The Impact of Social Diversity and Dynamic Influence Propagation for Identifying Influencers in Social Networks

Pei-Ying Huang, Hsin-Yu Liu, Chin-Hui Chen, and Pu-Jen Cheng
Department of Computer Science and Information Engineering
National Taiwan University, Taiwan
{d96004, b98013, r98922059, pjcheng}@csie.ntu.edu.tw

***Abstract*—There has been significant recent interest in using the aggregate information from social media sites (*e.g.,* Twitter) to identify influencers. To investigate this issue, one dynamic diversity-dependent algorithm is proposed for detecting the influencers by evaluating the influence of users throughout social networks. Comparative analyses with the existing methods on either synthetic social networks or real Twitter data show that our strategy performs best. It implies that the pattern of the influence propagation should be updated dynamically to reflect the flow of influence spread to better capture the rapidly changing dynamics of microblogs. Our proposed scheme is therefore practical and feasible to be deployed in the real world.**

Keywords—influencers; social diversity; information propagation; Twitter; social networks

I. INTRODUCTION

Owing to the real-time and fast growing features of social network, rendering it become a new media for advertising companies and politicians to spread of information and influence for target marketing and political thinking. It has been believed that target influencers usually lead to a vast propagation of information across the social networks. Identification of influencers by using Twitter follower graph has caught much attention from recent research studies.

Many approaches are proposed to solve this problem by considering different influence measures. However, no researches studied the problem of identification of influencers from the perspective of social diversities of mediators until a recent study [1]. Note that we call these followers who are influenced by the twitterer to share information with their friends as mediators. Their basic idea is that if a user exerts influence on people with high diversity, diverse people would contribute to independent information flows and trigger a large cascade effect to make propagation disperse further and faster. On the contrary, exert influence on people with less diversity might bring a looping effect which renders the information propagation end within the cluster of people without dispersing outside the loop. Furthermore, they believe that a user needs to exert more influence to target his/her mediators with high social diversity than those with low social diversity. Social diversities of user's mediators are thus hypothesized to have a great impact on the influence of a user.

They evaluated the social diversities of mediators by considering either community structure or static influence propagation. However, we believe the influence propagation should be updated dynamically, when calculating the social diversities of mediators, to better reflect the real interaction between users. To the best of our knowledge, there are no previous research studies targeted to issue of influencer detection from this new sight. In this paper, we therefore focus on understanding the impact of social diversities of mediators and dynamic influence propagation on the detection of influencers by aggregating Twitter follower-retweet graph. Our study provides clear evidences that the features of social diversity and dynamic influence propagation are essential factor in detection of influencers.

II. RELATED WORK

Several studies in social network focus on the issue of social influence. Considering the impact of social influence, Agarwal *et al.* [2] addressed the problem on identifying the influential bloggers in blogosphere. Their findings showed that the most influential bloggers were not necessarily the most active bloggers. Ye et al. [3] evaluated different social influences by considering their stabilities, assessments, and correlations. Bakshy et al. [4] reported that weak ties, responsible for the propagation of novel information, may play a more dominant role in the dissemination of information than currently believed.

Several approaches have been proposed to measure influence in social networks (*e.g.,* Twitter). Lee *et al.* [5] proposed a method to find influencers based on both the temporal order of information adoption and the link structure. Kwak *et al.* [6] compared three different measures of influence, namely followers, page-rank and retweets and

978-1-4799-2902-3/13 $31.00 © 2013 IEEE
DOI 10.1109/WI-IAT.2013.58

found that the ranking of the influencers differed by these three measures. And the analysis of network topology showed low reciprocity. Besides, a consistent argument of low reciprocity was found by Cha *et al.* [7]. They also made a comparative analysis of three different measures of influence, namely followers, retweets and user mentions, which were used to evaluate the social influence. Their results presented that the number of followers may not be the best measure of the influence. However, Weng *et al.* [8] contradicted the observation by Cha *et al.* [7] and reported that their study exhibited the phenomenon of high reciprocity which can be explained by homophily. Moreover, based on this finding, they proposed a TwitterRank measure, similar to the work by Haveliwala [9] to find topic-sensitive influential twitterers. Bakshy *et al.* [10] uncovered an interesting observation that the ordinary influencers would be the most cost-effective in many circumstances, only under some circumstances the most influential users are the most cost-effective. Note that the ordinary influencers are defined by them as individuals who exert average or even less-than-average influence.

