Determining Influential Users with Supervised Random Walks

Georgios Katsimpras
NCSR Demokritos
Ag. Paraskevi
Athens, Greece
gkatsibras@iit.demokritos.gr

Dimitrios Vogiatzis
NCSR Demokritos
Ag. Paraskevi
Athens, Greece
dimitrv@iit.demokritos.gr

Georgios Paliouras
NCSR Demokritos
Ag. Paraskevi
Athens, Greece
paliourg@iit.demokritos.gr

ABSTRACT

The emergence of social media and the enormous growth of social networks have initiated a great amount of research in social influence analysis. In this regard, many approaches take into account only structural information while a few have also incorporated content. In this study we propose a new method to rank users according to their topic-sensitive influence which utilizes a priori information by employing supervised random walks. We explore the use of supervision in a PageRank-like random walk while also exploiting textual information from the available content. We perform a set of experiments on Twitter datasets and evaluate our findings.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data Mining; H.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

Keywords

social influence analysis, supervised random walks, pagerank

1. INTRODUCTION

At recent years there has been an immense growth of on-line social communities and an overwhelming flow of information through them. Participating in these social networks not only enables users to share information or opinions but also offers more opportunities to develop social ties. Consequently, these interactions determine the circulation of attention through the network and hence, designate a user's relationship with others. Interestingly, these social bonds play an essential role in many sociological and information research studies. The former try to explain why and how these ties are formed while the latter to empirically provide some evidence. In this regard, the concept of influence has been long examined and its role in the dissemination of information has been pointed out in many studies [5]. However, despite the enormous amount of available on-line data, determining influential users is a challenging task.

Identifying influence denotes the need to comprehend and reveal a change in the social world. Such a change is a result of human choices and social forces which cannot be reproduced within a lab experiment or formalized in a mathematical equation. As chaotic the real world is, it is even more difficult to conceive the factors that can be used to quantify influence. An individual's choice of relations is heavily constrained by other aspects of his or her life, such as geographical location, choice of occupation, place of work, and so on. Moreover, social phenomena such as homophily, the tendency of individuals to form ties with similar others [10], constitute an even more complex statistical problem.

Nevertheless, the rise of social platforms like Twitter and Facebook, naturally lead to investigate the potential of influence in certain scenarios such as opinion propagation and viral marketing. For journalists, companies and other public entities determining and interacting with influential individuals is elemental. This need has initiated a considerable amount of research regarding the quantification of influence. As influence is constitutionally ambiguous, most studies give their own definition of influence and in regards to the context of their research direction. However, a human can clearly distinguish which users are influential to them and which are not. A simple scenario could be of a journalist maintaining a list of influential and a list of non-truthful users. Motivated by this fact, we believe that such a priori information can be employed to discover similar influential users.

In the present paper we are trying to tackle the problem of identifying influential users on a on-line social community like Twitter. Particularly, we introduce a new method which measures influence of users, based on the existence of a priori information of influential users. It utilizes structural information and the textual content affiliated to each node in network to measure topic-sensitive influence of nodes. Our method is based on a technique called supervised random walks and which was originally introduced in [1] to address the link prediction problem. In addition, we compare different measures of influence on their ability to rank users w.r.t. a topic. We conducted experiments on three different real-world datasets and illustrate the effectiveness of our approach compared to previous work.

The rest of the paper is organized as follows: Section 2 briefly summarizes related literature. In Section 3 we give a detailed analysis of our approach while in Section 4 we present our experimental findings. Finally, Section 5 con-

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2742472.

cludes with a discussion on our findings and directions for further research.

2. RELATED WORK

Structural approaches. A great number of studies has been conducted in the domain of social network analysis in recognizing important users in a social network. One of the most studied ranking algorithm in graphs and particular in social networks is PageRank [13]. A numerical value is assigned to each node that expresses its importance and its computation depends solely in the structure of the network. Many PageRank variations were proposed since its first appearance. The method proposed in [19], called TunkRank, is a PageRank variation adapted to the specific conditions of the social network of Twitter. The intuition of their approach is that the more followers you have the more probable your tweet is going to be re-tweeted and thus spread to the network.

