

# An Insight on Social Media Stream Mining

Rojina Deuja<sup>1</sup>, Krishna Bikram Shah<sup>2</sup>

Department of Computer Science and Engineering, Nepal Engineering College  
Changunarayan-5, Bhaktapur, Nepal

<sup>1</sup>rojinadeuja33g@gmail.com, <sup>2</sup>krishnabs@nec.edu.np



*Rojina Deuja has received her Bachelor of Engineering in Computer from Nepal Engineering College, Pokhara University, Nepal. She has been working on Data Warehousing and Business Intelligence for over a year. Her interests lie in Data mining, Machine Learning and Artificial Intelligence.*



*Krishna Bikram Shah is Assistant Professor at Department of Computer Science and Engineering, Nepal Engineering College. He holds M. Tech. in Computer Science and Engineering from Sikkim Manipal University, India. He has served at several computing departments in India. His research interest include Artificial Intelligence, Bioinformatics and Machine Learning.*

## Abstract

*Data stream mining is one of the realms gaining upper hand over traditional data mining methods. Transfinite volumes of data termed as Data Streams are often generated by Internet traffic, Communication networks, On-line bank or ATM transactions etc. The streams are dynamic and ever-shifting and need to be analysed online as they are obtained. Social media is one of the notable sources of such data streams. While social media streaming has received a lot of attention over the past decade, the ever-expanding streams of data presents huge challenges for learning and maintaining control. Dealing with billions of user's data measured in petabytes is a demanding task in itself. It is indeed a challenge to mine such dynamic data from social networks in an uninterrupted and competent way. This paper is purposed to introduce social data streams and the mining techniques involved in processing them. We analyse the most recent trends in social media data stream mining to translate to the detailed study of the matter. We also review innovative implementations of social media stream mining that are currently prevalent.*

## I. Introduction

Data Streams are high-speed, continuous flow of data. These streams are generated from an array of different sources like online searches, social media, business transactions, phone conversations, network traffic etc. These data are mined by programmers to maintain records, which is eventually used to acquire useful information. This process of extracting knowledge structures from data streams is termed as Data Stream Mining. Data stream mining operates under the common canopy of data mining, machine learning and knowledge discovery.

The process of deriving actionable trends and patterns by analysing raw social media data streams is called Social media mining. It may require either human data analysts or automated software programs to sift through huge amounts of social media data (social media usage, opinion mining, content sharing, online customer behaviour, etc.) in order to discern patterns and trends. These patterns and trends are usually of interest to businesses, companies, governments and

even non-profit organizations, as these systems can use the data patterns and trends to design their strategies or introduce new programs [1].

A major challenge is that social media data streams possess multiple unparalleled properties like transfinite length, concept-drift, concept-evolution, feature-evolution and limited labelled data. Moreover, the volume and speed of the stream itself stimulates added complexities that are not present in traditional data mining. Such streams may have different types of features in addition to the large volume of data. On top of that, the concepts and features tend to evolve throughout the stream. A concept-drift takes place when the underlying concept of data changes over a certain period. Feature-evolution is said to occur when a feature set varies over time. Similarly, whenever new classes evolve in the streams, concept-evolution takes place. There is also a lack of labelled data since manual labelling is infeasible at large scale. Due to these reasons, traditional data mining techniques are indeed incapable of addressing social data streams.

In this paper, we discuss different techniques used for Social Network Analysis (SNA). We also examine the implications for mining social media streams, and enlist four popular online services working on the discovery and analysis of useful information from social media.

## II. Tracking data flow

Unlimited units of data are produced in social media sites every second. The essential information that propagates between the users in a social network can be in the form of posts, messages, comments, links, tags and multimedia. There are billions of discrete users who have their own connections, behaviours, preferences and trends.

In order to mine such data, it is necessary to track units of information that correspond to pieces of information (posts, events, articles, etc.). These data vary within a very short time and need to be handled at a very large scale. Below are a few ways in which this can be done:

### A. Cascading links to article

A hyperlink is a reference to a document or a resource typically activated by clicking or tapping. Hyperlinks are reliable to trace since they travel relatively unchanged. They are widely used in blogs, microblogs, social networking sites and internet

forums. Users write posts that refer (link) to other posts and in this way the information propagates. The transfer from such hyperlinks can be identified and cascade graphs can be constructed by a time ordered propagation of information as shown in Fig. 1.

