Heliyon 7 (2021) e06472

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Identification of influencers in online social networks: measuring influence considering multidimensional factors exploration

Yun-Bei Zhuang [a,*], Zhi-Hong Li [b], Yun-Jing Zhuang [c]

[a] School of Economics, Shandong University of Technology, ZiBo, 255000, China
[b] School of Business Administration, South China University of Technology, Guangzhou, 510641, China
[c] Literature School, University of Jinan, Jinan, 255022, China

ARTICLE INFO

Keywords:
Online social networks
Social influence measurement
Multidimensional
Factors exploration
Topic-level network
Global-level network

ABSTRACT

Online Social networks exhibit heterogeneous nature with nodes playing far different roles in structure and function. To identify influencers is thus very significant, allowing us to control the outbreak of public negative opinion, to conduct advertisements for e-commercial products, to predict popular scientific publications, and so on. The identification of influencers attracts increasing attentions from both computer science and communication science, with multiple dimensional metrics ranging from structure-based to information-based and action-based. However, most work simply rely on one dimensional metrics. Therefore, in this paper, we analyze three dimensional characteristics (structure-based, information-based, and action-based factors) to develop the multidimensional social influence (MSI) measurement approach. With topic distillation and conditional expectation, the MSI approach can not only measure users topic-level influence, but also measure users global-level influence. Based on data collected from SinaWeibo.com, the experimental results show that the proposed framework outperforms two traditional methods (LeaderRank and FBI) both on the topic-level and the global-level. The proposed framework can be effectively applied to promote word-of-mouth marketing, and to steer public opinion in certain directions, even to support decisions during a negotiation process.

1. Introduction

With the integration and development of technologies and social networking services, Online Social Networks (OSNs) is becoming the decisive dissemination platform of information, knowledge, technology and other resources [1]. However, due to the scale-free property of social networks [2], the diffusion in OSNs hinges on a specific set of users, called influencers, which allows us to better control the outbreak of rumors conduct successful advertisements for e-commercial products, optimize the use of limited resources to facilitate information propagation [3]. Therefore, identifying influencers is becoming one of the most significant issues in both computer science and communication science.

Much effort has been put to detect influencers. However, there still lack of a universal criteria of influence. Since iterature on influencer detection is highly diverse with many different methods and variations on the methods proposed by different authors. According to [3, 4], the current influencer detection approaches can be divided into " influence maximization approaches" or" influence measurement approaches". The target of influence maximization is identifying a subset of influential nodes under a given diffusion model [5] to achieve global optimization. While the influence measurement is a microcosmic problem, which target on developing an influence measurement method to detect influencers individually. However, most influence measurement methods were essentially based on single dimension factors, while influence in OSNs is a complex force, which is determined by multiple attributes from multidimensions [3].

Our study contributes to the literature by developing a multidimensional social influence (MSI) measurement approach, which can detect influencers more accurately comparing to most existing approaches. Firstly, We analyze the characteristics of online social network from three dimension (structure-based, information-based, and action-based), and explore the crucial factors of users' influence. Secondly, with selection of appropriate metrics, calculation of weights, and acquisition of topic distribution, a topic-level multidimensional social influence measurement (TMSI) model and a global-level social influence measurement (GMSI) model are developed. Based on dataset retrieved from Sinaweibo.com, the experimental studies demonstrate the proposed MSI approach outperforms other

* Corresponding author.
E-mail address: zhuangyunbei@139.com (Y.-B. Zhuang).

https://doi.org/10.1016/j.heliyon.2021.e06472
Received 30 November 2020; Received in revised form 2 February 2021; Accepted 5 March 2021

Table 1. Representative methods for identifying individual influencers.

Classification	Study	Structure	Action	Information
Centrality	Degree centrality [23]	✓		
	Closeness centrality [32]	✓		
	Betweenness centrality [33]	✓		
	PageRank [16]	✓		
	HITs [17]	✓		
	LeaderRank [6]	✓		
Node operation	Node contraction method [34]	✓		
	Connectivity-sensitive [3]	✓		
	Stability-sensitive [3]	✓		
Machine learning	[35]	✓	✓	
	[27]	✓	✓	✓
	[24]	✓	✓	✓
	[36]	✓	✓	✓
Diffusion-based	[29]	✓	✓	
	[37]	✓		✓
	[38]	✓	✓	✓

traditional approaches [6, 7] in each topic-level networks and the global-level network.

The contributions of this work are summarized as follows:

Based on LeaderRank, MSI approach captures the generative process of the users' information-based and social influence by considering multidimensional (structure-based, information-based and action-based) factors.

② The social influence between users are different with topics. Thus in this paper, we take a topic-aware perspective to jointly learn topics characteristics and social influence. Thus, the TMSI model is developed to measure users' social influence on a specific topic. And utilizing knowledge of the conditional expectation of users' topic-level influence, the GMSI model is developed to measure users' social influence on the global-level network.

③ Experiments using real data collected from Sinaweibo.com show that MSI can find influential nodes with an acceptable computational complexity and effectiveness.

The paper is organized as follows. Section 2 reviews the related work in this area. Section 3 explains in detail the TMSI and GMSI social influence measurement model. Section 4 describes the experimental results of the proposed approach. Section 5 concludes the paper and shows future research directions.

2. Related work

Influencer detection has been raising research issues and receiving lots of attention in various domains, ranging from social psychology to network sciences [8]. And the conventional methods for identifying individual influencers can be divided into centrality based methods, node operation methods, diffusion-based methods and machine learning methods [3, 9].

2.1. Centrality based approaches

Centrality based approaches devoted to detect influencer only based on structural information [3]. The literature on network theory describes a large number of such centralities, ranging from neighborhood-based centralities (Such as, K-Shell decomposition method [10, 11], H-index method [12], cycle-based method [13]), path length-based centralities (Such as, Information Entropy method [14], gravity-based model [15]), and iterative centrality methods (Such as, PageRank [16], LeaderRank [6], HITS [17]).

However, Centrality methods have one obvious limitation, that is a centrality which is optimal for one application is often sub-optimal for a different application [3, 18]. Therefore, one centrality based method maybe sub-optimal for another different type network. Since different online social network have different topological structure, centrality based approaches cannot maintain their effectiveness and accuracy when used on OSNs [19].

2.2. Node operation approaches

A node is important if its removal would largely shrink the giant component (connectivity, stability, or agglomeration) of the network [20]. Therefore, node operation approaches are proposed to find influential user by node removal and contraction [21, 22]. In detail, Linyuan Lü *et al.* summarized those methods into connectivity-sensitive methods, stability-sensitive methods and node contraction methods [23].

Similar to centrality based methods, the node operation approaches also have limitation on OSNs. Since the accuracy of node operation methods depend on the global input of the topological structure. Therefore, node operation methods are lack of efficiency to be adapted onto large-scale OSNs.

2.3. Machine learning approaches

With the opening of the online data platform, automated classification methods have gained considerable importance for influencers identification [24, 25, 26]. For instance, Pajo etal. developed a Fast Lead User Identification (FLUID) approach with features extracted from activity measures, centrality measures and sentiment [27]. Fan et al. introduced a deep reinforcement learning framework FINDER for key players finding in complex networks [25]. However, although machine learning methods can realize fast detection of influencers, their effectiveness depends on the size of the data and the quality of learning set, thus will have the limitation of “light data start”.

2.4. Diffusion based approaches

Another common approach to the influencer detection problem is to simulate influence cascades through the network based on the existence of links in the network using diffusion models, where the influence spread of a node on given diffusion processes are treated as criteria for influence [28]. For example, for an arbitrary node, setting this node as the infected seed and then the total number of ever infected nodes in a

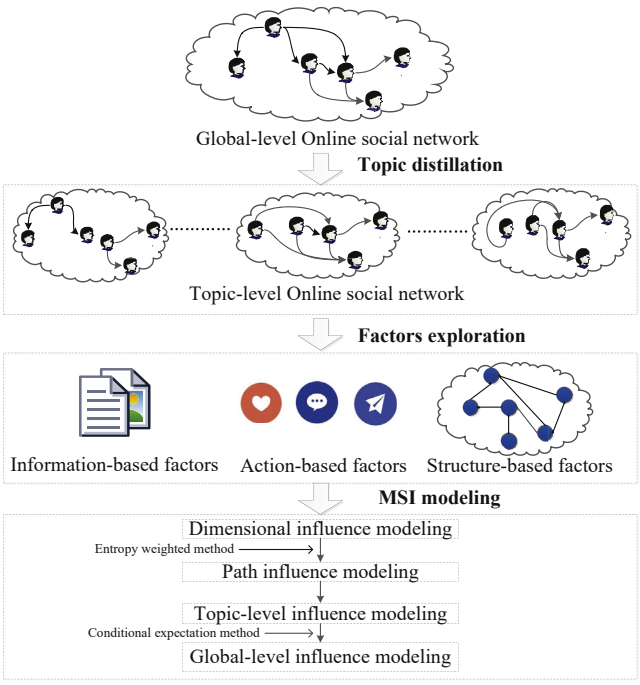


Figure 1. The framework of MSI approach.

"susceptible-infected-recovered (SIR)" process is widely used as a metric quantifying the importance of this node [3, 29].

However, based on extensive simulations, Šikić et al. [31] showed that for a given SIR process with two parameters (i.e., the spreading rate and recovering rate), the rank of nodes' influences largely depends on the parameters.

