

Influential post identification on Instagram through caption and hashtag analysis

Measurement and Control
2020, Vol. 53(3-4) 409–415
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0020294019877489
journals.sagepub.com/home/mac
 SAGE

Benyamin Bashari and Ehsan Fazl-Ersi

Abstract

Influencer marketing through social networks is becoming an important alternative to traditional ways of advertising. Various solutions have been proposed that often take advantage of graph-based approaches to discover influencers in social networks. This paper designs a new method for the discovery of influential users in Instagram, by focusing on user-generated posts as an alternative source of information, to potentially augment the existing solutions based on network topology or connections. The text associated with each Instagram post potentially consists of a set of hashtags and a descriptive caption. Various word embedding methods such as Co-occurrence and fastText are examined to represent captions and hashtags. These representations are combined within a support vector machines framework to distinguish influential posts from non-influential ones. Extensive experiments show that the text data can play a significant role in identifying influential posts, and further demonstrate the strength of the proposed method for discovering influencers on Instagram.

Keywords

Social networks, influencer marketing, natural language processing, word embedding, support vector machines

Date received: 31 December 2018; accepted: 8 August 2019

Introduction

The study of social-based recommender systems has gained significant attraction since the development of Web 2.0. While primary systems often ignored social interactions among users, the more recent approaches try to incorporate the social network information, in order to improve the quality of the recommendations and make them more personalized. When people are faced with multiple potentially confusing recommendations, they tend to turn to influential people, whether being part of the supply chain (e.g. retailers or manufacturers) or being value-added influencers, such as industry analysts or professional advisers. In general, influencers in a social network are individuals, whose impacts broadened through the network. This impact is usually not comprehensive, and influencers often focus on limited topics to advertise. Being able to predict or identify the influencers on social networks could be of great value since influencers can play significant role in the success of various social, political and viral marketing campaigns as well as the entertainment events.

In this paper, we attempt to identify influential users by examining their posts. Influential posts are identified through text analysis and natural language processing. The most significant contribution of the proposed

method is that unlike the vast majority of the existing solutions, it only relies on the analysis of the User Generated Content (UGC) and not the extent of user interactions (e.g. number of posts, number of followers, number of “likes” and so on) or the structure of their connectivities.

The rest of this paper is organized as follows: in the “Literature review” section, the relevant literature concerning the topic of influence evaluation in social networks is reviewed. The “Proposed method” section describes the proposed method and its various components. In the “Experiments” section, the conducted experiments are presented and their results are discussed. Finally, the “Conclusion and Future Works” section concludes the paper by giving a brief overview of the proposed methods and highlighting a few directions for future work.

Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

Corresponding author:

Ehsan Fazl-Ersi, Department of Computer Engineering, Ferdowsi University of Mashhad, Azadi Square, Mashhad 9177948974, Khorasan Razavi, Iran.
Email: fazlersi@um.ac.ir



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without

further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Literature review

According to a study by Sun et al.,¹ the existing solutions for identifying influencers in social networks can be categorized into three groups: (1) topological-based, (2) interaction-based and (3) topic-based.

By exploiting the fact that a social network can be represented using a large-scale graph, where nodes represent the users and the edges represent the follower–following relations between the users, the topological-based methods employ the graph analysis tools to identify influencers in social networks.^{2,3} Such methods preform either locally, where an influential node (i.e. user) is identified based on itself and its neighbors, or globally, where all the nodes in the graph contribute to the prediction of the level of influence for a given node.^{4–6} The local approaches, while very fast, often perform poorly. The global ones, on the other hand, lead to a much better accuracy but at the cost of higher computational complexity. In the case of a very large network, the global approaches might be intractable or even impossible to use.⁷

The second category of methods, which are interaction-based, mostly focuses on the behavior and interactions of users to predict their influence. User interactions such as “like,” “share,” “comment,” “retweet” and “mentions,” and sometimes their combination with the topological-based methods, have been used to identify influencers^{8–11} in a variety of social network such as Twitter.

As one would expect, using such simple and compact set of features often fails to reliably and accurately predict the extent of a user’s influence, particularly with the rise of social bots and fake accounts that automatically gather followers and generate messages.

The topic-based methods, on the other hand, attempt to use yet another source of information, ignored by the previous two groups of solutions, which is the content of the users’ posts (or UGC). These methods propose various ways of integrating this information into previous methods in order to provide a better evaluation of users’ influence.^{12–14} For instance, Xiao et al.¹⁵ have used hashtags in Twitter to identify influential users in the news-related communities and reported promising results.