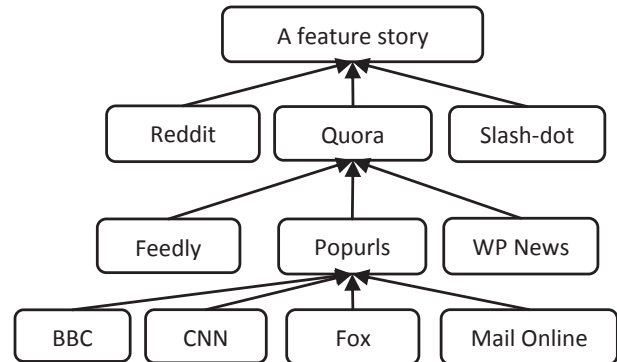


Figure 1: A cascade of articles tracing a feature story in the web

Previously, hyperlinks were limited to designing the structure of networks, where posts were modelled as nodes and hyperlinks as edges in a directed graph. Jackson [2] suggested that social network analysis (SNA) can be carried out using hyperlinks in websites. He contented hyperlink-based network analysis as a strong approach for studying representations of online networks. Adamic and Glance [3] analysed the connectedness of political blogs by studying the linking behaviour in the blog posts. Miller et al, [4] used a combination of graph theory and sentiment analysis fields to track sentiment flow in blog post networks traced through hyperlinks. Park [5] presented hyperlink network analysis (HNA) for the study of online social (or communication) networks. He also explained various data-gathering techniques used for HNA.

### B. Keywords and Hashtags

Keywords are words or phrases which are used as a metadata tag in websites. At one time, keywords were one of the most important elements used in web searching and text mining. Even today, search strings are matched with keywords.

Hashtags are similar to keywords, but they are preceded by a '#' (hash) symbol. Hashtags are mostly associated with driving engagements on a topic. For example, the hashtag #MeToo was used to spark up a movement against sexual violence. Hashtags have been incorporated in social media like Twitter, Facebook, Instagram, Snapchat etc.

The information network that is built with the help of keywords and hashtags can be very crucial in studying online information flow. Techniques like Double Ranking (DR) technique [6], Search set (S) retrieval [7], context-sensitive topical PageRank (CTPR) [8] etc. have been proposed to extract topical keywords in Twitter. Meanwhile, [9] analysed the motivation, goals and usage patterns of social media users through the hashtags that they used. A comparison of users' search behaviours between Web search and Twitter search was done in [10] while the difference between traditional search engines and social media search associated with health related information was explored in [11].

### C. Memes

Meme can be regarded as a unit of information that spreads between various users in any social media network through their social ties. Memes can be used to derive valuable data and also to study user or community behaviour in any social network. When a certain meme gets viral, it can be used to attract attention for marketing and advertisement.

Memes can be, thus analysed to follow iterations of any story as they evolve. Leskovec[12] introduced a meme-tracking approach to identify news cycles- which are daily rhythms in the online new outlets. The Memetracker clusters together short, distinctive phrases that characterize a specific topic. It then creates groups of news on topics that can be tracked over time to find diffusion patterns in the news cycle. One difficulty in clustering memes is that the actual granularity of a meme is very difficult to capture with the help of general entities such as tags and keywords. Thus, the messages are grouped on the basis of existing atomic entities called Protomemes. Protomemes are aggregated to place them into cohesive groups that represent a consistent topic of interest. Protomeme detection is a very common pre-clustering procedure used for isolating memes before processing them and has been utilized in works including [13] and [14].

