Summarization of Social Activity over Time: People, Actions and Concepts in Dynamic Networks

Yu-Ru Lin Hari Sundaram Aisling Kelliher
Arts Media and Engineering Program, Arizona State University
Tempe, AZ 85281
e-mail: {yu-ru.lin, hari.sundaram, aisling.kelliher}@asu.edu

ABSTRACT
We present a framework for automatically summarizing social group activity over time. The problem is important in understanding large scale online social networks, which have diverse social interactions and exhibit temporal dynamics. In this work we construct summarization by extracting activity themes. We propose a novel unified temporal multi-graph framework for extracting activity themes over time. We use non-negative matrix factorization (NMF) approach to derive two interrelated latent spaces for users and concepts. Activity themes are extracted from the derived latent spaces to construct group activity summary. Experiments on real-world Flickr datasets demonstrate that our technique outperforms baseline algorithms such as LSI, and is additionally able to extract temporally representative activities to construct meaningful group activity summary.

Categories and Subject Descriptors: H.3 INFORMATION STORAGE AND RETRIEVAL; H.3.3 Information Search and Retrieval; H.2 DATABASE MANAGEMENT; H.2.8 Database applications

General Terms
Algorithms, Experimentation, Measurement, Theory

Keywords
Community, Summarization, Social activity, Evolution, Non-negative Matrix Factorization, Social Network Analysis

1. INTRODUCTION
This paper focuses on the summarization of social activity in online social networks. Networks such as Flickr, Facebook and Myspace, allow for a diverse range of interactions amongst their members, resulting in large, temporal datasets relating users, media objects and actions. Automatic summarization of this collective activity will create great values: End users can make better use of shared media within their social groups. Business analysts can begin to understand the collective decision-making behavior in a group for marketing purposes. Finally, social science researchers can use the summary as a framework to understand how ideas emerge and diffuse. The problem is highly challenging since the social network activity changes rapidly over time, the interactions between users can have different semantics and finally due to the scale of the problem.

Recently, there has been work on analysis of social groups and their temporal dynamics [4], including mining the community network dynamics [5]. These works focus on dynamic but homogeneous networks, i.e. the edges represent homogeneous actions (e.g. posting). However in online social networks users can interact with each other though heterogeneous actions which have different semantics. Heterogeneous interrelated entities have attracted considerable interests recently [1,6,7]. It has been shown that combining multiple relationships such as the linkage and the document-term matrices is effective in text mining [8]. These multi-graph mining algorithms do not consider the temporal evolution of the interrelated entities. To summarize the dynamics and richness of social activity, we propose a unified temporal multi-graph framework for extracting activity themes over time. Experiments on real-world Flickr datasets demonstrate that our technique outperforms baseline algorithms such as LSI/SVD, and is additionally able to extract temporally representative activities to construct meaningful group activity summary.

CIKM'08, October 26–30, 2008, Napa Valley, California, USA.
ACM 978-1-59593-991-3/08/10.

2. PROPOSED METHOD
We propose a novel method for automatically summarizing social group actions over time. The major contribution in this work is two-folded, including a new summarization methodology and a unified mining algorithm.

The basic unit of analysis is an *activity* triple – *user*, *concept* and *action*. The concepts in our work refer to tags associated with photos e.g. "beach," "home." An *activity theme* for a group is a set of co-occurring activities for a certain time. Example activity themes include (also ref. Figure 1):

- "In October 2007, Alex, Becky and Colleen post photos about the Grand Canyon."
- "In November 2007, Alex, Daren and Eric comment on photos about the Kaibab trail in the Grand Canyon."

The summarization methodology identifies who (users), what (concepts), how (actions) and when (time) to represent the collective activities. We believe these are necessary aspects for understanding the collective activity in a social group. Our methodology is complementary to the concept summarization approach, e.g. tag cloud representation [3].

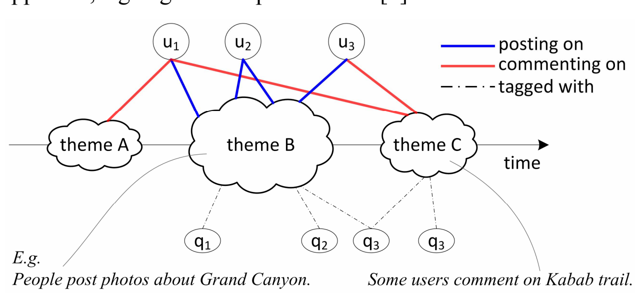

Figure 1: Illustration of temporal activity themes. An activity theme summarizes the collective activities in a group by identifying who (users u_1, u_2, etc.), what (concepts q_1, q_2, etc.), how (different actions, e.g. posting, commenting, etc.) and when

We propose a unified temporal multi-graph algorithm for extracting activity themes over time. Our algorithm differentiates action semantics by the type and the co-occurring time of actions. We combine four different networks (user-photo, user-comment, photo-tag, and comment-tag matrices) that represent different semantics of user actions. We use a non-negative matrix

