SOCIAL NETWORK ANALYSIS: TRENDS, TECHNIQUES AND FUTURE PROSPECTS

Aanchal gupta, Amit chhabra
student, assistant professor
Guru Nanak Dev University

Gurdaspur, India

Abstract. A social network is a set of people or organizations or other social entities connected by set of social relationships such as friendship, co-working or information exchange. Online social networking sites have become popular. The most powerful combination of methods and techniques makes a special breakdown of social networks. Social network analysis focuses on the pattern of relationships among people, organizations, states and social entities. This article discusses the challenges and opportunities arising from the evolution of social networks and provides the structure for a better understanding of the rapid growth of social networks. This paper reviews the state of play in the selected aspects of social networking and a survey of the works done in the field of social network analysis is done. This presents an open set of challenges in the field of social networking. Problem shows that more research is needed in social networks.

Keywords: Social Media, Social Network Analysis, Social Network Models, Community Detection.

1. INTRODUCTION

The growth and development of the Internet and the global network has provided a global network for exchanges and cooperation in trust in this relationship.
A social network is a social structure made up of individuals called nodes, which are tied by one or more specific type of inter-dependencies　　　　　　　　　　　　such as friendship
interest, dislikes, beliefs. With ease of accessibility, they are spread throughout our lives to destinations where users can access or share information anywhere.
The Internet is no more than the BBS, which allows users to share software, data, messaging and news together. In the late nineties, last century saw a surge in popularity of websites where general Internet users could share their own personal life: Today's equivalent of a Weblog or a blog. Sessions of the company's website and e-commerce begins with starting a new relationship, Amazon and eBay in 1995 and six years later was taktilizirane properly by the dot-com bubble in 2001. The social
network perspective encompasses theories, models, and applications that are expressed in terms of relational concepts or processes
The current trend towards social media can be seen as an evolution back to the roots of the Internet as it transforms the worldwide network of the UN which was created as a means to facilitate sharing between users.

This report outlines the challenges posed by the development of social networking. We start with definitions and social networking ratings. A review of the social network analysis, focusing on the techniques used in the social network analysis, was further discussed. We are finalizing the report on future challenges / areas of research.

2. HISTORY OF SOCIAL NETWORK ANALYSIS

Research in network analysis is found in educational psychology, and studies of child development. Network analysis also developed in fields such as sociology and anthropology. The third parties could affect the relationship between two individuals and examined how organizational structures were needed to coordinate interactions in large groups.

3. SOCIAL MEDIA AND NETWORKING OVERVIEW

In 1979, Tom Truscot and Jim Ellis of Duke University created Usenet to create a global forum for Internet users to publish public messages. Twenty years later, Bruce and Suleene Abel founded the Open Platform, a social networking site in the first logbook[1]. The word "weblog" was created at the same time and was called "block" a year later. Increasing the speed of access to the internet has contributed to the popularity of ideas that led to the creation of social networking sites such as MySpace (2003) and Facebook (2004). This is the turn of the word "social media".
Social networking sites allow people to maintain their relationships, engage with acquaintances, and create new relationships with others based on shared features such as community, interests and interests. The following section provides an overview of various types of social media.

3.1 TYPES OF SOCIAL MEDIA

Social networks can be classified on the dimensions of self-presentation/ self-disclosure and social presence media richness. It can be seen that with respect to social presence and media richness, blogs and collaborative projects score the lowest as they present simple text interfaces leading to simple exchange, whereas content communities like YouTube and social networking sites like Facebook, in addition to text based communication also allow for posting of videos, photos. On the highest level are the virtual games and virtual social worlds which allow face to face interaction in 3-D space leading to sophisticated user interaction. Similarly, moving on the dimension of self presentation, blogs score better than collaborative projects as the latter focuses on content specific domains.
On the same lines social networking sites, virtual social worlds allow for more self disclosure.
Each of the social media can be briefly explained as follows:-
Content Communities:-The main objective of content communities is the sharing of media content

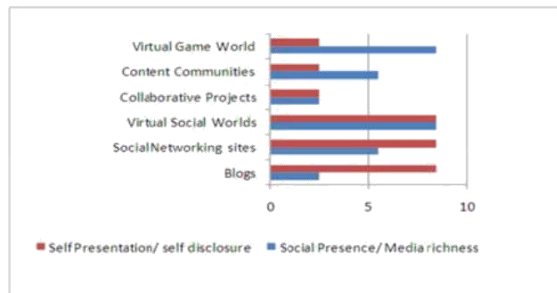between users. Content communities exist for a wide range of

F ig. 1. Types of Social Media

different media types, including text (e.g. Book crossing), photos (e.g. Flickr) ,videos (e. g. YouTube), power point presentations(e.g.SlideShare)

Blog:- Blog is a website or a webpage which is easy-to-update where authors write regular entries in a diary-like format. The most effective and interesting blogs allow readers to engage in conversations with the author and other readers.(e.g. word press, twitter).

