ISSN:2320-0790

An International Journal of Advanced Computer Technology

Empirical Analysis of Data Mining Techniques for Social Network Websites

S.G.S Fernando, Management and Science University, Malaysia
MdGaparMdJohar, Management and Science University, Malaysia
S.N. Perera, University of Colombo, Sri Lanka

Abstract: Social networks allow users to collaborate with others. People of similar backgrounds and interests meet and cooperate using these social networks, enabling them to share information across the world. The social networks contain millions of unprocessed raw data. By analyzing this data new knowledge can be gained. Since this data is dynamic and unstructured traditional data mining techniques will not be appropriate. Web data mining is an interesting field with vast amount of applications. With the growth of online social networks have significantly increased data content available because profile holders become more active producers and distributors of such data. This paper identifies and analyzes existing web mining techniques used to mine social network data.

Keywords: Social Networks, Web Data Mining, Data mining techniques, Social Network Analysis

1. Introduction

Social networks are defined as virtual spaces where people of all ages can make contacts, share information and ideas, and build a sense of community [1]. A social network community represents people and connects them. It has also provided a way of keeping in touch with friends, constructing personal profiles, view others profiles and connections, communicate and share personal information. Members of social network communities manage their identity through their profile, they meet new friends and like-minded people in the community, they connect with each other, rate peers and objects, ask questions, get answers and discuss topics [2]. Facebook, twitter, MySpace, and bebo, can be taken as most commonly accessed social network sites. Social networks can be used in many business activities like increasing word-of-mouth marketing, marketing research, General marketing, Idea generation & new product development, Co-innovation, Customer service, Public relations, Employee communications and in Reputation management [3].

Substantial quantities of data are available on online social network sites, blogs, knowledge sharing sites, collaborative filtering systems, news groups and email systems. According to the IEEE Spectrum, new data will come in great part from the social web. Already, Facebook users share more than 30 billion pieces of content web links, news stories, blog posts, photos each month. Twitter users generate more than 155 million tweets per day. These companies are establishing themselves as platforms for data aggregation, granting other companies access to the results through APIs, or application programming interfaces. For example, Facebook Connect, this allows sites to access public data from Facebook user. The growth of these interconnections, of course, spawns still more online data, in a widening spiral [4]. In studying the significances of this area researchers are faced with the opportunity to analyze social network data at unique levels of scale and in a timely manner. Data mining along with data warehousing supports business users to analyze a vast amount of data and discover patterns. Web Data Mining used in web environment by data mining technology. This research paper analysis the existing data mining techniques available to mine data in social network sites.

First the paper gives an overview of data mining and existing data mining algorithms. The second part of the paper explains the web data mining along with the taxonomy of web data mining. The third section of the paper presents the empirical analysis of available data mining techniques to mine social network data.

2. Data Mining

2.1. Overview of data mining

Data mining can be performed on large data sets in order to discover useful patterns and relationships

.

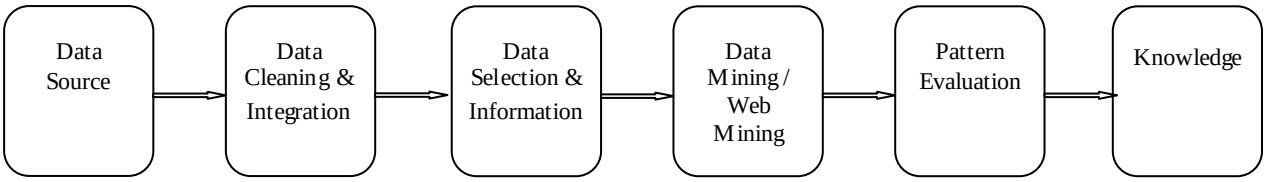

Figure 1: Steps in Data Mining Process

It is the process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems [5]. Data warehouses are being made use of in order to store large amounts of data. The growth of commercial databases has had a huge impact on the necessity of data mining in organizations. Data mining allows organizations to proactively respond to problems that may arise in future by forecasting about specific occurrences [6]. As illustrated in figure 1, the first step is data preparation. Data is selected, processed under the knowledge of a domain expert. Second, a data mining algorithm is used to process the prepared data. The third phase is to analyze whether important facts were generated by the data mining algorithms [5].

Data mining is becoming increasingly common in both the private and public sectors. Industries such as banking, insurance, medicine, and retailing commonly use data mining to reduce costs, enhance research, and increase sales. In the public sector, data mining applications initially were used as a means to detect fraud and waste, but have grown to also be used for purposes such as measuring and improving program performance [5].

Hart (2006) identifies that data mining can be directly effect on the higher growth and increased turnover in certain business areas. Data mining will reduce bad debts and healthier credit portfolios, improved sales leads. It can be used to improved client differentiation, customer profiling and customer retention and will reduced call-centre costs. By using this technology company can provide better quality and more accurate information and produce lot of operational and strategic advantages [7].