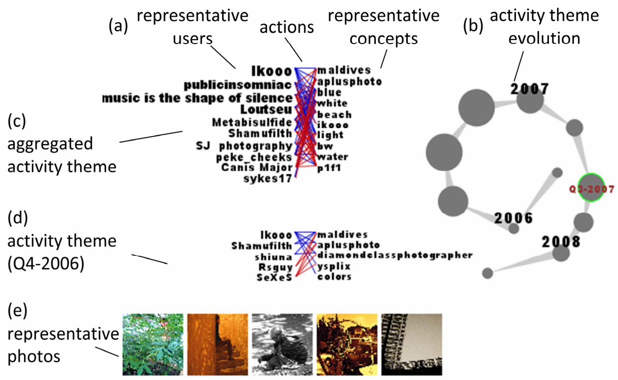

Figure 2: An example of visual representation of group activity summary. (a) An activity theme is represented by a bi-partite graph, where the left-side nodes represent users, the right-side nodes represent concept terms, blue edges represent "post-on" actions, and red edges represent "comment-on" actions. (b) The theme evolution is represented by a spiral map, where each bubble represents an activity theme. Consecutive activity themes are connected by an arrow with distance indicating the difference of two themes. When a bubble is highlighted, the corresponding activity theme (d) and related photos (e) are shown. (f) The aggregated activity theme shows representative activities aggregated by all activity themes.

factorization (NMF) approach to derive two interrelated latent spaces for users and concepts, where users and concepts are related to each other through different actions. We introduce time indicators (photo-time and comment-time matrices) as regulizer to enforce the temporal co-occurrence of actions. We provide an efficient iterative algorithm to solve the proposed objective function. Then, activity themes are extracted from the derived latent spaces to construct group activity summary. An example of the visual representation of group activity summary is shown in Figure 2.

Theme extraction. For each activity theme A_t at time t, we find a set of users, concept terms and actions as representative activities for A_t. We extract the top K users and terms from the solution matrices given by the matrix factorization, and then extract the observable actions among the top K users and terms.

Theme evolution. To determine how a theme evolves into another theme, we compute the similarity between themes by comparing the corresponding axes in both latent spaces. We use a cosine similarity measure to compare two axes.

A visual representation of theme evolution can be constructed as a spiral map, as shown in Figure 2(b). Each bubble represents an activity theme A_t, with size representing the amount of activities during t. Two adjacent themes A_t, A_{t+1} are connected by an arrow with distance proportional to the dissimilarity between A_t and A_{t+1}. When a bubble is highlighted, the corresponding activity theme (Figure 2(d)) and related photos (Figure 2(e)) are shown.

The visual representation help understand the collective social activity in a group. From the example shown in Figure 2(d), the users can easily observe that in Q4-2006, some users like "lkooo" and "Shamufilth" posted "aplushphoto" and got comments from others (e.g. "Rsguys").

3. EXPERIMENTS AND EVALUATION

We conduct extensive experiments using real-world datasets collected from a popular photo-sharing site, Flickr. We select 191 public Flickr groups and extract all users, tags, photos and comments for each group. The Flickr "groups" are used as a proxy for a realm of a social space.

We evaluate the quality of temporal group activity summarization based on two criteria, coverage and coherence, which are defined based on the topological relationship between the extracted activities and the total activities.

We compare our method with frequency based algorithms, "interestingness" measure [3] and the well-known latent semantic indexing method (LSI) [2]. Our method outperform baseline methods: the users extracted by baseline methods cover 23-64% of concepts that have appeared in this group, while our method has significant improvement with a coverage of 38-70% – our improvement is 27.2% on the average. Our method differentiates users/concepts by the types and times of actions, so the extracted users are diversely related to more different concepts. The results also suggest that our method gives a more coherent summary than baseline methods. Our method outperforms all baseline methods by consistently increasing the connections of extracted users and concept terms over time. This is because the users and concepts extracted by our method are based on their connections across different times, while other methods only consider the connection for a given time.

4. CONCLUSION AND FUTURE WORK

We propose a method for summarizing and representing social group activity over time. In our framework, we formulate the summarization problem as extraction of representative activity themes. This summarization framework help identify who (users), what (concepts), how (actions) and when (time) to represent the collective activities. We use a large collection of real-world datasets to show the effectiveness of our method. The reulsts demonstrate that our method is able to construct representative and meaningful group activity summary.

5. REFERENCES

[1] A. AGARWAL, S. CHAKRABARTI and S. AGGARWAL (2006). *Learning to rank networked entities*, SIGKDD, 14-23, 2006.

[2] S. DEERWESTER, S. DUMAIS, G. FURNAS, T. LANDAUER and R. HARSHMAN (1990). *Indexing by latent semantic analysis.* J. of the American Society for Information Science **41**(6): 391-407.

[3] M. DUBINKO, R. KUMAR, J. MAGNANI, J. NOVAK, P. RAGHAVAN and A. TOMKINS (2006). *Visualizing Tags over Time*, WWW, 193-202, 2006.

[4] R. KUMAR, J. NOVAK and A. TOMKINS (2006). *Structure and evolution of online social networks*, SIGKDD, 611-617, 2006.

[5] Y. LIN, Y. CHI, S. ZHU, H. SUNDARAM and B. TSENG (2008). *Facetnet: a framework for analyzing communities and their evolutions in dynamic networks*, WWW, 2008.

[6] X. WANG, J. SUN, Z. CHEN and C. ZHAI (2006). *Latent semantic analysis for multiple-type interrelated data objects*, SIGIR, 236-243, 2006.

[7] W. XI, B. ZHANG, Z. CHEN, Y. LU, S. YAN, W. MA and E. FOX (2004). *Link fusion: a unified link analysis framework for multi-type interrelated data objects*, WWW, 319-327, 2004.

[8] S. ZHU, K. YU, Y. CHI and Y. GONG (2007). *Combining content and link for classification using matrix factorization*, SIGIR, 487-494, 2007.