Another firmly relevant approach to PageRank is [16]. Romero et. al see influence as a twofold quantity where a user can be described by both the number of people they influence as well as their passivity, meaning the number of people who are not influenced by. In this regard, influence depends on the passivity of followers and passivity depends on the influence of the followees. In [4], they made a comparison of some simple measures of influence, expressly number of followers, number of re-tweets, and number of mentions. They showed that each measure alone cannot be a significant indicator of how influential a user is. However they suggest that a user should follow a particular pattern or effort in order to maintain their ability to influence other users. Also [11] performed experiments with the rankings of the three following measures: number of followers, PageRank in the following/follower network and number of re-tweets. They found the first two ranking methods to be highly correlated while the number of re-tweets does not reflect accurately the influence of users. In [6] they argue that the use of the structure of a network is not mandatory in order to identify influential users. They propose a new influence metric, $Velocity$, that depends only on the available information collected from the Twitter streaming API, namely the number of mentions and the number of followers, and can be computed in near-real time.

Topic-based approaches. Although link-based techniques have been proved to be effective in influence ranking, there exists a great number of studies in the literature that try to combine structural information with content. An early approach towards that direction is Topic-Sensitive PageRank (TSPR) [8]. As the name suggests, it is a PageRank variant which incorporates topic-specific information to calculate per-topic PageRank scores. To achieve that it replaces the classic PageRank's teleport vector with predefined topic-specific ones in a separate preprocessing stage. A more recent study [21] tries to bias the random walk towards similar users according to a topic. As an alternative to TSPR, it uses topic modeling [3] to extract topics from a users content. The distribution similarity on these topics between users is used then to compute the influence scores. In [18] they perform topical-level influence propagation (TAP) by incorporating a) an existing network structure and b) topic distributions on all nodes into a topical factor graph (TFG). Pal et al. in [14] make use of a list of features from Twitter graphs and perform clustering to identify topical authorities. On the other hand, Hu et al. in [9], utilize Topical Authority Propagation via re-tweeting to identify topical authorities in Twitter. The intuition behind their method is that a user is more likely to be an authority on a topic if they get retweeted by another topical authority. Ghosh et al. in [7] try to infer topical experts by leveraging crowd-sourced information from Twitter Lists. A very recent study [2] integrates topic modeling and influence analysis in the same model. Their proposed method called FLDA, it is a more complex mixture model than LDA which integrates probability distributions over topics with influence scores.

The work presented here builds on this prior knowledge but also lends many basic aspects from random walk theory and especially from a supervised version which was presented by Backstrom and Leskovec in [1]. While their technique was developed to address the link prediction problem, we propose a new method based on the original to deal with the identification of influential users. Our contribution is a method that in a supervised manner learns how to bias PageRank-like random walk so that it is headed to a more topic-sensitive influential nodes. Readers interested in a more detailed view on influence identification techniques may refer to surveys papers [15, 20] and their references.

3. TOPIC-SENSITIVE SUPERVISED RANDOM WALKS

In general the purpose of topic-based social influence analysis is to capture the following information: nodes' topic distributions, similarity between nodes, and network structure [18]. In order to assimilate this information into one model we propose Topic-Sensitive Supervised Random Walks (TS-SRW), a supervised algorithm for identifying topic-sensitive influential users. Our main idea is that given a graph and its content we would like to assign a score to each node which would represent the influence of that node in a specific topic derived from its content. Since PageRank has been proved to be significantly effective in graph ranking we adopt a similar approach. Nonetheless, we incorporate at the same time valuable node and edge features together with the structure and the content of the network so as to bias the random walk to step into more influential users.

Expressly, we can divide our method in three phases as depicted in Fig.1: i) topic extraction, ii) parameter learning and iii) PageRank-like random walk. The general setting of the proposed framework requires to extract the topics that users are interested in. Hereof, we handle this issue by applying LDA[3], a widely used topic modeling method to automatically infuse topics. Moreover, assuming that we are given a set of influential nodes $i_1, i_2, \ldots, i_n \in I$ and the topic distributions, we try to bias the transition probability of each edge with the purpose that the random walk will visit nodes from set I more often. To achieve this allocation we intend to learn a function that will specify the transition probability by taking into account the node and edge features. So, the task in the second phase is to learn a set of parameters π of an objective function that assigns a transition probability to each edge. Then, these parameters will be used by the random walk to rank nodes accordingly in the last stage. In the following we discuss each part of our approach in depth. We begin by giving some notation involved in our analysis and then examine each step in detail.

