Exploration of Social Webs Consuming the Practices of Web Mining

T. Prema, B. Namitha[*]

Department of ComputerApplications, Vellalar College of Engineering and Technology
Erode,Tamilnadu, India.

*Corresponding Author: T. Prema

E-mail: premaatkongu@gmail.com,

Abstract

Web 2.0 machineries have carried fresh behaviours of joining publics in social webs for collaboration in various on-line groups. Social web exploration deals with the collaborations between folks by seeing them as nodes of a web (graph) whereas their Relationships are planned as web edges. Learning of such constructions falsehoods on the attaching of dissimilar zones of investigation: sociology, graph theory and data mining. This paper revisions problem survey where exploration of social webs using web mining approaches. Approach and ideas of web mining and social webs exploration will be familiarized and swotted along with a debate about how to use web mining approaches for social webs exploration. This paper also sets out a procedure for social web exploration using web mining. Judgement of social webs with other webs is also deliberate. Deliberations of the encounters and future investigation are also involved.

Keywords: Web Mining, Social webbing, Social web Exploration, Association Rule

1. Introduction

Social webs exploration is an exciting investigation direction to appraise the constructions and relationships of social webs, such as appraises of density, importance and groups in social web constructions. In recent years, on-line social webbing has become a very prevalent solicitation in the age of Web 2.0, which allows users to attach, interact and share on the WWW.

Social webs, such as Facebook and Whatsapp, are fundamentally online clusters that allow users to come together, attach and share effects such as photographs, music or other records; and, most frequently, to create short mails, often in the elegance of a mobile phone text memo but shared among a group.

Individuals use the sites to ask their friends queries, say how they impression today and what they are up to, to remark on somewhat they have seen on someone's page. These sites are typical of the internet today in developing what are known as „Web 2.0 skills. Recently, the websites that are formed based on the idea of web 2.0

have becoming the main stream in WWW particularly for those social webbing websites, such as Blog, friends making website, web album, etc.

A social web is the web of relationships and collaborations among social entities such as folks, groups of folks, and organizations. Since the increase of Internet and the World Wide Web has allowed us to investigate large-scale social webs, there has been rising interest in social web exploration. A social web is usually shaped and constructed by daily and continuous communication between persons and therefore includes dissimilar relationships, such as the positions, between's and closeness among folks or groups.

In order to understand the social constructions, social relationships and social behaviour's, social web exploration therefore is an essential and essential technique. Investigation on social webs could be sketched back to sociology, anthropology and epidemiology. Initially, the studies of social webs exploration absorbed on small groups and small social webs.

However, it became harder and harder to manually appraises very broad social webs. Therefore, strong calculation power and information technology has become a very essential tool for social webs exploration and the direction of the investigation is therefore now moving from sociology to computer science.

For social webs exploration, the exploration targets are mainly absorbed on resources from the web, such as its content, constructions and the user behaviours. Application of data mining approaches to the World Wide Web, referred to as *Web mining*, can be used for the exploration of social webs. In web mining, main exploration aims are from the World Wide Web, in the form of web content mining, web constructions mining and web usage mining. The construction of this paper is organized as below:

In section 1, the introduction will be presented. Some related writings of social web exploration, the catalogue and approaches of web mining will be studied in section 2. Section 3 presents social web representations. In section 4, a study of how web mining approaches can be used for social webs exploration will be included. In section 5, a general technique for applying these approaches will be planned. In section 6, a comparison of various kinds of social webs with other webs is reviewed. In section 7, there is a conversation of the challenge of using web mining for social webs exploration and some proposals for future investigation guidelines.

2. Literature Review

In this section, related prose about social web exploration, the catalogue and approaches of web mining will be reviewed

2.1. Social webs Exploration

A social web is a social construction made up of folks (or organizations) called nodes which are tied (attached) by one or more specific types of interdependency, such as friendship, relationship, common interest, financial exchange, dislike, relationships of beliefs, knowledge. Most generally, social web exploration fetishizes social constructions as a web with ties fastening followers and channelling resources, absorbs on the features of ties

rather than on the features of the separate followers and views collections as „personal collections", that is, as webs of separate relations that folks foster, maintain, and use in the sequence of their daily lives. Social web Exploration (SWE) is used to evaluate the interpersonal relationships within a group or public and can provide rich and methodical descriptions and clarification of complex social relationships. SWE absorbs on the inter attractions of the actors, instead of on the individualities of the actors themselves.