### III. Maintaining data quality

After tracking data flow, the data from different sources are gathered using varying methods. Such data do not necessarily come in a pre-processed form. A huge amount of the collected data may be incomplete,

incorrect, misleading or inadequate for the required application. For social media streams, the quality of data that is being tapped is much more significant than the scale of the data being accumulated. In order to assure the quality of data, the following discrepancies need to be removed:

#### A. Noise

Noise is the distortion of the data that is present in the social networks. The signal-to-noise ratio is often very high in social media streams. There are four main sources of noise in social media data:

- Fake/Unused accounts: A lot of social media accounts are not owned or managed by a person of its true identity. Similarly, many accounts in different sites are created by the users but never used again.
- Spam: Spams are mainly characterized by malicious links, malware injection attempts, unsolicited and poorly targeted marketing messages, phishing etc. Spams are undesirable and their appearance impacts user's sentiment analytics and keyword/brand trends.
- Deceptive content: Deceptive content includes information that is produced with a motive to deceive the users and passed on as valid information. For example, in 2016, the news of the death of renowned celebrity Johnny Depp took over Facebook. Many fans, news channels alike were fooled by the news which was later found to be a hoax.
- Duplicate data: Duplicate data are formed when there are multiple instances with the exact same characteristic values. Multiple accounts of the same user, Retweets, Reposts, Shares, Forwards, Auto blogs, Splogs, Tweetbots, Cross-posting etc., can lead to duplicate copy of the same data. Duplicate data hampers the magnitude and validity of statistics in a sample stream.

Clearly, noise in data streams is undesirable and needs to be filtered out. Even so, there are a few filtering algorithms that are effective in combating the adverse effects of noise. In [15], the noise contributed by users was detected with the help of data collected from Twitter and forums. Several

classical strategies along with graph algorithm were used to filter out noises from the data. In another work, user interests were evaluated in order to address dynamic issues in noise web data reduction [16]. The authors contend that noise in online data should be isolated by taking user interest levels on web data into account. Here, a tool called the Noise Web Data Learning (NWDL) - capable of learning noisy web data in user profiles was presented. Similarly, in [17] a Noise Removal and Structured Data Detection (NR-and-SDD) algorithm was proposed, which reduces processing cost of noisy data and improves structured data detection in relevance to personality features.

### B. Missing Data

Missing data are those feature values that are missing in instances. The missing information may either be that of an actor (user) or the link with other users. For example, individuals may avoid reporting profile information on social media, such as their age, location, or hobbies. Missing values limit the capacity to describe the network context of the actors and their neighbours.

There are two distinct forms of missing data: node level and tie level. Node level is when the data of a particular person or object is missing, as if they were not part of a study at all. Node level missing data is characterized by unit non-response. Unit non-response is when the actor does not participate. Tie level missing data occurs when any unit connecting the social contacts of that unit to other units in the network is unavailable [18].

If not excessive, the instances containing the missing values can be removed. Another strategy is to entirely ignore the missing node or tie, as if it never existed. But ignoring missing data can have a large effect on descriptive analysis of social networks. An additional approach is to impute the data that is missing. Imputing involves assuming or guessing a missing node and its properties. Several methods have been suggested for this purpose: One is to use an ERG model to model the dataset, then filling in the missing data through maximum likelihood estimates which are based on the ERGM parameters. K-tree model [19] was used to accurately estimate properties of complete cascade with up to 90% data missing. Similarly,

Ward and Hoff [20] used Latent space models which are based on logistic regression to replace the missing data.

## IV. Analysing data streams

The data that retrieved from social streams presents new opportunities to study human behaviour, while still challenging our capability to analyse them. It is necessary to analyse the underlying mechanisms within the data streams to keep with the real-time spread of information through the on-line networks. Predictive models of social interactions can be constructed if we understand how messages flow through social networks and how we can identify networks over which important messages spread. Below is a list of techniques that have been widely implemented in analysing dynamic social media streams.

### A. Graph theory

Graph theory relates to the study of graphs using mathematical or pictorial representation of a network that models the relationship between various objects. A graph constitutes of a set of vertices or nodes, with certain pairs of these nodes connected by edges (undirected) or arcs (directed).

Graph theory is credibly one of the widely-used methods for social network analysis since the early times of social network. In such networks, a vertex denotes a user and an edge denotes any form of association between the corresponding vertices (e.g. Fig. 2). One major advantage of modelling relationships in a graph is that the edges and vertices can be easily added or removed.