Social Networking Sites:- Social networking sites are websites that connect people. In these online communities, people can join for free and can establish a page with their profile. The most popular sites, such as Facebook and MySpace, also have groups, which helps to communicate with other members of the groups by chatting or calling. Members can share comments, links, photos, videos, and more.

Collaborative Projects:-They enable the joint and simultaneous creation of content by many end-users. Wikis are a type of collaborative application which allows users to add, remove, and change text-based content.

Virtual Game worlds:-Virtual worlds are platforms that replicate a three dimensional environment in which users can interact with each other as in real life. They provide the highest level of social presence and media richness of all the applications and are thus the ultimate manifestation of social media.

Virtual social worlds:-A virtual world is an interactive simulated environment that allows multiple users to participate simultaneously via an online interface and has the following characteristics[5]. They provide shared space to many users for participating simultaneously. Interaction takes place predominately in real time and allows users to build, develop and submit customized content. (e.g. Disney's Toon town).

The dramatic growth in social multimedia and user generated content has revolutionized all phases of content value chain and has given rise to a new rapidly evolving research field. Social multimedia computing in which all well established computing and multimedia technologies are brought together with emerging social media research.

Types of Social Media	Nature/Description
Content Communities	Main is to share the media content between users for wide range including text and videos.
Blog	This is an easy-to-update website or webpage where they write regular entries in diary like format. Eg-twitter
Social Networking Sites	This connects people by online sites like facebook where people can join and communicate witn one another by establishing an intro in their profile page.
Collaborative Projects	In this,there is a creation of content by many end users as it allows users to add,remove and change text-based content.
Virtual Game Worlds	This shares a platform that replicate a 3-D environment in which users can interact with each other as in real life.
Virtual Social Worlds	This is an interactive environment that allows multiple users to participate via online.

(a) Table :explaining types of social media and its nature.

4. SOCIAL NETWORK ANALYSIS (SNA)

Social Network Analysis (SNA) is a set of powerful techniques to identify specific roles, important groups and hidden structures in organizations and groups. SNA has a long and successful history within sociology, networks are everywhere and related methodology that can be used to analyze a wide variety of problems.

A Social Network is modeled by a graph, where the nodes represent individuals, and an edge between nodes indicates a direct relationship between individuals.

An edge in a social network refers to a tie which can be directed (i.e. potentially one directional, as in giving advice to someone) or undirected (if both the entities are equally involved) or weighted (measured on a scale, as in strength of friendship).

Networks which are only made of one type of nodes are homogenous whereas the networks which are made up of different types of nodes are called heterogeneous Networks . Social Network Analysis (SNA) is the mapping and measuring of relationships and flows between people, groups, organizations or other information/knowledge processing objects[4]. In a typical social network, there always exist various relationships between individuals, such as friendships,

business relationships, and Common interest relationships. Social network Analysis aims at understanding the network structure by description, visualization and modeling .

In the analysis of complete networks, a distinction can be made between- (a)Descriptive methods involving network statistics and network visualization (b)analysis procedures, often based on decomposition of the adjacency matrix (c)statistical models based on probability distributions.

4.1 Social Networks Analysis: Techniques

Social Network Analysis can be termed as an application of Web Mining techniques with the focus on web content mining, web structure mining and web usage mining[7]. Mining techniques can further be considered as an application of data mining techniques such as graph mining, rule mining, link mining, clustering and classification.

