



How to find opinion leader on the online social network?

Bailu Jin¹ · Mengbang Zou¹ · Zhuangkun Wei¹ · Weisi Guo¹

Accepted: 28 March 2025
© The Author(s) 2025

Abstract

Online social networks (OSNs) provide a platform for individuals to share information, exchange ideas, and build social connections beyond in-person interactions. For a specific topic or community, opinion leaders are individuals who have a significant influence on others' opinions. Detecting opinion leaders and modeling influence dynamics is crucial as they play a vital role in shaping public opinion and driving conversations. Existing research have extensively explored various graph-based and psychology-based methods for detecting opinion leaders, but there is a lack of cross-disciplinary consensus between definitions and methods. For example, node centrality in graph theory does not necessarily align with the opinion leader concepts in social psychology. This review paper aims to address this multi-disciplinary research area by introducing and connecting the diverse methodologies for identifying influential nodes. The key novelty is to review connections and cross-compare different multi-disciplinary approaches that have origins in: social theory, graph theory, compressed sensing theory, and control theory. Our first contribution is to develop cross-disciplinary discussion on how they tell a different tale of networked influence. Our second contribution is to propose trans-disciplinary research method on embedding socio-physical influence models into graph signal analysis. We showcase inter- and trans-disciplinary methods through a Twitter case study to compare their performance and elucidate the research progression with relation to psychology theory. We hope the comparative analysis can inspire further research in this cross-disciplinary area.

Keywords Online social network · Social influence · Influence analysis · Influential user detection

1 Introduction

As far back as the 1940s, Paul F. Lazarsfeld, Bernard Berelson, and Hazel Gaudet conducted social influence experiments to understand the social network opinion dynamics towards a topic [1]. As a part of their research, *opinion leaders* were defined as individuals with a significant impact on the opinions, attitudes, and behavior of others. These studies were typically conducted in relatively small social circles.

Fast track to modern day, the rise of online social networks (OSNs) has seen a rapid expansion in social network size and the role of opinion leaders has become increasingly crucial in shaping public opinion and driving online discourse. It is widely recognized that developing algorithms that can detect opinion leaders is crucial. There are a range of application areas in business intelligence, social monitoring the spread of (mis)information and mitigating the negative impact on public discourse [2].

Over the past few decades, empirical research in psychology has explored the phenomenon of opinion evolution during interpersonal interactions. Studies have shown that people tend to modify their opinions to seek similarity with others in the group, highlighting the high interdependence of individual opinions. The combined effects of the influences from cultural norms, mass media and interactions are collectively known as social influence. The concept of opinion leader was first introduced in the hypothesis of *two-step flow of communication* [1]. It posited that the influence from mass media first reaches opinion leaders, who subsequently disseminate it to their followers or associates.

✉ Weisi Guo
weisi.guo@cranfield.ac.uk

Bailu Jin
bailu.jin@cranfield.ac.uk

Mengbang Zou
m.zou@cranfield.ac.uk

Zhuangkun Wei
zhuangkun.wei@cranfield.ac.uk

¹ Cranfield University, College Rd, Cranfield, Wharley End, Bedford MK43 0AL, UK

In recent years, numerous review papers have discussed the related research topics. Riquelme et al. provided an extensive survey on activity, popularity and influence measures that rank influential users in Twitter network [3]. Bamakan et al. categorised the characteristics of opinion leaders and the approaches for opinion leader detection [2]. A great deal of existing work focus on proxies for opinion-leaders which is to see how information diffuses on the social network statistically, without checking for: (i) whether this information has influence for a topic, and (ii) how is influence actually exerted and by whom. Part of the challenge is the lack of well labeled data sets (need to label topic-specific influence) of sufficient size across diverse topics. Panchendrarajan et al. conducted a comprehensive survey on topic-based influential user detection [4]. However, there is a lack of consensus between definitions and methods of what constitutes a holistic view of opinion leader across disciplines. In contrast to previous reviews, this review paper focuses on identifying influential nodes in OSNs and providing a cross-disciplinary definition of opinion leaders in relation to social psychology foundational knowledge.

In this paper, we categorise the opinion identification methods into four main categories, *Topology-based Centrality*, *Topic-sensitive Centrality*, *Control- and Sampling-based Centrality*. These categories define opinion leaders in distinct ways and ingest different data features. Topology-based centrality mainly concentrates on the network structure. In this context, opinion leaders are defined as individuals who occupy the most significant position within the social group. When user semantic content is taken into consideration, the Topic-Sensitive Centrality facilitates the identification of opinion leaders within specific topics. This approach helps identify influential users capable of disseminating topic-related information and influencing opinions within specific contexts. Additionally, real-time content can be utilised as

a representation of the dynamic opinion states of users, which can be used to build a mathematical model to describe the evolution of opinion states. Leveraging the dynamic influence model, control methodologies aim to identify individuals who can steer the direction of overall opinion. Finally, graph sampling methodologies focus on identifying a specific subset of opinion leaders who, despite their limited numbers, can be instrumental in reconstructing the comprehensive opinion network. As illustrated in Fig. 1, we show these four methodologies, and then go on in rest of paper to offer a deeper understanding of each method.

In Table 1, we provide the notations that are used in this paper.

2 Background

2.1 Opinion leader definition in social science

The concept of an 'opinion leader' in online social media, originally derived from social theory, plays a pivotal role in understanding the dynamics of digital communication networks. This notion aligns closely with the node influence metrics from graph theory. Such applications are crucial for gauging the influence of individuals in social networks and identifying key infrastructure nodes in transportation networks. However, since the early 2000s, social scientists have raised concerns about the adequacy of centrality indices in fully capturing the nuances of node influence in these contexts.

To clarify the distinctions between graph theory and social theory in defining 'opinion leader', we present an overview of the evolution of this concept within social theory. We categorize the definition of opinion leaders based on three main characteristics: spread, impact, and representativeness.

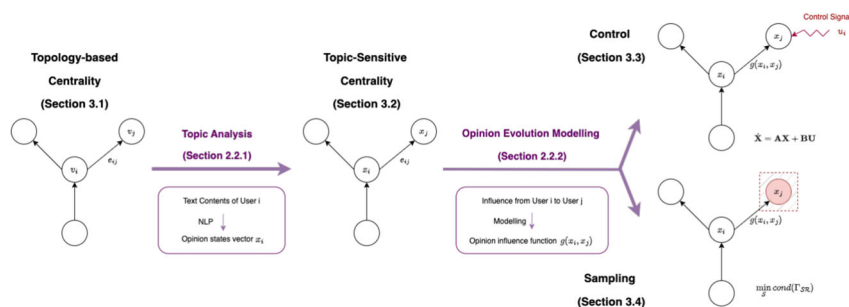


Fig. 1 Relationship Among Four Methodologies. The Topology-based Centrality methodologies (3.1) focus on the graphical structure of the network. Real-time content of user i can be represented as an opinion states vector denoted as x_i using Natural Language Processing (NLP). By incorporating Topical Analysis, the Topic-Sensitive Centrality methods (3.2) integrate topical information with the graph structure to identify opinion leaders within specific topics. In opinion evolution

modelling, the influence from User i to User j can be modelled as a function $g(x_i, x_j)$. Given the dynamic influence model, the Control methodologies (3.3) aim to find individuals who have the ability to steer the opinion direction by considering the control signals u . The Graph Sampling methodologies (3.4) seek to identify specific users who can accurately reconstruct the entire opinion network

