Ilkogretim Online - Elementary Education Online, 2021; Vol 20 (Issue 6): pp. 2213-2227
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.06.204

A Social Networking Services Based On Mining Information Flow

Ms. Sneha Dongre Computer Science And Engineering, Guru Nanak Institute Of Engineering And Technology Rashtrasant Tukdoji Maharaj Nagpur UniversityNagpur, India
mesnehadongre26897@gmail.com

Prof. Vijaya Kamble, (project guide)
Computer Science And Engineering,
Guru Nanak Institute Of Engineering And
Technology Rashtrasant Tukdoji Maharaj Nagpur University Nagpur,
India. sairamvijaya@gmail.com

Abstract— A social networking service (SNS) is an online platform for creating relationships with other people who share an interest, background, or real relationship. Social networking service users create a profile with personal information and photos and form connections with other profiles. Social networking services vary in format and the number of features. They can incorporate a range of new information and communication tools, operating on desktops and on laptops, on mobile devices such as tablet computers and smart phones. This may feature digital photo/video/sharing and diary entries online (blogging).Onlinecommunity services are sometimes considered social-network services by developers and users, though in a broader sense, a social-network service usually provides an individual centered service whereas online community services are groups centered. We propose a novel method to discover information diffusion processes from SNS data. The method starts pre- processing the SNS data using a user-centric algorithm ofcommunity detection based on modularity maximization with the purpose of reducing the complexity of the noisy data. Afterthat, the Info Flow miner generates information diffusion flow models among the user communities discovered from the data.The algorithm is an extension of a traditional process discovery technique called the Flexible Heuristics miner, but the visualization ability of the generated process model isimproved with a new measure called response weight, which effectively captures and represents the interactions among communities. The final constructed models allowed us to identify useful information such as how the information flows between communities and information disseminators and receptors within communities

Keywords—Information flow, social networking services,community detection, network modularity, Process mining.

I. INTRODUCTION

Social Networking service that facilitates social and special interest networking. Such services provide electronic social spaces or social network sites designed to facilitate communication, collaboration and content sharing across networks of contacts. The most famous SNSs include genuine social network sites such as MySpace, Face book, Tagged or Friendster and various kinds of special interest sites. Usually content-sharing sites and media communities, such asYouTube or Flickr are also included in this

category due totheir social networking features. There are number of likes,comments and sharing of various posts. Taking Facebook andTwitter as examples, a connection on Facebook is calledfriendship, with both users agreeing to establish the socialrelationship. In contrast, the connection on Twitter is represented by a following action, in which followed users donot have to approve the relationship, and the followed users donot have to follow their followers. When a user posts orpublishes content, other users can interact using comments orspread the content using different mechanisms of the SNS; for example, shares on Facebook and re-tweets on Twitter. Whenthese actions are repeated continuously, various processesoccur among the users, such as information dissemination. Itcreated huge data at every moment which may also containpersonal interaction with other users. There are two basic branches of process mining, one which mines process modelfrom stored event logs. In SNS we develop a method fordiscovering information diffusion model by applying processmining techniques. User centric clustering technique isperforming to reduce complexity of SNS filtered data. Theuser-centric approach was achieved using a novel measurecalled the user intimacy value, which measures the relationship level between users. The clustered data aretransformed into an event log and used as an input artifact fora newly developed process discovery algorithm named InfoFlow miner, which is based on the Flexible Heuristics miner.The algorithm uses another novel measure called responseweight, which is defined as the extent of influence or impactthat one user's actions has on another user's actions. The finalresult of the method is a graph representing the informationflow between user communities contained in the original data. The contribution of this research is twofold. First, thecomplexity of noisy SNS data can be reduced using a user-centric approach by introducing the user intimacy measure, sothat further analysis using the detected communities such asmarket segmentation can be performed. Second, an effectivevisualization method was developed for discovering information diffusion flows directly from SNS data with anextended process mining technique.

II. BACKGROUND

A. SOCIAL NETWORK ANALYSIS

Social network analysis (SNA) is the process of investigating social structures through the use of networks and graph theory.It characterizes networked structures in terms of nodes (individual actors, people, or things within the network) and the ties, edges, or links (relationships or interactions) that connect them. Examples of social structures commonlyvisualized through social network analysis include social media networks, memes spread, information circulation, friendship and acquaintance networks, business networks, knowledge networks, difficult working relationships, socialnetworks, collaboration graphs, kinship, disease transmission, and sexual relationships.

