the consequence of an interestingness measure, we compute the mutual information of the measure of interestingness with measures for activity, cohesiveness, and thematic interestingness.

The results of evaluation are shown in Fig. 5. The mutual information between the interestingness measure and the consequence measure is affected by δ, the time difference between a point in time in the future and the current time. Hence for each interestingness measure and consequence metric, we determine the optimal δ for that pair—using training data—and use this δ to compute mutual information for the pair. We observe that our method I_1 maximizes mutual information for all three consequence metrics (mean 0.83)—our computed measure of interestingness can successfully explain the three consequences in terms of mutual information, in contrast to the baseline methods (mean 0.41).

C. Case Study: Understanding Event Dynamics With Visual Memes

This case study addresses the final challenge on understanding how social media facilitates the reuse of shared content and the roles played by the participants in the social network. We present a method for tracking "visual quotes" (i.e., memes) in social media and use the outcome of meme tracking to make sense of real-world news events as well as understand user activity.

Remixing and reposting are prevalent on video-sharing platforms like YouTube, as it has been observed that remixing is a source of "vernacular creativity" [39]. A number of studies have covered quoting and remixing on text-based social platforms, especially Twitter [40]–[42]. We note, however, that the interaction and remixing traces are not unavailable for video content since video is a linear media, and no prior approach is available for tracking interactions therein.

Analyzing large-scale video propagation enables us to study the social dimensions of rich-media sharing, especially user influence and event dynamics. Real-world event traces have been an area of considerable interest,

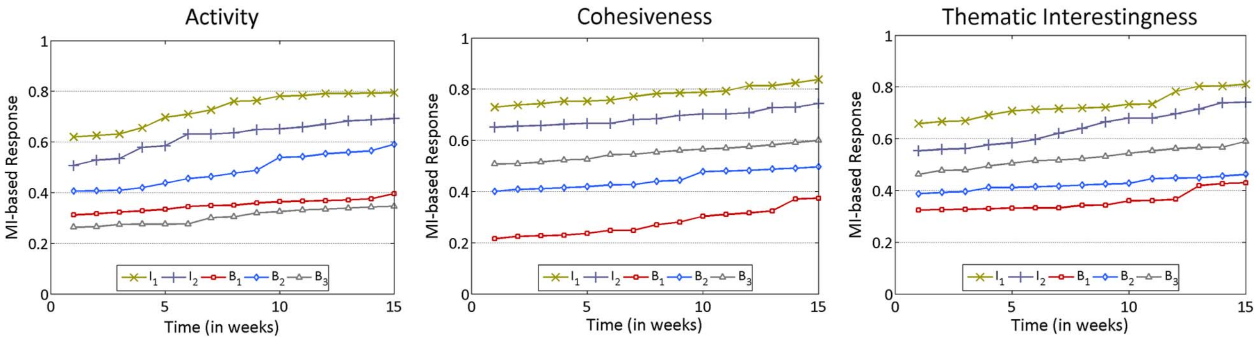


Fig. 5. ***Mutual Information between the interestingness measure and the three consequence metrics. Results are shown for two rich media-based datasets: YouTube and Flickr. We evaluate our computed interestingness I_1 and I_2 against baseline methods B_1 (comment frequency), B_2 (novelty of participation), and B_3 (coparticipation-based PageRank). Our method incorporating temporal smoothness (I_1) maximizes the mutual-information-based response metric for the three consequence-based metrics (activity, cohesiveness, and thematic diffusion).***

and several systems have covered event profiles [43]–[45] via tweet volumes and sentiments. The focus of our work is on user roles and influence.

1) Visual Memes and Video Content Duplication: We propose visual memes as a unit for analyzing visual quotes on YouTube. A *meme*[9] is defined as a cultural unit (e.g., an idea, value, or pattern of behavior) that is passed from one person to another in social settings. We define a *visual meme* as a short segment of video that is frequently remixed and reposted by more than one user. Fig. 6 shows examples of visual memes represented by keyframes. Despite variations in the videos that contain the memes (e.g., varying sizes, colors, captions, edits), each meme instance is semantically consistent.

Visual memes are shared and frequently reposted parts of remixed videos. They exist because users tend to create "curated selections based on what they liked or thought was important" [46]. News event collections are particularly suited for studying large-scale user curation since remixing is more prevalent here than on video genres designed for self-expression, such as video blogs. The unit of interaction appears to be video segments, consisting of one or a few contiguous shots. The remixed shots typically contain minor modifications that include video formatting changes (aspect ratio, color, contrast, gamma) and video production edits (the superimposition of text, captions, borders, transition effects). Most of these modifications are well known in the visual copy detection problem domain. In this paper, we will use *meme* to refer both to individual instances, visualized as representative icons (Fig. 6, top), and to the entire equivalence class of reposted near-duplicate video segments, visualized as clusters of similar keyframes (as in Fig. 6, middle).

Visual memes represent endorsements. Intuitively, remixing and reposting is a stronger endorsement requiring much more effort than simply viewing of, commenting on, or linking to the video content. A reposted visual meme is an explicit statement of awareness; a statement on a subject of mutual interest. Hence, memes can be used to study virality, lifetime and timeliness, influence, and (in)equality of references.

2) Scalable Extraction of Visual Memes: There are two main challenges in detecting visual memes in a large collection. The first challenge lies in reliably matching video segments despite the variations in their appearances. The second challenge is the overall computational complexity of the matching process. Finding all pairs of near-duplicates by matching all N shots against each other has a complexity of $O(N^2)$, which is infeasible for a typical collection containing millions of video shots. Our process for detecting video memes is outlined in Fig. 7 and summarized below; details are available in the corresponding paper [47].

Our solution to the first challenge is robust keyframe matching. We first temporally segment the video, and then extract a representative keyframe for each segment. We preprocess the keyframes by detecting and removing borders and normalizing aspect ratio. We then extract the *color correlogram* [48] for each frame to capture the local spatial correlation of colors. The color correlogram is rotation-, scale-, and, to some extent, viewpoint-invariant. The keyframe matching uses a globally tuned query-adaptive threshold to normalize the match radius based on the query frame and feature complexity.

Our solution to the scale challenge is to use an indexing scheme for fast approximate nearest neighbor (ANN)

