INTERNATIONAL JOURNAL OF RESEARCH
In Advanced Engineering Technologies

Volume 2, Issue 2 SEP 2014

SOCIAL NETWORKING SITES STRUCTURE AND FEATURES

[1]M.V. Mohan Babu PG Scholar in CSE

mohan09cse@gmail.com

[2]C.V. Chiranjeevi Kumar M.Tech, CSE

cvchiru@gmail.com

ABSTRACT: —

There has been came into sight social networking sites of online, 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 concentrate on civilizing the functionality of gathering data from the locality of a client in a vibrant social network. We initiate sampling depended algorithms to perfectly explore a client's social network with regard to its construction and to speedily expected quantities of interest. We initiate and examine variations of the fundamental sampling plan of correlations exploring all over trials. Distributed and centralized social networks models are under taken. We explain that our algorithms to rank items can be utilized in the neighborhood of a client, undertaking that data for each client in the network is accessible. Utilizing synthetic and real information sets, we authenticate the outputs of our analysis and show our algorithms effectiveness in estimating quantities of concern. The techniques we explain are common and can most likely be simply inherited in a diversity of policies targeting to proficiently gather data from a social graph.

I. INTRODUCTION

THE changing trends in the use of web technology that aims to enhance interconnectivity, self-expression, and information sharing on the web have led to the emergence of online social networking services. This is evident by the multitude of activity and social interaction that takes place in web sites like Facebook, Myspace, and Twitter to name a few. At the same time the desire to connect and interact evolves far beyond centralized social networking sites and takes the form of ad hoc social networks formed by instant messaging clients, VoIP software, or mobile geosocial networks. Although interactions with people beyond one's contact list is currently not possible (e.g., via query capabilities), the implicit social networking structure is in place.

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. To further motivate our research, we discuss in more detail the case of social search. 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 gathering and examining data from a client's implicit or explicit social network we 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.

The utility of social search has been established via experimental user studies. For example, in [1], Mislove et al. report improved result accuracy for web search when urls for a query are not ranked based on some global ranking criteria, but based on the number of times people in the same social environment endorsed them. Many ideas have been suggested to realize online social search; from entirely human search engines that utilize humans to sort the search outcomes and help client in expounding their search requests to social-influenced algorithms that exploit a user's web history log to influence result rankings, so that pages that she visits more often are ranked higher. In any case, the goal

INTERNATIONAL JOURNAL OF RESEARCH
In Advanced Engineering Technologies

Volume 2, Issue 2 SEP 2014

outcomes informed by human judgement, as disparate to long-established search engines that frequently return a huge numeral of outcomes that may not be relevant. These are all examples of tasks that require visiting and probing a large number of peers in the extended network of an individual for information that lies locally in their logs, and then using this information to improve the quality of search experience. Despite the fact that many algorithms and tools exist for analysis of networks, in general, these mainly focus on analysis of the properties of the network structure and not on the content of the nodes. They also typically not operate on user specific graphs (i.e, users' neighborhoods), but on the whole graph. Instead, for many modern applications, it would be beneficial to design algorithms that operate on a single node. For example, in the case of social search, it would be beneficial to design algorithms that starting from a specific user in the network, crawl its (extended) neighborhood and collect information that lies on their close peers. Such networks may consist of thousands of users and their structure may not be static, thus a complete crawl of all social peers is infeasible.

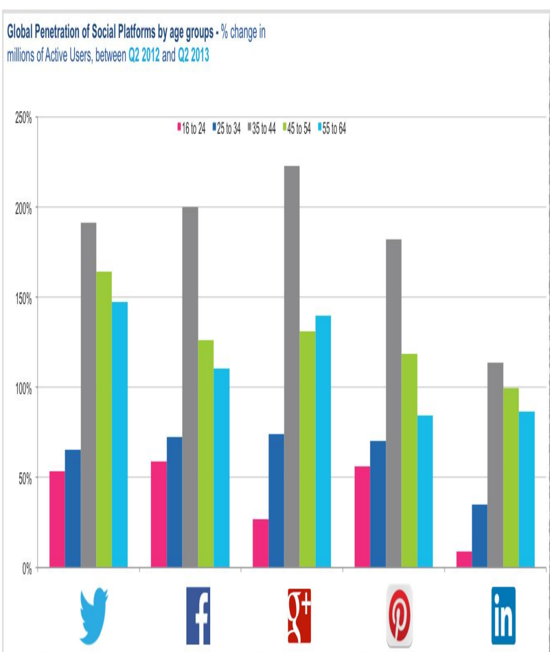

Figure: The social networking sites part in online

Therefore, efficient methods are required. Depending upon these observations we concentrate on humanizing the is to provide end clients with a restricted digit of relevant network of a client and describe the below contributions:

. Sampling-based algorithms are introduced that provided a client in a social network speedily gain a uniform of near nodes random sample. We utilize these algorithms to speedily fairly accurate the digit of client in a client's neighborhood that have authorized an item.

