An Improved Performance for Dynamic Online Social Network

1 Manali Patel,2 Ankur Shah

1 Manali Patel, Student,Parul Institute of Technology,Vadodara,Gujarat, India

2 Ankur Shah, Assistant Professor, Parul University, Vadodara, Gujarat, India

Abstract

Online social networks (OSNs) are an important 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. Our research aims to offer performance improvements, via sampling, to the process of (almost) uniformly collecting information from user logs by exploring the underlying graph structure of a social network. Sampling methods are proposed to approximate community structures in a social network. Our work is also related to work on personalized and social search. The premise of personalized search is that by tailoring search to the individual improved result accuracy may be brought off. A vast amount of literature on search personalization reveals significant improvement over traditional web search. Our contribution is to evaluate the behavior of these techniques on a real directed graph by considering sampling scenarios.

Key Words: *Sampling Online social network, social network, sampling methods*

---

1. Introduction:

Data mining refers to extracts or mining knowledge from large amounts data. Because of extensive accessibility of data, it is essential to convert this data into information and knowledge. This helpful information and knowledge can be used in applications similar to Business management, market analysis, Risk prediction etc. Data mining can be viewed as a outcome of the natural evolution of information technology.

Sampling is the process of selecting units from a population so by studying the sample we generalize the result in population from which they were chosen. Data sampling is analysis of statical technique that is used to select, manipulate and analyze subset of data points in order to identify patterns and trends in the larger data set being examined. There are several sampling methods that are listed below: Sampling methods are divided as probability and nonprobability.

The task of social network analysis has been limited. Sampling methods are used for estimating OSN properties. Sampling can allow exact inference that are relative small number of observations. The need of a sampling frame for OSNs makes sampling difficult. Because of this, recent work in this area has focused on sampling methods. 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.

New technologies are stayed in market because of its impact. That is by changing social environment and by sharing knowledge they get new idea about social networks. These social media technologies are more often now a day's used. So they also have competition so they share new ideas and gain impact. Today many organizations are devoted on in social media because of lack of proper understanding of what social media is, meaning of it and the benefit. The use of social media in organizations is also relevant as an increasing number of organizations. Social media gives an opportunity to study utilization using real time example.

In this paper, in section 2 there is related work, in section 3 there is literature review, in section 4 there is proposed work, in section 5 there is conclusion.

2. Related Work:

Previous work focuses on improving the performance of information collection from the neighborhood of a user in a dynamic social network. they introduce sampling-based algorithms to efficiently explore a user's social network respecting its structure and to quickly approximate quantities of interest.

They introduce and analyze variants of the basic sampling scheme exploring correlations across our samples. Models of centralized and distributed social networks are considered. They show that our algorithms can be utilized to rank items in the neighborhood of a user, assuming that information for each user in the network is available.

Social search or a social search engine is a type of search method that tries to determine the relevance of search results by considering interactions or contributions of users. The premise is that by collecting and analyzing information from a user's explicit or implicit social network they can improve the accuracy of search results. The most common social search scenario is the following:

1. A user v in a network submits a query to a search engine.
2. The search engine computes an ordered list L of the most relevant results using a global ranking algorithm.
3. The search engine collects information that lies in the neighborhood of v and relates to the results in L.
4. The search engine utilizes this information to reorder the list L to a new list L0 that is presented to v.[6]

They offers performance improvements, via sampling, to the process of (almost) uniformly collecting information from user logs by exploring the underlying graph structure of a social network. In that sampling theorem is used for obtain random uniform sample. But here main problem is that how to obtain the uniform or near-uniform random sample, for that issue they discuss the issue under assumption of static and dynamic network topologies.

By that they create the neighbourhood samples that are shown in figure. And here in this paper there is three algorithms are used, sampling dynamic social networks, Count estimation for separate samples and count estimation for same sample.