[9] http://wordnetweb.princeton.edu/perl/webwn?s=meme

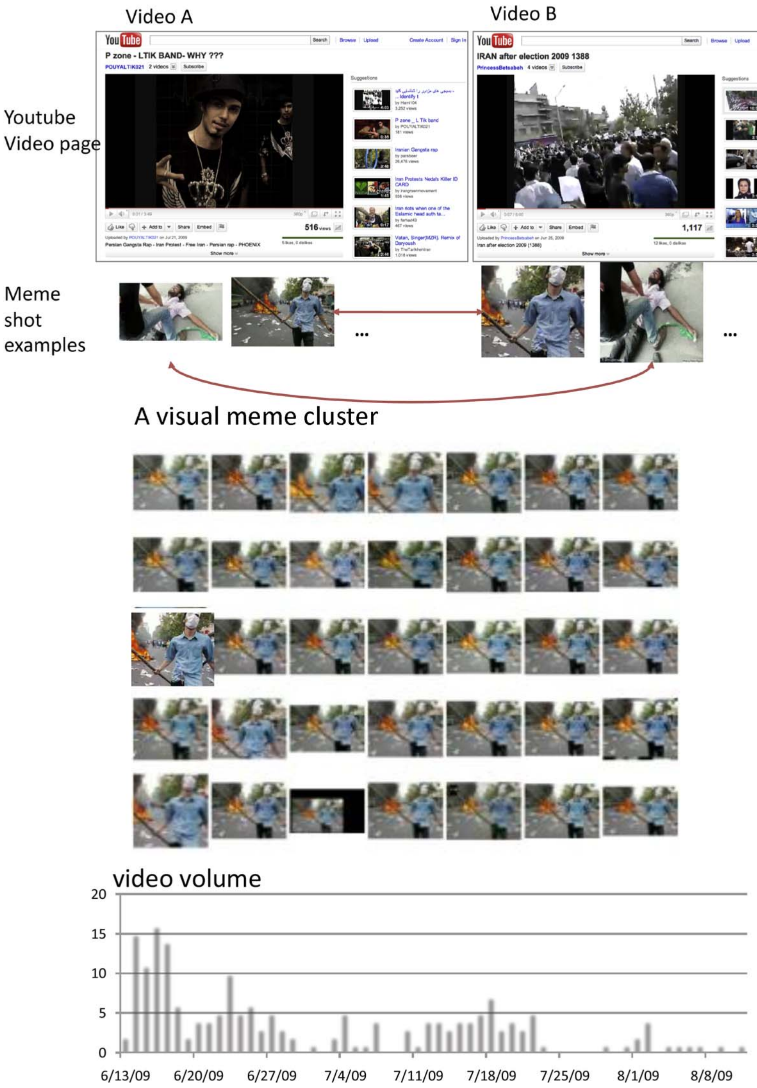


Fig. 6 ***Visual meme shots and meme clusters. (Top) Two YouTube videos that share multiple different memes. Note that it is impossible to tell from metadata or the YouTube video page that they shared content, and that the appearance of the remixed shots (bottom row) has large variations. (Bottom) A sample of other meme keyframes corresponding to one of the meme shots, and the number of videos containing this meme over time—193 videos in total between June 13 and August 11, 2009.***

lookup. We use the FLANN Library [49] to automatically select the best indexing structure and parameters for a given dataset. Our image features have over 300 dimensions, and we empirically found that ANN needs to traverse approximately $\sqrt{N}$ to obtain 0.95 precision in finding nearest neighbors, corresponding to two to three decimal orders of magnitude speedup over exact nearest neighbor search when $N \sim 10^6$. Furthermore, each FLANN query results in an incomplete set of near-duplicate pairs, so we perform transitive closure [50] on the neighbor relationship to find equivalence classes of near-duplicate sets.

We measure meme detection performance using ground truth from a YouTube dataset, which contains ~15 000 near-duplicate keyframe pairs and ~25 000 nonduplicate keyframe pairs. We compute the near duplicate equivalence classes as described above, and calculate precision (P) and recall (R) on the labeled pairs. The results are shown on Fig. 8 (left) for varying values of the match threshold parameter. We note that the performance is generally quite high, with $P > 95\%$. For the rest of our analysis, we use the operating point maximizing recall, which leads to $P = 96.6\%$, $R = 80.7\%$.

3) Meme Network for Influence Modeling: We can estimate the influence of content and of authors using visual memes. Visual memes can be viewed as *links* between creators videos that share the same visual segment. We

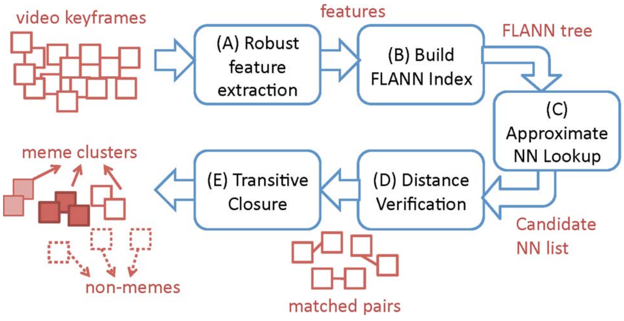


Fig. 7. ***Visual meme detection workflow. The input to this system is a set of video frames, and the output splits this set into two parts. The first part consists of a number of meme clusters, where frames in the same cluster are remixes with each other. The second part consists of the rest of the frames that are not considered near-duplicates with any other frame. Blocks A and D address the robust matching challenge using correlogram features and query-adaptive thresholding, and blocks B, C, and E address the scalability challenge using approximate nearest-neighbor (ANN) indexing.***

derive a link-based measure—*diffusion influence index*—to depict the influence of a meme and its author.

Denote a video (or multimedia document) as d_i in event collection $\mathcal{D}$, with $i = 1, \ldots, N$. Each video is authored (i.e., uploaded) by author $a(d_i)$ at time $t(d_i)$, with $a(d_i) \in \mathcal{A}$. Each video document d_i contains a collection of memes, $\{m_1, m_2, \ldots, m_{K_i}\}$, from a meme dictionary $\mathcal{M}$.

We compute the in-degree (resp., out-degree) of each meme m in video d_i as the number of other videos containing meme m that appeared before (resp., after) d_i

$$\begin{aligned}\zeta^{\text{in}}_{i,m} &= \sum_j I\big\{m \in d_j, t(d_j) < t(d_i)\big\}\\ \zeta^{\text{out}}_{i,m} &= \sum_j I\big\{m \in d_j, t(d_j) > t(d_i)\big\}\end{aligned} \tag{4}$$

where $I\{\cdot\}$ is the indicator function that takes a value of 1 when its argument is true, and 0 otherwise. Intuitively, $\zeta^{\text{in}}_{i,m}$ and $\zeta^{\text{out}}_{i,m}$ capture the number of videos that can serve as potential sources and potential followers for meme m in d_i. The video influence index χ_i is defined for each video document d_i as the ratio of its out-degree over its in-degree, aggregated over all memes (5). The smoothing constant in the denominator accounts for d_i itself. The total author influence index $\hat{\chi}_r$ is the aggregate χ_i over all videos from author a_r

$$\chi_i = \sum_m \frac{\zeta^{\text{out}}_{i,m}}{1 + \zeta^{\text{in}}_{i,m}} \tag{5}$$

$$\hat{\chi}_r = \sum_{\{i, a(d_i) = a_r\}} \chi_i. \tag{6}$$

Intuitively, our influence index gives a higher score to a video, which was uploaded early, containing a very popular meme.

4) *Observations From Youtube Event Datasets:* We monitored real-world events in YouTube by using a few generic, time-insensitive text queries to filter the content. The queries were designed to capture a generic topic or theme, including causative agents and consequences related to the topic. Our query on "iran election" topic, for example, consists of the terms *iran election, teheran protest, iran unrest*, and so on. We created queries to cover the invariant aspects of a topic; automatic time-varying query expansion is a natural extension of our work.

For each unique video, we segment shots, extract keyframes, and extract visual features from each keyframe. We also retrieve the associated metadata, including author, publish date, view-counts, and free-text title and descriptions.