Table 1 reclassifies the existing literature according to the dimensions of metrics in each paper. As illustrated in Table 1, centrality based methods are solely focusing on the topological structure, and do not take into account information about properties of the vertices/edges or edge weights [3, 29]. This may be one of the reasons why centrality based and node operation methods have limitations on OSNs [39], since the most difference between social network and other physical networks, is social network is connected by human behavior and information, therefore, as in [3, 30], a good influencer detection method has to integrate topological features and dynamical properties dynamical parameters into account.

Unlike centrality based methods and node operation methods, most machine learning methods and diffusion based methods consider users' behavior and information factors, which greatly improve the efficiency and accuracy on large scale network. However machine learning methods have the limitation of "light data start" and the performance of 4 diffusion based methods largely depends on the parameters.

Together, these studies provide important insights into the identification of influencers in complex systems, however, the influencer detection on large scale network still lack an accurate and efficient approach. In this paper, we are supposed to design a better-performed influencer detection method. The accuracy is improved compared to centrality based methods by taking into account 12 features from structure dimension, information dimension and action dimension. The efficiency is improved compared to node operation methods and machine learning methods by the entropy weighted modeling of the three dimensional metrics.

3. Multidimensional social influence measurement

In this section, we present a detailed multidimensional user influence measurement approach, with a Topic-level Multidimensional Social Influence (TMSI) measurement model and a Global-level Multidimensional Social Influence (GMSI) measurement model. Following the definition of Granovetter, Section 3.1 introduces factors relating to user influence measurement, and integrates these factors to model each dimensional social influence. Section 3.2 proposes the TMSI model of a given topic-level network, with a corresponding algorithm. Subsequently, Section 3.3 presents the GMSI model with information of the topic distribution. The Figure 1 shows the framework of MSI. With MSI approach, each user v in topic-level network $G_Z = (V_Z, E_Z)$ is assigned a topic-level social influence $UI_z(v)$ ($z \in T$) and a global-level social influence $UI(v)$.

Table 2. Notations of OSNs.

Symbol	Description
G = (V, E)	Online social network
N	Number of nodes in V
u,v	Index of nodes in V
(u, v)	Index of edge in E
z	Index of topics
$G_z = (V_z, E_z)$	The z_{th} topic-level network of G
g	The group node, which is double-directional with each node in each topic-level network
$G^g = (V^g, E^g)$	The z_{th} topic-level network with the added group node
$-v_d\}(d \leq 3)$	Set of node v's d-layer friends ($d \leq 3$)
$-v_{in}\}$	Set of nodes that points to node v
U I(v)	The global-level social influence score of v
U $I_z(v)$	The z_{th} topic-level social influence score of v
v.kw	The keywords of node v's text content
t_z.kw	The keywords of topic z
n_v	Number of messages of node v
n_v.media	Number of messages of node v that contains multimedia
n_v.URLs	Number of messages of node v that contains URLs
n_v.hot	Number of messages of node v that contains hot topics
-v.action}	Set of v's actions
-(u, v).action}	Set of (u,v)'s interactions
v.IP	The registered address of user v
-(u,v).action}.@	Set of (u,v)'s interactions that contain ``@''

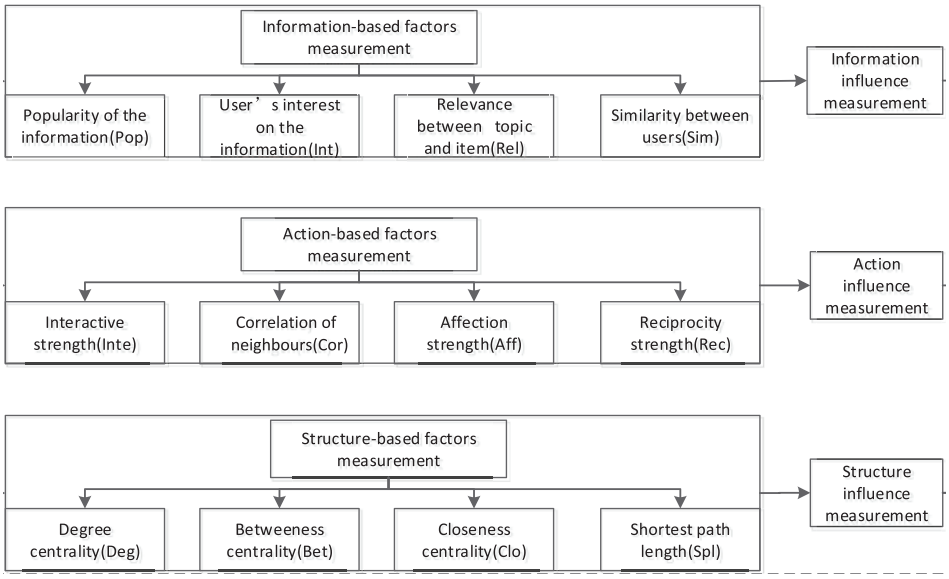


Figure 2. Factors related to user influence.

Table 2 below describes the notations used in this paper. All variables in the table are normalized in the range of [0, 1].

3.1. Factors exploration and measurement

Whether information will be spread depend on the connection influence between users [40, 41]. As argued by Granovetter [42], the strength of social relation is a function of "duration, emotional intensity, intimacy, and reciprocal services between people [43]". In detail, "duration" is a measurement of the amount of time of a tie between two nodes; "emotional intensity" is the degree, amount of strength or force that something has; "intimacy" is the state of being in a very personal or private relationship; and "reciprocal services between people" indicates actions carried out in common between two nodes in an OSN [40].

With this seminal work as a baseline, prior works started to consider the factors in a social-relation-based dimension to measure the connection influence between users. Structural factors, reciprocity, similarity, interaction activity (type, frequency, context), or social distance (socio-economic status, education level or political affiliation), etc. are respectively taken as predictors in [40, 44].

As illustrated by Figure 2, similar to [40, 44], to measure the connection influence between users, we use interaction frequency, reciprocity and other indicators as predictors of "duration, emotional intensity, intimacy, and reciprocal services between people". Then classify them into three dimensions, and correspondingly define the combined strength of these factors in each dimension as "information influence", "action influence" and "structure influence". The following section 3.1.1-3.1.3 in this section will exhaustively show the measurement of each factor and each dimensional influence.

3.1.1. Information factors measurement

As illustrated by Figure 2, information-based factors are used to estimate the contribution of users' posts, which includes the following four factors related to the text content [45]: the "popularity of the information", the "type of information", the "relevance of the message with respect to the topic", and "users' similarities". The following Eqs. (1), (2), (3), and (4) show the measurement of these factors.

3.1.1.1. Popularity of information. According to [46], some users may be more likely to rebroadcast popular content, therefore messages related to hot topics are more likely to be spread. To measure this effect, we define the "popularity of the information" factor, *Pop(v)*, by the frequency of messages that contain keywords related to hot topics.

$$Pop(v) = \frac{|n_v.hot|}{|n_v|} \tag{1}$$

3.1.1.2. Information type. Studies in [43, 47] conceive that the strength of connection partially depending on the type of the information. And Chen etal. further show, people tend to pay more attention to and spread specific/popular messages that with multimedia or URLs. To measure this effect, we build "information type" factor, *Typ(v)*, as Eq. (2) by the frequency of messages that contain multimedia (videos, photos or sounds) or URLs.

$$Typ(v) = \frac{|n_v.media| + |n_v.URLs|}{|n_v|} \tag{2}$$

3.1.1.3. Relevance between user's messages and a topic. This work aims to derive users' topic-level social influence, that is the social influence on a given topic. Therefore, we define a factor indicating the relevance of a user's messages with respect to the topic, *Relz(v)*, which can be measured by a Jaccard coefficient as Eq. (3).

$$Rel_z(v) = \frac{|v.kw \cap t_z.kw|}{|v.kw \cup t_z.kw|} \tag{3}$$

where $t_z.kw$ and $v.kw$ refer to the keywords of topic z and user v's messages, respectively. Text word segmentation is conducted on the data set to derive keywords.

3.1.1.4. Users' similarity. According to the social identity theory [48] and homogeneity theory [49], McPherson et al believes that users with similarity are more likely to connect and interact with each other [50]. The similarity can be reflected in many ways, such as age, hobbies, interested topics, etc. To capture the effect of "similarity", we define *Sim(u, v)* as Eq. (4).

$$Sim(u, v) = \frac{|u.kw \cap v.kw|}{|u.kw \cup v.kw|} \tag{4}$$

3.1.2. Action factors measurement

As shown in Figure 2, to determine the effect of users' actions (viewing, mentioning, tweeting, etc.) on their connection strength, we define the following action factors according to users' interaction records:

3.1.2.1. Interactive frequency. On OSNs, users frequently interact (post, review, comment) with others, always have better reputations and

influence than other users [51]. Therefore, we denote interactive frequency as Inte(u, v) (Eq. (5)), to measure the effect of users' interactions.

$$\mathrm{Inte}(u,v) = \frac{|\{(\mathrm{u},\mathrm{v}).\mathrm{action}\} \cup \{(\mathrm{v},\mathrm{u}).\mathrm{action}\}|}{|\{\mathrm{u}.\mathrm{action}\} + \{\mathrm{v}.\mathrm{action}\}|} \tag{5}$$

3.1.2.2. Correlation of neighbors. The correlation of users' friends is a key indicator of similarity. Richard Alba and Charles Kadushin support this observation and further verify the overlapping or similar friends among users will affect their intensity of connection [52]. In this paper, *Cor (u, v)* is defined by the correlation of users' first-layer neighbours' ID list, to measure the similarity of users' friends.

$$Cor(u,v) = \frac{\{\{u1\} \cap \{v1\}\}}{-\{u1\} \cup \{v1\}} = correl(\{u_1\}.ID, \{v_1\}.ID) \tag{6}$$

3.1.2.3. Affection strength. Affection strength refers to the degree of affection that arises between users who share the same relationships, such as consanguinity relationships, friend relation-ships, colleague relationships or location-based relationships. Thus, we use the frequency of "@" action to quantify affection strength based on consanguinity, friends and colleagues relationships, and denote this strength as *Aff@(u, v)*:

$$\mathrm{Aff}_{@}(\mathrm{u},\mathrm{v}) = \frac{|-(\mathrm{u},\mathrm{v}).\mathrm{action}\}.@|}{|-(\mathrm{u},\mathrm{v}).\mathrm{action}\}|} \tag{7}$$

where the *|-(u, v).action}.@|* simplifies the number of actions that contain the signal "@", as "@" is usually used between familiar users (friends, relatives or colleagues).

