Survey on Privacy Preserving Influencer Mining in Social Media Networks via Hypergraph

S. Narmatha, K. Janani
Computer Science and Engineering, Sri Shakti Institute of Engineering and Technology, Coimbatore, Tamilnadu, India

Abstract — A social networking service is an online service, platform, or site that focuses on facilitating the building of social networks or social relations among people who, for example, share interests, activities, backgrounds, or real-life connections. A social network service consists of a representation of each user, his/her social links, and a variety of additional services. Most social network services are web-based and provide means for users to interact over the Internet, such as e-mail and instant messaging. Online community services are sometimes considered as a social network service, though in a broader sense, social network service usually means an individual-centered service whereas online community services are group-centered. Social networking sites allow users to share ideas, activities, events, and interests within their individual networks. The large-scale user-contributed content contains rich social media information such as tags, views, favorites, and comments, which are very useful for mining social influence. The social links such as views, favorites, and re tweets, indicate certain influence in the community. Since the content of interest is essentially topic-specific, the underlying social influence is topic-sensitive. Novel Topic-Sensitive Influencer Mining (TSIM) framework aims to mine topic-specific influential nodes in the networks and find topical influential users and images. The influence estimation is determined with a hyper graph learning approach. 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. The influence estimation is determined with a hyper graph learning approach. 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.

Keywords — Hypergraph learning, Influencer mining, 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 behaviors of involved users in the networks. Mining the topic-sensitive influencers can enable a variety of applications, such as recommendation, social search, influence maximization for product marketing. While this paper focusing on analyzing the direct social influence is important, mining influence strength at the topic-level in heterogeneous networks has been largely ignored. To this end, mine the topic-level influence by utilizing textual information and link information. The rich social media information is used to quantitatively measure social relationship strength of users and to addressing the problem of influence strength measure with awareness of topics of interest and combination of heterogeneous data from multiple modalities simultaneously in social media networks. This problem can be estimated by hypergraph learning. Hypergraph is a simple graph, vertices are used to represent the samples, and an 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. Therefore, the hypergraph can be employed to model both various types of entities and complex relations, which is extremely suitable in social media modeling.

II. LITERATURE REVIEW

A. LEARNING A HIDDEN HYPERGRAPH

Basically, an independent covering family of a hypergraph is a collection of independent sets that cover all non-edges. An interesting observation is that the set of negative queries of any algorithm that learns a hypergraph drawn from a class of hypergraphs that is closed under the operation of adding an edge is an independent covering family of that hypergraph. Note both the class of r-uniform hypergraphs and the class of (r, Δ)-uniform hypergraphs are closed under the operation of adding an edge. This implies that the query complexity of learning such a hypergraph is bounded below by the minimum size of its independent covering families. In the opposite direction, we give subroutines to find one arbitrary edge from a hypergraph. With the help of the subroutines, we show that if we can construct small-sized
independent covering families for some class of hypergraphs, we are able to obtain an efficient learning algorithm for it. In this paper, we give a randomized construction of an independent covering family of size $O(r^2 2r \log n)$ for $r$-uniform hypergraphs with $m$ edges.

In this paper, first give an algorithm that finds an arbitrary edge in a hypergraph of dimension $r$ using only $r \log n$ edge-detecting queries. The algorithm is adaptive and takes $r \log n$ rounds. The success probability in the construction of independent covering families in the previous section can be easily improved by drawing more samples. Using the high-probability version of the construction, we obtain an algorithm using a number of queries that is quadratic in $m$ that learns an $r$-uniform hypergraph with $m$ edges with high probability. Although the first algorithm for finding one edge is deterministic and simple, the round complexity $r \log n$ might be too high when $n$ is much larger than $m$. We then improve the round complexity to $O(\log m + r)$ using only $O(\log m \log n)$ more queries.