We present sample observations from the Iran3 set, which contains videos related to the Iranian election in 2009 and related international reactions from June to August in 2009. This collection is representative since it has significant volume and it showcases event dynamics. The dataset comprises more than 23 000 YouTube videos,

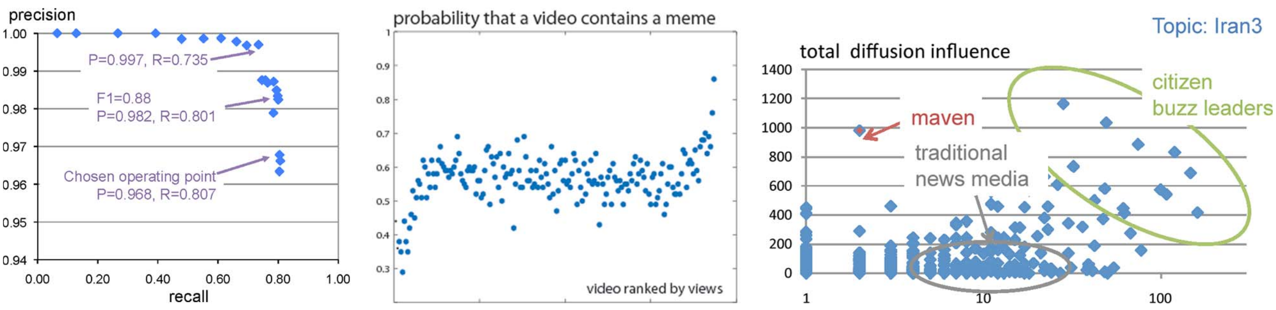


Fig. 8. ***(Left) Performance of visual meme detection method on the Housing dataset. (Middle) Video views versus meme probability on Iran3 set. (Right) Total diffusion influence versus the number of videos produced by each author on Iran3 dataset.***

with up to 1000 new videos uploaded per day at peak times, and it contains more than 1.25 million shots in total. It showcases rich real-world event dynamics—there are multiple national protests, large-scale conflicts between supporting parties, and various incidents, which generated massive international attention. More detailed dataset information, analysis, and observations on other topics are available in [47].

An examination of the dataset offers some interesting insights. The behavior of remixing and reposting is quite dominant—over 58% of the videos and 70% of the authors contain visual memes for Iran3. Video popularity, however, is a poor indicator of how likely a video is to be reposted. In the Iran3 set of more than 23 K videos, for example, the four most popular videos have no memes and have nothing to do with the topic, and likewise for 7 of the first 10. One has to get beyond the first 1600 most popular videos before the likelihood of having memes passes the average for the dataset, at about 0.58 (see Fig. 8, middle). There are several reasons for this mismatch. Among the video entries returned by the YouTube search API, the most viewed are often not related to the topic—the one with the highest view-count is a music video irrelevant to Iranian politics or the specific query words we used. Such videos also tend to be part of a production (e.g., promotion for a song), which bears lesser value for reposting and reinterpretation. Moreover, there is a strong "rich-get-richer" effect due to content recommendations on YouTube, which tend to promote popular videos regardless of their relevance to a query. In short, video view-counts are a poor proxy for importance to an event of interest, and visual memes provide a different metric for relevance and important.

Let us analyze users via the author influence index (6) on dataset Iran3. In Fig. 8 (right), we plot the total diffusion influence $\hat{\chi}_r$ versus the number of videos produced by each author. We can see a few distinct types of contributors. We call one type "maven" (marked in red), who post only a few videos that end up being massively remixed and reposted—this particular maven was among the first to post the murder of Neda Soltan[10] and one other instance of a student murder on the street. The former became the icon of the entire event and the face of Iranian struggle during this turbulent period. A second group can be dubbed "citizen buzz leaders" (circled in green), who tend to produce a large number of videos with high total diffusion factor, yet relatively low influence per video. They aggregate notable content and come relatively late in the timeline, which is penalized by the influence factor. We examined the YouTube channel pages for a few authors in this group, and they seem to be voluntary activists with screen names like "iranlover100." Some of their videos are slide shows of iconic images and provide good summaries. Note that traditional news media, such as Al Jezeera English, Associated Press, and so on (circled in gray), are ranked rather low for this topic, partially because the Iran government severely limited international media participation in the event, and most of the event buzz was driven by citizens.

In this section, we examined the interplay between content and community through multiple case studies: extraction of latent structure, interestingness of conversations, and tracking of visual memes. The first two analyze the effect of a rich media object on user activities, including tagging and commenting, while the last study uses content analysis to study how people influence each other. Analysis of photographs from the same event, similar queries issued at the same time, and similar click streams on retrieval results are some of the other methods to understand social interaction.

Understanding social interactions around media objects is beneficial. Social interactions derived from multimedia content can in turn help improve solutions to a large class of content analysis problems such as video annotation, image tagging, or search reranking. We can add the social interactions derived from multimedia content to the existing social graphs in other modalities (such as friendship, likes, and comments) and use them to better understand social structures. Example of such social analysis include estimating influence, ranking users, categorizing user roles.

III. RELATED WORK

We now discuss research in support of answering the three key questions: determining social interaction context, finding value for media objects, and what media object use in social context tells us about people.

The answers to questions on social interaction context are most closely related to work on community discovery. Identification of communities as cohesive subgroups of individuals within a network, where cohesive subgroups are defined as "subsets of actors among whom there are relatively strong, direct, intense, frequent, or positive ties" [51], is an important research topic in social network analysis. This is because social network analysis does not presume a prior solidary local bounds that organize people's interpersonal relationships. Newman [52] gives a broad review of important findings and concepts in network research, including degree-distribution, small-world effect, and community structure.

Community detection algorithms identify the modular structure of a network, where nodes represent individuals and where links represent the interaction or similarity between individuals. Intuitively, modules or communities are subsets of nodes within which the links are dense and between which the links are sparse [53], [54]. Many graph-based approaches, including those based on analysis of cliques, degree, and matrix-perturbation, have been proposed to extract cohesive subgroups from social networks [51]. Examples of detected communities range

[10] http://en.wikipedia.org/wiki/Death_of_Neda_Agha-Soltan

from communities of scientists working on similar areas of research [53] to authors of home pages who have some common interests [55]. See Fortunato [56] for a comprehensive review.

The algorithms for community identification are closely related to the family of algorithms for clustering. The goal of clustering is to discover groups of similar objects within the data. Each cluster (i.e., group) consists of objects that are similar to one another within the same cluster,and dissimilar to the objects in other clusters. Community identification can be considered to be clustering in the sense that it involves a distance function and a clustering objective function and generates a clustering assignment for each person and object to a set of clusters. While there are similarities between community extraction and clustering analysis, community extraction focuses on the pairwise relationship between network nodes and, more generally, the network topology.

Research on community discovery includes measures for quantifying community structure (including the clustering coefficient [52]) and techniques for community extraction. A variety of methods for extracting community structure have been proposed including *modularity*-based methods [54], flow- or *graph cut*-based methods [57], spectral clustering or graph Laplacian-based methods [58], and information-theoretic models [59]. Community extraction techniques have been used to study dynamic properties of communities in empirical networks [60].

Clustering-based methods for community detection need to account for interactions with the following characteristics: social context, temporal coherence, and contextual coherence. These characteristics are consistent with Garfinkel's observation on the necessity of mutual awareness [61] and Jones' work on the virtual community [62].

We can incorporate social context with two concepts: mutual awareness and transitive awareness. Mutual awareness refers to a relationship developed through observable interactions between two people. Mutual awareness can be asymmetric—the asymmetry can arise, for example, when one person is a celebrity or is in touch with more people than the other. In addition, mutual awareness strength can change over time. Transitive awareness refers to a relationship—computed via a mutual awareness measure—between two connected people on a network. We can compute transitive awareness between a connected pair of users on a social network graph, through mutual-awareness expansion. We can use a random-walk-based distance, with an efficient method for mutual awareness expansion, to extract communities [63].

