A Novel Approach for Finding Rank of the User based on User's Information Diffusion Region

Amrita Namtirtha
Department of IT
NIT Durgapur,India
namtirtha.asansol@gmail.com

Shaswat Gupta
Department of IT
NIT Durgapur,India
shaswatnit@gmail.com

Animesh Dutta
Department of IT
NIT Durgapur,India
animesh.dutta@it.nitdgp.ac.in

Biswanath Dutta
DRTC,ISI Bangalore
Bangalore,India
dutta2005@gmail.com

***Abstract*—Social media is one of the best platform to diffuse an emergency information within a very short time period. To diffuse an emergency information in a region, many journalists, government information vendors and companies keep track of influential nodes of that region. But, due to the enormous growth of social media volume, it is quite challenging to identify influential node. Many researchers have focused on user's micro-blogging activity to find influential node. We propose a new method to rank the node on the basis of user's probable information diffusion region rather than counting user's total micro-blogging activities. It is based on how many unique users are infected by the influence of targeted nodes. Using this ranking approach we can estimate users' probable diffusion network and it could be used for future diffusion reference. We perform a set of experiments on real friendship network and random information flow data-sets to evaluate our findings. Our algorithm successfully able to find out top-influential nodes in case of email and random dataset. Consequently comparison result shows that our algorithm outperforms PageRank and Highest follower algorithms.**

***Keywords*—Most Influential node, network reachability, information diffusion region**

I. INTRODUCTION

In today's globalization, social networking platform is mostly used to diffuse information worldwide. Social networking sites are powerful tool to diffuse an information among millions of users within a very short time across the network. Nowadays many companies, political organizations, advertising agencies use social media as e-marketing platform to maximize their profit and to become popular among millions of users. It is mainly used for e-marketing, political campaigning, announcing government initiatives, and most importantly diffusing emergency information during crisis situations [1], [2], [3], [4].

Social media platform works as the word of mouth [5]. Most of the people are willing to share their information with friends and are likely to be affected by the opinions of their friends and followers [6]. Like a chain reaction information transmits from one friend to another and an information diffusion network is generated. During the information diffusion in a particular network "whom should we choose from large volume of social network. The answer is the user who can maximize the information diffusion network most efficiently and effectively. These users are termed as influential user [6]. Many researchers proposed different definitions of most influential user in the context of their research direction. Many journalist, government information vendor and companies keep track of influential nodes to diffuse emergency information during emergencies like natural calamity, appearance of new diseases and urgent alertment from the government. Thus, importance of influential user is quite evident and detail study required to track such users in real time in order to spread the news across the network. It is important to know how many number of people known or aware about the fact rather counting the user's total number of conversions among them. It is not a straightforward method to find out top-influential users from a large volume of social media data, it requires depth study in several aspects. Formulating an efficient algorithm to identify the influential users from a large pool of data is also quite challenging. Moreover, considering the economic point of view spreading the required information throughout the network with the help of least number of influential user is also desirable. Hence, our goal is to maximize the number of information receivers during any emergency information diffusion from a targeted node.

Based on above explained scenario, we propose a ranking system to rank individual node's in terms of their information diffusion region or information spreading zone. Information diffusion region means how many neighbour's level information are transmitted (depth and breadth wise) and number of unique users are getting infected by that information. It evaluates the rank of users' based on their existing prior information. In the proposed ranking system, top influential users means, those who holds large information diffusion network. To determine the information diffusion region of nodes we need to know two types of network structural data i.e. users' friendship network structure and real information flow structure. Figure 1:a shows two layer network structure; lower layer is friendship network and upper layer is original information flow structure. Every individuals have their own power to influence others in a society via message passing. By this influencing power every active person constructs their own information diffusion region and this region we call as the infected region. Let's take an example as described in Figure 1:b where node 'S' spreads a message in the network and message transmits one neighbor's level to another. In that way, node 'S' constructs a diffusion region (green shaded zone) with 5 active influential nodes. This concept originally introduced by kartz [7] to address the node centrality in terms of influence flow. We have taken two types of datasets to conduct our experiment. First dataset is real-world dataset for friendship network and second is real information flow dataset generated randomly on friendship network in order to illustrate the effectiveness of our approach compare to previous work.

