International Journal of Technical Research and Applications e-ISSN: 2320-8163,
www.ijtra.com Volume 4, Issue 1 (January-February, 2016), PP. 164-167

SYNOPSIS OF SECURITY AND PRIVACY: NEW MODULE OF TRUST FOR ACQUAINTANCES IN ONLINE SOCIAL NETWORKS

Fadi Al-Ayed[1], Mohammed Alhaqbani[2]
School of Electrical Engineering and Computer Science
The Catholic University of America
Washington D.C. 20064. USA
[1]93alayed@cardinalmail.cua.edu
[2]90alhaqbani@cardinalmail.cua.edu

Abstract— **Online social networks (OSN's) such as Twitter, Facebook, and Nexopia have gained a lot of popularity in the last decade, and has gone through some drastic changes in recent years. These Online social networks, allow people to create profiles, blogs, articles, and share content virtually with their families as well as friends at a massive scale. However, some datasets are publicly available on various sites, i.e., socialnetworks.mpi- sws.org[6]; aiming to examine and analyze the behaviors and trends social networks to understand the complex behavior and to inform the design of human-centered systems.**

In this article, we built a real data set from nexopia.com; in order to conduct several experiments and measurements with the objective of gathering a large set of unique users and extracting the associated friends for each user. Furthermore, we propose a module of trust by indexing the communal friends from the extracted dataset.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – Retrieval models;D.1.3 [Programming Techniques]: Concurrent Programming – Parallel programming.

Index terms- **Information Dissemination, Social Networks, Network Structure, Distributed Graph Mining.**

I. INTRODUCTION

Online social networks have been exponentially growing with contents being disseminated and information sharing as a result of user interactions across social networks.

Today, with over 1.2 million registered members and over a billion pages served per month [3] ; the marketing aspect, with its aim to attract user attention, plays an important part in sites' behavior and user activity.

One of the shortcoming aspects of online social networks the is losing the potential activities of the site content dissemination, i.e., content propagation among friends in social networks. Nexopia, is an online social network for messaging system; founded in 2003 and acquired by digital ad network Ideon Media [5] in 2012. Since Nexopia was founded a year prior to Facebook, its' social network is still active. However, Nexopia users fluctuate from time to time over current online social network.

In this paper, we extract hidden data and analyze a large dataset of disseminated information from Nexopia's social network. The analysis was based on crawling of many users as well as their friends with the goal of formulating minding models. We crawled objects known as users in order to collect discrete information from over a 100k unique objects with an approximation of 23 million of their friends. This primary approach will assist to analyze large-scale of information in the social networks.

The rest of the paper is structured as follows. In section 2 we explore some of the related work and techniques used to disseminate information on OSN. In section 3, presents an overview of Nexopia and our data collection methodology. In section 4, we briefly illustrate some notations and descriptions applied in this paper. In section 5, we analyze our experimental results. Finally, we conclude in section 6.

II. RELATED WORK

An extensive amount of the research has be done on social networks which have used the concept of information dissemination and how content can be used to obtain and analyze online social networks Recently, researchers started to explore the area of social networking giving its rising popularity. The research mostly is aimed in examining users' interaction and analyzing the whole network in general.

Cha et al. [1] conducted a large scale of measurement on Flickr social network, they trace content propagation based on crawls of the favorite markings and targeted picture popularity across Flickr's social network based on characterization and role played by social links in information propagation.

Anagnostopoulos et al. analyzed the propagation of photo tags in Flickr and constructed a numerical assessment to distinguish social influence from correlation [2]. They concluded that tags used by Flickr users is not likely to be due to social influence. They found that the choice of tags used by Flickr users is not likely to be due to social influence.

Wilson et al. [4] discussed user interaction in social networks and their implications, focusing on studying if the social network connection means that there is a real life connection between the nodes. They used Facebook data to do their study. They used a crawler to collect their data for three months in 2008 which consist of wall posts and pictures. They used features that Facebook offers to select

10 random users for a giving regional network. They gathered around 10 million users' data to do the study. The authors claim that they cover 56% of the networks the crawl where the rest where having restrictive privacy settings. They found out that most users only interact with a small set of friends which indicate that social network ties does not always imply a real relationship between the two.