The above researches have been proposed different measures to find the influencers, but the impact of dynamic influence propagation of a user has yet to be investigated. Therefore, in the following, we attempt to fill this research gap.

III. TWITTER DATASET

For the purpose of this study, a set of Twitter data was prepared over the two weeks, from December 4 2012 to December 17 2012. First, 600 out of top 1000 users and their two levels of followers, covering 5,894 users, were obtained from http://twitaholic.com/ as seeds to construct a retweet graph of 75, 042 tweets and of 400,967 retweets. Second, we crawled all followers of every user. Subsequently, we placed those followers in a queue to be crawled, thereby finding their followers, who were then also placed in the queue, and so on. A follower graph is then constructed. The retweet and follower graph are combined to further construct a retweet-follower graph of 152,119 users, of 169,942 edges, of 192,437 message propagations. To eliminate ambiguity in our analysis, we removed the users who published excessive tweets. As a result, we had a total of 151,305 users for later analysis.

IV. INFLUENCERS IDENTIFICATION

A. Problem Statement

We adopt the diversity-dependent influence (DI) measure for estimating the influence of users. Second, a social network is modeled as a directed graph $G = (V, E)$, where V denotes users, E denotes the relationships of pairs of users. Each directed edge $(u, v) \in E$, where $u, v \in V$, is associated with a transition probability $TP(u, v) \in [0, 1]$, representing the probability of influence propagation that user v is influenced by user u through the edge of (u, v). For example, Twitter retweet-follower graph where vertices are Twitter users and edges represent message propagations between users. That is, if user u post a message (*i.e.,* tweet) and user v (the follower of user u) forwarding the message (*i.e.,* retweet) to his/her follower, user x, an edge between these two corresponding vertices exists, representing an influence propagation from user u to user v. Third, we define a outbound neighbor set of node i as $ON(i)$. For each node $j \in ON(i)$, we calculate the social diversity of j as $SD(j)$ describing to what extent, the diversity of j among $ON(i)$. Given a social network graph G, the problem of identification of influencers is defined as finding influential users according to their corresponding influence scores evaluated via mining social networks. The influencers expect to influence the most diverse users in the social graph. Influence propagates effectively and efficiently through these most diverse users and eventually spreads over the social network. Note that in this paper, we will use the term outbound neighbors and the influenced people interchangeably.

B. Transition Probability

Transition probability is measured as how much attention the twitterer could draw from his/her outbound neighbors? For example, user u_g receive messages from five other users besides u_a while user u_h is the only message receiver of u_a. That is, u_a need to compete with other five inbound neighbors of u_g to have u_g's attention. The transition probability from u_a to u_g is thus measured as 1/6 while that from u_a to u_h is measured as one. Note that, in this case, only one message propagation in each of directed edges. The transition probability $TP(i, j)$ for a directed edge $e(i, j)$ is defined as:

$$TP(i,j) = \frac{mp_{ij}}{\sum_{k \in IN(j)} mp_{kj}} \tag{1}$$

mp_{ij} is number of messages propagated from u_i to u_j, and $\sum_{k \in IN(j)} mp_{kj}$ sums up the number of received messages by u_j. $IN(j)$ is an inbound neighbor set of u_j.

This concept looks similar but different from the work by Weng *et al.* [8]. They consider following relationship, when measuring the transition probability, while we consider the following and retweeting relationship. Note that v follows u do not necessarily mean v will retweet the message from u to his/her follower x. If $e(u,v)$ only has following relationship, we consider these is no influence propagation from u to v.

