

# CLASSIFICATION OF PERSONALIZED NEWS TOPICS IN SOCIAL MEDIA USING SOCIRANK

<sup>1</sup>V. Sathwika, <sup>2</sup>N. Venkateswaran, <sup>3</sup>Dr. R. Jegadeesan <sup>4</sup>B. Sneha, <sup>5</sup>Dr. S. Prabaharan, <sup>6</sup>P. Swathi, <sup>7</sup>K. Ravali

<sup>1,4,6,7</sup>Final year Student Computer science and Engineering, <sup>2,3,5</sup>Associate Professor-CSE  
<sup>1,2,3,4,5,6</sup>Jyothishmathi Institute of Technology and Science, Karimnagar, India

**Abstract:** Social media has experienced growth in today's life. The interactions have been grown between traditional media and social media. As, social media being an open source, users exchange unrelated content to achieve importance, that is in the form of likes. For instance, Twitter alone generates 500 million tweets per day, information overload exist, it is important to categorize both media and filter the noise. It is important to find relevant topics in social media. To know, how the influences are made inversely when compared to news media we proposed an unsupervised framework socirank i.e., ranking important topics in social media. The factors taken into consideration to achieve prevalence are, the temporal factors: Media focus (MF) which is important, the temporal factor in social media is user attention (UA) and strength of the group discussing about a topic is user interaction (UI). Based on how the nature of the event considered, volume of activity over time and procedure will affect the quality of filtered topics. This method is proposed to detect the spread of viral topics and also improves the passivity of automatically identified news topics.

**Index Terms:** social Media, Ranking, News Topics

## I. INTRODUCTION

The Mining is the process of extracting valuable information from online sources has become a research area in recent years. Mass media sources provide public of daily events, specifically the news media. The news media sources are able to be trusted because they are published by professional writers. In social media regular users are posting unverified data and expressing their interests.

One of the most popular network site in social media like Twitter, is used by many people, providing large amount of user-created data. This unverified data is useless. The social media used for topic identification. We must find and filter the unverified information and capture only data. The news media, considered as valuable. The news media presents verified data, while social media posting the interests of audience. Unfortunately, after removal of unverified content, there is still information overload in the remaining data.

The news must be ranked based on Media Focus (MF), which is important. Secondly, user interests considered as User Attention (UA). Similarly, the number of users discussing a topic and interaction between the users indicates User Interaction (UI). By using these factors, it is able to rank the news topics. The main use of the paper is to improve the quality of news. The topic modelling and other topic detection techniques do not rank topics according to their popularity. We propose an unsupervised system to identify topics that are common in both social media and news media, and then ranks them by using degrees of MF, UA and UI.

To achieve its goal, this paper uses keywords and media sources to identify and build a graph. Whose nodes represent keywords and edges in social media. The graph is clustered to identify topics. Then, the factors that are calculated based on their importance: MF, UA and UI. Finally, the topics are ranked by measure that using of three factors. The excess of this paper as: Section II reviews previous research topics. Section III presents overall design and framework. Finally, we provide conclusion in Section IV.

## II. LITERATURE SURVEY

Aiello et al. [1] Online social and news media generate events of all categories. Information filtering is needed to find important topics and events by comparing six detection methods on three datasets which differ on time. Based on nature of event, volume and procedure, the preprocessing stage is completed. One of the detection method, based on n-grams co-occurrence and topic ranking, topic detection is done. Jie Tang et al. [2] large social networks like Twitter, users are influenced by others. Due to which social Influences arise among them. To address this issue, Topical Affinity Propagation model is used to model influences on large networks. By using this model, social influence of neighbouring users on a particular user is determined. Andriy Shepitsen [3] some applications allow users to create personalized tags like creating uncontrolled words can result in tag repetition. Clustering method is used to remove these problems by identifying important topics. Personalization algorithm is used in which cluster selection is an important step to suggest that topic selection is an important method. Fabian Abel et al [4] Filtering, Searching and analysing information about real world events on web streams is done using a method that is Twitcident. It automatically connects to services and start filtering information from web streams which allow users to retrieve prevalent information. Ding Zhou et al [5] to discover the relationship between documents shared in social networks, we used a method. Generally probabilistic method is used to associate actors. Markov transition matrix method is used to determine social interactions, Topics and also how actors, authors impact these topics and introduce new ways for other impact. J. Ratkiewicz [6] to identify the political Abuse in the social media we proposed a detecting and tracking algorithm with the help of machine learning frame work that combines topological, content-based and crowd source features. Some political individuals and community which are responsible for wide spread of political misinformation is detected at its early stages itself. Daniel M. Romero [7] The information from the social media which

