A Survey on Online Social Network Methodologies

Manali Patel
Department of Computer Science & Engineering
Parul Institute Technology, Vadodara, India

Prof. Ankur Shah
Department of Computer Science & Engineering
Parul Institute Technology, Vadodara, India

Abstract

Online social networks (OSNs) are an important source of information for scientists in different fields such as computer science, sociology, economics, etc. However, it is hard to study OSNs as they are very large. For these reasons, we argue that sampling techniques will be the best technique to study OSNs in the future. For this reason here we combine some papers and present some Online social network methodologies that are used for easily finding sampling techniques.
Keywords: Online Social Networks(OSN), Social links, Private matching protocol, Social networks, Lifestyle

---

I. Introduction

From the point of view of data mining, a social network is a heterogeneous and multi-relational data set represented by a graph. The graph is typically very large, with nodes corresponding to objects and edges corresponding to links representing relationships or interactions between objects. Both nodes and links have attributes. Objects may have class labels. Links can be one-directional and are not required to be binary. Social networks need not be social in context. In sampling there are so many methods.[5]

Twenty years ago, people typically made friends with who live or work close to themselves, such as neighbors or colleagues. We call friends made through this traditional fashion as G-friends, which stands for geographical location-based friends because they are influenced by the geographical distances between each other. With the rapid advances in social networks, services such as Facebook, Twitter and Google+ have provided us revolutionary ways of making friends.[1]Many of previous works on link prediction treat all users in the network equally and focus on predicting potential links that will appear among all the users, based upon a snapshot of the social network. However, in real-world social networks, many new users are joining in the service every day. Predicting social links for new users are more important than for those existing active users in the network as it will leave the first impression on the new users.[4]

In online social network there is given the large adoption of these networks, there has been increased interest to explore the underlying social structure and information in order to improve on information retrieval tasks of social peers. Such tasks are in the core of many application domains. Sampling methods are proposed to approximate community structures in a social network. [5]
In this paper we finalize you that which Online social network method is used for which technique. For example, we show you that for sampling online social network which method is used and so on.
The rest of the paper is organized as follow. We discuss the Online social network methodologies in Section 2 and conclude the paper in Section 3.

II. Online social network methodologies

A. Sampling Methodologies
The social network structure can be modeled as a graph G with individuals representing nodes and relationships among them representing edges. We model environments in which social peers participate in a centralized social network or distributed. Centralized graphs are typical in social networking sites in which complete knowledge of users' network is maintained (e.g., delicious, Flickr, etc.). Distributed graphs, where a user is aware only of its immediate connections, are more common. We discuss this issue under assumption of static and

B. Dynamic Network Topologies
1) Assuming Static Networks
We first consider the case where the topology of the social network is static, or changes only slowly over time. For this case, a straightforward solution exists where each node, in a pre-computation phase, performs a complete crawl of its neighborhood Dd(v) and selects a uniform random sample S of n nodes, whose addresses (or access paths) are then stored at the initiating node. However, this phase needs to be recomputed infrequently; once the social
network topology has undergone significant changes.
2) Assuming Dynamic Networks
We next consider the case where the topology of the network is dynamic, i.e., where the network structure changes frequently in addition to the data changes at each node. In such a case, it makes little sense to precompute samples of Dd(v) as such samples go

stale very quickly. Thus, the task of sampling from Dd(v) has to be deferred to runtime. This problem is challenging because we cannot crawl the entire neighborhood DdðvÞ at runtime (this will be prohibitively slow). It becomes even more challenging by the fact that we are constrained to simulate random walks by only following edges of the social network. So for that we have to create random samples that is brifly given in[1].

However, we can adopt our solution of sampling from trees to this specific scenario by ensuring that the union of all random walks made in collecting a sample always resembles a tree.Here in dynamic network we have to improve the performance of information collection from the neighborhood of a user so firstly we have to find SampleDyn, an algorithm that is able to compute a near-uniform sample of users in dynamic social networks. Then, we introduce two algorithms, EvalSinle and EvalBatch, that utilize SampleDyn in order to estimate counts for a set of items in user's vicinity.