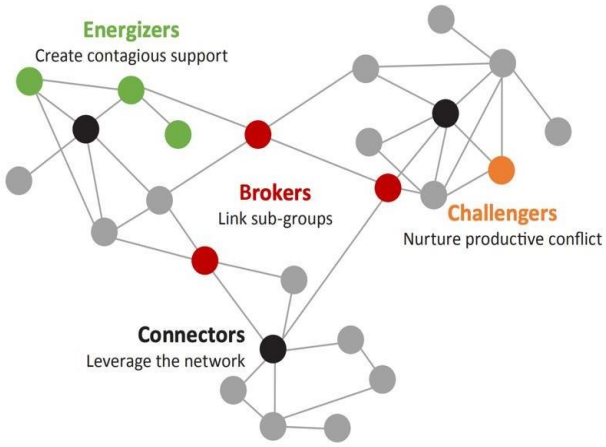

Fig. 1 Social network

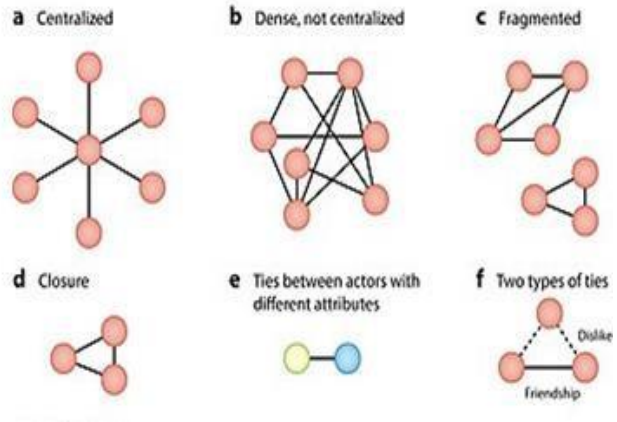

Fig. 2 Different Characteristics of Social Networks

These networks are often visualized through sociograms in which nodes are represented as points and ties are represented as lines. These visualizations provide a means of qualitatively assessing networks by varying the visual representation oftheir nodes and edges to reflect attributes of interest. To understand modularity, we briefly introduce the concepts of homophily and assortative mixing. These concepts are common in social network studies and refer to the fact that people generally have a strong tendency to associate withothers whom they perceive as being similar to themselves in some way. Moreover, assortative mixing can be quantified. Different characteristics of social networks. A, B, and C showvarying centrality and density of networks; panel D shows network closure, i.e., when two actors, tied to a common third actor, tend to also form a direct tie between them. Panel E represents two actors with different attributes (e.g., organizational affiliation, beliefs, gender, education) who tend to form ties. Panel F consists of two types of ties: friendship (solid line) and dislike (dashed line). In this case, two actors being friends both dislike a common third (or, similarly, two actors that dislike a common third tend to be friends). A network is considered assortative if a significant fraction ofthe edges in the network run between same types of node. Ifwe find the fraction of edges that run between nodes of the same type and then subtract the fraction of such edges that we would expect to find if edges were positioned at random without considering the type of node, we can quantify the networks associative mixing. This quantity is called modularity. The standard equation for calculating modularityis given by:

$$Q = 1/2m \ \Sigma ij(Aij - kikj/2m)\ \delta\ (ci,cj)$$

where m is the number of edges in the network, A is the adjacency matrix of the network, k is the degree of the node, c is the class or type of node, and δ(ci, cj) is the Kronecker deltafunction. The community detection algorithm used in our approach is based on modularity maximization, which moves nodes between communities inside the network with the objective of finding the maximum value of modularity.

B. PROCESS MINING

Process mining is a technique to analyze and track processes. In traditional business process management, it is done with process workshops and interviews, which results in anidealized picture of a process. Process mining, however, uses existing data available in corporate information systems and automatically displays the real process.

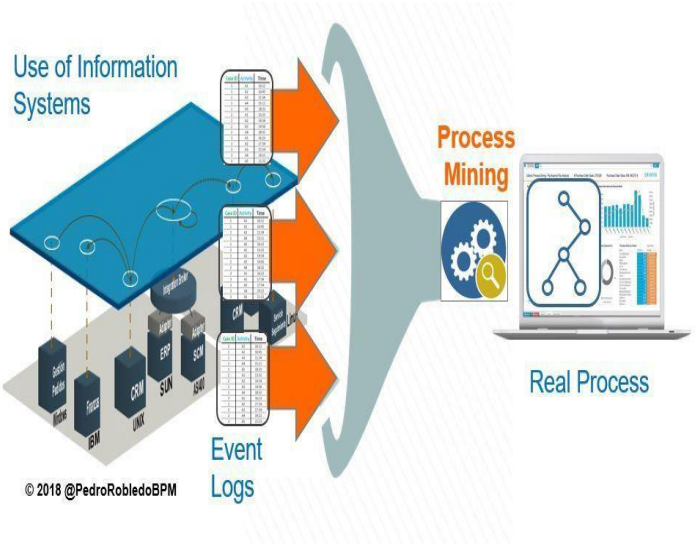

Fig. 3 Process Mining

Process mining is a research discipline that combines data mining techniques with business process modeling and analysis to discover, monitor, and improve real processes. There are three areas of process mining: process discovery, conformance verification, and enhancement. Process discovery algorithms generate a process model based only on actual process execution data found in event logs. Conformance verification deals with the analysis of process models and event logs with the objective of finding differences between the process model and the actual execution of the process. Finally, process enhancementextends and improves current process models usinginformation from the event log. A multi-dimensional quality assessment of several process discovery algorithms is presented. The algorithm using quality measures including accuracy, precision, recall, and comprehensibility. In one of their conclusions, works well with real-life context data in terms of accuracy, comprehensibility, and scalability.Therefore, it appears to be appropriate for data with considerable noise. This is a crucial characteristic that is desirable for analyzing SNS data. Therefore, we base our technique on the Heuristics miner algorithm adapted to our purpose.

