ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

MALLET-Privacy Preserving Influencer Mining in Social Media Networks via Hypergraph

Janani K[1], Narmatha S[2]

Assistant Professor, Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, L&T Bypass Road, Coimbatore, India[1].

PG Scholar, Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, L&T Bypass Road, Coimbatore, India[2].

ABSTRACT: Online Social Media Networks (OSNs) provide an online service for building social relations among users to share interests, images, audios and videos. A social network service represents each user's social links such as likes, comments, favorites and tags which are very useful for mining social influence. The social links indicate certain influence in the community. The existing system suffers from analyzing the generic influence but ignoring the more important topic-level influence. Since the content of interest is essentially topic-specific, the underlying social influence is topic-sensitive. To address these restrictions develop a Novel Topic-Sensitive Influencer Mining (TSIM) framework in social networks which aims to mine topic-specific influential nodes in the networks and find topical influential users and images. The influence estimation is achieved by using hyper graph learning approach in which the vertices represent users and images, and the edges represent multi-type relations include visual-textual content relations among images, and social link relations between users and images. Social influence mining is used in real applications like friend suggestion, photo recommendation, expert identification and social search. The proposed algorithm provides privacy framework for each user in Social Networks like Flickr.

KEYWORDS: Hypergraph learning, Influencer mining, Multiparty Access Control, Topic modeling, Topic influence, Topic distribution learning.

I. INTRODUCTION

The emergence and rapid proliferation of social media networks provides users an interactive sharing platform to create and share content of interest. For example, every minute of the day in 2012, there are 100,000 tweets sent on Twitter, 48 hours of videos uploaded to YouTube, and 3,600 photos shared on Instagram. In Flickr, users uploaded 1.54 million photos per day on average in 2012. In such interest-based social networks, users interact with each other through the content of interest. Such interactions forming the social links are well recognized forces that govern the behaviours of involved users in the networks.

In interest-based online social media networks, users can easily create and share personal content of interest, such as tweets, photos, music tracks, and videos. Photographers with popular photos in the network usually garner rich social links such as contacts, photo comments and favourites. The large-scale user-contributed content contains rich social media information such as tags, views, favourites, and comments, which are very useful for mining social influence. Mining such topic-sensitive influencers can enable a variety of applications, such as recommendation, social search, influence maximization for product marketing and adoption, etc.

II. SYSTEM ANALYSIS

The existing system includes two major contributions in topic sensitive influence mining for rich social media information. First, it proposes a TSIM framework for exploiting text information to learn the topic distribution and it

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

analyze the generic influence in homogeneous networks using Topic distribution learning algorithm. Second, OSNs often use user relationship and between trusted and set of group membership to distinguish untrusted users. For example, in Facebook, users can allow friends, friends of friends (FOF), groups, or public to access their personal authentication and data depending on privacy requirements. Although OSNs currently provide simple access control mechanisms allowing users to govern access to information have no contained in their own spaces, users, unfortunately, control over data residing outside their spaces. For instance, if a user posts a comment in a friend's space, she/he cannot specify which users can view the comment. In another case, when a user uploads a photo and tags friends who appear in the photo, the tagged friends cannot restrict who can see may have this photo, even though the tagged friends different privacy concerns about the photo. To address such a critical issue, preliminary protection mechanisms have been offered by existing OSNs. Fig.1 illustrates the proposed framework is Novel Topic- Sensitive Influencer Mining (TSIM) framework in interest-based social media networks aims to find topical influential nodes among users and images. The influence estimation is determined with hypergraph learning approach.

