www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 5 Issue 09 September 2016 Page No.18025-18028

Improving the Performance of Information Collection Using Social Networks with Online Sampling

Praveen Kumar Ramtekkar[1], Shyam P. Dubey[2], Mohammad Shahid Nadeem[3]

[1]Department of CSE, Nuva College of Engineering & Technology, Nagpur, India
pramtekkare@rediffmail.com

[2]Department of CSE, Nuva College of Engineering & Technology, Nagpur, India
shyam.nuva@rediffmail.com

[3]Department of CSE, Nuva College of Engineering & Technology, Nagpur, India
shahid4sam@rediffmail.com

Abstract: *This paper concentrates on improving the performance of information collection from the neighborhood of a user in a dynamic social network. By introducing sampling based algorithms to efficiently explore a user's social network respecting its structure and to quickly approximate quantities of interest. It introduces and analyzes variants of the basic sampling scheme exploring correlations across the samples. As online social networking emerges, there has been increased interest to utilize the underlying network structure as well as the available information on social peers to improve the information needs of a user. Models of centralized and distributed social networks are considered to implement this algorithm.*

This algorithm can be utilized to grade items in the neighborhood of a user, assuming that information for each user in the network is available. Using real and synthetic data sets, this work validates the results of analysis and expresses the efficiency of algorithms in approximating quantities of interest. The methods described are general and can probably be easily adopted in a variety of strategies aiming to efficiently collect information from a social graph.

Keywords: social network, facebook, twitter, online social network, depth first search.

1. Introduction

The changing scenarios 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 marked by the multitude of activity and social interaction that takes place in web sites like Facebook, Myspace and Twitter etc. At the same time the need 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 geo social 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 discover 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 motivate the work the case of social search is discussed in detail. Social search or a social search engine is a type of search method that tries to determine the significance 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 the accuracy of search results can be improved. The most common social search scenario is:

1. A user u 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 u and relates to the results in L.

The search engine utilizes this information to reorder the list L to a new list L' that is presented to u.

2. Review of Literature

A. Social Network

Social Networking, it's the way the 21st century communicates now. Social networking is the grouping of individuals into specific groups, like small rural communities or a neighborhood subdivision. Although social networking is possible in person, especially in the workplace, universities, and high schools, it is most popular online. This is because unlike most high schools, colleges or workplaces the internet is filled with millions of individuals who are looking to meet other people.

Social network is the mapping and measuring of relationships and flows between people, groups, organizations, computers, URLs and other connected information or knowledge entities. The nodes in the network are the people and groups while the links show relationships or flows between the nodes. Social network provides both a visual and a mathematical analysis of human relationships.

B. Online Social Network

Online social networks often feature a web interface that only allows local-neighborhood queries i.e., given a user of the online social network as input, the system returns the immediate neighbors of the user [1].

Online Social Networks (OSNs) have recently emerged as a new internet killer-application. The adoption of OSNs by internet users is off-the-charts with respect to almost every metric. In November 2010 Facebook, the most popular OSN, counted more than 500 million members; the total combined membership in the top five OSNs Facebook, QQ, Myspace, Orkut, Twitter exceeded one billion users. Putting this number into context, the population of OSN users is approaching 20% of the world population and is more than 50% of the world's Internet users. According to [2] users worldwide currently spend over 110 billion minutes on social media sites per month, which accounts for 22% of all time spent online, surpassing even time spent on email. According to [3] Facebook is the second most visited website on the Internet second only to Google with each user spending 30 minutes on average per day on the site more than the time spent on Google.

Sampling Online Social Networks focuses on improving the performance of information collection from the neighborhood of a user in a dynamic social network. It introduces sampling-based algorithms to efficiently explore a user's social network respecting its structure and to quickly approximate quantities of interest. It also introduces and analyzes variants of the basic sampling scheme exploring correlations across our samples. Models of centralized and distributed social networks are considered [4].

Efficient Sampling of Information in Social Networks introduces sampling based algorithms to quickly approximate quantities of interest from the vicinity of a user's social graph. This analyzes variants of basic scheme exploring correlations across our samples. Models of centralized and distributed social networks are considered. It also shows 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 [5].

Walking in Facebook: A Case Study of Unbiased Sampling of OSNs, expresses how do you collect a sample of nodes using crawling?, what can we estimate from a sample of nodes?, sampling techniques like Random Walks/BFS for sampling Facebook, Multigraph Sampling, Stratified Weighted Random Walk, What can we learn from a sample? Etc. [6]

Sampling of User Behavior Using Online Social Network, describes the performance of information collection in a dynamic social network is studied. By analyzing the results, we reveal that personalized search has significant improvement over common web search. The mixing time of thee sampling process strongly depends on the characteristics of the graph [7].

Overcoming Limitations of Sampling for Aggregation Queries, reveals the problem of approximately answering aggregation queries using sampling. It also observes that uniform sampling performs poorly when the distribution of the aggregated attribute is skewed. This also introduces a technique called outlier-indexing. Uniform sampling is also ineffective for queries with low selectivity. It also relies on weighted sampling based on workload information to overcome this shortcoming. This demonstrates that a combination of outlier-indexing with weighted sampling can be used to answer aggregation queries with significantly reduced approximation error compared to either uniform sampling or weighted sampling alone [8].

Distributed Algorithms for Depth First Search, Information Processing Letters presents distributed algorithms for constructing a depth first search tree for a communication network which are more efficient then other methods. A more efficient distributed algorithm for the DFS traversal of a network can help reduce the complexity of other distributed graph algorithms which use a distributed DFS traversal as their basic building block. The improvement over the best of other algorithm is achieved by dynamic backtracking with a minor increase in message length [9].