C. Social Diversity

Social diversity is measure as how diverse a user v is within a cluster, comparing to his/her neighbors. Social diversity of user $v \in ON(u)$, denoted as $SD(v)$, is measured as one divided by the size of cluster which v belongs to, describing to what extent v is diverse from $ON(u)$.

V. DIVERSITY-DEPENDENT ALGORITHMS

To solve the problem of identification of influencers, we consider the transition probability and social diversity to develop a dynamic diversity-dependent approach. Our goal aims to find the influencers by measuring the diversity-dependent influence (DI) score.

The iter-weighted spread scheme is then proposed to consider the diversity between outbound neighbors of user u by computing pair-wise cosine similarity of their corresponding influence spread distribution. If $cos(\theta) \geq \sigma$ (set to be 0.5), we think these two users are similar and has a link to each other. Star clustering algorithm [11] is then employed to $G_\sigma = (V_\sigma, E_\sigma)$ to find clusters. According to these clusters, social diversity of each outbound neighbor is determined. With the transition probability and social diversity, the diversity-dependent influence (DI) measure for each node u, can be calculated by iterations fast and easily as follows:

$$DI(i)^{(t)} = \begin{cases} \frac{1-d}{|V|} + d \times \left(\sum_{v \in ON(i)} \frac{TP(i,v) \times SD(v)}{\sum_{x \in IN(v)} TP(x,v)} \times DI(i)^{(t-1)} + Z \right), \textit{if u with outlink} \\ \frac{1-d}{|V|} + d \times Z \qquad , \textit{otherwise} \end{cases} \quad (2)$$

d is a damping factor (set around 0.85), $Z = \frac{\left(\sum_k DI(k)\right)}{n}$ where $IN(k) = \varnothing$.

To better capture the rapidly changing dynamics of microblogs, we proposed iter-weighted spread which considered dynamic influence spread vectors to measure the diversity of a user. That is, the social diversity of a user is updated iteratively by dynamically updating the vector of influence spread to reflect the flow of influence propagation. $\overrightarrow{IS(u)}$ is updated iteratively as $\overrightarrow{IS(u)}^{(t)} = \overrightarrow{IS(u)}^{(t-1)} + TP(u,v)\overrightarrow{IS(v)}^{(t-1)}$. The complete algorithm for dynamic influence propagation is shown in Algorithm 1.

VI. RESULTES AND DISCUSSION

A. Discrimination of Influencers of Different Influence Levels

In order to evaluate the performance of our approach in discrimination and discovery of the various levels of influencers, we conducted the experiment on a synthetic social network. BA model [12] was then employed to generate a synthetic social network in which node A directed to five independent groups of 1000 nodes while node B only directed to one of them. These two nodes have the same number of out-degree nodes and their out-degree nodes got similar rankings by PageRank. Table I presented that these two nodes have the similar scores and ranks by PageRank since they have the same number and similar ranking of out-degree nodes. Compared to PageRank and previous methods [1], our method showed that there was a significant difference between scores and ranks of these two nodes. Since node A has more diverse out-degree nodes compared to node B, the influence level by node A is thus higher than node B. Our method performed better and successfully discriminate the influence level via considering the social diversities of out-degree nodes (the influenced people).

B. Effect of Dynamic Influence Propagation

A comparative analysis was then performed in order to clarify the performance of our approach. BA model was then employed to generate a synthetic social network in which node A directed to five independent groups of 1000 nodes, denoted as G_0. Additional three social networks, named G_1 to G_3, were subsequently generated by adding 5000, 10000, 15000 links among five groups. The independence level of five groups decreased from G_0 to G_3, that is, the social diversities of nodes within the groups decreased with the adding links. Five groups might eventually be clustered into one by adding more enough links. In a real-world, the denser group means people within the group interact frequently to each other. This clustering effect is often driven by the social similarity, to a certain extent, loses social diversity. The influence scores of node A in our method fall when the social diversities of out-degree nodes decrease (Fig. 1a), but a similar relationship is also presented in PageRank.