2.1.1 Mutual concepts in social web exploration:

(i) Ties: Ties or links attach two and more nodes in a chart. Much human behaviour, such as advice seeking, information-sharing, and giving money to somebody are focused ties while co-followership are examples of undirected ties.

(ii) Between's: The range to which anode lies between other nodes in the web. This measure takes into account the attractive of the node's neighbours, giving a sophisticated value for nodes which bridge groups.

(iii) Uniqueness: The measures of uniqueness identify the most prominent actors, especially the star or the "key" players, that is, those who are extensively involved in relationships with other web followers. The most essential criticality measures are: Degree criticality,

(iv) Closeness: The degree a separate isnear all other folks in a web (directly or indirectly). It reflects the ability to access information through the "grapevine" of web followers. Thus, closeness is the opposite of the sum of the direct distances between each individual and every other person in the web.

(v) Clique: A clique in a graph is a sub-graph in which any node is directly committed to any additional node of the sub-graph.

(vi) Clustering co-efficient: A measure of the probability that two connections of a node are contacts them. A higher clustering coefficient indicates a greater 'cliquishness'.

(vii) Cohesion: The degree to which actors are attached straight to each other by cohesive promises. Groups are identified as clique if every separate is directly tied to every other separate, social circle if there is less strictness of direct contact, which is imprecise, or as anatomically cohesive blocks if exactness is wanted.

(viii) Density: Density is a measure of the closeness of a web. Given a number of nodes, the more associates between them, the larger the density. Its formal definition is as follows. If the number of nodes in a web is N, and the number of links L, then the density is,

$$D=2L/N*N-1 \qquad \textit{for directed graph}$$
$$D = L/N*N-1 \qquad \textit{for undirected graph}$$

(ix) Path distance: Nodes or actors maybe straight devoted by a line, or they may be secondarily devoted through an order of lines. An order of lines in a graph is a "walk", and a walk in which each point and each line are distinct is called a path. The length of a path is dignified by the number of lines which makes it up.

(x) Reach: The degree any associate of a web can reach other followers of the web. In social webs exploration the main job is typically about how to citation social webs from dissimilar communication resources. The data used for building social webs is relational. Data, which can be attained from dissimilar resources including the web, email

communication, Internet relay chats, telephone communications, etc. For example, email communications are a rich source for removing and building social webs. By means of email social webs extraction, the relationship between email sources and destinations can be altered by measuring the frequency of email communication and taking the communication performance (such as reply, forward, etc.) into account. In additional to social web extraction, there are other amounts that can be used for social web exploration as well. For example, degree criticality in a social web is used to measure the between's and closeness of the social web. The density measurement can be used to appraise the attachivity and the degree of nodes and links in a social web.

2.2. Web Mining

Web mining is an application of data mining, the method of discovering and extracting useful information from large data groups or records. Web mining therefore can be defined as to discover or extract useful information from the web.

Dissimilar Types of Web Mining The web mining approaches can be separated into three dissimilar types, which are

> *Web Content Mining,*
> *Web Structure Mining*
> *Web Usage Mining*

2.2.1. Web Content Mining

Is a web mining technique to examine the contents in the web, such as texts, graphs, graphics, etc. Recently, most of web content mining investigations are absorbed on the text data procedure and few are absorbed on other multimedia data. Natural language procedure (NLP) is therefore the main technology that used in this area. The idea and approaches of Semantic Web and Ontology are also relevant here.

2.2.2. Web Structure Mining

Is a system that can be used to examine the links and buildings of web sites. Graph theory is usually the main idea and theory for web buildings mining to examine and explain the buildings of websites. In addition, the extraction of the buildings of websites is always important in this investigation area. The main concern in this area is how to design and implement a crawler to extract and build the buildings of websites, such as in the examination area of Deep-web.

