INFORMATION RETRIVAL FROM SOCIAL NETWORKING

Twinkle Patel, Hiten Sadani
Computer Engineering Department
Silver Oak College of Engineering and Technology, Ahmedabad

***Abstract*- Large quantities of information are shared through online social networks, making them attractive sources of data for social network research. When studying the usage of online social networks, these data may not describe properly users' behaviours of online social networks. For instance, the data collected often include content shared by the users only, or content accessible to the researchers, hence obfuscating a large amount of data that would help understanding users' behaviours. there has been enlarged significance to uses the underlying network structure at the same time the available information on social peers to improve the data requires of a client. In this we focus on improve the performance of information collection from the neighbourhood of a user in a dynamic social network. We introduce and analyze variants of the basic sampling scheme exploring correlations across our samples. Models of centralized and distributed social networks are considered.**

***Index Terms*- Information retrival ,Social Network**

INTRODUCTION

The Internet and the World Wide Web (WWW) changed since the Internet became popular. Those early days our understanding and usage of the WWW significantly differs to that of these days. Those days a new special media, which was only available to scientists and computer experts. A recent large scale study showed a how the online world is approached by people [1]. Today, internet has become part of our everyday life. Internet is used routinely to reach other people through various applications ranging from email, chats, community platforms and social network sites. Therefore it has become an extension of our self in a way that allows us to construct and extend our social ties and form social networks within this new medium. Now, this is ongoing trend that, people are increasingly well connected on the internet through some social network sites. Social networking is virtual communication that allows people to connect each other[2]. It provide various ways for users to interact such as chat, video sharing, blogging, discussion group etc. Social networking sites such as MySpace , Facebook and LinkedIn are examples of wildly popular networks used to find and organize contacts. Other social networks such as Flickr, YouTube, and Google Video, are used to share multimedia content, and others such as Live Journal and BlogSpot are used to share blogs. In this user can create a virtual profile and connect old friends, family member, etc. Recently people increasingly update very personal information on social network sites[3]. This information more and more available from social network sites and social relationships between people can be identified. This publicly available information is extract from social network sites. Table shows history of social networking sites and features.

Year	Social Networking Sites	Description
1997	SixDegrees.com	First Network, that allowed public profile and friend list
1999	LiveJournal.com	Blogging system and blog hoster with addition buddy list functions
2000	Xanga.com	LiveJournal clone
2000	SkyBlog.com	LiveJournal clone
2001	CyWorld.com	Korean social network
2001	Ryze.com	First business oriented social network
2002	Friendster.com	Social complement to ryze.com
2003	LinkedIn.com	Business oriented social network
2003	Tribe.net	Designed for classifieds but adopted by the Burning Man community
2003	MySpace.com	To compete with friendster
2004	Orkut.com	Google's social network to compete with myspace.com, became very popular in Brazil
2004	Mixi.jp	Japan
2004	Lunarstorm.se	Sweden
2004	Grono.net	Poland
2004	Hi5.com	Smaller south American countries
2004	Xing.com (open)	Germany/Europe business oriented social network
2004	Bebo.com	United Kingdom, New Zealand, and Australia
2004	Facebook.com	Social network for students
2005	studivz.net	German social network for students
2006	Twitter	Micro blogging site with social network features

Table : History of Social Networking Sites

INFORMATION RETRIEVAL SYSTEMS DEFINITION

The origins of information retrieval can be traced way back during World War II as
massive quantities of documentation and reports about weaponry were produced[4].
At this time, representation of documents is difficult task. The growth of the Internet and the WWW has produced huge volumes of available information. These information are just a few clicks away, but the enormous amount of information, an ever growing demand for tools that assist users in the process of locating important information for their information needs. Systems that provide these tools are usually known as search engines.

Definition for *IR systems* is provided by Belkin(1984):
"The goal of an information [retrieval] system is for the user to obtain information from the knowledge resource which helps her/him in problem management".

Definition for *IR systems* by (*Kowalski, 1997*)
"An Information Retrieval System is a system that is capable of storage, retrieval, and maintenance of information. Information in this context can be composed of text (including numeric and date data), images, audio, video and other multi-media objects."