While the method proposed in this paper is closer to the topic-based approaches in that it mainly relies on the UGCs, existing topic-based approaches^{12–16} have all been designed for and tested on micro-blogging social networks such as Twitter. Instagram is nearly 2.5 times larger than Twitter, and as a result, many marketers and big brands put more effort toward Instagram. To the best of our knowledge, no major study has been conducted on identifying influencers in Instagram based on analyzing UGCs. Instagram is a social networking service, which unlike Twitter aims primarily at sharing visual content (i.e. photo and video), and not textual data. However, it would still be interesting, to say the least, to investigate the extent to which the text

associated with a given post on Instagram could be revealing of the post’s (or its owner’s) influence.

Proposed method

This paper focuses on identifying influential posts in Instagram, as a mean to discover influential user. Since the textual content of each post potentially consists of a set of hashtags and a descriptive caption, different techniques for processing each type of data are investigated. The result of such processing can then be combined together to produce a richer representation for posts. By evaluating the representations of a user’s posts, influential users can be identified as those with sufficient number of influential posts. Figure 1 shows the entire process of the proposed method.

Text preprocessing

Generally various types of preprocessings are needed to extract information from a given piece of text data, whether it being a set of hashtags or a caption. Since this paper only focuses on English posts, we start by removing non-English words from the text by discarding the words that contain non-ascii letters. The following steps are then performed to further prepare the text:

1. Removing punctuations and digits
2. Removing mentions
3. Changing uppercase letters to lowercase
4. Tokenizing text into a list of words¹⁷
5. Removing stopwords¹⁸
6. Stemming¹⁸

Excellent overviews of various techniques used for initial text processing (including the above-mentioned) are available in the works of Wagner¹⁷ and Srividhya and Anitha¹⁸

Word weighting

Once the text is preprocessed, the remaining constituent words are weighted using the TF-IDF (term frequency–inverse document frequency) approach, to facilitate the production of more effective text representations.¹⁹ Term frequency (TF) is defined as the total number of occurrences of a particular word in a document (i.e. post). Document frequency (DF) refers to the number of documents (i.e. posts) in the corpus that contain a particular word. The combination of term frequency and inverse document frequency, which is denoted by TF-IDF, for a given word w and document d is computed as follows