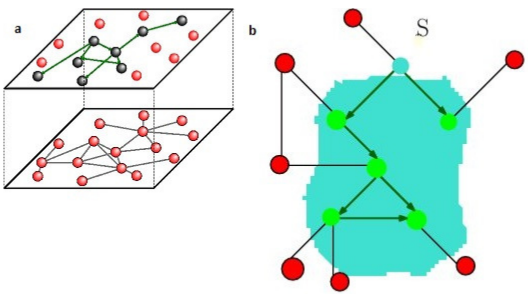

Figure 1:

a. An example of two layers structure, the lower layer displays social friendship network while the upper layer represents the information flow graph. b. An example of a diffusion region of node s and it shaded as green and it contain total 5 active unique node

The main contributions of this paper are as follows:
1) We propose a ranking system which will evaluate rank of users' in terms of individuals' information spreading region on the network.
2) Using this ranking system we can estimate users' probable information diffusion network and it could be used for future diffusion reference as diffusion epidemiological model [8].
3) Comparison result shows that our algorithm performs better than PageRank and Highest follower algorithms.

The rest of the paper is organized as follows: Section 2 briefs the summary of related work. Section 3 introduces the scope of the work. In section 4, we give a detailed analysis of our approach while in Section 5, algorithms are elaborated in detail. In section 6, experimental setup and result are discussed in detail. Section 7, evaluation of our model. Finally, Section 8 concludes with a discussion on our findings and directions for further research.

II. RELATED WORK

In today's e-marketing era, twitter or other social networking sites are becoming popular for providing platform of information diffusion and advertisement. In order to diffuse emergency information in the network most effectively we need to target those users who are most active influential. To identify most influential node from the network many researchers have focused on finding central node in network. Among them most popular ranking systems on graph are page rank [9] and HITS [10] algorithms. In PageRank algorithm, every node of a graph gets a numeric value and this value represents the importance of nodes in the network. PageRank of a node is based on number of in-degree of node's that means PageRank calculation depends on structure of the network. HITS [10] algorithm is sensitive towards out-degree of node's. In-degree and out-degree notion mainly indicate towards the popular social leaders, celebrities, and news channel hubs. With the flavors of page rank and HITS algorithm many researchers proposed different ranking systems to find out influential user. With the concept of page rank algorithm Kwak et al. [11] proposed a simple ranking system based on user in-degree or number of followers. TunkRank [12] is another flavor of PageRank. Author states that user's tweet getting retweeted more on the network if user's having more number of followers. But author Cha et. al [13] mentioned influence of twitters are measured by three major factors i.e. in-degree ,mention, and retweet. Author found that in-degree represents the user's popularity but it doesn't always imply that users have high influence capacity to engage audience by retweets and mentions. On the hand, few ranking systems adapted PageRank and Hits algorithms which are based on total micro-bloging activities such as users' total number of retweets and mentions activity [14], [15], [16], [17]. KHY Rank [18] determined influence based on graph structures. Khyrank calculated user influence by dividing the number of mentions or retweets done by user with the total number of mentions and retweets of an event. M. Zhang et al [19] have introduced the idea of pagerank with action based dynamic ranking model .They have proposed two types of weighted link structural transition probability on retweets and replies. It's yielded good result but it is unable to predict diffusion region of individual's. Twitter Rank [20] is the extended version of PageRank. Twitter Rank ranked the twitter user based on number of tweets published by user on a specific topic and how much topical similarities exist between user and their followers. But twitter rank has not considered retweet and mention activity of users. Many users actively participated in events by sharing or mentioning the information on tweets but twitter rank returns zero value for those users. Georgios et al. [21] extended the concept of Twitter Rank with the flavor of supervised random walk. As described in [21], Gayo-Avello et al. [17] state that link structure of network is not mandatory while measuring the influential users. They evaluated influential user on the basis of number of mentions and followers and it could be computed in real time. On the other hand, author Romero et al. [22] claimed that a majority of twitter users are might be passive. They measured user's influence based on two aspects i.e. user's passivity and influence rate on followers. They used HITS algorithm to evaluate users' rank.

