IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

EVENT BASED ANALYSIS IN SOCIAL NETWORK USING THE APPROACHES OF MACHINE LEARNING

Naga Hari Babu KV ,Swapna Vanguru ,V Bhavani

[1]Assistant Professor, [2]Assistant Professor, [3]Assistant Professor

[1]Computer Science and Engineering,

[1]Keshav Memorial Engineering College, Hyderabad, India

Abstract : The importance of machine learning for social network analysis is realized as an inevitable tool in forthcoming years. This is due to the rapid growth of data in social network, increased by the proliferation of social media websites and the embedded heterogeneity and complexity.

In recent times, Social Network Analysis has become a very important and interesting subject matter with regard to Artificial Intelligence in that a vast variety of processes, comprising animate and inanimate entities, can be examined by means of SNA. In this regard, SNA has been employed by security agencies for counter-intelligence and law enforcement purposes. SNA has been used in medicine and pharmaceuticals for gaining insights into protein-protein interactions. Also, SNA has been employed in the World Wide Web (WWW) for hyperlink analysis, cyber society analysis, sentiment analysis, etc. Furthermore, prediction tasks within social network structures have become significant research problems in SNA.

Thus, hidden facts and details embedded in social network structures can be effectively and efficiently harnessed for training AI models with the goal of predicting several missing components (such as links/ties, nodes/actors, structure type, etc.) within a given social network. Therefore, important factors such as the individual attributes of spatial social actors, and the underlying patterns of relationship binding these social actors must be taken into consideration; because these factors are 1 relevant in understanding the nature and dynamics of a given social network structure. SNA is a subdomain (or research topic) within the domain (or research area) of AI; and several open problems still exist with regard to SNA. Some of these open problems with respect to SNA are included .

This paper proposes two important things.

1) the various approaches which dealt by machine learning with concern to deep learning viz: Information Diffusion, Event-Based Analysis,. Thus, in this dissertation, we have proposed effective Machine Learning (ML) approaches toward resolving the following SNA (research) problems, namely: Breakup Prediction, Link Prediction, Node Classification, Event-based Analysis, and Trend/Pattern Analysis.

Keywords: Social Networks, Machine learning, Clustering, Classification, WWW

1. INTRODUCTION

Social Network is an emerging area of research today. The amount of information carried by Online Social Networks in the form of text and images is of immense value for data mining and knowledge extraction. There are many approaches to social network analysis including machine learning. In recent times, Social Network Analysis has become a very important and interesting subject matter with regard to Artificial Intelligence in that a vast variety of processes, comprising animate and inanimate entities, can be examined by means of Social Network Analysis .

In this regard, Social Network Analysis(SNA) has been employed by security agencies for counter-intelligence and law enforcement purposes. SNA has been used in medicine and pharmaceuticals for gaining insights into protein-protein interactions. Also, SNA has been employed in the World Wide Web (WWW) for hyperlink analysis, cyber society analysis, sentiment analysis, etc. Furthermore, prediction tasks within social network structures have become significant research problems in SNA.

The analysis of social media is now no longer limited to the scrutiny of the structure of the network, but also includes a deeper study of the characteristics of the users who are part of the network. Several approaches exist for studying and analysing these networks; however, machine learning algorithms have emerged as one of the most suitable approach, considering the volume and dynamic nature of the underlying data.

The very fundamental idea behind Machine Learning is that systems can learn from user's data; recognize patterns and draw conclusions with minimum human interference. Machine Learning approaches have the ability to evolve. The iterative aspect of machine learning exposes the models to new data, giving it the ability to adapt. They learn from previous computations to produce reliable, repeatable decisions and results. This increases the efficiency of the system, making it more robust.

Machine learning techniques have proven their usefulness in a variety of applications and Social Network Analysis (SNA) [1] is one of them. Social media analysis (SNA) involves congregation and investigation of data obtained from social networking sites. SNA allows us to study the network and the components comprising the network. An online social network (OSN) [2]comprises of users, and is characterized by the interaction between them. A detailed study of these interactions would provide us with intricate details about the users. What makes machine learning an attractive tool in this scenario is the sheer amount of data availability. OSN provides a truckload of information about the users and their interactions, which is used to train the machine learning model. This inherent reliability on data makes machine learning models an attractive prospect for performing SNA.