$$TF(w, d) = \frac{|\{w' \in d : w' = w\}|}{|\{w \in d\}|}$$

$$IDF(w, D) = \log \frac{|\{d' \in D\}|}{|\{d' \in D : w \in d'\}|} \quad (1)$$

$$TF - IDF(w, d, D) = TF(w, d) IDF(w, D)$$

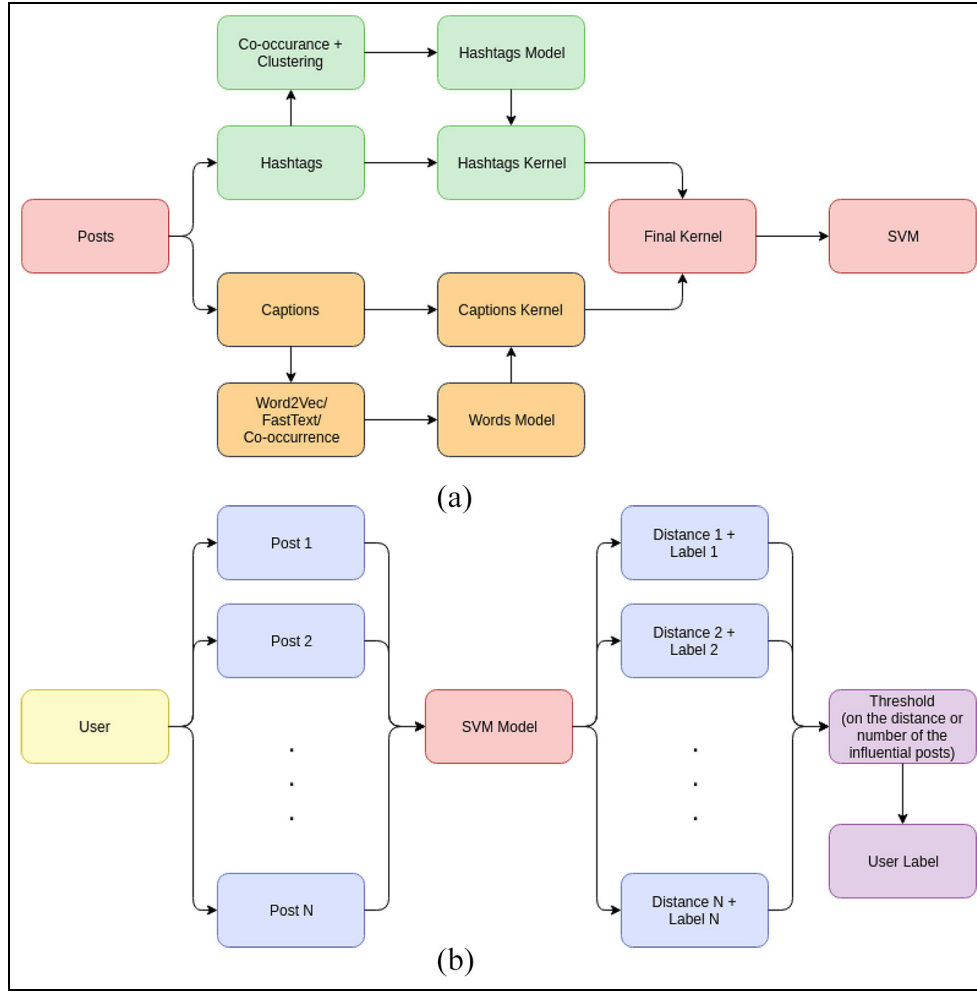


Figure 1. (a) Training phase and (b) testing phase of the proposed method, for identifying both influential posts and users.

where D denotes the corpus which contains all the documents.

Weighting the words inside a given post, the next step is to produce the post's representation. Given the different nature of captions and hashtags as the two types of available text data, their representation mechanisms are investigated separately, in the sections "Representing captions" and "Representing hashtags."

Representing captions

In this paper, two different approaches are proposed to represent captions, which are described in the following subsections.

Word2Vec and fastText. Word2Vec is a word representation method that incorporates words' semantics into a multidimensional vector of numbers.²⁰ It takes advantage of the fact that the position of a word in different sentences, along with its correlation with other words, can reveal the word's meaning.

The fastText works similarly as Word2Vec with the main difference being that it uses character n-grams to

build word vectors.²¹ Its main advantage over Word2Vec is that it can learn higher quality vectors even for words that are not sufficiently frequent in the corpus. However, fastText has a higher computational complexity.²¹

In order to produce the representation of a caption, the following two steps are performed: First, for each word occurred more than 10 times in the corpus, a 100-dimensional vector is learned using either Word2Vec or fastText. The second step produces the caption representation by taking a weighted average over the word vectors, where the weight for each vector is the TF-IDF value of its corresponding word

$$d = \{w_i \mid 1 \leq i \leq k\}$$

$$R(d, D) = \frac{\sum_{i=1}^k [r(w_i, D) \text{TF-IDF}(w_i, d, D)]}{\sum_{i=1}^k \text{TF-IDF}(w_i, d, D)}$$

In the above equation, $R(d, D)$ denotes the representation of the post d in corpus D , and $r(w_i, D)$ is the numerical vector of the word w_i generated using either fastText or Word2Vec.

Co-occurrence statistics. Similar to a bag-of-words (BoW) model, first a dictionary of keyword is formed by identifying the words that occur more than a certain number of times in the corpus. This results in a collection of N unique keywords. In the next step, the similarity between every pair of keywords (w, w') in the dictionary is calculated based on Jaccard similarity measure

$$Jaccard(w, w') = \frac{|\{d \in D : w \in d \text{ and } w' \in d\}|}{|\{d \in D : w \in d \text{ or } w' \in d\}|} \quad (3)$$

This produces a vector representation for each unique keyword that encodes its co-occurrences with other keywords in the documents. L1-norm is finally used to normalize the representations. Taking these vectors as the word representations and TF-IDF coefficients as their weights, the final representation of a document is obtained similar to the previous approach, by taking weighted average.

Representing hashtags

Previous studies have investigated hashtags in Instagram for purposes such as picture recommendation²² and automatic image annotation (AIA).²³ The method proposed for picture recommendation by Huang and Wang²² builds a correlation matrix based on the hashtags' co-occurrence and their synonymity. Then it uses a weighting scheme, which considers a hashtag more valuable, if it has more co-occurrence with other hashtags. Argyrou et al.²³ apply LDA (latent Dirichlet allocation) on hashtags in order to find the subject of an image, which can lead to AIA.