Algorithm 1. Dynamic Influence Propagation

Input: the social network $G = (V, E)$, a set of message propagations MP
Output: a set of influence scores DI

```
DI(i)^(0) ← 1/|V| for all i
t = 1
for i = 1 to |V| do
   for j = 1 to |V| do
      Compute the transition probability TP(i, j)
   Compute the influence spread vector IS(i)→
while (Σ_{i∈V} | DI(i)^(t) − DI(i)^(t−1) | ≥ ε) do
   for i = 1 to |V| do
      foreach j, v ∈ ON(i), j ≠ v do
         e_σ(j, v) ← (cos(IS(j)→, IS(v)→) ≥ σ)
      Find clusters C ← StarClustering(G_σ)
      Calculate social diversity SD(j) for all j ∈ ON(i)
      Compute diversity-dependent influence score DI(i)^(t)
     update IS(i)→
   t ← t + 1
```

TABLE I. SCORES AND RANKS FOR NODE A AND NODE B

	PageRank		Prior		Zero-one spread		Weighted spread		Iter-weighted spread	
Node	A	B	A	B	A	B	A	B	A	B
Score	1.099e-03	1.053e-03	1.009e-03	9.571e-04	1.521e-03	9.775e-04	1.270e-03	1.062e-03	1.422e-03	1.075e-03
Rank	117	126	118	133	91	194	110	152	95	153

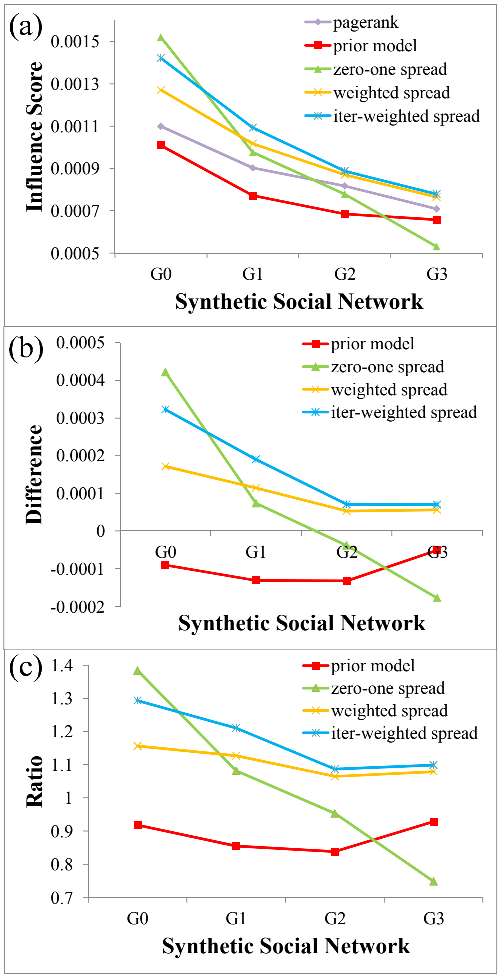

Fig. 1. Influence score falls with social diversities (a) The influence scores for our methods (b) The difference of influence scores (c) The ratio of influence scores.

In order to measure the effect of our method relative to PageRank, computed as either the difference or ratio between the influence scores by our methods and PageRank (Fig. 1bc). The difference in influence scores falls with the decrease of social diversities of out-degree nodes of node A, the relative ratio shows the similar results. This finding suggests that social diversities of out-degree nodes in our method are most likely to make an impact on influence score of node A. The relative impact on the influence is the highest for node A whose out-degree nodes are of the highest social diversities. That is, changes in the social diversities of out-degree nodes could alter the influence level of node A. Our method could better capture the variances in social diversities of out-degree nodes to discover and discriminate the influencers of different influence levels.