III. SCOPE OF THE WORK

Previous ranking systems [14], [15], [16], [17] are mostly implemented based on flavor of PageRank [9] and HITS [10] algorithms and evaluated the rank based on user's number of occurrence of microblogging activities such as number of sharing, mentioning or conversions. Few ranking systems [11], [12] computed rank based on number of followers. But, these ranking systems are not adequate enough to retrieve prior information about the information reachability on network or infection spreading capacity of nodes within neighbors level. Also previous ranking systems unable to visualize network enlargement growth from a targeted node or in other words what would be the probable infected area in the network. As per our studies, previous researchers have not evaluated the user's rank based on user's information spreading zone or diffusion region, which could be used for emergency information diffusion in future. During emergency it is important to know how many numbers of people know or alert about the incident rather counting the user's total number of conversations among them. In order to find the most influential node in the diffusion network we propose a ranking based system on user's probable information spreading capability on that network. Inspired by Katz [7], our ranking system first determines the probable information spreading area starting from a particular node. Then, rank each node based on user's influencing power on the network. And we choose the node with the maximum rank and that node is declare as top-influential node.

IV. PROBLEM FORMULATION

In order to diffuse emergency information in the network, the number of active receivers need to be maximized and optimize the diffusion process time. Rank of nodes' are evaluated in terms of node's information diffusion region in the network. Lets represent a social media network in the form of a directed graph $G[V,E]$ where V is the set of nodes and E is the set of edges which is denoted as follower/friend relationship between users. Now, consider an information I_r traverse in a network and thus constructs an active directed graph $G' = [V', E'] \subseteq G[V,E]$. $V' = \{v_i | i \in [1,n]\}$ Where V' is the number of users received the information I_r and perform some activities such as sharing or commenting. $E' = \{e_{ij} | i, j \in [1,m]\}$ where e_{ij} communication link between v_i to v_j. Now, consider user $v \in V'$ shares an information I_r in the network. First I_r is received by neighbors of v i.e. $N(v)$. If $N(v)$ shares the information I_r then it goes to immediate neighbors of $N(v)$ and neighbor's neighbors of v. Like this way a neighbor's chain reaction is created and diffusion network of node v enlarges like $v \longrightarrow N(v) \longrightarrow N_1(v) \longrightarrow N_2(v) \ldots \longrightarrow N_h(v)$ where, h mean Hop level or neighbor's level. Hence, network enlargement capacity of v not only depends on active followers directly but it also can affect non-follower nodes.

Now, consider a probability function to estimate affected area due to spread of information I_r .Probability of infection spread by a node v, is represented as the ratio of number of unique active node generated from followers of v i.e. N_{act} and total number followers of v i.e. $F_w[v]$.

$$P_{Ir}(v) = N_{act}/F_w[v] \tag{1}$$

Probability of number of active node generated by a single node in its immediate hop/neighbors level is the summation of active node generate by each active node of that hop.

$$[P_{Ir}(v)]_h = \sum [N_{acth}/F_w[v]]_{i_a} \tag{2}$$

Here, h means neighbor's level and i_a represents active node in the hop h. We have assumed information spreading path of user v will be saturated at some hop level say h_{max} and information influence power of v will decrease gradually at every hop level. Let's assume α is the Information influence decay factor. The value of decay factor is not an absolute value and it lies between $0 > \alpha > 1$. Now, consider $C_{Ir}(v)$ is the network reachability capacity of user v. Hence, we can write $C_{Ir}(v)$ as –

$$C_{Ir}(v) = \alpha * [[P_{Ir}(v)/2]_1 + [P_{Ir}(v)/2^2]_2 + \cdots + [P_{Ir}(v)/2^{h_{max}}]] \tag{3}$$

Hence we can rewrite the equation (3) as below

$$C_{Ir}(v) = \beta + \alpha \sum_{m=0}^{h_{max}} 2^{-m} [P_{Ir}(v)]_m \tag{4}$$

Where initial active node factor is β and $\beta < \alpha$. In our above discussed method a particular node may be counted more than ones which may create a looping situation in the network. As a result, we might get wrong estimation of the influential node. To get rid of that problem we have taken the following measure. We have counted a node at its first occurrence only and discarded every other occurrence of the same node. In the above formula (4), an active node might get zero value if node hop level reachability is zero $(h_{max} = 0)$ though user is an active participant. To overcome the null network reachability problem we define an initial factor β. Similarly, we derive network reachability for other nodes $V' - v$ in information flow graph G'. Now, lets consider an event where k is the number of post published in the network $I = \{I_r | r \in [1,k]\}$. As mentioned earlier, each Information I_r flow structure generates a graph $G'[V', E']$ and evaluates the nodes' network reachability value. Final rank of a node is determined by averaging all the network reachability values generated from each information flow graph. Like this way, rank of all participated node in the event are evaluated. Top ranking users are most-influencing users from the aspect of information diffusion region.

V. PROPOSED ALGORITHMS

To formulate our methodology we propose two algorithms 1 and 2. Algorithm 1 evaluates the network reachability value (NR Calculation) of each node's in the information flow graph. Algorithm 2 determines overall rank of each node in an event.

Algorithm 1 NRCalculation $((G'_{Ir}[V', E'], G[V,E], v))$

Input: Directed Acyclic information flow Graph $G'_{Ir}[V', E']$ per information $I_r \in I$. And Initialize node $v \in V'$ as Information publisher.
Output: Network reachability value $C_{I_r}[V']$ is for every participants V' in Graph $G_{Ir}[V', E']$.

1: $S_l[] = G'_{Ir}.S(v)$ ▷ Check, any successor link of node v is present or not
2: **if** $v.S_l == \phi$ **then**
3: $C_{Ir}[v] \leftarrow 0$
4: Return $C_{Ir}[v]$
5: **else** ▷ If node v consist any successor link then calculate reach-ability of their successor using recursion function call and bottom up approach.
6: Set $T_{net} \leftarrow 0$ ▷ This variable holds the network reachability value for a node
7: **for** v.S $\in length(S_l[])$ **do**
8: $c_{value} =$
9: NRCalculation
10: $(G'_{Ir}[V', E'], G[V,E], S(v))$ ▷ Calculate successor network reachable value
11: $T_{net} \leftarrow T_{net} + c_{value}/2$
12: $c_{hop} \leftarrow \alpha * length(S_l[])$
13: $T_{net} \leftarrow T_{net} + c_{hop}$
14: $C_{Ir}[v] \leftarrow (T_{net} * N_{act}[v]/F_w[v])$
15: **end for**
16: Return $C_{Ir}[v]$
17: **end if**

In algorithm 1 we pass the input as information flow graph, friendship network, and root node of information flow graph. Steps [1-4] check successor links of node v exist or not. If successor links of node v are not present then network reachability of node v is zero. If node v has successor links then, network reachability for all successor is calculated using recursion function call [steps 7-10]. Remaining part of the algorithm [steps 11-14] explain how to calculate network reachability value using the defined objective function.

Algorithm 2 RankCalculation

Input: Information set $I = \{I_1, I_2, I_3, I_4 \ldots I_k\}$ for a particular event and Friendship graph G[V,E]
Output: Rank value R[v] for every participant in the event

1: **for** I_r to I **do**
2: Generate Information flow graph $G'_{Ir}[V', E']$ for each information I_r
3: $C[I][V']$=NRCalculation $(G'_{Ir}[V', E'], G[V, E]v) + \beta$ ▷ Set $v \in V'$ as information publisher and $C[I][V']$ is the Network reachability matrix which hold user network reachable value per information I_r
4: **end for**
5: **for** v to Every participants in event[] **do**
6: $R[v] = \sum C[I][v]/k$ ▷ rank of v calculated by sum of network reachability value divide by total number of information flow graph.
7: **end for**

Algorithm 2 calculates rank of every participants of an event. For k number of information flow graphs network reach-ability values are calculated in steps [1-4]. And remaining part of the algorithm [steps 5-7] evaluate the rank of the user.

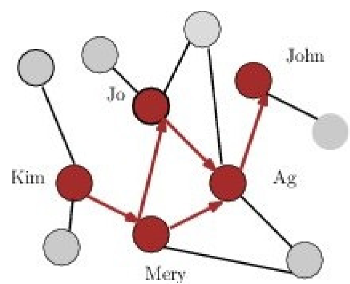

Figure 2: An Example to explain how information flow graph (mark as red) generate on network topology

Let's take an example to calculate network reachability value. In Figure 2 shown information flow structure(red mark) on friendship network. Now consider node kim is publisher of information I_r and information first received by kim neighbors . Like a chain reaction information transmitted from one neighbor's level to another. Now network reachability value of kim estimate by- $C_{Ir}(kim) = 0.5*1/3+0.25*2/4+0.125*1/5 = 0.3166$ where decay factor $\alpha = 0.5$.

VI. EXPERIMENTAL SETUP AND RESULT

We evaluated the rank of users on the basis of real dynamics information flow on the network topology (Figure 1). Two types of data-sets, i.e. network topology structure and real information flow structure are used to conduct our experiments. Table I summarizes the detail properties of the network topology and information flow data. We have collected real email and blog contact network data-set from [23], [24]. During the collection of real information flow data-set we faced some difficulties due to the company privacy policies of client's. Most of the popular social networking sites (e.g twitter,google+ etc.) impose rate limit on data and due to this important part of information is absent in public API[1] .

To simulate with more realistic information we generated random information flow graph on the network topology (Table I). Random information flow graph is generated using python random package [2] . We have set α = 0.5 according to Katz [7] and β signifies minimal arbitrary value which is less than α. To compute experimental results we generated information flow tree from information flow graph by discarding duplicate occurrence of nodes. Then, information flow tree is traversed in breadth first search (BFS) fashion to calculate network reach-ability of nodes. Finally, we optimized the computation time of individual network's reach-ability with the help of dynamic programming approach. Our experimental datasets are available in GitHub[3] for further research.

NetworkName	Type	$\|V\|$	$\|E\|$	$\|G'[V', E']\|$
Random	Directed	3000	4476527	1000
Email	Undirected	1134	5453	500
Blog	Undirected	1490	19090	373

Table I: Data Description

A. EVALUATION METRIC

Due to the lack of real information flow data and ground truth unavailability, ranking of top-influential node is very challenging and ambiguous even by manual survey. To verify our system and compare with other existing systems, we assume an evaluation metric in perspective of graph. Defined metric evaluate top-influential nodes in terms of their information diffusion region in graph, which means how many number of active nodes influenced by the top influence node. As well as to verify the correctness of the ranking methods, we consider Kendall Tau $\mathcal{T}$ rank correlation coefficient[4]. It measures the similarity between two ranking systems using their rank data. Tau correlation range lies between [-1, 1]. Two ranking systems give almost identical results if correlation value near to +1 and negative correlation value -1 means two system are not producing similar results.

B. EVALUATE OUR SYSTEM

Figures 3-5 demonstrate the experimental comparison based plot of our algorithm for each dataset. In Figures 3-5, horizontal axis represents top-influential node's ranking number and vertical axis represents count of infected node generated by each top-influential node. In Figures 3-5 we can observe top-influential nodes hold diffusion region as per their ranking number. Experiments with email and random datasets relatively meet our claim. Whereas the blog dataset result slightly differs from our expectation for certain users. Further analysis reveals that blog users' connection graph may not always be a connected graph i.e. a disjoint graph.