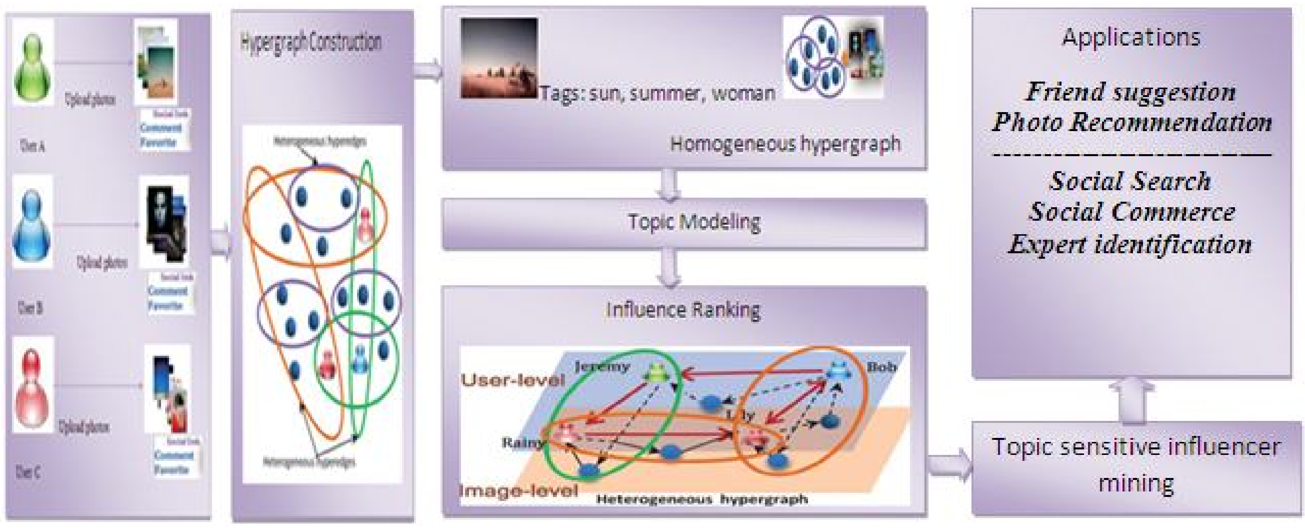

Fig.2. The Proposed Framework of TSIM

In the hyper graph, the vertices represent users and images, and the hyper edges are utilized to capture multitype relations including visual-textual content relations among images, and social links between users and images. Algorithm wise, TSIM first learns the topic distribution by leveraging user-contributed images, and then infers the influence strength under different topics for each node in the hyper graph. The System is used in real applications like friend suggestion, photo recommendation, expert identification and social search etc. The System pursues a systematic solution to facilitate collaborative management of shared data in OSNs. We begin by examining how the lack of multiparty access (MPAC) for data sharing in OSNs can undermine typical data sharing the protection of user data. Some patterns with respect to multiparty authorization in OSNs identified. Based on these sharing patterns, the core features of are also MPAC model is formulated to capture multiparty authorization requirements that have not been accommodated so far by existing access control systems. Model also contains a multiparty policy specification scheme. Meanwhile, since conflicts are inevitable in multi-party a voting mechanism is further provided to deal with authorization and privacy conflicts in the model.

III. Related Work

In this section, we briefly introduce the related work on hypergraph learning, topic modeling and influence ranking.

A. *Hypergraph Learning*

The hypergraph learning method combines the content of images and social links to determine the topic-specific influence for users and images in the network. In a simple graph, vertices are used to represent the samples, and an

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

edge connects two related vertices to encode the pairwise relationships. The graph can be undirected or directed, depending on whether the pairwise relationships among samples are symmetric or not. A hypergraph is a generalization of the simple graph in which the edges, called *hyperedges*, are arbitrary non-empty subsets of the vertex set [1]. The hyperedges are used to capture different high-order relations between users and images such as content similarity relations and social links. Therefore, the hypergraph can be employed to model both various types of entities and complex relations, which is extremely suitable in social media modeling.

B. *Topic Modeling*

Related work also includes topic modeling. We develop a topic model using MALLET tool. MALLET is a Java package "Machine Learning for Language Toolkit" and purpose of this tool is for people who want to do their own topic modeling. The topic model learns topics in a collection of documents, and tags each documents with a small number of topics. It includes routines for transforming text documents into numerical representations that can then be processed efficiently. A "topic" consists of a cluster of words that frequently occur together. Using contextual clues, topic models can connect words with similar meanings and distinguish between uses of words with multiple meanings.

C. *Influence Ranking*

Based on the learned topic distribution and the constructed hypergraph, we perform topical affinity propagation on the hypergraph with the heterogeneous hyperedges for measuring influence regarding topics for each user and image. The affinity propagation algorithm is originally used for clustering data to identify a subset of exemplars, which are used to best account for all other data points.

IV. MODULES

In this section, the proposed system consists of four modules: hypergraph construction, topic distribution learning and topic sensitive influence ranking and multiparty authentication.

A. *Hypergraph Construction*

The System use two types of vertices corresponding to the users and images, which constitute the vertex set. The edge set which contains homogeneous and heterogeneous hyperedges is introduced to capture the multi type relations among users and images.

Homogeneous hyperedges: The homogeneous hyperedges are used to represent the visual-textual content relations among image vertices. The system constructs two types of homogeneous hyperedges including visual content relation hyperedge and textual content relation hyperedge.

Heterogeneous hyperedges: The heterogeneous hyper edges are utilized to connect image vertices with user vertices to capture social link relations. The owner of the image and the users who post comments on or favourite the image are connected by the interest hyperedge. All the heterogeneous hyperedges weights are set to 1.

