



ISSN(Online): 2320-9801  
ISSN (Print) : 2320-9798

## International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

# The Role of Community Forum in Online Countering of Radicalization to Violence

Shahbaz S<sup>1</sup>. Baig, Kishor.P.Wagh<sup>2</sup>

PG Student, Dept. of CSE, Government College of Engineering Amravati, India<sup>1</sup>

Assistant Professor, Dept. of IT, Government College of Engineering Amravati, India<sup>2</sup>

**ABSTRACT:** Online Web Forum is one of the prime attribute of the of web 2.0 technology that has empowered users across the globe to raise their opinion on a virtual planet. The internet has become an essential part of our lives. But it comes with the risk that it is inadvertently used by many users as a model to share their ideologies in social media which may be positive or negative. As vehement Ideology is propagated through online medium a research need is urgent in this area. In this paper, a framework is proposed which has a new approach to include user's activity in online web forum to analyze and distinguish these users to find the real culprit among all users. The method involves use of term frequency and inverse document frequency to calculate Cosine Similarity and making use of this score in calculation of user ranking by use of adjacency matrix. A tree can be created showing relation among different users with the help of the comment that are being posted by user in online web forum. The implementation of User rank and weighted in degree algorithm to the above grid will compute User dominance.

**KEYWORDS:** Security; Terrorist; Grid; Dominance; Metrics; Ranking

### I. INTRODUCTION

Internet has been used remarkably applied on wide range of user's activity including browsing, downloading etc. Very few efforts are taken to address the question of extremist challenge that is present on day to day basis. However this approach is not new. However it is the responsibility of government to keep its view very secure even if the view of individuals raises the concern for the rising extremism. Even political extremism are to the concern of the individual society because this type of extremism brings polarization in the society and thus is very harmful as it turn out to be the reason for violent activity. Here we are left with the question of how we can protect our society without putting the liberty of an individual's at the risk. Here we come across the main question of extending the influence of government which is confronted everyday with subversion and political violence. However there is certain reversible procedure to counterattack the radicalization by aiming denial to access the information specially the material that is being produced by the political extremists. When any authority aims at removing content from webpage like forum or any discussion they are confronted with the very alarming questions that involves whether they are suppressing the freedom of speech, a voice against any measures. As we can see the reversible does not include argument conversation feature which can be the prime focus so as to have this question tackled. Hence we could aim to neutralize the content which are inspired by online radicalization with the help of alternative messages or increasing the number of enticing forum which will lead to reduction in the share of extremist audience. Here our question arises between whether our measures are political or ethical. Now our positive measure steers the ongoing radicalization debate.

The measures include:

- Deterring creator of extremist resources and create a more aggressive circumstances for extremist post:** Prosecution of users that are involved in spreading radicalised material in internet, this could prevent them from publishing such articles.
- Allowing online web forum to create their club so as to give users the power to report:** Creation of Online club that could empower the users in specific online web forum to check for any such kind of radicalised material and thus allow them to report any such kind of action to their club organisation.

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

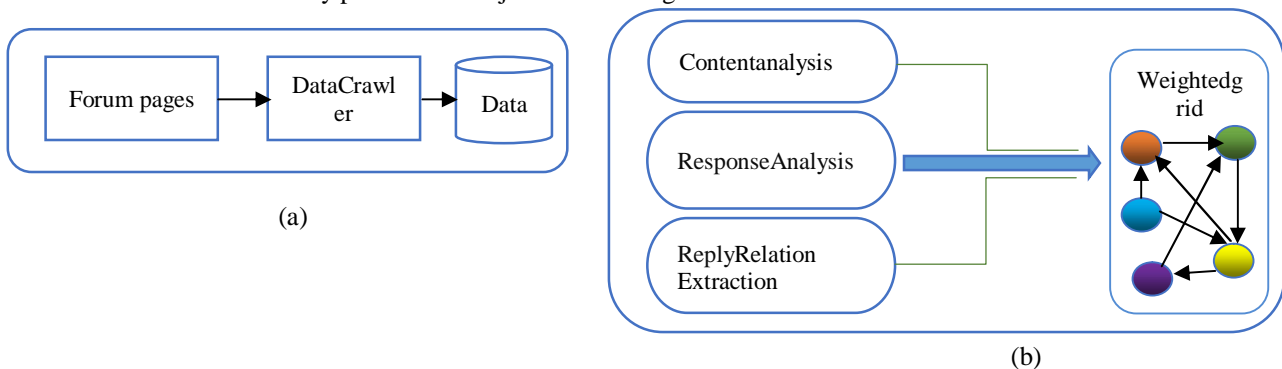
•**Sinking the call of extremist post:** Prime attention in the online community forum to juveniles so as to resist the extremist challenge that is being turned by the radicalised material, these lessons could help them to analyse the article manually.

