Asian Journal of Computer and Information Systems (ISSN: 2321 – 5658)
Volume 05– Issue 02, June 2017

Framework for Enhancement Performance of Heterogeneous System via Social Network Analysis

Abdulkareeem Merhej Radhi

Information Engineering College-Al-Nahrain University
Baghdad, Iraq
Email : kareem_m_radhi [AT] yahoo.com

ABSTRACT--- ***In recent decades, social network analysis arise as a powerful tool for describing social competitive systems. Vertices, or entities represent states or nodes or agents, while edges represent relations between different nodes. A graph theory is a suitable way for simulating complicated systems. This paper presents framework and architecture to improve system performance using social network analysis. System crawling then web blog analysis is a methodology to clustering and exploring community nodes. Graph and semantic network used for simulate agents and edges. The proposed architecture involves two consecutive modules affecting system performance. Clustering is a first module which simulate vertices and it's edges with user supervised for filtering nodes and attribute , while supervised and machine learning module stimulate the agents for optimal path and performance under complicated environment circumstances. Data prepared from social networks, relations extracted, and crawling with breadth first search with some small graphs as a base are maintained. The proposed framework affect the performance of learning in university as a sample of community through analyzing and maintaining relations and features enhancement which are achieved via supervised machine learning. The proposed system can be implemented to a various social institutes or organizations, using its webs or blogs with different queries from interviews and emails to enhance its performance.***

Keywords--- Vertex, edges, Semantic network, clustering, blog, attribute, machine Learning

1. INTRODUCTION

Today the Internet has become a convenient and ubiquitous platform for anyone with access to publish their thoughts and ideas, express their opinions, argue with their peers on various issues and, most importantly, organize and form online communities. Although video and image formats, and 3D environments are rapidly gaining popularity, the majority of user generated discussions on the Internet are still text-based. Many of these discussions are archived and readily available to organizers, developers and researchers of online communities [1].

The basic idea of a social network is very simple. A social network is a set of actors (or points, or nodes, or agents) that may have relationships (or edges, or ties) with one another. Networks can have few or many actors, and one or more kinds of relations between pairs of actors. To build a useful understanding of a social network, a complete and rigorous description of a pattern of social relationships is a necessary starting point for analysis. That is, ideally we will know about all of the relationships between each pair of actors in the population [1].
The amount of information that we need to describe even small social networks can be quite great. Managing these data, and manipulating them so that we can see patterns of social structure can be tedious and complicated. All of the tasks of social network methods are made easier by using tools from mathematics. For the manipulation of network data, and the calculation of indexes describing networks, it is most useful to record information as matrices [2].

2. CLASSIFYING COMMUNITY

A *community* is an amalgamation of living things that share an environment. In human communities, intent, belief, resources, preferences, needs…may be present and common…, but the definitive driver of community is that all individual subjects in the mix have something in common.” Recently, researchers have employed clustering algorithms to identify communities in blogs [3]. Determining community in blogs requires crawling and analyzing the blogs, along with a methodology to discover community patterns from existing blogs[3].

2.1 Features of sense of community

1. Feelings of membership.
2. Feelings of influence.
3. Reinforcement of needs.
4. Shared emotional connection [4].

2.2. Social network analysis

Social network analysis (SNA) is the mapping and measuring of relationships and flows between people, groups, organizations, computers, or other information/knowledge processing objects. Social network analysis as a theme has been studied for years [5].

Social Network Analysis (SNA) provides powerful methods to study the relationships between people expressed as binary or weighted adjacency matrices. It can be used to find influential or popular nodes, communities and informal hierarchies. [6].

2.3 Related work and Contribution of SNA

A vast number of studies have shown the importance of social networks in organizations in related areas such as power, individual performance, organizational performance cooperation, and job satisfaction,. While it is clear that social network analysis can be important in understanding the way in which individuals interact in organizations, there are still areas which need to be researched further in order for this field to mature [7].

2.4 Role of SNA

The Internet has spawned different types of information sharing systems, including the Web. Recently, online social networks have gained significant popularity and are now among the most popular sites on the Web. The resulting social network provides a basis for maintaining social relationships, for finding users with similar interests, and for locating content and knowledge that has been contributed by other users [8].