B. *Topic Distribution Learning*

The system utilizes the image vertices and homogeneous hyper edges in the hyper graph to learn the topic distribution. We propose to develop a hyper graph regularized topic model to fully leverage both content and context information of images to help learn the potential topics of interest. However, in real-world scenarios, user-contributed social media data is inevitably noisy and the textual information associated with images is usually sparse, which makes it difficult to use the hyper graph regularized topic model to learn topics of interest accurately.

C. *Topic Sensitive Influence Ranking*

Users share topical similarity which can be computed through the connected images and social links. The system can use affinity propagation to exchange influence messages between users and images along heterogeneous hyperedges in

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

the hypergraph. The final derived influence messages between users can be viewed as mutual social influence between users. In the system, the influence of users and images is recursively updated until it achieves the optimal condition.

D. *Multiparty Authentication*

This module represents privacy of each user in interest based social networks. For instance, In Social networks A, B and C are friends, if anyone of them upload his/her photo in it and they comment for that photo. The comments of the various users for that photo should hide in order to protect the privacy of each user.

V. EXPERIMENTS

In this section, we present various experiments to evaluate the effectiveness of the proposed topic-sensitive influencer mining approach on a dataset. In this section, we demonstrate its utility in friend suggestion and photo recommendation.

A. *Friend Suggestion*

Friend suggestion is useful and valuable in social media communities. Finding potential friends sharing similar interest for users can improve user experience. The ground truth of evaluation is the *comments* function which is generated by users. Based on the content of interest, the system suggests friends to each user.

B. *Photo Recommendation*

We also conduct the experiment on photo recommendation. We adopt the favourite list of each user. Considering the sparse favourites links between users and photos in the dataset. We filter out users who have less than 400 favourites in the dataset and we obtain 230 users. These users' favourites are taken as test data for evaluation purpose. In the learning stage, we remove all the corresponding information of these users' favourites. In addition to the methods compared for friend suggestion, we add two simple approaches: 1) Using social, visual, and textual signals of favourite photos for recommendation 2) A simple weighted sum of content similarity and social interactions the weighted parameter is empirically set to 0.5

VI. CONCLUSION AND FUTURE WORK

TSIM aims to find the influential nodes in the networks and the system use a unified hypergraph to model users, images, and various types of relations. The influence estimation is determined with the hypergraph learning approach. We have justified the motivation that (1) the content information of images contributes to the topic distribution, and (2) social links between users and images indicate the underlying social influence of users and images in the social media networks. We also demonstrate that TSIM can improve the performance significantly in the applications of friend suggestion, photo recommendation, expert identification and social search which reveal the potential of interest graph in reshaping the social behaviours of users in the networks. It is worth noting that the approach can be seamlessly generalized to many other interest-based network sharing platforms. We develop a systematic solution to facilitate collaborative management of shared data in OSNs. We begin by examining how the lack of multiparty access control (MPAC) for data sharing in OSNs can undermine the protection of user data. In the future, we will be working towards into two directions: 1) applying the proposed TSIM in more applications such as behaviour prediction and social commerce to verify its extensive effectiveness; and 2) investigating the topic-specific influence mining over time in the networks.

REFERENCES

1. Anagnostopoulos, R. Kumar, and M. Mahdian, "Influence and correlation in social networks," in Proc. KDD, pp. 7–15, 2008.
2. D. Angluin and Jiang Chen "Learning a Hidden Hypergraph," in proc. The Journal of Machine Learning Research, pp. 2215-2236, 2006.
3. D. Cai, Q.Mei, J. Han, and C. Zhai, "Modeling hidden topics on document Manifold," in Proc. CIKM, pp. 911–920 , 2008.
4. J. Weng, E. Lim, J. Jiang, and Q. He, "Twitterrank: Finding topic-sensitive influential twitterers," in Proc. 3rd ACM Int. Conf. Web Search and Data Mining, pp. 261–270, 2010.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

5. J. Yu, D. Tao, and M. Wang, "Adaptive hypergraph learning and its application in image classification," IEEE Trans. Image Process., vol.21, no. 7, pp. 3262–3272, 2012.
6. L. Liu, J. Tang, J.Han, M. Jiang and S.Yang, "Mining topic-level influence in heterogeneous networks," in Proc. CIKM, pp. 199–208, 2010.
7. Li Pu and Boi Faltings, "Hypergraph Learning with Hyperedge Expansion," in Proc. SPRINGER, pp. 410–425, 2012.
8. S. Agarwal, K. Branson, and S. Belongie, "Higher order learning with graphs," in Proc. ICML, pp. 17–24, 2006.
9. T. H. Haveliwala, "Topic-sensitive page rank," in Proc. WWW, pp. 517–526, 2002.
10. Yun Yang, Wenwu Zhu and Shigiang Yang, "User Interest and Social Influence Based Emotion Prediction for Individuals," in Proc. KDD, pp. 10–14, 2013.