1.1 Motivation and Objectives

Social Network, SN: This is a tuple comprising a graph, G(V,E); a metadata function, fV , which extends the definition of the vertices' set by mapping it to a given set of attributes, V ′; and a metadata function, fE, which extends the definition of the edges' set by mapping it to a given set of attributes, E′. Thus, G(V,E) ⊂ SN as expressed via equation 1.1. Also, Figure 1.1 depicts a SN structure.

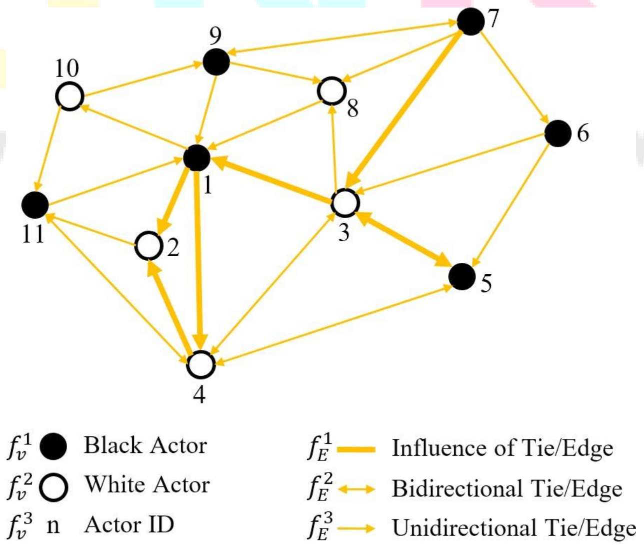


Figure 1.1: Social network (SN) structure.

SN = (G, fV , fE) ≡ (V,E, fV , fE)

fV : V → V ′ vertices' metadata function

fE : E → E′ edges' metadata function

II. LITERATURE REVIEW

2.1 The Development Of Social Network Analysis:

In social science, the structural approach that is based on the study of interaction among social actors is called social net-work analysis. The relationships that social network analysts study are usually those that link individual human beings. But important social relationships may link social individuals that are not human, like ants or bees or deer or giraffes or apes. Or they may link actors that are not individuals at all. Network analysts often examine links among groups or organizations—even among nation-states or international alliances. The social network approach is grounded in the intuitive notion that the patterning of social ties in which actors are embedded has important consequences for those actors.

All four of these features are found in modern social network analysis, and together they define the field:1. Social network analysis is motivated by a structural in-tuition based on ties linking social actors,2. It is grounded in systematic empirical data,3. It draws heavily on graphic imagery, and4. It relies on the use of mathematical and / or computation all models.

IIIProposedFrameWork:

Our paper mainly focus on the various approaches which dealt by machine learning with concern to deep learning viz: Information Diffusion, Event-Based Analysis, Thus, in this dissertation, we have proposed effective Machine Learning (ML) approaches toward resolving the above SNA problem, namely: Breakup Prediction, Link Prediction, Node Classification, Event-based Analysis, and Trend/Pattern Analysis.

3.1.1 Information Diffusion:

Information diffusion is a process by which information about new opinions, be- haviors, conventions, practices, and technologies flow from person-to-person through a social network. Studies on information diffusion primarily focus on how information diffuses in networks and how to enhance information diffusion.

Information diffusion is one of the popular research topics in social network analysis. Most of the these work address the following two questions:

(a) Analyze how information diffuses in OSNs and

(b) How to increase information diffusion

To address the above two questions, information diffusion has been modeled using two popular models, namely IC model and LT model. IC model is an information push/ sender-centric model, where each active user has an opportunity to activate his inactive neighbors with a given probability. LT model is an information pull/ receiver-centric model, where a node is activated by his neighbors if their aggregated weight surpasses his own influence limit. For example, spread of disease in a network can be modeled using IC model whereas spread of opinion in a network can be modeled using LT model.

Information diffusion is a process of disseminating information from an individual orcommunity to another in a social network. Understanding information diffusion in social networks can help to enhance business performance, increase audience engagement, improve personalized recommendation system, and develop a better opinion mining system. Diffusion is usually successful when information reaches to a largenumber of users in the network. Social networking services allow users to diffusethe information through various reactions such as like, comment, share, and retweet. Business, organizations, and individuals in social networks yearn to increase information diffusion by increasing the number of reactions. A post content with a large number of reactions can increase the visibility of the content, build the reputation of the content creator, and attract other users to give their reactions. Existing studieson information diffusion mainly focus on addressing two questions:

(a). how a piece of information spread in social networks (b). how to enhance the information diffusion

The main focus of this thesis is to enhance information diffusion by optimizing the factors that affect information diffusion.