Two weighted schemes perform better (*i.e.,* weighted spread and iter-weighted spread), but iter-weighted spread dynamically updates the vector of influence propagation and thus yield the best performance. The zero-one spread treats the number of message propagations as zero or one rather than considers its actual number. Besides, prior model evaluates the social diversities of out-degree nodes based on the network structure and not to consider the distribution of influence propagation. Thus, zero-one spread and prior model might underestimate the effect of social diversities of out-degree nodes and not to reflect the real influence level.

C. Effectiveness of Our Method

In order to demonstrate the effectiveness of our method, we conduct the experiment on a real social network (*i.e.,* Twitter) to make a comparative analysis with PageRank and two centrality-based methods. We use influence spread to measure to what extent a user could spread his/her influence all over the social network. The influence spread, for each ranked user, is thus computed as the total number of influenced people on the whole social network. As shown in Fig. 2, iter-weighted spread outperforms two centrality-based methods and slightly better than PageRank.

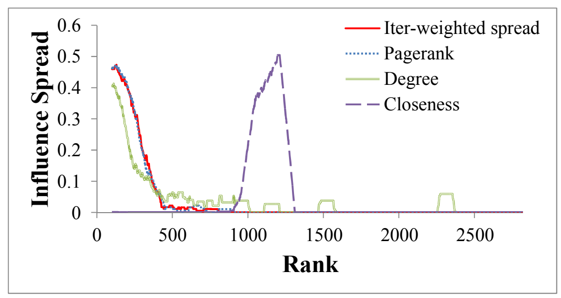

Fig. 2. Influence spread for each ranked user

Take a closer look at our method and PageRank. We investigated the rank correlation between the rankings computed by iter-weighted spread and by PageRank. As shown in Fig. 3, it separated by a correlation coefficient (R=0.7) into three sections. Of the 151,305 nodes, we found that only a small fraction of nodes (top ranked nodes) exhibited significant correlation level. Specifically, the proportions were 0.16% for the first section, 1.70% for the middle section, and 98.13% for the last one. Note that few of users in the last section had followers and thus were excluded from our following analysis.

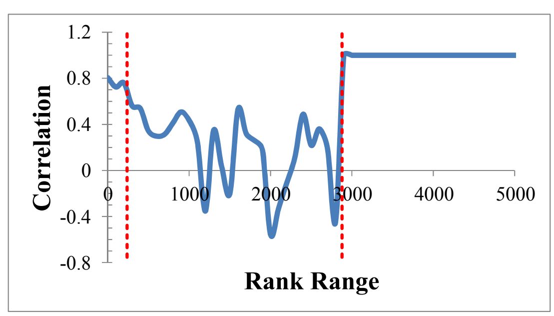

Fig. 3. The correlation of ranks on Twitter data

Since the rankings in the middle section by our method and PageRank reveal a clear low or even negative correlation. Next, we conduct a series of experiment to elucidate this inconsistent observation. We first utilize node-based centrality measures (degree and closeness) to examine the performance of our method. Our results indicated that the average scores evaluated by these two measures are the highest for the first section while the lowest for the last section. Specifically, average degree centralities were 596.2249 for the first section, 7.1477 for the middle section and 0.0209 for the last one. Average closeness centralities were 0.3376, 0.2186 and 0.0003, respectively. Note that node-based centrality is defined to measure the importance of a node in the network. A node with high centrality score is usually considered more highly influential than other nodes in the network [13]. According to the results, it confirmed that our method performs well on the real social networks.

Second, as shown in Fig. 4ab, we can find that higher ranked users have more out-degree nodes as well as the in-degree nodes. It suggests that higher influencers are often more popular and easy to draw more attention from others. This finding is supported by the work of Yeruva et al. [14], who found that the number of leaders in each level increases while the number of followers decreases. In addition, the study by Pal and Counts [15] presented that the name value of the popular users plays an important role in their authority. Their studies further offered an explanation why follower-based PageRank performs well in the first section while get worse performance in the middle section. On the other hand, these higher influencers are also gregariousness and often have more friends, comparing to the lower ranked users. Besides, the relative ratio shows the similar results (Fig. 4c).