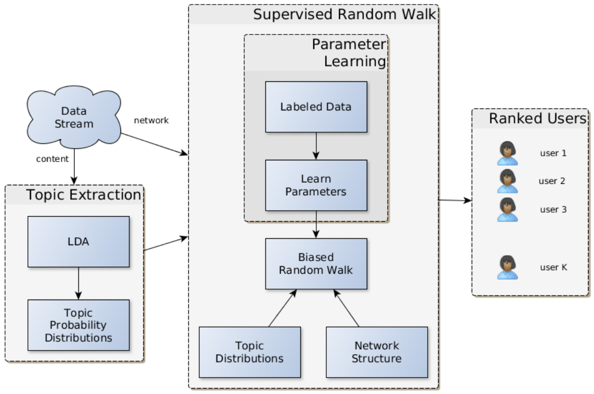

Figure 1. Flow of the proposed technique

3.1 Preliminaries

We consider a social network represented as a directed graph G = *(V, E)* where vertices correspond to users and edges to social interaction between them such that edge *(u, v)* implies that u *interacts with* v. For each edge we built a feature vector ϕ_{uv}which describes the relationship between the nodes that form the edge as well as the nodes themselves. Moreover, each node is assigned a topic vector t_u which contains the topic probability distributions for that specific node. Then, for each edge we can compute a topic similarity score s_{uv} which measures how related are the nodes that form the edge w.r.t a topic.

As for the learning stage, we define a set of influential nodes I=$\{i_1, i_2, \ldots, i_n\}$ and a set of non-influential nodes N=$\{n_1, n_2, \ldots, n_n\}$. We want the random walk to follow the nodes contained in I and not in N. Consequently we can assign each edge a weight $w_{uv} = f_\pi(\phi_{uv}) \times s_{uv}$ where f is a function which calculates the weight of the edge w.r.t parameters π. Supposedly, these edge weights will be used to guide the random walk. Therefore, the rankings of the nodes depends on weights and hence, on parameters π, but also on the topical similarity s_{uv} of the nodes.

3.2 Topic Extraction

To address the issue of topic extraction, we apply the Latent Dirichlet Allocation (LDA) model [3] which automatically and in a unsupervised way can discover a set of topics from large collections. LDA does not describe semantically the topics extracted, however represents each topic as a probability distribution over words and each document as a distribution over topics.

In this regard, the output of the LDA model will provide: θ_K, a K x D vector which contains the distribution of topics K for each document $d \in D$ where D is the set of available documents. To accommodate it to our needs we make the following assumption: each document in our case represents the total number of a user's posts. So for each user we are going to have a topic vector t_u which holds the probability distribution of the user's interest over the topics K. More precisely, each element of t_u, for instance t_u^i, encloses the probability that the user u is related to topic i.

3.2.1 *Topic Similarity.*

As we want to find influential users according to a topic, we measure the similarity between them w.r.t. the topics. More formally, given the topic distributions for all users, $t_u \forall u \in V$, we can compute topical similarity between users as follows:

$$topicalSimilarity = s_{uv} = \frac{1}{2} \times D_{KL}(t_u||M) + \frac{1}{2} \times D_{KL}(t_v||M). \tag{1}$$

where D_{KL} is the $KL - divergence$, which is defined as $D_{KL}(t_u||t_v) = \sum_i ln(\frac{t_u^i}{t_v^i})t_u^i$, and $M = \frac{1}{2}(t_u + t_v)$. The Eq.1 is in point of fact the $Jensen - Shannon\ divergence$ which is a popular method of measuring the similarity between two probability distributions [12].