Table 1 Notations Used In This Paper

Variable	Definition	Description
\mathcal{G}	A graph of a network	OSN represented as a graph with users as nodes and social connections as edges
\mathcal{V}	The set of nodes in the graph	The set of users in the OSN graph
\mathcal{E}	The set of edges in the graph	The set of connection relationships in the OSN graph
v_i	An arbitrary node (e.g. social account) in the graph $v_i \in \mathcal{V}$	User i in the OSN
$e_{i,j}$	An arbitrary link in the graph $e_{i,j} \in \mathcal{E}$	Social connection(e.g. follow) from User i to User j
N	Number of nodes in the graph	The number of users in the OSN graph
\mathbf{A}	The adjacency matrix of the graph	Matrix representation of the OSN network
$a_{i,j}$	The (i, j) th entry of the matrix \mathbf{A}	The presence or absence state of a connection from user i to user j
$x_i(t)$	The opinion state of user i at time t	Content posted by user i at time t transformed into numerical value $x_i(t)$
x_i	The opinion states vector of user i	Collection of opinions $x_i(t)$ integrated into a vector x_i
X	The opinion states matrix of all users	Each column represents the opinion states vector of a user
$f(\cdot)$	The self opinion dynamic function	It describes self opinion dynamic over time
$g(\cdot)$	The opinion influence function	It describes the influence between connected nodes
$\mathbf{u}(t)$	The control signals at time t	Input signals imposed on nodes to control the system.
\mathbf{B}	An input matrix	It identifies the nodes that are controlled by the input vector $\mathbf{u}(t)$
\mathbf{C}	The controllability matrix	The dimension of the controllable subspace of the system is given by the rank of \mathbf{C}
$C(i)$	The exact control centrality of node i	The dimension of controllable subspace of node i with exact values in \mathbf{A} and \mathbf{B}
$C_g(i)$	The structure control centrality of node i	The maximum dimension of controllable subspace of node i for varying free parameters.
Γ	The orthogonal subspace of graph signal	Determine the orthogonal subspace of time-varying graph signal.
\mathcal{R}	The graph frequency set	Represent the bandwidth of the networked dynamics to the orthogonal subspace.
S	The sampling node set	Critical node set that ensure the complete recovery of networked dynamics.

The spread, the foremost attribute, originates from the theory of “the two-step flow of communication” [1, 5], highlighting that information from mass media first reach opinion leaders who then disseminate it to less active segments. The impact dimension is also underscored in various studies, portraying opinion leaders as individuals capable of affecting others’ opinions, attitudes and behaviors in an appropriate way [6, 7]. Representative, as the third characteristic, positions opinion leaders as trusted and influential within their groups, reflecting collective viewpoints [8].

In the context of online social media, the role of opinion leaders has evolved beyond traditional communication models. Presently, they are recognized as individuals who significantly influence others’ opinions through online interactions. These leaders are pivotal in various sectors, including marketing, political science, and public health, effectively shaping public opinion [2].

2.2 Technical background

2.2.1 Topic analysis

Topic analysis involves the utilization of natural language processing (NLP) to detect the topic-related semantic struc-

tures from human language. In this paper, we mainly employ two types of topic analysis: topic modelling and opinion representation [9]. Topic modelling utilizes statistical modelling approaches to assign topic probability distributions to user-generated content. On the other hand, opinion representation is a task of classifying the content into opinion state vectors associated with specific topics.

2.2.2 Opinion evolution modelling

As the effect of “word-of-mouth”, people are likely to be influenced by the idea of their friends in the process of agricultural innovation, adoption of medical, and new product promotion. To explain how individuals develop their opinions towards various topics over time, a formal model of the opinion evolution in a group was proposed by French in 1956 [10]. In French’s formal theory, the discrepancy of opinions x_i^t and x_j^t determines the effect from influencer j to recipient i . So the influence effect is determined to be proportional to the size of the difference between their opinions $g(x_j^t, x_i^t) = (x_j^t - x_i^t)$. Beyond the function, there may include influence weights (w_{ij}) representing the strength of the effect. Formally, social pressure on the recipient i is the

sum of the effect from all influencers j conditioned by the weight (w_{ij}) of the tie between i and j . The self-weight (w_{ii}) of the recipient i represent to what degree the recipient is anchored on his previous position ($-1.0 \leq w_{ii} \leq 1.0$) [11]. The influence process takes place gradually, as the influencer changes its position over time and influences the recipient toward its position. For each recipient, the discrete-time interpersonal influence mechanism can be describe as a ordinary differential equation

$$x_i^{t+1} = w_{ii}x_i^t + \sum_{j, j \neq i} w_{ij}(x_j^t - x_i^t) \quad (1)$$

From there, social influence models were developed to explain social phenomena such as opinion clustering or controversy. To capture the complexity of opinion evolution, researchers have considered both linear and non-linear models. One example of a linear model is the French-DeGroot model, which introduced a more general form using Markov Chain processes to illustrate how social influence leads to opinion consensus [12]. However, opinion consensus is not the only outcome from group discussions [13]. Non-linear models, such as the Hegselmann-Krause model, have been proposed to incorporate a bounded confidence attribute that limits the influence of opposing opinions [14].

In general, the following (2) represents broader dynamic in both linear and Non-linear models. Here, $\dot{x}(i)$ denotes the rate of change of opinions for agent i , A symbolizes the graph structure, $f_i(x_i)$ represents the self dynamic, $g(x_i, x_j)$ indicates the influence from agent j to agent i , and u_i represents the external social signal influencing agent i .

$$\dot{x}_i = f_i(x_i) + \sum_{j/neqi}^n g(x_i, x_j). \quad (2)$$

3 Methodology

The definition of being influential point is ambiguous, leading to the development of various measures for identifying opinion leaders. Here, we categorise detection methods into four groups: topology-based centrality, topic-sensitive centrality, control and graph sampling.

Centrality is the most commonly employed method, operating under the assumption that opinion leaders are structurally important nodes within a social network. We introduce three traditional measures - Degree, Closeness and Betweenness - and explore their application in the context of online social network. Beyond considering the individual status of neighbours, we delve into a group of eigenvector-based centralities such as Eigenvector, Katz, PageRank. PageRank, for instance, is widely used in topic-sensitive centralities to

fit various social network assumptions. In addition to these measures, we also introduce two dynamical measures - Maximization and SIR - which account for the dynamic states of nodes.

Recognising that the influence of these leaders can fluctuate based on various topic field, the topic-sensitive centrality approaches incorporate topical attributes into the analysis. For instance, topic analysis can be employed to determine the novelty and similarity of content on OSN, which lead to InfluenceRank, TwitterRank, TopicSimilarRank, and ClusterRank. Simultaneously, the dynamics of opinion are analysed based on a particular topic, leading to the creation of OpinionRank, Dynamic OpinionRank, TrustRank and InfluenceModellingRank.