2.2.3 Web Usage Mining

Is a web mining procedure that can be used to examine how the websites have been used, such as the navigation performance of user? The server-side click stream data is the main sources that used for web usage mining. Client-side data (such as client-side is occasionally to be used due to some study concerns, such as in order to record more complete behaviour of users. Dissimilar web usage mining appraises include basic statistical exploration of the navigation behaviour of users in a website: such as how many times the website has been browsed, where the users come from, etc.

Fig.1. Catalogue of Web Mining

2.3. Web Mining Approaches

Out-dated data mining methods can also be used for web mining, such as organisation, clustering, Relationship rule mining, and imagining. In web mining, organisation procedures can be used to classify users into dissimilar classes according to their browsing behaviour, for example according to their browsing time. After ordering, a useful classification rule like "30% of users browse product/food during the hours 8:00-:900 AM" can be exposed. The criterion by which items are allocated to dissimilar clusters is the degree of parallel among them. The main purpose of grouping is to maximize both the likeness of the items in a cluster and the variance between clusters .The Relationship rule system can be used to specify pages that are most often referenced together and to determine the direct or indirect connections between web pages in users browsing behaviour.

3. Social Web models

3.1. Consuming official approaches to show Social webs

One reason for using scientific and graphical approaches in social web exploration is to represent the similes of webs compactly and scientifically. A related reason for using official approaches for representing social webs is that scientific representations allow us to apply PCs to the exploration of web statistics.

3.2. Consuming Charts to Represent Social Relations

Web exploration uses one kind of graphic display that contains node to represent performers and edges to represent ties. When sociologists lent this way of graphing things from the statisticians, they renamed their graphs as "sociograms". Imagining by displaying a sociogram of graph theoretical ideas provides a first explanation of social web data. For a small graph this may suit, but usually the data and/or enquiry questions are too complex for this moderately simple method.

3.3. Consuming Matrices to Signify Social Relatives

The most common form of matrix in social web exploration is a very simple one poised of as many rows and columns as there are performers in our data set, and where the elements represent the ties between the performers. The modest and most mutual matrix is binary. That is, if a tie is current, a one is entered in a cell; if there is no tie, a zero is arrived. This kind of a matrix is called an "adjacency matrix" (Figure2). Fig.2 Using Matrices to Represent Social relations.

	Bob	Carol	Ted	Alice
Bob	---	1	0	0
Carol	1	---	1	0
Ted	1	1	---	1
Alice	0	0	1	---

Fig.2 Using Matrices to Represent Social relations

4. Web Mining Approaches for Social Web Exploration

4.1. The Three Web Mining Types for Social webs Exploration

Web Content mining, text mining or natural language procedure are very valuable in social web exploration. For example, Web content mining can classify documents on an on-line social webbing website, particularly courses on blogs or text forums. Article categorization is typically the first task for many socil webs appraises or applications. Web content mining can also be used in social webs exploaration to examine users reading interests, and determine their favourite content. However, for most on-line social webs exploration tasks, it is usually necessary to utilize all three types of web mining and styles composed.

Web usage mining also shows a crucial role in social webs exploration. It is valuable for the exploration of social webs extraction deliberated in section 2 of this paper. The usage data and user communications on an online social webbing website can be altered into relational data for social-webs building. In addition, web usage mining is also a tool for measuring importance degree. For example, the closeness of blog users can be measured *by,*

$$Closeness: (f_c*(w_c*l_c)) \qquad (f_c*(w_c*r_c)) \qquad (f_c*(w_c*i_c)) \quad ... \text{ Equ. (1)}$$

In the equation Equ.(1) above, w_c is the weight of closeness for each b f_c denotes the frequency of a blog behaviour. The three blog behaviours are

$$l_c=looking, r_c=reading \text{ and } i_c=collaborations.$$

This is just a simple example of web usage mining, but the approaches allow many possible means of social webs exploration. Web buildings mining is the third kind of web mining and it is also useful for extracting and building social webs to extract the links from WWW, e-mail or other sources. It can also be used to analyze path length, reach ability or to find structural holes, which are very basic and traditional social webs appraises. Web buildings mining usually uses graphs and imagined means to represent the data about social webs, allowing the analyst to easily know and analyze social webs.

