Exploration and Optimization Of Friends’ Connections In Social Networks

Kashif Nisar and Edith N. Ekwem*

School of Computing (Information Technology), Utara Malaysia University, Malaysia

E-mail: kashif@uum.edu.my, ekwemedith@yahoo.com

* Tel.: +60135974880

Abstract: One paragraph only. Over the past few years, the rapid growth and the exponential use of social digital media has led to an increase in popularity of social networks and the emergence of social computing. In general, social networks are structures made of social entities (e.g., individuals) that are linked by some specific types of interdependency such as friendship. Most users of social media (e.g., Facebook, LinkedIn, MySpace, Twitter, Flickr, YouTube) have many linkages in terms of friends, connections, and/or followers. Among all these linkages, some of them are more important than others. This paper discusses related work on social networks and method use in crawling online social network graph.

Keywords Social network; social computing; analysis; social media; security; trust.

1. Introduction

Social networks are among the foremost widespread sites on the web since Internet has bred several varieties of information sharing systems [1]. The rapid advancement and exponential use of social digital media over the last decades has led to a proliferation in popularity of social networks and social computing emergence [1, 2, 3, 4]. According to [5, 6] a collection of Internet based applications that are constructed on the ethical and technological fundamentals of Web 2.0, and permit the design and exchange of user contend generated is referred to as social networks also known as social media or a friend-of-a-friends. It can also be defined as a description of the social structure between actors, mostly persons, groups or organizations. When the person uses known/unknown people to create new contacts, it forms “social networking”. The primary objective of social networking user’s is to make connections, communication and
maintain relationships. In addition, social networking sites accelerate the sharing of knowledge, increase teaming, augment communication, and information management between workforces in dissimilar sections.

Numerous sites are dedicated to finding and maintaining contacts and sharing different types of content. Social networks sites like Facebook, Twitter, MySpace and so on, represent a new kind of information network that differs significantly from existing networks like the Web 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 LiveJournal and BlogSpot are used to share blogs. For example, in the Web, hyperlinks between content form a graph that is used to organize, navigate, and rank information. The properties of the Web graph have been studied extensively and have led to useful algorithms such as PageRank [7] which is an algorithm used by Google Search to rank websites in their search engine results. Users join a network, publish their own content, and create links to other users in the network called “friends”. This basic user-to-user link structure facilitates online interaction by providing a mechanism for organizing both real-world and virtual contacts, for finding other users with similar interests, and for locating content and knowledge that has been contributed or endorsed by “friends”. However, there is still constraint in the areas of trust and security in social networking sites because its user’s data are not safe and their data can result in loss of properties, wealth, etc.

The present online social networks do not provide trust and security to users [8]. Their centralized architecture and techniques employed for online information sharing has left wide holes for online fraud, threatening users’ lives. Therefore, to ensure and enhance privacy and security architecture is needed to be model that incorporates privacy principles and provides secure mechanisms for information sharing, to its users. The issue of information privacy has been captivating with social networking users consider themselves victims because their information privacy has been compromised [9]. Users of social network are been cheated, given wrong identities, most of their information are not safe, etc. Program like Beacon, which is part of Facebook advertisement system that sent data from external website to Facebook, has triggered user’s protest over privacy issues. In addition, there are many other policies used by social networking sites where privacy and trust of the user’s may be violated.

For social networking site user’s, there are many privacy and trust consideration that needs to be addressed. For example, information revealed in a user’s profile can lead to risk like identity theft, online stalking, and cyber harassment [10]. However, social networking site operators have provided many security features for preserving the privacy of user’s. Despite all such features, the impact of security and trust on user’s willingness to share information with-in the social networking sites need to be addressed especially using Facebook context.
The rest of this paper is organized as follows: Section 2 presents the related works. Section 3 describes the method use in crawling online social network graph. Finally, Section 4 concludes this paper.

2. Related work

In this section, review of some of the related work done by previous scholars on social networks and social connections are discussed.

Social network is rooted in the field of Sociometry. Milgram and Travers [11] presented the famous theories of the “six-degrees of separation and the small globe”. Social network attracted the attention of numerous sciences including computer science. For instance, a study by Granovetter [12] argues that a social network can be divided into “strong” and “weak” ties, and that the strong ties are strongly clustered. In order to create friends’ connections with other members and start communication, Boyd and Ellison [13] used social network sites (SNSs) that permits users to register, make their own profile page comprising their information either be genuine or virtual. To improve the collaboration and wide-spread of knowledge, Zhou et al. [14] constructed a social network miming solution to determine the social network users’ relationship, major figures and impaction to the organization on bulletin board services (BBS) Website to apprehend the inner and outer link of an organization. In addition, Lewis et al. [15] presented a new social network dataset site Facebook.com with results to exemplify the scientific and academic possibility of this firsthand network source with recommendation views. However, it is difficult to ascertain the benchmarks for social networks since social network research represents a range of skills from anthropology to computer sciences [16].