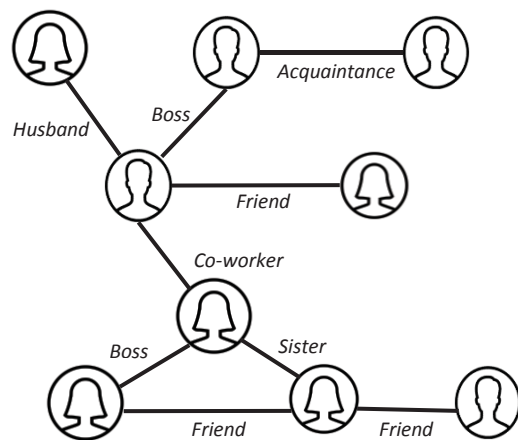


Figure 2: A graph showing different types of associations between users

Network models like Random graphs, Erdos–Rényi model, Barabási – Albert model, Agent based, Fitness model and Boolean Network are commonly used for the purpose of social analysis. Räsänen [21] employed random graphs as models of Peer-to-Peer social networks. Three random graph models along with their degree distributions were studied by performing numerical simulations for each model. Ioannides [22] wrote about ongoing research on networks by illustrating three main strands of literature on social networks. He outlined different graph theories, old and new, and then discussed Markov random graphs of social interaction. Hamill and Gilbert [23] compared the characteristics of four standard models (regular lattice, random, small-world and preferential attachment) to conclude that they do not reproduce social networks quite well. They proposed an agent-based model based on the ideas of social space and social distance termed as a social map to properly replicate the basic natures of large social networks.

### B. Community Extraction

Community can be categorized as a group of users sharing common interests in any social group. Communities showcase a distinct character of a user group in social networking sites. For mining the data from a particular social service, it is essential that the communities of users in that particular social network are identified. However, it is no easy take to detect and model the communities formed among the dynamic group of users. The majority of such methods are based on partitioning an entire network into communities, with the expectation of multiple ties within communities and a few ties between [24]. An example of such structure is shown in Fig. 3.

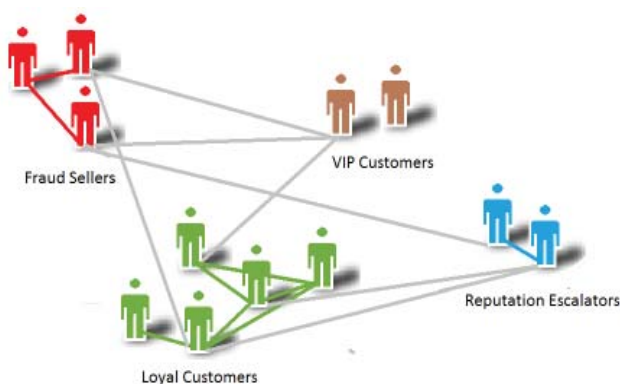


Figure 3: Example of community extraction in a retail business using clustering

Detecting the communities and extracting their behaviour is necessary to study their domain. Several popular link analysis techniques can be used for this purpose which includes HITS algorithm and several Graph Clustering techniques. Many clustering techniques have been applied in works like [25], [26] and [27] for the extraction of community from social network, with hierarchical clustering being mostly used. The hierarchical clustering approach is a combination of techniques where network nodes are grouped to uncover their individual strengths, which is then used to extract communities from the overall network [28].

### C. Link Prediction

The data produced in social media sites are complex and ever-changing. The association between specific nodes (user or objects) is comparatively easier to study. While studying social networks, the need of foreseeing the unobserved social links between different users may arise. This can be further described using a set of interesting questions such as: Can associations develop between any two nodes? What factors cause such associations? How do any patterns in the association between nodes evolve over time? Does the association between two nodes affect other nodes?

