ISSN:2231-0711

Preeti Narooka et al | IJCSET(www.ijcset.net) | June 2016 | Vol 6, Issue 6, 228-232

# Graph Search Process in Social Networks and it's Challenges

Ms. Preeti Narooka[1], Dr. Sunita Chaodhary[2]

*Ph. D Student, Banasthali University, Jaipur, India*[1]

*Former Associate Professor, Banasthali University, Jaipur, India* [2]

*Abstract*: - **Social networking sites are grabbing attention of millions of users and it resulted into creation of humongous data on these networking sites. Number of users on any networking sites denotes the complexity and size of the data. These developments have thrown a challenge in front of computer scientists to optimize the search operations on these social networking sites. This paper outlines the Graph theory as a tool for search operation on social networking sites & for optimization of search graph operation map reducing technology is discussed. In this paper on one side, different traditional system of social networking and graph algorithm are elaborated. On the other side, various application and different issues related to the graph algorithm and social networking are discussed in detail.**

*Keywords:* - **Social *Networks, Graph, Big Data, Cloud, Graph Searching***

## I. INTRODUCTION

"Social Network" provides a platform where friends, family, client etc can interact socially with the help of internet based social media. Social Networks have undergone a tremendous and superfast growth in recent time. And they are providing a powerful reflection of the dynamics society of the 21st century peoples and show the interaction of the Internet with both technology and people. Such networks are suitable for sharing text, videos, photos, audio and other information between peoples through their connections that is represented as Social Graph. Furthermore, it can be applied to many useful areas, such as marketing, storage capability, the detection of crime and terrorists etc. That is the reason that, recently, the web data are the main target in many researches about using information technology in the area of Social Networks [1]. But the rapid increment of the network size, day to day, requires quite attention toward the rising issues and challenges of it. For the robustness and effectiveness of the Social networks, issue of complex data, security, authentication and authorization needed to be solved. According to the today's research scenario of computer department most depending topic is Social Networks. Since it's all about internet, sharing, connections, socializing, where all this comes with number of issues and challenges. Security, huge size of data, exponential increment of networks, high performance of the resource and many more issues and challenges are highlights of the research. In such situation the combination of Graph Algorithms with Cloud Computing is emerge as a natural solutions for complex network where huge amount of data can be distribute and store easily.

Graph algorithm is the efficient tool which defines the relationships between all objects on social cloud. In today's scenario, users are increasing in Social Networks in leaps and bounds. With the increase of users, there is a rapid increment in number of objects on social graph, results in exponential growth of the graph size [12]. Accessing useful information from exponentially growing graph in fast manner is a very difficult challenging task. To provide solution for the aforesaid problem it is the need to optimize the graph for reducing the search time.

The objects of this paper is to show how the social networks are presented as Social Graphs, its different aspects, issues and challenges of social search graph and networks.

## II. GRAPHS

Graphs provide a powerful tool to model objects and relationships among objects. The study of graphs dates back to Euler's days in the 18th century, when he defined the Konigsberg bridge problem, and since then has been pursued by many researchers. Specifically, graphs can be used to model problems in many areas such as transportation, scheduling, networks, robotics, VLSI, compilers, mathematical biology, distributed computing and software engineering. Many optimization problems from these and other diverse areas can be phrased in graph-theoretic terms, leading to algorithmic questions about graphs.

*Graphs:* Graphs are defined by a set of vertices and a set of edges, where each edge connects two of its vertices. Graphs are further classified into directed and undirected graphs, depending on whether the edges are directed or undirected. An important subclass of directed graphs that arises in many applications, such as precedence-constrained scheduling problems, is directed graphs. Interesting subclasses of undirected graphs include trees and bipartite graphs.

## III. SOCIAL GRAPH

Social Graph can be understood by the relationship between "actors", and the term actor can be a person, an organization, an event or an object. In a Social Graph, each actor is presented as a node and each pair of nodes can be

connected by lines to show the relationships. The Social Graph is a graph that formed by those lines and nodes. In Social Graph, Graph analysis and Social Network Analysis are valuable tools for studying the web and human behaviors of the web users. Social Graph may be applied in any web field where a graph may be constructed such as Social Graph Expansion, Facebook Open Graph, etc.

a) *Social Graph Expansion:* Social Graph applied in the web by utilizing the interconnected Web 2.0 blogs and their comments. Social Graph created the tool that made relations easier to track and build. Now every user that has an account in a Social Networking Sites can "tag" information and propagate it to that network [11].