3.1.2 Determinants of Information Diffusion

An ample amount of content is generated every day in social media. One of the main goals of content creators is to disseminate their information to a large audience. Manyfactors affect the information diffusion which includes network connectivity, posting time, post content, location, sentiment, etc. In this thesis, we study three of the most important factors namely, network connectivity, posting time, post content that can highly affect the information diffusion and develop new methods to increase the information diffusion

Network Connectivity

A highly connected authoritative user of a social network can diffuse the information widely compared to an ordinary user of the network. To find authoritative users,we can model a social network as a graph, where nodes represent users and edges represent a relationship between users. For example, it can be seen in Figure 3.1, ifa few authoritative users (i.e., green nodes) pass the information to other connected users, it can widely spread across the network. Authoritative users can be usedto advertise a given topic through word-of-mouth (WoM) marketing as these usersget the most attention. WoM marketing is one of the trusted and cost-effectiveforms of marketing where products are advertised through friends, family, or known authorities. In this thesis, we present a novel algorithm to find authoritative users in online social groups (OSGs) such as Facebook Groups.

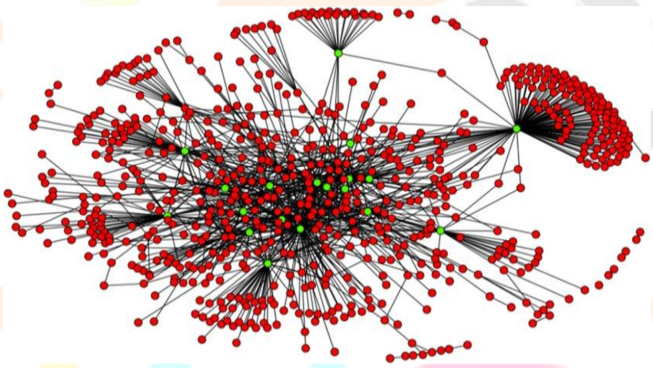


Figure 3.1: Network structure of a online social group

As authoritative position of users varies across the topics , it is more useful to find topic-sensitive authoritative users. To this end, we propose a topic-sensitive social interaction graph where edge weights are dynamically computed based on users' interactions and similarity of a post to advertising topic. The interactions couldbe in the form of likes, likes-on-comment, shares, and comments. To find prominent topic-sensitive authoritative users, we employ link analysis from social network analysis on the topic-sensitive social interaction graph.

Posting Time

Lifetime of social media content is very short, which is typically a few hours. As it can also be seen in Figure 1.2, 50% of reactions are received within four hours of posting on Facebook pages. The main obstacle in getting high information diffusion is that a post has to compete with many other posts within its very short lifetime. Ifa post content is posted at the time when audience are not online or not interested in interacting with the content, the content will not receive a large number of reactions. Thus a post with less number of reactions will not diffuse to a large audience. Onthe other hand, a post created at right time can lead to a higher number of reactions and thereby increase the information diffusion. The newsfeed ranking algorithms of social networks also use social interaction counts to determine the rank of the post in the audience feed. In this thesis, we look at the problem of finding the best posting time(s) for a given type of content for it to get a high audience reaction.

For our analysis, we use Facebook pages from five domains, namely e-commerce,traffic, telecommunication, hospital, and politics. To find the best time to post, we derive two classes of schedules: posting-based schedules and reaction-based schedules. The posting-based schedules are computed based on the post creation time. Many admins of Facebook pages may not be aware of when they should post to get maximumaudience reactions. However, a few admins with knowledge of Facebook News Feed ranking might have an intuition of when they should post to get maximum audience reactions. We therefore propose three posting based schedules that are based on frequent posting timings. Since our goal is to maximize the number of audience