Social networks are said to have evolved in sixties, with the history deep rooted in social sciences, when Milgram and Travers[8] proposed "The Small World Problem". Social Networks are characterized by three prominent key features i.e.-
 (i) "Small World Effect"
(ii) "Scale Free Distributions"
(iii)"Emergence of a community structure"[13].
"Small World Effect" puts into evidence that there exists a relatively short path connecting any pair of nodes within the network[12]. Another important feature of Social Networks is the degree distribution of nodes within the network which specifies how the nodes within a network are interconnected. In particular, many large scale networks like World Wide Web[9], Internet[10] follows power law and are free from any characteristic scale and hence are called scale free networks[11].
Another important feature is the Emergence of community structure where some nodes within the network are densely connected with respect to the rest of the nodes.
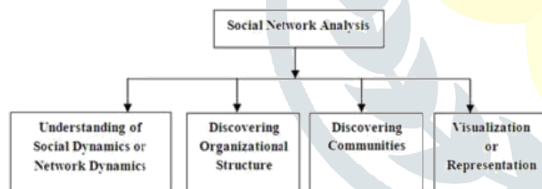

Fig. 2. Realm of Social Network Analysis

As per the above mentioned features, many models have evolved so as to model Social Networks[12].

Realm of Social networks analysis can be classified as per Figure 2. The task of understanding how a network is formed with actors and their behavior is called *Social Dynamics.*

Social dynamics can refer to the behavior of groups that result from the interactions of individuals, group members as well to the study of the relationship between individual interactions and group level behaviors[7]. While studying the dynamics of social networks, there are two types of dynamics that can be defined[17] i.e. **dynamics of the network** which involves the study of the evolving or changing structure of the network comprising of making or breaking of ties and the **dynamics on the network** which are a result of the action of nodes on the basis of influence of the neighboring nodes. .If the aim is to understand the organization of a social network with the identification code of key nodes and their impact on the overall network, the area known as *covering or understanding of Organizational structure* emerges. One might be interested in discovering the groups that emerge due to specific response / activity patterns among users in a network. This gives rise to an active and interesting area of research which is called "*Community Identification*".

Several studies have been conducted in order to investigate the community structure of real and Online Social Network[14][18][19]. Many aspects of social networks are best understood by the process of "*Visualization*" which presents a graphical view of the network. Many tools are available today for visualization.

The research on community detection algorithms which is a problem of considerable practical interest and has received great deal of attention[14][13][17] and will be a subject of interest in this paper.

The aim of a community detection algorithm is, to divide vertices of a network into groups, while maximizing the number of edges between the groups and minimizing the edges established between vertices in different groups where each group represents the desired community. Review of many of the community detection techniques are presented. The aim of detecting communities should be governed by the user's interest i.e. what kind of communities a person is interested[17]. Keeping this in view the search for community can be described as follows:-
Density based: The community can be described as a group in which there are many edges between groups **t**han the edges connecting different groups**.**

Vertex Similarity based: The communities may be assumed by comprising of group of vertices that are similar to each other[17]. The similarity between each pair of vertices can be computed with respect to some reference property.

Action based: The approach centers around the concept of grouping entities by the set of actions they perform.

Influence propagation based: According to this approach, the users that frequently perform the same actions due to the influence of a particular user are considered as part of same community.

4.2 The **community detection** algorithms can be classified as:-

1.Hierarchical Clustering algorithms:– These algorithms reveal the multilevel / hierarchical structure of graph. The definition of a similarity measure between vertices forms the basis of this method. After a measure is chosen, the algorithm computes similarity between each pair of vertices which results in a similarity matrix for the whole graph. After the creation of matrix, similar vertices are chosen and merged to form a cluster. These algorithms can be further classified into following categories:-

(a) *Agglomerative algorithms* are those algorithms in which each node of the graph forms a cluster and clusters are iteratively merged if their similarity is high. These type of algorithms follow bottom up approach as they start with the leaf nodes of the graph and form clusters till they reach the root of the graph.

(b) *Divisive algorithms* follow top down approach in which clusters are iteratively split on the basis o f low similarity between vertices. Famous algorithm of Newman-Girvan[16] falls under this category.

Hierarchical clustering has the advantage that it does not require any prior knowledge about the number and size of clusters present in graph. However, it also has a disadvantage that the algorithm always yields a hierarchical structure which may not be present in actual graph. Sometimes, some vertices are also mis-classified by the algorithm.

2. Partitional Clustering algorithms: – These algorithms tends to find a predefined set of clusters say 'k' in graph or set of data points. The goal is to separate in 'k' clusters so as to maximize or minimize a given cost function . Minimizing the diameter of the cluster, minimizing the distance between the centroid of the cluster and each node in cluster are some of the most common cost functions. K-means clustering[28] is the most common partitional algorithm which tries to minimize the total intra-cluster distance. Limitation of partitional clustering is that the number of partitions must be specified at the beginning which is not possible in most of the cases.

