

**International Journal of Advanced Research in Computer Science** 

**RESEARCH PAPER** 

Available Online at www.ijarcs.info

# Web Mining Concept on Social Network Analysis

Varsha Jotwani Assistant Professor of computer science SHGC, MP India Umika Jotwani Khatarkar Assistant Professor of Physic Bangalore,Karnataka India

*Abstract:* The purpose of the research paper is to analyze the features of users in social networking Web sites as well as the related contents of the Web sites. The social networking techniques has been used and related web mining. The analysis results will then be used to construct a recommendation mechanism and system and its impact on the society. Furthermore, we find out the social networking analysis by using web mining feature into it.

Keywords: World Wide web, Social Networking analysis, Web mining

## I. INTRODUCTION

World wide web has become a very powerful platform which present information and retrieve information as well as analyze information now a days a web has known a huge repository which consist of various types of data types as well as knowledge base which is hidden in web pages, users are often facing lots of problem regarding as information are increasing day by day and there is rapid growth of numbers of users too [1]. Social network seems to be the traditional topics which cover various features like physcology, sociology and behavioral science .Web social network analyzing the temporal changes of web communities and capturing the dynamic evolution pattern of web networked structure and an empirical study of identifying social sense from a web log archieve, Recent advances in storage property has opened avenues for a series of large web archieve.It has now increased a various challenges for the evolution of web and simultaneously increased the dynamic growth and dynamic changes in its structure .A Web community is the collection of web pages which was created by the individual and the association with a common interest on topic.

## II. EVOLUTION OF SOCIAL NETWORK ANALYSIS

Social network analysis is a key technique in modern sociology, anthropology, sociolinguistics, geography, social psychology, communication, studies, information science, organizational studies, economics, and biology as well as a popular among forecasting and study. It is Analyze the structure of a social network to infer knowledge about an individual or a group. Graphical way to represent relations. Graphs is a way that is mathematical structure that is used model pair wise relations between objects[2].

- a. Nodes- to represent objects
- b. Edges -to represent relations

To represent only symmetric data we uses undirected graph and to represent both symmetric and asymmetric graph we use directed graph and to represent intensities, cost and distance we use weighted graph.Static graph represent at a given point in time.Dynamic graph properties observed over a period of time.

## III. CONCEPTS OF SOCIAL NETWORK ANALYSIS

Social network analysis(SNA) brings an interpersonal relationship among organization and the communities and can provide rich and innovative description among these communities.SNA focuses on the interconnection between the nodes [3]

Concepts in Social Networking analysis:

- *a. Ties:* Ties or links connect two or more nodes in a graph.
- **b. Betweenness:** The extent to which a node lies between other nodes in the network. This measure takes into account the connectivity of the node's neighbors, giving a higher value for nodes which bridge clusters.
- *c. Centrality:* The most important centrality measures are: Degree centrality, Between-ness centrality and Closeness centrality.
- *d. Association:* The degree an individual is near 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 the sum of the shortest distances between each individual and every other person in the network
- e. Clique: A clique in a graph is a sub-graph in which any node is directly connected to any other node of the sub-graph.
- *f. Clustering coefficient:* A measure of the likelihood that two associates of a node are associates themselves. A higher clustering coefficient indicates a greater 'cliquishness'.
- *g. Path length:* Nodes or actors may be directly connected by a line, or they may be indirectly connected through a sequence of lines. A sequence 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 measured by the number of lines which makes it up. (x) Reach. The degree any member of a network can reach other members of the network.

#### IV. WORLD WIDE WEB IMPACT

# A. Democratisation + sharing of ideas = more accountable governance):

The internet has opened the ways we communicate both as individuals and groups. Communities are no longer limited by their location and this has had a profound impact on the exploration of ideas. Nowhere is this more visible than the democratizing impact social media has had on political activism. Beyond media outlets, social is the new medium for discovering information, especially from those at the forefront of major crises. This democratization will directly influence the way in which we better inform and shape personal opinions about issues [4][5].

# **B.** Accessibility+Vast quantity of information = greater Social mobility:

The notion of the existential millennial is a product of the internet. Young people are now acutely aware of their desire to make an infuence and this is a by-product of the proliferation of knowledge through the internet. The internet celebrates knowledge unashamedly and this will continue to empower people to expand their understanding of the world, helping refine their abilities to make the impact they desire. Web mining and its impacts on social networking [4][5].

#### V. WEB MINING

We make use of the web in several ways and extract information from the data mining techniques [6] and we interact with the web for the following purposes.