INFORMATION SOURCES FOR INFORMATION RETRIEVE FROM SOCIAL NETWORKS

Traditionally SNA is conducted by collecting data via questionnaires, interviews, observations, archival records, experiments, which is manual process of gathering and analysing available information[5]. In contrast the World Wide Web and diverse databases provide an increasing number of information sources, which exhibit characteristics that are interesting to SNA. Wellman proposes very broadly that *"a computer network is a social network"* and argues that humans use technologies to create communities[6]. Within these communities supportive and sociable relationships are built, which form sustainable community ties. Social ties can be created and maintained via various different media types, including face to- face contact, meetings, telephone, writing and other means of communication. The nodes in communication networks are the same as nodes in traditional social networks and that the content of social relation are the communication exchanges or the information flow. Following this he distinguishes among various types of networks: the social network, the computer network, the internet network and the hyperlink network. World Wide Web is a hyperlink network and as such can be seen as a special kind of social network, it should be emphasized that both, links and texts, on web pages in the World Wide Web reflect social interactions of users in the real world. This notion allows the extraction of social network structures not only from the hyperlinks but also from available textual information on web pages in the World Wide Web [7].

PROBLEMS FOR INFORMATION RETRIEVE

The social network structure can be modeled as a graph G where nodes representing user and edges representing relationships among them[8]. Social graph model environments in which social peers participate in a centralized social network(where knowledge of the network structure is assumed) or distributed (where network structure is unknown). Centralized graphs are typical in social networking sites in which complete knowledge of users's network is maintained(e.g., del.icio.us, flickr, etc.). Distributed graphs, where a user is aware only of its immediate connections. for example, the case in ad hoc social networks formed by typical instant messaging or VoIP protocols (e.g., MSN and Skype). The rate of change of the structure of these networks is also an important factor. The most typical case is for such networks to change rapidly as users join and depart from the graph by forming or destroying social connections. We focus on dynamic networks (either centralized or distributed) but also treat the relatively easier case of static networks[8].

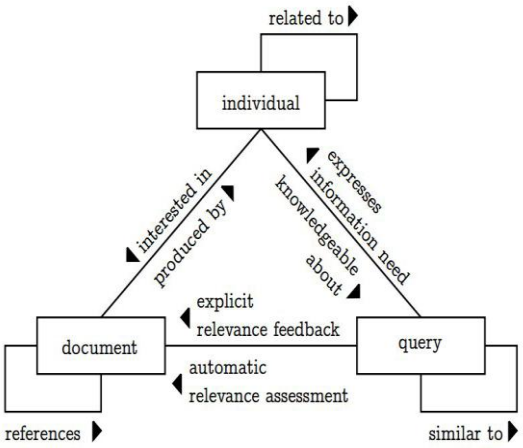
DOMAIN MODEL FOR SOCIAL IR

The traditional models for information retrieval concern themselves with documents, queries, and their relations to each other. A document is relevant to a query, a document references other documents, a

query is similar to other queries. In the social network analysis models individuals and their relations with each other[8]. In the figure Individuals appear in their role as information producers or information consumers, queries relate to an individual's information needs. In this way all are three modules related with each other. But information retrieve from social network steps are below

The most common steps for information retrieve from social network.

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 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.



A Domain Model For Social Information Retrieval

LITERATURE SURVEY

Manos Papagelis, Gautam Das in this paper[9], describe static and dynamic social network. Static network is change slowly over time. Information retrieve from this network is easy compare to dynamic social network. In this start from any specific node v and after complete crawl of neighborhood of v at specific depth d and select a uniform random sample S of n nodes. This process computed infrequently because social network topology has undergone significant changes. Now in dynamic social network nodes are add and links are add or delete from the network. So that computing sampling process is very hard. In this paper computing the sampling process using biased manner because some nodes are more likely to be destination of random walks than other nodes. After sampling process used ranking algorithm Estimating item count. Uses these algorithms all nodes are arrange in re-rank method in specific sample.