III. MEASUREMENT METHODOLOGY

In this section, we describe Nexopia's website objective, and illustrate our data collection methodology.

A. Nexopia

Nexopia is an individual photo sharing site with a remarkable social network characteristics, where users have the ability to create their own profiles, galleries, and add friends to share pictures, blogs, and articles. The site's activities and users behaviors play an essential role in attracting prospective users.

The most common objective that arises the interest of users are contents that feature certain attractive characteristics such as high quality and ranked amongst users. Currently, many websites do not have the feature and capability of ranking in order to display items as the highest level.

In other words, there exists some websites that determine the recent uploaded contents along with their hits, ranks, rates, and viewpoints, but do not have that functionality to sort the highest object on each different categories of photos, videos, music, and blogs.

Nexopia provides two types of users: regular users (Free) and premium service (Plus) users. The regular users have a limited control access to standard services such as friends profiles, comments, messages, forums and so forth. However, for the premium service(Plus) users have an unlimited access to the extra services such as having more photo upload features and additional storage capacity for the images or galleries as well as the creation of forums with no restrictions, advanced user search, customization options including layouts, and the elimination of unsolicited scripts or ads.

B. Data collection methodology

We demonstrate our Nexopia information collection because we were interested in learning the network connections of individual's data flow in a large-scale online social network, we aimed to gather the most recent information of that social network. Data gathering methods vary in many ways and here we illustrate our method for gathering data.

Dataset is retrieved by using an API website crawling technique. Website crawling is used for filtering and fetching hidden data based on the indexes of the URL's requests. We gather some users' hidden information by analyzing the HTML web pages to have output datasets such as users' ages, genders, hits, locations, and the total number for their friends as well as listing their friends in classified order. The core objective is to obtain a large number of non-repetitive unique users and their associated friends to construct a network topology.

In our approach, we came up with a technique to retrieve useful data behind the contents of the web pages by implementing a designated HTML tags as keyword search engine using advanced texts segmentations (indexes). The technique that has been used is API DOM for parsing strategy, and HTML tags representing as nodes indexing for information retrieval.

There are two types of APIs parsing , Simple API for XML (SAX) and Document Object Model (DOM). However, there are some advantages and disadvantages of implementing these techniques. In our scenario, we apply API DOM parsing, since it does handle Xpath whereas SAX does not. Furthermore, because we have a large dataset, DOM parsing supports navigation to an object as a tree structure. Thus, we crawled a vital portion of Nexopia's dataset by selecting a random users along with associated users(friends followed) list in a breadth first search (BFS) approach.

C. Data description

We crawled the Nexopia social network graph in a series of days started from February 28-March 19, 2013. In order to construct a graph, we crawled approximately 100k Nexopia users along with 2.1 million edges. For each vertex in a graph represents a unique user and for each edge exists when two users are friends.

D. Limitations

Despite the fact that this information provides us a distinctive chance to examine data propagation dynamically via social hyperlinks, it contains various limitations. Initially, our methodology does not look at to any deleted user from online social network. However, this does not impact our evaluation outcomes; due to the fact that deleted users can no longer distribute through a social network.

Second, we do not know the exact time when the users deleted their accounts in Nexopia social network. Lastly, we are able to only observe the network, but we cannot alter or adjust it in any manner. We are not in a position to create alterations towards the Nexopia web page or run tests within a controlled atmosphere.

IV. PRELIMINARIES

In this section, we briefly illustrate some notations and descriptions applied in this paper.

A. Graph notation

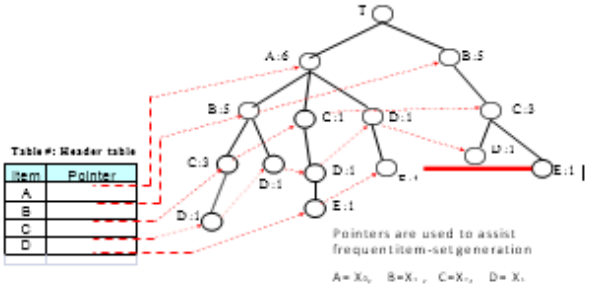
G = (Xn, Xk) is an graph where the Xn is a set of unique users which represent vertices and Xk⊆XnxXn is a subset of edges that represent when a pair of non-repetitive two users are connected. The mapping of graph is a pair M = (Xn, Xk)nk=1,2,3,..m. where Xn, Xk> 0 and n≠k.