One applicable field of SNA is Leadership development practitioners are increasingly interested in social networks as a way to strengthen relationships among leaders in fields, communities, and organizations. Evaluating leadership networks is a challenge for the field of leadership development. Social network analysis (SNA) is an evaluation approach that uses mathematics and visualization to represent the structure of relationships between people, organizations, goals, interests, and other entities within a larger system[9].

3. GRAPH THEORY

The natural mean to mathematically represent a network is through a graph model.

A graph G = (V; E) as an abstract representation of a set of objects V, namely vertices (or nodes), and a set E of edges which connect pairs of vertices together. V (G) such set of vertices V and, similarly, as E(G) the set of edges within the graph G. The amount of vertices V and of edges E, namely the cardinality of V and E, are commonly represented by n and m or denoted by $|V|$ and $|E|$. Two vertices connected by an edge e are called end vertices for e, and they are adjacent or neighbours[10].

3.1 Centrality Measures

There are four measures of centrality that are widely used in network analysis: degree centrality, closeness, betweenness, and eigenvector centrality [1].

3.2 Degree centrality

The most intuitive measure of centrality of a vertex into a network is called degree centrality. Given a graph G = (V; E) represented by means of its adjacency matrix A, in which a given entry $A_{ij} = 1$ if and only if i and j are connected by an edge, and $A_{ij} = 0$ otherwise, the degree centrality $C_D(vi)$ of a vertex $vi \in V$ is defined as :

$$C_D(v_i) = d(v_i) = \Sigma_j A_{ij} \ldots\ldots (1)$$

The idea behind the degree centrality is that the importance of a vertex is determined by the number of vertices adjacent to it, i.e., the larger the degree, the more important the vertex is.

3.3 Closeness centrality

A more accurate measure of centrality of a vertex is represented by the closeness centrality. The closeness centrality relies on the concept of average distance, defined as

$$D_{avg}(v_i) = \frac{1}{n-1}\sum_{j\neq 1}^{n} g(v_i, v_j) - - - - - (2)$$

where $g(v_i; v_j)$ represents the geodesic distance between vertices v_i and v_j .
The closeness centrality $C_C(v_i)$ of a vertex v_i is defined as

$$C_c(v_i) = \frac{1}{n-1}\sum_{j\neq 1}^{n} g(v_i, v_j) - - - - - - (3)$$

3.4 Betweenness centrality

It relies on the concept of shortest paths, previously introduced. In detail, in order to compute the betweenness centrality of a vertex, it is necessary to count the number of shortest paths that pass across the given vertex.
The betweenness centrality $CB(v_i)$ of a vertex vi is computed as

$$C_{B_{(v_i)}} = \sum_{r_s \neq v_i \neq v_t \in v} \frac{\sigma_{st}(v_i)}{\sigma_{st}} - - - - (4)$$

3.5 Eigenvector centrality

Another way to assign the centrality to a vertex is based on the idea that if a vertex has many central neighbors, it should be central as well. This measure is called eigenvector centrality and establishes that the importance of a vertex is determined by the importance of its neighbors.

The eigenvector centrality $C_E(v_i)$ of a given vertex v_i is

$$C_B(v_i) \propto \sum_{v_j \in N_i} A_{ij}\, C_e(v_i) - - - - (5)$$

where N_i is the neighborhood of the vertex v_i, being $x \propto Ax$ that implies $Ax = \lambda x$. The centrality corresponds to the top eigenvector of the adjacency matrix A.

4. DATA PREPARATION FOR SNA

4.1 Relation extraction

Relation extraction is the process of interact with the web and then extract relevant data for a user [3]. In this paper the extraction algorithm implemented with crawling process viewed in the proposed block diagram. The work proposed in this paper applied using machine learning crawling. Identifying community in blogs requires crawling and analyzing the blogs starting with small graph as a representation of community as a base of crawling.

4.2 Crawling Network (Snow ball Method)

Determining community in blogs requires crawling and analyzing the blogs, along with a methodology to discover community patterns from existing blogs [3]. In the case of online social networks, crawling the graph efficiently is important since the graphs are large and highly dynamic. This research presents crawling graphs include breadth-first search. In snow ball method the researcher use some graph as a biase. Case studies have used samples of 0.43% of nahrain users university [7].

5. CLUSTERING

The clustering coefficient of a node with N neighbors is defined as the number of directed links that exist between the node's N neighbors, divided by the number of possible directed links that could exist between the node's neighbors (N(N − 1)). The clustering coefficient of a graph is the average clustering coefficient of all its nodes [7]. Introducing classification in the proposed algorithm due to the actual need for the classification of several communities in the broader community, which will affect the improvement of relations within a single institution.