B. IMAGE RETRIEVAL VIA PROBABILISTIC HYPERGRAPH RANKING

The System proposes a hypergraph based transductive algorithm to the field of image retrieval. Based on the similarity matrix computed from various feature descriptors, we take each image as a ‘centroid’ vertex and form a hyperedge by a centroid and its $k$-nearest neighbors. To further exploit the correlation information among images, we propose a novel hypergraph model called the probabilistic hypergraph, which presents not only whether a vertex $v_i$ belongs to a hyperedge $e_j$, but also the probability that $v_i \in e_j$. In this way, both the higher order grouping information and the local relationship between vertices within each hyperedge are described in our model. To improve the performance of content-based image retrieval, we adopt the hypergraph-based transductive learning algorithm proposed in to learn beneficial information from both labeled and unlabeled data for image ranking. After feedback images are provided by users or active learning techniques, the hypergraph ranking approach tends to assign the same label to vertices that share many incidental hyperedges, with the constraints that predicted labels of feedback images should be similar to their initial labels. The contribution of this paper is threefold:

- The system proposes a new image retrieval framework based on transductive learning with hypergraph structure, which considerably improves image search performance
- The system propose a probabilistic hypergraph model to exploit the structure of the data manifold by considering not only the local grouping information, but also the similarities between vertices in hyperedges.
- In this work we conduct an in depth comparison between simple graph and hypergraph based transductive learning algorithms in the application domain of image retrieval, which is also beneficial to other computer vision and machine learning applications.

Presents an active learning framework, in which a fusion of semi-supervised techniques (based on Gaussian fields and harmonic functions) and SVM are comprised. In a pairwise graph based manifold ranking algorithm is adopted to build an image retrieval system. In a simple graph both labeled and unlabeled images are taken as vertices; two similar images are connected by an edge and the edge weight is computed as image-to-image affinities. Depending on the affinity relationship of a simple graph, semi-supervised learning techniques could be utilized to boost the image retrieval performance.

C. USER INTEREST AND SOCIAL INFLUENCE BASED EMOTION PREDICTION FOR INDIVIDUALS

Emotions are playing significant roles in daily life, making emotion prediction important. To date, most of state-of-the-art methods make emotion prediction for the masses which are invalid for individuals. In this paper, we propose a novel emotion prediction method for individuals based on user interest and social influence. To balance user interest and social influence, we further propose a simple yet efficient weight learning method in which the weights are obtained from users’ behaviors.

The problem of emotion prediction for individuals is not trivial. So far, there are fewer works on emotion prediction for individuals. Emotions have long been viewed as passions produced on their own interest. However from social aspect, has shown that how happy you’re is influenced by your social links to people in social networks.

More recently, Tang’s work quantitatively studies how an individual’s emotion is influenced by his friends in social network. It can be seen from the above that existing emotion prediction methods for individuals either focus on user interest or social influence. However, neither user interest or social influence alone can predict individual’s emotion accurately. In this work, we propose a novel method jointly considering user interest and social influence in social network platform to predict user’s emotion. We further propose a simple yet efficient weight learning method to balance the weights of user interest and social influence to figure out exactly what kind of roles they are playing in the final emotion prediction.

Fig.1. A simple graph of six points in 2-D space nearest neighbors form a hyperedge

Fig.2. A hypergraph is built, in which each vertex and its 2
The conceptual framework of this paper is emotion prediction for individuals. With the popularity of social network, i.e. Facebook, Twitter and Flickr, more and more people are willing to share their own feelings towards hot events or their experiences in daily life, whether delivering positive or negative emotions, which makes it easier for people to know others’ minds. More or less, users are getting increasingly easier to be influenced by others in social network. However, different friends may have different extents of influences on the user based on how close they are or how much similarity they have in common. We use the emotion similarity when treated the same microblog to measure the social influence.