B. Dataset Mining Estimation

According to the preceding studies on social networks, it is obvious that there are several statistical calculations to be enhanced in order to obtain service access and resourceful content. In this section, we estimate the candidate trusted users from leaf in tree by using the apriori algorithm that applies to both breadth first search and a Hash tree scheme. Given a threshold C, it recognizes the set of given unique users along with their friends which are subsets of a minimum of one C transactions in the main dataset.

Table 1: C Transaction - Trust Model

TID	Items
1	{X1,X3}
2	{X2}
3	{X0,X1,X2,X3}
4	{X0}
5	{X1,X2,X3}
6	{X0}
7	{X1}
8	{X0,X2}
9	{X0}
10	{X1}
11	{X0}



FP-Trust Tree

V. EXPERIEMENTAL RESULTS

Our dataset depends on the operation of <key, value> pair values of users. The overall execution can be simply described as the following streaming. Our dataset depends on the operation of <key, value> pair. The complete execution can easily be referred to as the following streaming:

Input | Neighbors | Mapping | Indexing | Ranking | Output

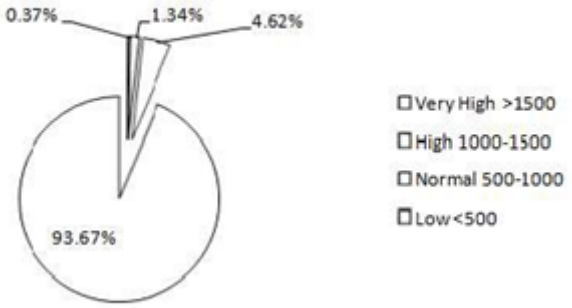


Figure 1:Data Scaling

We examine the module of trust into four levels such as very high, high, normal and low. After we apply the calculation of common repeated friends from the leaves in a tree, we classify the range of existing 500 edge in between each leave of trust.

Table 2: Summary of Nexopia data set

#	Level of Trust	Range of Edges	No. of Nodes	Total Repeated
1	Very High	>1500	2	3908
2	High	1000-1500	12	14082
3	Normal	500-1000	72	48432
4	Low	<500	68268	982653

Our experiments are performed on an Intel(R) Core(TM) i5-4210U CPU, 1.70GHz machine with 8GB main memory running Windows 8 operating system. The programs for the metrics are calculated in Java and Pajek.

VI. CONCLUSION

Social networks have gained a lot of popularity in the past decade. Some social network's websites have a huge amount of data, and has millions of users connected to each other forming a very large and complex network. In our paper, we retrieved unique users from Nexopia's social network by applying an API DOM parsing technique and crawled the dataset by selecting a random users along with their perspective associated users list in a breadth first search (BFS) approach. We analyzed the findings and proposed a trust model to illustrate the candidates with the highest level of trust among a portion of users. The apriori algorithm with breadth first search and a Hash tree scheme was used to attain our results. A number of results can be very promising and might help on developing social networks in exploring the level of trust amongst the users.

REFERENCES

] Meeyoung Cha , Alan Mislove , Krishna P. Gummadi, A measurement-driven analysis of information propagation in the flickr social network.

[2] A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and Correlation in Social Networks. In ACM SIGKDD, 2008

[3] Nexopia. http://www.nexopia.com/about

[4] Wilson, C.; Boe, B.; Sala, A.; Puttaswamy, K.P.; and Zhao, B.Y. 2009. User Interactions in social networks and their implications. In Proceedings of the 4th ACM European Conference on Computer System. (Nuremberg, Germany, April 01 - 03, 2009). EuroSys

'09. ACM, New York, NY, 205-218.

[5] R. Lewis. Canadian Social Networking Pioneer Nexopia Acquired by Digital Ad Network. In Proc. of WWW, 2012

[6] MPI. http://socialnetworks.mpi-sws.org .