In order to analyze the additional value brought on by the social interaction context, we need to understand prior work on dynamic properties of media objects, how to extract themes, and how to analyze communication.

There is significant prior work on the analysis of dynamic properties (e.g., associated tags on a media object) of media objects. In [38], 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 [64] leveraged the community-contributed collections of rich media (Flickr) to automatically generate representative views of landmarks. Their work suggests that community-generated media and tags can improve access to multimedia resources. In [65], Smith *et al.* explored the search methodologies of rich social media content by utilizing the social context of users, including both their personal social context (their friends and the communities to which they belong) and their community social context (their role and identity in different communities). Singh *et al.* [66] analyzed and modeled user contributions in social media sites to study associated dynamics. Their model was based on the idea of users as rational selfish agents and considered domain attributes like voluntary participation, virtual reward structure, and public sharing to model the dynamics of this interaction.

Theme extraction from dynamic web collections is a well-studied problem [37], [67]–[69]. In research related to topic models [70], the goal is to discover patterns in text corpora. In [70], the authors propose a generative model for documents using topics; topics in turn are represented with word distributions. Dynamic topic models [71] have been proposed to capture the evolution of topics in a sequentially organized corpus of documents. In [67], the authors study the problem of discovering and summarizing evolutionary theme patterns in a dynamic text stream. The authors modify temporal theme extraction in [37] by regularizing their theme model with timestamp and location information. In [68], the authors use a dynamic probability model to predict discussion topics in online social networks. Recently, Iwata *et al.* [72] developed an online topic model for sequentially analyzing the time evolution of topics in document collections.

Researchers have also studied topical and structural analysis of commentary on social websites as well as on social media communications platforms, including Twitter. Gomez *et al.* [73] analyze several social network properties of communication activity on the website Slashdot. They study the structure of discussion threads using a radial tree representation. The findings show that nesting of conversations on Slashdot exhibits strong heterogeneity and self-similarity. Honeycutt *et al.* [74] investigated conversations on Twitter, with special attention to the role played by the @ sign. Naaman *et al.* [75] analyzed message content from individuals on Twitter with the goal to categorize individuals into those who focus on the "self" versus those who are driven more by sharing information.

The analysis content in social media platforms, including how content has been shared and reused, has received significant attention. YouTube has been the focal platform for many social network monitoring studies. The first large-scale YouTube measurement study [76] characterized content category distributions and tracked exact duplicates of popular videos. Benevenuto *et al.* studied video response actions on YouTube [77] using metadata, and De Choudhury *et al.* monitored user comments to determine interesting conversations (see Section II-B). Recently, early views of YouTube videos have been used to predict ultimate popularity, characterized by view-counts [78]. Quoting, duplication, and reposting are popular phenomena in online information networks. One well-known example is the use of the RT (retweet) tag on Twitter [40], [41], where users often quote the original message verbatim, having little freedom for remixing and context changes within the 140-character limit. MemeTracker [79] is another example that tracks the lifecycles of popular phrases among blogs and news websites. Prior studies have shown that image copying and editing can be tracked over the Web and used to enhance retrieval results [80], the frequency of video reuse can be used as an implicit video quality indicator [81], and that segmenting videos into small pieces enhances media sharing experiences [82].

Tracking near-duplicates in images and video has been a problem of interest since the early years of content-based retrieval. Recent attention on this problem has been on user-dependent definitions of duplicates [83], speeding up detection on image sequence, frame, or local image points [84], and scaling out to web-scale computations using large compute clusters [85]. We note, however, that most prior work in this area is concerned with optimizing retrieval accuracy of individual frames or sequences rather than tracking large-scale duplication behavior.

IV. EMERGING RESEARCH AREAS

The difficulty in acquiring data implies that much of the research on social network analysis is from small datasets.[11] Robust sampling of social graphs that preserves the data characteristics of interest is therefore an important technical problem.

The volume of the data and the speed with which the data changes pose significant challenges for efficient data analysis. Furthermore, a framework for extracting useful information from social media data needs to scale also against the number of facets and diversity of facets. Consequently, a scalable framework for managing voluminous user data in a form amenable for large-scale data analysis is a key step to any significant technical advances in social media understanding.

In the next two sections, we review recent work to address these two issues. In Section IV-A, we discuss robust sampling of network data [86], and in Section IV-B, we discuss the use of compressive sampling in efficiently monitoring changes to social media streams [87].

A. Robust Data Sampling

Robust sampling of social networks is closely related to the problem of "subgraph sampling." *Snowball sampling* [88] is a subgraph sampling method commonly used in sociology studies; *random walk sampling* [89] is another well-known method. Recent work has investigated sampling of large-scale graphs, with a focus on recovering topological characteristics including degree distribution, and path length. Prior work [90] has also investigated the influence of missing data on measurement of social network properties. Leskovec *et al.* in [91] and [92], for example, focus on empirically observed static and dynamic graph properties such as densification and shrinking diameter. The authors study different sampling methods, including random node/edge selection and random walk, for recovering topological properties. They also introduce the *forest fire sampling* strategy, in which a forwarding probability is used to sample a subset of neighbors of the current traversed node.

There are two limitations of topology-based sampling methods. First, topology-based network sampling methods—application-independent and designed to recover the topological characteristics of the particular social graph—*ignore* information content. Online social media feature extensive activity dependent on the shared content. Additionally, social media activity exhibits correlation with external events [27]. Hence, pure topology-based sampling is unsuitable for studying social processes dependent on the relationship between the shared content and external user actions and events. Second, prior sampling research does not consider the *contextual information* of the users in the social graph, including geographical location, or rate of change of user status.

To understand the influence of different sampling strategies on social network analysis, we conducted a preliminary study [86] on the effects of different sampling methodologies on a well-studied social phenomenon: information diffusion. Diffusion has been studied in the context of medical and technological innovations [93], cultural bias [5], [94], and understanding information roles of users [95], [96].

Our approach comprises two steps. First, we utilize several popularly used sampling techniques such as random sampling, degree of user activity-based sampling, forest-fire, and location-attribute-based sampling to extract subgraphs from a social graph of users engaged in a social activity. Second, these subgraphs—one for each sampling method—are used to study diffusion characteristics with respect to the properties of the users (e.g., participation), structural (e.g., reach, spread), and temporal characteristics (e.g., rate), as well as relationship to events in the external world (e.g., search and news trends).

[11] While Twitter datasets in the millions of tweets are common, the number is dwarfed by the total number of tweets on Twitter: ~17 billion per year.

Our experiments [86] reveal that methods that incorporate both network topology and user-context—activity, attributes related to "homophily" (e.g., location)—are able to better explain diffusion characteristics compared to naïve methods (e.g., random or activity-based sampling) by a large margin of ~15%–20%. Moreover, we can show that for moderate sample sizes (30%), these hybrid methods can better approximate the measurements relating to information diffusion than can pure topology-based methods.

Our primary observation from the results is that sampling influences the discovery of diffusion in a nontrivial manner. Our observations are in contrast to prior empirical observations [92] that ignore contextual information. We found that topology-based sampling techniques that additionally incorporate user context (e.g., activity or location) perform better than pure topology-based sampling. Interestingly, pure context-based techniques leveraging location perform better than activity-based or random sampling. Themes vary in their diffusion characteristics. Hence, content has significant influence on sample quality. Studies of diffusion related to US political events, for example, would benefit more from samples chosen based on location than only on graph topology. On the other hand, if the interest is related to a recent technological event, such as release of an electronic gadget, one can benefit more from sampling techniques based on both topology and activity.