The method proposed in this paper for hashtag representation uses the co-occurrence statistics method described in the previous section, within a BoW framework. Given the fact that hashtags are often sparse, without any type of ordering, and not following any particular linguistic vocabulary, the previous co-occurrence statistics method is changed such that not only lower dimensional representations can be produced for hashtags, but also a better modeling of hashtags' semantics can be obtained.

The co-occurrence representations of all hashtags are used along with the Jaccard similarity metric to form an affinity matrix, where each element (i, j) indicates the similarity between hashtags i and j .

The Affinity Propagation algorithm²⁴ is then used to cluster hashtags into C groups (where $C = 280$ in our experiment). Using this setup, each post is finally represented based on its hashtags through a histogram of C bins, where each bin indicates the frequency of hashtags belonging to a certain cluster.

Classification

Once the appropriate features are extracted from each post's content, the task of identifying influential posts can be considered as a binary classification problem

where we have to decide whether or not a post is influential based on its representation. We use two support vector machine (SVM) classifiers, one for hashtags and the other for captions. The two classifiers are combined in the kernel space by computing a weighted average over the individual kernels. RBF (radial basis function) is used as the kernel function for caption representation methods and histogram intersection as the kernel function for hashtag representation method; the weight of each kernel is determined according to its accuracy on a validation dataset.

Experiments

Data gathering

Given the fact that no publicly available datasets exist for identifying influential Instagram users, one of the first goals of this work has been to produce a well-established dataset. For this purpose, we used a set of Instagram users listed at www.pro.iconosquare.com, who have been verified as influencers in various fields. To form the non-influencer's set, we took advantage of the fact that influencers barely follow one another on social networks, and therefore we chose non-influencers from the set of influencers' followers who are fairly active and have less than 200 connections (following and followed-by). The data for each of the influencer and non-influencer users were gathered using Instagram's API. User's data include bio, number of followers, number of followings and posts, where each post itself includes media ID, caption, hashtag and image URL.

Data preparation

In order to remove the effects that the posts' topic or individual writing styles could potentially impose over the experiments, three cleaning steps were performed, to make sure that the experiments are not being biased toward variables other than the posts' contents.

Taking advantage of the fact that hashtags are often revealing of the topics of the posts, we first identified the 10 most frequent hashtags used in our dataset and then performed the influencer versus non-influencer classification experiments for posts related to each hashtag, independently. The identified 10 hashtags are as follows: art, beautiful, fashion, food, instagood, love, nofilter, ootd, photooftheday and travel.

To remove the bias that individual users could impose on the experiments, we introduced a threshold, t_p , or the maximum number of posts a user can contribute to each experiment. For users with more relevant posts, t_p posts are randomly chosen and added to the relevant data subsets. In all our experiments, t_p is set 200.

In the last step, to ensure that the training and test splits for each experiment are almost balanced, and that the posts for each user are either in the training set or

the test set, a dynamic programming algorithm known as coin change²⁵ was used.

The final dataset consists of a total number of 16,038 posts that belong to 2272 users; of which, 346 are influencers and 1926 are non-influencers.

Results

Post classification. A set of experiments was conducted to investigate the performance of various solutions discussed in this paper, for the purpose of classifying an Instagram post into Influencer (Inf.) and Non-Influencer (NInf.) categories. Table 1 summarizes the average accuracy obtained by each method for each category. As expected, the accuracy of various representation methods is often improved through the use of TF-IDF, although the solution based on Word2Vec did not follow this trend.

Furthermore, according to the results shown in Table 1, hashtags appear to play a more significant role in distinguishing influential posts from non-influential ones, when compared to each method of caption representation. More specifically, our solution based on the BoW framework achieves a classification accuracy of 65.51%, which shows its effectiveness in representing hashtags and incorporating their semantics.

As for the classification solutions based on captions, the proposed co-occurrence statistics method achieves higher accuracy compared to other word embedding methods based on Neural Networks. However, the improved accuracy comes at the expense of higher computational and space complexities, since the word representations based on co-occurrence statistics are far larger in terms of dimensionality than Word2Vec and fastText.