The challenges that these questions present is to predict the likelihood of a future association between the nodes that is different from the current association between them. This problem is often termed as the Link Prediction problem [29]. For example, in any given social network at time  $t$ , it may be needed to predict the social link between actors at time  $t+1$ . According to the classical approach for link prediction, the social network is fit into a model and then the model is used for link prediction. Some of the notable models for link prediction are  $p^*$  model [30], Latent Space model [31] and Dynamic Latent Space model [32]. Besides these, other models have also been proposed that include [33]; [34].

### D. Semantic Web of Social Network

The Semantic Web (SW) can be regarded as an extension of the current World Wide Web. The SW environment allows data in different web pages to be structured and tagged in a way that the data can be shared and reused across different applications.

SW has applications in areas including Search Engine Optimization (SEO), Data Integration (DI) systems, Decision Support Systems (DSSs), Seamless applications and Intelligent Agents.

SW technologies are implemented with the help of ontologies to model user data. An ontology is simply a formal conceptualization of a domain. It specifies classes of objects, their instances and relationships between them. Noy [35] demonstrated the evolution of applications of ontology for user modelling and adaptation in relation to the Semantic Web. Plumbaum et al. [36] implemented 17 distinct social applications to identify requirements and attributes for developing a common user model called the Social Web User Model (SWUM), that aggregates information from distributed users. Mika [37] used a community-based ontology extraction that was implemented as part of the Flink system.

Today, one of the most widely used ontologies is FOAF (Friend of a Friend). The FOAF ontology specifies a vocabulary that can be used to define and maintain information describing users, their attributes and their connections with other users. The FOAF is further explained in works including [38], [39] and [40].

## **V. Examples of social media mining**

Trend analysis on social media provides a strong foundation for strategic recommendations. Many applications for processing massive volumes of user data generated by microblogging and social networking services are prevalent. We list a few examples of popular online mining services that use social media streams:

### **A. Google Hot Trends**

Google Trends is a Google Inc. product that is publicly available to everyone. It can be used to find out about how often a particular search is carried out of the total searches across different parts of the world. The popularity of a particular search-item is also categorized in terms of location and language. Then Google launched Google Hot Trends as an add-on to Google Trends.

Google Hot Trends can be installed as an iGoogle gadget. It is can also be utilized as an hourly Atom

web feed [41]. Google Hot Trends is slightly more versatile than Google Trends. Google Hot Trends provides insight with a 24-hour search-volume graph on online search results, news, articles, blogs etc. Google Hot Trends collects online data and models them to present them to its users. Hence, it is an excellent example of the implementation of social media data mining.

### **B. Cloud4Trends**

Cloud4Trends is a system that identifies trend and relevant topics in social network through clustering in real-time. It was built as a framework for collecting and analysing user generated content (UGC) in microblogging and blogging spaces. This trend detection application has a 3-tier design: (1) data collection in streaming fashion; (2) application of an online clustering technique on the data to detect recent trends; and (3) refinement and ranking of clusters for data visualization [42].

### **C. IceRocket**

IceRocket is another popular Internet search engine that caters its users with real-time search facilities. IceRocket was launched in 2004, originally with features purported to make online searching easier on a Personal Digital Assistant (PDA) [43]. Today, IceRocket specializes in blog searching. However, IceRocket is also widely employed in mining data streams from social networking websites like Twitter and Facebook. It also mines data from news and articles on the internet. The Big Buzz feature of Ice-Rocket provides all of the previously mentioned services under a single page. Ice-Rocket is a free public web facility that is ad supported. IceRocket owns an API that it licenses to firms and companies monitoring social media streams.

### **D. SPOONS**

SPOONS (Swift Perceptions of Online Negative Situations) are a Netflix service outage detection system via real-time stream analysis [44]. The SPOONS system was developed as an early warning mechanism for outages. It tracks Twitter for those situations when the users are complaining about the same thing at the same time. Apparently, the main motive of the SPOONS system is to use Twitter posts to determine when Netflix users are

reporting a problem with any of the Netflix services. SPOONS work with a combination of term-frequency measurement and sentiment analysis.