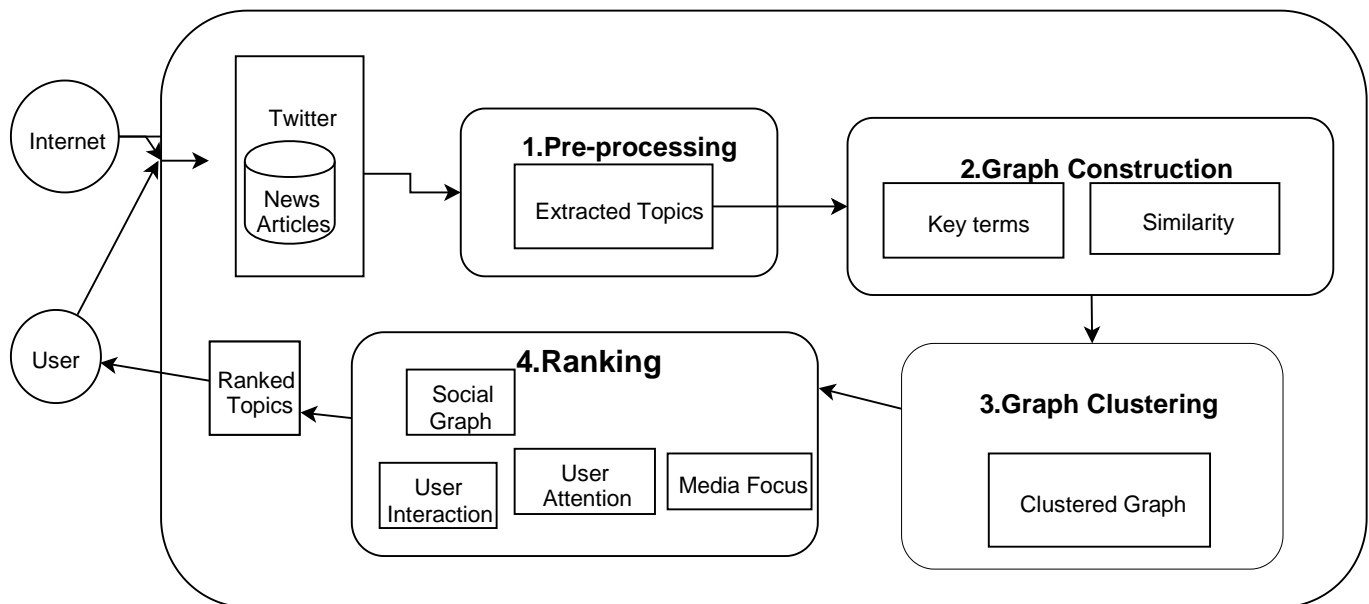
as high amount of information forces the users of twitters or the social media to pay attention for it. To overcome this the users must not influence and passivity for the popular information we propose an algorithm to determine how much amount of users influences and passivity to attain we performed with a 2.5 million user dataset shows that our proposed sequence measure is good predictor. Jianshu WENG et al [8]To know how users use the social media like twitter, “twitter Rank” was proposed based on the concept of following and followers using page Rank algorithm we calculated the users influence by twitter. [9]Based on factors like content source topic interest models and social voting, interested content for users is evaluated using best performing algorithm. CerenBudak et al [10]Trend detection methods used in past are not supposed in network topology. So using two models i.e., co-ordinated model to get best trend in news media and uncoordinated less trends in media. By this we found structural trends i.e., important trends. G Nandi and A Das [11] Social Media is a prominent research area. In this know the social media and news media topics. In this social media factors are Media Focus (MF), User Attention (UA) and User Interaction (UI) is proposed to collected twitter events. These factors ranked under unsupervised framework and article filtered and ranked social media events occurring twitter dataset. YogeshLonkar [12] in Social Media like twitter provide enormous amount of user generated data, this paper filter the noise and capture the content using social media factors. For implementation of this paper we use CNN algorithm.KlimisNtalianis [13]In this paper social media ranking scheme proposed and unsupervised architecture takes into user-content interactions. Social media receives likes, comments and shares from friends and other users. Ranks the posted items based social computing method.R.Balamurugan [14]In search engines, different users may seek different information by issuing similar query re-ranks the results. First the media focus, User Interaction and Use Attention. We propose an unsupervised framework approach which identifies any type of search query and ranks the frequent using topics and give their numbers. Sandhiya. R [15]These are many concepts of using algorithmic and data mining perspective of Online Social Networks(OSNs) are used in research are. It is introduced enormous several factors studied for OSN important by researchers. OSNs help the researchers to solve those challenges.PavanKapanipathi [16]Personalized filtering system involves understanding user interests. Information overload is challenge to the users in social media. Users have to analyze the data for topic identification, scalability according to their interests. Personalized filtering for user’s comes common, interesting information in social media. These are used to build a personalized information filtering system for social media to improve context, topics on social media. SrishankJaiswal et al. [17]the social and news media services are used to finding similar news and finding the trends. In this we use a hierarchical Bayesian model to capture different topics as well as topic influences on both media’s, based on techniques and knowledge. By using proposed model news and social media data, we will show the improvement in the real world. Ravneetkaur et al. [18]the survey discusses different anomalies and categorization of data mining approaches used to detect anomalies. Anomalous activities in social network represent unusual and illegal activities exhibiting different behaviours than others present in same structure. This paper presents number of data mining approaches for detecting anomalies. The anomaly detection techniques are categorized as behaviour, structure and spectral based. Eswari K E et al[19]to reach information ranking, we use three social media factors. The population over a period of time specific topic in news media is a factor considered as Media Focus (MF),second factor is the User Attention (UA), and interaction between social media users discusses about the topic indicates User Interaction (UI).This proposes an edge-centric clustering method to extract social dimensions. Matthew Bures et al. [20]Active users of social media are applied to information overload on twitter are not allowing users to place a priority. On content posted by friends, human effect is needed. We used a method that seeks help users for find more common content. Our model B UTTER WORTH. Naturally provides rankers of users producing the same content. The rankers providing human readable content and allow users to find and ranking the relevant topics.

