Journal of Engineering and Applied Sciences 12 (Special Issue 9): 8505-8512, 2017
ISSN: 1816-949X
© Medwell Journals, 2017

An Overview of Clustering the Virality of Memes

Shamsiah Abd Kadira, Anitawati Mohd Lokman and Ahmad Azran Awang
Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia

Abstract: The amount of information shared on online media is increasing at unprecedented rates in recent years. In the realm of social media, the internet users use hashtags, Tweets, posts or even funny images, not only to interpret specific events but also to promote ideas and discussions known as internet memes; this study thus provides an overview of an array of memes communication-viral memes, clustering and classification of memes. This study employs the narrative synthesis to analyse past literatures. In addition, this study attempts to provide an understanding on techniques that could classify and cluster viral memes and ultimately strives to provide evidence on the areas that could realise an integration of clustering framework and viral memes.

Key words: Cluster, memes, viral, analyse past literature, clustering frame work, integration

INTRODUCTION

A meme is a virally transmitted cultural symbol or social idea. A meme (rhymes with “team”) behaves like flu or a cold virus but transmits an idea instead of a life form. Memes cover a wide spectrum and can be classified into two categories: one is typically a snapshot with a caption on it (usually the proper name that becomes the vernacular for this type of meme) while the other category of memes is anything else relating to popular internet videos, hashtags, tweets, posts, jokes or trends (Baumgartner, 2008). In this situation, successful memes go on to cross cultures and live on, whilst unsuccessful memes become extinct. Accordingly, this study discusses the internet meme or cultural artefact, generally consisting of user-generated content such as hashtags or tweet posts that are shared widely and remixed in various ways. An expansive adoption of online social networks not only makes internet memes possible but also offers the community with valuable data on the spreading of memes and user behaviour (Vespignani, 2009; Lazer *et al*., 2009). A meme may become viral merely because it appeals to many (Berger and Milkman, 2012; Cataldi *et al*., 2010). The success of memes depends on timing, social network structure, randomness and many other factors (Centola, 2010; Weng *et al*., 2014; Pinto *et al*., 2013). The virality of memes has been scrutinised from various perceptions including their innate attractiveness and its influential role along with their adoption patterns (Weng *et al*., 2014). Basically, the virality’s contribution is derived from the innate appeal of a meme, for example, Berger and Milkman (2012) examined the content of new’s emotion and they discovered articles that suggested provocation to be more viral.

In addition, this study also discusses data clustering. Cluster analysis is an exploratory data analysis tool for solving classification problems. In this situation, the object is to sort cases into groups, so that, the degree of association is strong between members of the same cluster and become weak between members of different clusters. Performing the cluster analysis would help researchers to identify a classification scheme, suggest statistical model to describe populations, indicate rules for assigning new cases to classes for identification as a means of targeting as well as for diagnostic purposes. Moreover, it could also provide measures of definition, size and change in what were previously broad concepts, as well as find typical cases to represent classes.

Clustering can be categorised as unsupervised learning in social media mining. Unsupervised learning is the division of the instances into groups of similar objects. In clustering, the data is often unlabelled. Also, the clustering algorithm does not know the label for each instance (Zafarani *et al*., 2014). Consequently, the uncontrolled nature of social media makes Twitter, Facebook and Instagram vulnerable to manipulation as well as exploitation for spreading spam, rumours, slanders and other types of misinformation (Chew and Eysenbach, 2010; Metaxas and Mustafaraj, 2010). In this situation in the domain of politics, for example, the subtler phenomenon of astroturfing has received attention in the recent literature (Ratkiewicz *et al*., 2011). Aastroturfing arises when one or a few individuals make a coordinated effort to create the false impression of a spontaneous movement including users to deem the information reliable and feed its propagation (Ferrara *et al*., 2013).

Effective use of memes requires internet users to have knowledge on trending topics, to enable effective design

Corresponding Author: Anitawati Mohd Lokman, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia

and strategy of such memes. Trending topics will need to be based on the cluster of memes. However, this difference of cluster is not present and these trending topics are hard to visualise. Hence, this study strives to discuss memes, the virality of internet meme, clustering, as well as classification of memes. Narrative synthesis is employed as the method to realise the discussion whilst this study attempts to provide an overview and a discussion on clustering and the integration of clustering and virality of internet memes.