III. INFORMATION FLOW MINING

In this section, the overall framework of the proposed approach is described. Moreover, the common structure and main characteristics of SNS data are introduced, as are the assumptions that allow the use of process mining techniques for this type of social data.

A. Framework

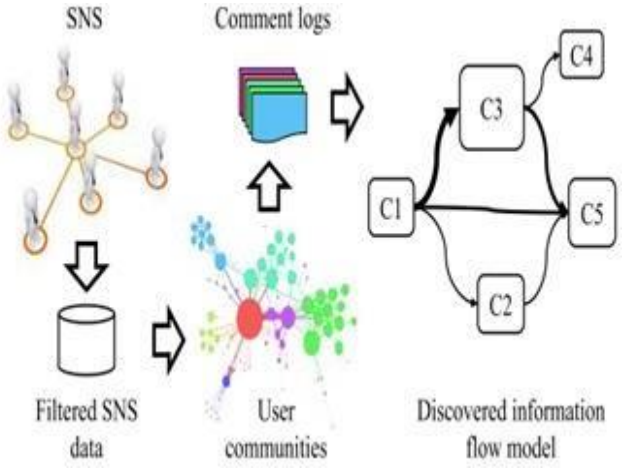

Fig. 4. Framework for mining information flow and SNS Data

Proposed methodology for mining information flow is shown in Fig. 4. SNS provide public interfaces for accessing data. Different SNS platforms exhibit similar basic data structures and attributes. Therefore, it is possible to apply our methodology to different SNS. Using the public interface, SNS data are gathered and filtered to reduce noise. The required data attributes are stored systematically for posterior use. A user clustering technique based on modularity maximization is applied to the SNS data to obtain communities of users. A comment log file is generated based on the communities found in the previous step. Finally, we apply the Info Flow miner, which is based on the FlexibleHeuristics Miner, to generate models of information diffusion.

Business data	SNS data
A process consists of business cases	A process consists of posts
A case consists of events related to precisely one case	A post consists of user actions related to precisely one post
Operation events within a case are ordered	User actions within a post are ordered
Operation events can have attributes such as time, cost, and resource	Users' actions can have attributes such as likes, timestamp, and location

Table 1. Models of Information Diffusion

B. Social Networking Service Data

Social network data are important for discovering knowledge about a community, which is critical in criminology, terrorism,public health, and many other applications. At the same time, there is a great deal of private information about individuals ina social network, which makes it sensitive when social

network data are shared across organizations. For SNS data, a process consists of a set of posts, which are created by users. Each post also contains user actions, such as likes and comments on Facebook and conversations and re-tweets on Twitter, which are precisely related to one post. User actions within a post are ordered by time. Actions can appear ordered by importance, but the timestamp of each action is recorded. Finally, user actions can have different attributes such as timestamp, likes, and shares on Facebook.

IV. USER COMMUNITY DETECTION

Communities are discovered based only on information found in the data log stored in the SNS. In other words, we do not use any prior knowledge about the social relationships between users. The interactions between users are measured with an intimacy function that quantifies how related any two users are based on the frequencies of their common actions in SNS. User communities are detected based on the intimacy values obtained from the data. The driving factors for data mining social networking sites is the "unique opportunity to understand the impact of a person's position in the network oneverything from their tastes to their moods to their health.". The most common data mining applications related to social networking sites include: 1. Group detection – One of the mostpopular applications of data mining to social networking sites is finding and identifying a group. In general, group detection applied to social networking sites is based on analysing the structure of the network and finding individuals that associate more with each other than with other users. Understanding what groups an individual belongs to can help lead to insights about the individual such as what activities, goods, and services, an individual might be interested in. 2. Group profiling – Once a group is found, the next logical question to ask is 'What is this group about' (i.e., the group profile)? The ability to automatically profile a group is useful for a variety purposes ranging from purely scientific interests to specific marketing of goods, services, and ideas. With millions of groups present in online social media, it is not practical to attempt to answer the question for each group manually. 3. Recommendation systems – A recommendation system analyses social networking data and recommends new friends or new groups to a user. The ability to recommend group membership to an individual is advantageous for a group that would like to have additional members and can be helpful toan individual who is looking to find other individuals or a group of people with similar interests or goals. Again, large numbers of individuals and groups make this an almost impossible task without an automated system. Additionally, group characteristics change over time. For those reasons, datamining algorithms drive the inherent recommendations made to users. From the moment a user profile is entered into a social networking site, the site provides suggestions to expand the user's social network. Much of the appeal of social networking sites is a direct result of the automated recommendations which allow a user to rapidly create and expand an online social network with relatively little effort on the user's part. Data mining is a powerful tool which will facilitate to seek out hidden patterns and various relationship between the data. Data processing discovers hidden facts from massive databases. The overall objective of the data mining technique is to extract 5 information from a huge data set and transform it into a comprehensible structure for more use. The different data Mining techniques are:

i) Characterization – used to generalize, summarize and possibly different data characteristics.