#### a. Finding Relevant Information:

We either browse or use the search service when we want to find the specific information on the web .we usually specify a simple keyword query and the response from a web search engine is a list of pages ,rank based on the their similarity to the query. However today's search engine have the following problems:

- a) Low accuracy: This is due to irrelevances of many search result
- b) Low recall: This is due to the inability to index all the information available on the web

#### b. Exploring Knowledge from the Web

We can call this term as data triggered process that presumes that we have already have a collection of web data and want to extract potentially useful knowledge out of it.

#### c. Personalized Web page Synthesis:

We may wish to synthesize a web page for different individuals from the available set of web pages .Individual have their own preferences in the style of the contents and presentation while interacting either the web. The information provider like to create a system which responds to user queries by potentially collecting information from several resources, in a manner which is dependent on the user.

#### d. Finding about Users needs:

Web mining techniques provides a set of techniques or we may call it a solution that can be used to solve the various problems related to know about the users and its interest such as mass customizing the information to the individual users ,the other problems which are related to the effective web site design and management ,problems related to marketing etc

#### A. Web Content mining:

Web content mining describe the useful information from the web content .We can observe in the recent years so many government organization are putting their information on the web pages and their was vast amount of increase in placing the digital library on the web sites. We cannot ignore the web applications so that user could easily access the application through web interfaces .Many application and systems are being migrated to the web and many types of information are emerging in the environment. A huge amount of multimedia data which is application of data mining concepts and techniques [6] are available in a web which include such vedio, sounds, graphs, images etc which are very helpful to retrieve multimedia information from it Web-based multimedia data are inserted on the Web page and are linked with text and link information. In a contentbased image retrieval system, there are often two kinds of queries: image sample-based queries and image feature specification queries. Image-sample-based queries find all of the images that are similar to the given image sample. This search compares the feature vector extracted from the sample with the feature vectors of images that have already been extracted and indexed in the image database. Based on this comparison, images that are close to the sample image are returned. Image feature specification queries specify or sketch image features like color, texture, or shape, which are translated into a feature vector to be matched with the feature vectors of the images in the database. Content-based retrieval has vast applications which are applied to the medical diagnosis, weather prediction, TV production, Web search engines for images, and e-commerce.

The textual parts of web content data consist of unstructured data such as free texts, HTMLS documents which semi structured data such as and more structured data such as data in a tables or database generated HTML pages .It is quite necessary to understands the categories and its subcategories in which documents belongs to the organization of categories, maximum number of categories per document in which we count dimension like time, subjects and locations.

From preprocessing content [7] from content preparation can be extracted from HTML, perform stemming and removing stop words, collect worldwide frequencies and by calculating per document frequencies.

#### B. Web Structure Mining:

Web structure mining is concerned with discovering the model underlying the link structure of the web It is used to study the topology of the hyperlinks with or without the description of the links this model can be used to arrange web pages and it generate important information such as common factors among different websites they are used to discover authority sites for the subjects and overview sites for the subjects that point to many authorities. The main aim of the web content mining to focuses on the structure of the inner document while web structure mining tries to discover the link structure of the hyperlinks at the inter document level .The structure of the typical web graphs consist of web pages as nodes and hyperlink as edges connecting between two related pages .It is a method of finding out the structure information from the web and this type of mining can be done at the document level or at the hyperlink level .The analysis which could be performed at the hyperlink level is called hyperlink analysis. This review of trends in research of web mining by Manoj Pandila [8] has given a brief idea about web structure mining. Web mining has a web graph which presents the web, it consist of node, link,in-degree.

- a. Node –Each web page is a node of the web graph
- b. Link –Each hyperlink on the web is a directed web graph
- c. In-degree- The in-degree of a node, p is the number of distinct links that point to p.
- d. Out-degree: The out degree of a node ,p,is the number of distinct link originating that point to the other nodes

Web page consist of directed path which has a direction pointing from 'p' to 'q', whereas the shortest path indicate all the paths between nodes p and q which has the shortest length that is numbers of link on it and at last the diameter which represent the maximum of all the shortest paths between a pair of nodes p and q for all the pair of nodes p and q in the web graph. Some algorithm have been proposed to model web topology such as HITS, Page Rank, Clever and these method is adopted to calculate the relevancy and the quality rank of each web page.

#### C. Web Usage mining:

Web usage mining[5] deals with studying the data generated by the web users session and their patterns mines the secondary data extract from the interaction of the user with the web. These kind of secondary data includes the data from the web server logs ,browser logs, proxy server logs, user profiles registration data ,user session or transaction, cookies, user queries, bookmark data mouse clicks and scrolls and any other data which are result of these interactions ,the goals of these to analyze the discovered pattern patterns and profiles of the users are usually represents as collection of pages ,objects that are randomly used by the users. This data can be collected by the web servers. Analyses of the web access logs of variety of the websites can be helpful for understanding the patterns and the web structure, improving the design of this collection of the information. There are two main approaches in web usage mining driven by the applications of the discoveries.