[1]https://dev.twitter.com/rest/public/rate-limiting
[2] https://docs.python.org/2/library/random.html
[3]https://github.com/shaswat770/NetworkReachabilityCalculationData.git
[4]https://en.wikipedia.org/wiki/Kendall..

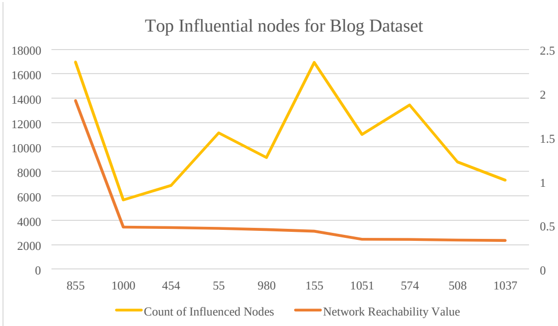

Figure 3: Line graph to show the comparison between count of influenced users and NR value of top-10 most influenced users for blog data-set

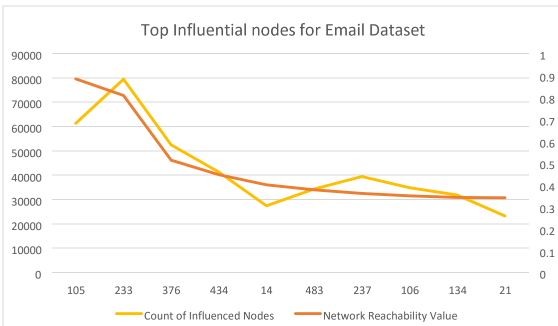

Figure 4: Line graph to show the comparison between count of influenced users and NR value of top-10 most influenced users for Email data-set

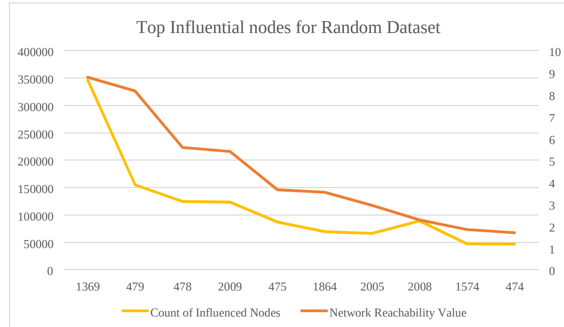

Figure 5: Line graph to show the comparison between count of influenced users and NR value of top-10 most influenced users for Random data-set

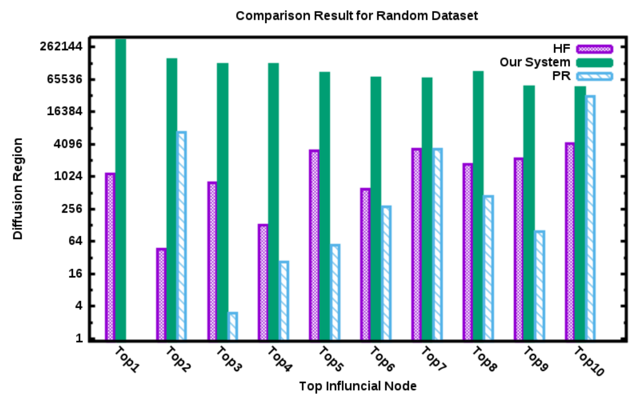

Figure 6: Comparison Result for Random data-set

VII. COMPARE WITH OTHER METHODS

To compare our system with other existing methods we consider two well known popular existing systems: PageRank(PR) and Highest followers of node's (HF). To estimate top-influential node from other methods we set some experimental setup as follows. We run the pagerank algorithm on each information flow graph. To evaluate top-influential page rankers we sum and average the pagerank value of individual node's participation in each information flow graph. In pagerank algorithm, α set to default value (0.85). We estimate highest followers (HF) of node's from followers graph. To make a proper comparison platform with other methods (HF, PR), we apply same defined metric i.e information diffusion region of top-influential node's. We can observe the results in Figures 6-8, our system (Network Reachability Calculation :NR) clearly outperforms all others systems in terms of information diffusion region of node's. As shown in Figures 7-8, pagerank output is not better than HF and our system. It may be due to the fact that pagerank works well on node in-degree notion and in case of strongly connected graph.

