Asian Journal of Information Technology 15 (15): 2663-2670, 2016
ISSN: 1682-3915


# Classification of Information Hubs Based on the Interest to Maximize Viral Marketing in Social Networks

A.N. Arularasan and S. Koteeswaran
Department of Computer Science and Engineering, Vel Tech University, Chennai, Tamil Nadu, India

**Abstract:** Social networks play an essential role in the online information diffusion. The understanding of the social relationship among users is a factor beneficial to various applications like viral marketing and advertising through social networks. The advertisers aim to select initially a small number of information hubs and provide free samples of a new product to influence a large number of people to buy the product in a short time. Estimating the influence of users, discussing different domains like daily charts, movies, celebrities and sports require analysis of short tweets posted to a social network. With this intent, the fundamental motive is to maximize the influence of a social network in a certain domain of interest. This work proposes support of self-centered network on domain specific information hub classification (speed-diffusion) technique with the aim of classifying information hubs in different domains. Speed-diffusion is an unsupervised classification model, where the output is based on the software analysis without using training samples. The speed-diffusion technique integrates the self-centered network and N-gram classification model to classify the domain specific information hubs. Speed-diffusion represents each domain as a self-centered network, automatically built using WordNet and Yago Ontology. Analysis of short tweets posted to a social network quantifies the general influence of each user with respect to the self-centered networks of various categories. The N-gram classifier leverages these short tweets posted by the user depending on how far its words are from the self-centered network to mine the information hubs specific to a domain. Finally, it evaluates the accuracy of the speed-diffusion technique than that of the existing approaches using a sample of the Twitter network.

**Key words:** Influence, twitter, classification, self-centered network and N-gram classification model, India

## INTRODUCTION

Recently, the proliferation of mobile devices and wireless technologies has increased the use of large scale social networks online. The social networks act as a popular and efficient medium for information diffusion (Domingos and Richardson, 2001). Understanding short tweets posted to a network has a huge impact on the business analysis, ranging from the study of online users' behavior to advertisement and security purposes. Viral marketing is one of the main application of influence maximization that aims to promote the diffusion of product information into the social network and get maximum profit from all the users. However, influence maximization is not always appropriate for viral marketing because some products, in fact, are useful only to specific users. Companies try to fit between the products and the online users to maximize the product sales (Goldenberg *et al*., 2009; Bakshy *et al*., 2012).

The company aims to target initially a set of information hubs and provides free samples of the new product to them. Companies hope that the information hubs recommend the new product and ultimately adopt the new product through the effect of word-of-mouth on the network in a short period (Chen *et al*., 2009, 2010). The social website invites the users to discuss different domains like daily chats, economics, movies, celebrities and sports with their followers in a same network. The long-standing challenge is to identify the highly influential users or information hubs in large-scale social networks.

The existing works aim at selecting a small number of users with respect to the domain of discussion and targeting them initially to maximize the diffusion of information on a community to cope with the challenges. They employ different classifiers for categorizing the information hubs under various domains based on hubs interest. The major constraint of these approaches is that they use the bag-of-words model for categorizing the texts. For each message, represented as feature vectors measure term frequency or term frequency-inverse document frequency of the words. However, short

**Corresponding Author:** A.N. Arularasan, Department of Computer Science and Engineering, Vel Tech University, Chennai, Tamil Nadu, India

messages posted to a community do not provide sufficient word occurrences and so it is better to use distance metric than the frequency to accurately classify the short messages. Instead of using a representative training set as a requirement to build the classification model, this research proposes the speed-diffusion technique that leverages the predefined set of categories and the knowledge about a particular domain in terms of entities, relations and events. The main contributions of the study are. The main objective of the proposed unsupervised classification methodology called speed-diffusion is to classify the users whose behavior have the most significant impact on the activities of others over the social network in a particular domain. Speed-diffusion employs the knowledge of the self-centered network in terms of centroids, concepts and entities extracted from the WordNet Domains and Yago Ontology. It significantly reduces the classification time of tweets into different categories of interests.

Speed-diffusion infers the general information hubs on social network using N-gram similarity model that can find the most relevant categories of a given tweet depending on the measured semantic score of each category and improves the efficiency of speed-diffusion. Instead of using word frequencies, speed-diffusion implements proximity measure between words and entities appearing in relevant categories of self-centered networks and successfully mining the information hubs from general to domain specific. The experimental results show that the speed-diffusion scale to networks with millions of users and efficiently to classify the information hubs under various domains of interests.

**Literature review:** Many techniques exist for classifying short messages posted on a social network. The supervised learning technique (Zhang *et al*., 2013) exploits labeled training set with transfer learning process that conveys the knowledge learned from one domain to another domain. It applies an iterative classification approach on test data by identifying the information paths over the data. To support dynamically changing relationship among users, it rebuilds the classifier on different training and test data set to achieve better accuracy in each iteration. However, the supervised classification model consists of many training and testing iterations that are time-consuming and expensive. The other techniques consider corpus and use the topic as features into account for classifying the information hubs.

For instance, a decreasing cascade model spreads the behavior in a cascading manner is proposed and it is based on the probabilistic rule, beginning with a set of initially "active" nodes (Kim *et al*., 2014). A two-stage mining algorithm (GAUP) (Zhang *et al*., 2011) is proposed for discovering most influential node on a particular topic. A new algorithm is suggested to determine the information hubs (Ilyas *et al*., 2013) and it can preserve user privacy. It identifies the information hubs without requiring a central entity to access entire community graph. It achieves the above objective by completely distributing the computation using the Kempe-Mcsherry algorithm and also addresses the user privacy concerns.

Based on the motivation for promoting the viral marketing, several works have been proposed on influence maximization in social networks (Cataldi *et al*., 2013; Wang *et al*., 2014; Lee and Chung, 2015). The target set selection problem for influential spreading under the scenario of mobile data offloading is difficult (Liu *et al*., 2014; Han and Srinivasan, 2012). However, the authentic determination of interactions between users is not provided on the social networks and it is a tough task in real scenarios.

Even though analysis of the topic correlation on training and test data improves the classification accuracy, it needs both a significant training set and an external auxiliary data. A method in (Sun, 2012) uses representative query words related to the short tweets posted by the user, achieves similar results with the accuracy of categorization of information hubs with respect to the domain of discussion. These works describe that some inherent features of short texts that play a critical role in the classification. However, it still requires a labeled dataset to train the classification, model. To overcome this issue, the proposed work takes into account a different domain knowledge from WordNet domains and Yago ontology and N-gram similarity for identifying the information hubs effectively over social networks.

**Problem statement:** Most of the research in viral marketing over social networks has been focused on selecting the information hubs for improving the influence maximization. Several supervised classification algorithms like support vector machine and naive bayes have been proposed to identify the information hubs in various topics of interests. These works need a representative training set to train the classification model. This type of training set is very expensive and not currently available in public. A further limitation of these works is that they consider the word frequencies in identifying the information hub and the interest or the influence of users in a particular domain. However, the short messages posted on a social network do not provide sufficient word occurrences. This study proposes a new technique that takes into account the distance metric instead of using word frequency to resolve these issues. It exploits the knowledge of self-centered network for identifying the domain specific information hubs more efficiently over social networks.

**Overview of the speed-diffusion:** The knowledge-based speed-diffusion technique aims at identifying the information hubs in different domains of interests and improving the diffusion of business ideas or innovations in social networks to maximize the viral marketing. It is necessary to quantify each user's overall influence and interest to classify the information hubs in a particular domain. Speed-diffusion relies on three phases with the aim of improving information diffusion over a network. They are:

- Building self-centered network
- N-gram similarity
- Information hub identification

**Building self-centered network:** The proposed speed-diffusion technique represents each domain as a self-centered network. To build a self-centered network, it connects each centroid to other words that are semantically related to it. The semantic relations in the self-centered network include the category and all those related entities derived from the Yago ontology (Suchanek *et al.*, 2008). Speed-diffusion initiates this process from the identification of centroids. Defining the category of textual contexts and senses of each text manually does not scale properly. Hence, speed-diffusion proposes the best strategy for automatic centroid selection using WordNet domains and Yago ontology.

## MATERIALS AND METHODS

**WordNet domains and Yago ontology:** WordNet domains are an extension of the lexical resource, WordNet. Each category in WordNet domains is modeled as a self-centered network using Yago ontology. All objects in the Yago ontology are represented as entities and two entities stand in relation. It derives the Yago class with a set of similar entities, related to a Synset of each category in WordNet domains. To design the self-centered network, the category in WordNet domains is connected to the related concepts/entities in Yago ontology.

Speed-diffusion uses the derivations of a concept or the related forms of a concept such as an economist, economical and the economy as centroids. Each Synset labeled with the economics category in WordNet domains is submitted as input to the Yago ontology. The Yago ontology groups set of semantically related entities to the given Synset into a same class. For example, in Fig. 1, entities related to the Synset of economics like Economic Science, Accounting, Financial, Banking and Clerking are grouped into a same class. It connects each entity to other concepts/entities in a same class semantically related to it. The speed-diffusion considers the related entities obtained from Yago ontology as concepts and repeats this step for each encountered concepts. Finally, it models the semantic structures of a category as self-centered network where concepts and edge labels such as Synonym, Hypernym, Hyponym and Pertainym represent the type of the semantic relationship between the concepts/entities.

**N-gram similarity:** The main aim of N-gram similarity implemented in speed-diffusion is to classify a short tweet into a category. If the words in a tweet provide high similarity with the entities in a certain category of a self-centered network, it is likely that tweet is related to that category. The N-gram similarity technique infers the categories of tweet having a relation between the words and entities appeared in self-centered networks of various categories.

Consider a set of self-centered networks of categories C = {c1, c2, ..., cm} and a set of short tweets T = {t1, t2, ..., tn} that needs to assign its semantically related categories using n-gram similarity method. The value of n is 3 (trigrams). For example, the phrase "mobile phone" consists of the following trigrams: mob, obi, bil, ile, le_, _ph, pho, hon and one. The speed-diffusion formalizes the entities in each category as weighted template vectors ($K_e$) of the n-gram. The speed-diffusion eliminates the stopwords in each tweet and then partitions it into a list of tokens. Each token is formalized into test vector ($K_t$) of N-gram. The N-gram similarity method compares test vectors with a set of template domain vectors:

$$Ke_j = \{e_{1,j}, e_{2,j}, e_{3,j}, ..., e_{v,j}\} \quad (1)$$

Where, j varies from 1-m.

$$Kt_j = \{t_{1,j}, t_{2,j}, t_{3,j}, ..., t_{v,j}\} \quad (2)$$

where, j varies from 1-m. For example, template vector of the jth entity of category, $e_j$ and test vector of the jth token in a tweet, $t_j$ are shown in the Eq. 1 and 2, respectively. Where, K is an N-grams vocabulary of the token/entity and v are equal to |K|. The N-gram

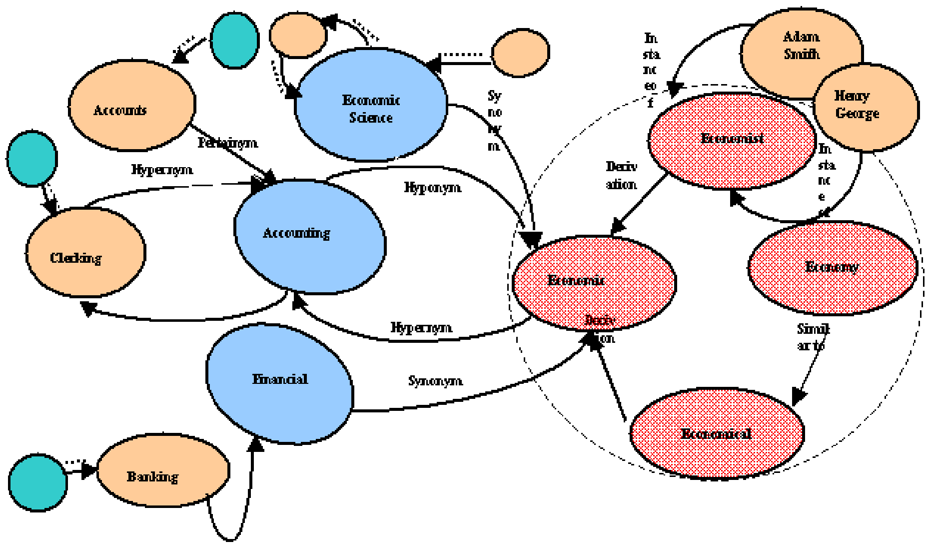


Fig. 1: Self-centered network of economic category

similarity method performs the string-matching process between a set of test and template domain vectors consisting concepts/elements in a self-centered network of various categories. The String similarity (Ss) measure for the two N-gram sets is calculated as shown in the Eq. 3. Where, $v_t$ and $v_e$ represent the total number of unique n-grams in a token of the tweet, t and an element in a category, c, respectively. The $C_{t,e}$ represents the total number of unique N-grams that are common to both tokens in a tweet and entity in the category. The Average of $S_s$ value (AVG $S_s$) of each tweet to a certain category represents the degree of relationship among tweets and that category. If the tweet has AVG $S_s \geq 0.6$ for a category c, it is considered as relevant one:

$$S_s = 2C_{t,e}/(v_w + v_e) \tag{3}$$

In case, if multiple categories to satisfy the condition of AVG $S_s \geq 0.6$, speed-diffusion measures the proximity of the high similarity or related entities from the centroid of the category. Thus, the distance metric is the most robust way to measure the semantically related categories for each tweet.

**Information hub identification:** Speed-diffusion uses the proximity measure to analyze how well the user diffuses information and estimates the influence of each user on the social network for each category that satisfies the condition of AVG $S_s \geq 0.6$. The smaller the distance between a related concept and centroid of a self-centered network of a certain category, the higher is the semantic similarity between the word and that category. The proximity between a concept and centroid is the shortest path. As the centroid is formed with a set of concepts, multiple shortest paths are created for a concept. Among them, the Minimum length of the Shortest Path (Min |SP|) is considered.

The distance between the concept and centroid, named as category distance is zero when the concept is one of the centroid and it is infinitive when the concept is not appearing in the self-centered network of a category. Thus, the category distance of an entity falls in the range of [0,8] as shown in the Eq. 4. Figure 1 shows economics and economic science are semantically more similar than economics and clerking because the 1st pair of entities is directly connected whereas the 2nd is connected through a path length of 2. Moreover, economics is semantically closer to the financial than Banking. The distance of entities Adam Smith and Henry George from the centroid, the Economist is the same and so they are equally semantically similar to the centroid. The semantic similarity score is maximum when the category distance is zero but it decreases when the category distance between the concept and centroid increases:

$$\left\{\begin{array}{ll} 0 & \text{If concept} \in \text{centroid} \\ \text{Category distance}_{\text{(concept-centroid)}} = \infty & \\ \text{If concept} \notin \text{SC} & \text{Min |SP| otherwise} \end{array}\right\} \tag{4}$$

**Semantic similarity score of a tweet:** Speed-diffusion measures the semantic similarity score of a tweet by considering the distance of category that satisfies the condition of $S_s$ = 0.6. Speed-diffusion employs a sum weighted function over the semantic scores of concepts/entities of all the words in a tweet with the idea that if category distance of the concept is small, the weight factor is high. Particularly, speed-diffusion assigns weight one to the concepts that have maximum category distance and for others the weight is equal to the ratio of Maximum distance metric ($MAX_{CE}$) with them:

$$\text{SS score}_{cn}\ (t) = \left(\sum_{cn \in CE} \{SS_{cn}/MAXSS\}\right) \Big/ \sum_{cn \in CE} MAXSS_{cn} \quad (5)$$

$$SS_{cn} = (1+\text{Category Distance}_{cn\text{-}CE})^{-1} \quad (6)$$

$$MAXSS_{cn} = (1+MAX_{CE})/(1+\text{category distance}_{cn\text{-}CE}) \quad (7)$$

Substituting Eq. 6 and 7 in Eq. 5, the Semantic Similarity (SS) score of a tweet is estimated. For each category extracted from N-gram similarity measure, speed-diffusion first ranks the users in descending order of SS $\text{score}_{cn}$ (t) and those users tweeted at least one message, categorized as related to a domain. The semantic similarity score represents the degree of influence for each user about each category extracted from the N-gram similarity measure. This score enables speed-diffusion to establish an ordering among users for each category based on their messages posted on a social network. Thus, the proposed information hub classification model with respect to the domain of discussion among online users improves the influence maximization in a particular domain of interest.

**Algorithm for speed-diffusion technique**

**Building a self-centered network**

**Algorithm 1; algorithm to classify the domain specific information hubs:**

Input: Set of Categories $C_g$ and Meaning of a category, $C_{gn}$ Cd ($C_{gn}$)
Output: The self-centered network of a category, $C_{gn}$
  for each $C_{gn}$ do
N <--- All the semantically related concepts to a $C_{gn}$ retrieved from the WordNet
    C <--- Derivations of Cd ($C_{gn}$)
    label <--- "Derivation of"
    for each C do
      Build (C, label, 0)
      VerifyYago (Cn ª C, 0)
    end
  end
Build (n, label, distance)
N <--- All the semantically related concepts to a C retrieved from Yago Ontology
    edge <--- (C, i ª N)
    E <--- edge;
    L <--- label;
    Distance <--- Distance (n, distance)
  end
VerifyYago (C, 0):
  For all concepts ª Yago's classes related to a C do
    for All entities e related to a C do
      label <-- Relation between e and C obtained from Yago Ontology
      Build( entity, label, distance+1);
    end
    Assign entity = C
    Distance = distance +1
    if C∉N then
      Continue;
    end
  End
    /*Classifying Domain Specific Information Hubs*/
Input: Set of self-centered networks of Categories C = {c1,c2,...cm} and Tweets posted by each user
Output: Information Hubs in each Category
Measuring General Influence of User ($C_g$, $Tri_{tweet}$)
  for each user (u) do
    for each $C_{gn}$ do
      Se = Set of entities in $C_{gn}$
        K = (Trigram of entity ($Tri_{entity}$) ª Se) U (Trigrams of all tokens of an input tweet ($Tri_{tweet}$))
        Measure Average $S_s$;
        if Average $(S_s)_{Cgn}$ > 0.6 then
          List $(C_{gn})_u$ <-- $C_{gn}$
        end
      end
    end
Measuring Influence of a User from General to Domain Specific (List $(C_{gn})_u$, $Tri_{tweet}$, $(Tri_{entity}$) ª $Se)_{Cgn}$):
    for each $C_{gn}$ do
    for each user do
      if $C_{gn}$ ª List $(C_{gn})_u$ then
        Measure Category Distance(C, Cd ($C_{gn}$));
          Measure Semantic Similarity Score (SS) of Tweets sent by a user
      end
    end
    Create List $(U)_{Cgn}$ in the descending order of SS value of user;
  End

The algorithm 1 demonstrates the process of speed-diffusion mechanism for classifying the domain specific information hubs in social networks. Building a self-centered network returns the network for each category using WordNet domains and Yago ontology. The verify Yago function expands the self-centered network using related entities retrieved from the Yago ontology. Measuring general influence of each user provides different related categories to the tweets posted to a social network. To further improve the performance of speed-diffusion, the distance between concept and origin and the centroid is measured in each related category of a tweet. The domain specific influence measurement function iterates over the list of interested categories of each user and ranks the domain specific information hubs successfully.

## RESULTS AND DISCUSSION

**Experimental evaluation:** The main aim of experimental evaluation is to test the performance of the proposed speed-diffusion on a real-time data set and to demonstrate the impact of parameters like number of tweets, number of users and number of available interested users in a particular category. On the other hand, comparing the effectiveness of the proposed classification approach, speed-diffusion in identifying the domain-based influence of Twitter users with real-time scenarios Estimating Domain-based User Influence (EDUI) (Cataldi *et al*., 2013). The proposed speed-diffusion consists of two algorithms like building a self-centered network for each category in WordNet domains and classifying domain-based information hubs in each category. These algorithms use Java and run on a Linux Ubuntu server with an 8×2.9 GHz CPU and 32 GB memory.

**Data set:** The performance of the proposed speed-diffusion method is evaluated on a Twitter dataset, consisting of 41.7 million user profiles, 1.47 billion social relations, 4.262 trending topics and 106 million tweets (Kwak *et al*., 2010). Some of the trending topics of the Twitter dataset are considered as categories of interest. This work builds a self-centered network for each category to classify the information hubs under different groups of interest. The distance measure based Semantic similarity score measurement is used to classify the domain specific information hubs in Twitter.

**Evaluation metrics:** The performance metrics of the proposed speed-diffusion technique are precision, recall, F-Measure and Tweet classification time per user.

**Precision:** The ratio between the No. of relevant labeled-information hubs and the total No. of retrieved labeled-information hubs in a certain domain of interest.

**Recall:** The ratio between the No. of retrieved labeled-information hub and the total No. of labeled information hub in a particular domain of interest.

**F-Measure:** F-Measure is defined as the harmonic mean of Recall and p recision:

$$\text{F-Measure} = 2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \qquad (8)$$

**Tweet classification time per user:** It is the time consumed to classify the user with respect to the domain of discussion.

**Experimental results:** The experimental results discuss the real time case scenarios in various tweets, users and categories.

**Impact of number of tweets:** Figure 2 illustrates the precision on considered Twitter dataset with speed-diffusion and EDUI. In this scenario, the number

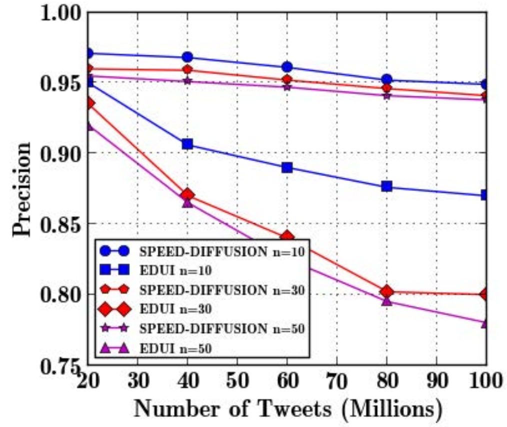


Fig. 2: No. of tweets vs precision

tweets is varied from 20-100 million and the number of considered Categories (n) is in the range of 10-50. Both the precision of speed-diffusion and EDUI regularly decreases when the number of tweets increases. The EDUI is less scalable than the proposed speed-diffusion on large data sets. It is because that the EDUI needs a much greater number of accesses to all the template vectors of the considered set of categories. Moreover, it is not compatible with the categorization of tweets when the number of tweets posted on a social network becomes large. However, the proposed speed-diffusion technique automatically provides built in a self-centered network using WordNet domains and Yago ontology to cope up with this fact. If the number of tweets increases from 8-40 million, the precision value drops to a range of 0.948 in speed-diffusion. It attains the precision value of 0.95 when the value of n is 10. However, it decreases the precision value of speed-diffusion in the range of 0.25% when the value of n reaches 50.

Figure 3 shows the quality of the domain-based information hub classification in speed-diffusion and EDUI algorithms in terms of recall with varying number of tweets posted on a social network, Twitter. Higher, the number of tweets and considered categories for classification, lesser the quality of classification approaches in both speed-diffusion and EDUI. For instance, the recall value increases in the range of 0.12% more than that in EDUI at the point of 100 millions of tweets with n = 10.

**Impact of number of users:** Figure 4 shows the F-measure achieved against the number of users on the social network. An increasing number of users and number of categories are considered for information hub classification that decays the F-measure

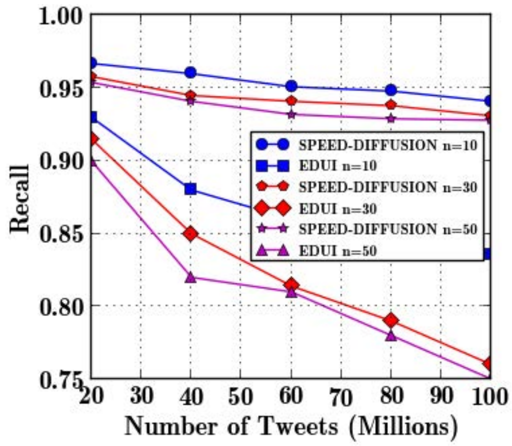


Fig. 3: Number of tweets vs recall

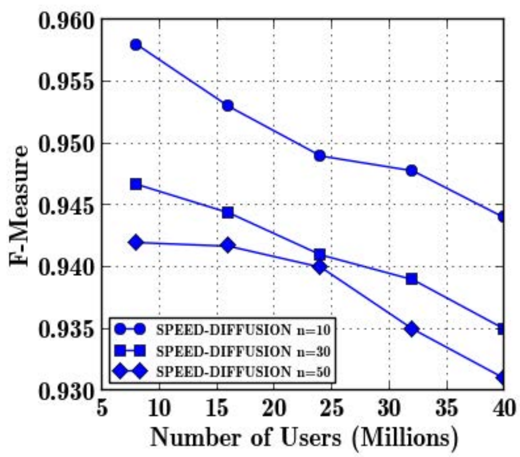


Fig. 4: Number of users vs F-measure

of speed-diffusion. It can classify the information hubs from general to a particular domain with high quality comparable to those classified by existing algorithms.

The speed-diffusion produces a better F- measure in the range of 0.93, even when the number of users in a social network is over 40 million. Figure 5 shows that with 50 categories to classify the information hubs, the F-measure decays below 0.943-0.93 with a maximum of 40 million online users in a social network

**Impact of number of categories:** Figure 5 demonstrates the Execution time with varying number of categories in the network. T represents the average number of tweets posted by a user on a social network and to vary the value of T, the number of users is ranged from 4-20 millions of users. The relation between the classification time of tweets posted by a user and the number of self-centered networks considered for

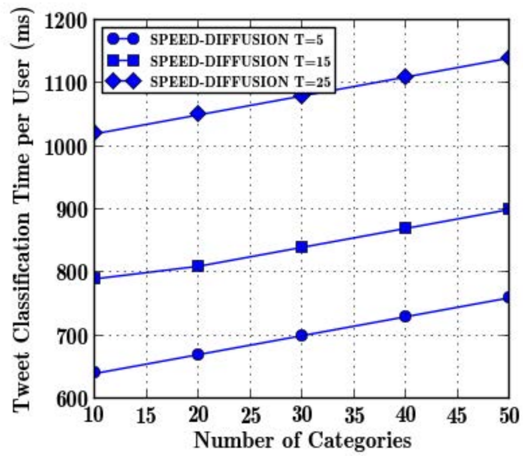


Fig. 5: Number of categories vs tweet classification time per user

information hub classification is linear. In all the instances, above 40% of the classification time is due to the built-in self-centered network in each category which the speed-diffusion depends on the WordNet domains and Yago ontology. When the number of categories is 10 and T is 5, Twitter classification time on Twitter dataset is 660 milliseconds and 460 milliseconds to classify all the tweets posted by a user. The time to build the self-centered network is increased to 651 sec when the number of classes in speed-diffusion reaches 50. Moreover, time to classify the tweets posted by the user is increased to 7.76 sec when the value of t reaches 25.

## CONCLUSION

This study has proposed an unsupervised and knowledge-based method for information hub classification in different domains of interests. The proposed classification approach, the speed-diffusion including self-centered network and N-gram classification model, categorizes the users into the different category of interests with respect to the domain of discussion or tweets posted into a social network. The speed-diffusion represents each domain as a self-centered network. It enables N-gram similarity measurement model to leverage the short tweets posted by the user depending on how far its words are from the self-centered network of a certain domain. It effectively mines the information hubs from general to domain specific. The self-centered network helps in reducing the accessing memory of template vectors of the corpus or any other resources unlike EDUI and thus, it identifies the domain-based information hubs in a short time. The experimental evaluation reveals that the proposed speed-diffusion outperforms the supervised EDUI classification approach in terms of recall and tweet classification time per user. The experimental results show that the speed-diffusion has improved the recall value by 0.2% as compared to EDUI.