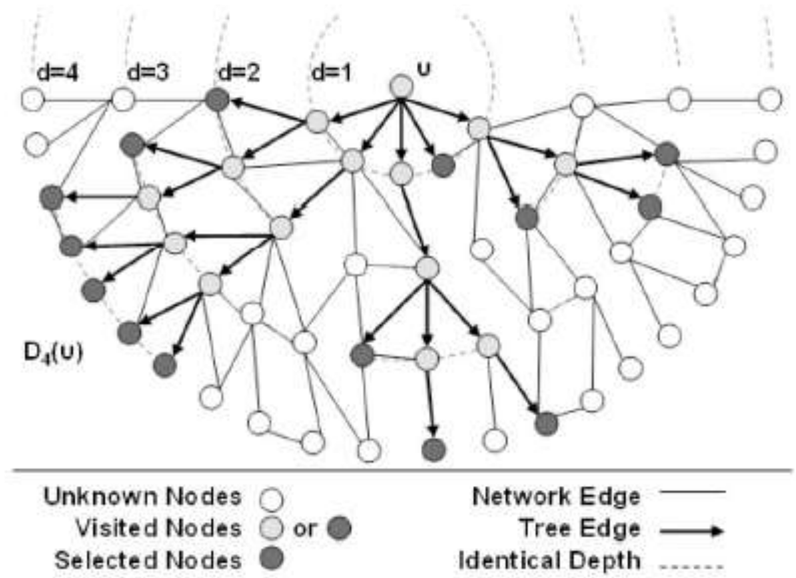

Fig 1: Random walks guided by SampleDyn

Here naïve method and Evalsingle sampling method is used for sampling accuracy and sampling cost. And here Evalsingle is succeeds in selecting samples in near-uniform way.

EvalSingle performs better in terms of accuracy than a naive sampling method, but as many of the performed random walks end up rejecting a selected leaf node, it can be expensive.

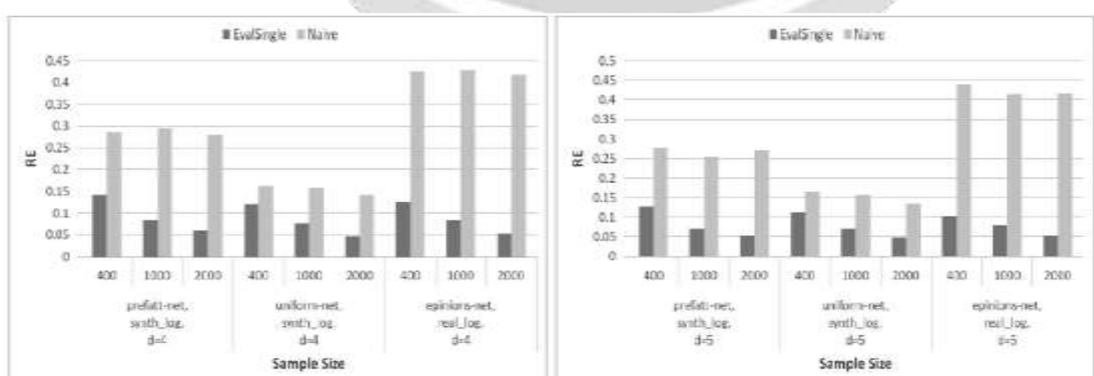

Fig 2: Sampling Accuracy: EvalSingle versus Naïve

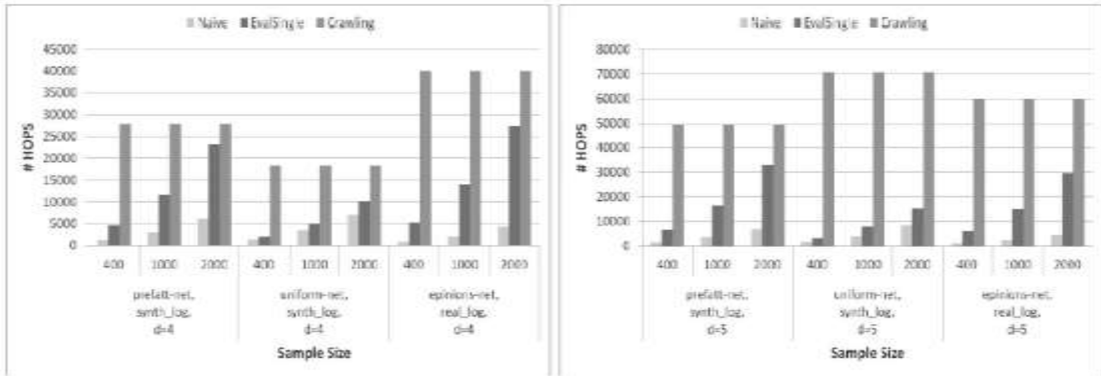

Fig 3: Sampling Cost: EvalSingle versus Naïve

2.1 Limitations of the this system

In the existing system performance of information collection from the neighbourhood of a user in a dynamic social network is not possible. Models of centralized and distributed social networks are not considered. The algorithms cannot be utilized to rank items in the neighborhood of a user, assuming that information for each user in the network is not available.

2.2 Problems in this system:

1. Using real and synthetic data sets, are not possible.
2. We cannot validate the results and the efficiency of our algorithms in approximating quantities of interest.
3. It is not possible to collect information from a social graph in an efficient manner.
4. The premise of collecting and analyzing information from a user's explicit or implicit social network we cannot have to improve the accuracy of search results.

3. Literature review:

3.1 Asymmetric Social Proximity Based Private Matching Protocols 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.

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. These schemes have some limitations and can still reveal users' privacy.

By using community structures to redefine the OSN model and propose a realistic asymmetric social proximity measure between two users. Then they design three private matching protocols, which provide different privacy levels and can protect users' privacy better than the previous works.