6. FRAMEWORK OF THE PROPOSED SYSTEM

This section presents outlines of the proposed system in a research paper presented: general framework designed to improve the performance of enterprise level through the collection of information available on that institution, which contain a large number of employees and presidents who vary their views about the way the organization of work and the services they provide.

That the views of the staff and their intentions can be obtained through a code or more or its own website. Where it is to ask questions and exchange ideas and information about the many issues that may go beyond the work of the institution to which they belong. So we must build an interactive forums where ideas and intentions pose questions and constructive conversation, and that lead to the creation of system can be utilized in the development work of the Foundation.

6.1 Proposed System

The proposed system consists of two modules : the first consists of a repository containing a package of blogs and websites which mimic the work of the institution and interact through a lot of clients and beneficiaries and experts in the field of work of the Foundation. It is directed questions and inquiries . Then note the reactions and ideas.

7. RESULTS AND DISCUSSIONS

Bottom-Top design used to extract cluster features starting from a specific characteristic to general one for different items in homogeneous environments with supervised learning mechanism. So in order to classify different items in these regions, labeled samples through linear programming (LP) for discriminating between two or more classes via parallel hyperplanes used in this research to distinguish different items which are belonging to different clusters.

Figure (2) illustrates clustering area with different items. First of all, and in order to classify these items, the proposed algorithm used two parallel hyperplanes F1F1 and F2F2 for this purpose. In this environment there are two classes of items, on the other hand there are a lot of them till now not be classified. So there is a need to another hyperplanes to recognize the rest items. S1S2 will be the second hyperplanes to achieve the classification. Since there are no items will be wait for classifying, T1T2 be the last hyperplanes. The algorithm in figure (1) describe the procedure for categorize two items.