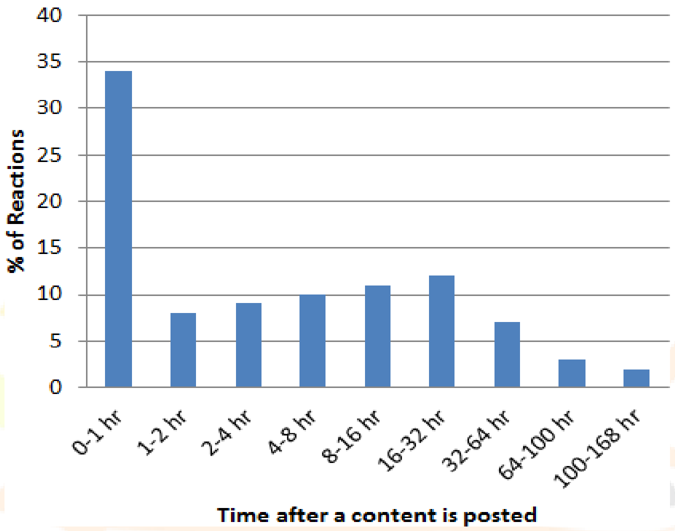


Figure 3.2: Average reactions per post in different time intervals

reactions, the reaction-based schedules are computed based on reaction timings of audience on posts. We analyze the audience reaction timings on created posts and recommend the best time to create a new post for getting a high audience reaction.

POST CONTENT

The content of a post plays a crucial role to determine its success or popularity in social networks. Naveed et al. showed that content of a post is one of the most important factors in its popularity. Popularity is measured in terms of audience reactions such as likes, comments, shares, retweets, etc. Among the three popular types ofaudience reactions in social media, namely likes, comments and shares, interaction in

the form of comments is the most informative. Through comments, users can expresstheir opinions. We say that a post has "high arousal" content if it can attract a huge number of comments from users. In other words, a post has high arousal content ifit is on some debatable topic. If a post is getting a huge number of comments, thenin social media it will get higher rank because many people would be reading thepost and also the corresponding comments. A high arousal content can be useful in various applications such as analyze how a user perceives the news, enhance existing post recommendation systems, increase information diffusion, and understand collective behavior of users' opinions. In this thesis, we predict high arousal news posts published in social media news pages

3.1.2 Event Based Analysis:

Event prediction in social network structures remains a very interesting research problem with respect to SNA. This impels understanding the intrinsic relationship patterns pre- serving a given social network structure, based on the study of several structural properties computed on the constituent social units, with respect to space and time. In this regard, tackling problems of this nature is considered NP-Complete . Consequently, herein in this dissertation, we proposed a unique DL approach based on deep-layer stacks of Multi- Layer Perceptron (MLP)s. The proposed model is capable of resolving event-prediction related problems about a target social unit, y, based on the intrinsic patterns of relation- ship learnt from one or more neighboring social units, x. Additionally, this model is appended with an adjustment-bias (ab) vector, at its output layer, so as to improve the accuracy and precision of predictions made with respect to the target unit (or node). This research was supported by International Business Machines (IBM) via the provision of computational resources necessary to carry-out our experiments herein.

3.1.3 Event-based Social Network Examples

On these web services, people may propose social events, ranging from informal get-togethers (e.g. movie night and dining out) to formal activities (e.g. technical conferences and business meetings). In addition to supporting typical online social networking facilities (e.g. sharing comments and photos), these event-based services also promote face-to-face offline social interactions. Till today, many of these services have attracted a huge number of users and have been experiencing rapid business growth. For example, Meetup has 9.5 million active users, creating 280, 000 social events every month; Plancast has over 100, 000 registered users and over 230, 000 visits per month.