2.2. Data Mining Algorithms

Data mining techniques is dividing in to two approaches; direct approach is used in prediction where it tries to predict a state of a new value by looking at the known values. The Second approach, non-direct approach is used to identify new patterns by looking at the past values. Before start creating mining models data should be cleaned and prepared. The Mining models can be created on following Algorithms [8].

Association Rules - This algorithm can be used in marketing base analysis like identifying cross-selling opportunities. This takes multiple items in a single transaction, scans the data and counts the number of times the items appear in the transaction [8] so it can be used to identify the relationships in the large data sets.

Clustering - This algorithm groups the data according to their similar characteristics. This can be used to identify the relationship of the characteristics among a group. When a new data is introduced, the characteristics of it can be mapped with the relationships, it can be used to predict the behavior of the new data. Clustering can be used to find anomalies of the data as well. This is commonly used in systems of fraud detection and Customer relationship Management [8].

Decision Trees - This is the simple and one of the most commonly used algorithm. This is used to predict discrete and continues variables [8].

Liner regression - This is predicting only continues variables using single multiple liner regression formula [8].

Logistic Regression – This algorithm uses a neural network without hidden layers [8].

Naïve Bayes –This can be used to calculate probabilities for each possible state of the input attribute when a state of a predictive attribute is given. This can be used as the starting algorithm of the predicting process [8].

Neural Networks – This algorithm has been adopted from artificial intelligence. This can be used to search nonlinear functional dependencies. This will perform non liner transformations on the data in layers from input layers to the hidden layers and finally to the output layer [8].

Sequence Clustering - This looks for cluster based models than the similarity if the data. The model use sequence of events by using hidden Markov chains. The states are models to a matrix and the probabilities of transiting from one state to another in the cells of the matrix. With these probabilities the probabilities for sequence of transition can be calculated by multiplying probabilities of state transitions in the sequence. The chains of highest probability can be used to model the clusters [8]

Time Series - This is used to forecast continues variables. This is a combination of two algorithms called auto regression trees and Auto regressive integrated moving average.

3. Web Data mining

3.1. Overview of Web Data Mining

Web mining is the process of analyzing and discovering patterns on web data. It can be defined as searching data automatically from various online resources [9]. Since the heterogeneity nature of the data on web, mining is very hard [10]. Applying the above mentioned data mining algorithms directly is not feasible. Special methodologies should be used to make web data structured and to mine. The web information retrieval tools like web crawlers do not do exactly web mining, they extract only text, and they do not extract information or knowledge from web data.

Web mining can be categorized it to three aspects, web usage mining, web content mining and web structure mining [11]. Web Usage to identify user browsing patterns of web sites by recording the URL visited or by accessing web server logs [12]. This can be used to identify users, figure our patterns in session creations, detect robot, filtering and deriving the sites that visited together by the user as well [13]. Web Structure Mining is identifying the structure of a particular web. This will focus on hyperlink information. By analyzing the structure, interesting connectivity information can be recognized [9]. This is ideal when applying web data mining in social network sites. By starting with one user's profiles page, the friends' network or the friends' cycles can be identified. This data can be very useful to cluster the profile page data and identify the relationships and interesting details regarding the connectivity. Web content mining is mining the content of the web pages. There are two approaches to do web content mining, as mentioned in paper [11]. Namely agent-based web mining systems having three variations like intelligent search agents, information filtering/categorization and personalized web agents. The second one is the databases approach with multilevel databases or web query systems. Content mining does not exactly means to search keywords on web pages; it is extracting information and discovering patterns buy analyzing web documents. This leads to discover new knowledge. This is more difficult that mining data in data warehouses since web data is semi structured [10].

3.2. Web Data Mining algorithms

There are various web data mining algorithms were introduce from 1996. Each of those is under the three categories mentioned above and many researches have been done on various fields. Figure 2 is a taxonomy of web mining highlighting the data and Figure 3 explains the algorithms used under different categories of web mining. Slightly modified form the one introduced in the paper [10].