3.2 Friendbook: A Semantic-based Friend Recommendation System for Social Networks:

Friendbook, 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, Friendbook 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.

For recommend friends they use: 1) habits or life style; 2) attitudes; 3) tastes; 4) moral standards; 5) economic level; and 6) people they already know.

3.3 Predicting Social Links for New Users across Aligned Heterogeneous Social Networks:

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. 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.

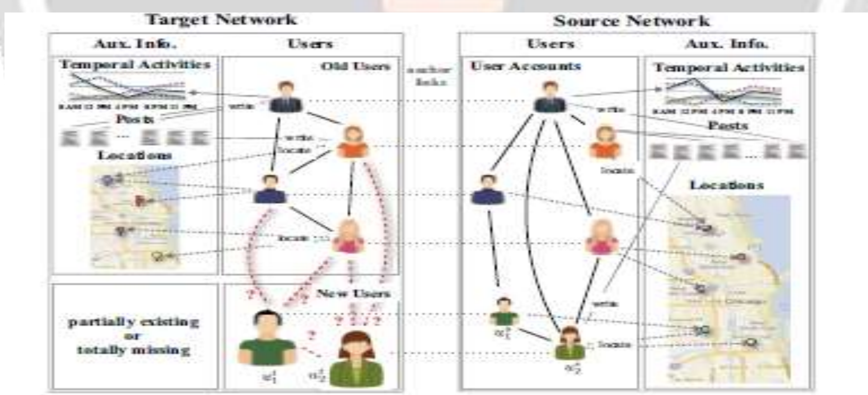

Fig. : Example of predicting social links across

two aligned heterogeneous online social networks

To solve these problems, they propose a novel supervised cross aligned networks link recommendation method, SCAN-PS. Twitter and Foursquare two networks are taken.

3.4 Sampling Online Social Networks via Heterogeneous Statistics:

The most sampling techniques for online social networks (OSNs) are based on a particular sampling method on a single graph, which is referred to as a statistic. However, various realizing methods on different graphs could possibly be used in the same OSN, and they may lead to different sampling efficiencies, i.e., asymptotic variances.

To utilize multiple statistics for accurate measurements, they formulate a mixture sampling problem, through which we construct a mixture unbiased estimator which minimizes the asymptotic variance. Given fixed sampling budgets for different statistics, they derive the optimal weights to combine the individual estimators; given a fixed total budget, they show that a greedy allocation towards the most efficient statistic is optimal.

They show that their two-stage framework is a generalization of 1) randomly choosing a statistic and 2) evenly allocating the total budget among all available statistics, and our adaptive algorithm achieves higher efficiency than these benchmark strategies in theory and experiment.

3.5 Sampling Online Social Networks Using Coupling from The Past

Recent research has focused on sampling online social networks (OSNs) using traditional Markov Chain Monte Carlo (MCMC) techniques such as the Metropolis-Hastings algorithm (MH). While these methods have exhibited some success, the techniques suffer from slow mixing rates by themselves, and the resulting sample is usually approximate.

An appealing solution is to apply the state of the art MCMC technique, Coupling From The Past (CFTP), for perfect sampling of OSNs. In this initial research, they explore theoretical and methodological issues such as customizing the update function and generating small sets of non-trivial states to adapt CFTP for sampling OSNs. Their research proposes the possibility of achieving perfect samples from large and complex OSNs using CFTP.

A number of existing techniques for crawling include Breadth First Search (BFS) and Random Walk (RW). While such techniques usually yield a bias toward the most highly connected nodes.

MH algorithm can be naturally overcome by using the state of the art Coupling From The Past (CFTP). In CFTP convergence is achieved by coalescence to a single state, which turns out to be a perfect sample from the stationary distribution.

4. Proposed Work:

4.1 Key Idea of CNRW:

In traditional random walks, the transition at each node is memoryless - i.e., when the random walk arrives at a node, no matter where the walk comes from (i.e., what the incoming edge is) or which nodes the walk has visited, the outgoing edge is always chosen uniformly at random from all edges attached to the node.

The key idea of CNRW is to replace such a memoryless transition to a stateful process. Specifically, given the previous transition of the random walk u → v, instead of selecting the next node to visit by sampling with replacement from N(v), i.e., the neighbours of v, we perform such sampling without replacement.

4.2 Flow of Proposed Work:

A change is demonstrated through an example in Figure 1:

1. When a CNRW transits through u → v for the first time, it selects the next node to visit in the same way as traditional random walk i.e., by choosing w uniformly at random from N(v).
2. Nonetheless, if the random walk transits through u → v again in the future, instead of selecting the next node from N(v), we limit the choice to be from N(v) – {W}.
3. One can see that, given a transition u → v, our selection of the next node to visit is a process of sampling without replacement from N(v). This, of course, continues until ∀w € N(v), the random walk has passed through u → v → w, at which time we reset memory and restart the process of sampling without replacement. Also, we introduce a notation b(u; v), which is defined as a set of nodes in N(v) that we have passed through.