In addition, we use the correlation of two users' registered address to represent their affection strength based on location, *AffIP(u, v):*

$$Aff_{IP}(u,v) = correl(u.IP, v.IP) \tag{8}$$

Then, the affection strength of *(u, v)*, Aff (u, v), can be quantified by the following formula:

$$Aff(u,v) = \alpha_1 Aff_{@}(u,v) + \alpha_2 Aff_{IP}(u,v) \tag{9}$$

where α_1 and α_2 are two weight coefficients that can be obtained using many evaluation methods, such as the entropy weight method, the analytic hierarchy process (AHP), or the gray correlation analysis.

3.1.2.4. Reciprocity. According to network exchange theory [53], people are more easier to interact with those who grant them rewards, these rewards can be emotional (such as affection) or behavioral (such as reciprocity). In this work, we define reciprocity strength, *Rec(u, v)*, as the bi-directional interaction frequency of two users:

$$\mathrm{Rec}(\mathrm{u},\mathrm{v}) = \frac{|-(\mathrm{u},\mathrm{v}).\mathrm{action}\} \cap -(\mathrm{v},\mathrm{u}).\mathrm{action}\}|}{|-(\mathrm{u},\mathrm{v}).\mathrm{action}\} \cup -(\mathrm{v},\mathrm{u}).\mathrm{action}\}|} \tag{10}$$

3.1.3. Structure factors measurement

According to the Social Capital Theory [42], an individual's structural position in the network, significantly affect users' information exchange behavior and users' status. Specifically, a node with high centrality score is usually considered more highly influential than other nodes in the network [3, 36]. To capture this effect, the following Eqs. (11), (12), (13), and (14) are defined:

3.1.3.1. Betweenness centrality. Betweenness centrality, Bet(v), measures the average degree to which a given node lies in the shortest paths of other nodes, and can be defined by the following Eq. (11) [32].

$$\mathrm{Bet}(v) = \sum_{r \neq v \neq w \in V} \frac{|\{g_{\mathrm{rw}}(v)\}|}{|\{g_{\mathrm{rw}}\}|} \tag{11}$$

where $|\{g_{rw}\}|$ denotes the number of shortest paths from node r to node w and $|\{g_{rw}(v)\}|$ denotes the number of the shortest paths from node r to node w through node *v*.

3.1.3.2. Closeness centrality. Closeness centrality, *Col(v)*, are length-based measure that counts the length of walks [32], and can be defined as Eq. (12).

$$\mathrm{Col}(v) = \frac{1}{N+1} \sum_{w \in V, w \neq v} \frac{1}{g_{v,w}} \tag{12}$$

3.1.3.3. Degree centrality. Degree centrality counts the number of paths of fixed length *k* ($k \in N^+$) that begin from a given node. Eq. (13) shows the most common expression of user *v*'s degree centrality as *Deg(v)*:

$$\mathrm{Deg}(v) = \frac{|\{v_k\}|}{|\mathrm{N}+1|} \tag{13}$$

where $k = 1$ and $|\{v_1\}|$ represents the number of *v*'s adjacent neighbors, i.e., the degree of *v*.

3.1.3.4. Inverse shortest path length. In principle, any path (direct or indirect) connecting nodes *v* and *w* can be used as a passageway to deliver v's influences onto w and vice versa. Generally speaking, the influence will decay as the increase of the path length [3, 54]. Therefore, we define a decaying function, as Eq. (14), to measure the indirect influence between two indirectly connected users.

$$\mathrm{Ispl}(\mathrm{u},\mathrm{v}) = \frac{1}{d(\mathrm{u},\mathrm{v})} \tag{14}$$

where *d(u, v)* is the smallest distance between user *u* and user *v*.

There are many centrality measures, such as "coreness", "eccentricity", "eigenvector" etc. [3]. However, among these measures, "degree" and "shortest path length", are the simplest metrics, with the computation time complexity as $O(m)$ and $O(mn)$ (n is the number of nodes in the network and m is the average number of neighbor nodes), when the network is unweighted. But "degree" only reflects the local characteristics of the node, thus "betweenness" and "closeness" are added to reflect the global importance of the node based on the whole topological structure. While the "coreness" and "closeness" have similar performance which are better than "eigenvector" in the same experiment [3]. And to measure the mutual effect of user's indirectly connected neighbor the "inverse shortest path length" is added.

3.2. Topic-level social influence measurement

According to [37], users' social influence differentiates with topics (items/domains). Therefore, in this section, we develop a TMSI model to measure users' topic-level social influence.

3.2.1. Dimensional influence measurement

According to the modelling of social impact [55], information influence, action influence, and structure influence are correspondingly defined as multiplicative functions of information factors, action factors and structure factors. The following Eqs. (15), (16), and (17) show the details:

Definition 3.1 (Information influence). Information influence refers to the influence derived from users' messages. And a content-oriented analysis is indispensable to obtain a better understanding of social influence. Therefore, according to the above analysis, a user's information influence can be formulated as follows:

$$Inf(u,v) = Pop(v) \cdot Typ(v) \cdot Relz(v) \cdot Sim(u,v) \tag{15}$$

The "·" is the inner product operator, which is used to incorporate factors' effect on social influence measurement.

Definition 3.2 (Action influence). Similar to the measurement of "information influence", the action influence *Act(u,v)* of *(u,v)* can be measured by the product of the four action-based factors as follows:

$$Act(u,v) = Inte(u,v) \cdot Inti(u,v) \cdot Aff(u,v) \cdot Rec(u,v) \tag{16}$$

Definition 3.3 (Structure influence). The structure influence *Str(u, v)* is defined as the structure importance of *(u,v)* in the network topological structure.

$$Str(u,v) = Bet(v) \cdot Col(v) \cdot Deg(v) \cdot Ispl(u,v) \tag{17}$$

where *Bet(v)*, *Col(v)*, *Deg(v)* and *Ispl(u,v)* are defined by Eqs. (11), (12), (13), and (14). Therefore, we can derive $I = (Inf(u,v))_{n\times n}$, $A = (Act(u,v))_{n\times n}$, $S = (Str(u,v))_{n\times n}$ as the $n \times n$ information, action and structure dimensional influence matrix, respectively.

3.2.2. Path influence measurement

Definition 3.4 (Path influence). Path influence, *pi(u,v)*, is defined to measure the influential strength of users' (direct or indirect) connection, which is the social influence on a path from user u to user *v* in the network:

$$pi(u,v) = \beta_1 Inf(u,v) + \beta_2 Act(u,v) + \beta_3 Str(u,v) \tag{18}$$

where β_1, β_2, β_3 are the weights that can be used to adjust the contribution of information dimensional, action dimensional, and structure dimensional influence value, which can be derived by "information entropy weighted method", and satisfying $\sum_{l=1}^{3}\beta_l = 1$.

Basically speaking, there are 3 steps to obtain the weights. Firstly, data standardization; Secondly, computing each factor's entropy based on information entropy definition; Thirdly, determining each factor's weight according to the information entropy it has. In this paper we utilize the Shannon's information entropy [56] "$E_l = -\frac{1}{\ln(n)}\sum_{i,j} q_{ij}\ln(q_{ij})$" ($q_{ij} = S(i,j), q_{ij} = I(i,j), \text{or} q_{ij} = A(i,j)$ in this study) to calculate factors' entropy and employ the classic weight calculate formula "$\beta_l = \frac{1-E_j}{\sum l(1-E_l)}$" to determine the weights.

Therefore, the initial social influence can be measured as the summation of the path influence between users and their neighbors. For each *v* in $G_z^g = (V_z^g, E_z^g)$ (*g* is the group node [6], $G_z^g = (V_z^g, E_z^g)$is the z_{th} topic-level network with the added group node *g*), the initial TMSI value can be derived as Eq. (19).

$$UI_z^0 = \frac{1}{|V_z^g|}\sum_{u\in V_z^g} pi(u,v) \tag{19}$$

3.2.3. Model adjustment

In Eq. (19), each friend's contribution is assigned evenly with the same weight $\frac{1}{|V_z^g|}$, However, in fact, these contributions should be different, the stronger connection, the more contribution. Similar to LeaderRank, we argue that the contribution from *u* to *v* should be determined by their path influence (in [6], the edge importance) and can be measured as $p_{u,v}$ by the following equation:

$$pi(u,v) = \frac{pi(u,v)}{\sum w\in V_z^g pi(u,w)} \tag{20}$$

with *p(u,v)* as the contribution proportion, the initial TMSI value (Eq. (19)) should be adjusted as follows.

$$UI_z^{t+1}(v) = \sum_{u\in -v_{in}\}} p_{u,v} UI_z^t(v) \tag{21}$$

when $t = 0$, $UI_z^0(v)$is the initial social influence obtained by Eq. (19).

