iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** III **Month of publication:** March 2017

DOI: http://doi.org/10.22214/ijraset.2017.3054

Volume 5 Issue III, March 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Survey on Community and Social Networks in Big Data

Jabeen H Patel[1], Prof. Mohammed Azharuddin(Guide)[2]
[1,2]*VTU University of Belgaum*

Abstract: Today's world is interconnected through many types of links. These links include Web pages, blogs, and emails. This consider community mining and the mining of social networks as important topics. Big data has become an important issue for a large number of research areas such as data mining, machine learning, computational intelligence, information fusion, the semantic Web, and social networks. In big data mining the main Information and Communication Technology (ICT) has a great impact on social well-being, economic growth and national security in today's world. The Community structures are important properties of social networks and in this paper we are mainly focusing on such properties. The survey also considers the network setting with commonly implemented factors.
Index Terms: Notion Of Community, Involvement Of Entity towords Network, Nature And Frequency Of Local Interactions, Network's Static Structures Behavior

I. INTRODUCTION

The Big data plays vital role in social networks through Web pages, blogs, and emails. The analysis of social networks has recently experienced a surge of interest by researchers, due to different factors, such as the popularity of online social networks (OSNs), their representation[1] and analysis as graphs, the availability of large volumes of OSN log data, and commercial/marketing interests. An interesting property to investigate, typical to many networks, is the community structure, i.e. the division of networks into groups (also called clusters) having dense intra-connections, and sparse inter-connections [4][11].

The recent proliferation of online social networks such as MySpace, Facebook, Twitter, and so on has attracted attention of computer scientists, as well [2][12]. In 2003, another form of online community acquired stunning popularity:"online social networking services". In addition to descriptive personal profiles, members of such communities publicly articulate mutual "friendship" links[3] with other members, creating a browseable network of social relations. The identification problem of social networking in itself is a challenging one and we are mainly focusing on those challenges. Those challenges are[7] First, it's critical to have the right characterization of the notion of "community" that is to be detected. Second, the entities/nodes involved are distributed in real-life applications, and hence distributed means of identification will be desired. Third, a snapshot-based dataset may not be able to capture the real picture; what is most important lies in the local relationships (e.g. the nature and frequency of local interactions) between the entities/nodes. Under these circumstances, our challenge is to understand (1) the network's static structures (e.g. topologies and clusters) and (2) dynamic behavior (such as growth factors, robustness, and functional efficiency). Email exchanges within an organization or in one's own mailbox over a long period of time can be mined to show how various networks of common practice or friendship[7] start to emerge. In this paper we did survey on all the challenges related to the social network which plays vital role in communication of online social networks to provide more effective and efficient result. We study this challenges one by one to understand the behavior of big data.

A. Notion Of Community

Our notion of (anti)social behavior in the source-based social networks, and the reliance of this notion on 'communities' in those networks, suggests the relevance of so-called community detection methods from the social network[8] literature in developing detection algorithms. There are many such methods of this sort. All of them essentially take as input a graph topology and they output a partitioning of the graph into subsets of nodes, for which nodes within subsets are more heavily linked than nodes between subsets. A popular method in this area is that of Newman, Clauset and Moore, based on the concept of modularity.

Modularity: During the elicitation process, the graph is successively divided in components,[1] and the correctness of the community partitions is measured. The quality metric used for a given community is called the modularity. For a graph divided into

Volume 5 Issue III, March 2017
ISSN: 2321-9653

k communities, a symmetrical matrix e is defined of order k2 whose elements eij are the subset of edges from the total graph which connect the nodes of communities i and j. High values of modularity correspond to community assignments with greater numbers of intra-community links than expected at random (with respect to a particular null model [9]. Although numerous other community detection methods are also available, modularity optimization is perhaps the most popular way to detect communities and it has been successfully applied to many applications [9]. One might also consider using a method that includes a resolution parameter to avoid issues with resolution limits. However, our primary focus is on global organization of the networks, so we limit our attention to the default resolution of modularity. This focus arguably biases our study of communities to the largest structures, such as those influenced by common class year, making the observed correlations with other demographic characteristics even more striking.

To try to ensure that the communities we detect are properties of the data rather than of the algorithms that we used, we optimize modularity (with default resolution) using 6 different combinations of spectral optimization, greedy optimization, and Kernighan and Lin [1970] (KL) node-swapping steps (in the manner discussed by Newman [2006b]). Specifically, we use (1) recursive partitioning by the leading eigenvector of a modularity matrix [Newman, 2006a], (2) recursive partitioning by the leading pair of eigenvectors (including the Richardson et al. [2009] extension of the method in Newman [2006a]), (3) the Louvain greedy method [Blondel et al., 2008], and each of these three supplemented with small increases in the quality Q that can be obtained using KL node swaps. Each of these 6 methods yields a community partition, and we obtain our comparisons by considering each of these 6 partitions.

Involvement Of Entity Towords Network

Entering a community[8] is an important issue which shows that to which one does not belong – requires us to identify when one ‘belongs’ (and when one does not ‘belong’) to a ‘community.’ Both of these terms are relational in nature, and so it is natural to employ network-based representations of our data for our work. Generally speaking, any choice of network representation needs to be made carefully and can be expected to influence the level of success one achieves with high-level tasks utilizing such networks[5]. For that reason, one of the goals of our study is to examine certain alternatives in choice of network representation and understand the implications of this choice on our ability to detect intrusions.

In this section we will consider two specific algorithms for automatic community extraction from a complete graph: that of Newman and Girvan , and that of Blondel et al[1]. Newman and Girvan’s algorithm was effective but slow, whereas that of Blondel et al., designated as the Louvain Method, was developed four years later and is much more efficient in computational terms, having now become an ‘industry standard’. In this section we will briefly describe both algorithms, and discuss some of the results of community extraction for two benchmark datasets.

Newman and Girvan’s algorithm focuses on how to extract a community structure from social network graph data. Two main approaches are defined: (i) the identification of groups around a prototypic nucleus defined in terms of the ‘most central’ edges, an adjacency matrix being used as the basis to calculate the weights; and (ii) the identification.

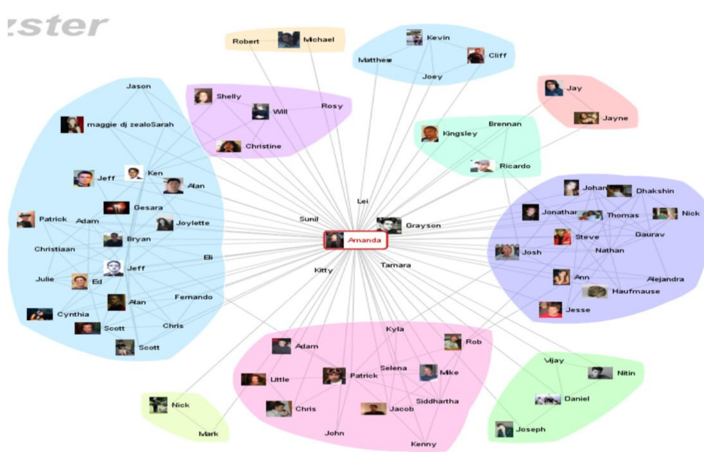

Figure 1: Vizster presentation of network

B. Nature and Frequency of Local Interactions

In social network applications, each user is typically defined by a profile, together with a functionality which facilitates searching for and aggregating contacts in a contact list. For each contact to be established, both parties have to mutually accept to create the

Volume 5 Issue III, March 2017
ISSN: 2321-9653

'link'. Other functionality is provided such as a 'chat', 'photo albums' and a 'wall' in which the user can publish messages and content which are 'broadcast' to the contact list. Online applications, such as games, allow the user to participate, compete and collaborate with other users. An online social network can be generically[1] understood to be some kind of computer application which facilitates the creation or definition of social relations among people based on acquaintance, general interests, activities, professional interests, family and associative relations, and so on.

To better navigate the myriad design decisions we faced, we turned to an ethnographic study of the Friendster service. Friendster was designed to be an online dating site, complete with profiles, demographic and interest driven search, and a private messaging system[3]. What made Friendster unique was its articulated social networking component and testimonial feature. Users were asked to declare "friends" on the system whose pictures would also appear on the profile when the friends confirmed the relationship. Friends could write testimonials that would also appear on the profile. Both the friends and testimonials were intended to signal additional information about the person's character for those interested in dating the person.

Vizster presents social networks using a familiar node-link representation, where nodes represent members of the system and links represent the articulated "friendship" links between them .

In Figure 1 Vizster presentation,network members are presented using both their self-provided name and, if available, a representative photograph or image. The networks are presented as egocentric networks: networks consisting of an individual and their immediate friends. Users can expand the display by selecting nodes to make visible others' immediate friends as well. To the right of the network display is a panel presenting a person's profile. As discussed later, the profile panel also provides direct manipulation searches over profile text.

Network's Static Structures Behavior

In this section we consider three interrelated aspects: modelling (simulation) of OSN graphs, how they evolve over time, and their structure.

Aspects such as clustering, 'characteristic path length' and 'connectedness' are also mentioned, as well as 'exponential random graph models' and their simulations, which are studied in some detail. A probabilistic formula is given which relates a random graph to an observed graph, in terms of the links defined in the corresponding adjacency matrices. However, in social networks the assumption of independent ties is stated as being generally implausible. In order to model an OSN graph, we have to understand what are its basic building blocks and characteristics.

Clustering refers to unsupervised learning algorithms which do not require pre-labeled data to extract rules for grouping similar data instances. Although there are different types of clustering techniques. The difference between regular clustering and co-clustering is the processing of rows and columns. Regular clustering techniques such as k-means[10] clusters the data considering the rows of the data set where as the co-clustering considers both rows and columns of the data set simultaneously to produce clusters.

C. Future Work

Compared to other surveys, this paper provide a discussion on the challenges related to the social networking in big data this can also be improved by implementing the algorithms related to the social network structure such as re-identification algorithm and by implementing efficient techniques in the online social networks to achieve the efficient and effective results in big data networking.

II. CONCLUSION

This paper analysis challenges of representing and measuring the social network and community in big data, and shows the social behavior of network in big data. The survey also includes the important and effective challenges of social network and study of the social structure of Friendster service with "friendship" networks. The case study represented a design of Vizster, a visualization system for end-user exploration of online social networks. The survey include how the social network can be improved agains the challenges come across it.

REFERENCES

[1] David F. Nettleton,Universitat Pompeu Fabra, Barcelona, Spain Iiia-Csic, Bellaterra, Spain "Data Mining Of Social Networks Represented As Graphs" Elsevier, 20 December 2012.

[2] Narayanan and Vitaly Shmatikov The University of Texas at Austin "De-anonymizing Social Networks Arvind", ieee, arXiv:0903.3276v1 [cs.CR] 19 Mar 2009.

[3] Jeffrey Heer Computer Science Division University of California, Berkeley danah boyd School of Information Management and Systems University of California, Berkeley "Vizster: Visualizing Online Social Networks" 2009.

[4] Clara Pizzuti ICAR-CNR Via Pietro Bucci, 41C 87036 Rende (CS), Italy pizzuti@icar.cnr.it, "Community Detection in Social Networks with Genetic

Algorithms", July 12–16, 2008

[5] Eric D. Kolaczyk. Statistical Analysis of Network Data: Methods and Models. Springer, New York, 2009.

[6] Maitrayee Mukherjee Department of Computer Science and Engineering University of Texas at Arlington Box 19015, Arlington, TX 76019 mukherje@cse.uta.edu Lawrence B. Holder Department of Computer Science and Engineering University of Texas at Arlington Box 19015, Arlington, TX 76019 holder@cse.uta.edu, "Graph-based Data Mining on Social Networks", August 22–25, 2004.

[7] QIANG YANG Department of Computer Science Hong Kong University of Science and Technology Clearwater Bay, Kowloon, Hong Kong, China XINDONG WU Department of Computer Science University of Vermont 33 Colchester Avenue, Burlington, Vermont 05405, USA xwu@cs.uvm.edu, "10 Challenging Problems In Data Mining Research" International Journal of Information Technology & Decision Making Vol. 5, No. 4 (2006) 597–604 c_ World Scientific Publishing Company.

[8] Qi Ding Boston University qiding@math.bu.edu Natallia Katenka Boston University nkatenka@math.bu.edu Paul Barford University of Wisconsin pb@cs.wisc.edu Eric Kolaczyk Boston University kolaczyk@math.bu.edu Mark Crovella Boston Univeristy crovella@cs.bu.edu, "Intrusion as (Anti)social Communication: Characterization and Detection", August 12–16, 2012.

[9] Amanda L. Traud1,2, Peter J. Mucha1,3, and Mason A. Porter4,5 1Carolina Center for Interdisciplinary Applied Mathematics, Department of Mathematics, "Social Structure of Facebook Networks", arXiv:1102.2166v1 [cs.SI] 10 Feb 2011.

[10] A survey of network anomaly detection techniques Mohiuddin Ahmed,Abdun Naser Mahmood,Jiankun Hu School of Engineering and Information Technology,UNSW Canberra,ACT2600,Australia 19 November 2015.

[11] S. Lozano, J. Duch, and A. Arenas. Analysis of large social datasets by community detection. European Physical Journal ST, 143:257–259, 2007

[12] G. Kossinets, J. Kleinberg, and D. Watts. The structure of information pathways in a social communication network. In KDD, 2008.