### III. FRAMEWORK

The goal of our project is to identify and rank the most common topics in both social media and news media. Fig1: shows the framework for classifying and rankingtopics.

1. *Pre-Processing*: System collects all news articles and tweets from database. We extract the top keywords from each news article, which does not contain duplicate terms. Tweets are relevant and unique, any tweet that is not in English, and also tweets that are less than 3 characters length are rejected. Terms tagged as hash tag are selected.
2. *Graph Construction*: N is keyword in news and T is keyword in tweets. To extract the topics that are common in both social and news media, new set R is taken.  

$$R=N \wedge T$$
Graph is constructed for common topics. Vertices represent the identified terms, edges represent the term similarity.
3. *Graph Clustering*: In this stage, well-defined sub graphs are identified and separated from constructed graph.
4. *Ranking*: News sub graphs are selected and ranked from two media sources. Related items from news media represent MF of the topic. Related items from social media represent User Attention.



**Fig1: Framework for classifying and ranking topics**

#### IV. CONCLUSION

In this paper, we suggest an unsupervised method. We comparing social media and news media identify the frequent newstopics using three factors they are MF,UA and UI. These three factors give ranking the frequent news topics in social media. The frequent particular topics are changes in news media is taken as Media Focus(MF), the frequent particular topics are changes in social media, especially Twitter,indicated user interest is taken as User Attention(UA) and last one is who are in social media to interact with that users and indicates the topic and discussing with group and develop the understanding way it's taken as User Interaction(UI).Filtered and ranked news topics are used to professional news provider and individuals. In this one of the most important pointis increasing the quality and variety of news guidance system as well as identify the hidden popular topics. Socirank can be external such as sports,science,technology and other things. The evaluation provides evidence that our method is selecting frequent news topics and ranking them based on the previously mentioned three factors. Our result is ready a clear distinction between ranking topics MF only and then include UA and UI.

Future work is to perform experiments SociRank on different areas and datasets and we plan to include a other forms of UA. Additional experiments also be performed in different stages of the approaches. Lastly, we intend to develop a professional version of SociRank, where topics are presented differently to each other.