Given the opinion dynamic modelling, social influence can be analysed via control and sampling methodologies. Control methodologies identify influential nodes by assessing their ability to influence the states of others in the network.

Another approach is graph sampling, which determines the most influential nodes by studying whether the samples on these nodes can ensure the complete recovery of the whole networked dynamics. This is generally achieved by determining an orthogonal subspace of the networked opinion vector, and then evaluating the importance mapping from the orthogonal basis to the node set. This section reviews how to build the orthogonal subspace from the spatial and spatial & temporal dynamics correlation, and the graph sampling-based ranking strategies.

3.1 Topology-based centrality

Centrality in graph theory and network analysis is a fundamental concept that refers to the importance of a node within a network. In the context of social network, centrality measures help identify users who have extensive connections with other members of a network. Bavelas first introduce the idea of centrality to human communication network, aiming to explain the influence in group processes [15]. Bavelas proposed that an individual strategically positioned on the shortest communication path connecting pairs of others within a group occupies a central position. Subsequently, various methods for detecting opinion leader based on centrality have been proposed.

3.1.1 Degree centrality

Degree centrality is the number of connections a node has in a network. Freeman presented Degree Centrality in social network, which is rooted in the belief that an individual's significance within a group is tied to the number of people they are connected to or interact with [16]. In real-world case, the node with the highest degree is the user that directly interacts with many other users within the network. This method

is intuitive to the definition of influence, whereas the global structural of the graph is not considered.

3.1.2 Betweenness centrality

Betweenness centrality is based on the number of times a node lies on the shortest path between two other nodes in the network [16]. In online social network, user with high Betweenness centrality operates like a bridge in the shortest paths between possible user pairs. Closeness and Betweenness centrality are challenging to apply in large-scale networks, and have been proved to be unstable in some cross-sectional and temporal networks.

3.1.3 Closeness centrality

Taking consideration of indirect link using the path length, the closeness centrality extends the local centrality to global centrality. The basic idea of closeness centrality is that the node with high closeness centrality can spread the information to other nodes quickly. In this case, the position of one point in the network is more essential than the number of links it own. In online social network, users with high closeness centrality have been proved to be effective spreaders of information by measuring the diffusion effect [17]. However, Closeness centrality is very sensitive to a large distance or missing link due to considering the distance of each pair.

3.1.4 Eigenvector centrality

Eigenvector centrality of a node is calculated as the weighted sum of the centralities of its neighbors, with the weights determined by the strength of the connections between the node and its neighbors. Therefore, this measure can be used to measure the level of influence of each node, where the higher score the greater level of influence. Eigenvector centrality is designed to differ from the former measures when the network contains high-degree nodes connected to many low-degree nodes or low-degree nodes connected to a few high-degree nodes. The disadvantage of Eigenvector Centrality is that it has limitations when applied to directed networks. A node can receive a score of zero in the absence of incoming links, resulting in no contribution to the centrality metric of other nodes.

3.1.5 Katz centrality

Katz and PageRank are variants of the eigenvector centrality. Katz centrality takes into account both the number of direct connections a node has and the connections of its neighbours, which can be less sensitive to the size of the network and proved stable ranking [18]. The limitation of Katz centrality is that it can be influenced by new links to a particular group of nodes.

3.1.6 PageRank centrality

To mitigate the impact of spendthrift nodes on centrality scores, PageRank reduce the weight of ingoing links from these nodes. In PageRank, the weight of an incoming link is proportional to the PageRank score of the node it originates from [19]. Compared to Katz centrality, PageRank add the scaling factor which gives it the ability to penalise nodes that are linked to from many low-quality nodes and reward nodes that are linked to from high-quality nodes. In this way, the PageRank centrality mitigates the impact of nodes with many outgoing links, and instead focuses on the quality of the incoming links, rather than the quantity.

3.1.7 HITS

As PageRank, Hyperlink-Induced Topic Search(HITS) is also a link-based ranking algorithm to determine the importance of the node [20]. The intuition of HITS is that Authority score and Hub score are both allocated to each web page. Assuming that high-quality Hub usually point to high-quality Authorities, and high-quality Authority is pointed by high-quality Hubs. As a result, the Authority score is proportional to the total hub scores of the Hubs that link to it. In online social network, the algorithm search the influential accounts by collecting the query-related accounts and then ranking only by the network structure instead of textual contents.

3.1.8 SPEAR

Yeung et al. introduced the terms experts and expertise for resource discovery [21]. Assuming that a user's expertise depends on the quality of the resources they have collected and the quality of resources is depend on the expertise of other users who have assigned relevant tags, Spamming-Resistant Expertise Analysis and Ranking(SPEAR) was introduced to rank users in online knowledge communities. SPEAR is a graph-based algorithm similar to the HITS algorithm implementing the concept of expertise. Later in 2016, Shinde and Girase proposed the modified SPEAR algorithm [22] where the expertise of user is based on different topics. In the topic-specific SPEAR algorithm, the credit score function considers not only time, but also number of comments, number of likes, word count and all.

3.1.9 TunkRank

Tunkelang introduced TunkRank, a measure of user influence based on PageRank [23]. TunkRank operates on three assumptions: 1)influence power of a certain influencer corresponds to expected number of audiences who read a tweet from the influencer, 2) the probability of audience reading a tweet based on the number of accounts they follow, and 3)

the audience has a constant probability to retweet the seen tweet. The expected number of people who read the tweet can be recursively calculated based on the equally distributed probability of each follower read the tweet and the constant probability that user will retweet the tweet.

3.1.10 Dynamical influence

Dynamical influence is a centrality measure that takes into account the interplay between network structure and the dynamical state of nodes. This is a departure from classical centrality measures which rely solely on topology. In the context of social network, the dynamical influence process can be used to explain the dynamics of idea adoption. In this scenario, opinion leaders are defined as the key individuals who can trigger a significant cascade of influence. The challenge lies in identifying these key individuals, which is essentially an influence optimization problem. The goal is to target an initial set of nodes with the greatest influence spread, thereby promoting information to a large fraction of the network. The maximization of information flow was first considered as a discrete optimization problem by Kempe et al. [24]. They discussed models for how influence propagates through online social networks, and proposed a greedy hill climbing approach of identifying the most influential nodes which provide provable approximation guarantees. Zhao [25] built a statistical model SEISMIC building on the theory of self-exciting point processes to model the information cascades. SEISMIC provides an extensible framework for predicting information cascades. It requires no feature engineering and can scaling linearly with the number of observed reshapes.

3.1.11 SIR

The Susceptible-Infected-Recovered(SIR) model is another algorithm that considers the dynamical state of nodes. The SIR mathematical model was originally designed to describe the spread of infectious diseases in a population. The model divides the dynamical state of population into three categories: Susceptible(S), Infected(I), and Recovered(R). The SIR model has also been used to model the spread of information in a network, where nodes can be thought of as either susceptible to influence, influenced, or recovered from the influence. When applied to identify opinion leaders, the opinion leaders are set as the initially infected nodes, and the probability of an infection depends on the influence from the opinion leader.