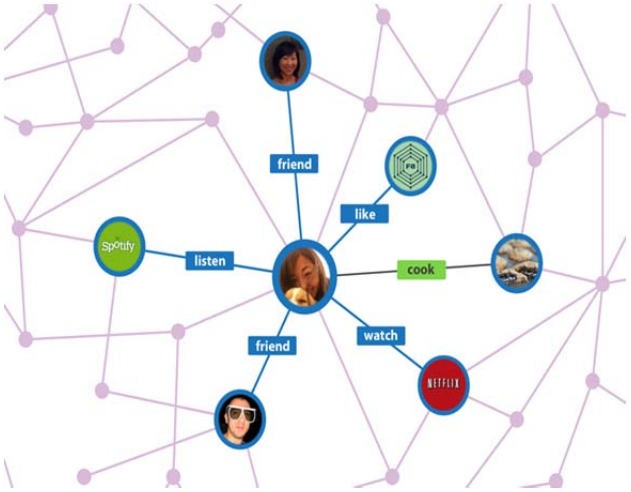


***Fig. 1****: Social Graph Expansion*

*b) Facebook Open Graph:* Facebook Open Graph [8] provides an interface for interconnecting web pages with the Facebook social graph. The most common practice is to add a "Like" button near a media object in a web page and let users share their "Likes". When a user clicks the "Like" button outside the Face-book platform, a new connection is formed in the user's profile.

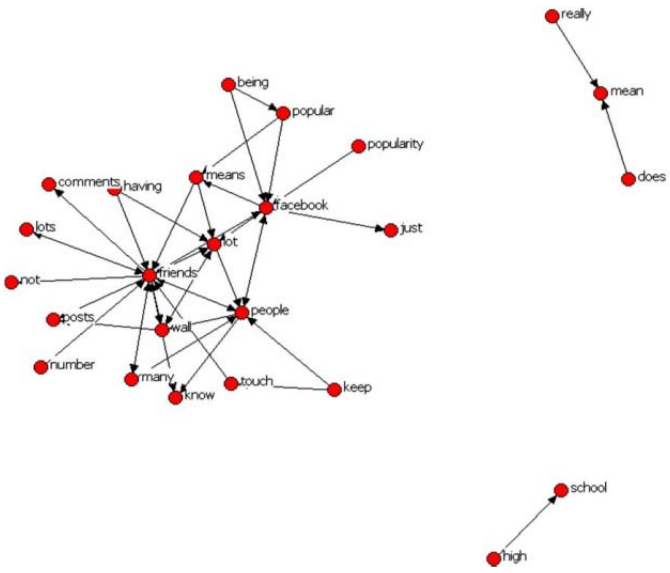


***Fig. 2****: Facebook Open Graph*

## IV. SOCIAL GRAPH SEARCH

The Social Graph in the Internet context is a graph that depicts personal relations of internet users. The social graph has been referred to as "the global mapping of everybody and how they're related". Social Graph have large amount of data & accessing useful information from large amount of data is very difficult.

In the past years, problems emerged from real applications related to the structure of modern (or projected) networks that are expected to be large scale and dynamic, and where agents' behavior can be probabilistic & decentralized. This is one of the reasons why the field of Graph Searching [9] is nowadays rapidly expanding. Several new models, problems or approaches have appeared relating it to diverse fields for example random walks, game theory, logic, probabilistic analysis, complex networks, motion planning, and distributed computing and one most difficult problem is to increase the efficiency of Social Graph Search.

## V. ISSUES AND CHALLENGES:

### *A. Graph Searching*

Everyday lot of people joins the social network which increases the size of graph and whenever the size of graph increase search become too difficult. Search Graph having some problems that make them poorly matched to current problem-solving approaches. In particular, the following properties which mention below are graph problems; present the significant challenges for efficient parallelism.

- Efficiency- In social network connections are complex, unstructured and big in size which decrease the efficiency of search graph means timing to find out the required key takes long search.
- Unstructured problem- these search graphs are highly irregular and unstructured. The irregularity of the graph makes difficulty to parallelism process.
- Very poor locality- Because graphs represent the relationships between entities and because these relationships may be irregular and unstructured, the computations and data access patterns tends not to have very much locality. This is particularly true for graphs that come from data analysis. Performance in contemporary processors is predicated upon exploiting locality. Thus, high performance can be hard to obtain for graph algorithms, even on serial machines.
- Computations of high data - Algorithms of Graph are based on exploring the structure of a graph to perform large numbers of computations on the graph data. As a result, needs a higher ratio of data access for computation. Since these accesses tend to have a low amount of exploitable locality, runtime is dominated by the wait for memory fetches.
- Complex- When the graph is too complex to be processed, it is also necessary to design efficient and effective distributed algorithms. This requires that graph search should take dynamic changes and temporal factors into consideration.