Our results are promising, but are limited by the scope of our dataset, which itself is based on a sample, and also limited to a single phenomena—diffusion. Nevertheless, we believe that our empirical observations can form the basis of investigation of other phenomena as well, including community discovery, because most social processes are affected by both topology and context.

Sampling is a well-known problem in the signal processing community. However, many open questions remain for social networks, including sampling strategies to preserve properties of content in the sample as well as determining a relationship between the social dynamics and the sampling strategy—akin to the relationship between bandwidth and sampling frequency.

B. Real-Time Temporal Monitoring

To facilitate large-scale data management and analysis, we proposed SCENT—*Scalable Compressed Domain Analysis of Evolving Tensors*—a framework for monitoring the evolution of multifaceted (also called multirelational) social network data [87]. In particular, we focused on the problem of *change detection*, a key step in understanding the evolution of patterns in multirelational social media data.

We model the social data in the form of tensor[12] streams [97], [98] and consider the problem of tracking the changes in the spectral properties of the tensor over time. Tensor decompositions have been used in a variety of applications in data mining. Chi *et al.* [97] applied the high-order singular value decomposition (HOSVD, a version of the Tucker decomposition) to extract dynamic structural changes as trends of the blogosphere. Sun *et al.* [98] proposed methods for dynamically updating a Tucker approximation, with applications ranging from text analysis to network modeling. Tensor decomposition is computationally expensive, and scalable solutions have been proposed [99], [100]. Incremental tensor decomposition, while being faster than regular tensor decomposition, has exponential complexity [98].

There is a significant body of work devoted to web data reduction. Methods based on ranking algorithms [101], [102], spectral clustering or graph partitioning [103], and probability mixture model [104] have been proposed for web data reduction. Existing work in lossless social graph compression [105] can reduce storage complexity to 3–4 bits per edge. Work on graph sampling seeks to reduce the data while preserving network statistics and topological properties. [90], [106].

There are two major differences between prior research and the techniques for monitoring social media data, especially within the context of change detection. First, at each time instance, the relevant data forms a (multirelational) graph—in contrast to a vector or matrix of intensity values. Second, the graph is very large and dynamic. Therefore, change detection in social media data requires techniques that can efficiently detect structural changes in very large graphs.

In our work [87], we introduced an innovative *compressed sensing* mechanism—thereby reducing the computational cost of detecting significant changes in tensor streams—to encode the social data tensor streams in the form of compact descriptors. We have extended recently developed Compressive Sensing (CS) theory to tensor stream analysis. The CS theory shows that, under certain conditions, a signal can be faithfully reconstructed using fewer number of samples than predicted by the well-known Nyquist sampling theorem [107]–[109]. CS theory has been primarily used in the analysis of 1-D and 2-D continuous time signals, and the basic compressed sensing theory assumes the availabilities of a sparse data basis and a constant-time random sensing mechanism. Neither assumption holds for social media tensors. The key contribution of our work is to create and use *random sensing* ensembles to transform a given tensor into a compressed representation that implicitly captures the spectral characteristics of the tensor (i.e., the so called core tensor coefficients [110]). The length of this compressed representation is only $O(S \cdot \log N/S)$, where N is the size of data tensor and S is a small approximation rank. We show that the descriptors support very fast detection of significant spectral changes in the tensor stream, which also reduce data collection, storage, and processing costs.

[12] A tensor is a multidimensional array.

We proposed three recovery strategies to incrementally obtain core tensor coefficients either from the input data tensor or from the compressed representation. These strategies can be used in different situations based on the availability of data and resources. We demonstrated the efficiency of SCENT on monitoring time-varying multi-relational social networks on real-world social media data as well as synthetic datasets. The experimental results show that our SCENT monitoring procedure is able to maintain an approximated tensor stream with high accuracy (above 0.9 in terms of F1-score), low errors (under 1.1 relative to baseline tensor decomposition in real-world datasets), and low time cost (17 X–159 X faster for change detection).

V. CONCLUSION

Social networks, with attractive interactive design and sophisticated media sharing and notification features, make it easy for people to stay in touch with friends and family. The result is an unprecedented growth in social networks with an accompanying increase in information. Social networks like Twitter and Facebook, for example, have a significant real-world impact, including impact on offline behavior such as retail products consumption and political participation.

In this paper, we have examined the interrelationship between content and social context through the prism of three key questions. First, how do we extract context in which social interactions occur? Second, does social interaction provide *value* to the media object? Finally, how does social media facilitate the repurposing of shared content and engender cultural memes? To understand these questions, we examined three cases. First, to understand the context in which social interactions occur, in Section II-A, we examined the idea of *interaction semantics*. Interaction semantics, which arise from social interaction around and on media objects, are distinct from media semantics—instead of asking about the meaning of the media object, we are interested in the semantics arising from social interaction around media objects. We can use interactional semantics to address familiar problems such as tagging new media objects and media organization. Second, in Section II-B, we examined the *interestingness* of conversations. We were motivated by the observation that community members return to repeatedly participate in a conversation around the same video. The key idea we explored is that interesting people make for interesting conversations. A measure of conversational interestingness can be applied to the familiar problem of video ranking. Finally, we defined and studied properties of *visual memes*. Visual memes help us understand community participants share and repurpose content. The analysis can reveal relationships between community members and their influence. In particular, visual memes help us understand the different roles that people play in content-sharing networks.

We briefly examined two emerging research areas: sampling bias and data diversity. The volume of the data and the speed with which the data changes pose significant challenges for efficient data analysis. Furthermore, a framework for extracting useful information from social media data needs to scale also against the number of facets and diversity of facets.

Robust sampling methods have primarily focused on topological sampling to recover the topological characteristics of the particular social graph, while ignoring the information content, or pertinent contextual information. Hence, pure topology-based sampling is unsuitable for studying social processes dependent on the relationship between the shared content and external user actions and events. We discussed that real-time monitoring of social media data is very challenging due to several reasons. First, at each time instance, the relevant data forms a (multi-relational) graph—in contrast to a vector or matrix of intensity values. Second, the graph is very large and dynamic. Therefore, change detection in social media data requires techniques that can efficiently detect structural changes in very large graphs.

Social networks are a fertile ground for interesting questions related to media semantics. For the first time—via social networks including Twitter—we are able to instrument social activity at a very large scale. The unprecedented data scale and our newfound ability to collect data over extended time periods result in new questions—*emergence* is one of several interesting questions. Much of the current work on multimedia semantics implicitly assumes that multimedia semantics are static—that is, the meaning of concepts remains unchanged. New concepts, however, constantly appear such as "iPhone" as an example. Prior to 2007, this word would have been meaningless, yet in 2011, the word is ingrained in our contemporary culture. "Pwn" (to own) is another word that is new—it emerged in online social networks and is used by young people.

While we recognize that semantics are an emergent artifact of human activity, data from social networks have helped us examine this issue in great detail, for the first time. Through data collected from social networks, we are in a position to understand not only emergence, but also how meaning evolves with time. These research directions complement substantial current work on concept learning. ■

REFERENCES

[1] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, "Information diffusion through blogspace," in *Proc. 13th WWW*, 2004, pp. 491–501.

[2] D. Liben-Nowell and J. Kleiberg,, "Tracing information flow on a global scale using internet chain-letter data," *Proc. Nat. Acad. Sci.*, vol. 105, no. 12, pp. 4633–4638, 2008.