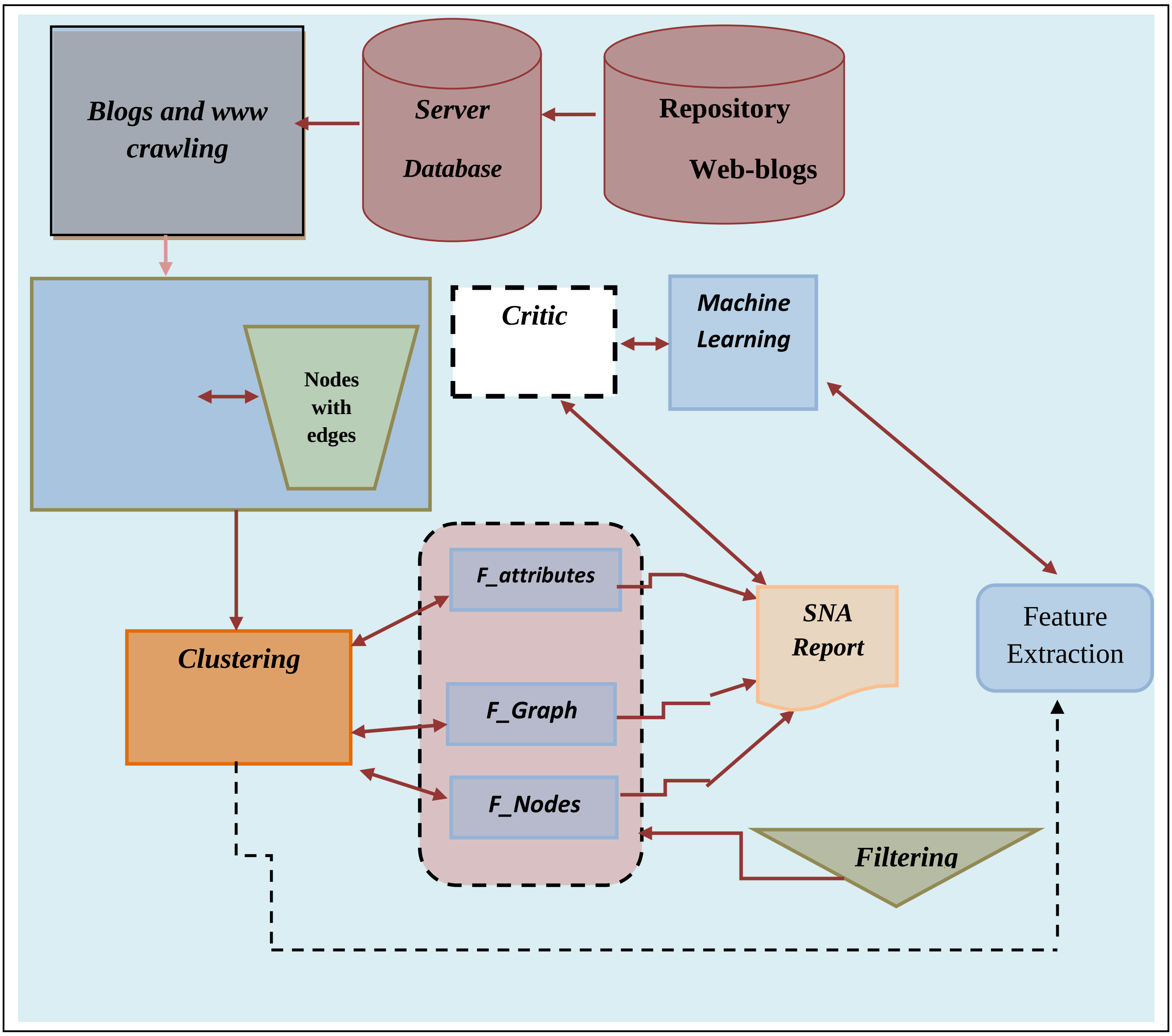

Figure (1) block diagram of the proposed algorithm

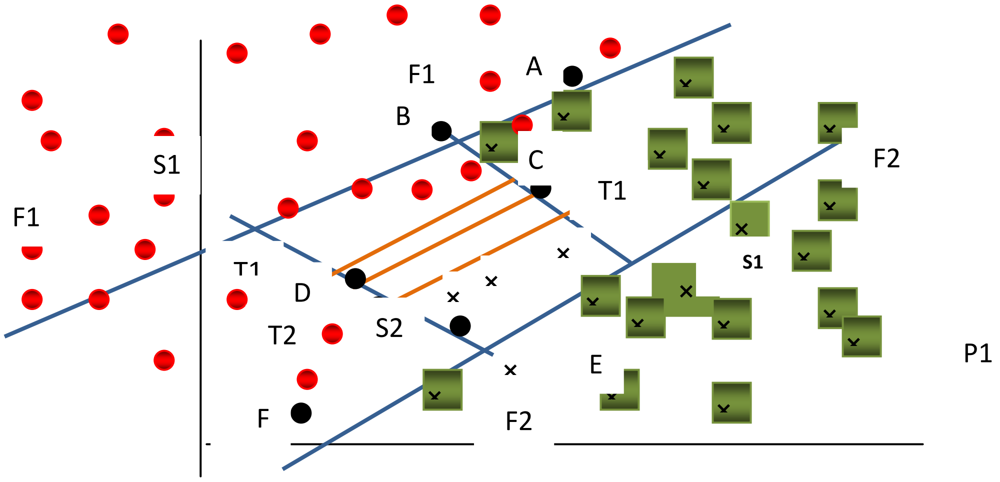

Figure (2) Clustering area with different items

Piecewise discrimination Algorithm

This discriminating algorithm implemented in a recursion algorithm
with two labeled classes and two hyperplanes.
Let F1F1,F2F2 be a hyperplanes
P ← F1F1
q← F2F2
Function piecewise(item,C1,C2) :
While d(q,p) = [$\sqrt{(q1-p1)^2}+\sqrt{(q2-p2)^2}$] > 0
If $item_{ij}$ ϵ C1 then
begin
C1[i,j] ← $item_{ij}$ else C2[i,j] ← $item_{ij}$
i = i+1 ;
end;
j = j+1 ;
Return
End
piecewise
Return

7.1 Data Format and preparing data

The data format in the data processed was RDF which is suitable for mapping edges and relation between nodes in the network where the linking between nodes (resources) represented by edges such that these edges are named to be predicates of nodes. Therefore RDF will be more understandable with visual conceptualization. The researcher select this structure of format due to its properties as a language of defining web ontology and communities. Data universities were selected as the best way to test the proposed algorithm and for that was obtained that data from the site: [http://archive.ics.uci.edu/ml/datasets.html][14].