D. HYPERGRAPH LEARNING WITH HYPEREDGE EXPANSION

Many tasks require clustering in a graph where each edge represents a similarity relation. Often, it is a co-occurrence relation that involves more than two items, such as the co-citation and co-purchase relations. The co-occurrence relation can be represented by a hyperedge that connects two or more vertices in a hyper graph. But most clustering algorithms, such as k-means, or spectral clustering, are defined for graphs but not hypergraphs. Therefore, hyperedge relations are often transformed into another graph that is easier to handle. For classification and clustering tasks, the hyperedges are usually transformed into cliques of edges. This category of techniques includes clique expansion, star expansion. We shown a simple example of such transformation from a hypergraph to a graph (the induced graph). Since the transformations are carried out on the vertex level, we call them vertex expansions. With a vertex expansion, evaluating the goodness of clustering is done on the induced graph. For example, in a hyperedge of k vertices, a cut that separates the hyperedge into 1 and k − 1 vertices would cut k − 1 pairwise edges, while a cut that splits the vertices in two equal halves would have k 2 /4 cut edges. Thus the vertex expansion would prefer an unbalanced clustering. To mitigate the problem of unbalanced clustering, it is proposed in star expansion and NHC to use the cluster volume as a normalizer for balancing the cluster sizes. But such normalization can not completely eliminate the problem.

We present the following example of vertex embedding to explain why the problem still exists.

By computing the eigenvectors of the normalized Laplacian LN HC of the induced graph, it is possible to project the vertices into an Euclidian space, which is called embedding in spectral graph learning.. Although the hyperedges have the same weight and the cluster volume normalizer is applied, the overlapping part is still biased to the side with less vertices (in this case e2 side). This means that the optimal clustering of two clusters should assign the overlapping part and other vertices in e2 to one cluster. Such bias might be a problem when the hyperedge sizes are unbalanced, e.g. co-citation relations with a lot or a few citations. Moreover, the behavior of the artificial normalization (or “correction”) could be undesirable when many hyperedges intersect with each other, because the cost of the clustering would depend on how a hyperedge is split into the clusters. An even split would introduce a different cost compared to an uneven split. As any hyperedge that is not entirely within the same cluster represents a relation that is violated by the clustering, it would be natural to have the learning result independent of the hyperedge sizes and only depend on the hyperedge connectivity and hyperedge weights. We present a new transformation called hyperedge expansion (HE) based on a network flow technique so that the learning result is invariant to the distribution of vertices among hyperedges. HE expansion is first carried out on the hyperedge level. Then the learning results on hyperedges are projected back to the vertices through the adjacency information between hyperedges and vertices.

III. RELATED WORKS

Quan Fang and Jitao Sang present various experiments to evaluate the effectiveness of the proposed topic-sensitive influencer mining approach on a real-world dataset from Flickr. Jitao Sang evaluate the performance of the hypergraph regularized topic model on learning topic distribution. They then qualitatively and quantitatively verify the effectiveness of TSIM and indirectly demonstrate its utility in friend suggestion and photo recommendation.

IV. CONCLUSION

The System study the problem of Topic-Sensitive Influence Mining(TSIM) in Flickr. TSIM aims to find the influential nodes in the networks. The System use a unified hypergraph to model users, images and various types of relations. The system justify 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 study that TSIM can improve the performance significantly in the applications of friend suggestion and photo recommendation, which reveals the potential of interest graph in reshaping the social behaviors of users in the networks. It is worth noting that our approach can be seamlessly generalized to many other interest-based network sharing platforms. In the future, the System will be working towards two directions. (1) Applying the proposed TSIM in more applications such as expert identification and social search to verify its extensive effectiveness. (2) Investigating the topic-specific influence mining over time in the networks.

ACKNOWLEDGEMENT

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organization. I would like to extend my sincere thanks to all of them.

I am highly indebted to Sri Shakthi institute of engineering and technology and Mrs. Janani for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project. I would like to express my gratitude towards my parents & member of Sri Shakthi institute of engineering and technology for their kind co-operation and encouragement which help me in completion of this project. I would like to express my special gratitude and thanks to industry persons for giving me such attention and time.

My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities.

REFERENCES