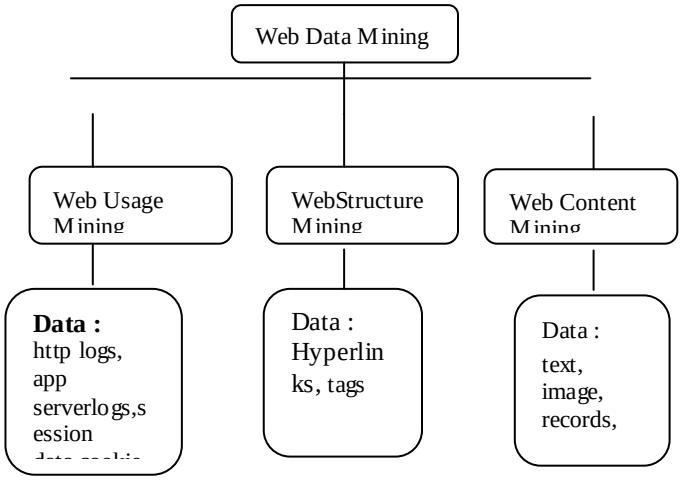

Figure 2 : Taxonomy of Web Mining (Data)

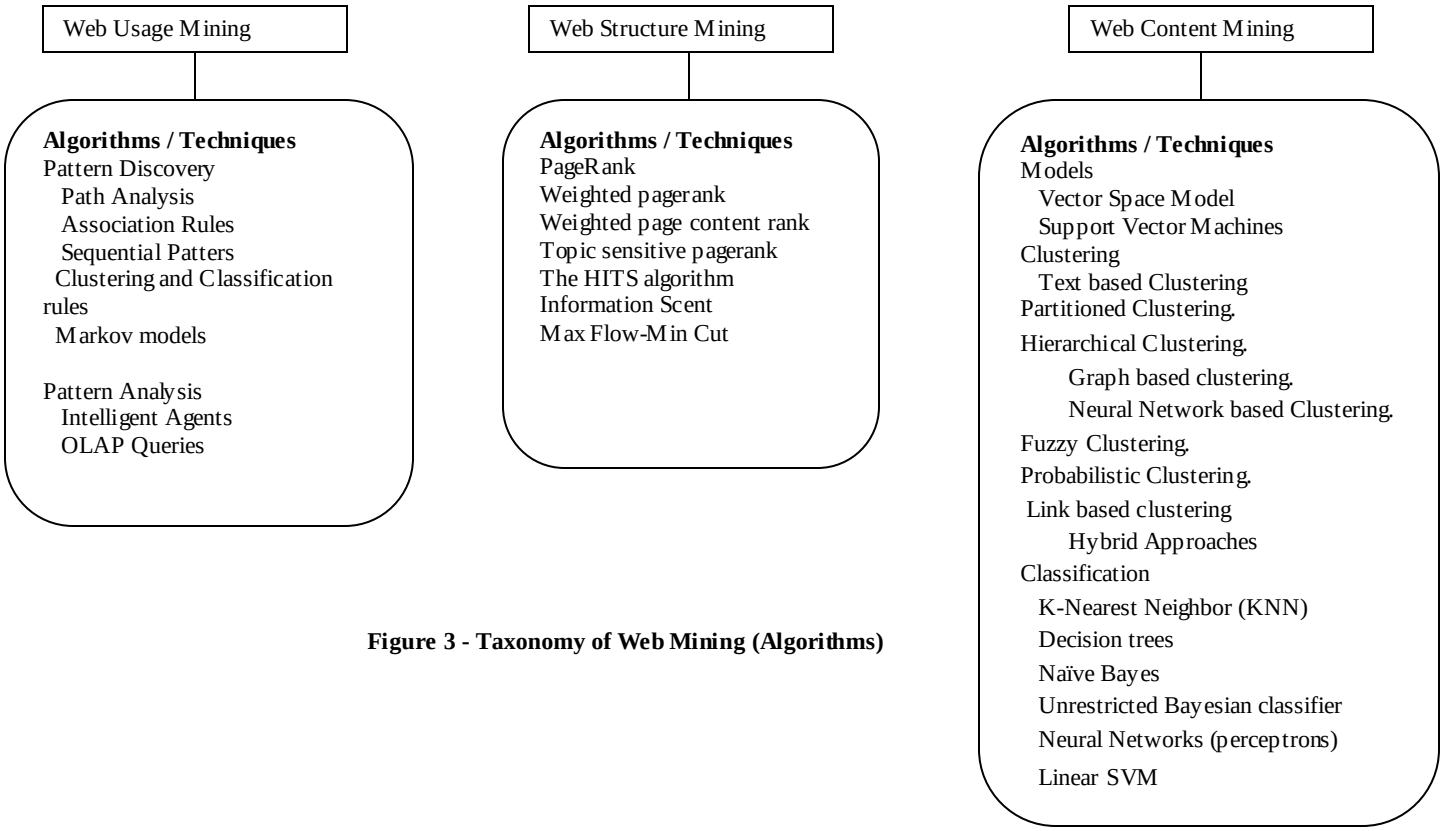

Figure 3 - Taxonomy of Web Mining (Algorithms)

4. Social network data mining Algorithms

4.1. Overview of the Algorithms

When mining social network data it should be a combination of web structure mining and web content mining. Analyzing the structure of the Social network is known as Social Network Analysis. Social Network analysis where was a hot topic among the researchers from 1994 especially in the fields of psychology, anthropology, economics, geography, biology and epidemiology, anthropology, economics, geography, biology and epidemiology [14]. Several tools has been introduced in the area of social network analysis like Graph Characterization Toolkit [26], TweetHood [27] , Meerkat [28], NetDriller [29], HiTS/ISAC Social Network Analysis Tool [30] ,D-Dupe [31] and X-RIME, a cloud-based library forlarge scale social network analysis [45]. Web content mining was more popular in marketing and advertising research [1, 9, 10].