3.3 Learning Parameters

For the learning phase of our method, we follow a similar approach to [1]. We are dealing with a optimization problem where we want the objective function to learn a set of parameters. Specifically, we target on learning the optimal parameters π so as to be used on calculating the edge weights that will hopefully lead the random walk to pick influential nodes. The optimization problem then, can be formulated as:

$$\underset{\pi}{\text{minimize}} \quad f(\pi) = ||\pi||^2 + \lambda \sum_{i \in I, n \in N} h(p_i - p_n) \ . \tag{2}$$

where λ is a regularization parameter, h is a non-negative loss function and p_u with $u \in \{I, N\}$ is the pagerank score for u. To solve this optimization problem we follow the same steps as in [1]. We first derive the gradient of $f(\pi)$ with respect to π, and then use a gradient based optimization method (L-BFGS) to find π that minimize $f(\pi)$. As loss function h we use Wilcoxon-Mann-Whitney loss function $h(x) = \frac{1}{1+\exp(-x/b)}$ and as edge weight function we use the logistic function $w_{uv} = \frac{1}{1+\exp(-\phi_{uv} \cdot \pi)}$. Moreover we choose $\lambda = 1$[1]. For a more detailed analysis we advise the reader to go through the aforementioned publication.

3.4 Topic-Specific Random Walk

Similar to PageRank, we perform a random walk on the directed graph G. The random walker follows the edges with a certain probability which resembles the transition probability from one node to another. This transition, however, is biased by both the edge weight and the topical similarity of the nodes. Consequently, we can define the transition matrix T and the transition probability P_T from node u to node v as:

$$P_T(u,v) = \begin{cases} \frac{w_{uv}}{\sum_{j \in Neighbors_u} w_{uj}} & \text{if } (u,v) \text{ exists} \\ 0 & \text{otherwise} \end{cases} \ . \tag{3}$$

and

$$T = \{P_T(u,v) \forall u, v \in V\} \ . \tag{4}$$

The intuition here is that the higher the topical similarity s and the edge weight w, the higher the transition probability and thus will lead to higher influential nodes. Furthermore, we assign a restart probability γ for the random walk i.e. the probability the random walk to jump back to start node n, and thus the final transition probability will be:

$$P'_T(u,v) = (1-\gamma) * P_T(u,v) + \gamma * (v = n) \ . \tag{5}$$

We conclude in the following algorithm which iteratively calculates the PageRank-like scores for each node in the net-

[1]as proposed in [1]

work. Convergence of the algorithm is similar to those of power-iteration.

Algorithm 1: Topic-Specific Supervised Random Walk Algorithm

Data: V, E, I, N, ϕ_{uv}, objective function f
Result: learned parameters π, node rankings ρ
begin
  Initialization;
  for $u \in V$ **do**
    $p_u = \frac{1}{|V|}$;
    $t_u = LDA()$
  for $(u, v) \in E$ **do**
    s_{uv}= $topicSimilarity(t_u,t_v)$
  Learning Stage;
  $\pi = L\!-\!BFGS(f,\phi_{uv}, I, N)$;
  return π
  Biased Random Walk;
  repeat
    for $u \in V$ **do**
      $p_u^t = \sum_{i \in Neighbors_u} p_u^{t-1} T_{iu}$;
  until *converged*;
  return $\rho = p_u \forall u \in V$

4. EXPERIMENTS

4.1 Experimental Setup

Data description. In order to conduct our experiments we gathered data from Twitter using the Twitter Streaming API. Specifically, we aggregated tweets for three different topics; i) politics, ii) World Cup 2014 and iii) Scala. Apart from these, we also used the Snow 2014 Data Challenge test dataset[2]. As Twitter Streaming API allows to collect data w.r.t. some predefined keywords, we represented each topic as a set of keywords. For each dataset we built an interaction network based on the mentions of the users. Presumably, an edge (u, v) in this network implies that u *mentions* v. The details of each dataset can be viewed in Table 1.

Training data. In the absence of ground truth data, we decided to feed our learning algorithm with labeled data that were produced with the following way. We run the *PageRank* algorithm on the retrieved mention network and obtain a *PageRank* score per node. From this ranked list we use the top 200 as our set of influential nodes I and the last 200 as our set of non-influential nodes N. Moreover, each edge (u, v) of the network is described with the following six features for each node that form them[3]: a) number of followers, b) number of tweets, c) number of friends, d) number of favorites, e) number of lists is listed in and f) if is verified. For the experiments presented in this study, we trained our algorithm with the sets I, N that were obtained from Snow dataset. Then, we use the learned parameters π_{snow} to evaluate our method on the rest of the datasets.

Evaluation metrics. Evaluating a ranking is a very subjective/ambiguous procedure even for humans, unless there is a ground truth available. In our case, we do not consider to have a objective ranking to compare it with our results.