ii) Classification – is a process in which thegiven data is classified into different classes.

iii) Regression – is process similar to classification, the major difference is thatthe object to be predicted is continuous rather than discrete.

iv) Association – discovers the association between various data bases and theassociation between the attributes ofsingle database.

v) Clustering – involves grouping of data into several new classes such that it describesthe data. It breaks large data set into smaller groups to make the designing and implementation process to besimple.

vi) Change Detection – this method identifies the significant changes in the data from the previously measured values.

vii) Deviation Detection – focuses on the major deviations between the actual values of the objects and its expected values. This method finds out the deviation accordingto the time as well the deviation among different subsets of data.

viii) Link Analysis – traces the connectionsbetween the objects to develop models based on the patterns in the relationships by applying graph theory techniques.

ix) Sequential Pattern Mining – involves the discovery of the frequently occurring patterns in the data

A. Community Detection Using Hierarchical Clustering

A community is a smaller compressed group within a larger network (as shown in Fig.5). Community formation is known to be one of the important characteristics of social network sites. Users with similar interest form communities on social network thereby displaying strong sectional structure. Communities on social networks, like any other communities in the real world, are very complex in nature and difficult to detect. Applying the appropriate tools in detecting and understanding the behaviour of network communities is crucial as this can be used to model the dynamism of the domain they belong . Different authors have applied diverse clustering techniques to detect communities on social network,with hierarchical clustering being mostly used. This technique is a combination of many techniques used to group nodes inthe network to reveal strength of individual groups which is then used to distribute the network into communities. Vertex clustering belongs to hierarchical clustering methods, graph vertices can be resolved by adding it in a vector space so that pairwise length between vertices can be measured. Structural equivalence measures of hierarchical clustering concentrate on number of common network connections shared by two nodes.Two people on social network with several mutual friends are more likely to be closer than two people with fewer mutual friends on the network. Users in the same social network community often recommend items and services to one another based on the experience on the items or services involved.

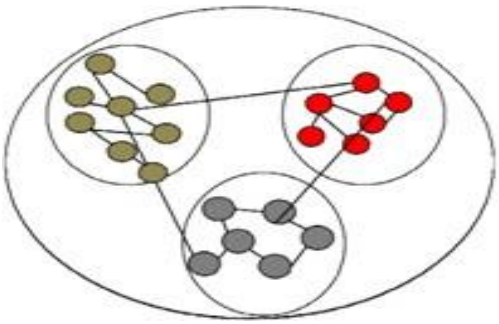

Fig. 5. Social Network Community Structure

V. SEMANTIC WEB OF SOCIAL NETWORK

The Semantic Web platform makes knowledge sharing and re-use possible over different applications and community edges. Discovering the evolvement of Semantic Web (SW) enhances the knowledge of the prominence of Semantic Web Community and envisages the synthesis of the Semantic Web. The work in employed Friend of a Friend (FOAF) to explore how local and global community level groups develop and evolve in large-scale social networks on the Semantic Web. The study revealed the evolution outlines of social structures and forecasts future drift. Likewise application model of Semantic Web-based Social Network Analysis Model creates the ontological field library of social network analysiscombined with the conventional outline of the semantic web toattain intelligent retrieval of the Web services. Furthermore, Voyeur Server improved on the open-source Web-Harvest framework for the collection of online social network data in order to study structures of trust enhancement and of online scientific association. Semantic Web is a relatively new areain social network analysis and research in the field is still evolving.

i) Opinion Analysis on Social Network

According to Technorati, about 75,000 new blogs and 1.2million new posts giving opinion on products and services are generated every day. Also massive data generated every minute on common social network sites are laden with opinionof users as regards diverse subject ranging from personal to global issues. Users opinions on social network sites can be referred to as discovery and recognition of positive or negativeexpression on diverse subject matters of interest. These opinions are often convincing and their indicators can be used as motivation when making choices and decisions on patronage of certain products and services or even endorsement of political candidate during elections. Even though online opinions can be discovered using traditionalmethods, this form is conversely inadequate considering the large volume of information generated on social network sites.This fact underscores the relevance of data mining techniques in mining opinion expressed on social network site.

ii) Aspect-Based/Feature-Based Opinion Mining

Aspect-based also known as feature-based analysis is the process of mining the area of entity customers has reviewed. This is because not all aspects/features of an entity are often reviewed by customers. It is then necessary to summarise the aspects reviewed to determine the polarity of the overall review whether they are positive or negative. Sentiments expressed on some entities are easier to analyse than others, one of the reason being that some reviews are ambiguous. According to aspect-based opinion problem lies more in blogsand forum discussions than in product or service reviews. The aspect/entity (which may be a computer device) reviewed is either 'thumb up' or 'thumb down', thumb up being positive review while thumb down means negative review. Conversely,in blogs and forum discussions both aspects and entity are not recognized and there are high levels of insignificant data which constitute noise. It is therefore necessary to identify opinion sentences in each review to determine if indeed each opinion sentence is positive or negative . Opinion sentences can be used to summarize aspect-based opinion which enhances the overall mining of product or service review. An opinion holder expresses either positive or negative opinion onan entity or a portion of it when giving a regular opinion and nothing else. However, put necessity on differentiating the twoassignments of finding out neutral from non-neutral sentiment,and also positive and negative sentiment. This is believed to greatly increase the correctness of computerised structures.