## REFERENCES


Bakshy, E., I. Rosenn, C. Marlow and L. Adamic, 2012. The role of social networks in information diffusion. Proceedings of the 21st international conference on World Wide Web, April 16-20, 2012, ACM, New York, USA., ISBN: 978-1-4503-1229-5, pp: 519-528.

Cataldi, M., N. Mittal and M.A. Aufaure, 2013. Estimating domain-based user influence in social networks. Proceedings of the 28th Annual ACM. Symposium on Applied Computing, March 18-22, 2013, ACM, New York, USA., ISBN: 978-1-4503-1656-9, pp: 1957-1962.

Chen, W., C. Wang and Y. Wang, 2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 25-28, 2016, ACM, New York, USA., ISBN: 978-1-4503-0055-1, pp: 1029-1038.

Chen, W., Y. Wang and S. Yang, 2009. Efficient influence maximization in social networks. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, June 28-July 01, 2009, ACM, New York, USA., ISBN: 978-1-60558-495-9, pp: 199-208.

Domingos, P. and M. Richardson, 2001. Mining the network value of customers. Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 26-29, 2001, ACM, New York, USA., ISBN: 1-58113-391-X, pp: 57-66.

Goldenberg, J., S. Han, D.R. Lehmann and J.W. Hong, 2009. The role of hubs in the adoption process. J. Marketing, 73: 1-13.

Han, B. and A. Srinivasan, 2012. Your friends have more friends than you do: Identifying influential mobile users through random walks. Proceedings of the 13th ACM International Symposium on Mobile Ad Hoc Networking and Computing, June 11-14, 2012, ACM, New York, USA., ISBN: 978-1-4503-1281-3, pp: 5-14.

Ilyas, M.U., M.Z. Shafiq, A.X. Liu and H. Radha, 2013. A distributed algorithm for identifying information hubs in social networks. IEEE. J. Select. Areas Commun., 31: 629-640.

Kim, J., W. Lee and H. Yu, 2014. CT-IC: Continuously activated and time-restricted independent cascade model for viral marketing. Knowl. Based Syst., 62: 57-68.

Kwak, H., C. Lee, H. Park and S. Moon, 2010. What is Twitter, a social network or a news media? Proceedings of the 19th International Conference on World Wide Web, April 26-30, 2010, Raleigh, NC., USA., pp: 591-600.

Lee, J.R. and C.W. Chung, 2015. A query approach for influence maximization on specific users in social networks. IEEE. Trans. Knowl. Data Eng., 27: 340-353.

Liu, B., G. Cong, Y. Zeng, D. Xu and Y.M. Chee, 2014. Influence spreading path and its application to the time constrained social influence maximization problem and beyond. IEEE. Trans. Knowl. Data Eng., 26: 1904-1917.

Suchanek, F.M., G. Kasneci and G. Weikum, 2008. Yago: A large ontology from wikipedia and wordnet. Web Semant. Sci. Serv. Agents World Wide Web, 6: 203-217.

Sun, A., 2012. Short text classification using very few words. Proceedings of the 35th International ACM Sigir Conference on Research and Development in Information Retrieval, August 12-16, 2012, ACM, New York, USA., ISBN: 978-1-4503-1472-5, pp: 1145-1146.

Wang, G., W. Jiang, J. Wu and Z. Xiong, 2014. Fine-grained feature-based social influence evaluation in online social networks. IEEE. Trans. Parallel Distrib. Syst., 25: 2286-2296.

Zhang, S., X. Jin, D. Shen, B. Cao and X. Ding *et al*., 2013. Short text classification by detecting information path. Proceedings of the 22nd ACM. International Conference on Information Knowledge Management, October 27-November 01, 2013, ACM, New York, USA., ISBN: 978-1-4503-2263-8, pp: 727-732.

Zhang, Y., J. Zhou and J. Cheng, 2011. Preference-based top-k influential nodes mining in social networks. Proceedings of the IEEE. 10th International Conference on Trust, Security and Privacy in Computing and Communications, November 16-18, 2011, IEEE, Changsha, China, ISBN: 978-1-4577-2135-9, pp: 1512-1518.