At last, we compare two groups of users. Users with ranks evaluated by iter-weighted spread better than those by PageRank are classified into group A. By contrast, group B represents the users with ranks by iter-weighted worse than those by PageRank. Experimental results report that the node centrality measures, which determine how influential a person is within a social network, for group A are all larger than those for group B. Specifically, the degree centralities were 58.9306 for group A and 15.3585 for group B. The closeness centralities were 0.4909 for group A and 0.0446 for group B. From the above comparative analyses, we could conclude that our proposed method, iter-weighted spread, performs better than PageRank.

VII. Case Study

Table II presents the list of top 20 influencers as recommended by our method (i.e., iter-weighted spread) and those by PageRank and two rankings are found to be similar. Take a closer look, we observe that users identified by our method are mostly classified into entertainment or marketing. People belong to these two classes aim to target diverse folks from all walks of life. Reciprocally, people outside the classes often pay a close attention to these classes. In a word, to a certain extent, the top users often have more and higher diverse outbound neighbors. This phenomenon might be partly driven by their name value [15]. It is worth to note that PageRank offers a similar ranking for top users to that by our method while performs worse for users of middle ranks.

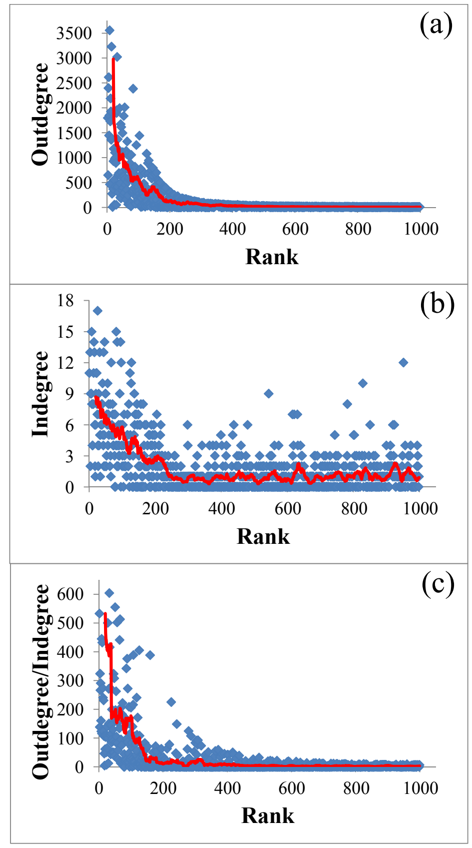

Fig. 4. Users with higher rank are more popular and gregarious. (a) Out-degree (b) In-degree (c) The multiplicative impact of out-degree for top 1000 users

TABLE II. List of top 20 influencers as computed by our method and PageRank

Rank	Pagerank	Iter-weighted spread	Description	Var.
1	SonReflexiones	SonReflexiones	Government	0
2	comedytexts	comedytexts	Advertising	0
3	TwitsImagenes	LogicaDeMujeres	Organization	-1
4	LogicaDeMujeres	LaVidaEnLetras	Entertainment	-5
5	MundoDeLaRisa	FIirting	Advertising	-3
6	TrueStoryPage	TheFunnyTeens	Advertising	-9
7	ItsLifeFact	ItsLifeFact	Marketing	0
8	FIirting	TrueStoryPage	Entertainment	2
9	LaVidaEnLetras	TwitsImagenes	Entertainment	6
10	Hilarious_Dude	MiFraseTipica	Advertising	-3
11	MenHumor	MenHumor	Entertainment	0
12	TheFunnyTeens	TheFunnySayings	Entertainment	0
13	SimpleLoveTweet	ItsFunnyLife	Entertainment	new
14	FunnyOrTruth	SimpleLoveTweet	Marketing	-2
15	GooglePics	Hilarious_Dude	Movie	5
16	ithinkthatway	TheGoogleImages	Entertainment	2
17	chistea	MundoDeLaRisa	Entertainment	12
18	ItsFunnyLife	EsDeNoviosQue	Entertainment	new
19	TheOfficialTed	CAJADEVERDADES	Advertising	new
20	teenagernotes	iMundoLoco	Entertainment	new