iii) Opinion Extraction

Sentiment analysis deals with establishment and classification of subjective information present in a material. This might not necessarily be fact-based as people have different feelingstoward the same product, service, topic, event or person. Opinion extraction is necessary in order to target the exact part of the document where the real opinion is expressed. Opinion from an individual in a specialised subject may not countexcept if the individual is an authority in the field of thesubject matter. Nevertheless, opinion from several entities necessitates both opinion extraction and summarization. In opinion extraction, the more the number of people that give their opinion on a particular subject, the more important that portion might be worth extracting. Opinion can aim at a particular article while on the other hand can compare two or more articles. The formal is a regular opinion while the latteris comparative. Opinion extraction identifies subjective sentences with sentimental classification of either positive or negative.

VI. UNSUPERVISED CLASSIFICATION OF SOCIAL NETWORK DATA

A straightforward unsupervised learning algorithm can be used to rate a review as 'thumbs up' or 'thumbs down' . This can be by way of digging out phrases that include adjective or adverbs (part of speech tagging) . The semantic orientation of every phrase can be approximated using PMI-IR and then classify the review using the average semantic orientation of the phrase. Cogency of title, body and comments generated from blog post has also been used in clustering similar blogs into significant groups. In this case keywords played veryimportant role which may be multifaceted and bare. EM-based and constrained-LDA were utilized to cluster aspect phrases into aspect categories. In two unsupervised frameworks basedon link structure of the Web pages, and Agglomerative/Conglomerate Double Clustering (A/CDC)was used to find group of individuals on the web. The result proves to be more accurate than those obtained by traditional agglomerative clustering by more than 20% while achieving over 80% F-measure. Other unsupervised learning used insentiment analysis in products rating and reviews include POS (Part of Speech) tagging. In POS adjectives are tagged todisplay positive and negatives ones. Sentiment polarity is the binary classification of an opinionated document into a largelypositive and negative opinion. In review this is commonly termed with the 'thumps up' and 'thumps down' expressionsas mentions earlier. The polarity of positive against negative isweighed to give an overall analysis of sentiment expressed on issue under review. Bootstrapping also forms part of the unsupervised approaches. It utilizes obtainable primary classifier to make labelled data which a supervised process canbuild upon. Semantic orientation is also an unsupervised approaches currently used for sentiment analysis on socialnetwork. It attaches different meaning to a single word – synonym. This could either be positive or negative (for example 'the party is bad' may in actual fact mean the party is fun). Direction and intensity of words used can determine the semantic orientation of the opinion expressed. Semi-supervisedand supervised classifications are more structured techniques as discussed next in Sections.

A. Semi-supervised Classification

Semi-supervised learning is a goal-targeted activity but unlike unsupervised; it can be specifically evaluated. Authors of worked on a mini training set of seed in positive and negative expressions selected for training a term classifier. Synonym and antonym comparatives were added to the seed sets in an online dictionary. The approach was meant to produce the extended sets P' and N' that makes up the training sets. Other learners were employed and a binary classifier was built using every glosses in the dictionary for both term in P'U N' and translating them to a vector. Their approach discovers theorigin of information which they reported was missing in earlier techniques used for the task. Semi-supervised

lexical classification proposed by integrated lexical knowledge into supervised learning and spread the approach to comprise unlabelled data. Cluster assumption was engaged by grouping together two documents with the same cluster basically supporting the positive - negative sentiment words as sentiment documents. It was noted that the sentiment polarity of document decides the polarity of word and vice versa. In

[72] semi-supervised learning uses polarity detection as semi supervised label propagation problem in graphs. Each node representing words whose polarity is to be discovered. The results shows label propagation progresses outstandingly above the baseline and other semi supervised techniques like Min cuts and Randomized Min cuts. The work of [38] compared graph-based semi-supervised learning with regression and [65] proposed metric labelling which runs SVM regression as the original label preference function comparable to similarity measure. Their result shows that the graph-based semi-supervised learning (SSL) algorithm as per PSP (positive-sentence-percentage) comparison (SSL+PSP) proved to perform well.