3. Modularity based algorithms:– The methods in this category tend to optimize 'modularity' which is a quality function and a higher value of modularity indicates good quality of cluster. It has been proved that modularity optimization is an N-P complete problem[14], so it is probably impossible to achieve a very high value of modularity. Practically all the algorithms in this category tend to achieve the maximum possible optimization of modularity. Modularity can be easily extended to weighted and directed graphs. The methods in this category can be broadly classified as:-

(a) *Greedy algorithms*:- These algorithms start with a single node of a graph which represents a cluster as it maximizes the value of modularity. Algorithms of Newman-Girvan[15], Clauset, fall under this category. The greedy optimization of modularity tends to form quickly large communities at the expenses of small ones.

(b) *Simulated annealing* :- This is a probabilistic method which tries to form new clusters from old ones by either joining vertices from neighboring clusters or splitting existing cluster after computing the probability that the process of 'merge' or 'split' would result in increase in modularity.

(c) *External optimization:-* It is a heuristic search procedure proposed by Boettcher and Percus[24], and tries to achieve higher values of modularity by successively replacing a single vertex of a cluster which has high connectivity with vertices of other cluster with a randomly chosen vertex of neighboring cluster. In other words, vertices with low value of modularity are replaced. The method is more accurate than simulated annealing but takes more time.

4.Spectral Clustering: - This technique includes all methods and techniques that partition the set into clusters by using eigen vectors of matrices. The algorithms transform the initial set of vertices into a set of points in space, whose coordinates are elements of eigen vectors. These points are then classified using standard clustering techniques like k-means clustering. The most important reason for such a transformation is that the change in representation makes the clustering properties of the data elements more evident. Hence, when the graph is very large, such process reveals clear clusters.

It has been observed by Fortunato[14] that if a clear cluster structure is present in graph, it can easily be revealed by the quality function 'modularity'. However, it has also been observed that modularity also has a resolution limit which may prevent the detection of small clusters in a graph.

The problem of community detection can be termed as an optimization problem whose precise formulations are known to be computationally intractable. With the increasing penetration of social networks in our lives, and with the increasing complexity of online social networks, there is a great need of fast algorithms for detection of communities in large networks[18]. Over the time, many algorithms have been presented to detect community in large scale networks by utilization of minimum resources[19]. Many works have been presented in static communities detection, however, detection of communities in dynamic network pose challenges of developing cost, time and memory effective solutions that need to identify changes in community structure over time.

Table(b):   **COMMUNITY**   **DETECTION ALGORITHMS**

Partitional Clustering algorithm	This algo tends to find a predefined set of clusters in a graph and a goal is to separate k clusters to minimize or maximize a given cost function.
Hierarchical Clustering algorithm	This reveals the multilevel structure of a graph. A tree like structure with nodes, vertices. Further classified as:- (a)Agglomerative algorithm- Each node of a graph forms a cluster. It's a type of bottom-up approach. (b)Divisive algorithm- It's a type of top-down approach.
Modularity Clustering algorithm	This tends to optimize the modularity which is a quality function and has a higher value which indicates the good quality of cluster. Further classified as:- (a)Greedy algorithm-Starts with a single node of a graph which represents a cluster. (b)Simulated algorithm-A probalistic method used to form a new cluster from the old ones. (c)External algorithm-This a heuristic search which tries to achieve higher values of modularity.
Spectral Clustering	This algo includes all techniques that partition the set into clusters by eigen vectors of matrices.

5. SOCIAL NETWORK ANALYSIS METRICS

- **CLOSENESS:-** An individual is near to all other individuals in a network (directly or indirectly). It reflects the ability to access information through the "grapevine" of network members. Thus, closeness is the inverse of sum of the shortest distances between each individual and every other person in the network.
- **BETWEENNESS:-** The extent to which a node lies between other nodes in the network. The measure reflects the number of people which are connected indirectly through their direct links.
- **NETWORK CLOSURE:-** A measure of the completeness of relational triads. An individual's assumption of network closure is called transitivity.

- **CENTRALITY:-** Centrality refers to a group of metrics that aim to quantify the "importance" or "influence" of a particular node within a network.
- **CLUSTERING COEFFICIENT:-** A measure of the likelihood that two associates of a node are associated. A higher clustering coefficient indicates a greater 'cliquishness'.