Literature review: Apart from characterised viral memes and clustering framework, there is also another social network analysis that has been used dynamically in the social media analysis as well as in the area of web intelligence, particularly on internet memes. Some researchers attempt to identify influential members in a community, so as to contain the spread of misinformation or rumours. Other than that, several researchers suggest models of how events could disseminate through online communities and use these track memes through specific social media (Adar and Adamic, 2005; Lin *et al.*, 2010) as well as investigate the interplay between social media and traditional media (Leskovec *et al.*, 2009).

Although, Wu and Huberman (2007) and Romero *et al.* (2011) claim that time series analysis is the best method to detect viral memes, it is still lacking the ability to identify the viral memes at their early stages. Viral memes could be most influential at their early stages-as what has been said by Weng *et al.* (2013) Lindzey and Aronson (1985) study because it may gain attention as well as spread the memes based on their potential audience. As mentioned earlier, this study has given an example on how Bauckhage (2011) study identified internet memes by using time series analysis. In the study, the researcher describes internet memes as evolving content that rapidly gains popularity on the internet. In addition, memes are spread voluntarily, rather than in a compulsory manner which fact, although true, does not describe the full picture. In this consequence, this study scrutinised that not every meme is produced by advertising or community campaigns as what has been mentioned by Bauckhage (2011). In fact, these campaigns are expected to have different behaviours on organically and non-strategically created memes.

There are related studies that employ the time series analysis in order to detect the Internet memes. The former cluster time series is obtained from a micro-blogging service in order to predict future interest in a topic (Yang and Leskovec, 2011; Kubo *et al.*, 2007). In its final investigation, the time-based evolution of content in bulletin boards and report which is a simple stochastic compartment model, gives a good account of the process. However, concerned with internet viral memes, this study could not support these findings and employ them as the possible integration that could cluster the memes.

Leskovec *et al.* (2009) suggest the meme-tracking approaches that could be a platform which tracks memes produced in online media such as mainstream news sites as well as blogs. The study groups together short, distinctive phrases which act as signatures of specific topics and identify small variations of them. Nevertheless, the meme-tracker only identifies and aggregates disjoint memes on the basis of textual similarity, yet no systematic evaluation of the quality of the retrieved memes is provided. Thus, this current study finds that, Ferrara *et al.* (2013) study focuses more on the assessment of the quality of the memes clustering process and allows for overlapping memes.

Simmons *et al.* (2011) also tackled the problem of tracking news for meme extraction. According to the meme-tracker dataset which has been used in the study, they examine the extent to which information evolves and mutates due to collective processing of social media users. However, while defining protomemes, Ferrara *et al.* (2013) root their investigation on the findings of both Leskovec *et al.* (2009) and Simmons *et al.* (2011) studies, increasing on the combination of memes variants based not only on textual similarity but also on other network and meta-data features.

Aggarwal and Subbian (2012) share some similarities framework with Ferrara *et al.* (2013) study with another line of research on event detection systems. In the study, a clustering algorithm is presented that exploits both content and network-based features to detect events in social streams. Nonetheless, the algorithm employed could only assume a pre-existing knowledge about the follower network of Twitter users. In this situation, especially on streaming scenario such information is expensive to obtain, specifically when encountering popular users. Likewise, time-based features in addition to social and topical ones can be exploited by employing an event classification system designed for Twitter (Becker *et al.*, 2011). These features are adopted to train a classifier that consumes manually annotated clusters of data points representing specific events on Twitter. As a result, they represent a starting point in the task of classifying different types of memes in Twitter.

MATERIALS AND METHODS

This study is based on a systematic analysis of 48 research studies in pertaining to memes and its types as

Table 1: Synthesised studies on social network analysis

Title	Researchers
Tracking information epidemic in blogspace	Adar and Adamic (2005)
The political blogosphere and the 2004 US election	Adamic and Glance (2005)
Event detection in social stream	Aggarwal and Subbian (2012)
Quantifying influence on Twitter	Bakshy *et al*. (2009)
Insight into internet memes	Bauckhage (2011)
The effects of digital political satire evaluations	Baumgartner (2008)
Real-world event identification on Twitter	Becker *et al*. (2011)
Viral online content	Berger and Milkman (2012)
Misinformation in social network	Budak *et al*. (2010)
Topic detection on Twitter based on temporal and social terms evaluation	Cataldi *et al*. (2010)
Behaviour in an online social network experiment	Centola (2010)
Complex contagions and the weakness of long ties	Centola and Macy (2007)
Content analysis of Tweets during the 2009 H1N1 outbreak	Chew and Eysenbach (2010)
Early warning analysis for social diffusion events	Colbaugh and Glass (2012)
Epidemics and rumours	Daley and Kendall (1964)
Clustering memes in social media	Ferrara *et al*. (2013)
Trajectory clustering with regression models	Gaffney and Smyth (1999)
Using blogs to provide context for news articles	Gamon *et al*. (2008)
Generalisation of epidemic theory transmission of ideas	Goffman and Newill (1964)
Large-scale sentiment analysis for news and blogs	Godbole *et al*. (2007)
Threshold models of collective behaviour	Granovetter (1978)
Description and prediction of slashdot activity	Kaltenbrunner *et al*. (2007)
Comment mining, popularity prediction and social network analysis	Jamali amd Rangwala (2009)
Identification of influential spreaders in complex networks	Kitsak *et al*. (2010)
Academic meme analogy for web community	Kubo *et al*. (2007)
Computational social science	Lazer et al. (2009)
Trajectory clustering: a partition-and-group framework	Lee *et al*. (2007)
Dynamical classes of collective attention in Twitter	Lehmann *et al*. (2012)
Non-parametric time series classification	Lenser and Veloso (2005)
A model of social dynamics to predict popularity of news	Lerman and Hogg (2010)
Meme-tracking and the dynamic of the news cycle	Leskovec *et al*. (2009)
Group psychology and the phenomena of interaction	Lindzey and Aronson (1985)
A statistical model for popular events tracking in social media	Lin *et al*. (2010)
A nearest trajectory strategy for time series prediction	McNames (1998)
Homophily in social networks	McPherson *et al*. (2001)
Political speech and real-time search	Metaxas and Mustafaraj (2010)
Early view patterns and the popularity of youtube videos	Pinto *et al*. (2013)
Detecting and tracking political abuse in social media	Ratkiewicz *et al*. (2011)
Mechanics of information diffusion across topics on Twitter	Romero *et al*. (2011)
Rumours in a network	Shah and Zaman (2009)
Memes online	Simmons *et al*. (2011)
Predicting the popularity of online content	Szabo and Huberman (2010)
Predicting behaviour of techno-social system	Vespignani (2009)
Virality prediction and community structure in social networks	Weng *et al*. (2013)
Predicting successful memes using network and community structure	Weng *et al*. (2014)
Novelty and collective attention	Wu and Huberman (2007)
Patterns of temporal variation in online media	Yang and Leskovec (2011)
Social media mining	Zafarani *et al*. (2014)

well as the social network analysis methods. In addition, this study focuses on synthesising data in relation to viral memes, clustering memes and the integration of viral memes technique from the literatures. This enables the current study to realise the analysis on the feasibility of clustering the virality of memes (Table 1).

No strict timeframe is imposed for the gathered data but any work after 2000 is given additional attention. However, any work before the year 2000 may still be relevant to provide greater comprehension of the discussion. The data search strategy of this study involves obtaining literatures from online journal databases, then data abstraction strategy takes place, in which this study extracts the significance of the data-from each past literature yielded-in pertaining to the selected topic.

Finally, findings from data abstraction strategy are synthesised to develop a summary of each past study. This study utilises narrative synthesis: a method of synthesising evidence relevant to an extensive assortment of questions including a systematic review. However, mathematical data and statistical tools are not present in this study as it simply strives to seek for areas of clustering memes and stipulates the possibility of integrating clustering and the virality of memes.

Clustering the viral memes: The spread of memes is often considered as social contagion or "infection", generally distinguished as the spread of information or behaviour on social networks where an individual serves as the stimulus for the imitative actions of another (Lindzey and Aronson, 1985; Goffman and Newill, 1964; Daley and Kendall, 1964; Weng *et al*., 2014). Nevertheless, information contagion may spread differently from diseases as multiple exposures could significantly increase the chances of adoption (Centola, 2010; Granovetter, 1978; Romero *et al*., 2011). In this situation, time series analysis is one of the common methods to detect viral memes (Wu and Huberman, 2007; Romero *et al*., 2011). For example, Bauckhage (2011) uses time series analysis to seek the temporal dynamics of internet memes. Characteristic similarities and differences among the data from different sources are analysed such as from Google Insight and social bookmarking services. The analysis discovers user communities of the considered services appear to have different interests and manifest behaviours that reflect different aspect of viral memes.

On the other hand, temporal patterns of memes could be well summarised into a few categories, and they have predictive power to spot trendy or bursty memes (Yang and Leskovec, 2011; Lehmann *et al*., 2012). The classification of temporal patterns could be seen as a lengthy application of route clustering (Gaffney and Smyth, 1999; Lee *et al*., 2007). However, existing virality prediction algorithms try to forecast time series based on past values (McNames, 1998; Lenser and Veloso, 2005; Kaltenbrunner *et al*., 2007; Weng *et al*., 2013). Also, some event detection methods group memes together to form topics and use temporal activity to detect trending topics (Cataldi *et al*., 2010; Becker *et al*., 2011).

Weng *et al*. (2013) identified signatures of viral memes at their early stages in terms of three characteristics: network topology, community diversity and growth rate. They have demonstrated the information on early adopters, particularly in the perspective of social network structure which are powerful enough to identify young viral memes. For the network topology, the position of an adopter in the network determines the size of the potential audience (Kitsak *et al*., 2010). In this situation (Weng *et al*., 2014), estimated the growth of the potential audience in time by examining the distance between consecutive adopters in the network. They also examined the characteristics of social contagions which are known to possess two distinctive characteristics such as social reinforcement (Centola, 2010; Romero *et al*., 2011; Pinto *et al*., 2013; Centola and Macy, 2007; Bakshy *et al*., 2009) and homophily or social relationships (Centola, 2010; McPherson *et al*., 2011). Therefore, community structure has been shown to help quantify the strength of these effects (Colbaugh and Glass, 2012; Weng *et al*., 2013).

Other than that, DESPIC is the platform that could be applied to classify different types of communication, specifically on memes (Ferrara *et al*., 2013). Types of communication-memes could be identified in the DESPIC Model and the model possesses two components: a message-clustering algorithm which takes a stream of tweets and groups them into memes and a meme classification algorithm that labels these memes according to categories of interest or communication. Ferrara *et al*. (2013) also focused on the clustering framework using Twitter as a test-bed scenario. Therefore, they propose a strategy that leverages various sources of available meta-data in addition to text. For example, Tweets may contain hashtags, confidentially described textual tokens-used to identify topics of discussion. Then, they easily group messages based on atomic entities called protomeme.

Protomemes encompass several similarity measures, which could leverage various features including content and network-based ones to build clusters of semantically and structurally related Tweets (Ferrara *et al*., 2013). The proposed diffusion similarity measure uses mention and re-Tweet information that could be reconstructed in real-time from the observed data, considering each protomeme diffusion set. Nevertheless, this approach performs as well as methods that assume prior knowledge of the data and better than methods that assume knowledge of the underlying social network.

RESULTS AND DISCUSSION

The integration of viral meme and clustering framework: An integration between viral memes and a platform to classify types of communication memes is believed feasible-the integration of viral memes which could predict the future popularity of a meme with three intuitive classes of features as well as formalise the framework to cluster memes from social media.

As at the early stages, it is suggested from the literatures to adopt a network which provides information on the size of potential audience groups that may affect the future popularity as well as to investigate the community diversity: as a good predictor of virality and to investigate the early growth rate of a meme usage which could be generalised to predict its future popularity (Weng *et al*., 2014).

By using the technique, this study discovers community based features perform the best among the three classes of viral memes. The community based feature includes social reinforcement which could drastically increase the probability of the adoption and involves homophily that are more likely to be formed between people who share similar characteristics such as interest culture and language-these increase the chances of adopting similar memes. However, there is a problem to predict a viral meme, namely in clustering and classification task (Weng *et al*., 2014) because the early popularity of online content is strongly correlated with its future popularity of memes. Moreover, there are recent literatures that confer related problems of clustering memes on online social media such as the identification of topics on memes as well as emerging events in social streams (Jamali and Rangwala, 2009; Szabo and Huberman, 2010; Lerman and Hogg, 2010). In conjunction with that, Ferrara *et al*. (2013) suggest a clustering framework that adopts an innovative pre-clustering procedure-the protomeme detection-that aims at identifying atomic tokens of information inside Tweets. Due to its ability, this strategy is appropriate to work in streaming scenario, for example, the viral memes.

Figure 1 indicates the possible integration between the viral meme: community-based features with the clustering framework which could classify the types of communication shared on social media.

In the community-based features, the dense connectivity inside a community could spread the chances of multiple revelations which consequently improves the contagion that is sensitive to social reinforcement. In that community-based, the groups with the same interests certainly create more edges among them, thus forming the communities. Subsequently, members of the same community are more probable to share similar interests. These two effects are solid, upon which the communities would facilitate the internal circulation of memes while preventing diffusion across communities, causing strong concentration or low community diversity. For example, detested memes tend to be focused in a small number of communities, whilst few viral memes have excessive community variety, spreading widely across communities like widespread outbreaks.

In clustering framework, several similar protomemes have been defined, leveraging various features including content and community-based to build clusters of semantically and structurally protomemes in related tweets. With such integration, this study could measure the viral memes through re-Tweet information which could reconstruct in real time from the observed data, considering each protomeme's dispersion set. Thus, to evaluate the promising performance of the clustering framework, it may use a ground truth dataset to curate the datase probable to share similar interests. These two effects are solid, upon which the communities would facilitate the internal circulation of memes while preventing diffusion across communities, causing strong concentration or low community diversity. For example, detested memes tend to be focused in a small number of communities, whilst few viral memes have excessive community variety, spreading widely across communities like widespread outbreaks t manually. The best trade-off between quality, number and size of clusters is obtained by pre-clustering using protomemes and combining similarity measures exploiting heterogeneous features, with simple pairwise maximisation strategy (Ferrara *et al*., 2013).

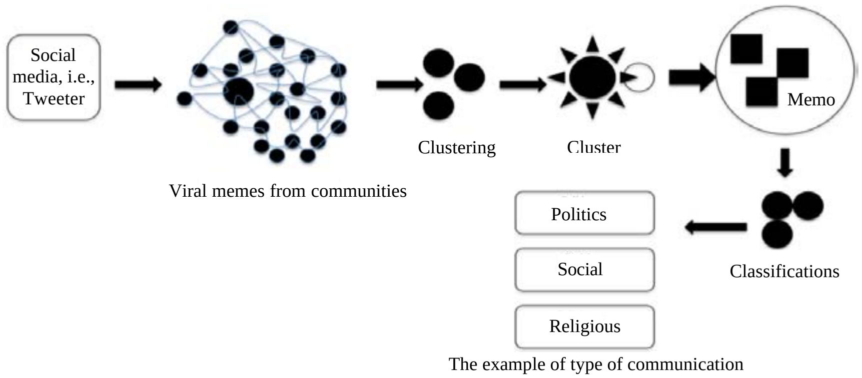

Fig. 1: Clustering the viral memes from the community-based features

The successfulness of clustering viral memes would help internet users or even data scientists to recognise the types of meme communications. Therefore, opinions are mixed: people cannot agree on whether this is a good thing or not. For instance, some believe the spread of political memes represents an epic win for crowd-sourced democracy whereas for others, they are signs of an intellectual apocalypse. By using these approaches, this study is also able to identify topics of discussion, mentions of other users that are used to address messages to their attention as well as help to identify the contributors of conversations.

As a result, the combination of the approaches with information about the political orientations (as an example) of the different news media and blog sources, could show the particular threads moving within and between opposed groups (Adamic and Glance, 2005; Gamon *et al*., 2008; Godbole *et al*., 2007). Nevertheless, this approach would be useful to further understand the roles of different participants play in the process as their collective behaviour leads directly to the ways in which individuals experience news and its consequences.

From now on, this method could be extended to the set of features combined by the clustering framework for instance, images. On the other hand, this study suggests that the initiation of time series as a feature may yield significant performance improvement. Memes clustering framework could also be integrated with the meme classifier to distinguish engineered types of social media communication from spontaneous ones. This framework or platform may perhaps adopt supervised learning technique to classify memes and determine their legitimacy with the aim of early detection of attempts to spread misinformation and deceiving campaigns. By using this technique, the work on the real-time and high volume streams of messages can be optimised.

CONCLUSION

The spread of memes is often considered as social contagion or 'infection' as the spread of information or behaviour on social networks where an individual serves as the stimulus for the imitative actions of another (Lindzey and Aronson, 1985; Goffman and Newill, 1964; Daley and Kendall, 1964). The community-based feature (Becker *et al*., 2011) adopted in this study, proves its ability to be the best among three classes of viral memes. However, a problem in predicting a viral memes exists, for example in the clustering and classification task. Ferrara *et al*. (2013) demonstrate a clustering framework that adopts a novel pre-clustering procedure-the protomeme detection-that aims at identifying atomic tokens of information inside tweets. Due to its proficiency, this strategy could be particularly appropriate to use in streaming scenarios, for example, in the viral memes context.

Ultimately, this study provides areas of possible integration of the viral memes with the meme-clustering framework to differentiate engineered types of social media communication from viral communication or even spontaneous ones. An integration between viral memes and a platform to classify types of communication memes is believed feasible-the integration of viral memes which could predict the future popularity of a meme with three intuitive classes of features as well as formalising the framework for clustering memes from social media. By using these approaches, researchers could also identify topics of discussion, mentions of other users that are used to address messages to their attention as well as identify the contributors of conversations.

ACKNOWLEDGEMENTS

This research is partially supported by Research Management Centre, Universiti Teknologi MARA and Ministry of Higher Education, Malaysia under the Research Entity Initiative Grant Scheme (Project Code: 600-RMI/DANA 5/3/REI (6/2013)) and the Exploratory Research Grant Scheme (Project Code: 600-RMI/ERGS 5/3 (32/2012)).

REFERENCES

Adamic, L. and N. Glance, 2005. The political blogosphere and the 2004 U.S. election. Proceedings of the 3rd International Workshop on Link Discovery, August 21-25, 2005, ACM, Chicago, Illinois, ISBN:1-59593-215-1, pp: 36-43.

Adar, E. and L.A. Adamic, 2005. Tracking information epidemics in blogspace. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, September 19-22, 2005, IEEE, Washington, DC, USA., ISBN:0-7695-2415-X, pp: 207-214.

Aggarwal, C.C. and K. Subbian, 2012. Event detection in social streams. Proceedings of the SIAM International Conference on Data Mining, April 26-28, 2012, SIAM, Anaheim, California, ISBN:9781622760947, pp: 1-150.

Bakshy, E., B. Karrer and L.A. Adamic, 2009. Social influence and the diffusion of user-created content. Proceedings of the 10th ACM Conference on Electronic Commerce, July 06-10, 2009, ACM, Stanford, California, ISBN:978-1-60558-458-4, pp: 325-334.

Bauckhage, C., 2011. Insight into internet memes. Proceeding of the 5th International AAAI Conference on Weblogs and Social Media. July 17-21, 2011, AAAI, Barcelona, Catalonia, pp: 1-9.

Baumgartner, J.C., 2008. Polls and elections: Editorial cartoons 2.0: The effects of digital political satire on Presidential candidate evaluations. Presidential Stud. Q., 38: 735-758.

Becker, H., M. Naaman and L. Gravano, 2011. Beyond trending topics: Real-World event identification on Twitter. Proceedings of the AAAI International Conference on Weblogs and Social Media (ICWSM), July 05, 2011, AAAI, San Francisco, California, pp: 438-441.

Berger, J. and K.L. Milkman, 2012. What makes online content viral?. J. Marketing Res., 49: 192-205.

Cataldi, M., L.D. Caro and C. Schifanella, 2010. Emerging topic detection on Twitter based on temporal and social terms evaluation. Proceedings of the 10th ACM International Workshop on Multimedia Data Mining. July 25-25, 2010, ACM, New York, USA., ISBN:978-1-4503-0220-3, pp:4-10.

Centola, D. and M. Macy, 2007. Complex contagions and the weakness of long ties 1. Am. J. Sociology, 113: 702-734.

Centola, D., 2010. The spread of behavior in an online social network experiment. Science, 329: 1194-1197.

Chew, C. and G. Eysenbach, 2010. Pandemics in the age of Twitter: Content analysis of Tweets during the 2009 H1N1 outbreak. PloS One, 5: e14118-e14118.

Colbaugh, R. and K. Glass, 2012. Early warning analysis for social diffusion events. Secur. Inf., 1: 1-26.

Daley, D.J. and D.G. Kendall, 1964. Epidemics and rumours. Nat., 204: 1118-1119.

Ferrara, E., M.J. Asbagh, O. Varol, V. Qazvinian and F. Menczer *et al*., 2013. Clustering memes in social media. Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), August 25-28, 2013, IEEE, Bloomington, Indiana, ISBN:978-1-4503-2240-9, pp: 548-555.

Gaffney, S. and P. Smyth, 1999. Trajectory clustering with mixtures of regression models. Proc. 5th ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining, 10: 63-72.

Gamon, M., S. Basu, D. Belenko, D. Fisher and M. Hurst *et al*., 2008. BLEWS: Using blogs to provide context for news articles. Proceedings of the AAAI International Conference on Weblogs and Social Media ICWSM, March 30-April 2, 2008, AAAI Press, Menlo Park, California, pp: 60-67.

Godbole, N., M. Srinivasaiah and S. Skiena, 2007. Large-scale sentiment analysis for news and blogs. Proceedings of the International Conference on Weblogs and Social Media ICWSM, March 26-28, 2007, WSM, Boulder, Colorado, pp: 219-222.

Goffman, W. and V.A. Newill, 1964. Generalization of epidemic theory. Nat., 204: 225-228.

Granovetter, M., 1978. Threshold models of collective behavior. Am. J. Sociology, 83: 1420-1443.

Jamali, S. and H. Rangwala, 2009. Digging digg: Comment mining, popularity prediction and social network analysis. Proceedings of the IEEE International Conference on Web Information Systems and Mining WISM, November 7-8, 2009, IEEE, Fairfax, Virginia, ISBN:978-0-7695-3817-4, pp: 32-38.

Kaltenbrunner, A., V. Gomez and V. Lopez, 2007. Description and prediction of slashdot activity. Proceedings of the Web IEEE Conference on LA-WEB Latin American 2007, October 31-November2, 2007, IEEE, Barcelona, Spain, ISBN:0-7695-3008-7, pp: 57-66.

Kitsak, M., L.K. Gallos, S. Havlin, F. Liljeros and L. Muchnik *et al*., 2010. Identification of influential spreaders in complex networks. Nat. Phys., 6: 888-893.

Kubo, M., K. Naruse, H. Sato and T. Matubara, 2007. The possibility of an epidemic meme analogy for web community population analysis. Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning, December 16-19, 2007, Springer, Berlin, Germany, ISBN:978-3-540-77225-5, pp: 1073-1080.

Lazer, D., A.S. Pentland, L. Adamic, S. Aral and A.L. Barabasi *et al*., 2009. Life in the network: The coming age of computational social science. Sci., 323: 721-723.

Lee, J.G., J. Han and K.Y. Hwang, 2007. Trajectory clustering: A partition and group framework. Proceedings of the ACM SIGMOD International Conference on Management of Data, June 11-14, 2007, Beijing, China, pp: 593-604.

Lehmann, J., B. Goncalves, J.J. Ramasco and C. Cattuto, 2012. Dynamical classes of collective attention in Twitter. Proceedings of the 21st ACM International Conference on World Wide Web, April 16-20, 2012, ACM, Lyon, France, ISBN:978-1-4503-1229-5, pp: 251-260.

Lenser, S. and M. Veloso, 2005. Non-parametric time series classification. Proceedings of the IEEE International Conference on Robotics and Automation ICRA, April 18-22, 2005, IEEE, Pittsburgh, Pennsylvania, ISBN:0-7803-8914-X, pp: 3918-3923.