In order to combine the result of different methods, which are based on different types of data (i.e. captions and hashtags) in the kernel space, we need a proper way to weight individual kernels. The method employed in our experiments is to use the normalized accuracy of each individual kernel (i.e. method) on a validation set as its weight.²⁶ Equation (4) shows this relation, where ACC_i is the accuracy of method i and a_i is its coefficient

$$a_i = \frac{ACC_i}{\sum_{i=1}^n ACC_i} \quad (4)$$

$$Kernel_{total} = \sum_{i=1}^n a_i Kernel_i, \quad \sum_{i=1}^n a_i = 1$$

Figure 2 shows the overall accuracy of influential posts identification for each data type and across all topics (i.e. captions and hashtags). (The dataset for this experiment was formed by putting together all the posts consisting at least one of the hashtags listed in Table 1 and splitting them equally into train and test set. The final dataset consist of 16,038 posts from 2272 users). As can be seen, the integrated solution clearly improves the accuracy of both methods.

Table 1. The accuracy of different methods for influential posts identification. In normal weighting scheme, all the words have equal weights (e.g. 1).

| Categories | Number of users | | | | Caption | | Co-occurrence | | | | Hashtag (%) |
|---------------|-----------------|----------|------|----------|------------|------------|---------------|------------|------------|------------|-------------|
| | Train | | Test | | Word2Vec | | fastText | | | | |
| | Inf. | Non-Inf. | Inf. | Non-Inf. | Normal (%) | TF-IDF (%) | Normal (%) | TF-IDF (%) | Normal (%) | TF-IDF (%) | |
| | | | | | | | | | | | |
| Art | 39 | 201 | 57 | 190 | 64.50 | 62.36 | 63.95 | 64.45 | 67.73 | 68.32 | 65.74 |
| Beautiful | 49 | 270 | 34 | 272 | 60.18 | 59.28 | 62.44 | 61.31 | 59.73 | 61.09 | 70.14 |
| Fashion | 54 | 218 | 53 | 227 | 64.27 | 64.68 | 64.82 | 65.37 | 64.82 | 64.27 | 61.22 |
| Food | 30 | 193 | 43 | 194 | 57.86 | 58.53 | 62.88 | 63.55 | 66.56 | 66.56 | 61.20 |
| Instagood | 40 | 238 | 39 | 239 | 71.62 | 66.73 | 68.30 | 68.88 | 72.21 | 72.02 | 67.91 |
| Love | 82 | 446 | 83 | 428 | 63.64 | 57.91 | 65.42 | 64.62 | 67.09 | 67.29 | 65.42 |
| Nofilter | 71 | 269 | 64 | 254 | 59.87 | 54.28 | 56.58 | 57.57 | 56.74 | 56.91 | 57.73 |
| Ootd | 46 | 169 | 61 | 177 | 60.66 | 57.05 | 60.00 | 61.31 | 60.00 | 60.00 | 61.31 |
| Photooftheday | 20 | 181 | 40 | 166 | 62.22 | 57.61 | 58.12 | 58.12 | 59.83 | 60.00 | 75.90 |
| Travel | 47 | 174 | 49 | 160 | 61.66 | 60.38 | 57.19 | 57.35 | 60.54 | 61.02 | 68.53 |
| Average | | | | | 62.65 | 59.88 | 61.97 | 62.25 | 63.53 | 63.75 | 65.51 |

Inf.: Influencer; NInf.: Non-Influencer; TF-IDF: term frequency-inverse document frequency.

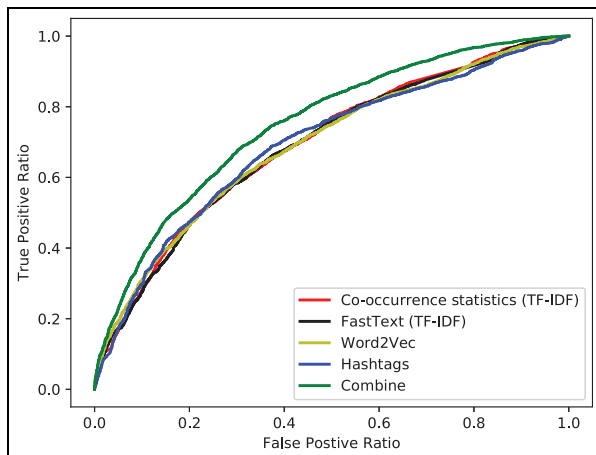


Figure 2. ROC curves of all posts, comparing the accuracy of each method.

