

# A Study on Designing User Web-Page Traversal Patterns Methods Using Ann

Poonam Rani<sup>1\*</sup> Dr. M. K. Bisht<sup>2</sup>

<sup>1</sup> Research Scholar, Pacific University, Rajasthan, India

<sup>2</sup> Assistant Professor, Department of Computer Science, SSJ Campus, Almora, UK, India

**Abstract – The availability of all types of huge data and growing size of users for social networking sites, it would seem this area is the perfect environment for extensive data mining, research, and designing new algorithm to improve data mining on various facts. While the usage rates, public availability and uploading of huge amount of data on social media provide us a new area for research, there are a number of impediments to capitalizing on data mining algorithms with best fit strategies for this area. Studying social area networks, growth rates, and social implications of social networking sites is likely to draw strong willing for design and test new approaches of data mining from social area network sites either by owners or by users.**

**Keywords: Web Page, Social Networking Sites, Networks**

-----X-----

## 1. INTRODUCTION

The problem of online shopping market prediction in the new context, namely, the use of social media and social networking. Social media is a new form of content on the Web. One of its major characteristics is the timely provision of new content and quick interaction among users. Social media is transforming the way information are provided and propagated over traditional media. We put the assumption based on our observation that information on social media, especially popular topics and news, are propagated quickly and can attract vast amount of attentions. If such information has an impact on online shopping behavior of many users .

This thesis will presents a hybrid recommendation system for matching people in social networks with the use of clustering. Two incremental constraint based clustering algorithms will proposed and developed by the researcher to produce fine clustering solutions that improve the accuracy of the recommendation. The proposed clustering algorithms combine the similarity measurement and constraints validation in one step and execute it at the cluster level, which makes these algorithms viable for large datasets, such as social networks. Two personalized ranking methods also suggested by the researcher that employ the users' constraints and past communications to rank the recommendations in the social networks. The proposed system and methods are evaluated using the data collected from a real life social network. This study is an approach towards process of Web Usage Mining like Data Collection, Pre-processing, Pattern

Discovery and Pattern Analysis using soft computing methods. The study focus is on several approaches of web mining such as statistical analysis; clustering, association rules and pattern finding to discover hidden patterns of online users.

The research work has a limitation that our research is not aiming at implementation of the proposed schemes. Therefore, the validation is verified only by numerical simulations and MATLAB simulations. All the simulations test and results will base on simplified mathematical model of social web sites users and groups without consideration to real practical details.

The Main research problem covered in this study is how to prepare good-quality social network data for data analysis and mining by collecting web data from online social networks such as Twitter, Facebook , Micro blogging sites , Mobile social media sites like Whatsapp etc. that are rapidly grown in popularity. As we studied that old mining algorithms cannot operate or generate accurate results on the vast and messy data. Thus social network data preparation deserves special attention as it processes raw data and transforms them into usable forms for data mining and analysis tasks. , in Our work we emphasizes the importance of data preparation for social network analysis and mining tasks and further designing some new approaches with the help of soft computing methods..”

## 2. REVIEW OF LITERATURE

In recent years, social media have experienced tremendous growth in their user base. For example, there are more than one billion members belonging to the “Facebook network (Facebook 2013), while Twitter now has more than 280 million monthly active users “ (GlobalWeb- Index 2013). There are a large number of different social media applications or platforms which in general can be categorized as weblogs, microblogs, social network sites, location-based social networks, discussion forums, wikis, podcast networks, picture and video sharing platforms, ratings and reviews communities, social bookmarking sites, and avatar based virtual reality spaces. In a broader sense, social media refers to “a conversational, distributed mode of content generation, dissemination, and communication among communities” (Zeng et al. 2010, p. 13).

The mainstream adoption of social media applications has caused a paradigm shift in how people communicate, collaborate, create, and consume information. In particular, the process of information consumption and dissemination is closely interrelated with the process of generating and sharing information (Zeng et al. 2010). Furthermore, while of course huge societal power differentials persist, the curation and diffusion of publicly available information is no longer as easily controlled by a small number of institutional “gatekeepers”.

In addition to the personal, everyday uses that are the source of social media’s mass adoption, social media have also been increasingly used as communication channels in business, political, and other contexts. For example, companies have started to adopt internal as well as external (public) social media platforms for a number of purposes. While the use of internal social media applications should improve communication and collaboration among employees, knowledge management, and product/service innovation, various companies have started to establish social media-based networks with business partners and have also begun to engage in public social media activities for the purposes of marketing, public relations (PR), customer relationships, reputation management, and recruitment.

In the political domain, social media are believed to have the potential to increase political participation by citizens and voters (Wattal et al. 2010).

While Twitter is an ideal public platform for disseminating political information and opinions quickly and widely (Stieglitz and Dang-Xuan 2013a, 2013b; Bruns and Highfield 2013), political actors (e.g., politicians, parties, foundations, etc.) have also begun to use Facebook to enter into dialogues with citizens and to encourage more political discussions.

Academic research from various disciplines of the social and even the natural and applied sciences has recently devoted more attention to social media. Social

networks and social media have also become an important domain in information systems (IS) research. This recent interest in “Big Social Data” (Manovich 2012; Burgess and Bruns 2012) has been driven in part by facilitated access to large-scale empirical datasets from popular online social networking platforms such as Twitter, Facebook, and LinkedIn, as well as from other platforms that facilitate mass collaboration and self-organization such as weblogs, wikis, and user tagging systems.

As boyd and Crawford (2012) point out, “the era of Big Data is underway. Computer scientists, physicists, economists, mathematicians, political scientists, bio-informaticists, sociologists, and other scholars are clamoring for access to the massive quantities of information produced by and about people, things, and their interactions” (boyd and Crawford 2012, p. 663). Defined as a cultural, technological, and scholarly phenomenon that rests on the interplay of technology, analysis, and methodology,

Big Data is “less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets” (boyd and Crawford 2012, p. 663). From a research perspective, social media can be understood as a kind of living lab, which enables academics to collect large amounts of data generated in a real-world environment.

There is a significant interest in analyzing Big Social Data from social media not only for research but also for practical purposes. For example, in analyzing social media data, companies see the opportunity for targeting advertising, PR, social customer relationship management (CRM), and business intelligence (BI). In particular, the primary interest behind corporate activities in social media is how to effectively use them as an additional channel for marketing. Further, B2B companies have also started to use social media analytics to identify new potential customers.

Recently, political institutions have also shown an interest in monitoring the public opinion on policies and political positions, detecting trending political topics, and managing their own reputation in the social web. Public officials could potentially use social media to identify situational information created by citizens in times of natural disasters (Bruns and Burgess 2012; Bruns and Liang 2012).

Furthermore, based on social media data, health organizations could establish an early warning system for disease outbreaks that should help provide timely response measures. Also, individuals and consumers seek to make use of information and opinions from diverse sources in order to make more informed decisions.

An additional use case for research can be found in conducting intercultural studies: by analyzing social media content, academics are able to directly

compare how people in different countries or cultures react to certain global events, for example.

### **3. RESEARCH METHODOLOGY**

#### **Data collection**

The data for this third progress report was primarily collected through offline and online survey form collection from various users of NCR region of India, which took place during "2013-15. We did some interview also with IT Expert and Social Media Users in form of closed and open questions. Main Questions framed by us was based on following points ":

1. Internet access on PC or Mobile Phone
2. Experience of social networking sites in years/months
3. Accessibility of social networking sites
4. Use of social networking sites
5. Use of social networking sites for study, marketing or other purpose
6. Accessibility of mobile phones.

#### **Sample Collection**

Local authorities, mainstream schools, colleges, university teachers government employee, business man, within NCR region of Delhi and nearby areas were part of this survey . Some IT service providers were approached also , and in all, positive responses (and eventual questionnaires) were received .They were requested to complete the questionnaire and return it to us by online or offline . The questionnaire focused on the use of social media by random collection of users and their pattern findings

#### **Survey**

The results of the most recent data capture in Summer 2015 are presented in this analysis report. This consisted of telephone interviews with 33 young users of social media 17-35 which typically took 15-30 minutes. The interview consisted of a combination of closed and open questions in the following sections:

1. Internet access
2. Experience of social networking sites
3. Accessibility of social networking sites
4. Use of social networking sites
5. Use of mobile phones

6. Accessibility of mobile phones

This report also presents data relating to social media or social site activity done by the users in past one month, which was obtained in previous data collections.

#### **Survey Analysis strategy**

A number of tables are referred to throughout the report. The data is presented in terms of both numbers and percentages. As our sample numbers are low, the percentages should be interpreted cautiously and numbers within individual cells of tables referred to at all times. Percentages are rounded to the nearest whole number and so the sum of percentages will not always be exactly 100.

#### **Sample demographics and representativeness**

Taking the sampling frame we included students of schools , colleges , university , government and private employee , housewives and businessman who were initially identified by the local authority services in phase one of the project as our population, the sample was examined to assess its representativeness.

The small sample size must be taken into account. Even so, overall the sample does appear to be a good reflection of the underlying population.

It was possible to make contact with 70 of the participants for the interview in relation to social networking and mobile phone use.

### **4. DATA ANALYSIS**

In the survey, the young people were asked to list three things they spend their time doing outside of school /office/business/home hours. This produced a wide range of responses. As they were only listing three activities (although some listed four), this list cannot be considered exhaustive. Nevertheless, it does give a good indication of the types of activities that blind and partially sighted young people are getting involved in. At the time cohort 1 were in Year 9 and cohort 2 were in Year 11.

**Table 1: We are interested in how you spend your time outside of school/office/home/business hours.**

	1	2	3	4	Total	Per%
Sport	15	48%	21	45%	36	47%
Socialising with friends	7	23%	23	49%	30	39%
Playing music	5	17%	14	30%	19	25%
Using computer / internet	8	27%	10	21%	18	23%
Watching TV	7	23%	9	19%	16	21%
Listening to music	3	10%	12	26%	15	20%
Social Sites Reading	5	17%	9	19%	14	18%
Playing computer games	5	17%	5	11%	10	13%
Youth group / youth theatre	3	10%	5	11%	8	10%
Social media posting	1	3%	5	11%	6	8%
Time with family	3	10%	2	4%	5	7%
Part time job or voluntary work	1	3%	3	6%	4	5%
Photography	1	3%	2	4%	3	4%
Another type of club	1	3%	1	2%	2	3%
Other	8	27%	9	19%	17	20%

**Computer and Internet Access**

When asked by us, 77% of users said that they had access to either their own computer, or a shared one in the household. All households had broadband connection.

All but one participant described having access to the internet at home. The majority (57%) also had access to the internet through their own personal computer at home. This is a higher percentage than found by Ofcom where only 68% of participants had internet access (although it should be noted that this was surveying people aged 25-34).

**Table 2: Do you or does anyone in your household have access to the internet at home?**

	Total (N)	Total (%)
Yes	69	77%
No	1	1%
Not sure	0	0%
Total	70	100%

**Table 3 where do you access the internet from?**

	Total (N)	Total (%)
Home	69	99%
School/College	62	89%
Friend’s house	35	50%
Library	20	28%
Other	26	37%

The young people were asked to identify where they typically access the internet from. As would be anticipated, many access the internet from home (99%) and from school or college (89%). Half accessed the internet from friends’ houses, whilst a few (29%) go to the library to use the internet. Some of those who said that they would not go to the library explained that this was not possible because those computers did not have access technology installed

**5. CONCLUSION**

The extraction and designing algorithms for better mining on social media sites to frame relationships among users and groups on various social media sites. Since the data for the proposed work has to be obtained from social media Web only, Web mining techniques for Social Network extraction will be studied as well. In this proposed work, relationship information between various groups , social media communities and users entities obtained from multiple user profiles of social websites data sources have to be extracted which may lead to various ambiguities. As we studied that Soft Computing techniques can be used to solve ambiguity problems. We intend to develop some Soft Computing techniques based algorithms using ANN, Fuzzy sets and we will use ACO tools also. We also intend to perform a comparative study to evaluate the effectiveness of these algorithms in social network extraction, particularly in a academic social network extraction, where user profile data of teachers, students or other workers of a University or college will be used for testing our results.

**6. REFERENCES**

1. Global WebIndex (2013) Twitter now the fastest growing social platform in the world. <https://www.globalwebindex.net/twitter-now-the-fastest-growing-socialplatform>
2. Bruns A. & Highfield T. (2013) Political networks on twitter: tweeting the Queensland state Election. Information, Communication & Society 16(5): pp. 667–691
3. Bruns A. & Liang E. Y. (2012) Tools and methods for capturing twitter data during natural disasters. First Monday 17(4).
4. Boyd D. & Crawford K. (2012). Critical questions for big data. Information, Communication & Society 15(5): pp. 662–679
5. Larson K. & Watson R. T. (2011). The value of social media: toward measuring social media strategies. In: Proc of 32nd international conference on information systems.

**Corresponding Author**

**Poonam Rani\***

Research Scholar, Pacific University, Rajasthan, India