Repeat Eq. (21), when the steady state is attained at time t_c, we have

$$UI_z(v) = UI_z^{tc}(v) + \frac{UI_z^{tc}(g_z)}{N} \tag{22}$$

Symbolizing the spread of information on OSNs as a finite Markov chain, and using $p_{u,v}$ as the transfer probability. Thus, the transfer matrix of each topic network $G_z^g = (V_z^g, E_z^g)$:

$$P = \begin{pmatrix} 0 & p(1,2) & \cdots & p(1,N+1) \\ p(2,1) & 0 & \cdots & p(2,N+1) \\ \vdots & \vdots & \cdots & \vdots \\ p(N+1,1) & p(N+1,2) & \cdots & 0 \end{pmatrix} \tag{23}$$

We can define a $N + 1$ dimension vector $UI_z^t = (UI_z^t(1), UI_z^t(2), \cdots, UI_z^t(v), \cdots, UI_z^t(N+1))(v \in V_z, N = \| V_z \|)$, where each dimension refers to the social influence score of its corresponding node at the t_{th} iteration. In addition, $UI_z^0 = (UI_z^0(1), UI_z^0(2), \cdots, UI_z^0(v), \cdots, UI_z^0(N+1))$. Then,

$$UI_z^{t+1} = UI_z^t \cdot P$$

After the c-step iteration,

$$UI_z^c = UI_z^{c-1} \cdot P = UI_z^{c-2} \cdot P^2 = \cdots = TMSI_z^0 \cdot P^c$$

and the final topic-level social influence is at the steady time t_c is

$$UI_z^{t_c} = UI_z^0 \cdot P^{t_c} + UI'^{t_c}_z$$

where, $UI'^{t_c}_z = \frac{UI_z^{tc}(g)}{N} \cdot \vec{1}$, $\vec{1}$is a unit column vector.

3.3. Global-level social influence evaluation

Assume topics are independent, with the topic-level social influence $UI_z(v)$, the topic set T = {z}[1] and the topic distribution Θ[2], the global-level social influence of node *v* can be calculated by the conditional expectation of his topic-level social influence.

$$UI(v) = E(UI_z(v)) = \sum_{z\in T} p(z|v) \bullet UI_z(v)$$

where $p(z|v)$ is the probability of topic *z* disseminated by user *v*, and can be calculated as $p(z|v) = \frac{p(v|z)\cdot p(z)}{p(v)}$. $p(v|z)$is the probability of user *v* in topic-level network G_z, and $p(z)$ is the probability of topic *z*. Besides, for calculation convenience, we assume users (in the same topic-level network) and the topics are independently distributed to U(0, 1), then $p(v|z) = p(v|z,u) = \frac{1}{|V_z|}$ ($v, u \in V_z$) and p(z) = p(z|z') = $\frac{1}{|T|}$ ($z \in T$). Therefore, with the total probability formula, global-level social influence can be calculated with the following Eq. (24):

[1] In general, the topic set can be obtained by many different ways. For example, the predefined categories or the user-assigned tags on OSNs might be used.

[2] To derive the topic distribution, topic modelling methods in [47, 57] can be utilized.

$$
\begin{aligned}
UI(v) &= E(UI_z(v)) \\
&= \sum_{z\in T} p(z|v) \bullet UI_z(v) \\
&= \sum_{z\in T} \frac{p(v|z) \bullet p(z)}{p(v)} \bullet UI_z(v) \\
&= \sum_{z\in T} \frac{p(v|z) \bullet p(z)}{\sum_{z'} p(v|z') \cdot p(z')} \bullet UI_z(v) \\
&= \sum_{z\in T} \frac{1}{\sum_{z'} \frac{1}{|V_{z'}|} \cdot |T|} \bullet UI_z(v)
\end{aligned} \tag{24}
$$

4. Empirical study

In this section, experiments are presented to evaluate the effectiveness and accuracy of MSI approach, with dataset crawled from Sina Weibo. Firstly, we analyze the correlation of different dimensional factors. Secondly, we quantitatively evaluate the performance of the topic-level TMSI by two evaluation metrics-"Duplication percentage" and "Influence spread". Thirdly, we verify the effectiveness and accuracy of the global-level GMSI model by three evaluation metrics- "Duplication percentage", "Influence spread" and "Ranking accuracy"[3]. The results show that both the topic-level TMSI model and the global-level GMSI model can achieve better influencer detection, in comparing with the FBI and LeaderRank approaches.

Algorithm 1. The calculation of users' social influence.

Input:

A topic set, T;

Topic distribution, $p(z)$

m topic-level networks, $G_z^g = (V_z{}^g, E_z{}^g)(z = 1,2,\cdots,m), m = |T|$

m factor sets, $F^z = -f_1{}^z, f_2{}^z, \cdots, f_{12}{}^z\}, z \in T$;

A constant, ε, to control the iteration times;

Output:

The global level social influence value for all nodes in G, and the topical social influence value for all nodes in G_z^g.

/∗ Let t indicate the number of iterations and ε control the convergence condition.$t \to 0$, after each iteration, $t \to t+1$; $\rho \to 1$ ∗/

1: **for** each topic $z \in T$ **do**

2: **for** each node $v \in V_z^g$ **do**

3: **for** each node$u \in -v_d\}(d \le 3)$ **do**

4: Calculate the information influence-$Inf(u,v)$, the action influence $Act(u,v)$, the structure influence$Str(u,v)$, and the path influence-$pi(u, v)$, according to equations (15), (16), (17) and (18);

5: Calculate the transfer probability from node u to node v, $p(u,v)$, according to Equation (20);

6: Calculate the initial TMSI, $UI_z{}^0(v)$, according to Equation (19);

7: **while** $\rho > \mathcal{E}$ **do**

8: Calculate the initial TMSI, $UI_z{}^t(v)$, according to Equation (19);

9: $\rho = \frac{\sum v\in V_z^g |UI_z{}^t(v) - UI_z{}^{t-1}(v)|}{N+1}$

10: **end while**

11: **if** $v \neq g$ **then**

12: The final TMSI value of v, $UI_z(v)$, according to Equation (22);

13: **end if**

14: **end for**

15: **end for**

16: $UI_z = (UI_z(1), UI_z(2), \cdots, UI_z(v), \cdots, UI_z(n))(v \in V_z, n = \| V_z \|)$;

17: **end for**

$$UI = E(UI_z) = (E(UI_z(1)), E(UI_z(2)), \cdots, E(UI_z(v)), \cdots, E(UI_z(n)))$$

18: **return** UI

The Algorithm 1 displays the complete process of the MSI construction for all users in both the topic-level networks and the global-level network. As shown by Algorithm 1, the optimization objective constructed is a unimodal optimization problem, which can be solved by iterative algorithm in a finite number of iterations, and can always reach a steady state. Besides, the time complexity of Algorithm 1 is a^3mN , where a is the maximum degree of nodes, m is the number of topics, and N is the number of nodes, which demonstrates that our method is an efficient approach.

[3] Actually, among the above three evaluating metrics, the "influence spread" is the most important metric. According to [28, 58], the nodes which maximize the influence spread" have the most probability to be the best influential users, since marketers are far more interested in this metric. Therefore, in this paper, we focus more on the approach performance on the "influence spread" metric.

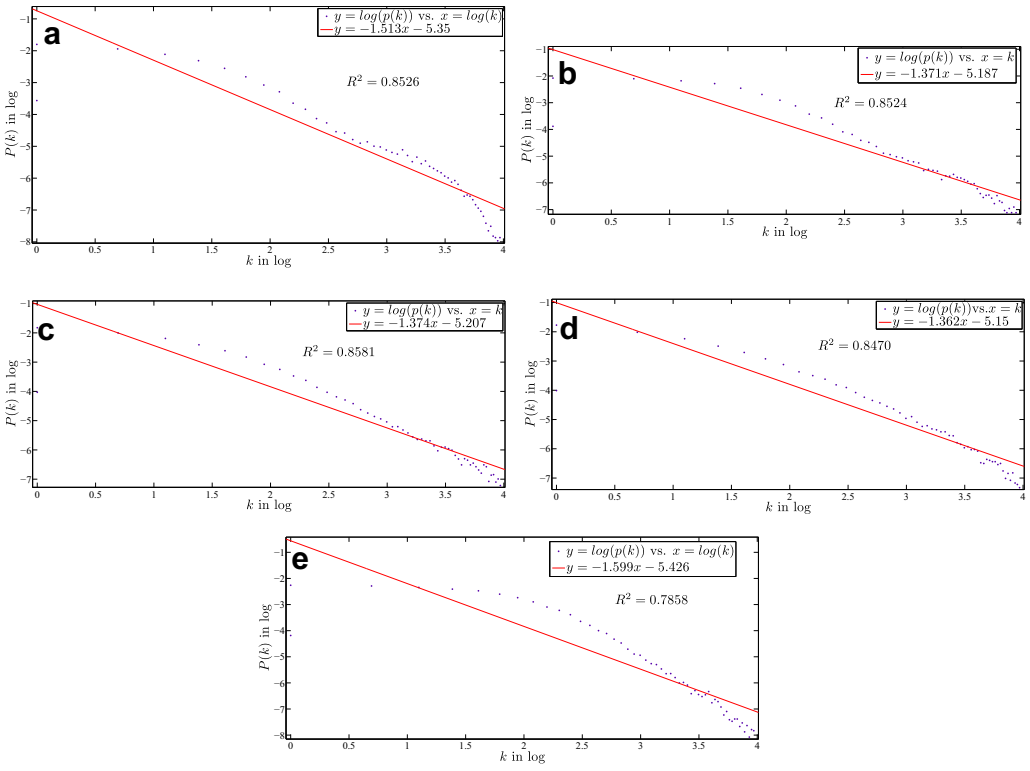


Figure 3. Degree distribution of topic networks. (a) My Old Classmate; (b) MI; (c) House Price; (d) Corrupt officials; (e) Smog.