### B. Social Networking

The current adoption of Social Networking graph is associated with numerous challenges [4] because users are still skeptical about its authenticity. The major challenges are as follows:

*i. Avoiding fragmentation of the Social Graph through open cross-platform interactions[7]*

A major hindrance to exploitation of Social Network data is the fragmentation of the population of Social Network users into numerous proprietary and closed Social Networks. This issue is compounded by the fact that each media application tends to build its own Social Network around it rather than building upon the rich data available about existing social relationships. Also applications are often restricted to execute within the confines of specific Social Network platform. A major research challenge, therefore, that would benefit the exploitation of Social Network Graphs for future media networking, is in finding solutions to open up Social Network platforms to allow cross-platform information exchange and usage. Of course, reliable mechanisms to preserve privacy are an essential prerequisite.

*ii. Personalization for Social Interaction*

In order to improve Social Interaction and enhance social inclusion, personalization engines that locate peers with possibly common likes, dislikes or developing trends should be engineered. Towards more efficient search engines that will be able to serve the users only with relevant content, personalization algorithms have to be studied in a greater extent.

*iii. Social and Ethical Issues in a Networked World*

As in every small or large community, online social communities face also critical social and ethical issues that need special care and delicate handling. Sharing of personal information, protection of child exploitation and many other problems have to be studied and answered appropriately.

*iv. Use of Social Networks for business and marketing*

Social Networking introduced novel collaboration paradigms between network users and serious study is conducted on the use of such platforms for internal business purposes. However, one of most prominent research challenges is how to use Social Networking for external communications, customer support and of course targeted marketing.

*v. Data Management*

Data is stored on multiple sites administered by several organizations. Cloud service providers have already collected petabytes of sensitive personal information stored in data centers around the world. This addressed the privacy issue [2].

All these together make it extremely challenging task to develop a big graph search and network with a friendly, interactive, accurate answers and highly efficient.

## VI. RELATED WORK

### A. Graph Search

Investigation has occurred about the graph searching [10] models in large scale networks because the focus of the graph searching researchers has been on deterministic problems and their associated complexity and algorithmic issues. Recent advances in techniques for understanding `large' graphs bode well for making progress in investigating graph searching on networks.

*1) Luczak and Pralat [LP], (2009)*

Using probabilistic methods, they showed that the cop number behaves in a regular but unexpected way. In this they consider the random graph, G(n,p) with expected degree pn=nx, and shows that as x decreases the expected number of cops rises then falls regularly with local extrema at x=1/2,1/3,1/4,.... Result suggests that the large networks will have enough edges and local structure to smooth out the irregularities inherent in small graphs.

*2) Bollobas, Kun and Leader [BKL], (2009)*

Using probabilistic methods, essentially proved Meyniel's bound in random graphs (up to the logarithmic factor). Recently, Pralat and Wormald showed that the logarithmic factor can be eliminated (from both random binomial graphs as well as random d-regular graphs) and so the Meyniel's conjecture is verified for these models. For deterministic graphs, we are still far away from proving the conjecture. Up until recently, the best known upper bound was given by Frankl in 1987 [F87], who showed that the cop number is always of order at most nloglogn/logn=o(n). After twenty years of attacking this problem we have three recent independent proofs (Lu, Peng [LP]; Scott, Sudakov [SS]; and Frieze, Krivelevich, Loh [FKL]) of the same result, namely, that the cop number is at most n/2(1+o(1))sqrt(log2n) (which is still n1−o(1)). We do hope that the Workshop will bring us a bit closer to the solution.

*3) Messinger, Nowakowski, Pralat, Wormald, (2008)*

The cleaning of large scale networks (the Brush problem, Alon) [MNP08, APW08, P09] was also investigated using Wormald's differential equations method for random regular graphs.

*4) Meyniel, (1985)*

He conjectured that if G is a connected graph of order n, then the number of cops needed to capture a robber is of order at most n1/2. This would be best possible because of a construction of a bipartite graph based on the finite projective plane.

