



---

## **WEB DATA MINING TECHNIQUES FOR ON-LINE SOCIAL NETWORKS: A STUDY**

**Tarminder Singh<sup>1</sup>, Dr. Vinod Sharma<sup>2</sup>**

**Department of Computer Science**

**<sup>1</sup>Shri Venkateshwara University, Gajraula (Uttar Pradesh)**

**<sup>2</sup>University of Jammu, Gujrabasti, Jammu (Jammu and Kashmir)**

### *Abstract*

This article main focussed on web data mining techniques for on-line social networks. Social networks analysis is an intriguing exploration direction to break down these structures and connections of social networks, for example, the investigations of thickness, centrality and factions of social network structure. A social network is typically framed and developed by day by day and constantly correspondences of individuals and it in this way incorporates various connections, for example, the positions, between and closeness among people or gatherings. So as to comprehend the social structure, social connections and social practices, social networks analysis hence is a fundamental and significant method that must be considered. As of late, on-line social networking is a hot and well-known application in the period of web 2.0, which enables user to convey, cooperate and share in the World Wide Web. Some on-line social networking websites presently even become the most mainstream sites on the web.

### **1. OVERVIEW**

The historical backdrop of social networks analysis is over hundred years from around 1900's, and for the most part in the examination areas of human science. Amid this period, the investigations of social networks analysis were concentrating on little gatherings and little social networks. Be that as it may, it ends up increasingly tough to investigate physically for those wide social networks, for example, the World Wide Web. Along these lines, the solid PC capacity and information technologies has turned out to be significant device for social networks analysis and the inquiry direction is in this manner presently moving from humanism to software engineering. For on-line social networks analysis, the analysis targets are mostly centered on resources from the web, for example, the substance of the web, the structures of the web and the utilization practices of users in the web. Among the information techniques that can be utilized for the analysis of on-line social networks, web mining is professed to be the most reasonable one[1-3].



---

Web mining is an application of Data Mining and the analysis focuses of web mining are for the most part from the World Wide Web, for example, web content mining, web structure mining and web utilization mining. In the examination area of social networks analysis, it is generally the fundamental errand about how to separate social networks from various correspondence resources. The data that utilized for structure social networks is social data, which can be gotten and exchanged from various resources including the web, email correspondence, web transfer visits, phone interchanges, association and business occasions, and so on. For instance, the email correspondence is a rich hotspot for removing and developing social networks[4, 5].

In the issue of email social networks extraction, the connection between email senders and recipients can be changed by estimating the recurrence of email correspondence with take the correspondence conduct, (for example, answer, forward, and so forth.) into record. The changed social data would then be able to be utilized for social networks development. In the previous three decades, social network analysis has built up a scope of ideas and methods for detecting auxiliary patterns, recognizing patterns of various sorts of relationship interrelate, breaking down the suggestions that basic patterns for the conduct of network individuals, concentrating the effect on social structures of network individuals and their social connections[6-8].

## **2 TYPES OF SOCIAL NETWORK ANALYSIS**

A social network has a set of relations of ties, which can be viewed in two different ways. One approach focus on an individual, called ego-centered network, and put it at the centers of the network. Members of the network are defined by the relations with the ego. Ego-centered network analysis can show the range and breadth of connectivity for individuals and identify those who have access to diverse pools of information and resources. The ego-centered approach is useful when the population is large, or the boundaries of the population are hard to define. The second approach considers the whole network based on some specific criterion of population boundaries such as a formal organization, department, club or kinship group. Whole network analysis can identify those members of the network who emerge as central figures or who act as bridges between different groups. This approach requires responses from all members on their relations with all others in the same environment, such as the extent of email and video communication in a workgroup. Network analysis provides a rich and systematic means of assessing such network by mapping and analyzing relationships among people, teams, departments or even the entire organization. A network is composed of three elements:



- (1) Actors
- (2) Relations between actors, and
- (3) The linkages among actors.

Actors and their actions are viewed as interdependent rather than independent, autonomous units. Actors can be persons, organizations, or groups, or any other set of related entities. Relations between actors are depicted as links between the corresponding nodes. A tie connects a pair of actors by one or more relations. Pairs may maintain a tie based on one relation only or a multiplex tie based on many relations. Thus, ties also have characteristics like content, direction and strength, but they are often referred to as weak or strong.

Social network analysts have found that multiplex ties are more intimate, voluntary, supportive and durable. In addition, the linkages among actors have several characteristics, which are direction, degree, and content. The direction of linkages covers symmetrical and asymmetrical relations; the degree of linkages means the strength of relations, and the content of linkages includes friendship, information, power, and influence, etc. Owing to complex properties of nodes, relations, and linkages, scholars utilizing the concept of network in their studies have different definitions of network.

### **SNA techniques**

Visualization is additionally an intriguing issue of social network analysis, and it is a reasonable technique in this area. Through the visualization of social networks, the characters of social networks can be seen effectively, for example, the structure of networks, the circulation of nodes, the (connections) among nodes and bunches and gatherings in the social network. In extra to social network extraction and visualization, there are different estimations that can be utilized for social network analysis too. For instance, centrality level of a social network is an estimation that is utilized to gauge the between and closeness of the social network. Between centrality demonstrates the degree to which a hub lies on the briefest way between each other pair of nodes. Closeness centrality dissects centrality structure of a network based on geodesic separations among nodes in a social network. Group coefficient is an estimation to find bunches in a social network and to gauge the coefficient of bunches.



## **Web mining**

In the presentation section of this part, a short clarification of web mining has been given. It is an application of data mining, and data mining is a technique to find and concentrate helpful information from enormous data sets or databases. For web mining, the definition in this manner can be disclosed as to find or concentrate helpful information from the web According to various analysis targets and resources; the web mining techniques can be partitioned into three distinct sorts, which are Web Content Mining, Web Structure Mining, and Web Usage Mining. Web content mining is a web mining technique to break down the substance in the web, for example, writings, charts, illustrations, and so on.

## **Web mining techniques**

Traditional data mining techniques can likewise be accommodated web mining, for example, order, clustering, affiliation rule mining, and visualization. In web mining, the arrangement calculations can be utilized to order users into various classes as per their perusing conduct. For instance, an order application characterizes their users as per their perusing time. After order, a helpful characterization rule like "30% of users peruse item/nourishment amid the hours 8:00-10:00 PM" can be found. The contrast among order and clustering is that the classes in the arrangement are predefined (directed), however in clustering are not predefined (unsupervised). The rule by which things are doled out to various bunches is the level of comparability among them. The principle motivation behind Clustering is to boost both the likeness of the things in a bunch and the contrast between groups.

## **3. WEB MINING AND SOCIAL NETWORKS**

A social network is a social structure made of individuals (organizations, company etc.) also called nodes, which are connected by links represent relationships and interactions between individuals, a rich relational interdependency and content for mining. Online social network focuses on building on Internet communities of people who share interests and/or activities, who are interested in exploring the interests, and activities of others or who are interested to communicate, interact and share. So, they are very popular Web 2.0 application. Some well-known social networking websites are: Facebook as general network, LinkedIn and Viadeo as business social network, Flickr about photo sharing etc. Thus, social network is a relevant part of human life.



---

- **Data and Analysis Task**

The data comprise of films and related items including individuals (who act in, produce, or direct the motion pictures), organizations (studios), occasions (arrivals of the motion picture), and different articles (grants). These items are associated in the manners in which that you would expect (e.g., on-screen characters are connected to motion pictures they act in) and in some sporadically sudden ways (e.g., films are connected straightforwardly to different motion pictures that are revamps). Notwithstanding the abnormal state structure of the database, the database contains properties related with each item, including the titles and sorts of motion pictures, the names and times of people, and the nations and film industry receipts of motion picture discharges.

#### **4. TRAFFIC ANALYSES IN SOCIAL MEDIA NETWORK OF TWITTER THROUGH DATA MINING TECHNIQUES**

Social media goes about as a significant job for all around the globe. Social media offer numerous sorts of data, which causes everyone to contact other individuals in public and offer their data. Social media offers, chat, and associate with other identity or with huge onlookers. Numerous individuals utilize these Social networking websites on the Internet every day to interface with a huge number of individuals a portion of the Social networking websites on the Internet are LinkedIn, Facebook, Twitter and Google Plus. At the present time, for every single bit of data partook in these sites like Twitter or Facebook, divider post, sharing a photographs, status, and video, the user who is transferring need to choose to which of these companions or gatherings must be sent. Twitter assumes an imperative job in the general public and regular media.

#### **5. WEB DATA MINING AND SOCIAL MEDIA ANALYSIS FOR BETTER COMMUNICATION**

The utilization of new nourishment that is polluted with pathogens like growths, infections, or bacteria can cause food contamination that prompts extreme wellbeing harms or even demise. As an outcome, actors in the agri-nourishment system are tested to plan quality and wellbeing affirmation plots that are proficient, solid, and globally perfect. This requires authoritative structures, suitable action plans, and technical arrangements that ensure a proficient accumulation and conveyance of information to improve correspondence techniques and to abbreviate the season of basic leadership.



---

## **6 APPLICATION OF TEMPORAL DATA MINING TECHNIQUES**

As of late, social media have encountered gigantic development in their user base. For instance, there are more than one billion individuals having a place with the Facebook network, while Twitter presently has more than 280 million month to month active users. There are an enormous number of various social media applications or stages which when all is said in done can be arranged as weblogs, micro blogs, social network sites, area-based social networks, dialog gatherings, wikis, digital broadcast networks, picture and video sharing stages, evaluations and surveys networks, social bookmarking sites, and symbol-based computer generated simulation spaces.

Late investigations and studies have uncovered a developing need to ceaselessly gather, screen, dissect, abridge, and imagine significant information from social interactions and user-created content in different areas, for example, business, public organization, politics, or purchaser basic leadership. These activities, be that as it may, are viewed as troublesome assignments because of the enormous number of various social media stages just as the huge sum, elements, and multifaceted nature of social media data. All the more explicitly, social media correspondence creates an enhanced and dynamic arrangement of data and Meta data, which have not been dealt with systematically in the data-and content mining writing. Tomoyuki NANNO presents a system that endeavors to consequently gather and screen Japanese blog accumulations that incorporate ones made with blog software's as well as ones composed as would be expected web pages.

## **7. E-LEARNING USING DATA MINING**

E-learning (likewise alluded to as web-based education and teaching), and another setting for education where lot of information portraying continuum of the teaching-learning interactions are unendingly created and universally accessible. This could be viewed as a gift: a lot of information promptly accessible only a tick away. Be that as it may, it could similarly be viewed as an exponentially developing bad dream, in which unstructured information stifles the educational system without giving any well-spoken learning to its actors, Data Mining was destined to handle issues this way. As a field of research, it is practically contemporary to e-learning. It is, however, fairly hard to characterize.



---

## 8. CONCLUSION

A number of Data Mining techniques (such as association, clustering, classification) are developed to mine this vast amount of data. Data mining is a powerful tool that can help to find patterns and relationships within our data. Data mining discovers hidden information from large databases. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. 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. There are various data mining techniques:

**Characterization:** Characterization is used to generalize, summarize and possibly different data characteristics.

**Classification:** Data classification is a process in which the given data is classified in to different classes according to a classification model.

**Regression:** This process is similar to classification the major difference is that the object to be predicted is continuous rather than discrete.

**Association:** In this process the association between the objects is found. It discovers the association between various data bases and the association between the attributes of single database.

**Clustering:** Clustering involves grouping of data into several new classes such that it describes the data and it breaks large data set into smaller groups to make the designing and implementation process to be simple. The task of clustering is to maximize the similarity between the objects of classes and to reduce the similarity between the classes.

**Change Detection:** This method identifies the significant changes in the data from the previously measured values. **Deviation Detection:** Deviation detection focuses on the major deviations between the actual values of the objects and its expected values. This method finds out the deviation according to the time as well the deviation among different subsets of data.

**Link Analysis:** It traces the connections between the objects to develop models based on the patterns in the relationships by applying graph theory techniques.



---

**Sequential Pattern Mining:** This method involves the discovery of the frequently occurring patterns in the data. Social networks are important sources of online interactions and contents sharing, subjectivity, assessments, approaches, evaluation, influences, observations, feelings, opinions and sentiments expressions borne out in text, reviews, blogs, discussions, news, remarks, reactions, or some other documents.

Fortunately, new and sophisticated techniques have been developed in the area of data mining (also known as knowledge discovery), which can aid in the extraction of useful information from the web. Data mining and the Web developed as independent technology areas in the mid-1990s. While it was felt that mining data on the Web would be useful to help the information overload problem, the extent to which Web mining would help key areas such as E-commerce was not well understood until recently.

## REFERENCES

- [1]. Borgatti, S. P., Everett, M. G.: A graph-theoretic perspective on centrality. *Social networks* 28, 466- 484, 4, 2006.
- [2]. Borgatti, S P.: "2-Mode concepts in social network analysis." *Encyclopedia of Complexity and System Science*, 8279-8291, 2009.
- [3]. Asur, S., and Huberman, B.: "Predicting the future with social network." *Web Intelligence Agent Technology (WIAT), 2010 IEEE/WIC/ACM International Conference on*. Vol. 1. IEEE, 2010.
- [4]. Castellanos, M., Dayal, M., Hsu, M., Ghosh, R., Dekhil, M.: U LCI: A Social Channel Analysis Platform for Live Customer Intelligence. In: *Proceedings of the 2011 international Conference on Management of Data*. 2011
- [5]. Adedoyin-Olowe, M., Gaber, M., Stahl, F.: A Methodology for Temporal Analysis of Evolving Concepts in Twitter. *Proceedings of the 2013 ICAISC, International Conference on Artificial Intelligence and Soft Computing*. 2013
- [6]. Becker, H., Naaman, M., Gravano, L.: Beyond Trending Topics: Real-World Event Identification on Twitter. *ICWSM*, 11, 438- 441, 2011.



- [7]. Y. K. Mathur, abhayanand , “soft computing techniques and its impact in data mining”, international journal of emerging technology and advanced engineering (issn 2250-2459, iso 9001:2008 certified journal, volume 4, issue 8, august 2014).
- [8]. Ting, i-hsien 2008, “web mining techniques for on-line social networks analysis.” In proceedings of 2008 international conference on service systems and service management, july 2008, pages 1-5.