4.1. Experiment design

To validate the proposed MSI approach, we compare its performance with the commonly used LeaderRank approach [6] and the FBI approach [7]. The Equation (25) and Equation (26) below show the main equations in LeaderRank and FBI, respectively.

Where $Inf^L(v,t)$ is the t_{th} iterative LeaderRank value of user v, g is the group node, ρ_{uv} is the edge weight of (u, v),$|V^{out}(u)|$ is the number of users which followed by user u, and t_c is the iterative end time of LeaderRank; $Inf_{initial}(v)$ is the initial FBI influence value of user v, A^d_{uv} and A^{id}_{uv} are the direct affinity and indirect affinity between user u and v, f_{vi} is the i_{th} factor value of user v.

$$\mathrm{Inf}^L(v,t) = \sum_{u \in V^{in}(u)} \frac{\rho_{uv}}{|V^{out}(u)|} Inf(u,t) \tag{25}$$

$$\begin{cases} FBI^{in}(v) = q \cdot A^d_{uv} + (1-q) \cdot A^{id}_{uv} \\ FBI^{out}(v) = \dfrac{\sum f_{v_i}}{i \in [1,5]} \\ FBI^{initial}(v) = p \cdot FBI^{out}(v) + (1-p) \cdot \sum_{u \in V^{in}(u)} path = (u,v) \end{cases} \tag{26}$$

1) Duplication percentage. A suitable detection approach should distinguish all users' influences to the greatest extent, in other words, less duplication means greater effectiveness [7]. Suppose ξ is the number of nodes with the same influence values, then the "duplication percentage", η, is defined as follows:

$$\eta = \begin{cases} 100\% & IF\ \xi = 0 \\ \dfrac{\xi}{N} \cdot 100\% & ELSE \end{cases}$$

2) Influence spread. This metric measures how many users can be influenced by k seed users. A larger influence spread means a higher level effectiveness [7, 54, 58]. To evaluate the performance of MSI, leaderRank and FBI, ranked top 100 users are chosen as seed nodes under the independent cascade (IC) model.
3) Ranking accuracy. A higher ranking accuracy indicates a more accurate approach. In this paper, we define rank accuracy as similarity to the official ranking:$\lambda = 1 - \frac{\sum_{v \in V} |\mathrm{rank}(c_v) - \mathrm{rank}(o_v)|}{\sum_{v \in V} |\mathrm{rank}(c_v) + \mathrm{rank}(o_v)|}$, where c_v refers to rank of v under the approach needed to be

Table 3. Maximum-likelihood fitting and K–S test on topic-level network.

Topic-level network	Alpha	X_{min}	L	p value of KS test	gof
My classmates	1.68	14	-60	**0**	0.5698
Mi PHONE	3.5	3	-2.6162e+03	**0**	0.5570
Smog	3.25	3	-3.2203e+03	**0**	0.5378
House Prices	2.49	9	-1.814e+3	**0**	0.6342
Corrupts network	3.5	3	-2.669e+3	**0**	0.5575

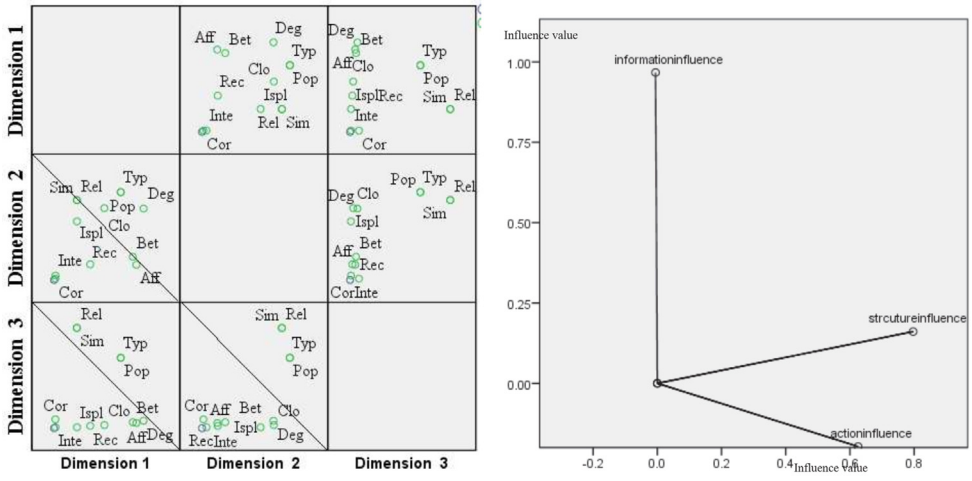
"My Old Classmate" is a 2014 Chinese drama-romance film.
"MI" is a kind of smartphone produced by Xiaomi company.

Table 4. Description of data set.

Topics	Number of nodes	Number of edges	Number of messages
My Old Classmate	54808	389783	10887
MI	57828	475560	11576
House Price	54465	440250	8936
Corrupt officials	58567	432749	6829
Smog	57918	454584	5946
Global-level	62444	1615111	40651

Table 5. The total variance explained of factors analysis for the 12 different dimensional factors.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	Variance	Cumulated	Total	Variance	Cumulated
1	4.361	36.338	36.338	4.361	36.338	36.338
2	3.064	25.534	61.872	3.064	25.534	61.872
3	1.832	15.265	77.137	1.832	15.265	77.137
4	1.279	10.655	87.792			
5	0.810	6.751	94.543			
6	0.462	3.847	98.390			
7	0.148	1.230	99.621			
8	0.026	0.216	99.837			
9	0.019	0.157	99.994			
10	0.001	0.006	100.00			
11	0.000	0.000	100.00			
12	0.000	0.000	100.00			



(a) Correlation analysis of the 12 factors (b) Correlation analysis of the dimensional influence

Figure 4. Factors correlation analysis ("Pop, Typ, Rel, Sim, Inte, Cor, Aff, Rec, Bet, Col, Deg and Ispl" are average values of corresponding factors in the global-level network.).

compared, and o_v represents the rank of v under the official ranking approach[4].

4.2. Data source

Due to Sina Weibo, the most popular microblogs in China, we successfully collected a dataset during the period from May 3rd, 2014 to May 11th, 2014. As described in Table 4, the data set contains 63,641 users, 1,391,718 follower relationships, 12 topics and 84,168 messages. Without loss of generality, five topics-"My Old Classmate Film", "MI Phone", "House Prices", "Corrupt officials" and "Smog" are chosen as examples among the 12 topics, which represent for "Entertainment", "Technology", "Economy", "Politics" and "Environment", respectively. Figure 3 manifests users' degree distributed to the power-law distribution with good fitness in log-log graphs. Besides, Table 3 reports the output of maximum-likelihood fitting (where suppose the degree of the network follows the distribution of power law: $p(x = k) = x^{-alpha}$ *for* $x \geq x_{min}$) and the goodness-of-fit test based on the Kolmogorov-Smirnov test (K–S test) [59], all the p value ($p < 0.05$) of K–S test all verify the power-law distribution of the five topic-level networks in our paper, which imply these topic-level networks are scale-free.

Te preprocessing of the data set includes three steps. Firstly, each of the five topic-level networks is retrieved from the global-level network

[4] Actually, in this paper we apply the ranking in "weiboreach.com" as the official ranking. Similar to "Klout", "http://www.weiboreach.com/" is an official social influence measurement website.

Table 6. The duplication percentages comparison of TMSI, FBI, and LeaderRank on topic-level network.