6. CHALLENGES IN SOCIAL NETWORK ANALYSIS

Social Networks are changing and evolving in the real world. New members join the network , some existing members may quit the network or change their associations and collaborate with new communities or groups. Thus social networks have an element of dynamism associated with them. As the penetration of social network is increasing in our lives, so is the interest of researchers in this field as many interesting areas of research have emerged. However, there are certain challenges associated with this area, some of which are listed in below:-

Availability of live/ Dynamic data has been a desire for researchers. Indeed, dynamic data are very difficult to collect and analyze. Many researchers have been sharing their datasets on the web, however, these datasets are static and cater to specific problem and the availability of live data from many social networking sites like facebook poses certain restrictions.

Community discovery is an important problem in social network analysis (SNA), where the goal is to identify related groups of members such that intra-community associations are denser than inter-communities associations [14]. Many researchers have been actively working in the area of dynamic community detection [19][20]. One of the new areas that has been of interest is topic based community discovery [21] which aims to compute community structure by considering the user involvement in a particular topic on community.

Spam detection and advertisement detection are research challenges that need extra attention from the research community. Since users and data production increase, spam (irrelevant information) and advertisements will continue growing [31].

Security by means of Social Networks Analysis The information is extracted from Social Networks analysis has proved to be a useful tool towards security. One example of an application related to security is the Analysis of terrorism, as for instance, the Analysis of the 9-11 [22]. This study was done by gathering public information from major newspapers on the WWW and analyzing it by means of Social networks. Further, many researchers have emphasized the use of techniques for social network analysis in discovering hidden groups [29], identifying key nodes in terrorist networks [30],detecting cyber crimes [23]. Recent trends have shown that social networking sites are used by extremist groups for spreading propaganda, hatred, recruitments etc. [27] . So, there is a growing need to monitor/track/detect extremist groups active on various social networking sites. An important application of social network analysis techniques combined with the use of text mining approach can be used to detect communities of extremist groups engaged in serious conversations, propagation of their ideas. Thus we can say topic based cyber crime detection in real world is another area of concern. However, the problem requires a dynamic and less complex solution as the active

groups may not tend to operate in large numbers in a social networking site.

CrimeNet Explore [29] is a system for Criminal Network knowledge discovery that incorporates several techniques including the concept of space approach, hierarchical clustering, social network analysis and multidimensional scaling. However, the systems does not caters to incorporate dynamic characteristics of criminal networks. Hence, Dynamic Analysis of Criminal Networks is again an ongoing research area.

The increasing use of social networks by masses have given way to cyber criminals known as hackers/ hacking groups which use these sites for spreading botnets. Thus an important aspect of security is **detection of botnets in online social networks.**

J.Qiu and Z. Lin [20] in their research efforts have described the use of community tree data structure to represent organizational structure and presented a framework for exploring organizational structure in dynamic social networks. However, the development of **time effective and optimized solution for dynamic social network analysis** is the need of the day and thus emerges as a challenge.

There is vast amount of unstructured data available over the internet which if linked with open source data and social network data, can present a new fold research outlook resulting in development of brilliant mining techniques / framework which can be further incorporated to develop a socio-semantic analyser to draw out effective hidden information. This finds its linkages and applications to the semantic web.

7. CONCLUSION

Real World Social networks and the online communities have evolved as an exceptionally useful collaborative and communication tools for millions of users. As the penetration of the social networks in our lives increase, we may see abuse as well as the benefits of the technology in future. In this paper we have tried to present an overview of social networks with an overview of the techniques used in the analysis of social networks. We have also tried to discuss the challenges and the emerging research areas associated with the analysis of social networks.

REFERENCES

[1] Kaplan, Andreas M., Haenlein Michael.: Users of the world unite! The challenges and opportunities of Social Media. Business Horizons 53, 59-68 (2010), www.elsevier.com/locate/bushor

[2] Wasserman, S., Faust K.: Social Network Analysis: methods and applications (1994). Cambridge University Press, Cambridge.

[3] Newman M., Barabasi A., Watts D.J (eds): The structure and dynamics of networks. (2006) Princeton University Press, Princeton.

[4] Scott J.: Social Network Analysis, A handbook. Sage, 2000.