Other studies include PageRank [17]. Travers and Milgram [11] conducted an experimental study on the essential set of “small world” constraints which offers awareness about the network design rather than reconstructing the real networks. The authors in this study, attempts to probe the distribution of path lengths in a friend network by passing a document to one of their first name acquaintance to allocated individual. Many document got missing while processing it and only six people could effectively targeted and passed on average which consequently led to the “six degrees of separation”, coined by Guare [18].

Marsden [19] discussed a review of the issues concerning controlling the feasibility source of variation in social network data grouped directly by making use of questionnaires and interviews, citing reasons why scholars tried to adopt other potential methods for searching social network and collaborations. In this type of connection network, participants work together in groups of different types and connections among pairs of individuals are recognized by collective group relationship. Networks of co-authorships amongst academicians, where linking of individual is possible if they are co-authors in one or more papers as explained in [20-21] are examples of such network. Carrington, Scott, and Wasserman [5] conducted a study to determine the position
of nodes in the network graph using singular value decomposition as used in the field of social network data.

More so, a study can be seen in Golbeck and Hendler [22] that described the advancement of trust relationships among friends in online social networks to investigate the behavior of users.

Privacy issues have a lot to do with social network structure as well. Zhou and Pei [23] back this up by highlighting an example of where privacy is implemented but this does not stop data being leaked. In their example, they published a social network of close friends after removing the identities of the people on the graph to preserve privacy. This concept is known as anonymity where data cannot be traced back to an individual. The problem comes when an enemy of the social network decided to disclose the information because they know the neighbours of somebody. Also, the link privacy problem raised by Korolova et al. [24] concerns how an attacker discovers the social graph. The goal of the attacker is to maximize the number of nodes/links it can give the number of users it bribes (crawls). Several attacks evaluated in [24] actually correspond to node selection algorithms for crawling, such as Breadth-Search-First (BFS), Depth-First Search (DFS), Forest Fire (FF) and Snowball Sampling (SBS), Re-Weighted Random Walk (RWEW) and Metropolis-Hastings Random Walk (MHRW) [25-27].

3. Methodology

Online social networks can be represented as graphs, whereby nodes denote users, and edges represent connections. Most Crawlers (which are programs that exploit the graph structure of the Web to change from one page to another) make use of sampling techniques [28]. Our methodology will be to obtain a representative sample of social network sites like Facebook, Twitter, MySpace, etc., users by Crawling its social graph. So that users frequencies attributes such as age, name, privacy settings, etc., will be estimated. Additionally, probability sample of users will allow estimation of certain local topological properties such as node degree distribution, clustering and assortativity [25, 27, 29, 30, 26].

Presently, the algorithms for crawling online social networks can be divided into two main classes: graph traversal techniques and random walks techniques. In graph traversal techniques, nodes are sampled without replacing them and once a node is visited, it can’t be visited again. Examples include Breadth-Search-First (BFS), Depth-First-Search (DFS), Forest Fire (FF) and Snowball Sampling (SBS) [25, 26]. Previous scholars [25, 31, 26] uses BFS as a basic technique used extensively for sampling online social networks. A motive for this popularity is that (even incomplete) BFS sample gathers full view (all nodes and edges) of some certain region in the graph. However, BFS has been shown to lead to a bias towards high degree nodes in numerous artificial and real world topologies [30, 32]. Random walks on graphs are a well-studied topic used for sampling the World Wide Web (WWW) [33], peer-to-peer networks [34], and other large graphs [35]. Similarly to traversals, random walks are usually biased towards high-degree
nodes. However, by using classical results from Markov Chains, random walks bias can be investigated and corrected.

Metropolis-Hastings Random Walk (MHRW) algorithm is the overall Markov Chain Monte Carlo (MCMC) technique in sampling from a probability distribution that is difficult to sample directly [36]. Alternatively, we can re-weight the sample after it is collected which will result to Re-Weighted Random Walk (RWEW). Expected results will show that MHRW and RWRW work remarkably when compare to BFS.

3.1 Process for Crawling a graph

The process for Crawling a graph can be outlined as follows as depicted in Figure 5.

1) Put seeds into a queue.
2) Select a node from the queue.
3) Crawl the node.
4) Add the newly found nodes into the queue.
5) Go to Step 2 or terminate if the stop conditions are met.

Figure 5, shows the flow of a basic sequential crawler. The crawler maintains a list of unvisited URLs called the frontier (list of a crawler that contains the URLs of unvisited pages). The list is initialized with seed URLs which may be provided by a user or another program. Each crawling loop involves picking the next URL to crawl from the frontier, fetching the page corresponding to the URL through HTTP, parsing the recovered page to extract the URLs and application specific information.
Finally, the unvisited URLs are added to the frontier. When a certain number of pages have been crawled, the crawling process may be terminated. The scenario signals a dead-end if the crawler is ready to crawl additional page and the frontier is vacant [28]. For instance, the Web is seen as a large graph with pages at its nodes and hyperlinks as its edges. A crawler normally begins at few of the nodes (seeds) and then follows the edges to reach other nodes.
4. Conclusions

We have presented an analysis of the structural properties of social networks. Related works on social networks and method use in crawling online social network graph have been discussed. In our future work, we will implement properties of social network.

References


© 2014 by the authors; licensee Asia-Pacific Advanced Network. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).