Topic	TMSI	FBI	LeaderRank
My classmate	0.06%	0.06%	66.79%
MI	0.05%	0.00%	60.35%
House Price	0.00%	0.00%	54.02%
Corrupt officials	0.00%	0.00%	44.90%
Smog	0.00%	0.00%	12.95%

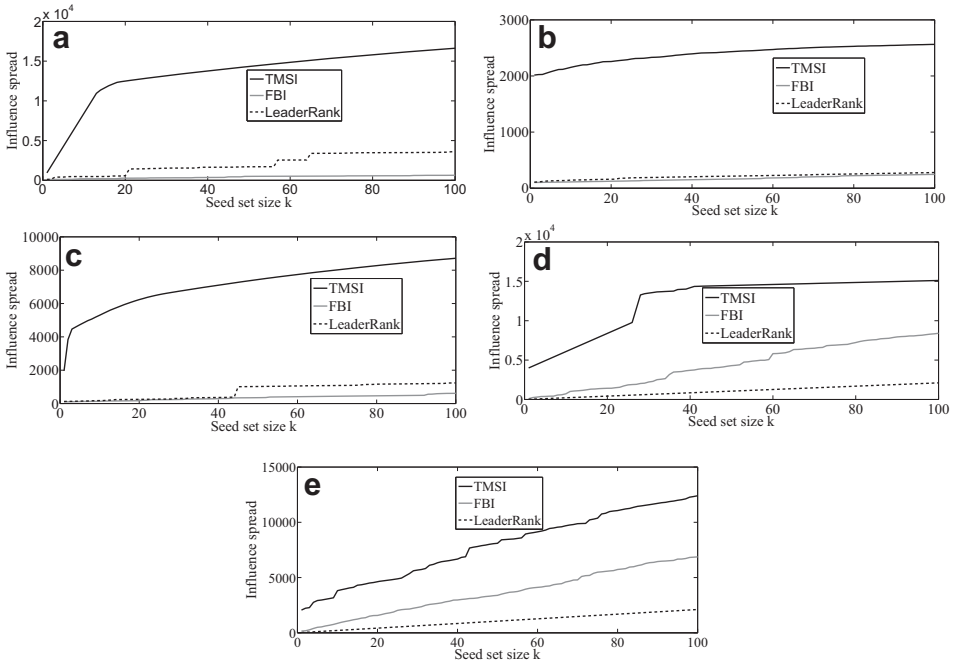


Figure 5. Top *k* influencers' accumulated influence spread comparison of TMSI, FBI and LeaderRank. (a) My Old Classmate; (b) MI; (c) House Price; (d) Corrupt Officials; (e) Smog.

with the topic distribution of the global-level network, based on the topic context contained (this extraction includes node, edge, messages and the attributes they contained). Secondly, irrelevant or abnormal data is excluded, normal data is standardized. Finally, for each topic-level network, factor set is obtained according to Section 3.1. Take information-based dimensional factors as an example, we first perform word segmentation for each user's messages using NLPIR (ICTCLAS2015)[5], and then establish a keyword matrix and calculate information-based dimensional factors with the keyword matrix.

4.3. Evaluation results

In this section, factors correlation analysis is first presented. Then, the experimental results of the TMSI model and GMSI model are reported.

4.3.1. The correlation analysis of factors

In Section 3, factors are classified into three dimensions-"Information", "Action" and "Structure". In order to support this classification, the results of "factor analysis" is first proposed in Table 4, in which, the cumulated variance can be explained 77.137 % of the total 12 factors by three components. Besides, the variance explained decreases monotonically with the number of components (as shown by the third column in Table 5). Therefore, it's relative reasonable to divide the 12 factors into three categories. Next, in order to support the definition of the three components (Information Influence, Action Influence, and Structure Influence). The results of "Multidimensional Scaling(MDS)" analysis is shown by Figure 4. Figure 4.a clearly show "Pop, Typ, Rel, Sim", "Inte, Cor, Aff, Rec, and "Bet, Col, Deg and Ispl" belong to dimension 3, dimension 1 and dimension 2, respectively. Therefore, we can define the dimension 1,2,3 as "Action dimension", "Structure dimension", and "Information dimension", respectively. Furthermore, the relation among the three dimensional influence is shown by the Figure 4.b. It clearly reveal the orthogonal relation among the three dimensional influence.

In summary, through the analysis of Figure 4, the 12 factors can be divided into "Information factors", "Action factors" and "Structure factors". Besides, The three dimensional influence are orthogonal and can not be replaced or explained by each other. Therefore, to construct a more reasonable influence measurement model, all of the three dimensional influence must be considered simultaneously. This conclusion can support the modelling of path influence in Section 3.2.2.

4.3.2. Comparison on topic-level networks

(1) Compared with FBI and LeaderRank approaches, the topic-level TMSI model performs better than LeaderRank on "duplication percentage" (Table 6), and slightly inferior to FBI's performance. Table 6 shows the duplication percentage of TMSI, FBI and LeaderRank for the five topic-level networks. As shown in this table, the "duplication percentage" value of TMSI is approximate to that of FBI and is much lower than that of LeaderRank. Besides, the "duplication percentage" of LeaderRank changes a lot on

[5] A classic Chinese word segmentation system at http://ictclas.nlpir.org/.

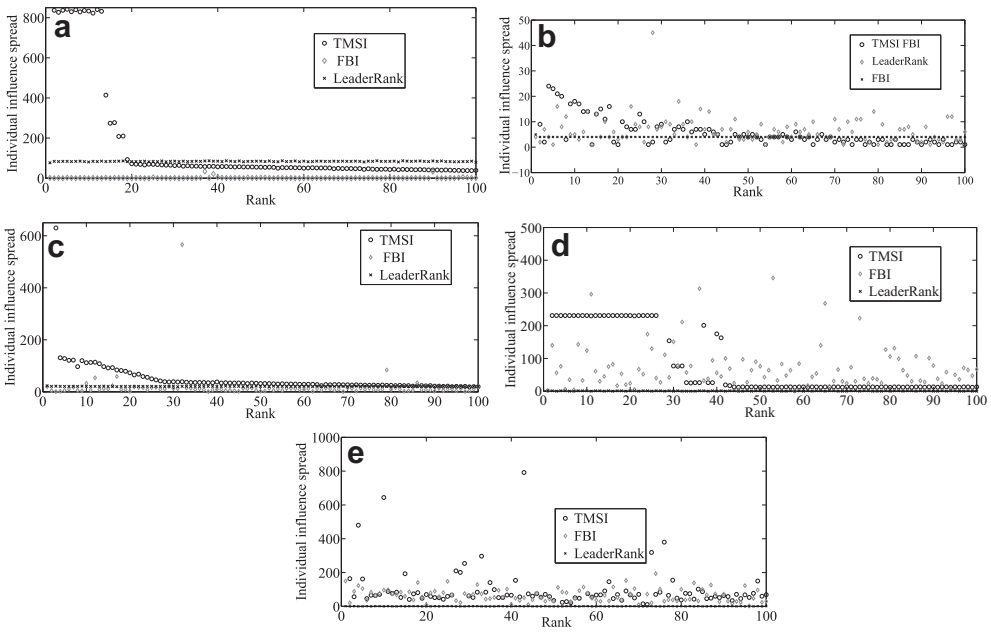


Figure 6. Top k influencers' individual influence spread comparison of TMSI, FBI and LeaderRank. (a) My Old Classmate; (b) MI; (c) House Price; (d) Corrupt Officials; (e) Smog.

different topic-level networks. Therefore, among the three approaches, TMSI and FBI approach are also the most stable two approaches on "duplication percentage", in comparison with LeaderRank approach.

(2) Compared with FBI and LeaderRank, the top k ($k = 1, 2, \cdots, 100$) influencers of TMSI achieve the greatest influence spread for each topic-level network.

As Section 4.1 noted, we apply an IC model to propagate influence. To conduct the experiment, the activation probabilities of each edge should first be determined. In this paper, we utilize pi(u, v), the path influence between node u and v, as the activation probability of TMSI (for FBI and LeaderRank approaches, we utilize the path(u, v) in Eq. (30) and $\frac{\sum vInf(v)}{N} rand()$, (respectively as the activation probability). To obtain the performance of each approach on the influence spread metric, we select the top k influencers as the seeds, and progressively compared the accumulated and individual influence spread of each approach as shown by Figures 5 and 6.

As shown in Figure 5, the top k TMSI influencers' accumulated influence spread is much larger than that of the other two approaches, on each of the five topic-level networks. Therefore, the TMSI approach demonstrates the best and most stable performance on the accumulated influence spread of the top k influencers.

Besides, intuitively, the individual influence spread of an effective user influence measurement approach should approximately be a decreasing function of the influencer's rank. As Figure 6 illustrates, for each topic-level network, the individual influence spread of TMSI is decreasing with the decreasing of the rank, from the overall trend. While the top k influencers' individual influence spread of LeaderRank and FBI approaches are similar to a constant function. Hence, TMSI is more effective and robust than the other two approaches with respect to influence spread. In summary, in terms of the most important metric, "Influence spread", TMSI model achieved the most strong and stable validation in measuring topic-level social influence compared with the FBI and LeaderRank. And in terms of the "duplication percentage", the proposed TMSI performs better than LeaderRank and similar to the FBI on most topic-level networks.

4.3.3. Comparison on the global-level network

To validate the effectiveness and accuracy of global-level MSI measurement approach, we compare it with the FBI and LeaderRank approaches based on the three metrics in this section.

The results are as follows:

(1) With respect to the "duplication percentage", MSI performs better than LeaderRank, while slightly inferior to the FBI. However, this inferior is not significant, since the zombie users 7 are count for almost 17%.

Next, we compared the "duplication percentage" performance of MSI, LeaderRank and FBI on the global-level. As the table indicates, on the global-level network, the duplication percentage under the MSI is 44.91%, which is much less than the ratios of LeaderRank (62.37%), but inferior to the FBI (32.06%).