[5] Papp Raymond : Virtual worlds and social networking: reaching the millennials (2009) , Journal of Technology Research.

[6] Clauset A., Newman, M. and Moore, C.: Finding community structure in very large networks (2004) , Physical Review E, 70:066111 .

[7] Kosala R., Blockeel H.: Web Mining Research: A Survey, (2000), SIGKDD Explorations, Volume 2, Issue 1.

[8] Travers J., Milgram S.: An experimental study small world problem, (1969), Sociometry, vol. 32, no. 4, pp. 425-443, 1969.

[9] Albert R., Jeong H., Barabasi A.: Diameter of the World Wide Web (1999), Nature, vol. 401, no. 6749, p. 130.

[10] Dunbar, R: Grooming, gossip, and the evolution of language (1998). Harvard University Press.

[11] Barabasi A., Albert R.: Emergence of scaling in random networks (1999), Science, vol. 286, no. 5439, p.509.

[12] Watts : Six Degrees: The Science of a Connected Age (2003). Norton, New York.

[13] Karrer B., Levina E., Newman M.: Robustness of community structure in networks (2008), Phy. Rev. E, vol. 77, no. 4, p. 46119.

[14] FortunatoS.: Community detection in graphs (2010), Physics Reports, vol. 486, pp. 75-174.

[15] ClausetA., Newman M., Moore C.: Finding community structure in very large networks (2004), Physical Review E, 70:066111.

[16] Girvan M. , Newman M.: Community structure in social and biological networks (2002), Proceedings of the National Academy of Sciences of the United States of America, 99 (12):7821-7826.

[17] CosciaM., Giannotti F., Pedreschi D.:A classification for Community Discovery Methods in Complex Networks (2011) , Wiley online library DOI:10.1002/sam.10133.

[18] Peters S., Jacob Y., Denover L., Gallinari P.: Iterative Multi-label Multirelational Classification Algorithm for complex social networks (2011). Soc. Netw. Anal. Min. DOI 10.1007/s13278-011-0034-8 .

[19] NguyenN.P, Dinh T.N, Xuan Y., Thai M.: Adaptive Algorithms for Detecting Community Structure in Dynamic Social Networks (2011), IEEE Infocomm, 2011 .

[20] QiuJ., Lin Z.: Framework for exploring organizational structure in dynamic social networks (2011). Decision support systems, doi:10.1016/j.dss.2011.01.011.

[21] Branting K.: Context-sensitive detection of local community structure (2011). Soc. Netw. Anal. Min. DOI 10.1007/s13278-011-0035-7

[22] Valdis K.: Uncloaking Terrorist Networks (2002). First Monday 7, no. 4 (2002). http://www.firstmonday.org/ issues74/krebs.

[23] Chen H., Chung W., Xu J., Qin G.W.Y, Chau M.: Crime Data Mining: A General

Framework and Some Examples (2004). IEEE Computer Society.

[24]Boettcher S., Percus A.G.: Extremal Optimisation, an evolutionary local-search algorithm (2002), http://arxiv.org/pdf/cs.NE/0209030.pdf

[25]KirkpatrickS., Gelatt C.D., Vecchi M.P.: Optimisation by Simulated Annealing (1983), Science, Volume 220, Number 4598, 13 May 1983.

[26]Luxburg V.: A Tutorial on Spectral Clustering (2007). Statistics and Computing, 17 (4), 2007. Springer.

[27]Gerstenfeld P.B, Grant D.R., Chau-Pu C.: Hate Online : A Content Analysis of Extremist Internet Sites (2003). Anal. of Social Issues and Policy, Vol 3, No. 1, 2003, pp. 29-44.

[28]Kanungo T.,Mount D., Netanyahu N.S., Piatko C.D., Silverman R., Wu A.Y.: An Efficient k-Means Clustering Algorithm:Analysis and Implementation 2002,IEEE Trans. On Pattern Ana. and Mach. Intell. Vol.24, July 2002.

[29]Xu J., Chen H.: CrimeNet Explorer: A Framework for Criminal Network Knowledge Discovery (2005). ACM Transactions on Information Systems, Vol. 23, No. 2, Pages 201-226.

[30]Shaikh M.A., Jiaxin W.: Investigative Data Mining: Identifying Key Nodes in Terrorist Networks (2006). IEEE.

[31]Luo W., Liu J., Liu Jing, Fan C. : An Analysis of Security in Social Networks (2009). Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing.