



**International Journal of Allied Practice, Research and Review**  
Website: [www.ijaprr.com](http://www.ijaprr.com) (ISSN 2350-1294)

# **Social Network Topic Diffusion for Influential Node Cluster Identification**

**Reena Pagare<sup>1</sup> and Dr. Akhil Khare<sup>2</sup>**

<sup>1</sup>Research Scholar, Faculty of Engineering, Udaipur, Rajasthan, India

<sup>2</sup> MVSR Engineering College, Hyderabad India,

**Abstract** - The social networking and e-commerce is the backbone of information diffusion. This paper contributes an analysis of information influence over the internet. Paper also focuses on proposed Modified Topic Diffusion Model Algorithm. The quantitative analysis is usually carried out for information diffusion study; hence for the proposed study, we analyzed twitter and Flipkart web portals visitor responses. Data clusters are optimized to know the most influential information spreader. Based on such information the analysis is done for identification of nearest influential network nodes. The analysis shows the positive results for third-party web clients for the unsupervised network.

**Keywords**—Social network, influential nodes, information diffusion.

## **I. INTRODUCTION**

Information Diffusion is a phenomenon of spreading information while each person (as an element of society) comes with a limited awareness, we illustrate that restricted awareness can certainly increase the competition, resulting in heterogeneous mechanics of occurrence acceptance [1]. Higher topical variety of first adopters tag words can be proven to indicate substantial upcoming virility of information; although minimal topical variety assists anyone builds up cultural influence [2].

Topic closure carries a powerful influence; however traffic-based techniques tend to be one more essential element in interpretation of community development. We also analyze exactly how individual's strong and weak ties affect the information diffusion. This challenge is designed to find about information diffusion in online networks via 2 facets: Individuals who discuss information and the communication involving community structure and diffusion. The next element investigates components of information, specifically to the topic area. Numerous intricate design of individual community could possibly be much better by continuing exploration of community framework. It is observed that network areas capture data streams in general, however, not the viral phenomenon. This paper focus on the information spread, the most influential node and neighbors of the influential nodes analysis.

## II. LITERATURE REVIEW

Micro-blogging has become common and is recurrent kind of social networking in current years since it is easy to use. Nevertheless, discovering hot topics through these kinds of small blogs creates completely new problems [3]. Processing these types of blogs effectively can also be an issue because people interact using social network more often and thus generate huge data. Consequently, conventional topic investigation approaches normally crash to offer the preferred accuracy and reliability. Contrary to standard documents, these kind of communications are often fluctuating, fewer topic focused and far smaller. Numerous scientific studies are carried out to remedy the problems stated earlier. Many others produce to leverage the potential of data propagate in intricate networks [4, 5].

Applying these kinds of characteristics has revealed optimistic enhancement within the hot topic discovery. Hot topic removal can be an essential process for examining social networking [6]. Existing studies indicate that the additional information about topics can definitely improve the efficiency [7, 8, 9].

Additionally, much fewer communications on micro blogs tend to be important for scorching topic discovery due to the fact huge communications range between coverage of considerable social activities and everyday life, for instance, climate, meals, feelings, etc. Additional information, like Wikipedia or maybe media content, is required to develop and improve the framework in the small communications [10]. Like the research of data progression within, another approach computes the temporary submission of conditions, subsequently identify the hot topics. There has been research work going on in the domain of social network analysis mainly in the field of influence maximization.[11]. So far, there are primarily a couple of strategies of time-stamped information evaluation.

With even more extensions, the task is to run a new time-series [12] kind of susceptible-infectious-information model to keep track of microblog rising breakouts, where the SI design continues to be commonly employed in distribution in intricate networks. Nevertheless, this technique is associated with small utilization in discovering precise hot topics. Along with a growing number of groups by using this design, it is progressively considerable to draw out hot topics by using this small amount of information to ensure that end users can understand intriguing hot topics easily and rapidly [13]. To get acceptable outcomes, numerous traditional models are widely-used, like single-pass clustering [14], LDA [15] and full-space clustering or subspace clustering algorithm. It has been pointed out that the characteristics of micro blogs, like published time, document time and also repost count number and review count number, incorporate huge amount of beneficial data

## III. MODIFIED TOPIC DIFFUSION MODEL

As social media has become more prevalent, its influence on business and society has become significant. Due to easy access and interaction between large numbers of users, information diffuses in an epidemic-style on the web.

According to topic modeling, a document is defined as a mixture of various topics. A topic is defined as a set of words that frequently occurred together. The probabilistic model assumes that documents are mixtures of topics, where a topic is a probability distribution over words. The probabilistic topic model works to find out the best set of words and to explain the shown-up words in

documents. The probabilistic topic model specifies a simple probabilistic procedure by which documents can be generated. To make a new document, one chooses a distribution over topics. Then, for each word in that document, one chooses a topic at random according to this distribution and draws a word from that topic. This process is inverted using statistical methods inferring the set of topics that were responsible for generating a collection of documents. The adopted model clusters messages according to probable keywords.

In this paper, we model ‘Modified Topic Diffusion Model Algorithm’ in web forums using the live data. The model was evaluated on a large longitudinal dataset from the web forum of a major retail company and from a general political discussion forum. The fitting results showed that the proposed model is a plausible model to describe the diffusion process of a topic. This research shows that epidemic models can expand their application areas to the topic discussion on the web, particularly social media such as web forums.

First, we set the number of topics, then algorithm assigns every word to a temporary topic according to a probability distribution. Since each word is assigned in a somewhat random manner, a word shown up more than once in different documents can be assigned to different topics. Then the algorithm takes an iterative topic assignment as follows. A word is assigned to the topic where this word is the most prevalent and a document is assigned to a topic where the words in the document are mostly assigned. After the iteration converges, topic modeling is done. For topic clustering, we used unstructured data gathered from twitter and Flipkart. When the number of topics is set to too low, words in a topic are general and a topic includes the words that are not close semantically. When the number of topics is set to too high, words in a topic become too specific, and words distribute to various topics. This causes semantic overlapping of topics. Thus, we varied the number of topics and examined whether topic modeling generates proper semantic clusters. We incrementally set the number of topics by ten and observed the semantic generation. The algorithm works in an iterative way as follow to analyze information spreader and influential nodes in the unsupervised network.

### Steps for Modified Topic Diffusion Model Algorithm

- Step1 a) Initial number of network nodes,  $N_i = 2$   
 (b) Set step size as per standard deviation  $\sigma_k$  and area  $a \cap (y_k)$   
 (c) Set  $\sigma_s$  as the referenced standard deviation of topic selection for  $\omega \cap (x_i y_i)$   
 (d) Record topic and diffusion element by matching node coordinates of similar topic  
 (e) Store values of first topic as a reference and store active nodes of network
- Step-2. Iterative approximation to identify diffusion state:
- (a) Record nodes covering topic coordinates and rate of diffusion  
 (b) Deduct redundant topic elements  
 (c) Recalculate network nodes involved in topic diffusion

If  $Wc \otimes F_a \cong$  total number of topics

go to step-1

else

Consider next topic for information diffusion

3. Identify network nodes and locate information diffusion root

We selected the social blog- twitter and Flipkart , which provide a longitudinal dataset.

Consider, ' $P$ ' and ' $Q$ ' are the total numbers of nodes and links, respectively. ( $F$ ) And  $F_{max}$  denote the average and the maximum degree. The average shortest distance is considered ( $d_{short}$ ). Clustering coefficient and assertive coefficient are  $C_c$  and  $R_c$  respectively. The degree heterogeneity, calculated as  $H_{total} = F^2 / ((F_{max})^2)$ .

In this work, we just showed the structural soundness of the model over the topic diffusion in the web forum twitter and Flipkart. However, after testing structural soundness, we can perform forecasting. Additionally, we can estimate how many authors have the latent interest in each topic at the initial phase of the diffusion process. We can also estimate the expected duration and the intensity of diffusion process at an initial stage. Even though these estimation measurements become reliable when it closer to the peak, we can estimate them in an adaptive way. We can compare those values for topics that lead the peak. The next section provides two sets of observations about information diffusion based on the topic selection.