## II. RELATED WORK

In the literature, there are a number of works related to dominant user identification in online web forums. Since ground-breaking idea, thoughts can proliferate by “word-of mouth” effect; online media is a perfect podium of understanding how the information is diffused in online community forum. The advertising that is being used by many online web forum takes recommendation from the data mining analyses and thus present the result to customers with the help of customer review analyses. Analyses on the statistical data collected from online forums helped Abbasi and Wang to find that the environment in the forum is responsible to give birth the kind of conversation that take place on the respective forum [1]. For a new product to be used by the customer it’s the duty of mining system to focus on the set of users that are affected by the use of product as such. This problem affecting was studied taking into account the data mining aspect from Domingo’s and Richardson [6]. Many aspects of estimation and likelihood were used by them to recognize the set of initial users that are affected by the advertisement. Kemp et al. overcame this problem by considering this matter in a discrete way and applying the greedy approximation technique to different models [2]. Experimental analysis of their result showed that their greedy approach showed better results than the classical heuristic based methods; however this procedure suffers severely from the cause of efficiency. To overcome this problem Chen et al. suggested new heuristic measures approximating near greedy method which turned to six time the order of magnitude than greedy measure [3]. The prime aim of this study is the finding of dominant users in the community whose review or comment as a function of input could yield influence on community. The study presented by Jeffery Kaplan gave wide-ranging analysis of the late-20th century political and social factors that instigate right wing radicals from all the sides of Atlantic closer together [7]. The needed inputs are not always readily available in the real scenario issue. KishorWagh and Satish Kolhe presented a similarity matching concept which made use of information content using the hierarchy of Word Net [8]. The study of various page rank algorithm helped [9] to conclude at User Rank Algorithm to position the most dominant user with the pattern of matching similarity and response. The data mining technique [10] is applied on the data that is being obtained by collecting all the related web pages and various online websites. How various group of terrorist organization interact with each other is graphed from east to west that is across the globe in a detailed study by [1].

## III. PROPOSED ALGORITHM

**A. Framework of Recognizing Dominant Users in Online Forum:** The procedure to get close to detection of the influential online web forum users is summarized into three steps which are (a) Availability of Data set (b) Tree relation of user’s creation and (c) User dominance computing. The data is treated with the algorithm that finds the number of threads or the post which has a word that is matched to a threat list. We get an output that has the frequency of occurrence of word in every post that is subjected to this algorithm.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

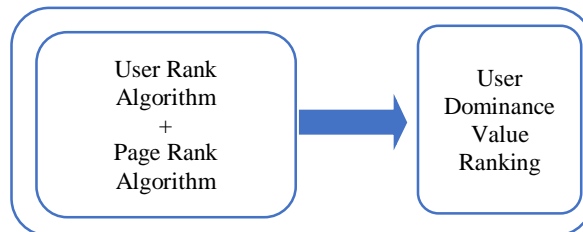


Fig. 1 Framework of identifying dominant users(a) Forum data gathering stage, (b) Community grid constructing stage, and (c) User dominance computing and ranking stage

## B. Extracting Community Grid from an Online Forum

Community Grid Topology: Community grid represent the way user interact with the other users. The aim here is creation of tree where each vertex represents a user .Whenever a user is destined to interact with the other users then this tree is generated with the initiator as a root node. However when the communication is exhaustive the nodes and the corresponding tree generated thus needs to be dealt with special details.

First for every letter of the alphabet the directory is created. This directory is then treated with preprocessing of the text. Here the text files are renamed with the help of keys.

Now using this keys we will correlate the corresponding files in the network graph. On the basis of creation of term document frequency vector, we will map every text files of corresponding directory to the file of its own directory. This directory is then subjected to get the most similar document from their graphs .Then this documents of respective directory are merged at last with all the high frequency documents of other alphabets directory. The merged documents of all the directory are further preprocessed and they are ranked based on the creation of indexes and text traffic.

## IV. COMPUTING USER INFLUENCE

The observation reveals the impact of user of various category on different users with the corresponding impact. Here different color circle describe different users with corresponding edges from source node to destination as the impact of the respective user to different user .This impact can be calculated with the help of ranking algorithm. Assenting to these interpretations, calculating user dominance step will include two steps: Weighted in-degree and User Rank.

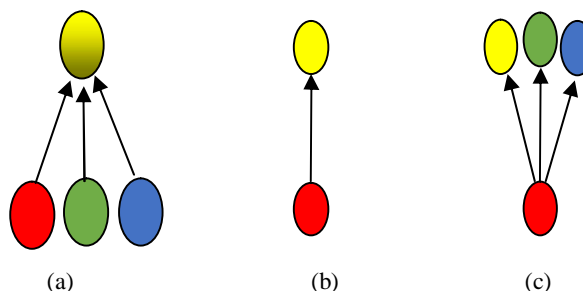


Fig. 3 Structure showing user A's participation in activity of forum

### A. Weighted in-degree

In a directed graph, In-degree of a node is the number of head endpoints neighbouring to this node. In graph Concept, In-degree is broadly utilized as a centrality measurement to measure the significance of a node in a grid.

$$ID(u_j) = \sum_{\forall u_i: e_{ij} \in E} 1(1)$$



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Using edge weights calculation, weighted in-degree is a direct way of calculating user dominance. Given a weighted community grid, user's dominance score is equal to the sum of weights on entire in-link edges of the grid.

$$WID(u_j) = \sum_{\forall u_i, e_{ij} \in E} W(u_i, u_j) \quad (2)$$

## B. User Rank

Since the user influence within a community grid is like the web page status in a hyperlink network, we stretched Page Rank algorithm and projected a User Rank algorithm to measure user dominance in a weighed community grid that we created. In page rank  $Pr(u_j/u_i)$  equals to  $1/(\text{out-degree } u_i)$  where out-degree ( $u_i$ ) is number of out link from  $u_i$ . In Page Rank algorithm all web pages are given a page rank score, which is unique over all the available hyperlinks available in the grid. The scores are updated until it is converged.

$$PR(u_j) = (1 - d) + d \sum_{\forall u_i, e_{ij} \in E} Pr\left(\frac{u_j}{u_i}\right) PR(u_i) \quad (3)$$

To include the content match and response, in our suggested User Rank algorithm we deploy a weighted community grid. As a end result, the transition probability can be calculated as follow:

$$Pr\left(\frac{u_j}{u_i}\right) = \frac{W(u_i, u_j)}{\sum_{\forall u_k, e_{ik} \in E} W(u_i, u_k)} \quad (4)$$

The User Rank scores will then be calculated by the following equation:

$$UR(u_j) = (1 - d) + d \sum_{\forall u_i, e_{ij} \in E} \frac{W(u_i, u_j)}{\sum_{\forall u_k, e_{ik} \in E} W(u_i, u_k)} UR(u_i) \quad (5)$$

We calculate Term Frequency and Inverse Document Frequency for the required term and frequency of word on the basis of presence of term in various documents. This value of term frequency and Inverse Document Frequency results in the calculation of Cosine Similarity between documents. This value of Cosine Similarity is then checked if it is greater than 0.5, if this value is greater than 0.5 then it is termed as 1 in adjacency matrix .We apply customized page rank algorithm as an input of this matrix and identify the most influential users. As a results we find the most influential users in web forum and calculate the correlation among users.

## V.RESULTS

The directory of every alphabets with corresponding key is made, here the alphabet z is taken into consideration. The graph made by their corresponding association is shown in figure 4. In this way we find effective users for directory of respective alphabets. In the end we collaborate this result for single directory and we get the result of most dominating user. The input of user node in the form of adjacency matrix helps us to reveal the rank of various nodes in graph. This node resemble the number of users in the given graph or network .Thus we rank the users on the basis of presence of words and number of post frequency .The graph with respect to presence of various nodes are shown below in the figure 4. Correlation Coefficient thus measure the similarity between the ranking generated as a result of post frequency and radicalness. To measure this comparison we calculate pearson ,kendall and spearman coefficients, the value generated for this coefficient is given below in figure 5.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016



Fig 4.General graph formed after finding association

The Correlation value between rankings formed by post frequency, radicalness, user ranking yield the value as shown in figure 5.

Pearson's correlation measure is : 0.9411764705882353  
Spearman's footrule measure is : 0.8076923076923078  
Kendall's tau distance measure is : 0.6666666666666666

Fig 5.Correlation value between different attributes

## VI.CONCLUSIONS

Our method makes use of parallel computing at first, by simultaneously compiling every csv file of each alphabets into their corresponding graph. As this file is later merged to get the rank users of various directories of corresponding alphabets it is beneficial when the size of text files or data set is extremely bourgeois. In this work we had proposed a novel method to calculate user dominance by the help of creation of grid. User reply relation is used to get their association from data of different forum. In this way ranking enables us to find the most dominant user in online community forum effectively.

## VII.FUTURE WORK

In the future, we shall be collecting a database of the overall sampled dna and age of every online users so as to match them with the data of their online repository to increase the efficiency in the field relating to calculation of user dominance.

## REFERENCES

- [1] GJ.-H. Wang, T. Fu, M. Abbasi, and H. Chen, "A framework for exploring Gray Web forums: Analysis of forum-based communities in Taiwan," in *Proc. IEEE Int. Conf. ISI*, May 2006, pp. 498–503.
- [2] D. Kempe, J. Kleinberg, and E. Tardos, "Influential nodes in diffusion model for social networks," in *Proc. 32nd Int. Conf. ICALP*, 2005, pp. 1127–1138.
- [3] J. Qin, Y. Zhou, G. Lai, E. Reid, M. Sageman, and H. Chen, "The Dark Web portal project: Collecting and analyzing the presence of terrorist groups on the Web," in *Proc. IEEE Int. Conf. ISI*, May 2005, pp. 623–624.
- [4] S. Sizov, J. Graupmann, and M. Theobald, "From focused crawling to expert information: An application framework for Web exploration and portal generation," in *Proc. 29th Int. Conf. VLDB*, 2003, pp. 1105–1108.
- [5] J. Glaser, J. Dixit, and D. P. Green, "Studying hate crime with the Internet: What makes racists Advocate racial violence?" *J. Soc. Issues*, vol. 58, no. 1, pp. 177–193, 2002.
- [6] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *Proc 8th ACM SIGKDD Int. Conf. KDD*, 2002, pp. 61–70.
- [8] J. Kaplan and L. Weinberg, *The Emergence of a Euro-American Radical Right*. New Brunswick, NJ, USA: Rutgers Univ. Press, 1998.



ISSN(Online): 2320-9801  
ISSN (Print) : 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

- [9] Kishor Wagh and Satish Kolhe ,”Information retrieval based on semantic similarity using information content ,” in *IJCSI International Journal of Computer Science Issues*, Vol. 8, Issue 4, No 2, July 2011
- [10] C. C. Yang, X. Tang, and B. M. Thuraisingham, “An analysis of user influence ranking algorithms on Dark Web forums,” in Proc. ISI-KDD, 2010, Art. ID 10
- [11] S. Sizov, J. Graupmann, and M. Theobald, “From focused crawling to expert information: An application framework for Web exploration and portal generation,” in Proc. 29th Int. Conf. VLDB, 2003, pp. 1105–1108.