[3] A. Anagnostopoulos, R. Kumar, and M. Mahdian, "Influence and correlation in social networks," in *Proc. 14th ACM KIGKDD KDD*, 2008, pp. 7–15.

[4] E. Sun, I. Rosenn, C. Marlow, and T. Lento, "Gesundheit! modeling contagion through facebook news feed," in *Proc. 3rd Int. Conf. Weblogs Social Media*, San Jose, CA, May 2009.

[5] E. Bakshy, B. Karrer, and L. A. Adamic, "Social influence and the diffusion of user-created content," in *Proc. 10th ACM EC*, 2009, pp. 325–334.

[6] Cisco, San Jose, CA, *Cisco Visual Networking Index: Forecast and Methodology, 2010–2015*, Jun. 2011. [Online]. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html

[7] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval," *Trans. Multimedia Comput., Commun., Appl.*, vol. 2, no. 1, pp. 1–19, Feb. 2006.

[8] A. Hauptmann, R. Yan, and W.-H. Lin, "How many high-level concepts will fill the semantic gap in news video retrieval?" in *Proc. 6th ACM Int. Conf. Image Video Retrieval*, 2007, pp. 627–634.

[9] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE CVPR*, 2001, vol. 1, pp. 511–518.

[10] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Proc. IEEE CVPR*, 2003, pp. 264–271.

[11] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba, "SUN database: Large-scale scene recognition from abbey to zoo," in *Proc. IEEE CVPR*, 2010, pp. 3485–3492.

[12] Y. Zheng, M. Zhao, Y. Song, H. Adam, U. Buddemeier, A. Bissacco, F. Brucher, T. Chua, and H. Neven, "Tour the world: Building a web-scale landmark recognition engine," in *Proc. IEEE CVPR*, 2009, pp. 1085–1092.

[13] M. Marszaek and C. Schmid, "Semantic hierarchies for visual object recognition," in *Proc. IEEE CVPR*, Minneapolis, MN, 2007, pp. 1–7.

[14] Y. Wu, B. Tseng, and J. Smith, "Ontology-based multi-classification learning for video concept detection," in *Proc. IEEE ICME*, 2005, vol. 2, pp. 1003–1006.

[15] M. Naphade, J. Smith, J. Tesic, S. Chang, W. Hsu, L. Kennedy, A. Hauptmann, and J. Curtis, "Large-scale concept ontology for multimedia," *IEEE Multimedia*, vol. 13, no. 3, pp. 86–91, Jul.–Sep. 2006.

[16] D. Shamma, R. Shaw, P. Shafton, and Y. Liu, "Watch what i watch: Using community activity to understand content," in *Proc. ACM Int. Workshop Multimedia Inf. Retrieval*, 2007, pp. 275–284.

[17] N. Garg and I. Weber, "Personalized, interactive tag recommendation for Flickr," in *Proc. ACM RecSys*, 2008, pp. 67–74.

[18] B. Sigurbjörnsson and R. van Zwol, "Flickr tag recommendation based on collective knowledge," in *Proc. 17th WWW*, 2008, pp. 327–336.

[19] R. Negoescu and D. Gatica-Perez, "Analyzing Flickr groups," in *Proc. CIVR*, 2008, pp. 417–426.

[20] A. Zunjarward, H. Sundaram, and L. Xie, "Contextual wisdom: Social relations and correlations for multimedia event annotation," in *Proc. 15th MULTIMEDIA*, 2007, pp. 615–624.

[21] L. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury, "How Flickr helps us make sense of the world: Context and content in community-contributed media collections," in *Proc. 15th MULTIMEDIA*, 2007, pp. 631–640.

[22] N. Ramzan, M. Larson, F. Dufaux, and K. Clüver, "The participation payoff: Challenges and opportunities for multimedia access in networked communities," in *Proc. ACM Int. Conf. Multimedia Inf. Retrieval*, 2010, pp. 487–496.

[23] M. De Choudhury, H. Sundaram, Y. Lin, A. John, and D. Seligmann, "Connecting content to community in social media via image content, user tags and user communication," in *Proc. IEEE ICME*, 2009, pp. 1238–1241.

[24] S. Gupta, D. Phung, B. Adams, T. Tran, and S. Venkatesh, "Nonnegative shared subspace learning and its application to social media retrieval," in *Proc. 16th ACM SIGKDD KDD*, 2010, pp. 1169–1178.

[25] E. Adar, D. S. Weld, B. N. Bershad, and S. S. Gribble, "Why we search: Visualizing and predicting user behavior," in *Proc. 16th WWW*, New York, 2007, pp. 161–170.

[26] M. De Choudhury, H. Sundaram, A. John, and D. D. Seligmann, "Can blog communication dynamics be correlated with stock market activity?" in *Proc. 19th ACM HT*, 2008, pp. 55–60.

[27] D. Gruhl, R. Guha, R. Kumar, J. Novak, and A. Tomkins, "The predictive power of online chatter," in *Proc. 11th ACM SIGKIDD*, 2005, pp. 78–87.

[28] Y. Liu, X. Huang, A. An, and X. Yu, "ARSA: A sentiment-aware model for predicting sales performance using blogs," in *Proc. 30th Annu. ACM SIGIR*, 2007, pp. 607–614.

[29] W. Antweiler and M. Z. Frank, "Is all that talk just noise? The information content of Internet stock message boards," *J. Finance*, vol. 59, no. 3, pp. 1259–1294, Jun. 2004.

[30] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, and B. L. Tseng, "Analyzing communities and their evolutions in dynamics networks," *Trans. Knowl. Discovery Data*, vol. 3, no. 2, Apr. 2009, Article no. 8.

[31] Y.-R. Lin, H. Sundaram, M. De Choudhury, and A. Kelliher, "Discovery multirelational structure in social media streams," *Trans. Multimedia Comput., Commun., Appl.*, vol. 8, no. 1, 2011, Article no. 4.

[32] Z. Xiao, Z. Hou, C. Miao, and J. Wang, "Using phase information for symmetry detection," *Pattern Recogn. Lett.*, vol. 26, no. 13, pp. 1985–1994, 2005.

[33] Z. Liu and R. Laganière, "Phase congruence measurement for image similarity assessment," *Pattern Recogn. Lett.*, vol. 28, no. 1, pp. 166–172, 2007.

[34] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 959–973, Aug. 2003.

[35] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[36] M. De Choudhury, H. Sundaram, A. John, and D. D. Seligmann, "What makes conversations interesting?: Themes, participants and consequences of conversations in online social media," in *Proc. 18th WWW*, 2009, pp. 331–340.

[37] Q. Mei, C. Liu, H. Su, and C. Zhai, "A probabilistic approach to spatiotemporal theme pattern mining on Weblogs," in *Proc. 15th WWW*, 2006, pp. 533–542.

[38] M. Dubinko, R. Kumar, J. Magnani, J. Novak, P. Raghavan, and A. Tomkins, "Visualizing tags over time," in *Proc. 15th WWW*, 2006, pp. 193–202.

[39] J. Burgess and J. Green, *YouTube: Online Video and Participatory Culture*. Cambridge, U.K.: Polity, 2009.

[40] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi, "Measuring user influence in Twitter: The million follower fallacy," in *Proc. 4th Int. AAAI ICWSM*, 2010.

[41] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in *Proc. WWW*, 2010, pp. 591–600.

[42] D. Boyd, S. Golder, and G. Lotan, "Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter," in *Proc. 43rd IEEE HICSS*, 2010, pp. 1–10.