Kwak et al. [6] confirmed this observation, they ranked users by three different measures of influence, number of followers, PageRank and number of retweets. They found that the first two rankings are similar but differ from ranking by retweets. Their findings suggest that followers might make an impact on PageRank rather than retweets. Consequently, it could better elaborate why our method, based on the retweet and follower graph, yields a better performance than PageRank.

To give a sense of how well our approach works, we further compare the users with middle influence level. The difference between the ranks computed by our method and by PageRank for celebrity was measured, as shown in Table III. Since both Robbie Williams and Emma Roberts are popular singers, they might catch attention from people from various walks of life. Thus, as our expectation, their ranks progressed a lot detected by our method compared to PageRank. On the contrary, AKON is a hip hop singer and he might only influence the people who like hip hop, resulting in a worse rank. In particular, Dr. Sanjay Gupta is a neurosurgeon and he might merely appeal to the persons of medical background or of interest in medicine. His rank, therefore, dropped a lot to response to the lower effect of social diversities from his influenced persons on his influence. Also, Ryan Sheckler and Blake Griffin got worse ranks because they are professional players. Ryan Sheckler would only attract people who like playing skateboard while Black Griffin would only attract people who like playing basketball. As for the users in the first section, such as PSY and Recep Tayyip Erdoğan, their ranks are similar in our method and PageRank. PSY is a global pop singer and Recep Tayyip Erdoğan is a politician, their influence could reach a wide range of persons.

According to above observations and findings, our proposed method, iter-weighted spread, is demonstrated to be dynamically sensitive to the change of social diversities evaluated from the influenced people and make it an attractive alternative in identification of influencers to better capture the rapidly changing dynamics of microblogs.

VIII. CONCLUSION

In this paper, we address the problem of identification of influencers in social networks by introducing the concept of social diversities of the influenced people and dynamic influence propagation. We extend the previous work and then proposed a dynamic diversity-dependent scheme to identify influencers via measuring the influence scores of users. It differs in the calculation of social diversity. In the previous methods, they considered the social diversity from either community structure or static influence propagation. Yet, we believe influence propagation should be updated dynamically to reflect the real interaction between users and the iter-weighted spread is thus proposed.

Comparative analyses on synthetic social networks suggest that the social diversity of the influenced people may play an important role in the identification of influencers. Comparative analysis between our method and the previous approaches shows that iter-weighted step is superior to the other methods. It implies that the pattern of the influence propagation should be updated dynamically to reflect the flow of influence propagation to better capture the rapidly changing dynamics of microblogs. We also apply our method to Twitter data. Our results show that our strategies perform well in detecting influencers. Our proposed scheme is therefore practical and feasible to be deployed in the real world.