3.1.12 Meta-centrality and learning meta-ranks

There are a multitudes of other meta-centrality approaches such as Centripetal Centrality, combining multiple centrality approaches [26]. Such functional combination approaches open up the more reasonable method of using deep learning

to learn new centrality measures [27]. However, recent work recognises that understanding the topic context is important to not only directing centrality measures to be more precise, but also incorporating knowledge of influence behaviour into the centrality measures [28].

3.2 Topic-sensitive centrality

Opinion leaders are identified based on various characteristics that align with diverse social groups. Analysis of dynamic influence across topics and time has demonstrated that ordinary users can gain influence by focusing on a single topic [29]. Recognising that the influence of these leaders can fluctuate based on various topic field, the topic-sensitive centrality approaches incorporate topical attributes into the analysis. Several topic-sensitive ranking methods have been developed to determine the topical influence of users and their capacity to disseminate information or influence opinion on specific topics. Simultaneously, the dynamics of opinion can be analysed based on a particular topic.

3.2.1 InfluenceRank

Topical analysis can be used to quantify the novelty of certain content by representing each content as a document and reducing the dimensionality using Latent Dirichlet Allocation (LDA). The InfluenceRank algorithm uses the topical analysis to measure the importance and novelty of a blog in comparison to other blogs [30]. With the feature vectors that represent the topic distribution, the dissimilarity can be calculated using cosine similarity. InfluenceRank outperforms other algorithms in terms of coverage, diversity and distortion.

3.2.2 TwitterRank

TwitterRank is a variant of the TunkRank algorithm that incorporates topical similarity in the calculation of influence. The phenomenon of “homophily” has been observed in various network ties, including information transfer, friendship, and marriage [31]. Weng et al. demonstrated that “homophily” also exists in the context of Twitter, where users tend to follow those who share similar topical interest [32]. TwitterRank was proposed based on this finding, measuring influence by considering both topical similarity and link structure. However, users’ topical interests can change over time, and as a result, the freshness of their activities needs to be taken into account. Dhali et al. [33] addressed this issue by proposing TemporalTwitterRank, a modified algorithm that estimates transition probabilities using topic profile vectors. By emphasizing the temporal dimension of users’ activities, TemporalTwitterRank provides a more comprehensive assessment of influence.

3.2.3 TopicSimilarRank

Wang et al. proposed the TopicSimilarRank algorithm considering the user's own influence and difference in influential values caused by responses from others. The TopicSimilarRank algorithm is inspired by TwitterRank and takes into account topic similarity, user attributes, interactive information, and network structure. To construct the weighted network, the users can be seen as a set of weighted nodes, and the reposts and comments can be seen as edges with weights represented by similarity values between users. Then the directed and weighted graph can reflect the influential relationships between users. The experiment analysis indicates that TopicSimilarRank is well-suited for mining opinion leaders in topic domains. Similarly, Eliacit et al. [34] developed three metrics - User Trust (degree of friendship, expertise and activity), Influence Period and Similarity - to construct a weighted influence network. Influence rank was calculated based on the PageRank Algorithm. The empirical experiment demonstrates that considering the ranking of users enhances the accuracy of sentiment classification in the community.

3.2.4 ClusterRank

To identify the most influential authors for a specific topic, Pal and Counts proposed a set of features, including both nodal and topical metrics, to describe the authors in various topic fields [35]. To reflect the impact of users with respect to one topic, various features are selected for original tweets, conversational tweets and repeat tweets. ClusterRank process includes using probabilistic clustering on this feature space, within-cluster ranking procedure and producing a list of top authors for a given topic. The experiment showed that topical signal and mention impact are two critical features to determine the ranking.

3.2.5 OpinionRank

OpinionRank considers both the dynamic of information influence and the dynamic of forming opinions. In 2009, Zhou and Zeng introduced the concept of opinion networks and OpinionRank algorithm to rank the nodes based on their opinion scores [36]. In this context, a weighted link in the opinion networks represents the opinion orientation from opinion sender to opinion receiver. For instance, in a review website, the opinion receiver is the original review writer and the opinion sender is the comment writer under the review. The opinion orientation can be calculated as the average opinion score after assigning an opinion score to each word. Experimental results have indicated that sentiment factors significantly influence social network analysis.

3.2.6 Maximization

Huang et al. introduced the Positive Opinion Leader Detection (POLD) to track the public formation process [37]. POLD constructs multiple opinion networks on comment networks rather than user networks. The comment network takes into account the time interval between comments, assuming that influence weakens with increasing intervals. Applying POLD to the comments of news reveals that the most influential comments and users vary over time. Dong et al. further hypothesised that influence only occurs when a recipient posts within a certain time interval after the influencer [38]. The weight of edges in this network is modelled based on the time gap between the posts by influencer and recipient.

3.2.7 TrustRank

Chen et al. proposed the TrustRank considering both positive and negative opinions [39]. TrustRank constructs a network with direct and indirect sentiment labelled links. The construction has 4 phases: 1) set up a basic network, 2) label the links, 3) infer the sign, and 4) transform the post network to user network. During the construction of network, both explicit link and implicit link are considered. The explicit link is denoted by reply and citation, and implicit link infers the semantic similarities between posts. TrustRank outperforms other PageRank-like models on the online comments of a real forum.

3.2.8 InfluenceModellingRank

In our previous work, we proposed a method to model the evolution of personal opinions as an ordinary differential equation (ODE) [40]. To account for the influence of influencers on one recipient's opinion, we employed French's formal theory [10] to model the social influence effect. This effect is determined by the discrepancy of their opinions and the influence weight representing the strength of the effect. To compute the influence weight, we utilized a collection of following links and posts. By assigning the influence weights as link weights and using the PageRank algorithm, we were able to rank the users based on their influence weight. The resulting *InfluenceModellingRank* provides a metric for understanding the opinion influence dynamics in social networks.

3.3 Influence based on control centrality

Social Influence is roughly defined as follows: Given two individuals u, v in a social network, u exerts the power on v , that is, u has the effect of changing the opinion of v in a direct or indirect way [41]. The influence of an individual in the

social network is affected by the self-dynamics of the individual's behavior, coupling dynamics between individuals, and the network structure of the social network. Metrics for influence based on the previous centrality measures mainly consider the network topology of the social network. When we consider both the social network structure and the dynamics of each node, it is natural for us to ask the following questions:

1. whether it is possible for a node to influence other nodes to any desired state
 2. how many nodes' states can be influenced by one nodes
- Therefore, it is reasonable to introduce controllability in complex network to quantify the influence of each node and detect the influential node.

Here we introduce the concept of controllability in complex networks to identify influential nodes. The analysis framework we introduce here to identify influence nodes can be generally applied in social networks, which reflects in following perspectives: 1) this framework can be used in any linear dynamics and does not need to know the specific dynamic functions; 2) only the network topology of the social network is needed, and even the weights of connections are not necessary to know.