The brief explanation of algorithm is in[1].

3) Algorithm 1

Sampling in Dynamic Social Networks

```
procedure SAMPLEDYN (u, n, d, C)
T = NULL, samples = 0, Sample array of size n
while samples <= n do
if (v = randomWalk (u, d, C, T))! = 0 then
Sample[samples ++] = v
end if
end while
end procedure
procedure RANDOMWALK(u, d, C, T)
depth =  0, ps = 1
while depth < d do
pick v € children(u) U u with pv = 1/(degree(u)+1)
if T U v has no cycle then
add v to T
ps = ps . pv
if v =  u then
accept with probability C/ps
if accepted then
return v
else
return 0
end if
else
u =  v, depth++
end if
end if
end while
return 0
end procedure
```

4) Algorithm 2

Counts Estimation—Separate Samples

```
procedure EVALSINGLE (v, d, C,  n, X)
S array of size n
Count array of size jXj
for all x € X do
S = SampleDyn(v, n, d, C)
for all i € S do
Count[x] = Count[x] + countix
end for
end for
return Count
end procedure
```

5) Algorithm 3

Counts Estimation—Same Sample

```
procedure EVALBATCH(v,  d, C,  n, X)
S array of size n
Count array of size jXj
for all x € X do
```

5) S = SampleDyn (v, n, d, C)
6) for all i € S do
7) Count[x] = Count[x] + countix
8) end for
9) end for
10) return Count
11) end procedure

And then by using random walks we compare these two methods as per sample accuracy and sample cost. We finalize that EvalSingle is best for sample cost and EvalBatch for accuracy. So in future we combine these two methods and generate new one that is best for both.

C. Asymmetric Proximity for Online Social Networks

The explosive growth of Online Social Networks (OSNs) over the past few years has redefined the way people interact with existing friends and especially make new friends. Some works propose to let people become friends if they have similar profile attributes. However, profile matching involves an inherent privacy risk of exposing private profile information to strangers in the cyberspace. The existing solutions to the problem attempt to protect users' privacy by privately computing the private set intersection or private set intersection cardinality of the profile attribute sets of two users. So here we design three private matching protocols, which provide different privacy levels and can protect users' privacy better than the previous works. And the asymmetric social proximity based private matching protocols are

1) Protocol for Level 1 Privacy (L1P)

The protocol ensuring level1 privacy is suitable for users who decide to make friends with each other simply based on the common communities of their overall community sets. In this protocol, we first let the Responder learn the mutual communities and the size of the Initiator's input set (CI) (i.e., the Initiator's overall community set), while let the Initiator learn nothing but the size of the Responder's input set (CR) (i.e., the Responder's overall community set). Then, the Responder securely sends the common communities to the Initiator, if she confirms the request from the Initiator. The whole descriptions of this protocol is given in[2].

2) Protocol for Level 2 Privacy (L2P)

In the protocol for level 1 privacy (L1P), the Responder determines whether or not to accept the Initiator's request for a social friendship only based on their common overall communities, which may not characterize the social proximity well.

L2P, utilizing the proposed community based asymmetric social proximity measurement. This protocol is suitable for the case when the Initiator is willing to establish a friendship relation with the Responder but the Responder accepts the relationship only if her requirement on the friendship is fulfilled. In particular, in L2P, the Responder accepts the friendship request from the Initiator if the social proximity measured by her, i.e, $\Psi_{R\leftarrow I}$, is greater than a threshold predefined by herself, denoted by Ψ_{R_τ}. The whole description of this protocol is given in[2].