#### REFERENCES

- [1] Luca Maria Aiello, Georgios Petkos, David Corney, Symeon Papadopoulos, Ryan Skraba, AyseGoker, YiannisKompatsiaris, and Alejandro Jaimes, "Sensing trending topics in Twitter".
- [2] Jegadeesan,R.,Sankar Ram,N. "Energy Consumption Power Aware Data Delivery in Wireless Network", Circuits and Systems, Scientific Research Publisher,2016 (Annexure-I updated Journal 2016)
- [3] AndriyShepitsen, Jonathan Gemmell, BamshadMobasher, and Robin Burke, "Personalized Recommendation in Social Tagging Systems Using Hierarchical Clustering", 2008, Lausanne, Switzerland.
- [4] Jegadeesan,R.,Sankar Ram, R.Janakiraman *September-October 2013* "A Recent Approach to Organise Structured Data in Mobile Environment" R.Jegadeesan et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (6) ,Page No. 848-852 ISSN: 0975-9646 Impact Factor:2.93
- [5] Ding Zhou, HongyuanZha, C. Lee Giles, "Topic Evolution and Social Interactions",2006.
- [6] Jegadeesan,R., Sankar Ram, M.S.Tharani (*September-October, 2013*) "Enhancing File Security by Integrating Steganography Technique in Linux Kernel" Global journal of Engineering,Design & Technology *G.J. E.D.T., Vol. 2(5): Page No:9-14* ISSN: 2319 – 7293
- [7] Daniel M Romero, WojciechGaluba, SitaramAsur, "Influence and Passivity in Social Media", 2007.
- [8] Jegadeesan,R., Sankar Ram "Defending Wireless Sensor Network using Randomized Routing "International Journal of Advanced Research in Computer Science and Software Engineering Volume 5, Issue 9, September 2015 ISSN: 2277 128X Page | 934-938
- [9] "Experiments on approved Content from Information Stream", 2009.
- [10] CerenBudak, Divyakant Agrawal, Amr El Abbadi "Structural Trend Analysis for Online Social Networks", Proceedings of the VLDB Endowment, Vol. 4, No.10, 2011.
- [11] G Nandi, A Das, "A Survey on Using Data Mining Techniques for Online SocialNetwork Analysis", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 6, No 2, November 2013.
- [12] JianshuWENG,Ee Peng LIM, Jing JIANG, Qi HE, "Twitterrank: Finding topic-sensitive influential Twitterers", Research

Collection School Of Information Systems, 2010

[13] KlimisNtalianis, Abdel-Badeeh, M. Salem b, andIbrahiem El Emary, “Social Media Content Ranking based UserInfluence”,International Conference on Communication Management and Information Technology, 2015.

[14] Jegadeesan,R.,Sankar Ram M.Naveen Kumar JAN 2013 “Less Cost Any Routing With Energy Cost Optimization” International Journal of Advanced Research in Computer Networking,Wireless and Mobile Communications.Volume-No.1: Page no: Issue-No.1 Impact Factor = 1.5

[15] Jie Tang, Chi Wang and Zi Yang, “Social Influence in Large-scale Networks”, 2009.

[16] Jegadeesan,R.,Sankar Ram October -2013 “ENROUTING TECHNICS USING DYNAMIC WIRELESS NETWORKS” International Journal of Asia Pacific Journal of Research Ph.D Research Scholar <sup>1</sup>, Supervisor<sup>2</sup>, VOL -3 Page No: Print-ISSN-2320-5504 impact factor 0.433

[17] Fabian Abel, Claudia,Ke Tao, “Fighting Fire with Information from Social Web Streams”, WWW Companion, 2012.

[18] Ramesh,R., Vinoth Kumar,R., and Jegadeesan,R., January 2014 “N<sup>TH</sup> THIRD PARTY AUDITING FOR DATA INTEGRITY IN CLOUD” *Asia Pacific Journal of Research Vol: I Issue XIII, ISSN: 2320-5504, E-ISSN-2347-4793* Vol: I Issue XIII, Page No: Impact Factor:0.433

[19]Vijayalakshmi, Balika J Chelliah and Jegadeesan,R., February-2014 “SUODY-Preserving Privacy in Sharing Data with Multi-Vendor for Dynamic Groups“ Global journal of Engineering,Design & Technology. G.J. E.D.T.,Vol.3(1):43-47 (January-February, 2014) ISSN: 2319 -7293

[20] Jegadeesan,R.,SankarRam,T.Karpagam March-2014 “Defending wireless network using Randomized Routing process” International Journal of Emerging Research in management and Technology

[21] J. Ratkiewicz, M. Meiss, A. Flammini, M. D. Conover “Detecting and Tracking Political Abuse in Social Media”, Association for the Advancement of Artificial Intelligence, 2011.

[22] Jegadeesan,R., Sankar Ram,N. “Energy-Efficient Wireless Network Communication with Priority Packet Based QoS Scheduling”, Asian Journal of Information Technology(AJIT) 15(8): 1396-1404,2016 ISSN: 1682-3915,Medwell Journal,2016 (Annexure-I updated Journal 2016)

[23] YogeshLonkar, Dattatray Savant, Prashant Nikam, SurajHalkude, KameshPatil, “A Review on Hybrid Approach for Searching and Ranking Large Scale Web Data using Social Media Factors”, International Journal of Innovative Research in Engineering and Technology,Vol 6, Issue 11, 2017.

[24] Jegadeesan,R., Sankar Ram , and J.Abirmi “Implementing Online Driving License Renewal by Integration of Web Orchestration and Web Choreography“ International journal of Advanced Research trends in Engineering and Technology (IJARTET) ISSN:2394-3785 (Volume-5, Issue-1, January 2018