3.3.1 Kalman's criterion of controllability

Consider a complex system described by a directed weighted network of N nodes, the dynamics of a linear time-invariant (LTI) system can be described as

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \quad (3)$$

where $\mathbf{x}(t) = (x_1(t), x_2(t), \dots, x_N(t))^T \in \mathbb{R}^N$ captures the state of each node at time t . $\mathbf{A} \in \mathbb{R}^{N \times N}$ is an $N \times N$ matrix describing the weighted connection of the network. The matrix element $a_{ij} \in \mathbb{R}$ gives the strength that node j affects node i . $\mathbf{B} \in \mathbb{R}^{N \times M}$ is an $N \times M$ input matrix ($M \leq N$) identifying the nodes that are controlled by the time-dependent input vector $\mathbf{u}(t) = (u_1(t), u_2(t), \dots, u_M(t)) \in \mathbb{R}^M$ with M independent signals imposed by the controller. The matrix element $b_{ij} \in \mathbb{R}$ represents the coupling strength between the input signal $u_j(t)$ and node i . The controllability of the LTI system can be checked by the best known Kalman's rank condition [42] which states that the LTI system is controllable if and only if the $N \times NM$ controllability matrix

$$\mathbf{C} \equiv [\mathbf{B}, \mathbf{A}\mathbf{B}, \mathbf{A}^2\mathbf{B}, \dots, \mathbf{A}^{N-1}\mathbf{B}] \quad (4)$$

has full rank, i.e.,

$$\text{rank } \mathbf{C} = N. \quad (5)$$

When the system (\mathbf{A}, \mathbf{B}) is not controllable, the dimension of the controllable subspace is $\text{rank } \mathbf{C}$, where $\text{rank } \mathbf{C} < N$.

3.3.2 Exact control centrality

One thing we are interested in is how many dimensions of the subspace of the system can be controlled by a single node. Here, we use $\text{rank } \mathbf{C}^{(i)}$ to capture the ability of i in controlling other nodes in the networked system. Mathematically, $\text{rank } \mathbf{C}^{(i)}$ captures the dimension of the controllable subspace or the size of the controllable subsystem when we only control node i . The exact control centrality of node i is defined as

$$C(i) \equiv \text{rank } (\mathbf{C}^{(i)}), \quad (6)$$

where the \mathbf{B} in matrix \mathbf{C} reduces to the vector \mathbf{b}^i with a single nonzero entry, e.g. $\mathbf{b}^i = [0, 0, \dots, b_i, \dots]^T$. By calculating the exact control centrality of each node in the networked system, the most powerful nodes in controlling the whole networked system can be identified. In a social network, with exact parameters, users with higher $C(i)$ can affect more users' opinions. Therefore, we can find the most influential nodes according to exact control centrality.

3.3.3 Structural controllability

When we know the exact network structure and the weight of each connection in the social network, the influence of each user can be ranked by the exact control centrality. However, there exist some limitations when the exact control centrality method is applied to analyze the influence of each user. $C(i)$ is sensitive to the perturbations of elements of matrix $\mathbf{C}^{(i)}$, especially in a large matrix. Usually, the social network contains a large number of users, so estimating the influence of each user by the exact control centrality has a high requirement of the accuracy of the weight of connections. The second limitation is that in social networks, the system parameters are not precisely known, e.g. the elements in matrix \mathbf{A} are not exactly known. We only know whether there exists an influence between two users but are not able to measure the weights of the influence between them. Hence, it is difficult to numerically verify Kalman's controllability rank condition using fixed weights. To solve this problem, the concept of structural control [43] can be introduced to measure the influence in social networks. The power of structural controllability comes from the fact that if a system is controllable the it is controllable for almost all possible parameter realizations [44].

An LTI system (\mathbf{A}, \mathbf{B}) is a structured system if the elements in \mathbf{A} and \mathbf{B} are either fixed zeros or independent free parameters. Apparently, $\text{rank } (\mathbf{C})$ varies as a function of the free parameters of \mathbf{A} and \mathbf{B} . It achieves the maximal value

for all but an exceptional set of values of the free parameters. This maximal value is called the generic rank of the controllability matrix \mathbf{C} , denoted as $\text{rank}_g(\mathbf{C})$, which represents the generic dimension of the controllable subspace. The system (\mathbf{A}, \mathbf{B}) is structurally controllable if we can set the nonzero elements in \mathbf{A} and \mathbf{B} such that the resulting system satisfies $\text{rank}_g \mathbf{C} = N$. The minimum number of nodes which control the state of the full system can be identified by mapping this problem to a pure graph-theoretical problem called maximum matching [45, 46]. In social influence networks, the subset of these nodes are the most influential nodes which can influence the state of all nodes in the network.

3.3.4 Structural control centrality

Correspondingly, to measure the dimension of the controllable space of one node without the exact information of system parameters, the concept of structural control centrality has been introduced, which can be defined as [47]

$$C_g(i) \equiv \text{rank}_g(\mathbf{C}^{(i)}). \quad (7)$$

The structure control centrality is an upper bound of exact control centrality for all admissible numerical realizations of the controllable matrix \mathbf{C} . So, in an influence network, the structure control centrality is a method to estimate the largest number of users that can be affected by one user with clearly known connections between users. To calculate $C_g(i)$, we need to introduce some concepts in graph theory. A node j is called accessible if there always exists at least one directed path from the input nodes to j . A stem is a directed path starting from an input node, so that no nodes appear more than once in it. $C_g(i)$ can be calculated according to Hosoe's controllable subspace theorem [48]:

$$\text{rank}_g(\mathbf{C}) = \max_{G_s \in G} |E(G_s)|, \quad (8)$$

where G_s is the subgraph of the accessible part of G only consists of stems and cycles and $|E(G_s)|$ represents the number of edges in G_s . The action space may take many forms from inserting control signals to rewiring the graph structure [49], this falls outside this review.

3.4 Graph sampling & recovery

Another idea to determine the most influential nodes over a network leverages whether these nodes can be sampled to recover the whole networked dynamics. This refers to as graph signal sampling and recovery techniques, which aim to compress the time series of high-dimensional and dependent networked dynamics via a subset of critical nodes, whose dynamics can guarantee the recovery of the whole networked data. From the theoretical perspective, this includes low-rank

matrix completion, optimization with Laplacian constraints, the spatial, and the temporal dependency analysis, whereby the former studies the correlation or hidden high-dimensional dependency among the set of nodes, and the latter focuses on the events at which time steps would be the trigger or with higher importance.

3.4.1 Low-rank approximation

Low-rank matrix completion aims to recover the whole matrix from the known entries (samples) [50–52]. The typical approach is to minimize the rank of the recovered matrix by the singular values, constrained by the values of the samples [53]. CUR [54] serves another popular method family, leveraging the sampled rows and columns to recover the whole data matrix. In the context of opinion leader identification, the rows of the matrix are the time-serial language embedding of different users, while the columns represent the sampling time indices. In this view, the aim is to find the minimum set of users whose embedding can recover the whole data matrix. This set of users constructs a low-rank core of the data matrix, which contributes significantly to the whole information entropy of the data. By capturing their tendency of the topics, the information flow of all the users can be approximately obtained.