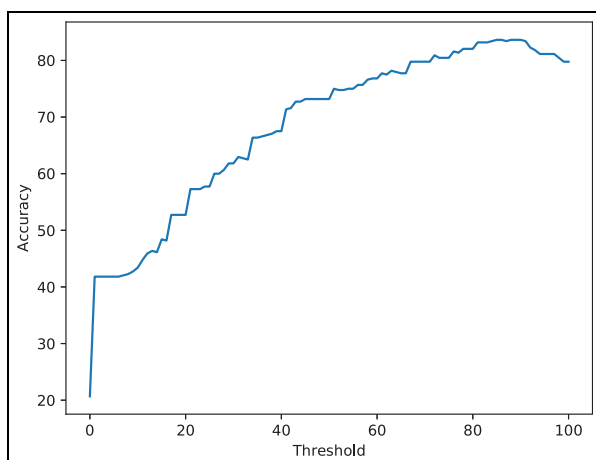


Figure 3. Accuracy of influencer identification based on different threshold on number of posts.

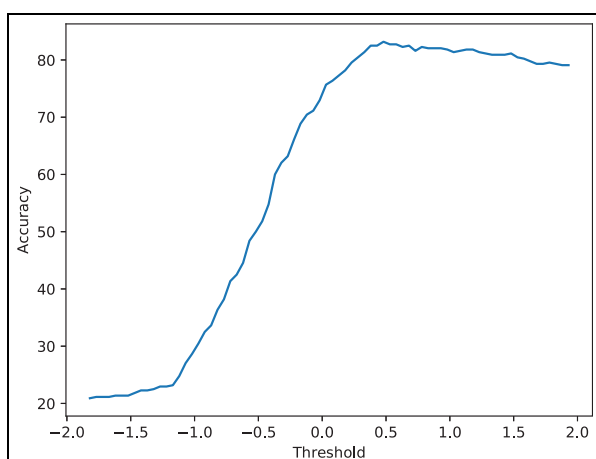


Figure 4. Accuracy of influencer identification based on different threshold on summation of scores.

Influencer identification. In order to investigate the extent to which of the proposed method can help identify influencers on Instagram, we setup two experiments. In the first

experiment, an influencer is identified as a user with a significant number of influential posts (framework (a)). In the second experiment, influencers are users for whom the average likelihood of their post being influential reaches a certain threshold (framework (b)).

Figures 3 and 4 demonstrate the accuracy of the two frameworks, respectively. As it can be seen, both frameworks perform similarly to one another. The best accuracy for framework (a) is 83.64%, which occurs when the threshold is 85% of a user's posts. In addition, the highest accuracy for framework (b) is 83.18% at the threshold set to 0.48.

Conclusion and future works

Identifying influential users in social networks is becoming of crucial importance for influencer marketing. This paper presented a novel method for identifying influential users by examining their posts. Several techniques based on text analysis and natural language processing were explored for discovering influential posts, resulting in a promising level of accuracy. The most significant contribution of the proposed method is that it only relies on the analysis of the UGC, a significant source of information that has often been ignored by previous solutions which rather focused on the extent of user interactions and the structure of their connectivities.

Integrating our approach with other methods which employ the topological structure of the network or the interactions of users could be a future research direction. Furthermore, in our method, we ignored any media content (i.e. images or videos) of the posts. Therefore, another potential direction for future work would be devising a method that gives a complete representation of the content of user's posts using all the information contained in them and attempts to evaluate the user influence based on those representations.



Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Benyamin Bashari  <https://orcid.org/0000-0002-6984-9032>
Ehsan Fazl-Ersi  <https://orcid.org/0000-0001-7384-1618>

References

1. Sun Q, Wang N, Zhou Y, et al. Identification of influential online social network users based on multi-features. *Int J Pattern Recogn* 2016; 30(6): 1659015.