4.2. Analysis of the Algorithms

In social network data mining, existing data mining algorithms cannot be used directly because of the dynamic behavior [16, 17].When analyzing the literature on social network data mining techniques; it was found that each algorithm has strengths and weaknesses. The following section explains about the existing algorithms in detail.

4.2.1. Graph mining algorithms

Most popular data mining technique in Social Network Analysis is using Graph mining algorithms. World Wide Web including social networks is a collection of interconnected hypertext documents. These are interconnected by hyperlinks. So web can be considered as a directed graph, where nodes will be the hypertext documents and edges will be hyperlinks [15]. Web structure analysis based on graph algorithms has been analyzed in many researches in past years [18-20]. Lahiri and Berger-Wolf (2008) have created and tested methods combining network, quantitative, semantic, data processing, conversion and visualization-based components. They have introduce a new graph mining algorithm “periodic subgraph mining, or the discovery of all interaction patterns that occur at regular time intervals” taking into consideration of the dynamic behavior of Social Networks. The Algorithm is based on frequent pattern mining in transactional and graph databases with periodic pattern mining in unidimensional and multidimensional sequences [16].
Bourqui et.al (2009) presented a framework which is based on dynamic graph discretization and graph clustering. This framework is capable of detecting the dynamic changers of the social network structure and

identifies events analyzing temporal dimension and exposes command hierarchies in social networks [17].

The particular algorithms treat the network as a graph but it minimize the clustering problems and graph partitioning problems. As a solution minimum spanning trees can be used to identify users having similar profile pages and strong relationships [21,22, 23]. Zhang et.al (2010) have conducted an experiment on the applicability of general greedy, hill-climbing and centrality-based algorithms on dynamic social network data to identify key users for target marketing by mapping the network to a graph.They have proposed a new approximation searching algorithm based on the heuristics information from the above algorithms [24].

Even though the graphs map the connection or the relationship between the nodes it does not show the relationship strength. One interesting tools has been developed called SocialViz to provide h frequency information on social relationship among multiple entities in the networks by using a Frequent Pattern Visualization Approach [25].

4.2.2. Classification

Classification is the method of categorizing data in to one of many categories. This can be apply in web data mining to classify user profiles based on profile characteristics. Most popular classification algorithms in data mining are decision trees, naïve Bayesian classifier and neural networks [15].

Surma and Furmanek (2010) introduced an interesting algorithm called C&RT, combining classification and regression tree algorithms to determine rules to identify target groups to market. This can be used in real social network data [44].

4.2.3. Clustering

Clustering is grouping a set of items such a way that items in the same group are more similar to each other than to those in other groups. These groups are known as a cluster. Clustering is mainly used in information retrieval in web mining. Based on past research clustering will increase the efficiency in information retrieval [15]. Graph based clustering is comely used in web structure mining as explained in early section. Text based clustering is most commonly used in web content mining whether you create clusters based on the content of the web document [15]. Bartal et.al (2009) introduced an interesting method combing social network analysis and text based clustering to predict the nodes of a social network would be linked next.

4.2.4. Associations

Association rule mining is used to find frequent patterns and correlation among data set [15]. Nancy et.al (2013) had use association rules to mine social network data using 100 Facebook university pages. The research focused on the formulation of association rules using which decisions can be made and uses Apriori Algorithm to derive association rules [46].

4.2.5. Semantic Web and Ontology

Semantic Web is a new research area where it tends to give meaning to Web data. This enables machine and humans to interact intelligently and exchange information [37]. There are many researchers has been carried out in this filed like using semantic geo catalogues [35] and recovery in mental health information [36]. Zhou et.al (2008) explains applying statistical learning methods on semantic web data. It has used an extended FOAF (friend-of-a-friend) ontology applied as a mediation schema to integrate Social Networks and a hybrid entity reconciliation method to resolve entities of different data sources [39].Tushar et.al (2008) explains the usage of Semantic Web technology to detect the associations between multiple domains in a Social Network [40].Opuszko and Ruhland (2012) introduced a novel approach of using semantic similarity measure based on pre-defined ontologies for classify social network data [37]. Ostrowski (2012) has developed an algorithm to retrieve information in social networks to identify trends. The Algorithm has use semantics for determine the relevancy of networks using unstructured data. The algorithm was tested on twitter messages [38].