. We initiate and examine variants of these fundamental sampling methods in which we target to decrease the whole nodes number in the network trip by discovering correlations from corner to corner samples.

. Sampling-based algorithms are estimated in terms of efficiency and accuracy utilizing synthetic and real data and exhibit the usefulness of our approach. We demonstrate that our fundamental sampling methods can be used for a multiplicity of strategies in a network targeting to rank items, assuming that data for every client is available in the network.

II PROBLEM STATEMENT

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 (where knowledge of the network structure is assumed) or distributed (where network structure is unknown or limited). 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, are more common. Consider for example, the case in ad hoc social networks formed by typical instant messaging or VoIP protocols (e.g., MSN and Skype). There are also cases that the model of the graph is between the centralized and the fully distributed allowing limited knowledge of a node's neighborhood typically controlled by the node in terms of privacy settings (e.g., LinkedIn and Facebook). Our methods apply to these models as well, with slight modifications. 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. Although one can make a case for relatively static social networks (in which the graph structure changes less frequently) in general such graphs are expected to be highly dynamic. We focus on dynamic networks (either centralized or distributed) but also treat the relatively easier case of static networks. In any case, we assume that the rate of change of the content in these functionality of data collection from nearby nodes in a social following two problems of interest in this paper.

III SYSTEM DEVELOPMENT

Description of Data Sets

For the needs of our experimental evaluation, we consider real and synthetic network topologies along with real and synthetic user search history logs (i.e., clickthrough data). To come up with social search logs suitable for our experiments we had to combine these sources. Below we provide details on data characteristics and the data collection process.

Network Topologies. In our experiments, we consider one real and two synthetic network topologies. The real network topology, e pinions-net, is an instance of the Epinions'3 real graph. The first synthetic topology, uniform- net, simulates a uniform random network topology, whiles the second synthetic topology, prefatt-net, simulates a preferential attachment network topology. When generating the synthetic networks, we set the parameters of the graph generators so that the formed networks have similar number of vertices and edges to the real network of Epinions (The JUNG4 tool has been used to generate the networks.).

User search history logs. We experiment with real and synthetic user search history logs. For the needs of our experiments we create real_log by randomly selecting 75,888

users from the AOL data set along with their search history logs (about 4M queries, 3M urls) [13]. The synthetic log, synth_log, consists of the same users as the real_log but we populate user's history logs with high numbers of queries and url counts. A summary of these logs is provided in Table 4.

Formation of final data sets. Thus far, we have come up with proper network topologies and user search history logs. In order to generate suitable final data sets for our experiments

we randomly map each of the 75,888 AOL users in a search history log to the 75,888 nodes of a network topology. Note that since the focus of our work is on performance we do not require that "similar" users are placed in adjacent network nodes. Thus, we do not care for any semantics of the data destroyed, such as the fact that friends may have similar interests and so. The objective of our work is to efficiently collect information in social networks. In doing so, we consider all users in the network to be equal, independently on whether they lie a few or

many hops away from the initiator node. In a real setting, one may make use of such semantics to design a better re-

networks is high. Given such an environment we define the this paper.

Cost Analysis:

Our sampling algorithms provide an alternative to performing an exhaustive search or crawling on the network of a user using a depth-first-search or breadth-first-search. Both DFS and BFS assume that there is a designated initiator node from which the search starts and define a DFS or BFS tree. At the end, nodes at distance d from the initiator appear at level d of the tree. In this paragraph, we present we simple cost model that helps to analyze and compare the complexity of our basic sampling algorithms to that of crawling.

IV RELATED WORK

Our work is related to work on sampling large graphs via random walks. Generating a uniform random subset of nodes of a graph via random walks is a well studied problem; it frequently arises in the analysis of convergence properties of Markov chains or the problem of sampling a search engine's index [19], [20]. The basic idea is to start from any specific node, say v, and initiate a random walk by proceeding to neighbors selected at random at every iteration. Let the probability of reaching any node u after k steps of this walk be pðuÞ. It is known that if k is suitably large (the value of k depends on the topological properties of the graph), this probability distribution is stationary (i.e., it does not depend on the starting node). However, this stationary distribution is not uniform; the probability associated with each node is inversely related to its degree in the graph. This stationary distribution can be made uniform using techniques such as the Metropolis Hastings algorithm (see [21]), or using rejection sampling (where, after reaching a final node, the node is included in the sample with probability inversely proportional to its degree). This process can be repeated to obtain random samples of a desired size. Similar approaches have been employed in [22] where Hastings describes sampling-based methods to efficiently collect information from users in a social graph and in [23] where sampling techniques are used to collect unbiased samples of Facebook. Likewise, in [24] Katzir et al. design algorithms for estimating the number of users in large social networks via biased sampling, and in [25] sampling methods are proposed to approximate community structures in a social network. Our research presents ways to improve upon these generic random walk methods on graphs by leveraging the fact that we need to sample from the ranking algorithm, but this is orthogonal to the objective of 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. In [26], the CubeSVD approach was developed to improve Web search by taking into account click through data of the type "user, query, url." Further studies showed that taking into account such data and building statistical models for user behavior can significantly improve the result ranking quality [27], [28]. Other approaches exist, as well, that utilize some notion of relevance feedback to rerank web search results [29], [30].