## V. Conclusion

Social media has become an inseparable part of the lives of people around the world. The data streams from social sites possess great value for companies, scientists and individuals the same. If we are able to accurately analyze the social media streams and acquire useful information from them, then data stream mining will be a hero in true terms. To fully realize the potential of social media streams, new data mining systems need to be developed. Incremental improvements will, of course, continue to help. This will assure that we get huge amount of information from social media users and reap benefits from it.

## References

- [1] A. Weiler and M. Grossniklaus R. Popovici, Online Clustering for Real-Time Topic Detection in Social Media Streaming Data in 23rd International World Wide Web Conference, Seoul, Korea, 2014.
- [2] M.H. Jackson, "Assessing the Structure of Communication on the World Wide Web," *Journal of Computer-Mediated Communication* (3:1), 1997.
- [3] L.A. Adamic and N. Glance, The political blogosphere and the 2004 U.S. election in Proceedings of the 3rd international workshop on Link discovery - LinkKDD '05, pp. 36–43, 2005.
- [4] M. Miller, C. Sathi, D. Wiesenhal, J. Leskovec, J. and C. Potts, Sentiment Flow Through Hyperlink Networks in ICWSM, 2011.
- [5] H.W. Park, "Hyperlink Network analysis: A New Method for the Study of Social Structure on the Web", 2003.
- [6] W. Shuai, C. Zhiyuan Chen, L. Bing and S. Sherr, Identifying Search Keywords for Finding Relevant in AAAI'16 Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence pp. 3052-3058, 2016.
- [7] G. King, P. Lam and M. Roberts, "Computer-Assisted Keyword and Document Set Discovery from Unstructured Text," *American Journal of Political Science*, 61, 4, pp. 971-988, 2017.
- [8] W.X. Zhao, J. Jiang, J. He, Y. Song, P. Achananuparp, E.P. Lim and X. Li, "Topical keyphrase extraction from twitter in ACL," pp. 379–388, 2011.
- [9] J. Sang, C. Xu, and R. Jain. Social multimedia mining: From special to general in *Multimedia (ISM), 2016 IEEE International Symposium on*, pp 481–485. IEEE, 2016.
- [10] J. Teevan, D. Ramage, and M M. Ringel, Twittersearch: a comparison of microblog search and web search in Proceedings of the fourth ACM international conference on Web search and data mining, pp. 35–44. ACM, 2011.
- [11] M. D. Choudhury, M. R. Morris, and R. W. White, Seeking and sharing health information online: Comparing search engines and social media in Proceedings of the 32nd annual ACM conference on Human factors in computing systems, pp. 1365–1376, ACM, 2014.
- [12] J. Leskovec, L. Backstrom, and J. Kleinberg, Meme-tracking and the dynamics of the news cycle in Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 497–506. ACM, 2009.
- [13] E. Ferrara, M. J. Asbagh, O. Varol, V. Qazvinian, F. Menczer, A. Flammini, Clustering Memes in Social Media in IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2013.
- [14] H. Sayyadi, M. Hurst, and A. Maykov. Event detection and tracking in social streams in Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media, 2009.
- [15] A. Stoica, Filtering Noisy Web Data by Identifying and Leveraging Users' Contributions in Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, 2012.
- [16] J. Onyancha, V. Plekhanova, Noise Reduction in Web Data: A Learning Approach Based on Dynamic User Interests in *International Journal of Computer and Information Engineering* Vol:12, No:1, 2018.
- [17] M. F. Uddin, Noise Removal and Structured Data Detection to improve search for personality features in IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, USA, 2016.
- [18] M. Huisman, "Treatment of Missing Data in Longitudinal Network Studies," in SIENA workshop, 2009.