4.2.6. Markov models

Markov chains are is a mathematical algorithm that undergoes transitions from one state to another, among a finite or countable number of possible states. It is a random process where the next state depends only on the current state and not on the sequence of events that preceded it [47]. Markov models can be used in web mining to predict users next action. The Social network can be mapped to a where nodes will be users previous visits. So based on the node information by using Markov models users next visit can be predicted [15]. The social networks can be mapped as a Markov chain [32, 33]. Crnovrsanin et.al (2009) introduces a new procedure for social network discovery based on the Markov centrality of a node [34]. Peng et.al (2011) has used the Hidden Markov model along with SVM classifier to predict sports, weather and social activities on Twitter messages [41].

Table 1: Existing web mining Algorithms

WebMining Category	Web Mining Algorithm	Usage in Social network data mining					
		Research Details	**Year**	**Success rate /Results**	**Tested on Real data**	**Static (S)/ Dynamic(D)**	**Handle temporal data**
	Clustering	Different clustering algorithms including K-Means, Make Density Based, Farthest First, EM and Filtered on Face book 100 university dataset.	2012	According to the results of their experiment average accuracy rates for 10 university set are as follows: K-means (64%), Make Density ased(55%),Farthest First(60%), EM(60%) and Filtered (57%) [47].	X	S	X
		An algorithm based on Business System Planning clustering algorithm along with Linked list data structure [51].	2010	The edges between node is assumed to have the same weight, but in real world it might get changed .	X		
	Classification rules	C&RT algorithm to improve market response by predicating target groups. This has used on real data from the Biznes.net social network [44].	2010	When analyzing the results of the algorithm it proves that applying simple data mining can be used to improve marketing responses.	✓.	S	
		Analysisof different clustering algorithms to mine data in facebook and twitter. Algorithms includes C4.5 Algorithm: RndTree (Random Forest): C-RT, CS-CRT, and CS-MC4[48]	2011	The RndTree Algorithm produced less error rate of 0.0004 when compared to all other Algorithms.	X	S	X
	Markov models	Social Network Discovery based on Sensitivity Analysis [34]introduce an algorithm that computes the Markov property of nodes and the sensitivity parameter. By encoding the pair of nodes depending on magnitude of sensitivity property to predict an edge between the nodes. that pair.	2009	The algorithm has been tested in the VAST challenge social network data set and the MIT Reality data set. The results proves that the algorithm can be used along with an visualization technique can be used to discover patterns in social networks.	X	D	X
		Hidden Markov model along with SVM classifier to predict sports, weather and social activities on Twitter messages [41].	2011				
	Intelligent Agents	There are no algorithm presents on					

		social network data yet.					
	OLAP Queries	OLAP queries can be used in Social Network analysis since it can be mapped to a Graph structure [15],[50]. No algorithm has been developed directly to mine Social network data					
Web Structure Mining	PageRank	TwitterRank as an extension of PageRank algorithm, is proposed to measure the influence of users inTwitter [53].	2010	The experimental results show thatthe proposed TwitterRank outperforms other related algorithms.This tries to fix the disadvantages of in-degree and PageRank by consideringlink structure and topical similarity amongtwitterers [53]. .	X	D	X
	Weighted pagerank	Has been proposed to use in SNA but not been used directly to mine Social network data [52].					
	Weighted page content rank	No algorithm has been developed directly to mine Social network data					
	Topic sensitive pagerank	Has been proposed to use in SNA but not been used directly to mine Social network data [15],[55].					
	The HITS algorithm	Has been proposed to use in SNA but not been used directly to mine Social network data [15].					
	Graph Algorithms	Most commonly used algorithm in SNA. All SNA research used this algorithms					
Web Content Mining	Vector Space Model	Mining social networks of person entities using Improved Vector Space Model (IVSM) [55].	2010	Based on the testing that has been done to analyze Social networks, they authors state that their approach is effective.	✓.	S	X
		New mining approach to discover social networks by integrating dialog thread structure association with message content similarity. They have modified the vector space model on heuristics in PieSpy and introducednew rules [56].	2011	Has not been tested on real situation.			
	Support Vector Machines	Neural Networks (NNs) and support vector machines (SVMs), in classification and regression tasks [57].	2010	Experiments have proved that it can be used effectively.	✓.	D	X
	Text based Clustering	Combination of social network analysis and text based clustering to predict the nodes of a social network would be linked next [43].	2009	Experiments have proved that it can be used effectively.	X	D	X

	K-Nearest Neighbor (KNN)	Used mostly in Social Network Analysis in web structure mining.					
	Decision trees						
	Naïve Bayes						
	Unrestricted Bayesian classifier						
	Neural Networks (perceptrons)						

When analyzing the literature it proves that most of the research has been carried out in web structure mining and less in web content mining. Most of the researches are tested on static networks; they do not consider the dynamic behavior.