3) Protocol for Level 3 Privacy

In the L2P protocol, the Responder determines whether or not to be friends with the Initiator based on the community based social proximity, while the Initiator still can only make his final decision based on their common communities. Besides, in terms of privacy, in L2P the Responder will know CI ∩ CR If $\Psi_{R\leftarrow I} > \Psi_{R_\tau}$, no matter whether the social proximity measured by the Initiator is large enough or not.

L3P is suitable for users with very high privacy requirements. In this protocol, both the Initiator and the Responder make sure their requirements on friendship are fulfilled before revealing any matching information to each other. If either of the requirements is not satisfied, neither of them knows the matching profile information, i.e., the common communities CI ∩ CR. The whole descriptions of this protocol is given in[2].

D. A Sementic-based friend Recommendation System for Online social networks

A novel semantic-based friend recommendation system for social networks, which recommends friends to users based on their life styles instead of social graphs. By taking advantage of sensor-rich smartphones, This discovers life styles of users from user-centric sensor data, measures the similarity of life styles between users, and recommends friends to users if their life styles have high similarity. Inspired by text mining, we model a user's daily life as life documents, from which his/her life styles are extracted by using the Latent Dirichlet Allocation algorithm. This algorithm is directly exapted. We further propose a similarity metric to measure the similarity of life styles between users, and calculate users' impact in terms of life styles with a friend-matching graph. Upon receiving a request, Friendbook returns a list of people with highest recommendation scores to the query user. Here we called this system Friendbook.

E. Predicting Social links

Nowadays, many new users are keeping joining in the online social networks every day and these new users usually have very few social connections and very sparse auxiliary information in the network. Prediction social links for new users is very important. Different from conventional link prediction problems, link prediction for new users is more challenging due to the lack of information from the new users in the network.

Meanwhile, in recent years, users are usually involved in multiple social networks simultaneously to enjoy the specific services offered by different social networks. The shared users of multiple networks can act as the “anchors” aligned the networks they participate in. Here we propose a link prediction method called SCAN-PS (Supervised Cross Aligned Networks link prediction with Personalized Sampling), to solve the social link prediction problem for new users. SCAN-PS can use information transferred from both the existing active users in the target network and other source networks through aligned accounts.
In addition, SCAN-PS could solve the cold start problem when information of these new users is total absent in the target network. Extensive experiments conducted on two real-world aligned heterogeneous social networks demonstrate that SCAN-PS can perform well in predicting social links for new users.

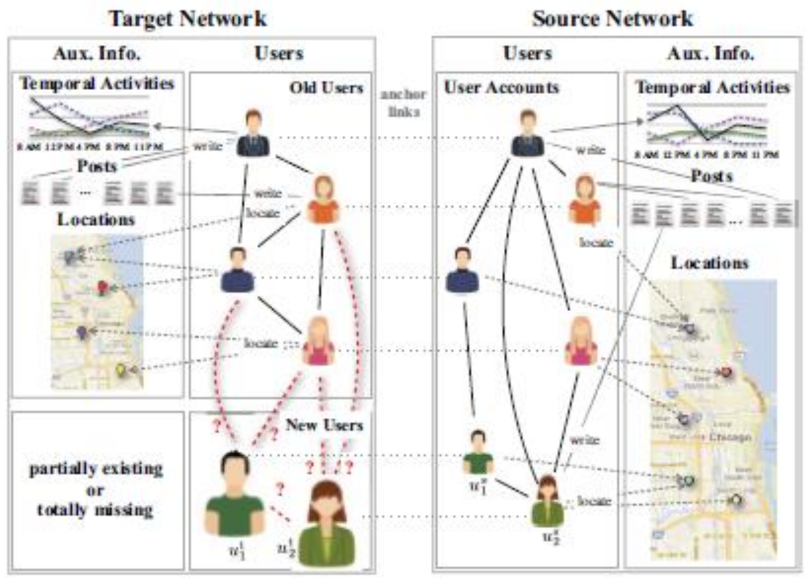

Fig. 1: Example of predicting social links across two aligned heterogeneous online social networks

By sampling the old users’ sub-network, we want to meet the following objectives:
- Maximizing Relevance: We aim at maximizing the relevance of the old users’ sub-network and the new users’ sub-network to accommodate differences in information distributions of new users and old users in Gt.
- Information Diversity: Diversity of old users’ information after sampling is still of great significance and should be preserved.
- Structure Maintenance: Some old users possessing sparse social links should have higher probability to survive after sampling to maintain their links so as to maintain the network structure.