[2]publicly available at `http://figshare.com/articles/SNOW_2014_Data_Challenge/1003755`
[3]twelve features in total

Table 1. Details of each Twitter dataset

Dataset	#tweets	#nodes	#edges	keywords
Snow	1.089.909	560.009	963.685	Syria, terror, Ukraine, bitcoin
World Cup	264.616	164.651	257.854	worldcup14
Politics	249.060	122.147	202.765	politics
Scala	10.423	4.469	8.157	scala

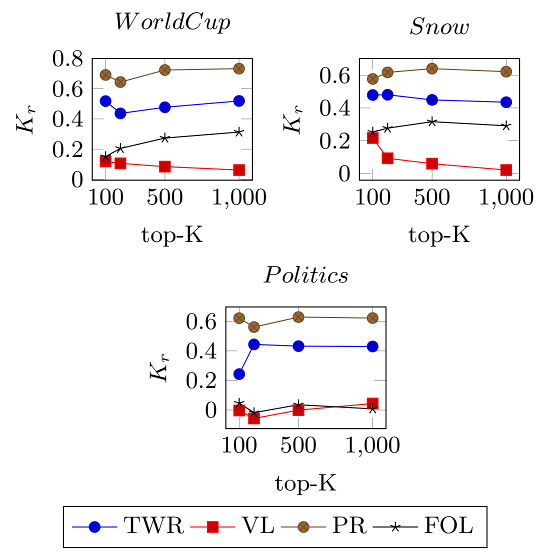

Figure 2. The Kendall Tau Correlation K_r of TS-SRW against all other metrics for the first top-K ranked results

Except for Scala dataset where we obtained a top-35 influential users list according to [17]. Consequently, we study the correlation between the rank lists computed by the different algorithms. For this purpose we consider *Kendall Tau Correlation* measure K_r, which is a measure of how similar two rankings are. The output of K_r is a value in the range of $[-1, 1]$, where a value close to 1 indicates high correlation and -1 that one ranking is the reversed of the other.

4.2 Comparison of different methods

Here we perform a comparison study of the proposed method, abbreviated as TS-SRW with other related metrics. These include: *TwitterRank*(TWR), *Velocity*(VL), *PageRank*(PR), and *number of followers*(FOL).

We setup the experiments as follows. We run the different techniques on the mention graph obtained from the collected tweets for each dataset. For TWR and TS-SRW, we use LDA, assuming that a document corresponds to the total number of tweets a user published in the corresponding dataset. As Dirichlet hyper-parameters we choose $a = 0.5, b = 0.5$ and set the number of topics to $T = 10$. In addition, for PR, TWR and TS-SRW we set the restart probability to $\gamma = 0.85$.

4.2.1 Correlation.

Table 2 lists the correlation values between the rankings that where produced by the aforementioned metrics. It is observed that PR and TS-SRW have the highest correlation

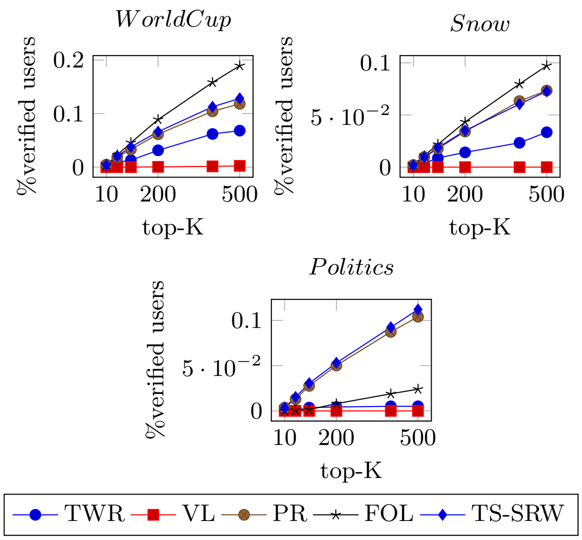

Figure 3. Percentage of verified users in top-K ranked results

which is expected as PageRank was used to train the TS-SRW model.

A closer look on the correlation between TS-SRW and the other metrics is depicted in Figure 2. In these diagrams we illustrate the correlation between TS-SRW and the rest of the methods for the first top-K results of their rankings. In contrast to the results presented in Table 2, where the whole ranked list is assumed, we observe different correlations between rankings.