5. Conclusion

Data mining is an interesting field which can be used to produce new knowledge by analyzing large collection of data. In order to apply the traditional data mining techniques the data should be stored in data warehouses in a structured manner. Web 2.0 has leaded the web to store massive collection of data. Even though interesting patterns can be identified by mining web data applying traditional data mining techniques is not practical because of the unstructured and the dynamic behavior of web data. This has lead may researchers to find special algorithms to mine web data. This paper has specifically focused on the techniques used to mine social network data. Most of the algorithms are developed to mine the structure of the social network where mapping the network to a graph. Fewer researches have been conduct in the category of content mining even fewer in web usage mining. For the future research it would be beneficial to focus more on the content mining where lot of human behavior patterns can be identified by analyzing the social network profile pages. Researching on an efficient hybrid approach by combing social network analysis (web structure mining) with content mining would be more useful. The statistical methods like Markov Models can be adopted to resolve the temporal behavior of web data and as well as to introduce personalization. This would really valuable to marketing and advertising fields since social network marketing is an emerging technique in business world.

Acknowledgment

I sincerely thank my supervisor Professor Dato' Dr. MdGaparMdJohar for all the advice and the guidance given. I would also like to extend my gratitude to my local supervisor Dr. NishanthaPerera. Also thank you very much Professor Dr. Ali Khatibi for all the encouragement given.

References

[1] E.K Clemons., " The Future of Advertising and the Value of Social Network Websites: Some Preliminary Examinations". Minneapolis, Minneosta, USA, 2007.

[2] D. Boyd., "Social Network Sites: Definition, History, and Scholarship",*Journal of Computer-Mediated Communication 13 (1)*, 2007.

[3] "Social Network Marketing: The Basics" Available: http://www.labroots.com/Social_Networking_the_Basics.pdf [Aug 01,2013].

[4] J. Rennie, G. Zorpette, "The Social Era of the Web Starts Now," IEEE Spectrum, June 2011. Available: http://spectrum.ieee.org/telecom/internet/the-social-era-of-the-web-starts-now [Aug 01, 2013]

[5] J.W. Seifert. "Data Mining: An Overview", CRS Report for Congress, 2004 [Online] Available: http://www.fas.org/irp/crs/RL31798.pdf [Sep 05 2013]

[6] H. Kob, G, Tan, "Data Mining Applications in Healthcare" [Online] Available: http://www.himss.org/content/files/Code%20109_Data%20mining%20in%20healthcare_JHIM_.pdf [Nov 10 2010]

[7] M Hart. "Progress of organisational data mining in South Africa", Department of Information Systems, University of Cape Town, South

Africa, 2006. [Online] Available: http://intranet.inria.fr/international/arima/006/pdf/arima00601.pdf [Jun 5, 2011]

[8] E.Veerman, T.Lachev, D.Sarka, Microsoft SQL Server 2008- Bussiness Intelligence Development and Manitenance, Microsof, PHI Learning Private Limited, pp 372 – 380, 2009.

[9] C. Yu, X. Ying, "Application of Data Mining Technology in E-Commerce" . in IEEE Int. Conf. on International Forum on Computer Science-Technology and Applications, pp.291-293, 2009.

[10] G. Lappas G. "From Web Mining to Social Multimedia Mining.", IEEE International Conference on Advances in Social Networks Analysis and Mining , pp 336 – 343, 2011.

[11] J. Srivastava, Cooley R., Deshpande M., and Tan P.N., "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," SIGKKD Explorations, 2000.

[12] S. Madria, S. Bhowmick , Research issues in web data mining, DataWarehousing and Knowledge Discovery Lecture Notes in Computer Science Volume 1676, pp 303-312 , 1999.

[13] J. Sirivastava, Web Mining :Accomplishments & Future Directions, Available : http://www.ieee.org.ar/downloads/Srivastava-tut-paper.pdf [Nov 06 2013]

[14] F. Bonchi, C. Castillo, A. Gionis, A. Jaimes ,"Social Network Analysis and Mining for Business Applications", Yahoo! Research Barcelona.

[15] M Eirinaki , "Web Mining : A Road trip", 2008. Available : http://www.ualberta.ca/~golmoham/SW/web%20mining%2023Jan2008/WEB%20MINING%20A%20ROADMAP.pdf [Nov 01 2013]

[16] M. Lahiri, T.Y. Berger-Wolf, "Mining Periodic Behavior in Dynamic Social Networks", Eighth IEEE International Conference on Data Mining, pp 375-382, 2008.