B. Supervised Classification

While clustering techniques are used where basis of data is established but data pattern is unknown, classification techniques are supervised learning techniques used where the data organisation is already identified. It is worthy of mention that understanding the problem to be solved and opting for the right data mining tool is very essential when using data miningtechniques to solve social network issues. Pre-processing and considering privacy rights of individual (as mentioned under research issues of this paper) should also be taken into account. Nonetheless, since social media is a dynamicplatform, impact of time can only be rational in the issue of topic recognition, but not substantial in the case of network enlargement, group behaviour/ influence or marketing. This is because this attributes are bound to change from time to time. Information updates in some Social network such as twitters and Facebook present Application Programmers Interfaces (APIs) that makes it possible for crawler, which gather new information in the site, to store the information for later usage and update. In a supervised learning algorithm used the combination of multiple bases of facts to label couple of adjectives having similar or dissimilar semantic orientations. The algorithm resulted in a graph with nodes and links which represents adjectives and similarity (or dissimilarity) ofsemantic orientation respectively.

VII. TOPIC DETECTION AND TRACKING ONSOCIAL NETWORK

Topic Detection and Tracking (TDT) on social network employs different techniques for discovering the emergent of new topics (or events) and for tracking their subsequent evolvements over a period of time. TDT is receiving high level of attention recently. Many researchers and authors are conducting experiments on TDT on social network sites, especially on Twitter. In support vector machine (SVM) was found to be efficient in training Twitter hash tags metadata when predicting the political alignment of twitter users. Authors of used an incremental online clustering algorithm to cluster a stream of Twitter messages in real time. They trained a Naïve Bayes-Text classifier to distinguish between fastest-growing real-world events contents and non-events contentson Twitter. The performance of the training set shows the precision of all classifier computed in 10-fold cross-validation.The experiments in used a range of query-building approaches to automatically enhance user-contributed information for planned events with robustly generatedTwitter contents. Their approach used browser plug-in script and a customizable web interface to identify relevant Twitter content for planned events. The experiments in proposed a combination of six techniques namely; LDA (Latent Dirichlet Allocation),

Doc-p (Document-Pivot Topic Detection, GFeat- p (Graph-based Feature-Pivot Topic Detection), FPM (Frequent Pattern Mining), SFPM (Soft Frequent Pattern Mining) and BN gram for real-world event detection on Twitter network. The techniques were verified on tweets relating to three major events (English FA Cup Finals, US Super Tuesday Primaries and US Elections 2012) with variations in time scale and topic mix level. The algorithms revealed that dataset pre-processing and sampling process affects the quality of topic retrieved. Conversely, the algorithms performed optimally on the three datasets considered. Similarly, proposed an algorithm for detecting and tracking breaking news in Twitter. The application named "Hot stream" was built to afford its users the opportunity of detecting and tracking breaking news from Twitter timeline. Authors of 63 proposed a state-of-the-art First Story Detection (FSD) technique to detect predictable and unpredictable events using real-time indication from Wikipedia and Twitter data streams. The result of the experiments recorded about 2-hour delay for Wikipedia in real- world events. Authors in used ED CoW (Event Detection with Clustering of Wavelet-based Signals) to cluster words to form events with a modularity- based graph partitioning method. On the other hand employed lightweight event detection using wavelet signal analysis of hashtags occurrences in Twitter public stream. The experiments used Latent Dirichlet Allocation topic inference model based on Gibbs Sampling. The outcome of the experiments shows that peak detection using Continuous Wavelet Transformation realized impressive outcomes in the ascertaining abrupt increases on the mention of specific hashtags. In the abruptness in hashtags usage is labelled unexpected rule evolvement which is discussed under TRCM in next section.

A. TRCM for TDT

Twitter as a social network and hashtags as tweet labels can beanalysed in order to detect changes in event patterns (TDT) using Association Rules (ARs). Twitter data can be used to analyse patterns associated with events by detecting the dynamics of the tweets. Association Rule Mining (ARM) can find the probability of co- existence of tweets' hashtags. Firstly, in ARM was used to analyse tweets on the same topic over consecutive time periods t and t+1. Rule Matching (RM) was later employed to detected changes in patterns such as 'emerging (EM)', 'unexpected consequent (Unx Cs)' and 'unexpected conditional (Unx Cn)', 'new (N)' and 'dead (D)' rules in tweets. This is obtained by setting a user-defined Rule Matching Threshold (RMT) to match rules in tweets at time t with those in tweets at t + 1 in order to ascertain rules that fall into the different patterns (as presented in Fig. below). The proposed methodology was coined TRCM (Transaction-based Rule Change Mining). All the detected rules in TRCM were linked to real life events and news reports. Subsequently, TRCM was utilized to discover the rule trend of tweets' hashtags over a consecutive period. Time Frame Windows (TFWs) was created (as shown in Fig.6) to describe different rule evolvement patterns which can be applied to evolvements of news and events in reality. This concept of TRCM was named TRCM-RTI (Transaction-based Rule Change Mining-Rule Type Identification). Time frame window(Fig.7) also serve as a means of calculating the lifespan of specific hashtags on Twitter. Using the experimental study result in, it was substantiated that the lifespan of tweets' hashtags can be related to evolvements of news and events in reality. TheTRCM techniques and can be said to be the first time ARM was used to mine Twitter data. This therefore opens up thearea for further research as ARM can be fine-tuned to be used

to mine data on other social networks sites for information retrieval and knowledge acquisition.