4.2.2 *Percentage of verified users.*

A feature that is offered by Twitter Streaming API is the characterization of a user as verified or not. In Twitter, verification is used to establish authenticity of identities of key individuals or brands. More precisely, of highly sought users in music, acting, fashion, politics, journalism, sports and other key interest areas. In other words, these users acquire a lot of attention and thus can be seen as influential users in their corresponding area. Although there is not a direct connection between verification and influence, this feature can give us information to perform relative comparisons.

We conduct the following experiment. For each dataset, we find the total number of verified users. Then, we compute the percentage of verified users that is contained in the top-K results of each method. The results are reported in Figure 3. As shown, FOL identifies a larger percentage of verified users in all cases except in Politics dataset. Closely follows TS-SRW and PR with the first being slightly better. Specifically, TS-SWR performs best in Politics dataset. TWR and VL report the lowest percentages.

4.2.3 *Recommendation Task.*

Next we compare the predictive performance of the different methods on a recommendation task. The task is conducted in the same pattern as in [21] but using the mention relationship graph. In this regard, L is the set of existing mention relationships in a dataset, U is the set of randomly chosen users that s_0 does not mention. For our experiments we generate L by randomly picking a mention from the mention graph. We choose $|L| = 40$ and $|U| = 20$. Ten rounds of evaluation are performed for each experiment. The averaged quality over all evaluation rounds is depicted in Figure 4. TS-SRW outperforms all other methods in 2 out of 3 cases. The improvement may not seem significant however it states that our method can provide better recommendations.

Table 2. Rankings correlation of different methods

Dataset		TWR	VL	PR	FOL	TS-SRW
WorldCup	TWR	-	-0.28	0.24	0.11	0.29
	VL	-0.28	-	0.06	-0.06	0.11
	PR	0.24	0.06	-	-0.06	0.46
	FOL	0.11	-0.06	-0.06	-	-0.02
	TS-SRW	0.29	0.11	0.46	-0.02	-
Snow	TWR	-	0.09	0.18	0.11	0.32
	VL	0.09	-	0.06	0.27	0.06
	PR	0.18	0.06	-	0.11	0.73
	FOL	0.11	0.27	0.11	-	0.02
	TS-SRW	0.32	0.06	0.73	0.02	-
Politics	TWR	-	0.06	0.11	0.06	0.24
	VL	0.06	-	-0.28	0.14	0.15
	PR	0.06	-0.28	-	0.11	0.6
	FOL	0.11	0.14	0.06	-	0.06
	TS-SRW	0.24	0.15	0.6	0.06	-

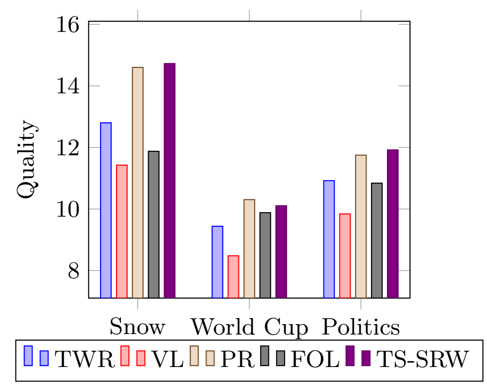

Figure 4. Averaged quality of each ranking method.

4.2.4 *Scala Dataset*

In [17] they have managed to manually produce a ranking of influential users of the Scala community in Twitter. As already mentioned, we gathered data from twitter using the keyword *scala*. So, our next evaluation task is to compare the rankings produced by the different algorithms against the manually produced one. To be precise, we obtained the first 35 users from the annotated ranking and then compute the Kendall tau correlation of this ranking against the rankings produced by the other metrics. We can observe the results in Figure 5. TS-SRW clearly outperforms all other methods.