*5) Sen Wu, Jie Tang, and Bo Gao (2012 )*

In this paper, they study a new problem of instant social graph search, which aims to find a sub graph that closely connects two and more persons in a social network. This is a natural requirement in our real daily life, such as “Who can be my referrals for applying for a job position?” . In this paper, they formally define the problem and present a series of approximate algorithms to solve this problem:

Path, Influence, and Diversity. To evaluate the social graph search results, we have developed two prototype systems, which are online available and have attracted thousands of users. In terms of both user's viewing time and the number of user clicks, we demonstrate that the three algorithms can significantly outperform (+34.56%-+131.37%) the baseline algorithm. [13]

6) Cormen et al. (1989); Even (1979); Tarjan (1983)

Graph-searching procedures such as depth-first search (DFS) and breadth-first search (BFS) form the basic preprocessing steps for most graph algorithms. Algorithms based on DFS have been known for a long time for the problem of searching mazes. However, it was the work of Hopcroft and Tarjan (for which they received the ACM Turing Award in 1986) that illustrated the full algorithmic power of DFS. They demonstrated efficient algorithms for several problems, such as finding bi-connected components and bridges of a graph and testing tri-connectivity and planarity. DFS on directed graphs can be used to classify its vertices into strongly connected components, to detect cycles, and to find a topological order of the vertices of a DAG.

### *B. Social Networks*

In recent years, the large number of online Social Networks Cloud is increased, many of which have attracted hundreds of millions of users. In this session narrow down discussed some current research paper related to Social Networks Cloud.

*1) JiwonSeo, Stephen Guo and Monica S. Lam, 2013*[3]*:*

SociaLite is a high-level graph query language based on Datalog. As a logic programming language, Datalog allows many graph algorithms to be expressed succinctly. With SociaLite, users can provide high-level hints on the data layout and evaluation order; they can also define recursive aggregate functions.

*2) J.Gobinath and D.Revathi, (2013)*[1]*:*

In this, they proposed a data warehousing and analyzing system which is based on the concept of Cloud Computing. The system has also been implemented and evaluated under the proposed environment with different Cloud Computing approaches**.**

*3) Abedelaziz Mohaisen, Huy Tran, Abhishek Chandra and Yongdae Kim, (2011)*[4]*:*

In this paper they investigate a new Computing paradigm, called SocialCloud, in which computing nodes are governed by social ties driven from a bootstrapping trust possessing Social Graph. They investigate how this paradigm differs from existing computing paradigms, such as Grid Computing and the conventional Cloud Computing paradigms.

*4) Kyle Chard, Simon Caton, Omer Rana, and Kris Bubendorfer, (2010)*[5]*:*

In this research paper, they propose leveraging the pre-established trust formed through friend relationships within a Social Network to form a dynamic "Social Cloud", enabling friends to share resources within the context of a Social Network.

## VII. ACCORDING TO STUDY FURTHER RESEARCH REQUIREMENT IN SOCIAL NETWORKS AND GRAPHS

### *A. Social Networks*

The following section is about most crucial challenges of Social network and all important aspect according to organizations and company aspects.

I. Storage of Big Data:

Current disk technology [6] needs 4 terabytes per disk. Means, 1 Exabyte requires around twenty five thousand disk space. One Exabyte cannot be processed on single system either it will hang or will take month to process. These types of database require huge number of storage space to perform properly [15]. Similarly transferring an Exabyte would take huge amount of hours. Since today's technology requirement is to maintain storage and transferring of data for fast and fluent performance.

TABLE 1:
SOME EXAMPLE OF BIG DATA [15]

| Data Set/Domain | Description |
|---|---|
| Large Hadron (CERN) | Approx. 15 PB found in 2010 serve |
| Social Networking Sites | • 12+ TB on tweeter<br>• 15+ TB on Facebook per day. |
| Human Digital Universe | In 2011 size was 1.7 Zbytes and it increased by 7.9 Zbytes in 2015 |
| British Library where Website Crawling | approx 100 TB per domain crawl to be archived |
| Other | RFIDS, smart electric meters, 4.6 billion camera phones w/ GPS |

II. Data Management:

Data Management is the most difficult problem to address with big data [6]. This problem first came in the UK eScience initiatives where data was distributed geographically and "owned" and "managed" by multiple entities. Resolving issues was impossible with traditional database system process. Hence, they used combination of Distributed and Graph database to handle the volume of data.

There is two way for collection of data it can be from manually, and by digital. Where digital are audio and video recording, photos, texts, applications, User interface data, software, etc. This adequate data describe where, why, what, how and when data can be collected. Yet, it is so raw data form for efficient result.