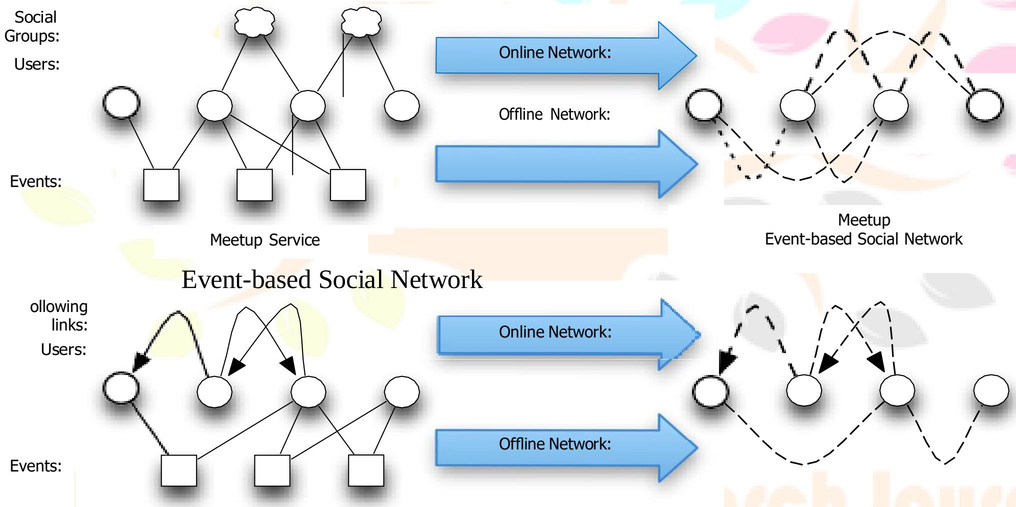


Figure 1: Event-based Social Network Examples

As these event-based services continue to expand, we identify a new type of social network event-based analysis in social network (EBASN) – emerging from them. Like conventional online social networks, EBASNs provide an online virtual world where users exchange thoughts and share experiences. But what distinguishes EBASNs from conventional social networks is that EBASNs also capture the face-to-face social interactions in participating events in the offline physical world. Fig. 1 depicts two example EBSNs from Meetup and Plan- cast. In Meetup, users may share comments, photos and event plans with members in the same online social groups (e.g. "bay area photographers", "Nevada county walkers"). In Plancast, users may directly "follow" others' event calendars. Bi-directional co-memberships of online social groups in Meetup or uni-directional subscriptions in Plancast ultimately constitute an online social network represented as the dashed lines on the right side of Fig. 1. Meanwhile, in both cases, users' co-participations of the same events derive their offline social connections. These connections collectively form an offline social network denoted as dotted lines in Fig. 1. The online and offline social interactions jointly define an EBASN.

As to be shown in our analysis, social events present very regular temporal and spatial patterns. In addition, both online and offline social interactions in EBASNs are extremely local. For example, we found that 70.65% of Meetup online friends and 84.61% of Meetup offline friends live within 10 miles of each other. To our surprise, the degree distributions of the Meetup EBASN do not follow the usual power law distribution, but are more heavy- tailed than power law. Furthermore, we found that the on- line and offline social interactions in an EBASN are positively correlated, implying a synergistic relationship between the two parts.

Community structure detection is a very useful approach for analyzing social networks. However, to correctly detect communities in an EBASN, one has to consider both online and offline social interactions. In this paper, we analyze the performance ratio of online event based and offline event based, we demonstrate the advantage of this method to other approaches. We also observed that the detected communities in the Meetup EBASN are more cohesive than those of the Gowalla LBSN.

Due to the short life time of an event, the event participation recommendation problem significantly differs from the usual recommendation problem for movies or places. Recommendation of an event is only valid after the event is created and before the event starts. This leads to a cold- start problem. In this paper, we design a number of diffusion patterns that capture the information flow over the heterogeneous EBASNs.

IV **EVENT-BASED SOCIAL NETWORKS:**

In this section, motivated by popular event-based social services, we define event-based social networks an describe how to construct the networks from collected datasets.

Event-based social services

As various online social networking services become prevalent, a new type of event-based social service has emerged. These web services help users to create social event proposals, disseminate the proposals to related people, and keep track of all participants. To foster efficient communication and sharing, these event-based services also provide online social networking platforms to connect users with others with similar interests. Below, we describe two examples of such event-based social services: online and offline.

4.1 Online Social Events:

online social event service that helps people publish and participate in social events. In online, a social event is created by a user by specifying when, where and what the event is. Then, the created social event is made available to selected users or public, controlled by the event creators. Other users may express their intent to join the event by RSVP (“yes”, “no” or “maybe”) online. To facilitate online interactions, we use one URL (Website) also allows users to form social groups (e.g. “bay area single moms”, “Nevada county walkers”) to share comments, photos and event plans.