Data include: Number of Instances: 285 including different controlled universities. The researcher coded attributes to minimize space used to allocate data, where the number of attributes was seventeen : beginning from university name and ended with academic emphasis. Note that there are two kinds of university controlling, such that , private coded with (1) and state coded with two. The attributes presented in table (1) are as follows : 1.University-name 2. Location 3. Control 4. number-of-students in thousands 5. male: female (ratio) 6. student: faculty (ratio) 7. sat-verbal 8. sat-math 9.Expenses 10. percent-financial-aid 11. number-of-applicants 12. percent-admittance 13. percent-enrolled 14. academics 15. Social 16. quality-of-life 17. academic-emphasis

.

Tablet (1) sample of universities coded attribute data

UN	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	10	5:10	20:1	450	500	4-7	50	17	80	60	1-5 2	1-5 2	1-5 2	4
2	2	20+	50:50	20:1	450	500	4-7	50	17+	80	60	1-5 3	1-5 4	1-5 5	4
3	1	5-10	40:60	20:1	500	550	10+	60	10-13	50	40	1-5 4	1-5 5	1-5 3	3
4	1	10-15	45:55	12:1	550	575	10+	60	13-17	60	40	1-5 4	1-5 4	1-5 3	3
5	1	5-	50:50	11:1	625	650	10+	40	10-13	20	50	1-5 5	1-5 5	1-5 5	3
6	1	5-	70:30	10:1	650	780	10+	70	4-	15	90	1-5 5	1-5 1	1-5 3	1
7	1	5-	60:40	10:1	600	650	10+	70	4-7	40	50	1-5 4	1-5 4	1-5 3	1
8	1	5-	70:30	9:1	550	650	10+	65	4-	85	35	1-5 3	1-5 3	1-5 2	3
9	2	10-15	60:40	15:1	4-	-	4-	80	4-	80	60	1-5 3	1-5 2	1-5 2	7

The first row start with university names which are coded by integers as they sorted in the repository, while the others are attributes excluding university location. So the last one begin from three and ended with seventeen.
To see the impact of Social Network Analysis for improve the quality of academic effectiveness, parameters of SNA which are explained previously are studied and tested in as shown in the presented data in table (1).

Table (2) SNA Tools for tested data

Seq	No. of nodes	degree	Betweenness	Eigen Vector	Centrality
1	2^{16}	1.637	0.2316	2.1	3.341
2	2^{10}	1.210	0.1532	1.92	3.123
3	2^{31}	2.1	0.1107	1.826	3.182
4	2^{17}	1.6	0.235	2.101	3.221
5	2^{23}	2^{33}	2.33	1.821	4.217