3.4.2 Optimization with laplacian penalty

The eigenvalues and eigenvectors of the Laplacian matrix of the graph structure serve as the graph frequency domain. Under the assumption that critical users capture a large portion of the information entropy of the networked dynamic data, the Laplacian penalty [55] can be set to constrain the graph bandwidth of the data for critical user identification. In this view, the critical users found by this approach represent the approximated data with minimum graph bandwidth concerning the graph structure-based Laplacian matrix, which, however, does not involve the specific dynamic patterns (e.g., ODE or PDE) that govern the evolution of the language embedding propagation [56]. As such, the selected users can better represent the graph structure from the frequency domain (other than the topological node domain as stated in Section 2.1), but lack the propagation information for different topic-sensitive patterns.

3.4.3 Spatial correlation analysis

Spatial correlation analysis tries to determine an orthogonal signal subspace (matrix), e.g., the operational matrix in compressed sensing (CS), or the graph Fourier transform (GFT) operator [57–59]. Then, leveraging the orthogonal subspace, the highly correlated networked data can be compressed by the linear combinations of the subset of the orthogonal bases,

which can be mapped to the critical nodes for sampling and recovery purposes.

Compared to the Laplacian penalty strategy, this approach takes into account both the graph structure and the dynamic pattern directly extracted from the data, and is therefore better to provide the critical opinion leader set for different topic and network -sensitive topics. One drawback is the overlook of temporal correlations between different time stamps, which will underestimate the appearance of opinions in the evolution process.

3.4.4 Spatial & temporal dependency analysis

Spatial and temporal dependency analysis aims to determine the critical nodes by considering both the node level and temporal level correlations [56]. By combining the temporal correlation information, a more compact dynamic subspace can be derived, which gives rise to a reduction in the number of sampling nodes, and leads to a node importance rank.

The derivation of the dynamic subspace contains the model-based and data-driven approaches. The model-based approach is to generate an orthogonal subspace leveraging the linearized dynamic model, e.g., via dynamic mode decomposition (DMD) or extended DMD (E-DMD) [60, 61]. Such a model-based subspace compresses the networked dynamics via the spatial and temporal correlations. The opinion leader identification is then converted to the sampling and recovering problem that selects the critical nodes to make truncated subspace full column rank.

When the model is unavailable (e.g., difficult to pursue a linear regression), the data-driven methods are well-suited to derive the dynamic subspace. To be specific, by pursuing a compact singular value decomposition of the data, the dynamic subspace can be constructed by the Kronecker product of the left and right singular vectors with non-zero singular values. After the derivation of the dynamic subspace, the important users can be derived by the greedy selection of rows to maximize the least singular value of the subspace.

3.5 Evaluation method

The evaluation of Opinion Leader Detection methods is not straightforward, and various papers use different evaluation methods. Unfortunately, there is no agreement on which evaluation method is the best. Nonetheless, some evaluation methods are still commonly used and will be discussed in this section.

3.5.1 Descriptive methods

In social sciences, descriptive methods are often utilized to identify opinion leaders. One of the most popular descriptive methods is using experts to rank the opinion leaders in

one group network. In this approach, an expert is asked subjectively to rate the comments from users having either a strong or weak influence. The ratings of comments are then combined to determine the influential rate of each user. However, descriptive methods require creating questionnaires and conducting interviews, which are costly and challenging to implement. These descriptive measures have been criticized because they do not consider the role of ordinary users in the information flow process [62].

3.5.2 OSN metrics

For OSN platforms, the number of followers is the most commonly used metric to determine a user's influence. This approach assumes that each tweet by a user is read by all of their followers. Other metrics such as likes, shares, or mentions are also used to measure user engagement and influence [63]. On Twitter, these public metrics are accessible through the Twitter Application Programming Interface (API), which is built on communication data and metadata.

3.5.3 Kendall's τ

Kendall's τ is a statistical measure to determine the similarity between the ranking orders of two variables, regardless of their magnitudes. Kendall's τ coefficient ranges from -1 to 1 , with a value of -1 indicating complete disagreement between the rankings, and 1 indicating perfect agreement between the rankings. Kendall's Tau correlation method is often used in social science research, including the opinion leader detection task, to assess the degree of agreement or disagreement between two rankings.

4 Case study

This section presents a case study of the application of different ranking methods to 2 Twitter datasets widely used by the research community: (1) COVID-19, and (2) feminism debate. In order to gather and prepare data for our pilot study, we relied on the methodology outlined in our previous study [40], which is we identified topic specific active users who posted a minimum level of topic-specific tweets over a specified period. Based on our process, we identified 98 active users for the COVID-19 topic and 180 active users for the feminism debate topic. We then crawled and analyzed 85,946 COVID-19 related tweets and 69,088 feminism related tweets. To pre-process the tweets and capture the vibration of opinion, we used compressed word-embedding vectors [40]. For the validation of the different centrality rankings, we selected four topic-filtered matrix rankings: Retweet, Reply, Like, and Quote. These filtered matrix rank-

Table 2 Kendall’s τ Score on ordinal categorical COVID-19 dataset

τ Score	Betweenness	Eigenvector	PageRank	InfluenceModel	Control	MGFT	DGFT
Retweet	0.025	0.055	0.066	0.032	0.274	0.046	-0.046
Reply	0.107	0.03	0.012	0.084	-0.002	0.172	0.206
Like	0.125	0.020	0.02	0.083	0.013	0.156	0.177
Quote	0.121	0.015	-0.002	0.080	0.033	0.141	0.161

ings were calculated by considering only the topic-related tweet matrix posted by the group of users.

4.1 Proof of concept

We first computed the previously reviewed three centrality rankings commonly used directly or as part of meta-centrality: Betweenness, Eigenvector, PageRank, as well as one topic-sensitive ranking: InfluenceModel, one control theory ranking, and two previously reviewed compressive graph sampling rankings: MGFT, DGFT.

The Kendall τ correlation results are illustrated in Tables 2 and 3. Since the absolute value of all Kendall τ are lower than 0.3 and most of them are close to 0 which means these ranking vectors are likely to be independent.

Our observations backed by logic and topic-sensitive evidence are as follows:

- The three classic centrality rankings and topic-sensitive ranking exhibit greater performance similarity with each other. Conversely, the Control ranking and two graph sampling rankings yield distinct results due to their different definitions of influence. This demonstrates that indeed the new definitions offer alternative value.
- In Table 2, the Control rank exhibits the highest similarity score with Retweet rank, and both GFT rank approaches demonstrates the highest similarity score with Reply rank. This shows key control points is a better indication of retweet relays, whereas graph compression is a better indication of response.
- In Table 3, the MGFT rank achieves the highest similarity scores with all four validation ranks.

A general non topic-sensitive horizontal comparison among different ranking strategies is challenging, due to the different criteria utilised by the methods. For instance, in an extreme case where someone that replies to many posts

may indicate that they are incorporating and modifying the context, then the graph sampling theory may select it as an influential node, given its potential to contribute to data recovery of all other contexts. Such a user, on the other hand, may not rank highly from a control perspective, meaning they do not control conversation. Then, there may be some overlap that a good representative or control user may have good topographical or topic-sensitive properties, yet their correlation and causality require further studies.