A. CORRELATION

Table II shows the correlation result between PR, HF and our system (NR). From the definition of kendall tau ranking correlation[5], we can observe in Table II that our ranking system (NR) is not similar to PR and HF though very low percentage of similarity is found between NR and HF for blogs and email datasets.

Correlation Result				
Dataset	Method	PR	HF	NR
Email	PR	-	-0.13	0.35
	HF	-0.13	-	0.33
	NR	0.35	0.33	-
Random	PR	-	0.2	-0.46
	HF	0.2	-	-0.55
	NR	-0.46	-0.55	-
Blog	PR	-	-0.25	-0.35
	HF	-0.25	-	0.44
	NR	-0.35	0.44	-

Table II: Kandell Tau Correlation with different methods

VIII. CONCLUSION AND FUTURE WORK

To maximize the information diffusion network during emergency situation, we have proposed a new ranking system which can find top-influential nodes from social network. In contrast, previous work unable to evaluate the rank of users based on user's information diffusion region, whereas proposed algorithm calculates user's diffusion region based on how many unique users are informed about an information from a information source node. We have used dynamic programming

[5]https://en.wikipedia.org/wiki/Kendall..

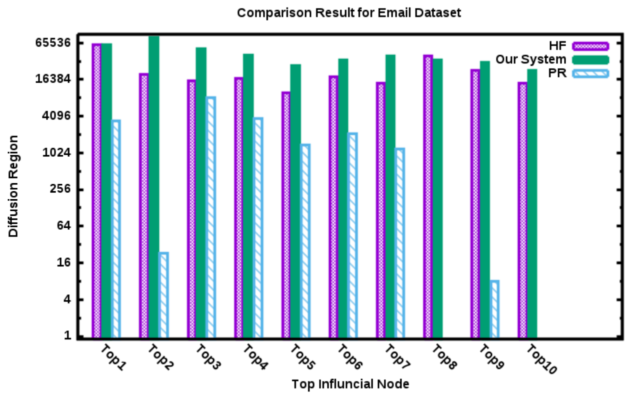

Figure 7: Comparison Result for Email data-set

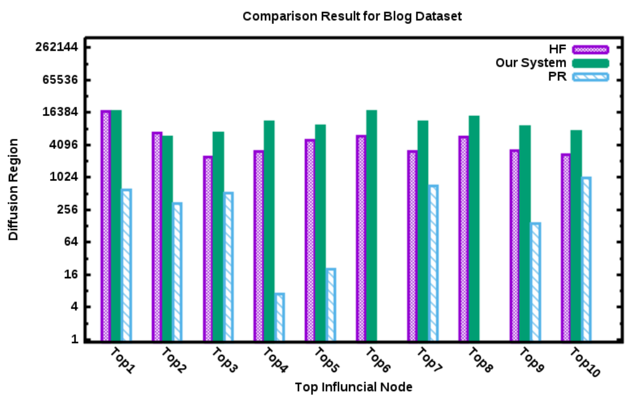

Figure 8: Comparison Result for Blog data-set

approach to optimize the time. Evaluation result proves that our algorithm can efficiently rank influential nodes. Comparison results show that our algorithm outperforms pagerank and highest follower algorithms. Limitation of our proposed system could be that it might not work efficiently for the case of information spreading by news channels. In future, we will plan to implement the current work with epidemiological diffusion model.

IX. Acknowledgments

This research work is funded by Visvesvaraya PhD scheme of DeitY (Department of Electronics & Information Technology), Govt. of India.