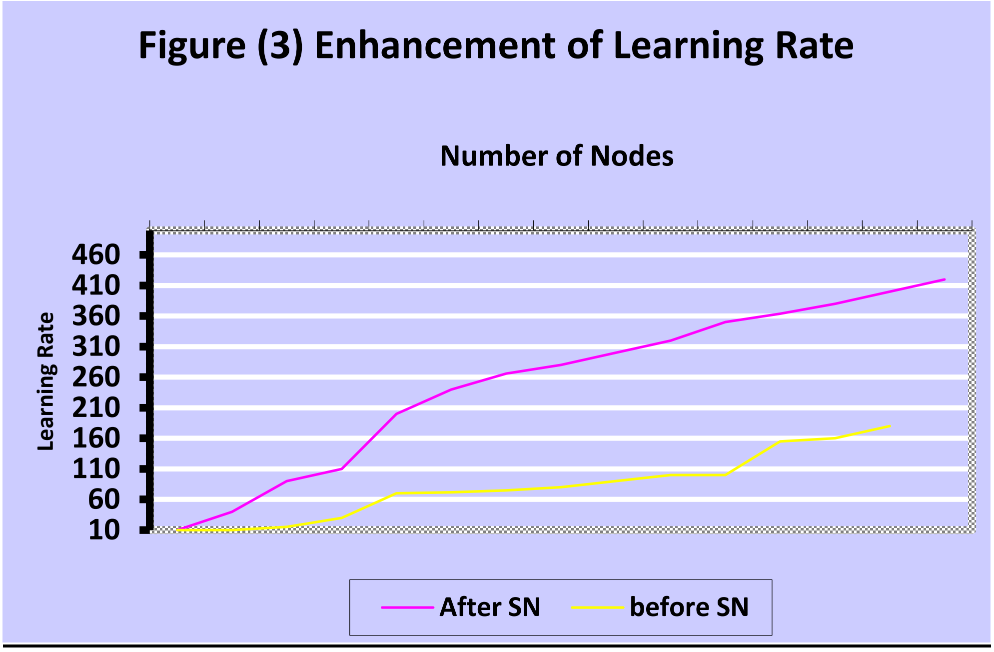

Figure (3) presents comparison of enhancement learning rate before and after SNA with variable number of nodes. The scale between 10 to 460 which represents learning rate which was the outcome of assessments . The tested nodes was 520 from form different clusters from different departments.

8. CONCLUSIONS

This system and the proposed algorithm is effective in improving the performance of any system or institution level, through node classification into groups with each other and share the same qualities, and be a strong relationship between them is measured through the use of social analysis tools. It has been identified the affecting nods, which are usually at the center and which have a significant impact on the system or the whole group. The algorithm was used to note the increase in the impact of the contract influential in decision-making for each group to be more effective impact. The use of the qualities or properties and the training of university education system using the data, the presence of expert or supervisor , the system to be adapted to the circumstances of each and every case.

9. ACKNOWLEDGMENT

My thanks and appreciation to all who contributed in the speed of response of the replies and observations on social networks, at the university and departments concerned, and so are planning for the success of this work, which serves institutions in general and educational institutions in particular.

10. BIOGRAPHY

The author is Assist. Prof. and Doctorial Philosophy in Artificial Intelligence. Supervisor of many M.Sc. students in Information Engineering Colleges rather than science college. Lecturer in Al-Nahrain University. Manager of Computer Center and Avin Cina for E-Learning. Interested in Data Security, Soft Computing, Distributed Database, Engineering Analysis, and Data Mining .

11. 11. REFERENCES

1. Gruzd Anatoliy. (2008) The Analysis of Online Communities using Interactive Content-based Social Networks , Graduate School of Library and Information Science, University of Illinois at Urbana- Champaign.
2. Hanneman, Robert A. (2005) Introduction to Social Network Methods, Department of Sociology, University of California.
3. Chin, Alvin, Chignell Mark (2005) Finding Evidence of Community From Blogging Co- Citations :A Social Netwok Analytic Approach , University of Toronto.

4. Chin, Alvin, Chignell, Mark (2007) Identifying Communities in blogs : roles of Social Network Analysis and Survey Instruments , *Int. J. Web Based Communities, Vol. 3, No. 3*.
5. Bohn, Angela, Feinerer, Ingo, Hornik , Kurt, and Patrick Mair (June 2001) Content Based Social Network Analysis for Maling Lists, The R Journal Vol. 3/1.
6. Waldstrorm,Christian (August 2003) Understanding Intra-Organizational Relations Through Social Network Analysis , Department of Organization and Management " Aarhus School of Business.
7. Mislove, Alan, Marcon, Massimiliano, Gummadi, Krishna P, Druschel, Peter, and Bhattacharjee, Bobby (2008) Measurement and Analysis of Social Network Analysis.
8. Hoppe, Bruce, Reinelt, Claire (2010) Social Network Analysis and the Evaluation of Leadership Networks, journal homepage: www.el sevier.com/ locate/ leaqua.
9. Ferrara Emilio (2012) Mining and Analysis of Online Social Networks, A thesis submitted for the degree of Philosophi_Doctor (PhD) in Mathematics 2012 February. Department of Mathematics ,University of Messina.
10. Cai, Deng, Zheng Shao, He, Xiaofei,Han, Jiawei (2005) Mining Hidden Community in heterogeneous Social Networks , Department of Computer Science, University of Chicago.
11. Golbeck, Jennifer (2006) The Dynamics of Web-Based Social Network : Membership, Relationships, and Change ", University of Myer land.
12. Shi, Xiaodong (2006) Social Network Analysis of Web Search Engine Query Logs , School of Information , University of Michigan.
13. Kazienko, Przemysław, Michalski , Radosław , Palus , Sebastian (June 29-July 1,2011) Social Network Analysis as a Tool for Improving Enterprize Architecture, Proceedings of the 5th International KES Symposium on Agents and Multi-agent Systems, KES-AMSTA, Manchester .
14. Lichman, M (2013) UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.