Similar to online, offline is also used by many people that helps users create and organize events offline

Users also RSVP to express their intent to join social events. In contract to online users are authorized to follow the scheduled events

4.2 Offline Social Events:

offline social event service that helps people to follow and participate in social events. In offline, a social event is followed by a user by specifying when, where and what the event is. Then, the social event is made available to selected users or public, controlled by the event creators. Other users may express their intent to join the event by RSVP (“yes”, “no” or “maybe”) online.

Similar to online, offline is also used by many people that helps users create and organize events offline

Users also RSVP to express their intent to join social events. In contract to online users are authorized to follow the scheduled events

V Event based analysis social network :

Based on the event based Based on the event-based social services described above, we formulate a new type of social network, called an

event- based social network (EBASN). Like any social network, EBASNs capture social interactions among users.

However, different from others, ESABNs incorporate two forms of social interactions:

a) Online social interactions b) Offline social interactions

a) Online social interactions. In EBASNs, users can interact with each other online without the need of physical contact.For example, people can share thoughts and experiences with those in the same social group in Meetup. In Plancast, user comments and event plans are pushed to those who "follow" the user.

b) Offline social interactions. Social events play a major role in EBASNs. In a social event, people physically get together at a specific time and location, and do something together. Therefore, the social events in EBASNs represent the offline social interactions among event participants.

Definition: Formally, we define an EBASN as a heterogeneous network $G = \langle U, A^{on}, A^{off} \rangle$, where U represents the set of users (vertices) with $|U| = n$, A^{on} stands for the set of on-line social interactions (arcs), and A^{off} denotes the set of offline social interactions (arcs). The online social interactionsof an EBSN form an online social network $G^{on} = \langle U, A^{on} \rangle$, and the offline interactions of an EBASN compose an offline social network $G^{off} = \langle U, A^{off} \rangle$.

Note that the online social network or the offline social network of a EBASN can be either directed or undirected.For simplicity, we only focus on undirected online and offline networks in this paper.

The online social network or the offline social network alone is not new and has been studied extensively before. But the co-existence of both is what makes EBASNs special. As shown later in this paper, these two forms of social networks in EBASNs are intertwined but also have their own distinct characteristics at the same time.

VI. FLOW OF DATA IN EBASN'S:

In this section, we study how information flows over this unique network structure. A good scenario that can be used to examine the information flow on EBSNs is the problemof recommending users to participate in social events *only* based on the topological structure of EBSNs. With this application, we can study how information flows from one user to the online/offline friends and how the information flow pathways latently drive the social event participationprocess.

Unlike classic movie/book recommendations, event participation recommendation is more challenging due to theshort life time of social events. An event is non-existent until its creation time t_C. And after the start time t_S of an event, participation recommendation becomes meaningless. Due to the very limited history of an event from time t_C to t_S, event participation recommendation suffers from the cold-start problem heavily.

Now, let's formally define the event participation problemas follows: given an event e, at time t ($t_C < t < t_S$), the task is to predict users who will RSVP "yes" to event e between$t$ and t_S. The EBASN built upon the collective data before t will serve as the network structure and all the users who responded "yes" to e between tc and t are the positive training examples for the prediction.

VII. INFORMATION FLOW EVALUATION

(a) Experimental Settings

When an event is created, except for the creator, it is unknown to all the other users. To simplify the problem, we treat the event creator as the first user who responded "yes" to the event. In evaluation, we can start the recommendation process immediately after the event creation, or wait for a while until there are a few responded users. We first focus on the latter case: given a testing event, we set the first k responded participants as the seed users, where k is randomly determined. The former case is a much harder problem and is examined at the endof the evaluation.

We split the Online data into two sequential parts The first part of data (on or before Mar 2011, take up 60%) are used for training and the second partof data (after Mar 2012, take up 40%) are used for testing. Given a testing event, we recommend top 5, 10, 20, 50, 100,200, 400, 800 users to it respectively. We choose to recommend a large number of users, because 1) in practice eventorganizers often broadly advertise their events to the public;and 2) we want to see the long-term trend of such a recommendation system. For the recommended top N users, we compute *recall* to evaluate the performance. *recall* is defined as the percentage of users who would respond "yes" to thetesting event that are covered by the top N recommendations. Finally, we average the *recall* for all testing events under the same top N.