- [19] E. Sadikov, M. Medina, J. Leskovec, and H. G. Molina, Correcting for missing data in information cascades, WSDM'11, Hong Kong, China, 2011.
- [20] P. D. Hoff and M. D. Ward, "Modeling Dependencies in International Relations Networks," Cambridge University Press, Vol. 12, No. 2 (Spring 2004), pp. 160-175
- [21] J. Räsänen, Random graphs as models of Peer-to-Peer social networks in U.U.D.M. Project Report, 2012
- [22] Y.M. Ioannides, "RANDOM GRAPHS AND SOCIAL NETWORKS: An Economics Perspective," Department of Economics, Tufts University yannis, 2015.
- [23] L. Hamill and N. Gilbert, "Simulating Large Social Networks in Agent-Based Models: A Social Circle Model," Centre for Research in Social Simulation, University of Surrey, UK, 2010.
- [24] Y. Zhao, E. Levina and J. Zhu, Community extraction for social networks in Proc Natl Acad Sci USA. 2011 May 3; 108(18): 7321–7326. Published online 2011.
- [25] M. Girvan, M. E. Newman, Community structure in social and biological networks in Proceedings of the National Academy of Sciences, 2002.
- [26] S. Papadopoulos, Y. Kompatsiaris, A. Vakali, P. Spyridonos, "Community detection in social media," Data Mining and Knowledge Discovery, pp. 515-554, 2012.
- [27] M. Newman, "Networks: An introduction," Oxford University Press, 2010.
- [28] M.A. Olowe, M.M. Gaber, F. Stahl, "A Survey of Data Mining Techniques for Social Network Analysis," pp. 4-7.
- [29] D.L. Nowell and J. Kleinberg, The Link-Prediction Problem for Social Networks in Journal of the American society for information science and technology, 2007, pp. 1019-1031.
- [30] S. Wasserman, P. Pattison, "Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and  $p^*$ ," Psychometrika, 1996, Volume 61, Number 3, p 401.
- [31] P.D. Hoff, A. E. Raftery and M. S. Handcock, Latent Space Approaches to Social Network Analysis in Journal of the American Statistical Association Volume 97- Issue 460, pp. 1090-1098, 2002.
- [32] P. Sarkar and Andrew W. Moore Dynamic social network analysis using latent space models in ACM SIGKDD Explorations Newsletter Volume 7 Issue 2, pp. 31-40, 2005.
- [33] M. Giatsoglou and A. Vakali, Social networking trends and dynamics detection via a cloud-based framework design in Mining Social Network Dynamics Workshop (MSDN 2012), WWW 2012 Conference, 2012.
- [34] T. Murata, S. Moriyasu, Link prediction of social networks based on weighted proximity measures in IEEE/WIC/ACM international conference on web intelligence. IEEE, pp. 85-88, 2007.
- [35] N. F. Noy Semantic integration: a survey of ontology-based approach in ACM Sigmod Record 33(4), pp. 65–70, 2004.
- [36] T. Plumbaum, S. Wu, E.W.De Luca, S. Albayrak, "User modeling for the social semantic web," SPIM, pp. 78–89, 2011.
- [37] Mika, Peter. "Ontologies are us: A unified model of social networks and semantics." J. Web Sem, pp. 5-15, 2007.
- [38] L. Ding, T. Finin, and A. Joshi, "Analyzing Social Networks on the Semantic Web," IEEE Intelligent Systems, 2005.
- [39] J. Golbeck and M. Rothstein, Linking Social Networks on the Web with FOAF: A Semantic Web Case Study in Twenty-Third AAAI Conference on Artificial Intelligence, Chicago, USA, 2008.
- [40] D. Kavitha, SURVEY OF DATA MINING TECHNIQUES FOR SOCIAL NETWORKING WEBSITES in International Journal of Computer Science and Mobile Computing, Vol.6 Issue.4, April 2017.
- [41] (2019) Google Trends [Online]. Available: [https://en.wikipedia.org/wi-ki/Google\\_Trends](https://en.wikipedia.org/wi-ki/Google_Trends)
- [42] S. Volkova, Data Stream Mining: A Review of Learning Methods and Frameworks, 2012.
- [43] (2017) IceRocket [Online]. Available: <https://en.wikipedia.org/wiki/-IceRocket>
- [44] E. Augustine, E., C. Cushing, A. Dekhtyar, K. McEntee, K. Paterson and M. Tognetti, Outage detection via real-time social stream analysis: leveraging the power of online complaints in Proceedings of the 21st international conference companion on World Wide Web, WWW '12 Companion, pp. 13–22, New York, NY, USA, 2012.

\* \* \*