5. CONCLUSION

This paper addresses the problem of determining topic-sensitive influential users in social networks. To model topic influence we introduce a supervised algorithm, called Topic-Specific Supervised Random Walks.TS-SRW assimilates at the same time valuable structural and textual information

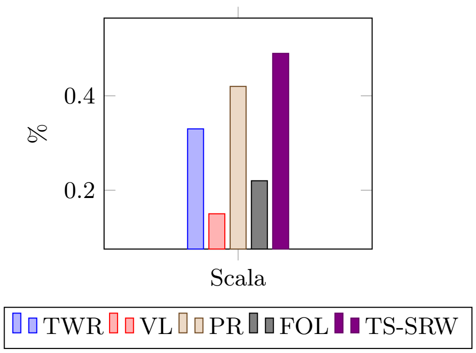

Figure 5. The Kendall Tau correlation of each method with the manually produced ranking

into one unified model to measure the topic-sensitive influence of a user. Through experiments in three real-world Twitter datasets, is shown how the proposed TS-SRW method is correlated with different metrics. In addition, we demonstrate that TS-SRW can produce better rankings in different scenarios. In the absence of labeled data, an extension of the proposed technique to a semi-supervised version would be an interesting future work.

6. ACKNOWLEDGMENTS.

The work described in this article is related to the research project REVEAL (REVEALing hidden concepts in Social Media), Grant agreement no: 610928 which is funded by the European Commission, under the FP7 programme.

7. REFERENCES

[1] L. Backstrom and J. Leskovec. Supervised Random Walks: Predicting and Recommending Links in Social Networks. Nov. 2010.

[2] B. Bi, Y. Tian, Y. Sismanis, A. Balmin, and J. Cho. Scalable topic-specific influence analysis on microblogs. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, WSDM '14, pages 513–522, New York, NY, USA, 2014. ACM.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, Mar. 2003.

[4] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, May 2010.

[5] R. B. Cialdini and M. R. Trost. *Social influence: Social norms, conformity, and compliance*, volume 2, pages 151–192. McGraw-Hill., 1998.

[6] D. Gayo-Avello, D. J. Brenes, D. Fernández-Fernández, M. E. Fernández-Menéndez, and R. García-Suárez. De retibus socialibus et legibus momenti. *CoRR*, abs/1012.2057, 2010.

[7] S. Ghosh, N. K. Sharma, F. Benevenuto, N. Ganguly, and P. K. Gummadi. Cognos: crowdsourcing search for topic experts in microblogs. In W. R. Hersh, J. Callan, Y. Maarek, and M. Sanderson, editors, *SIGIR*, pages 575–590. ACM, 2012.

[8] T. H. Haveliwala. Topic-sensitive PageRank. In *Proceedings of the 11th international conference on World Wide Web*, WWW '02, pages 517–526, New York, NY, USA, 2002. ACM.

[9] J. Hu, Y. Fang, and A. Godavarthy. Topical authority propagation on microblogs. In Q. He, A. Iyengar, W. Nejdl, J. Pei, and R. Rastogi, editors, *CIKM*, pages 1901–1904. ACM, 2013.

[10] G. Kossinets and D. J. Watts. Origins of Homophily in an Evolving Social Network. *The American Journal of Sociology*, 115(2), 2009.

[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*, WWW '10, pages 591–600, New York, NY, USA, 2010. ACM.

[12] C. D. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing*. The MIT Press, 1 edition, June 1999.

[13] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford Digital Library Technologies Project, 1998.

[14] A. Pal and S. Counts. Identifying topical authorities in microblogs. In *Proceedings of the fourth ACM international conference on Web search and data mining*, WSDM '11, pages 45–54, New York, NY, USA, 2011. ACM.

[15] F. Probst, L. Grosswiele, and R. Pfleger. Who will lead and who will follow: Identifying influential users in online social networks. *Business and Information Systems Engineering*, 5(3):179–193, 2013.

[16] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman. Influence and Passivity in Social Media. In *Proceedings of the 20th international Conference on World wide web*, Aug. 2010.

[17] D. Simmie, M. G. Vigliotti, and C. Hankin. Ranking twitter influence by combining network centrality and influence observables in an evolutionary model. *Journal of Complex Networks*, 2014.

[18] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '09, pages 807–816, New York, NY, USA, 2009. ACM.

[19] D. Tunkelang. http://thenoisychannel.com/2009/01/13/a-twitter-analog-to-pagerank/. Technical report, 2009.

[20] D. Vogiatzis. Influential users in social networks. In *Semantic Hyper/Multimedia Adaptation*, pages 271–295. 2013.

[21] 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*, WSDM '10, pages 261–270, New York, NY, USA, 2010. ACM.