#### IV. DISCUSSION AND ANALYSIS

We mainly focus on the operation of the social mechanisms of link malformation on the epidemic persuasion but it is worth tone that temporal activity model can also move the dispense dynamics and persuasion of people.

A reckoning of social mechanisms for connection establishments has been proposed in sociology. The endowments of this process are a manageable applicability for any kinds of the social web since one can gain the probabilistic actions of agents by second-hand survey from a population. Thus, we proposed a likely generalship for a biased prospect of a whole population. We also find that people with the massive crowd to concatenate distinct communities are more likely to be an effectual spreader for the case when a cobweb is composed of powerfully connected modules.

Through the analysis of large-ascend develop reticulation; we recognize the result of the tiny link form on minor consequences in disseminate. Thus, we design a possible generalship for identifying the message spreaders by characteristics of community's behavior basic the maneuver of familiar networks. Specifically, crowd advice of relationships among individuals is certainly imperfect and straying, since it cannot but be generalship for a biased prospect of a whole population. But collecting this information is closely infeasible in genuine social systems. We also find that people with the massive crowd to concatenate distinct communities are more likely to be an effectual spreader for the case when a cobweb is composed of powerfully connected modules.

## V. CONCLUSION

We proposed an integrated and novel methodology to model opinion/idea diffusion in web forums.

The present modified topic diffusion model algorithmic approach is used to analyze topic selection and knowledge diffusion for information spreader analysis and subsequent node analysis. The model was evaluated on a large dataset from twitter and Flipkart customer review dataset. The analysis results revealed that the proposed model performed well in modeling topic diffusion for selected web forums.

| Study Network | $p$ | $q$ | $F$ | $F_{max}$ | $C_c$ | $(d_{short})$ | $R_c$ | $H_{total}$ |
|---------------|-----|-----|-----|-----------|-------|---------------|-------|-------------|
| Twitter       | 5   | 7   | 5.4 | 66        | 0.5   | 5.29          | -     | 0.09        |
|               | 9   | 3   | 9   |           | 61    |               | 0.92  | 26          |
|               | 1   | 4   |     |           |       |               | 3     |             |
| Flipkart      | 7   | 9   | 3.1 | 97        | 0.3   | 3.18          | -     | 0.00        |
|               | 5   | 3   | 9   |           | 28    |               | 0.17  | 10          |
|               | 3   | 4   |     |           |       |               | 65    |             |