Random Walks in Peer to-Peer Networks: Algorithms and Evaluation, Performance Evaluation, quantifies the effectiveness of random walks for searching and construction of unstructured peer-to-peer networks. Two cases have identified, where the use of random walks for searching achieves better results than flooding: (a) when the overlay topology is clustered, and (b) when a client re-issues the same query while its horizon does not change much. Related to the simulation of random walks is also the distributed computation of aggregates, such as averaging. For construction, we argue that an expander can be maintained dynamically with constant operations per addition. The key technical ingredient of our approach is a deep result of stochastic processes indicating that samples taken from consecutive steps of a random walk on an expander graph can achieve statistical properties similar to independent sampling. This property has been previously used in complexity theory for construction of pseudorandom number generators. We reveal another facet of this theory and translate savings in random bits to savings in processing overhead [10].

3. Experimental Evaluation

Having presented our sampling methods and algorithms we now turn to evaluation. For the needs of our experiments we make a case of a social search application. Let G be a graph depicting connections between users in a social network, where each node in the graph represents a user.

For each user we assume availability of a click through log accumulated over time via browsing. The log, in its most simple form, at node, has the form q, url_q, $count^u_{url_q}$ where q is a query, url_q is the url clicked as a result of q and $count^u_{url_q}$ is the number of times url_q has been clicked by u. A few years ago, it would be difficult to assume that such a log exists due to privacy issues. However, recently, a number of Web2.0 services gather such kind of information.

The most notable example might be Google's web history service. But also, Bing's and Facebook's attempts to incorporate in search results a feature that shows you the opinion of your friends as it relates to that search, through the Facebook Instant Personalization feature. Therefore, is

reasonable to assume existence of such information. Now, consider the scenario where a query is submitted to a popular search engine by a user u and a set of urls r_{q_v} is returned. A social search algorithm would try to personalize this result. Intuitively, an algorithm might collect information from u's social network and use this information to re-rank the results according to a re-ranking strategy. Using G and starting at u we can obtain the total count of the number of times that each $url_q \in r_{q_v}$ has been clicked by consulting the neighborhood of u at some specific depth (number of hops) d Then a re-ranking r'_{q_v} of r_{q_v} is possible that incorporates the behavior of the users with which u has some social relationship.

4. Conclusion

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 this approach was demonstrated by running experiments on real and synthetic data sets. Further, it has been showed that these 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, this work inherits limitations of the sampling method itself and is expected to be inefficient for quantities with very low selectivity. A similar problem arises in approximately answering aggregation queries using sampling. Solutions there rely on weighted sampling based on workload information. However, in this context where data stored at each node are rapidly changing, this method is not directly applicable. This algorithm assumes that information for each user in a network, such as web history logs, is available. Access to personal information in fringes on user privacy and, as such, privacy concerns could serve as a major stumbling block toward acceptance of our algorithms. Systems that utilize this algorithms should hold to the social fineness approach to designing social systems that entail a balance of visibility, awareness of others and accountability.

References

[1] Azade Nazi, Zhuojie Zhou, Saravanan Thirumuruganathan, Nan Zhang, and Gautam Das, "Walk, Not Wait: Faster Sampling Over Online Social Networks," Proceeding of the VLDB Endowment, vol. 8, issue, 6, pp. 678-689, 2015.

[2] Nilesensratistics, June 2010, http://blog.nielsen.com/nielsenwire/online mobile/social-media-accounts-for-22-percent-of-time-online/.

[3] "Alexa traffic statistics for Facebook," June 2010, http://www.alexa.com/siteinfo/facebook.com.

[4] M. Papagelis, Gautam Das, and N. Koudas, "Sampling Online Social Networks," Knowledge and Data Engineering, Issue No. 3, vol. 25, pp. 662-676, 2013.

[5] G. Das, N. Koudas, M. Papagelis, and S. Puttaswamy, "Efficient Sampling of Information in Social Networks," Proc. ACM Workshop Search in Social Media (SSM), 2008.

[6] M. Gjoka, M. Kurant, C. T. Butts, and A. Markopoulou, "Walking in Facebook: A Case Study of Unbiased Sampling of OSNs," Proc. INFOCOM, 2010.

[7] J. Amarnath, "Sampling of User Behavior Using Online Social Network," International Journal of Computer Applications technology and research, vol. 4(9), pp. 648-654, 2015.

[8] S. Chaudhuri, G. Das, M. Datar, R. Motwani, and V. R. Narasayya, "Overcoming Limitations of Sampling for Aggregation Queries," Proc. 17th Int'l Conf. Data Eng. (ICDE), 2001.

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

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

Author Profile

Praveen Kumar Ramtekkar received the Master of Computer Applications degree from Government Engineering College, Raipur.He is pursuing M. Tech. in Computer Science & Engineering from Rashtrasant Tukdoji Maharaj Nagpur University. He has 14 Years of Teaching and industrial experience.

Shyam P. Dubey received the BE in Computer Science and Engineering degree from Rajiv Gandhi Prodyogiki Vishvavidyalay in 2006 and M. Tech.in Software Engineering from Madan Mohan Malviya National Institute of Technology in 2009. He is a Head of Department CSE & IT in Nuva College of Engineering and Technology, Nagpur. He has eight years of teaching experience.

DOI: 10.18535/ijecs/v5i9.40

Mohammad Shahid Nadeem received the BE and M. Tech. in Computer Science and Engineering from Rashtrasant Tukdoji Maharaj Nagpur University and Rajiv Gandhi Prodyogiki Vishvavidyalaya in 2009 and 2015 respectively. He has six years of teaching experience.