F. Cold-Start Link Prediction across Aligned Networks
In our problem settings, we have two aligned social networks. Link prediction methods porposed based on one single network can suffer from the cold start problem a lot. Here, we will propose two methods to utilize the aligned source network to help solve the problem.

1) NAIVE:

NAIVE just use the pseudo labels as the final prediction results of links in the target network. NAIVE is very simple and could work well in our task even when these new users are brand new, which means that we could overcome the cold start problem by using this method. However, it may still suffer from some disadvantages: (1) the social structures of different networks are not always identical which will degrade the performance of NAIVE a lot; (2) NAIVE only utilizes these new users’ social linkage information in the source network but ignores all other information.

2) SCAN-PS:

To overcome all these disadvantages, a new method SCAN-PS (Supervised Cross Aligned Networks Link Prediction with Personalized Sampling) is proposed. SCANPS could use heterogeneous information existing in both the target network and the aligned source and it is built across two aligned social networks. By taking advantage of the anchor links, we could locate the users’ aligned accounts and their information in the aligned source network exactly. If two aligned networks are used simultaneously, different categories of features are extracted from aligned networks. A more detailed description about the extracted features is available in[8].

To use the information in multiple networks, feature vectors extracted for the corresponding links in aligned networks are merged into an expanded feature vector. The expanded feature vector together with the labels from the target network are used to build a cross-network classifier to decide the existence of social links related to these new users in the target network.

This is how method SCAN-PS works. SCANPS is quite stable and could overcome the cold start problem for the reason that the information about all these users in the aligned source network doesn’t change much with the variation of the target network and we get the information showing of these new users ‘preferences from the information he/she leaves in the aligned source network.

-SCAN-PS to solve this problem by using information in multiple aligned heterogeneous social networks. A within-network personalized sampling method is proposed to address the differences in information distributions of new users and old users. Information from the aligned source network and that owned by the old users in the target network is transferred to help improve the prediction result. Extensive experiments results show that SCAN-PS works well for users of different degrees of novelty and can also solve the cold start problem.

III. Conclusion

In this paper we study online social network methodologies. By using this methodology we can easily understand social network and its techniques. These methodologies are helpful to online social networks.

References

[1] Manos Papagelis,Gautam Das,Nick Koudas "Sampling Online Social Networks" VOL. 25, NO. 3, MARCH 2013
[2] Arun Thapa,Ming Li,Sergio Salinas and Pan Li "Asymmetric Social Proximity Based Private Matching Protocols for Online Social Networks" 2013
[3] Zhibo Wang, Jilong Liao, Qing Cao, Hairong Qi and Zhi Wang "Friendbook: A Semantic-based Friend Recommendation System for Social Networks" 2013
[4] Jiawei Zhang,Xiangnan Kong, Philips S. Yu "Predicting Social Links for New Users across Aligned Heterogeneous Social Networks" 2013
[5] Book "Data mining : Concepts and techniques By Jiawei Han and Micheline Kamber"
[6] Kenton White, Guichong Li, Nathalie Japkowicz "Sampling Online Social Networks Using Coupling From The Past" 2012.
[7] Facebookstatistics.http://www.digitalbuzzblog.com/facebook-statistics-stats-facts-2011/.
[8] J. Zhang, X. Kong, and P. Yu. Predicting social links for new users across aligned heterogeneous social networks. CORR arXiv:1310.3492,2013.