4.2 Improvements for community to make

In our current approach, a progressive research flow on ranking/identifying opinion leaders is provided, whereby different ideas leveraging the uses of topology, static topic-sensitive, and opinion differential evolution are reviewed and evaluated. The main challenge would be integrating diverse contents, features and dynamics as input data to find the opinion leaders. Diverse contents refers to incorporation of Cross-platform content analysis-for instance, text content on Twitter, graphical content on Instagram, and video content on TikTok. Various features can include user features, content features and network features. Dynamic nature of social network and human behavior presents another challenge. The continuous flow of data offers opportunities to identify emerging influencers who have the potential to influence,high dimensional data resources.

In current studies, the original content is pre-processed via the word embedding and compression to represent the dynamics of opinions. However, this compression process inevitably lead to information loss, including the delays, the hidden dependency from spatial and temporal perspectives. In future work, to reduce information loss, more complicated opinion representation may be generated to describe the opinion evolution, which will challenge the current raking strategies that leverage the networked dynamics. Consequently, how to design nonlinear sampling and control spaces

Table 3 Kendall’s τ Score on ordinal categorical Feminism dataset

τ Score	Betweenness	Eigenvector	PageRank	InfluenceModel	Control	MGFT	DGFT
Retweet	0.041	0.083	0.008	0.052	0.098	0.105	0.031
Reply	0.014	0.043	0.010	0.035	0.108	0.116	0.081
Like	0.017	0.046	0.01	0.042	0.104	0.122	0.065
Quote	0.01	0.082	-0.013	0.048	0.096	0.102	0.021

Table 4 Comparative inter-disciplinary analysis of opinion leader definitions and their detection methodologies

Psychology Definition	Concept Type	Method
Unbiased dynamic based spread in large groups. [1] (Section 3.1)	Spread	Large Topology -based
Individuals who disseminate topic-related information effectively. [5] (Section 2.2.2)	Receptiveness	Topic Sensitive
Repeat the information to maximize influence based on observed context. [64] (Section 3.3)	Control	Graph signal control
Respond to information to capture diverse contexts [8] (Section 3.4)	Representation	Graph signal compression

may be worth studying in the future. Furthermore, it is also noteworthy that even with the sophisticated ODE construction with graph signal evolution and control layer inputs, only the correlation/dependency overlap with the opinion leader set can be identified. In this view, on one-hand, how to build a causality model that represents the causal relations from opinion leaders to the dynamic evolution data requires further studies; on the other hand, the reverse flow from the networked dynamics to infer the opinion leader from the causality perspective remains untouched and is worth studying.

5 Discussion

The purpose of this review is to provide an overview of the development of "opinion leader" concept and detailed comparative analysis of the corresponding detection techniques.

The key novelty is to review technical connections and cross-compare different approaches that have origins in: social theory, graph theory, control theory, and graph sampling - with the eventual goal to more holistically describe trusted opinion leaders and their strategies in influence [8]. Here, we conclude with a comparative analysis of opinion leader definitions and their detection methodologies, as shown in Table 4.

When it comes to topic-sensitive centrality, text content of a user’s conversation can be transferred into topic probability distribution or opinion states vector. In the case of the former, topic analysis can be leveraged to determine the novelty and similarity of content. Here, the identified opinion leader exhibits the ability to generate novel content and disseminate topic-related information effectively. In the case of the latter, opinion dynamics are analyzed under a specific topic. The identified opinion leaders in this scenario are users who possess the capacity to influence opinions of others [64].

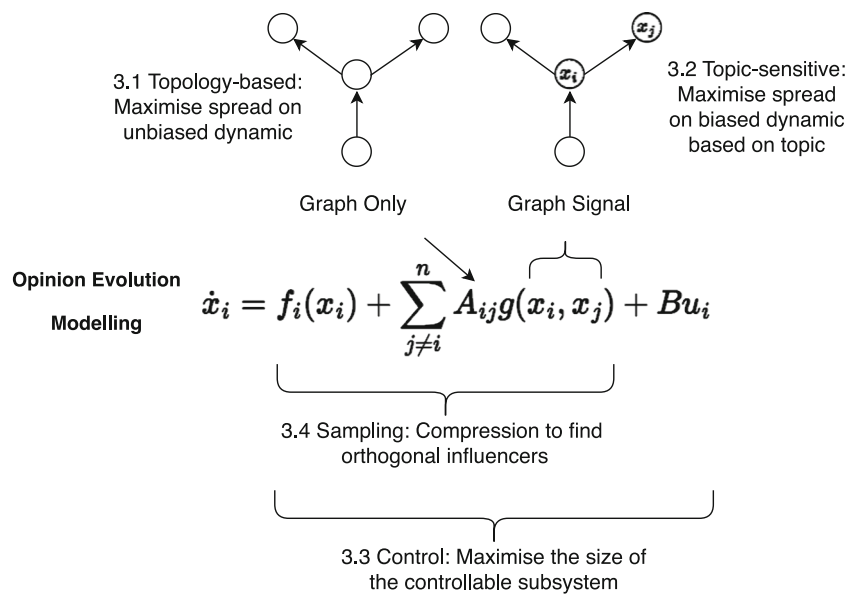


Fig. 2 Opinion Evolution Modeling, interlinking four methods: The Opinion Evolution Modeling formula at the center delineates the mathematical model underpinning opinion dynamics. Topology-based detection emphasizes the maximization of spread on an unbiased dynamic represented solely by graph structure A_{ij} . Topic-sensitive detection, where the maximization of spread on a biased dynamic is contingent on the topical relevance of the content, signified by the graph

signal x_i based on topic. By constructing opinion evolution models as $\dot{x}_i = f_i(x_i) + \sum_{j \neq i}^n A_{ij}g(x_i, x_j)$, graph sampling methods identify orthogonal influencers to reduce redundancy in message dissemination across the network. Control theory approaches aim to maximize the size of the controllable subsystem in the network, by incorporating control signal u_i

Here, the receptiveness of the recipients is more important than graph structure. Classical models include psychology informed differential equations in small group psychology experiments [10, 12, 14], which we reviewed in Section 2.2.2.

Opinion evolution modeling was developed to explain how individuals develop their opinions towards various topics over time. Researchers have considered both linear and non-linear models to capture the complexity of opinion evolution. With the linear dynamics of opinion modeling, such as the French-DeGroot model, control theory can be broadly applied to dynamic opinion networks, eliminating the need of knowing the specific dynamic functions or the weights of connections. The application of control theory aims to find opinion leaders who can steer the overall direction of public opinion, which we reviewed in Section 3.3.

Graph sampling theory can be applied to identify multiple opinion leaders that minimize redundant information and influence pathways and maximize the overall efficiency of the system in spreading influence. Using either linear model-based and data-driven manners, where the latter does not require the awareness of the dynamic model nor its linearity assumption. By deriving the orthogonal subspace from the data, the data-driven graph sampling method can obtain the opinion leaders who are the representative users in the dynamic opinion networks, which we reviewed in Section 3.4.