TABLE III. VARIATION OF THE RANKS FOR CELEBRITY

Celebrity				
Screen name	Description	PageRank	Iter-weighted Spread	Var.
Lea Michele	American actress and singer	321	316	-5
Dianna Agron	An American actress, singer, dancer	366	351	-15
PSY	South Korean singer	503	429	-74
Recep Tayyip Erdoğan	Prime minister of Turkey	459	465	6
Phillip Schofield	English broadcaster and television personality	484	477	-7
Adam Schefter	American sports writer and television analyst	532	535	3
Mark Wright	An English reality television personality and radio presenter	606	639	33
Hugh Hefner	American magazine publisher	668	696	28
Luisana Lopilato	An Argentine actress, model and singer	878	857	-21
Linkin Park	American rock band	850	881	31
Gusttavo Lima	A Brazilian singer of the sertaneja(Brazilian country) genre	1097	1196	99
Samuel L. Jackson	American film and television actor and film producer	1183	1280	97
Ian Poulter	English professional golfer	1184	1281	97
Kendra Wilkinson	American television personality	1358	1367	9
Robbie Williams	English pop singer	1564	1469	-95
Ryan Sheckler	American professional skateboarder	1381	1484	103
Shashi Tharoor	The Indian minister of state for human resource development	1472	1567	95
Yalın	A Turkish pop singer and songwriter	1932	2117	185
Blake Griffin	American professional basketball player	1935	2114	179
Ozzy Osbourne	English heavy metal vocalist and songwriter	1997	2184	187
David Pogue	American technology writer and TV science presenter	2002	2179	177
Novak Djokovic	Professional tennis player	3360	3442	82
Emma Roberts	American actress, model and singer	3717	3635	-82
Dr. Sanjay Gupta	American neurosurgeon	3631	3713	82
AKON	R&B and hip hop singer	5619	5632	13

References

[1] P.-Y. Huang, H.-Y. Liu, C.-T. Lin, and P.-J. Cheng, "A Diversity-dependent Measure for Discovering Influencers in Social Networks," in *Information Retrieval Technology, 9th Asia Information Retrieval Societies Conference*, R. E. Banchs, F. Silvestri, T.-Y. Liu, M. Zhang, S. Gao, and L. Jun, Eds., ed, 2013.

[2] N. Agarwal, H. Liu, L. Tang, and P. S. Yu, "Identifying the influential bloggers in a community," presented at the Proceedings of the 2008 International Conference on Web Search and Data Mining, Palo Alto, California, USA, 2008.

[3] S. Ye and S. F. Wu, "Measuring message propagation and social influence on Twitter.com," presented at the Proceedings of the Second international conference on Social informatics, Laxenburg, Austria, 2010.

[4] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of social networks in information diffusion," presented at the Proceedings of the 21st international conference on World Wide Web, Lyon, France, 2012.

[5] C. Lee, H. Kwak, H. Park, and S. Moon, "Finding influentials based on the temporal order of information adoption in twitter," presented at the Proceedings of the 19th international conference on World wide web, Raleigh, North Carolina, USA, 2010.

[6] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?," presented at the Proceedings of the 19th international conference on World wide web, Raleigh, North Carolina, USA, 2010.

[7] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in Twitter: The million follower fallacy," presented at the ICWSM '10: Proceedings of international AAAI Conference on Weblogs and Social, 2010.

[8] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "TwitterRank: finding topic-sensitive influential twitterers," presented at the Proceedings of the third ACM international conference on Web search and data mining, New York, New York, USA, 2010.

[9] T. H. Haveliwala, "Topic-sensitive PageRank," presented at the Proceedings of the 11th international conference on World Wide Web, Honolulu, Hawaii, USA, 2002.

[10] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, "Everyone's an influencer: quantifying influence on twitter," presented at the Proceedings of the fourth ACM international conference on Web search and data mining, Hong Kong, China, 2011.

[11] J. A. Aslam, E. Pelekhov, and D. Rus, "The Star Clustering Algorithm For Static And Dynamic Information Organization," *Journal of Graph Algorithms and Applications,* vol. 8, pp. pp. 95-129, 2004.

[12] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics,* vol. 74, pp. 47-97, 01/30/ 2002.

[13] J. Sun and J. Tang, "A Survey of Models and Algorithms for Social Influence Analysis," in *Social Network Data Analytics*, C. C. Aggarwal, Ed., ed: Springer, 2011, pp. 177-214.

[14] S. Yeruva, B. S. Ilango, and Y. S. Reddy, "Article: A study on Social Influence Analysis in Social Networks," *IJCA Proceedings on Amrita International Conference of Women in Computing - 2013,* vol. AICWIC, pp. 6-12, January 2013.

[15] A. Pal and S. Counts, "Identifying topical authorities in microblogs," presented at the Proceedings of the fourth ACM international conference on Web search and data mining, Hong Kong, China, 2011.