4.2. Web Mining Approach for Social webs Exploration

There are many other dissimilar kinds of web mining approach. In this section, samples will be given of two web mining approach used for social webs exploration. The two approaches are clustering and Relationship rule mining. *Grouping:* In social webs exploration, finding the group of nearby individuals in a web is usually the main

task, and is normally completed by using a imagining method in a small social web. Thus, the grouping system can be used with a large social web in finding more groups and clusters. Moreover, it can provide more detailed info than just using imagining can, including the closeness of a group, detailed facts of the followers in a group and the relationship between groups in a social web.

4.3. Association rule mining:

In social web exploration, Association rule mining can help notice the hidden relationships among nodes in a social web or even cross webs.

5. The Technique of Web Mining For Social Web Exploration

A general technique to use web mining for social webs exploration is shown in figure 3. Its steps contain selection of exploration goals, selection of social webs exploration, data preparation, web mining approach collection, results presentation and interpretation, recommendation and achievement.
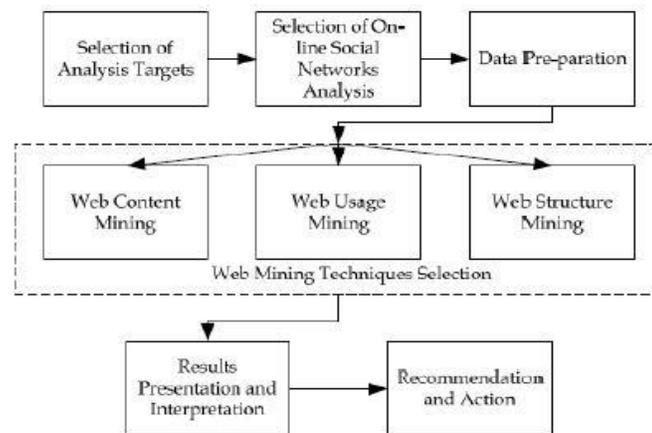

Fig.3 The general procedure of social web exploration

The Initial step is the selection of the exploration targets, such as web, email, telephone communication, etc. Sometimes, more than one target will be designated. After, we then hand-picked what of kind social webs exploration we will proceed with. Once the survey targets and the social webs survey approach have been selected, next is data preparation. In this stage related data will be together for examination, then cleaned and formed as the final plan to store in folder or database. Next is selecting the web mining method to be used and then scheduled with them. More than one system may be selected and sometimes a grouping of approach is needed. The selected suitable approaches are then used to analyze the facts collected and organised in the third step of the procedure.

The results of the exploration after web mining are then offered and construed either manually or mechanically. Conception approaches are sometimes used to assist the performance of the results of the exploration, such as the extracted social webs.

Web content mining is differentiated from two different points of view: Information Retrieval View and Database View. Summarized the research works done for unstructured data and semi-structured data from

information retrieval view. It shows that most of the researches use bag of words, which is based on the statistics about single words in isolation, to represent unstructured text and take single word found in the training corpus as features.

For the semi-structured data, all the works utilize the HTML structures inside the documents and some utilized the hyperlink structure between the documents for document representation. As for the database view, in order to have the better information management and querying on the web, the mining always tries to infer the structure of the web site to transform a web site to become a database.

Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site. Web usage mining itself can be classified further depending on the kind of usage data considered:

1. Web server data,
2. Application server data,
3. Application level data.

Web structure mining Terminologies:

1. Web graph: directed graph representing web.
2. Node: web page in graph.
3. Edge: hyperlinks.
4. In degree: number of links pointing to particular node.
5. Out degree: Number of links generated from particular node.

The last step of the general process to use web mining for social webs exploration is reference and action. This is an optional step, and the procedure may be ended after the exploration results have been made.

The reference and action step deals with the results of the exploration; for example, if buildings holes in a social web have been exposed references about how to fill the hole may be manually or mechanically made and then suitable action can be taken. This general process of using web mining for social webs can be a non-stop procedure. In some study projects, it starts again after references have been made and action occupied. The procedure starts again to do the action presentation evaluation or to do a new exploration.

PROS –

Web usage mining essentially has many advantages which makes this technology attractive to corporations including the government agencies. This technology has enabled e-commerce to do personalized marketing, which eventually results in higher trade volumes. Government agencies are using this technology to classify threats and fight against terrorism.