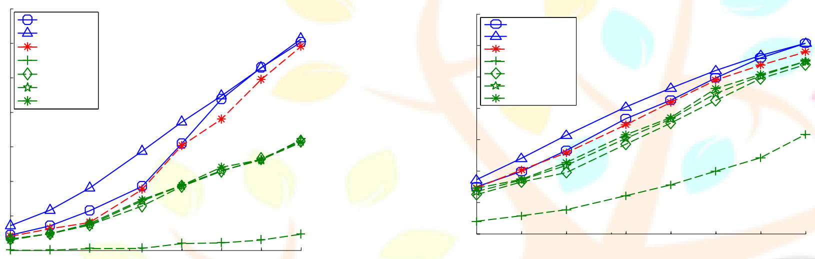
(b) Compare Event-Centric Diffusion Models withClassic Baselines

There are two popular baselines found in the prior art that can be efficiently applied to such an event participation recommendation problem. One is Collaborative Filtering (CF) and the other is the random walk model. Note that due to the extremely short life time of events, mostsupervised recommendation (link prediction) methods suffer from severe sparsity of labeled data. As a result, they do notapply to the event participation recommendation problem.

For the baseline CF, the users who ever participated in similar groups or events in the Meetup training data are recommendation candidates. They are then ranked by their Jaccard similarities to the responded users. The Jaccard similarity between two users is simply based on their past group or event participation count vectors.

For the baseline random walk model, we applied the *random walk with restart* (RWR) model. In the RWR baseline, there is a certain chance (probability β) with which the in formation will flow back to the starting users at each step of information flow. By setting various β, we have various RWR baselines with names like RWR (0.3). When $\beta = 0$, RWR downgrades to the basic random walk model.

As both CF and RWR were initially designed for homgeneous networks, we compared them with the basic event-centric diffusion models on individual G^{on} and G^{off} in Fig. 9. From all diffusion models on G^{on} in Fig. 9(a) and G^{off} inFig. 9(b), *DIF-com* outperforms *DIF* and *CF*, and RWR models perform the worst. By soft-restricting information flow in the same user communities, *DIF-com* can guarantee most closely related friends are recommended. The weight-ing strategies of *DIF* and *CF* differ only slightly, thus they yield similar prediction results. The poor performance of *RWR* indicates that identified network hubs are not rele- vant to the testing event. By raising return probabilities of RWR, the prediction performance does not improve much even with β as high as 0.6. In addition, by comparing Fig. 9(a) and Fig. 9(b), we find the offline EBSN has better prediction power when N is small but online EBSN gradu- ally catches up and even surpasses the offline EBSN as N grows large. This is because offline social interactions are able to capture closely related friends who are very likely toparticipate in the same events, but the recommended userstend to be regulars to similar events. In comparison, online social interaction can introduce non-regulars to the events and increase the coverage of the recommendation.



VIII CONCLUSION

In this paper, we have identified and formally defined a new type of social network, Event based social network and we alanysed the performance of online event based interactions and offline event based interactions By using the one dataset, we studied the unique features of EBASNs including basic network properties, community structures and information flow over EBASNs.

Acknowledgement

[11] B. Gao, T.-Y. Liu, X. Zheng, Q.-S. Cheng, and W.-Y. Ma. Consistent bipartite graph co-partitioning for starstructured high-order heterogeneous data co-clustering.In *KDD*, 2005.

[12] G. Golub and C. Loan. *Matrix Computations*. JohnsHopkins Univ. Press, 1996.

[13] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall Prentice-Hall advanced reference series,1988.

[14] S. Lattanzi and D. Sivakumar. Affiliation networks. In *STOC*, 2009.

[15] R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla. New perspectives and methods in link prediction. In *KDD*, 2010.

[16] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of onlinesocial networks. In *SIGCOMM*, 2007.

[17] M. Newman. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality". *Physical ReviewE*, 2001.

[18] M. E. J. Newman, D. J. Watts, and S. H. Strogatz.Random graph models of social networks. In *NationalAcademy of Sciences*, 2002.

[19] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil. An empirical study of geographic user activity patterns infoursquare. In *ICWSM*, 2011.

[20] J. F. Padgett and C. K. Ansell. Robust Action and the Riseof the Medici, 1400-1434. *The American Journal of Sociology*, 1993.

[21] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. 1999.

[22] T. Sander and S. Seminar. E-associations? using technology to connect citizens: The case of meetup.com. In*Annual Meeting of the American Political Science Association*, 2005.