2. Weitzel L, Quaresma P and De Oliveira JPM. Measuring node importance on Twitter microblogging. In: *Proceedings of the 2nd international conference on web intelligence, mining and semantics*, Craiova, 13–15 June 2012, p. 11. New York: ACM.
3. Wang P, Tian C and Lu J-A. Identifying influential spreaders in artificial complex networks. *J Syst Sci Complex* 2014; 27(4): 650–665.
4. Lü L, Zhang YC, Yeung CH, et al. Leaders in social networks, the delicious case. *PLoS ONE* 2011; 6(6): e21202.
5. Chen DB, Gao H, Lü L, et al. Identifying influential nodes in large-scale directed networks: the role of clustering. *PLoS ONE* 2013; 8(10): e77455.
6. Wang J, Hou X, Li K, et al. A novel weight neighborhood centrality algorithm for identifying influential spreaders in complex networks. *Physica A* 2017; 475: 88–105.
7. Liu Y, Tang M, Zhou T, et al. Identify influential spreaders in complex networks, the role of neighborhood. *Physica A* 2016; 452: 289–298.
8. Zhou G, Zhang X and Yamana H. Identifying topics and influential users based on information propagation in Twitter (データ工学). *IEICE Tech Rep: Data Eng (電子情報通信学会技術研究報告=)* 2013; 113(105): 29–33.
9. Romero DM, Galuba W, Asur S, et al. Influence and passivity in social media. In: *Proceedings of the joint European conference on machine learning and knowledge discovery in databases*, Skopje, 18–22 September 2017, pp. 18–33. Berlin; Heidelberg: Springer.
10. Al-garadi MA, Varathan KD and Ravana SD. Identification of influential spreaders in online social networks using interaction weighted K-core decomposition method. *Physica A* 2017; 468: 278–288.
11. Kwak H, Lee C, Park H, et al. What is Twitter, a social network or a news media? In: *Proceedings of the 19th international conference on World wide web*, Raleigh, NC, 26–30 April 2010, pp. 591–600. New York: ACM.
12. Kong S and Feng L. A tweet-centric approach for topic-specific author ranking in micro-blog. In: *Proceedings of the 7th international conference on advanced data mining and applications*, Beijing, China, 17–19 December 2011, pp. 138–151. Berlin; Heidelberg: Springer.
13. Weng J, Lim EP, Jiang J, et al. TwitterRank: finding topic-sensitive influential twitterers. In: *Proceedings of the third ACM international conference on Web search and data mining*, New York, 4–6 February 2010, pp. 261–270. New York: ACM.
14. Liu X, Shen H, Ma F, et al. Topical influential user analysis with relationship strength estimation in Twitter. In: *Proceedings of the 2014 IEEE international conference on data mining workshop (ICDMW)*, Shenzhen, China, 14 December 2014, pp. 1012–1019. New York: IEEE.
15. Xiao F, Noro T and Tokuda T. Finding news-topic oriented influential Twitter users based on topic related hashtag community detection. *J Web Eng* 2014; 13(5–6): 405–429.
16. Kardara M, Papadakis G, Papaoikonomou A, et al. Large-scale evaluation framework for local influence theories in Twitter. *Inform Process Manag* 2015; 51(1): 226–252.
17. Wagner W. Steven Bird, Ewan Klein and Edward Loper: natural language processing with python, analyzing text with the natural language toolkit. *Lang Resour Eval* 2010; 44(4): 421–424.
18. Srividhya V and Anitha R. Evaluating preprocessing techniques in text categorization. *Int J Comput Appl* 2010; 47(11): 49–51.
19. Salton G and Buckley C. Term-weighting approaches in automatic text retrieval. *Inform Process Manag* 1988; 24(5): 513–523.
20. Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. *Arxiv*, 2013, <https://arxiv.org/pdf/1301.3781.pdf>
21. Bojanowski P, Grave E, Joulin A, et al. Enriching word vectors with subword information. *Arxiv*, 2016, <https://aclweb.org/anthology/Q17-1010>
22. Huang YF and Wang PL. Picture recommendation system built on Instagram. In: *Proceedings of the 2017 international conference on artificial intelligence, automation and control technologies*, Wuhan, China, 7–9 April 2017, p. 23. New York: ACM.
23. Argyrou A, Giannoulakis S and Tsapatsoulis N. Topic modelling on Instagram hashtags: An alternative way to automatic image annotation? In: *Proceedings of the 2018 13th international workshop on semantic and social media adaptation and personalization (SMAP)*, Zaragoza, 6–7 September 2018, pp. 61–67. New York: IEEE.
24. Frey BJ and Dueck D. Clustering by passing messages between data points. *Science* 2007; 315(5814): 972–976.
25. Bocker S and Lipták Z. A fast and simple algorithm for the money changing problem. *Algorithmica* 2007; 48(4): 413–432.
26. Xiao J, Hays J, Ehinger KA, et al. Sun database: large-scale scene recognition from abbey to zoo. In: *Proceedings of the 2010 IEEE conference on computer vision and pattern recognition (CVPR)*, San Francisco, CA, 13–18 June 2010, pp. 3485–3492. New York: IEEE.