6. Contrast of Several Factors of Social Webs

6.1. Contrast of Social web with Extra Webs

Kind of Web	Idyllic Definition	Operative Measure	Content of relation/Link
Social web	A set of persons committed by a set of relationships	Separate, Group, Organization, Nation-State	Any Kind of Social Relation
Communication Web	A web collected of inter attached folks linked by spotted flows of info	Similar as above, but generally absorbs on separate individuals	Communication and Information
Internet Web	A communication web involved by the Internet among computer systems	Same as above, but absorbs on Internet users	Same as above, but limited to Internet as channel of info flow

6.2. Decision and Forthcoming Investigation

Social Network Analysis is the study of social structure. The social network analysts are interested in how the individual is inserted within a construction and how the structure occurs from the micro-relations between separate parts. This paper revisions the application of the concept and methods of web mining for social networks analysis, and evaluations the related works about web mining and social networks analysis.

Furthermore, how to custom web mining and a general process of by web mining for social networks analysis have also been studied. Social networks analysis approved out using the methods of web mining is an interesting area of study. But, there are several challenges in this research area to be overcome. For example, data sampling is one of the big problems. One more challenges include finding people in social networks, finding designs in social networks and analysing covering society.

The future research focus to overwhelmed the challenges deliberated above, such as the how to reduce the data size etc. In addition, we will focus to apply the web mining methods to websites, such as blogs and on-line photo albums.

References

[1] Adamic L.A and Adar,E."Friends and Neighbours on the web" Social networks,vol.25,2007,pp,211-130.

[2] Joshi .A,and Joshi.k "On mining web acess logs",in Technical Report,CSEE,Department,UMBC,1999.

[3] Cai,D.,Shao.Z.,He.x,Yan.X,and Han.J "Mining Hidden Community in Hetrogeneous Social Networks" in Processing of LinKDD 2005 Conference, August 2005,Chicago,USA,pp.558-567.

[4] Linton Freeman, The Development of Social Network Analysis. Vancouver: Empirical Press, 2006.

[5] Kilduff, M., Tsai, W. (2003). Social networks and organisations. Sage Publications.

[6] Hanneman, Robert A. & Riddle, Mark (2011). "Concepts and Measures for Basic Network Analysis". The Sage Handbook of Social Network Analysis. SAGE. pp. 364–367. ISBN 978-1-84787-395-8.

[7] Xu, Guandong et al. (2010). Web Mining & Social Net: Techniques and Applications. Springer. p. 25. ISBN 978-1-4419-7734-2.

[8] Tsvetovat, Maksim & Kouznetsov, Alexander (2011). Social Network Analysis for Startups: Finding Connections on the Social Web. O'Reilly. p. 45. ISBN 978-1-4493-1762-1.

[9] Hansen, Derek et al. (2010). Analyzing Social Media Networks with NodeXL. Morgan Kaufmann. p. 32. ISBN 978-0-12-382229-1.

[10] De Nooy, W (2003). "Fields and networks: Correspondence analysis and social network analysis in the framework of Field Theory". Poetics 31: 305–27. doi:10.1016/s0304-422x(03)00035-4.

[11] Baraglia, R. Silvestri, F. (2007) "Dynamic personalization of web sites without user intervention", In Commu of ACM 50(2): 63-67

[12] Dr.K.R.Ananth, Dr.P.Navaneetham, "Modified Mutliresolution Medical Image Segmentation Based on Wavelet Transform", International Journal of Computer Science and Mobile Computing, Vol. 3, Issue. 6, June 2014

[13] Bing Liu, "Web Data Mining: Exploring Hyperlinks, Contents and Usage Data", Springer, 2007

[14] Web Mining and Web Usage Analysis 2004 - revised papers from 6 th workshop on Knowledge Discovery on the Web, Bamshad Mobasher, Olfa Nasraoui, Bing Liu, Brij Masand, Eds., Springer Lecture Notes in Artificial Intelligence, 2006

[15] Advances in Web Mining and Web Usage Analysis 2005 - revised papers from 7th workshop on Knowledge Discovery on the Web, Olfa Nasraoui, Osmar Zaiane, Myra Spiliopoulou, Bamshad Mobasher, Philip Yu, Brij Masand, Eds., Springer Lecture Notes in Artificial Intelligence, LNAI 4198, 2006