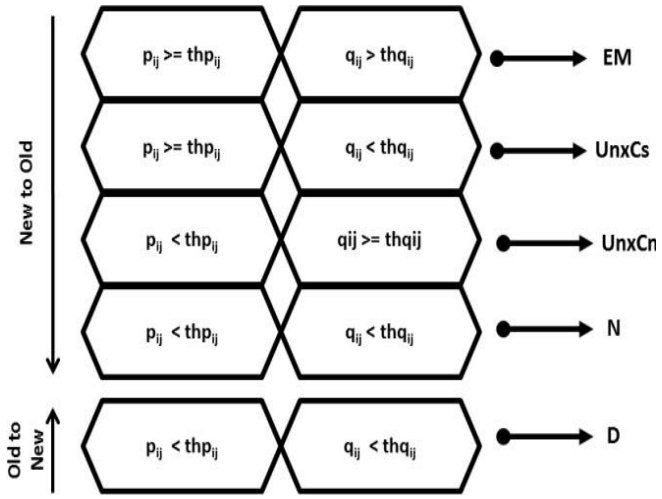

Fig.6 Time Frame Windows

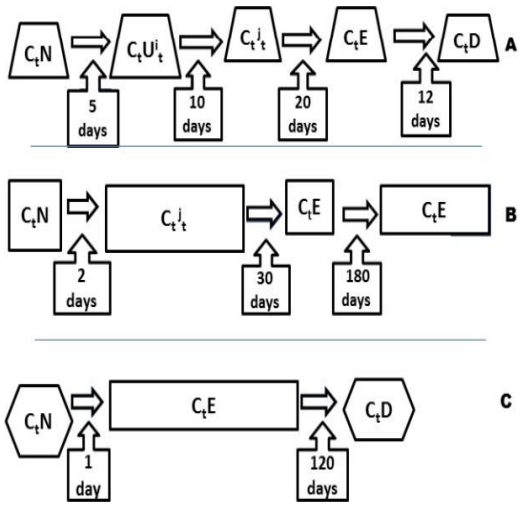

Fig.7 Time Frame Window

VIII. CONCLUSION AND FUTURE WORK

Different data mining techniques have been used in socialnetwork analysis as covered in this survey. The techniques range from unsupervised to semi-supervised and supervised learning methods. So far different levels of successes have being achieved either with solitary or combined techniques. The outcome of the experiments conducted on social network analysis is believed to have shed more light on the structure and activities of social network. The diverse experimental results have also confirmed the relevance of data mining techniques in retrieving valuable information and contents from huge data generated on social network. Future surveywill tend to investigate novel state-of-the-art data mining techniques for social network analysis. The survey will compare similar data mining tools and recommend the most suitable tool(s) for the dataset to be analysed. Different data mining techniques

covered in this survey are listed in the tablealso contains the approaches employed, the experimental results and the dates and authors of the approaches.

REFERENCES

[1]. A. Guille, H. Hacid, C. Favre, and D. A. Zighed, ``Information diffusion in online social networks: Asurvey,'' ACM SIGMOD Rec., vol. 42, no. 2, pp. 17_28, 2013.

[2]. W. M. P. van der Aalst, Process Mining: Data Science in Action. Heidelberg, Germany: Springer, 2016.

[3]. M. De Choudhury, ``Discovery of information disseminators and receptors on online social media,'' in Proc. 21st ACM Conf. Hypertext Hypermedia, Toronto,ON, Canada, 2010, pp. 279_280.

[4]. J. L. Moreno, Who Shall Survive?: A New Approach tothe Problem of Human Interrelations. Washington, DC, USA: American Sociological Association, 1934.

[5]. M. E. J. Newman, Networks: An Introduction. New York, NY, USA: Oxford Univ. Press, 2010.

[6]. M. Bastian, S. Heymann, and M. Jacomy, ``Gephi: An open source software for exploring and manipulating networks,'' in Proc. Int. Conf. Weblogs Social Media, San Jose, CA, USA, 2009, pp. 361_362. Mech.

[7]. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, ``Fast unfolding of communities in large networks,'' J. Stat. Mech., Theory Exp., vol. 2008, no. 10, 2008, Art. no. P10008.

[8]. A. J. M. M. Weijters and J. T. S. Ribeiro, ``Flexible heuristics miner (FHM),'' in Proc. IEEE Symp. Comput. Intell. Data Mining (CIDM), Paris, France, Apr. 2011, pp. 310_317.

[9]. J. De Weerdt, M. De Backer, J. Vanthienen, and B. Baesens, ``A multidimensional quality assessment of state- of-the-art process discovery algorithmsb using real-life event logs,'' Inf. Syst., vol. 37, no. 7, pp. 654_676, 2012.

[10].M.-S. Song, M. Günther,W. M. P. van der Aalst, and J.-Y. Jung, ``Improving process mining with trace clustering,'' J. Korean Inst. Ind. Eng., vol. 34, no. 4, pp. 460_469, 2008.

[11].S. Fortunato, ``Community detection in graphs,'' Phys. Rep., vol. 486, nos. 3_5, pp. 75_174, 2010.