Sampling cost : Computing the sampling process, many nodes are reject from a selected node during random walk. So it can be expensive.

Sampling accuracy : sampling process pick a node in a biased manner because nodes are close to root node. Experiment so that a fixed network depth d the sampling accuracy increase with the sample size n. But in fixed sample size n the sampling accuracy decrease as depth d increase.

M.V. Mohan Babu PG Scholar in CSE and C.V. Chiranjeevi Kumar M.Tech, CSE[10] in this paper, Sampling large graph via random walks. For this start from any specific node v, and initiate random walk by proceeding to neighbors selected at every iteration. This process can be repeated to obtain random samples of a desire size. Hasting describes sampling based methods and techniques to efficiently collect information from users in a social graph. In this improve random walks methods on graph and need sample from the ranking algorithm. Performance improve via sampling, process of collecting information from user logs by exploring the graph structure in social network.

The CubeSVD approach was developed to improve Web search by taking into account click through data of the type "user, query, url.". for improve ranking quality use account such data and building statistical models for user behavior. Utilize above approach, Many ideas have been suggested to realize online social search; from search engines that utilize humans to filter the search results, to systems that utilize real-time temporal correlations of user web history logs, to tag-based social search systems. Analyses suggest that integration of social search models improves the overall search experience.

Mr. Yogesh P Murumkar and Prof. Yogesh B. Gurav[11] in this paper used algorithm is Distributed Outlier Detection. In network suppose node and each

node having a subset of objects in the whole database. First each node computes its set of local outlier by using centralized algorithm. After local outlier generated, all the nodes communicate to compute the global outliers. At the end of algorithm each node will have its subset of the actual global outlier. In this way outlier node consider for information retrieve using the Distributed Outlier Detection algorithm.
Sampling accuracy nd sampling cost is high compare to naïve sampling method.

CONCLUSION AND FUTURE WORK

we focused on improving the performance of information collection from the neighborhood of a user in a social network. Our research suggests methods for quickly collecting information from the neighborhood of a user in a dynamic social network. In all this use sampling method we avoid visiting all nodes in neighborhoods of a user and thus attain improved performance. we showed that our algorithms are able to efficiently estimate the ordering of a list of items that lie on nodes in a user's network providing support to ranking algorithms and strategies.
Further study is improve efficiency of information collection from neighborhood using re-rank strategies.

REFERENCE

[1] Jeffrey Boase University of Toronto, John B. Horrigan Associate Director Pew Internet Project, Barry Wellman University of Toronto "The Strength of Internet Ties" 2006

[2] Dr. Biswajit Das Associate Professor - Marketing Management & Chairperson Doctoral Programme School of Management, KIIT University, Jyoti Shankar Sahoo Masters in Business Administration School of Management, KIIT University "A Critical Analysis of Its Impact on Personal and Social Life".

[3] ADAM N. JOINSON "Self-disclosure in computer-mediated communication:The role of self-awareness and visual anonymity".

[4] Heting Chu "Information representation and retrieve in the digital age" Second edition.

[5] Ann Dozier, PhD (Chair), Karen Hacker, MD, MPH, Mina Silberberg, PhD, Linda Ziegahn, PhD "The Value of Social Networking in Community Engagement"

[6] Berry Wellman "Pysical place and cyberplace : the rise of personalize network."

[7] Lada A. Adamic , Eytan Adar "Friends and neighbors on the Web" 2003.

[8] Sebastian Marius Kirsch, Melanie Gnasa, and Armin B. Cremers "Beyond the Web: Retrieval in Social Information Spaces".

[9] Manos Papagelis, Gautam Das," Real Time Enhanced Random Sampling of Online Social Networks"2013

[10] 1M.V. Mohan Babu PG Scholar in CSE, 2C.V. Chiranjeevi Kumar M.Tech, CSE "SOCIAL NETWORKING SITES STRUCTURE AND FEATURES"2014.

[11] Mr. Yogesh P Murumkar , Prof. Yogesh B. Gurav "SAMPLING ONLINE SOCIAL NETWORKS USING OUTLIER INDEXING" 2014