[43] D. A. Shamma, L. Kennedy, and E. F. Churchill, "Tweetgeist: Can the Twitter timeline reveal the structure of broadcast events?" in *Proc. CSCW Horizons*, 2010.

[44] N. Diakopoulos, M. Naaman, and F. Kivran-Swaine, "Diamonds in the rough: Social media visual analytics for journalistic inquiry," in *Proc. IEEE VAST*, 2010, pp. 115–122.

[45] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller, "Twitinfo: Aggregating and visualizing microblogs for event exploration," in *Proc. Annu. Conf. Human Factors Comput. Syst.*, 2011, pp. 227–236.

[46] P. Snickars and P. Vonderau, *The YouTube Reader*. Stockholm, Sweden: Nat. Library of Sweden, 2010.

[47] L. Xie, A. Natsev, M. Hill, J. R. Kender, and J. R. Smith, "Visual memes in social media: Tracking real-world news in YouTube videos," in *Proc. ACM MULTIMEDIA*, Nov. 2011, pp. 53–62.

[48] J. Huang, S. Kumar, M. Mitra, W. Zhu, and R. Zabih, "Spatial color indexing and applications," *Int. J. Comput. Visio*, vol. 35, no. 3, Dec. 1999.

[49] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *Proc. Int. Conf. Comput. Vision Theory Appl.*, 2009.

[50] B. A. Galler and M. J. Fisher, "An improved equivalence algorithm," *Commun. ACM*, vol. 7, no. 5, pp. 301–303, 1964.

[51] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[52] M. Newman, "The structure and function of complex networks," *SIAM Rev.*, vol. 45, no. 2, pp. 167–256, 2003.

[53] M. Girvan and M. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci.*, vol. 99, no. 12, p. 7821, 2002.

[54] M. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, p. 26113, 2004.

[55] L. Adamic and E. Adar, "Friends and neighbors on the Web," *Social Netw.*, vol. 25, no. 3, pp. 211–230, 2003.

[56] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, no. 3–5, pp. 75–174, 2010.

[57] G. Flake, S. Lawrence, and C. Giles, "Efficient identification of Web communities," in *Proc. 6th ACM SIGKDD KDD*, 2000, pp. 150–160.

[58] M. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Phys. Rev. E*, vol. 74, no. 3, p. 036104, 2006.

[59] M. Rosvall and C. Bergstrom, "An information-theoretic framework for resolving community structure in complex networks," *Proc. Nat. Acad. Sci.*, vol. 104, no. 18, p. 7327, 2007.

[60] J. Leskovec, K. Lang, A. Dasgupta, and M. Mahoney, "Statistical properties of community structure in large social and information networks," in *Proc. 17th WWW*, 2008, pp. 695–704.

[61] H. Garfinkel, *Studies in Ethnomethodology*. Cambridge, U.K.: Polity, 1984.

[62] Q. Jones, "Virtual-communities, virtual settlements & cyber-archaeology: A theoretical outline," *J. Comput. Mediated Commun.*, vol. 3, no. 3, pp. 35–49, 1997.

[63] Y.-R. Lin, H. Sundaram, Y. Chi, J. Tatemura, and B. Tseng, "Blog community discovery and evolution based on mutual awareness expansion,, 2007.

[64] L. S. Kennedy and M. Naaman, "Generating diverse and representative image search results for landmarks," in *Proc. 17th WWW*, 2008, pp. 297–306.

[65] M. Smith, V. Barash, L. Getoor, and H. W. Lauw, "Leveraging social context for searching social media," in *Proc. ACM SSM*, 2008, pp. 91–94.

[66] V. K. Singh, R. Jain, and M. S. Kankanhalli, "Motivating contributors in social media networks," in *Proc. 1st ACM SIGMM WSM*, 2009, pp. 11–18.

[67] Q. Mei and C. Zhai, "Discovering evolutionary theme patterns from text: An exploration of temporal text mining," in *Proc. 11th ACM SIGKDD KDD*, 2005, pp. 198–207.

[68] Y. Zhou, X. Guan, Z. Zhang, and B. Zhang, "Predicting the tendency of topic discussion on the online social networks using a dynamic probability model," in *Proc. WebScience, Hypertext Workshop Collab. Collective Intell.*, 2008, pp. 7–11.

[69] X. Ling, Q. Mei, C. Zhai, and B. Schatz, "Mining multi-faceted overviews of arbitrary topics in a text collection," in *Proc. 14th ACM SIGKDD KDD*, 2008, pp. 497–505.

[70] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[71] D. M. Blei and J. D. Lafferty, "Dynamic topic models," in *Proc. 23rd ACM ICML*, 2006, pp. 113–120.

[72] T. Iwata, T. Yamada, Y. Sakurai, and N. Ueda, "Online multiscale dynamic topic models," in *Proc. 16th ACM SIGKDD KDD*, 2010, pp. 663–672.

[73] V. Gómez, A. Kaltenbrunner, and V. López, "Statistical analysis of the social network and discussion threads in Slashdot," in *Proc. 17th ACM WWW*, 2008, pp. 645–654.

[74] C. Honeycutt and S. C. Herring, "Beyond microblogging: Conversation and collaboration via Twitter," in *Proc. 42nd HICSS*, 2009, pp. 1–10.

[75] M. Naaman, J. Boase, and C.-H. Lai, "Is it really about me?: Message content in social awareness streams," in *Proc ACM CSCW*, 2010, pp. 189–192.

[76] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system," in *Proc. ACM IMC*, 2007, pp. 1–14.

[77] F. Benevenuto, F. Duarte, T. Rodrigues, V. A. Almeida, J. M. Almeida, and K. W. Ross, "Understanding video interactions in YouTube," in *Proc. ACM MULTIMEDIA*, 2008, pp. 761–764.

[78] G. Szabo and B. A. Huberman, "Predicting the popularity of online content," *Commun. ACM*, vol. 53, pp. 80–88, Aug. 2010.

[79] J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-tracking and the dynamics of the news cycle," in *Proc. 15th ACM SIGKDD KDD*, 2009, pp. 497–506.

[80] L. Kennedy and S.-F. Chang, "Internet image archaeology: Automatically tracing the manipulation history of photographs on the Web," in *Proc. ACM MULTIMEDIA*, 2008, pp. 349–358.

[81] P. Schmitz, P. Shafton, R. Shaw, S. Tripodi, B. Williams, and J. Yang, "International remix: Video editing for the Web," in *Proc. ACM MULTIMEDIA*, 2006, p. 798.

[82] P. Cesar, D. Bulterman, D. Geerts, J. Jansen, H. Knoche, and W. Seager, "Enhancing social sharing of videos: Fragment, annotate, enrich, and share," in *Proc. ACM MULTIMEDIA*, 2008, pp. 11–20.

[83] M. Cherubini, R. de Oliveira, and N. Oliver, "Understanding near-duplicate videos: A user-centric approach," in *Proc. ACM MULTIMEDIA*, 2009, pp. 35–44.

[84] H.-K. Tan, X. Wu, C.-W. Ngo, and W.-L. Zhao, "Accelerating near-duplicate video matching by combining visual similarity and alignment distortion," in *Proc. ACM MULTIMEDIA*, 2008, p. 861.

[85] T. Liu, C. Rosenberg, and H. A. Rowley, "Clustering billions of images with large scale nearest neighbor search," in *Proc. IEEE Workshop Appl. Comput. Vision*, 2007, p. 28.