[12].W. M. P. van der Aalst, T. Weijters, and L. Maruster, ``Work_ow mining: Discovering process models from event logs,'' IEEE Trans. Knowl. Data Eng., vol. 16, no. 9, pp. 1128_1142, Sep. 2004.

[13].A. Rozinat and W. M. P. van der Aalst, ``Conformance testing: Measuring the _t and appropriateness of event logs and process models.,'' in Business Process ManagementWorkshops. Berlin, Germany: Springer, 2006, pp. 163_176.

[14].W. M. P. Aalst, van der and M. Song, ``Mining social networks: Uncovering interaction patterns in business processes,'' in Proc. Int. Conf. Bus. Process Manage. (BPM), in Lecture Notes in Computer Science, vol. 3080, J. Desel, B. Pernici, and M. Weske, Eds. Berlin, Germany: Springer, 2004, pp. 244_260.

[15]. R. J. C. Bose, E. H. M. Verbeek, andW. M. P. van der Aalst, ``Discovering hierarchical process models using ProM,'' in IS Olympics: Information Systems in a Diverse World. Berlin, Germany: Springer, 2012, pp. 33_48.

[16]. R. J. C. Bose and W. M. P. van der Aalst, ``Context aware trace clustering: Towards improving process mining results,'' in Proc. SDM, 2009, pp. 401_412.

[17].A. K. A. de Medeiros et al., ``Process mining based on clustering: A quest for precision,'' in Business Process Management Workshops. Berlin, Germany: Springer, 2008,pp. 17_29.
[18].J.-Y. Jung, ``PROCL: A process log clustering system,'' J. Soc. e-Bus. Stud., vol. 13, no. 2, pp. 181_194, 2008.
[19].D. Ferreira, M. Zacarias, M. Malheiros, and P. Ferreira, ``Approaching process mining with sequence clustering: Experiments and _ndings,'' in Business Process Management. Berlin, Germany: Springer, 2007, pp. 360_374.
[20].C. Di Francescomarino, A. Marchetto, and P. Tonella, ``Cluster-based modularization of processes recovered from web applications,'' J. Softw., Evol. Process, vol. 25, no. 2, pp. 113_138, 2013.
[21].A. Bogarín, C. Romero, R. Cerezo, and M. Sánchez- Santillán, ``Clustering for improving educational process mining,'' in Proc. 4th Int. Conf. Learn. Anal. Knowl., Indianapolis, Indiana, 2014, pp. 11_15.
[22]. J. De Weerdt, S. vanden Broucke, J. Vanthienen, and B. Baesens, ``Active trace clustering for improved process discovery,'' IEEE Trans. Knowl. Data Eng., vol. 25, no. 12, pp. 2708_2720, Dec. 2013.
[23].K. Kim, J.-Y. Jung, and J. Park, ``Discovery of information diffusion process in social networks,'' IEICE Trans. Inf. Syst., vol. E95-D, no. 5 pp. 1539_1542, 2012.
[24].K. Kim, J. Obregon, and J.-Y. Jung, ``Analyzing information _ow and context for Facebook fan pages,'' IEICE Trans. Inf. Syst., vol. 97, no. 4, pp. 811_814, 2014.
[25].B. Carrera, J. Lee, and J.-Y. Jung, ``Discovering information diffusion processes based on hidden Markov models for social network services,'' in Asia Paci_c Business Process Management (Lecture Notes in Business Information Processing), vol. 219, J. Bae, S. Suriadi, and L. Wen, Eds. Cham, Switzerland: Springer, 2015, pp. 170_182.

26].D. Li, S. Zhang, X. Sun, H. Zhou, S. Li, and X. Li, ``Modeling information diffusion over social networks for temporal dynamic prediction,'' IEEE Trans. Knowl. Data Eng., vol. 29, no. 9, pp. 1985_1997, Sep. 2017.
[27].C. Jiang, Y. Chen, and K. J. R. Liu, ``Evolutionary dynamics of information diffusion over social networks,'' IEEE Trans. Signal Process., vol. 62, no. 17, pp. 4573_4586, Sep. 2014.
[28].Y. Hu, R. J. Song, and M. Chen, ``Modeling for information diffusion in online social networks via hydrodynamics,'' IEEE Access, vol. 5, pp. 128_135, 2017.
[29].K. Zhang, J. Wang, C. Jiang, Z. Wei, and Y. Ren, ``Bigdata driven information diffusion analysis and control in online social networks,'' in Proc. IEEE Int. Conf. Commun. (ICC), Paris, France, May 2017, pp. 1_6.
[30].J. Wang, C. Jiang, T. Q. S. Quek, X. Wang, and Y. Ren, ``The value strength aided information diffusion in socially-aware mobile networks,'' IEEE Access, vol. 4, pp. 3907_3919, 2016.
[31].V. Arnaboldi, M. Conti, M. La Gala, and A. Passarella, and F. Pezzoni, ``Information diffusion in OSNs: the impact of nodes' sociality,'' in Proc. 29th Annu. ACM Symp. Appl. Comput., Gyeongju, South Korea, 2014, pp. 616_621.