Lerman, K. and T. Hogg, 2010. Using a model of social dynamics to predict popularity of news. Proceedings of the 19th ACM International Conference on World Wide Web, April 26-30, 2010, ACM, Raleigh, North Carolina, ISBN: 978-1-60558-799-8, pp: 621-630.

Leskovec, J., L. Backstrom and J. Kleinberg, 2009. Meme-tracking and the dynamics of the news cycle. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, June 28-July 01, 2009, ACM, Paris, France, ISBN:978-1-60558-495-9, pp: 497-506.

Lin, C.X., B. Zhao, Q. Mei and J. Han, 2010. PET: A statistical model for popular events tracking in social communities. Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 25-28, 2010, ACM, Washington, DC, USA., ISBN:978-1-4503-0055-1, pp: 929-938.

Lindzey, G. and E. Aronson, 1985. Handbook of Social Psychology: Group Psychology and the Phenomena of Interaction. Lawrence Erlbaum Ass Publisher, Salt Lake City, Utah,.

McNames, J., 1998. A nearest trajectory strategy for time series prediction. Proceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling, July 8-10, 1998, KU Leuven Belgium Publisher, Leuven, Belgium, pp: 112-128.

McPherson, M., L.L. Smith and J.M. Cook, 2001. Birds of a feather: Homophily in social networks. v 27: 415-444.

Metaxas, P. and E. Mustafaraj, 2010. From obscurity to prominence in minutes: Political speech and real-time search. Proceedings of the Conference on Web Science: Extending the Frontiers of Society On-Line, April 26-27, 2010, Wellesley College, Raleigh, North Carolina, pp:1-7.

Pinto, H., J.M. Almeida and M.A. Goncalves, 2013. Using early view patterns to predict the popularity of You Tube videos. Proceedings of the 6th ACM International Conference on Web Search and Data Mining, February 04-08, 2013, ACM, New York, USA., ISBN:978-1-4503-1869-3, pp:365-374.

Ratkiewicz, J., M. Conover, M.R. Meiss, B. Goncalves and A. Flammini *et al*., 2011. Detecting and tracking political abuse in social media. Proceedings of the 5th International AAAI Conference on Weblogs and Social Media ICWSM, July 05, 2011, AAAI, San Francisco, California, pp: 297-304.

Romero, D.M., B. Meeder and J. Kleinberg, 2011. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter. Proceedings of the 20th ACM International Conference on World Wide Web, March 28-April 01, 2011, ACM, Hyderabad, India, ISBN:978-1-4503-0632-4, pp: 695-704.

Simmons, M.P., L.A. Adamic and E. Adar, 2011. Memes online: Extracted, subtracted, injected and recollected. Proceeding of the 5th International AAAI Conference on Weblogs and Social Media ICWSM, July 17-21, 2011, AAAI, Menlo Park, California, pp:1-8.

Szabo, G. and B.A. Huberman, 2010. Predicting the popularity of online content. Commun. ACM., 53: 80-88.

Vespignani, A., 2009. Predicting the behavior of techno-social systems. Sci., 325: 425-428.

Weng, L., F. Menczer and Y.Y. Ahn, 2013. Virality prediction and community structure in social networks. Sci. Rep., 3: 1-18.

Weng, L., F. Menczer and Y.Y. Ahn, 2014. Predicting successful memes using network and community structure. Proceeding of the 8th International AAAI Conference on Weblogs and Social Media, May 23-26, 2010, AAAI, Washington, DC, USA., ISBN:978-1-57735-659-2, pp:1-10.

Wu, F. and B.A. Huberman, 2007. Novelty and collective attention. Proc. National Acad. Sci., 104: 17599-17601.

Yang, J. and J. Leskovec, 2011. Patterns of temporal variation in online media. Proceedings of the 4th ACM International Conference on Web Search and Data Mining, February 09-12, 2011, ACM, Hong Kong, China, ISBN:978-1-4503-0493-1, pp: 177-186.

Zafarani R., M.A.Abassi and L. Huan, 2014. Social Media Mining. Cambridge University Press, Cambridge, England, UK., ISBN:978-1-107-01885-3, Pages: 105.