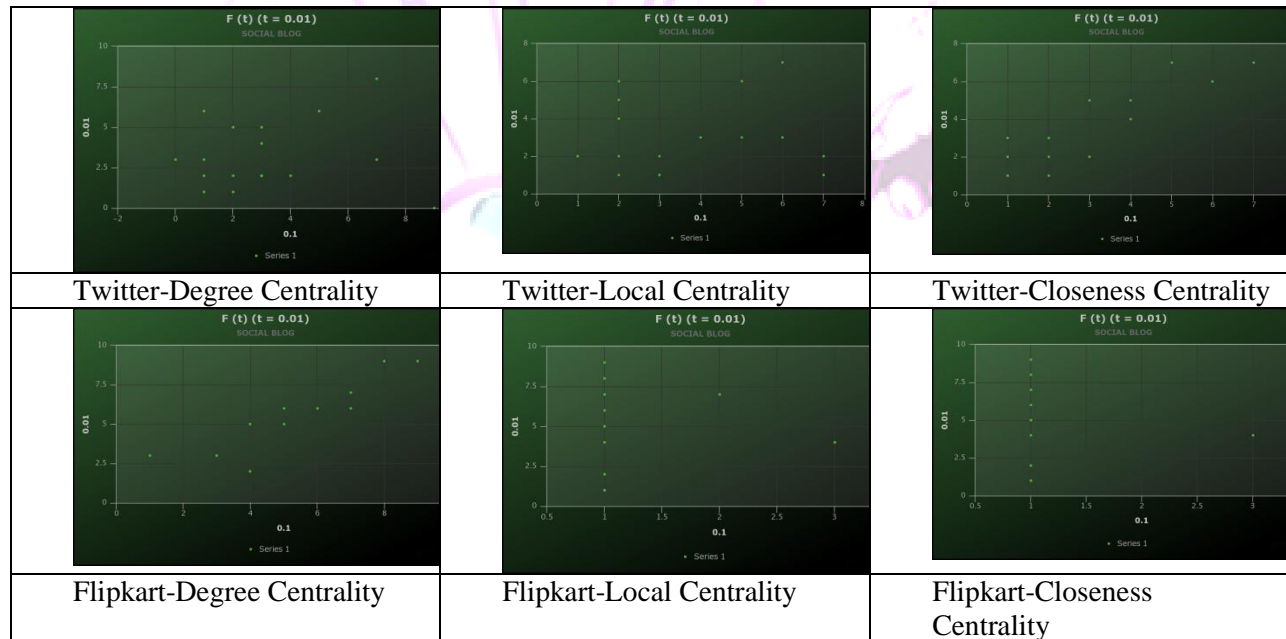


Fig 1: Analysis of Twitter and Flipkart web blogs.

## VI. REFERENCES

- [1] Huang, Chongfu, and Yundong Huang. "An information diffusion technique to assess integrated hazard risks." *Environmental research* 161 (2018): 104-113.

- [2] Liang, Yuhua, and Kerk F. Kee. "Developing and validating the ABC framework of information diffusion on social media." *New Media & Society* 20.1 (2018): 272-292.
- [3] Peng, Sancheng, et al. "Social influence modeling using information theory in mobile social networks." *Information Sciences* 379 (2017): 146-159.
- [4] Tu, Hong T., and Khu P. Nguyen. "Differential Information Diffusion Model in Social Network." *Asian Conference on Intelligent Information and Database Systems*. Springer, Cham, 2018.
- [5] Olanrewaju, Abdus-Samad Temitope, Rahayu Ahmad, and Massudi Mahmudin. "Influence Maximization Towards Target Users on Social Networks for Information Diffusion." *International Conference of Reliable Information and Communication Technology*. Springer, Cham, 2017.
- [6] Gao, Sheng, et al. "A novel embedding method for information diffusion prediction in social network big data." *IEEE Transactions on Industrial Informatics* 13.4 (2017): 2097-2105.
- [7] Wu, Yaning, et al. "A novel Bayes defect predictor based on information diffusion function." *Knowledge-Based Systems*(2017).
- [8] Si, Xia-Meng, et al. "A topic evolution model with sentiment and selective attention." *Physica A: Statistical Mechanics and its Applications* 471 (2017): 480-491.
- [9] Rui, Xiaobin, et al. "SPIR: The potential spreaders involved SIR model for information diffusion in social networks." *Physica A: Statistical Mechanics and its Applications* (2018).
- [10] Yang, Dong, et al. "True and fake information spreading over the Facebook." *Physica A: Statistical Mechanics and its Applications* (2018).
- [11] Yang, Dingda, et al. "Dynamic node immunization for restraint of harmful information diffusion in social networks." *Physica A: Statistical Mechanics and its Applications* 503 (2018): 640-649.
- [12] Gray, Caitlin, Lewis Mitchell, and Matthew Roughan. "Super-blockers and the Effect of Network Structure on Information Cascades." *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 2018.
- [13] Zhang, Yaming, et al. "Rumor and authoritative information propagation model considering super spreading in complex social networks." *Physica A: Statistical Mechanics and its Applications* (2018).
- [14] Yang, Shih-Feng, and Julia Taylor Rayz. "An event detection approach based on Twitter hashtags." *arXiv preprint arXiv:1804.11243* (2018).
- [15] Zheng, Liu, Zhang Caiming, and Chen Caixian. "MMDF-LDA: An improved Multi-Modal Latent Dirichlet Allocation model for social image annotation." *Expert Systems with Applications* 104 (2018): 168-184.