Social search informed by online social networks has actually gained attention as an approach toward personalized search. In a sense, utilizing information from one's social environment to improve on user satisfaction is a form of "extended" personalization, with the extent being defined as a function of the neighborhood of an individual in the network. Many ideas have been suggested to realize online social search; from search engines that utilize humans to filter the search results [31], to systems that utilize real-time temporal correlations of user web history logs [32], [33], to tag-based social search systems [34]. Analyses suggest that integration of social search models improves the overall search experience. Our research is complementary as we aim to offer performance improvements, via sampling, to the process of collecting information from user logs by exploring the graph structure offered by a social network.

V CONCLUSION AND FUTURE WORK

Social networking sites have its own mark in this time period throughout the world using World Wide Web. Our research suggests methods for quickly collecting information from the neighborhood 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. The utility of our approach was demonstrated by running experiments on real and synthetic data sets. Further, 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. Despite its competence, our work inherits limitations of the sampling method itself and is neighborhood of a node v (i.e., a few links away from v).

selectivity. A similar problem arises in approximately answering aggregation queries using sampling. Solutions there rely on weighted sampling based on workload information [35]. However, in our context where data stored at each node are rapidly changing this method is not directly applicable. Our algorithms assume that information for each user in a network, such as web history logs, is available. Access to personal information infringes on user privacy and, as such, privacy concerns could serve as a major stumbling block toward acceptance of our algorithms. Systems that utilize our algorithms should adhere to the social translucence approach to designing social systems that entail a balance of visibility, awareness of others, and accountability [36].

REFERENCES

[1] A. Mislove, K.P. Gummadi, and P. Druschel, "Exploiting Social Networks for Internet Search," Proc. Fifth Workshop Hot Topics in Networks (HotNets), 2006.

[2] W.G. Cochran, Sampling Techniques, third ed. John Wiley, 1977.

[3] D.E. Knuth, "Estimating the Efficiency of Backtrack Programs," Math. of Computation, vol. 29, no. 129, pp. 121-136, 1975.

[4] G. Cornujols, M. Karamanov, and Y. Li, "Early Estimates of the Size of Branch-and-Bound Trees," INFORMS J. Computing, vol. 18, pp. 86-96, 2006.

[5] P. Kilby, J. Slaney, S. Thie´baux, and T. Walsh, "Estimating Search Tree Size," Proc. Nat'l Conf. Artificial Intelligence (AAAI), 2006.

[6] H. Kautz, E. Horvitz, Y. Ruan, C. Gomes, and B. Selman, "Dynamic Restart Policies," Proc. 18th Nat'l Conf. Artificial Intelligence (AAAI), 2002.

[7] L. Katzir, E. Liberty, and O. Somekh, "Estimating Sizes of Social Networks via Biased Sampling," Proc. 20th Int'l Conf. World Wide Web (WWW), 2011.

[8] R.G. Miller, Simultaneous Statistical Inference. Springer Verlag, 1981.

expected to be inefficient for quantities with very low

[9] M.N. Garofalakis and P.B. Gibbons, “Approximate Query Processing: Taming the Terabytes,” Proc. 27th Int’l Conf. Very Large Data Bases, Nov. 2001.

[10] S.A.M. Makki and G. Havas, “Distributed Algorithms for Depth- First Search,” Information Processing Letters, vol. 60, no. 1, pp. 7-12, 1996.

[11] T.-Y. Cheung, “Graph Traversal Techniques and the Maximum Flow Problem in Distributed Computation,” IEEE Trans. Software Eng., vol. SE-9, no. 4, pp. 504-512, July 1983.

[12] B. Awerbuch and R.G. Gallager, “A New Distributed Algorithm to Find Breadth First Search Trees,” IEEE Trans. Information Theory, vol. 33, no. 3, pp. 315-322, May 1987.

[13] C.T.G. Pass and A. Chowdhury, “A Picture of Search,” Proc. First Int’l Conf. Scalable Information Systems (InfoScale), 2006.

[14] R. Albert and I. Barabasi, “Statistical Mechanics of Complex Networks,” Modern Physics Rev., vol. 74, p. 47, 2002.

[15] C. Gkantsidis, M. Mihail, and A. Saberi, “Random Walks in Peerto- Peer Networks: Algorithms and Evaluation,” Performance Evaluation, vol. 63, no. 3, pp. 241-263, 2006.