In order to dig out the reason of MSI's inferior performance on "duplication percentage" metric, we randomly choose 500 users as examples, manually view their homepage on Sinaweibo, and find out there are almost 17% zombie users, who barely have any action after their signing-up and following [60]. These zombie users only contribute on structure-based factors, and barely contribute on information-based factors or action-based factors. Since MSI approach is based on the combination of structure-based factors, information-based factors and action-based factors, while FBI approach is merely based on structure-based factors, these zombie users' FBI values are different, and their TMSI values are almost the same. Therefore, there is no significant difference between the MSI's performance and FBI′ performance on the "duplication percentage", with the existence of zombie users.

(2) In a comparison of the "influence spread", the MSI outperforms the other two approaches (FBI and LeaderRank).

Figure 7.a displays the top k ($k = 1, 2, \cdots, 100$) influencers' accumulated and individual "influence spread" under the MSI, FBI and leaderRank, respectively. As shown by Figure 7, the top k influencers' "accumulated influence spread" of MSI is much more than the other two

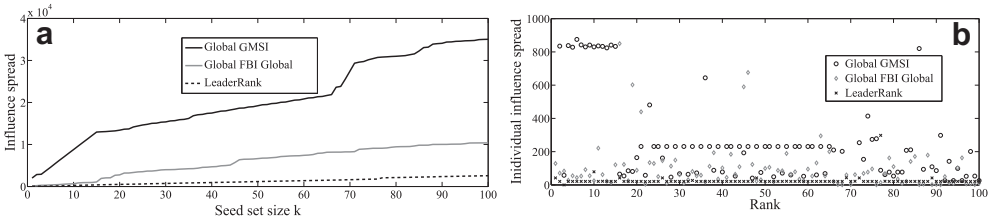


Figure 7. Top *k* influencers' influence spread comparison of TMSI, FBI, and LeaderRank on the global-level network. (a) accumulated influence spread; (b) individual influence spread.

approaches on the global-level network. Besides, as the Figure 7.b shown, the top *k* influencers' "individual influence spread" of MSI manifests a decline trend, while the top *k* influencers' "individual influence spread" of other two approaches manifests a constant trend. The results illustrate that the global-level MSI model can identify the most influential top k influencers. Therefore, in comparison with FBI, and LeaderRank, MSI is the most effective on the global-level network based on the "influence spread" metric.

(3) MSI performs the best in "ranking accuracy", comparing to FBI and LeaderRank.

In order to manifest the measurement accuracy of the proposed MSI on the global-level network, the ranking accuracy is computed according to Eq. (4.1). The results imply that our MSI model (71.89%) outperforms FBI (66.17%), LeaderRank (67.61%).

In summary, synthesizing the analyses of Section 4.3.2, we can draw the following conclusions: 1) in terms of the most important metric,"influence spread", the proposed MSI achieved the most strong and stable validation than the FBI and LeaderRank, both in the topic-level networks and the global-level network; 2) in terms of the "duplication percentage", the proposed TMSI performs better than LeaderRank and similar to the FBI on most topic-level networks; 3) in terms of the "ranking accuracy", the global-level TMSI model achieved the greatest accuracy on the extended global-level network. Therefore, MSI is the most effective and accurate approach, in contrast with LeaderRank and FBI, both on the topic-level networks and the global-level network, combining their performance on the "influence spread", the "duplication percentage" and the "ranking accuracy" metrics.

5. Conclusion and discussion

Information-based, action-based and structure-based metrics are common predictors for identifying influencers. However, a single dimensional method might not accurate. Besides, prior studies [3, 61, 62, 63, 64] have noted the importance of multidimensional factors (attributes) analysis in influencer detection of online social networks, however, very little was found in the literature on the modeling of multidimensional influence. This study set out with the aim of assessing the influence of users in online social networks by multidimensional factors analysis and multidimensional influence modeling. In detail, we construct a comprehensive approach in which three dimensions of factors, information-based, action-based and structure-based, are considered to measure users' social influence and identify influences in OSNs. Capturing the effect of topics on OSNs, MSI is developed to measure both users' topic-level and global-level social influence, based on the knowledge of topic distillation and distribution. The experimental results utilizing a real network data set collected from Sina Weibo (Section 4), demonstrate that in terms of "influence spread", "duplication percentage" and "ranking accuracy", our proposed methods achieved a better performance compared to LeaderRank and FBI, both on the topic-level networks and the global-level network.

LeaderRank is a improved structural method based on PageRank, and similar to our method FBI approach is based on factor analysis and modeling, what's different in FBI is the feature they extracted are solely structural. Therefore, a potential advantage of our method in terms of accuracy is that our method integrating 12 factors from three different dimension. Besides, by the entropy weighted modeling of the three dimensional metrics, the computation time complexity of MSI is close to the other two method, which is less than most node operation methods and machine learning methods. Therefore, MSI can stably find influential nodes with an acceptable computational complexity and effectiveness in large scale online social networks.

From a practical perspective, by identifying influencers, marketers could take them as the marketing nodes and launch marketing strategies to enhance the effectiveness and availability of viral marketing. Besides, influencers can be particularly influential in encouraging the trial and adoption of novel products and services. Therefore, the proposed MSI approach can be applied to improve word-of-mouth marketing, to steer public opinion in certain directions, even to support decisions during a negotiation process. Besides, by factors exploration, this study allows us to understand better the social activities taking place in OSNs, and providing an effective strategy to improve the reputation/influence of OSN users.

However, several avenues remain to be explored in future research. First, the TMSI model and GMSI model we presented are only static descriptive models rather than dynamic predictive models. Predicting social influence at a given time will require further examination. Second, it would be interesting to investigate how negative or positive attitudes affect users' social influence. Third, trust between users also plays a significant role in social influence diffusion. For example, if thousands of spammers are spreading spam information on OSNs, normal information diffusion will be disturbed, and user's social influence will undoubtedly be affected. Finally, the proposed MSI can only be utilized in online social networks, where "visiting, commenting, tweeting, mentioning" is their basic characteristics, and not appropriate for offline networks (such as bank network). Therefore how to extend it to explore influencers in large scale social networks, like bank network, is still a great challenge.

Declarations

Author contribution statement

Yun-Bei Zhuang: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Zhi-Hong Li: Conceived and designed the experiments.

Yun-Jing Zhuang: Performed the experiments; Wrote the paper.

Funding statement

This work was supported by Natural Science Foundation of Shandong Province (Grant NO. ZR2019QG002) and the National Natural Science Foundation of China (Grant No. 71801145).