References

[1] M. Vergeer and L. Hermans, "Campaigning on twitter: Microblogging and online social networking as campaign tools in the 2010 general elections in the netherlands," *Journal of Computer-Mediated Communication*, vol. 18, no. 4, pp. 399–419, 2013.

[2] N. L. Chan and B. D. Guillet, "Investigation of social media marketing: how does the hotel industry in hong kong perform in marketing on social media websites?" *Journal of Travel & Tourism Marketing*, vol. 28, no. 4, pp. 345–368, 2011.

[3] A. L. Kavanaugh, E. A. Fox, S. D. Sheetz, S. Yang, L. T. Li, D. J. Shoemaker, A. Natsev, and L. Xie, "Social media use by government: From the routine to the critical," *Government Information Quarterly*, vol. 29, no. 4, pp. 480–491, 2012.

[4] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: real-time event detection by social sensors," in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 851–860.

[5] M. Trusov, R. E. Bucklin, and K. Pauwels, "Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site," *Journal of marketing*, vol. 73, no. 5, pp. 90–102, 2009.

[6] D. Kempe, J. Kleinberg, and É. Tardos, "Influential nodes in a diffusion model for social networks," in *Automata, languages and programming*. Springer, 2005, pp. 1127–1138.

[7] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953.

[8] H. Kim, K. Beznosov, and E. Yoneki, "A study on the influential neighbors to maximize information diffusion in online social networks," *Computational Social Networks*, vol. 2, no. 1, pp. 1–15, 2015.

[9] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: bringing order to the web." 1999.

[10] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne, "Finding high-quality content in social media," in *Proceedings of the 2008 International Conference on Web Search and Data Mining*. ACM, 2008, pp. 183–194.

[11] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 591–600.

[12] D. Tunkelang, "A twitter analog to pagerank," *The Noisy Channel*, 2009.

[13] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi, "Measuring user influence in twitter: The million follower fallacy." *ICWSM*, vol. 10, no. 10-17, p. 30, 2010.

[14] R. Cappelletti and N. Sastry, "Iarank: ranking users on twitter in near real-time, based on their information amplification potential," in *Social Informatics (SocialInformatics), 2012 International Conference on*. IEEE, 2012, pp. 70–77.

[15] A. Silva, S. Guimarães, W. Meira Jr, and M. Zaki, "Profilerank: finding relevant content and influential users based on information diffusion," in *Proceedings of the 7th Workshop on Social Network Mining and Analysis*. ACM, 2013, p. 2.

[16] I. Anger and C. Kittl, "Measuring influence on twitter," in *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*. ACM, 2011, p. 31.

[17] D. Gayo-Avello, D. J. Brenes, D. Fernández-Fernández, M. E. Fernández-Menéndez, and R. García-Suárez, "De retibus socialibus et legibus momenti," *EPL (Europhysics Letters)*, vol. 94, no. 3, p. 38001, 2011. [Online]. Available: http://stacks.iop.org/0295-5075/94/i=3/a=38001

[18] A. Yu, C. V. Hu, and A. Kilzer, "Khyrank: Using retweets and mentions to predict influential users," 2011.

[19] M. Zhang, C. Sun, and W. Liu, "Identifying influential users of micro-blogging services: A dynamic action-based network approach," in *Pacific Asia Conference on Information Systems (PACIS) 2011 Proceedings*, 2011, pp. Paper–223.

[20] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 2010, pp. 261–270.

[21] G. Katsimpras, D. Vogiatzis, and G. Paliouras, "Determining influential users with supervised random walks," in *Proceedings of the 24th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 2015, pp. 787–792.

[22] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman, "Influence and passivity in social media," in *Machine learning and knowledge discovery in databases*. Springer, 2011, pp. 18–33.

[23] L. A. Adamic and N. Glance, "The political blogosphere and the 2004 us election: divided they blog," in *Proceedings of the 3rd international workshop on Link discovery*. ACM, 2005, pp. 36–43.

[24] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas, "Self-similar community structure in a network of human interactions," *Physical review E*, vol. 68, no. 6, p. 065103, 2003.