- a. General Access pattern Tracking: This method analyze the web logs to understand the patterns
- b. Customized Usage Tracking: this tracking helps in finding out the patterns of individual users. The pattern which has been discovered are usually represented as groups of pages, objects or resources that are randomly accessed by the groups of users with common interest factor.

#### VI. DESCRIPTION OF MODEL

Many complex network model study the question of whether the structure observed in social network could be explained by the network dependent interaction of nodes without intrinsic properties of nodes. Such models are based on the assumption about the local mechanism of formation such as people meeting friends of friends and thus forming a connection with other network neighbor [9].Following are the model which are discussed below:-

#### A. Nodal Attribute model (NAMs):-

We adopt the term nodal attribute model in which the probability of edge eij of nodes i and j being present explicitly stated as function of attributes of the nodes i and j only and the evolutionary aspect is absent.NAM are often based on the concept of homophily the tendency for like to interact like which is known to structure network ties of various types including friendship, marriage work information transfer and other form of relationship[10][11].

#### B. Exponential Graph Random Model (ERGMs):

Exponential random graph models (ERGMs) [12] also called p\* models, are used to test to what extent nodal attributes (exogenous factors) and local structural dependencies (endogenous factors) explain the observed global structure. Consider a random graph **X** consisting of N nodes, in which a possible tie between two nodes *i* and *j* is represented by a random variable Xij, and denote the set of all such graphs by  $\alpha$ .

#### VII. CONCLUSION

World Wide Web has the huge information of knowledge which contains a massive data which is not exploited properly, Web user having difficulty in finding the desired information properly due to two problems low precision and low recall.Web users to locate the needed information accurately and promptly and making the Internet use as friendly and conveniently as possible is always kept as a first priority research question to Web research academia and industrial partners. Analyzing the social networking data is able to present the underlying social and societal characteristics of human behavior and real world structure. Thus, the emerging of Web has put forward a great deal of challenges to Web researchers for Web-based information management and retrieval. Above are the models which helps in finding out the possible way for social networking, Such models are based on the assumption about the local mechanism of formation such as people meeting friends of friends and thus forming a connection with other network neighbor.

#### **VIII. REFERENCES**

- Lin li, Zhang Yanchun Guandung Xu," Web Information system Engineering and Internet Technologies Book Series"pp.3-4
- [2]. Hilbrich Robert, Department of Computer Science Humboldt University at Berlin, "Social Network analysis using graphs Metrics of web based social network", November 27, 2007,pp.4-8
- [3]. Raju.E,Sravanthi.K Department of Computer Science and Engg.SR Engineering, Warangal, "Analysis of social networks using the techniques of web mining" International journal of Advanced Research in computer Science and Software Engineering Volume 2,Issue 10 Oct 2012,pp.443-445
- [4]. G. Plickert, R. Cote, and B. Wellman. "It's Not Who You Know. It's How You Know Them: Who Exchanges What with Whom?" Social Networks, Vol. 29, No. 3, 2007, pp. 405–429.

- [5]. bolaky keshan,"The social impact of World Wide Web" years 12<sup>th</sup> march 2014
- [6]. Han Jiawei, Kamber Micheline "Data Mining: Concepts and Techniques" Second Edition, University of Illinois at Urbana-Champaign, pp. 607-608
- [7]. Srivastava Jaideep,"Web Mining :Accomplishments & Future Directions" University of Minnesota USA ,pp.2-34.
- [8]. Pandia Manoj, kumar Pani Subhendu, kumar Padhi Sanjay, Panigrahy Lingaraj, R. Ramakrishna, "A review of trends in research on web mining", JJICA, vol 1, Issue 1, 2011, pp 37-41.
- [9]. Toivenen Riita,Kovanen lauri,Kivela Mikko,Pekka Jukka,Jari saramaki Onnela,Kaski Kimmo,"A comparative

study of social networking models:Network evolution Models and nodal attributes models"

- [10]. Boguna M. Pastor Satorass ,"Epidemics spreading in complex correlated networks", Physical Review E-66 047 104
- [11]. McPherson, M., Smith-Lovin, L., Cook, J., 2001." Birds of a feather: homophily in social networks". Annual Reviews of Sociology 27", pp. 415–444.
- [12]. Frank ,O. Strauss,D, 1986 "Markov graphs", Journal of the American statistical Association(JASA)81(395),832