Data availability statement

Data included in article.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

[1] Peng Luo, Kun Chen, Chong Wu, Measuring social influence for firm-level financial performance, Electron. Commer. Res. Appl. 20 (2016) 15–29.
[2] JJ Chen Yun-Bei Zhuang, Zhi-hong Li, Modeling the cooperative and competitive contagions in online social networks, Phys. Stat. Mech. Appl. 484 (2017) 141–151.
[3] Linyuan Lü, Duanbing Chen, Xiao Long Ren, Qian Ming Zhang, Yi Cheng Zhang, Tao Zhou, Vital nodes identification in complex networks, Phys. Rep. 650 (2016) 1–63.
[4] Yu Yang, Zhefeng Wang, Jian Pei, Enhong Chen, Tracking influential individuals in dynamic networks, IEEE Trans. Knowl. Data Eng. 29 (11) (2017) 2615–2628.
[5] Siwar Jendoubi, Arnaud Martin, Ludovic Liétard, Hend Ben Hadji, Boutheina Ben Yaghlane, Two evidential data based models for influence maximization in twitter, Knowl. Base Syst. 121 (2017) 58–70.
[6] Linyuan Lü, Yi-Cheng Zhang, Chi Ho Yeung, Tao Zhou, Leaders in social networks, the delicious case, PloS One 6 (6) (2011), e21202.
[7] Guojun Wang, Wenjun Jiang, Jie Wu, Zhengli Xiong, Fine-grained feature-based social influence evaluation in online social networks, IEEE Trans. Parallel Distr. Syst. 25 (9) (2014) 2286–2296.
[8] Jun zhang, Research on the Mechanism of Time-Varying Evolution of Internet Public Opinion and It's Countermeasures, China Social Sciences Press, Beijing, 2020.
[9] I.A. Kovács, A.L. Barabási, Network science: destruction perfected, Nature 524 (7563) (2015) 38–39.
[10] Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H Eugene Stanley, Hernán A. Makse, Identification of influential spreaders in complex networks, Nat. Phys. 6 (11) (2010) 888–893.
[11] Flaviano Morone, Gino Del Ferraro, Hernán A. Makse, The k-core as a predictor of structural collapse in mutualistic ecosystems, Nat. Phys. 15 (1) (2019) 95–102.
[12] Linyuan Lü, Tao Zhou, Qian-Ming Zhang, H Eugene Stanley, The h-index of a network node and its relation to degree and coreness, Nat. Commun. 7 (1) (2016) 1–7.
[13] Tianlong Fan, Linyuan Lü, Dinghua Shi, Towards the Cycle Structures in Complex Network: A New Perspective. arXiv Preprint arXiv:1903.01397, 2019.
[14] Xiang Xu, Cheng Zhu, Qingyong Wang, Xianqiang Zhu, Yun Zhou, Identifying vital nodes in complex networks by adjacency information entropy, Sci. Rep. 10 (1) (2020) 1–12.
[15] Zhe Li, Tao Ren, Xiaoqi Ma, Simiao Liu, Yixin Zhang, Tao Zhou, Identifying influential spreaders by gravity model, Sci. Rep. 9 (1) (2019) 1–7.
[16] Lawrence Page, Sergey Brin, Rajeev Motwani, Winograd Terry, The PageRank Citation Ranking: Bringing Order to the Web, Technical report, Stanford InfoLab, 1999.
[17] Jon M. Kleinberg, Authoritative sources in a hyperlinked environment, J. ACM 46 (5) (1998) 604–632.
[18] Alexis Arnaudon, Robert L. Peach, Mauricio Barahona, Graph Centrality Is a Question of Scale. arXiv Preprint arXiv:1907.08624, 2019.
[19] Haewoon Kwak, Changhyun Lee, Hosung Park, Sue Moon, What is Twitter, a social network or a news media ?, in: Proceedings of the 19th International Conference on World Wide Web, 2010, pp. 591–600.
[20] R. Albert, H. Jeong, A.L. Barabasi, Error and attack tolerance of complex networks, Nature 406 (6794) (2000) 378.
[21] Xin Chen, Critical nodes identification in complex systems, Comp. Int. Syst. 1 (1-4) (2015) 37–56.
[22] Cai Bao Xue, Sheng Dai Fu, Wei Zhan Han, Evaluation method of network invulnerability based on disjoint paths in topology, Syst. Eng. Electron. 34 (1) (2012) 168–174.
[23] Duanbing Chen, Linyuan Lü, Ming-Sheng Shang, Yi-Cheng Zhang, Tao Zhou, Identifying influential nodes in complex networks, Phys. Stat. Mech. Appl. 391 (4) (2012) 1777–1787.
[24] Suppawong Tuarob, Conrad Tucker, Automated discovery of lead users and latent product features by mining large scale social media networks, J. Mech. Des. 137 (7) (2015) 1–13.
[25] Changjun Fan, Li Zeng, Yizhou Sun, Yang-Yu Liu, Finding key players in complex networks through deep reinforcement learning, Nat. Mach. Int. 2 (6) (2020) 317–324.
[26] Jiezhong Qiu, Jian Tang, Hao Ma, Yuxiao Dong, Kuansan Wang, Jie Tang, Deepinf: social influence prediction with deep learning, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 2110–2119.
[27] Sanjin Pajo, Dennis Vandevenne, R. Joost, Duflou, Automated feature extraction from social media for systematic lead user identification, Technol. Anal. Strat. Manag. 29 (6) (2016) 1–13.
[28] I. Roelens, P. Baecke, D.F. Benoit, Identifying influencers in a social network: the value of real referral data, Decis. Support Syst. 91 (2016) 25–36.
[29] Chanhyun Kang, Sarit Kraus, Cristian Molinaro, Francesca Spezzano, V.S. Subrahmanian, Diffusion centrality: a paradigm to maximize spread in social networks, Artif. Intell. 239 (2016) 70–96.
[30] Jain Gu, Sungmin Lee, Jari Saramäki, Petter Holme, Ranking influential spreaders is an ill-defined problem, EPL (Europhysics Letters) 118 (6) (2017) 68002.
[31] Mile Sikić, Alen Lančić, Nino Antulov-Fantulin, Hrvoje Stefančić, Epidemic centrality -is there an underestimated epidemic impact of network peripheral nodes? Eur. Phys. J. B 86 (10) (2013) 1–13.
[32] Linton C. Freeman, Centrality in social networks conceptual clarification, Soc. Network. 1 (3) (1979) 215–239.
[33] Linton C. Freeman, Stephen P. Borgatti, Douglas R. White, Centrality in valued graphs: a measure of betweenness based on network flow, Soc. Network. 13 (2) (1991) 141–154.
[34] W. U. Jun Yue Jin Tan, Hong Zhong Deng, Evaluation method for node importance based on node contraction in complex networks, Syst. Eng. Theory Prac. 11 (2006) 79–101.
[35] Sanjin Pajo, Paul Armand Verhaegen, Dennis Vandevenne, R. Joost, Duflou, Fast lead user identification framework, Proc. Eng. 131 (2015) 1140–1145.
[36] Feng Li, Timon C. Du, Listen to me | evaluating the influence of micro-blogs, Decis. Support Syst. 62 (2) (2014) 119–130.
[37] Nicola Barbieri, Francesco Bonchi, Giuseppe Manco, Topic-aware social influence propagation models, Knowl. Inf. Syst. 37 (3) (2013) 555–584.
[38] Chuan Hu, Huiping Cao, Aspect-level influence discovery from graphs, IEEE Trans. Knowl. Data Eng. 28 (7) (2016) 1635–1649.
[39] Dawei Zhao, Lixiang Li, Shudong Li, Yujia Huo, Yixian Yang, Identifying influential spreaders in interconnected networks, Phys. Scripta 89 (1) (2014), 015203.
[40] F. Liberatore, L. Quijano-Sanchez, What do we really need to compute the tie strength? An empirical study applied to social networks, Comput. Commun. 110 (2017) 59–74.
[41] Chunxiao Jiang, Yan Chen, KJ Ray Liu, Evolutionary dynamics of information diffusion over social networks, IEEE Trans. Signal Process. 62 (17) (2014) 4573–4586.
[42] Mark S. Granovetter, The strength of weak ties, Am. J. Sociol. 78 (2) (1973) 105–130.
[43] T.Q. Phan, X. Chen, R. van der Lans, Uncovering the importance of relationship characteristics in social networks: implications for seeding strategies, J. Market. Res. 54 (2) (2017) 187–201.
[44] Johannes Stauder, The Social Structure of Opportunities for Contact and Interaction and Strategies for Analysing Friendship Networks, Springer Fachmedien Wiesbaden, 2014.
[45] Yen Liang Chen, Kwei Tang, Chia Chi Wu, Ru Yun Jheng, Predicting the influence of users' posted information for ewom advertising in social networks, Electron. Commer. Res. Appl. 13 (6) (2014) 431–439.
[46] Wanqiu Guan, Haoyu Gao, Mingmin Yang, Li Yuan, Haixin Ma, Weining Qian, Zhigang Cao, Xiaoguang Yang, Analyzing user behavior of the micro-blogging website Sina Weibo during hot social events, Phys. Stat. Mech. Appl. 395 (2014) 340–351.
[47] Yuchi Zhang, Wendy W. Moe, David A. Schweidel, Modeling the role of message content and influencers in social media rebroadcasting, Int. J. Res. Market. 34 (1) (2016).
[48] Henri Tajfel, John C. Turner, The Social Identity Theory of Intergroup Behavior, 2004.
[49] Daniel J. Brass, A social network perspective on human resources management, Res. Person. Hum. Resour. Manag. 13 (1) (1995) 39–79.
[50] Miller McPherson, Lynn Smith-Lovin, James M. Cook, Birds of a feather: homophily in social networks, Annu. Rev. Sociol. 27 (1) (2001) 415–444.
[51] Yung-Ming Li, Cheng-Yang Lai, Ching-Wen Chen, Discovering influencers for marketing in the blogosphere, Inf. Sci. 181 (23) (2011) 5143–5157.
[52] C. Kadushin, R.D. Alba, The intersection of social circles: a new measure of social proximity in networks, Socio. Methods Res. 5 (1) (1976) 77–102.
[53] Karen S. Cook, Richard M. Emerson, Power, equity and commitment in exchange networks, Am. Socio. Rev. 43 (5) (1978) 721–739.
[54] Sancheng Peng, Aimin Yang, Lihong Cao, Shui Yu, Dongqing Xie, Social influence modeling using information theory in mobile social networks, Inf. Sci. 379 (2016) 146–159.
[55] Bibb Latane, The psychology of social impact, Am. Psychol. 36 (4) (1981) 343.
[56] C.E. Shannon, A mathematical theory of communication, Bell Labs Tech. J. 5 (4) (1948) 3–55.
[57] Thomas Hofmann, Probabilistic latent semantic indexing, in: International ACM SIGIR Conference on Research and Development in Information Retrieval, 2004, pp. 56–73.
[58] Domingos Pedro, Matt Richardson, Mining the network value of customers, in: Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2001, pp. 57–66.
[59] Yogesh Virkar, Aaron Clauset, Power-law distributions in empirical data, SIAM Rev. 51 (4) (2009) 661–703.
[60] Hongfu Liu, Yuchao Zhang, Hao Lin, Junjie Wu, Zhiang Wu, Xu Zhang, How many zombies around you?, in: Data Mining (ICDM), 2013 IEEE 13th International Conference on IEEE, 2013, pp. 1133–1138.
[61] Chunhua Ju, Wanqiong Tao, A novel relationship strength model for online social networks, Multimed. Tool. Appl. 76 (16) (2017) 17577–17594.
[62] Shixi Liu, Cuiqing Jiang, Zhangxi Lin, Yong Ding, Rui Duan, Zhicai Xu, Identifying effective influencers based on trust for electronic word-of-mouth marketing: a domain-aware approach, Inf. Sci. 306 (2015) 34–52.
[63] Yun-Bei Zhuang, User Influence Evaluation in Online Social Networks Considering Noise Existence, Economics and Management Press, Beijing, 2020.
[64] Jian-Guo Liu, Jian-Hong Lin, Qiang Guo, Tao Zhou, Locating influential nodes via dynamicssensitive centrality, Sci. Rep. 6 (1) (2016) 1–8.