[17] R. Bourqui, F. Gilbert, P. Simonetto, F. Zaidi, U. Sharan, F. Jourdan, " Detecting structural changes and command hierarchiesin dynamic social networks", IEEE Advances in Social Network Analysis and Mining, pp 83-89, 2009.

[18] K. Borgwardt, X. Yan: "Graph Mining and Graph Kernels", KDD, 2008.

[19] J. Kleinberg, et al.: "The web as a graph: Measurements, models andmethods", COCOON, 1998

[20] J. Leskovec and C. Faloutsos, "Mining Large Graphs", TutorialECML/PKDD, 2007

[21] R.M. Assunção, M.C. Neves, G. Câmara, and C.C.Freitas. Efficient regionalization techniques for socioeconomicgeographical units using minimum spanningtrees, In International Journal of GeographicalInformation Science, v. 20, n. 7, pp 797–811,2006.

[22] V.S.A. Menezes, R.T. Silva, M.F. Souza, J. Oliveira,C.E.R. Mello, J.M. Souza, G. Zimbrao. "Mining andAnalyzing Organizational Social Networks UsingMinimum Spanning Tree". COOPIS, 2008.

[23] V. Stroele, J. Oliveira, G. Zimbrão, J.M. Souza, "Mining and Analyzing Multirelational Social Networks", IEEE International Conference on Computational Science and Engineering, pp 711-716, 2009.

[24] Y. Zhang,Z. Wang,C. Xia "Identifying Key Users for Targeted Marketing by MiningOnline Social Network", IEEE 24th International Conference on Advanced Information Networking and Applications Workshops, pp 644-649,2010.

[25] C. K. Leung, C. L. Carmichael, "Exploring Social Networks: A Frequent Pattern Visualization Approach", IEEE International Conference on Social Computing / IEEE International Conference on Privacy, Security, Risk and TrustPp 419- 424, 2010.

[26] D. Ediger, K..Jiang ,J. Riedy D. A. Bader, C. Corley, R. Farber, W. N. Reynolds, " Massive Social Network Analysis:Mining Twitter for Social Good",IEEE 39th International Conference on Parallel Processing, pp 583-593, 2010.

[27] S. Abrol, L. Khan, "TweetHood: Agglomerative Clustering on Fuzzy k-Closest Friends with Variable Depth for LocationMining", IEEE International Conference on Social Computing / IEEE International Conference on Privacy, Security, Risk and Trust, pp 153-160, 2010.

[28] J. Chen, J. Fagnan, R. Goebel, R.Rabbany, F. Sangi, M.Takaffoli, E.Verbeek, O.Za¨ıane, "Meerkat: Community Mining with Dynamic Social Networks" , IEEE International Conference on Data Mining Workshops, pp 1377-1380, 2010.

[29] N. Koochakzadeh, A. Sarraf**,**K. Kianmehr**,** J. Rokne**,** R. Alhajj**, "**NetDriller: A Powerful Social Network Analysis Tool",11th IEEE International Conference on Data Mining Workshops, pp 1235-1238, 2011.

[30] H. Asadi, C. M˚artenson, P. Svenson, M. Sk¨old,, "The HiTS/ISAC Social Network Analysis Tool", IEEE European Intelligence and Security Informatics Conference, pp 291-296, 2012.

[31] M. Bilgic , L. Getoor . B. Shneiderman, "D-Dupe: An Interactive Tool for Entity Resolution

in Social Networks", IEEE Symposium on Visual Analytics Science and Technology, pp 43-50 2006.

[32] S. White, P. Smyth, "Algorithms for estimating relativeimportance in networks," *International Conference on KnowledgeDiscovery and Data Mining*, vol. Proceedings of theninth ACM SIGKDD international conference on Knowledgediscovery and data mining, no. 2, pp. 266 – 275, 2003.

[33] J. Kemeny and J. Snell, *Finite Markov Chains*. Springer, Verlag, 1976.

[34] T. Crnovrsanin, C.D. Correa, K. L. Ma, "Social Network Discovery based on Sensitivity Analysis", IEEE Advances in Social Network Analysis and Mining , pp 107-112, 2009.

[35] F. Farazi, V. Maltese, F. Giunchiglia, and A. Ivanyukovich, "Afaceted ontology for a semantic geo-catalogue," in *Proceedings ofthe 8th extended semantic web conference on The semanicweb:research and applications - Volume Part II*, ser. ESWC'11. Berlin,Heidelberg: Springer-Verlag,, pp. 169–182, 2011.

[36] M. Hadzic and E. Chang, T. S. Dillon,E. Chang, R. Meersman, and K. Sycara, Eds. Berlin, Heidelberg:Springer-Verlag, Web Semantics for Intelligent and DynamicInformation Retrieval Illustrated within the Mental Health Domain,pp. 260–275, Available: http://dx.doi.org/10.1007/978-3-540-89784-2 10 [Sep 03 2013].