[86] M. D. Choudhury, Y.-R. Lin, H. Sundaram, K. S. Candan, L. Xie, and A. Kelliher, "How does the data sampling strategy impact the discovery of information diffusion in social media?" in *Proc. 4th Int. AAAI Conf. Weblogs Social Media*, Washington, DC, May 23–26, 2010, pp. 34–41.

[87] Y.-R. Lin, K. S. Candan, H. Sundaram, and L. Xie, "Scent: Scalable compressed monitoring of evolving multirelational social networks," *Trans. Multimedia Comput., Commun., Appl.*, vol. 7S, pp. 29:1–29:22, Nov. 2011.

[88] O. Frank, "Sampling and estimation in large social networks," *Social Netw.*, vol. 1, no. 91, p. 101, 1978.

[89] A. Klovdahl, Z. Dhofier, G. Oddy, J. O'Hara, S. Stoutjesdijk, and A. Whish, "Social networks in an urban area: First Canberra study," *J. Sociol.*, vol. 13, no. 2, pp. 169–172, 1977.

[90] P. Rusmevichientong, D. Pennock, S. Lawrence, and C. Giles, "Methods for sampling pages uniformly from the world wide Web," in *Proc. AAAI Fall Symp. Using Uncertainty Within Comput.*, 2001, pp. 121–128.

[91] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: Densification laws, shrinking diameters and possible explanations," in *Proc. 11th ACM SIGKDD KDD*, 2005, p. 187.

[92] J. Leskovec and C. Faloutsos, "Sampling from large graphs," in *Proc. 12th ACM SIGKDD KDD*, 2006, p. 636.

[93] M. E. J. Newman, "Spread of epidemic disease on networks," *Phys. Rev. E*, vol. 66, no. 1, p. 016128+, Jul. 2002.

[94] W. Zachary, "An information flow model for conflict and fission in small groups," *J. Anthropol. Res.*, vol. 33, pp. 452–473, 1977.

[95] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proc. 9th ACM SIGKDD KDD*, 2003, pp. 137–146.

[96] D. J. Watts and P. S. Dodds, "Influentials, networks, and public opinion formation," *J. Consumer Res.*, vol. 34, no. 4, pp. 441–458, Dec. 2007.

[97] Y. Chi, B. Tseng, and J. Tatemura, "Eigen-trend: Trend analysis in the blogosphere based on singular value decompositions," in *Proc. 15th ACM Int. Conf. Inf. Knowl. Manage.*, 2006, p. 77.

[98] J. Sun, D. Tao, and C. Faloutsos, "Beyond streams and graphs: Dynamic tensor analysis," in *Proc. ACM SIGKDD KDD*, 2006, pp. 374–383

[99] T. G. Kolda and J. Sun, "Scalable tensor decompositions for multi-aspect data mining," in *Proc. ICDM*, 2008, pp. 363–372.

[100] P. Drineas and M. Mahoney, "A randomized algorithm for a tensor-based generalization of the singular value decomposition," *Linear Algebra Appl.*, vol. 420, no. 2–3, pp. 553–571, 2007.

[101] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, no. 1–7, pp. 107–117, 1998.

[102] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[103] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[104] T. Hofmann, "Probabilistic latent semantic indexing," in *Proc. ACM SIGIR*, 1999, pp. 50–57.

[105] F. Chierichetti, R. Kumar, S. Lattanzi, M. Mitzenmacher, A. Panconesi, and P. Raghavan, "On compressing social networks," in *Proc. 15th ACM SIGKDD KDD*, 2009, pp. 219–228.

[106] J. Leskovec and C. Faloutsos, "Sampling from large graphs," in *Proc. 12th ACM SIGKDD KDD*, 2006, pp. 631–636.

[107] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Prob.*, vol. 23, pp. 969–985, 2007.

[108] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[109] E. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[110] L. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.

ABOUT THE AUTHORS

Hari Sundaram (Member, IEEE) received the Ph.D. degree in electrical engineering from Columbia University, New York, NY, in 2002.

He is an Associate Professor with the School of Arts Media and Engineering, Arizona State University, Tempe. His research interests include analysis of social network activity and the design of adaptive multimedia environments.

Dr. Sundaram is an Associate Editor for the *ACM Transactions on Multimedia Computing, Communications, and Applications* and *IEEE Multimedia*. His research has won several best paper awards from the IEEE and the ACM. He also received the Eliahu I. Jury Award for best Ph.D. dissertation in 2002.

Lexing Xie (Member, IEEE) received the B.S. degree from Tsinghua University, Beijing, China, in 2000, and the M.S. and Ph.D. degrees from Columbia University, New York, NY, in 2002 and 2005, respectively, all in electrical engineering.

She is a Senior Lecturer with the Research School of Computer Science, Australian National University, Canberra, Australia. She was with the IBM T. J. Watson Research Center, Hawthorne, NY, from 2005 to 2010. Her recent research interests are in multimedia, machine learning, and social media analysis.

Dr. Xie has won several awards: the best conference paper award in IEEE SOLI 2011 and the best student paper awards at JCDL 2007, ICIP 2004, ACM Multimedia 2005, and ACM Multimedia 2002. She also received the 2005 IBM Research Josef Raviv Memorial Postdoc fellowship in computer science and engineering.

Munmun De Choudhury received the B.Tech. degree in computer science from the National Institute of Technology, Bhopal, India, in 2005 and the Ph.D. degree in computer science from Arizona State University, Tempe, in 2011.

She is currently a Postdoctoral Researcher with Microsoft Research, Redmond, WA. Her work has appeared in various media outlets: *The Wall Street Journal*, *Mashable*, and *Technology Review*. Her research interest spans the area of computational social science: at the intersection of data mining, social science, and human–computer interaction.

Dr. De Choudhury has been a finalist of the Facebook Fellowship program in 2010, a best paper nominee at CSCW 2012, and a winner of the Grace Hopper Scholarship.

Yu-Ru Lin received the Ph.D. degree in computer science with a concentration in arts media and engineering from Arizona State University, Tempe, supervised by Dr. H. Sundaram.

She is a Postdoctoral Researcher with IQSS, Harvard University, Cambridge, MA, and CCIS, Northeastern University, Boston, MA. Her research interests have been in analysis and visualization of interpersonal activities in social networks—in particular, large-scale community dynamics, high-dimensional social information summarization, and representation. She has developed matrix and tensor-based techniques as well as visualization for analyzing community structures and evolutions in time-varying heterogeneous social networks.

Apostol (Paul) Natsev received the M.S. and Ph.D. degrees in computer science from Duke University, Durham, NC, in 1997 and 2001, respectively.

He is currently a Staff Software Engineer and Manager with the Video Content Analysis Group, Google Research, Mountain View, CA. Previously, he was a Research Staff Member with IBM Research, Hawthorne, NY, from 2001 to 2011, and Manager of the Multimedia Research Group from 2007 to 2011. He is an author of more than 70 publications and 18 patents (granted or pending). His research agenda is to advance the science and practice of systems that enable users to manage and search vast repositories of unstructured multimedia content. His research interests span the areas of image and video analysis and retrieval, computer vision, and large-scale machine learning.

Dr. Natsev's research has been recognized with several awards, including the 2004 Wall Street Journal Innovation Award, Multimedia category (for the IBM MARVEL system), 2005 IBM Outstanding Technical Accomplishment Award, 2005 ACM Multimedia Plenary Paper Award, 2006 ICME Best Poster Award, and 2008 CIVR Public's Choice Award (for the IBM IMARS system).