JASON has defined [14] that "there is no universally accepted way to store raw data, reduced data, and the code and parameter choices that produced the data." Further, they note: "We are unaware of any robust, open source, platform independent solution to this problem." As far as technology is going this thing still remains true. In summarize, still lacking with the perfect solution for big

data management. This represents a gap between actuality and the research literature of big data which should be filled in a proper way.

III. Processing Issues:

Assume that an EB (Exabyte) can be processed in network. For this data need to be chunked into small block, for example one EB is equal to 1000 PB. Here needs number of system to distribute data equally. Then here parallel computing concept is applied to achieve robust result. But, if data size is rapidly increases, processes need more number of nodes which also increase the complexity of the system and affect the timing of process. It can take days to process. Thus, effective processing of big data requires more proper parallel processing and best algorithms to have timely result.

### *B. Social Graphs*

The current scenario of Social Graph still needed improvement in process and also faces open research challenges related to Social Graph. The suggested challenges for further research are the following areas:

- Social ranking and opinion sites
- Social Media Management
- Architectures for open and federated social network platforms
- Social Media Search
- Social graph analysis
- Business and social networking
- Mobile social networks

## VIII. CONCLUSION

This paper investigate Social Network and Graph, a novel promising research for Distributed and Parallel computing in the big data era. First, is discussion about what exactly Social Network, Graph, and Graph Search with example and the evolution history. Second, we have pointed out the challenges and issues of Social Network and Graph Search Likewise Ethical issues, large data management, High data access to computation ratio, Efficiency, Unstructured Graph many more. Finally, we have defined the most critical issues which required more attention is Performance and Storage of Data. And this is always most challenging issue for research in past, present and will also be in future too.

## REFERENCES:


[1] J.Gobinath and D.Revathi, (2013), *"Performance of Data Warehousing Based Social Network Analysis in Cloud Computing Using Crawler Algorithm"* International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 3.

[2] Dan C. Marinescu, (2012), *"Cloud Computing: Theory and Practice"* in University of Central Florida, Orlando, FL 32816.

[3] Jiwon Seo, Stephen Guo, Monica S. Lam, (2013),*"SociaLite: Datalog Extensions for Efficient Social Network Analysis"* in ICDE.

[4] Abedelaziz Mohaisen, Huy Tran, Abhishek Chandra, Yongdae Kim, (2011), *"SocialCloud: Using Social Networks for Building Distributed Computing Services"*,University of Minnesota.

[5] Kyle Chard, Simon Caton, Omer Rana, and Kris Bubendorfer, (2010*),"Social Cloud: Cloud Computing in Social Networks"*, Victoria University of Wellington.

[6] Rahm E, Do H H. D (2000),0" Data cleaning: Problems and current approaches",.IEEE Data Engineering Bulletin, 23(4): 3–13

[7] nextMEDIA, CSA, (2012), *"Social Networks Overview: Current Trends and Research Challenges"*.

[8] Feng K, Cong G, Bhowmick S S, Ma S, (2014), "In search of influential event organizers in online social networks". In: SIGMOD.

[9] Ma S, Li J, Liu X, Huai J, (2012), "Graph search: A new searching approach to the social computing era. Communications of CCF", 8(11): 26–31

[10] Stefan Edelkamp, Eric Hansen, Peter Sanders, (2009), Banff International Research Station on Mathematical Innovation and discovery, Grasta 2012 – 5th Workshop on Graph Searching, Theory and Applications.

[11] Lewis W.,Chu T., Salehi-Abari W., (2010), "Media Monitoring Using Social Networks", Social Computing (Social Com), IEEE Second International Conference on, vol., no., pp.661-668, 20-22 Aug. 2010 doi: 10.1109/SocialCom.2010.102.

[12] YipingKe, James Cheng, and Wilfred Ng, (2008), "Efficient Correlation Search from Graph Databases", IEEE Transactions on Knowledge and Data Engineering, Vol. 20, No. 12.

[13] Sen Wu, Jie Tang, and Bo Gao, (2012), "Instant Social Graph Search*", Publisher-Springer Berlin Heidelberg, 16th Pacific-Asia Conference, PAKDD 2012, Kuala Lumpur, Malaysia.

[14] JASON. (2008). "Data Analysis Challenges", The Mitre Corporation, McLean, VA, JSR-08-142

[15] Tom White, 2009 Book: MapReduce for the Cloud Hadoop –The Definitive Guide, Publications are O'reilly and Yahoo! Press. USA