[37] M. Opuszko, J. Ruhland, "Classification Analysis in Complex Online SocialNetworks Using Semantic Web Technologies", IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp 1032-1039, 2012.

[38] D. A. Ostrowski, "Semantic Social Network Analysis for TrendIdentification", IEEE Sixth International Conference on Semantic Computing, pp 178-185, 2012.

[39] Zhou, C. Chen, H. Tong , "Learning aProbalisitic Semantic Mdoel from Heterogeneous SocialNetworks for Relationship Identification Tools withArtificial Intelligence",ICTAI 08. 20th IEEEInternational Conference Vol 1, 2008.

[40] A.K.Tushar, , P.S.Thilagam, "An RDF Approach forDiscovering the Relevant Semantic Associations in a SocialNetwork Advanced Computing and Communications",16th International, 2008.

[41] Peng, Huan-Kai; Zhu, Jiang; Piao, Dongzhen; Yan,Rong; Zhang, Ying, "Retweet Modeling Using Conditional Random Fields ",Data Mining Workshops (ICDMW), 2011 IEEE 11thInternational Conference,pp 336 – 343, 2011.

[42] K. Sugiyama, H. ,.; Ohsaki, MImase, T. Yagi,J. Murayama, "NMF: Network Mining Framework UsingTopological Structure of Complex Network"s, pp 210 – 211,2008.

[43] A. Bartal, E. Sasson, G. Ravid, "Predicting Link s in Social Networks using Text Mining and SNA" ,Advances in Social Network Analysis and Mining, pp 131-136 , 2009.

[44] J. Surma , A. Furmanek, "Improving marketing response by data mining insocial network", International Conference on Advances in Social Networks Analysis and Mining, pp 446-451, 2010.

[45] W. Xue, J. Shi, Bo Yang , "X-RIME: Cloud-Based Large Scale Social Network Analysis", IEEE International Conference on Services Computing , pp 506-513, 2010.

[46] P. Nancy, R. G. Ramani, S. G. Jacob, " Mining of Association Patterns in Social Network Data (Face Book 100 Universities) through Data Mining Techniques and Methods ", Advances in Computing and Information TechnologyAdvances in Intelligent Systems and Computing Volume 178, pp 107-117 , 2013.

[47] "Markov chains", [Online]http://en.wikipedia.org/wiki/Markov_chain [Nov 6 2013]

[48] P. Nancy, R. G. Ramani, "Discovery of Patterns and evaluation of Clustering Algorithms in SocialNetwork Data (Face book 100 universities) through Data Mining Techniques and Methods ",International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, No.5, PP 71-82, 2012

[49] P. Nancy, Dr. R.G. Ramani: A Comparison on Performance of Data Mining Algorithms in Classification of Social Network Data", International Journal of Computer Applications (IJCA),32(8): 47-54, October 2011. Published by Foundation of Computer Science, New York, USA.

[50] A. Khan, Y. Wu, X. Yan, "Emerging Graph Queries In Linked Data" Avaialbe : http://www.cs.ucsb.edu/~arijitkhan/Papers/ICDE12_graph_queries.pdf [Nov 30]

[51] S. Sharma, R.K. Gupta , "Improved BSP Clustering Algorithm for Social Network Analysis",International Journal of Grid and

Distributed Computing ,Vol. 3, No. 3, September, pp 67-76, 2010 .

[52] Z. Markov, D. T. Larose, Data Mining the Web: Uncovering Patterns in Web Content, Structure and Usage, Wiley, 2007.

[53] J. Weng ,E. Lim ,J. Jiang ,Q. He ,"TwitterRank: finding topic-sensitive influential twitterers", WSDM '10 Proceedings of the third ACM international conference on Web search and data mining , pp 261-270 ,2010.

[54] T. H. Haveliwala ,"Topic-sensitive PageRank" , Proceeding WWW '02 Proceedings of the 11th international conference on World Wide Web, pp 517-526 ,2002.

[55] F. Yang ,Z. Xu, S. Li, X. Li ,"Social network mining based on improved vector space model",Proceeding ICIMCS '10 Proceedings of the Second International conference on Internet Multimedia Computing and Service pp 118-121 , 2010

[56] F. Huang, N. Xiao, X. Cheng, R. Xiao, "An Approach to Mining Social Networks in Chat Room",Journal of Computational Information Systems 7:1 (2011) 135-143, Available : http://www.Jofcis.com, 2011

[57] P. Cortez, Data mining with neural networks and support vector machines using the R/rminer tool Published in:· ProceedingICDM'10 Proceedings of the 10th industrial conference on Advances in data mining: applications and theoretical aspects pp 572-583,2010