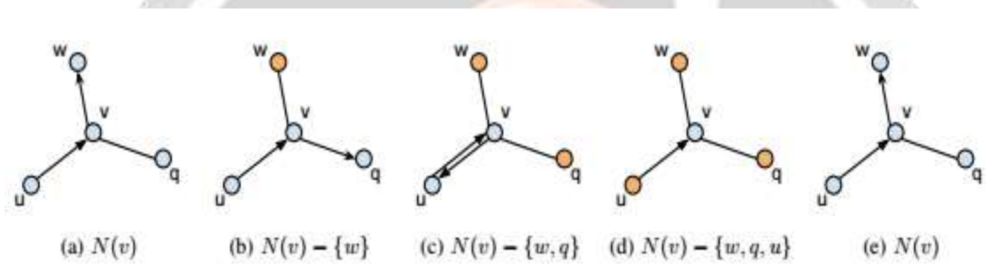

Fig: Chooses the next candidate from the set N(v) – b(u,v)

Thus, generalized idea of CNRW is:

1. Each time when the random walk travels from u to v, we uniformly choose the next candidate node w from N(v) → b(u; v).

2. Let b(u, v) ← b(u, v) ∪{w}. If b(u, v) = N(v), let b(u; v) = ∅.

In social network there is information in social network, login and registration, sending and receiving friend request and displaying friend list. By using that there is user can search detail and review the comment.

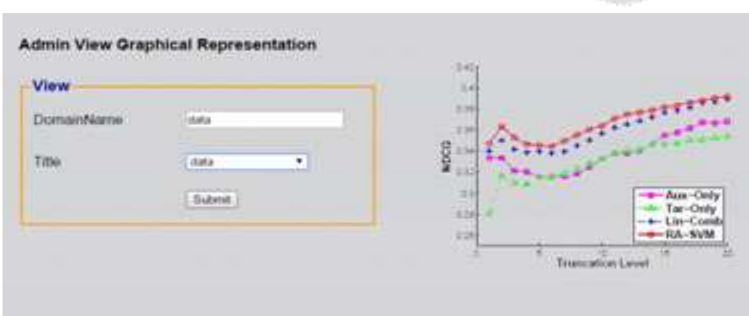

Fig. :Graphical representation

Here there is no graph because it has 0 views for graphical representation we need at least 7 views.

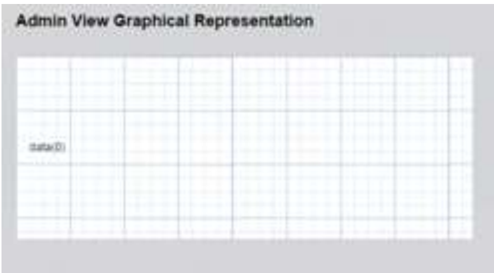

Fig. : Views with Graphical Representation

Fig. : Domain insertion

Fig. :User Login

Fig. : Different Data types User has Searched

5.Conclusion:

Our research suggests methods for quickly collecting information from the neighbourhood of a user in a dynamic social network when knowledge of its structure is limited or not available. Our methods resort to efficient approximation algorithms based on sampling. By sampling we avoid visiting all nodes in the vicinity of a user and thus attain improved performance.

Our algorithms assume that information for each user in a network, such as web history logs, is available to access personal information infringes on user privacy and, as such, privacy concerns could serve as a major stumbling block toward acceptance of our algorithms..

References:

1. Manos Papagelis,Gautam Das,Nick Koudas "Sampling Online Social Networks" VOL. 25, NO. 3, MARCH 2013

2. Kenton White Girih, Guichong Li, Nathalie Japkowicz "Sampling Online Social Networks Using Coupling From The Past" 2012

3. Arun Thapa,Ming Li,Sergio Salinas and Pan Li "Asymmetric Social Proximity Based Private Matching Protocols for Online Social Networks" 2013

4. Zhibo Wang, Jilong Liao, Qing Cao, Hairong Qi and Zhi Wang "Friendbook: A Semantic-based Friend Recommendation System for Social Networks" 2013

5. Jiawei Zhang,Xiangnan Kong, Philips S. Yu "Predicting Social Links for New Users across Aligned Heterogeneous Social Networks" 2013

6. Kenton White, Guichong Li, Nathalie Japkowicz "Sampling Online Social Networks Using Coupling From The Past" 2012

7. www.wikipedia.com/sampling methods

8. https://en.wikipedia.org/wiki/Data_mining

9. http://arxiv.org/abs/1505.00079

10. Book "Data mining : Concepts and techniques By Jiawei Han and Micheline Kamber"