What we have demonstrated across this review is how they offer different insight, but more importantly how they can be combined together. As illustrated in Fig. 2, we use opinion evolution formula to interlink the four methods. Topology-based centrality maximizes unbiased influence spread using graph information alone, represented by A_{ij} . In contrast, topic-sensitive methods rank users based on their ability to influence biased dynamics, using opinion states vector x_i as the graph signal. By constructing opinion evolution models as $\dot{x}_i = f_i(x_i) + \sum_{j \neq i}^n A_{ij}g(x_i, x_j)$, we can apply graph sampling to identify orthogonal influencers, thereby minimizing redundant messaging. Further, by incorporating control signal u_i , we can determine which influencers maximally impact the controllable subsystem size.

In summary, our review highlights the diverse methodologies for identifying opinion leaders across disciplines. By integrating these different approaches, we can better understand the complex dynamics of opinion formation and influence in large networks.

6 Conclusion

Through this review and a case study, we performed a comparative analysis of multiple methodologies across disciplines. Some of the analysis we focus on is intra-disciplinary, showing connections and differences between graph central-

ity, control theory and sampling theory. Our initial contribution is cross-disciplinary in nature, reviewing qualitatively from the perspective of different disciplines. The greatest contribution is that we go forward and build connections to psychology and develop psychology-informed differential equation signals that can be combined with aforementioned graph signal analysis. This shows a trans-disciplinary contribution to knowledge.

The results show that a horizontal comparison among different ranking strategies is challenging, due to the disparate criteria utilised by the methods. There may be some overlap between the identified opinion leaders through various methods, yet their correlation and causality require further studies. It is our hope that this survey will help researches in gaining better understanding of the development of opinion leader detection methods and inspire them to address the remaining challenges in this field.

For future work, the main challenge would be integrating diverse contents, features and dynamics as input data to find the opinion leaders. Diverse contents refers to incorporation of cross-platform content analysis-for instance, text content on Twitter, graphical content on Instagram, and video content on TikTok. Various features can include user traits, content and network features. Dynamic nature of social network and human behavior presents another challenge. The continuous flow of data offers opportunities to identify emerging influencers who have the potential to influence public opinion. We also wish to consider how we can create synthetic test environments using emerging large language model agents [65], which can create ethical environments to validate experiments. This can aid the development of larger diverse data sets with topic-specific influence labels is also important, as we have so far been limited to two widely used data sets.

Overall, a comprehensive approach is essential for identifying influential users using the high dimensional data resources. It is our hope that this survey will help researches in gaining better understanding of the development of opinion leader detection methods and inspire them to address the remaining challenges in this field.

Acknowledgements The work is supported by "Networked Social Influence and Acceptance in a New Age of Crises", funded by USAF OFSR under Grant No.: FA8655-20-1-7031, and is partly supported by the Engineering and Physical Sciences Research Council grant number: EP/V026763/1

Author Contributions All authors contributed to the conception and design of the study. Bailu Jin was responsible for data collection. Bailu Jin, Mengbang Zou, and Zhuangkun Wei conducted the literature search, data analysis, and wrote the first draft of the manuscript. Weisi Guo supervised the project, and provided critical reviews and commentary on the work. All authors have read, revised, and approved the final manuscript.

Funding The work is supported by "Networked Social Influence and Acceptance in a New Age of Crises", funded by USAF OFSR

under Grant No.: FA8655-20-1-7031, and is partly supported by the Engineering and Physical Sciences Research Council grant number: EP/V026763/1

Data Availability The data is published at <https://github.com/AlminaJin/OpinionLeaderDetection.git>.

The code is published at <https://github.com/AlminaJin/OpinionLeaderDetection.git>.

Declarations

Conflicts of interest/Competing interests The authors have no relevant financial or non-financial interests to disclose.

Ethics approval and consent to participate Not applicable.

Consent for publication All authors of this paper have given their consent for its publication.

Materials availability Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Lazarsfeld PF, Berelson B, Gaudet H (1968) *The people's choice*. Columbia University Press
2. Bamakan SMH, Nurgaliev I, Qu Q (2019) Opinion leader detection: a methodological review. *Expert Syst Appl* 115:200–222
3. Riquelme F, González-Cantergiani P (2016) Measuring user influence on twitter: A survey. *Inf Process Manag* 52(5):949–975
4. Panchendrarajan R, Saxena A (2023) Topic-based influential user detection: a survey. *Appl Intell* 53(5):5998–6024
5. Katz E (1957) The two-step flow of communication: An up-to-date report on an hypothesis. *Public Opin Q* 21(1):61–78
6. Hellevik O, Bjørklund T (1991) Opinion leadership and political extremism. *Int J Public Opin Res* 3(2):157–181
7. Rogers EM, Cartano DG (1962) Methods of measuring opinion leadership. *Public Opin Q* 435–441
8. Corey LG (1971) People who claim to be opinion leaders: identifying their characteristics by self-report. *J Market* 35(4):48–53
9. Tkachenko N, Guo W (2020) Conflict detection in linguistically diverse on-line social networks: A russia-ukraine case study. In: *ACM international conference on management of digital EcoSystems*, pp 23–28
10. French JR Jr (1956) A formal theory of social power. *Psychol Rev* 63(3):181
11. Myers DG (1982) Polarizing effects of social interaction. *Group Decis Making* 125:137–138
12. DeGroot MH (1974) Reaching a consensus. *J Am Stat Assoc* 69(345):118–121
13. Dong Y, Ding Z, Martínez L, Herrera F (2017) Managing consensus based on leadership in opinion dynamics. *Inf Sci* 397
14. Hegselmann R, Krause U, et al (2002) Opinion dynamics and bounded confidence models, analysis, and simulation. *J Artif Soc Soc Simul* 5(3)
15. Bavelas A (1948) A mathematical model for group structures. *Human Organ* 7(3):16–30
16. Freeman LC et al (1979) Centrality in social networks: conceptual clarification. *Social network: critical concepts in sociology*. Londres: Routledge 1:215–239
17. Yang L, Qiao Y, Liu Z, Ma J, Li X (2018) Identifying opinion leader nodes in online social networks with a new closeness evaluation algorithm. *Soft Comput* 22:453–464
18. Katz L (1953) A new status index derived from sociometric analysis. *Psychometrika* 18(1):39–43
19. Page L, Brin S, Motwani R, Winograd T (1999) The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab
20. Kleinberg JM (1999) Authoritative sources in a hyperlinked environment. *J ACM (JACM)* 46(5):604–632
21. Yeung C-mA, Noll MG, Gibbins N, Meinel C, Shadbolt N (2011) Spear: spamming-resistant expertise analysis and ranking in collaborative tagging systems. *Comput Intell* 27(3):458–488
22. Shinde M, Girase S (2016) Identification of topic-specific opinion leader using spear algorithm in online knowledge communities. In: *2016 International conference on computing, analytics and security trends (CAST)*. IEEE, pp 144–149
23. Tunkelang D (2009) A twitter analog to pagerank
24. Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. In: *Proceedings of the Ninth ACM SIGKDD international conference on knowledge discovery and data mining*, pp 137–146
25. Zhao Q, Erdogdu MA, He HY, Rajaraman A, Leskovec J (2015) Seismic: a self-exciting point process model for predicting tweet popularity. In: *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pp 1513–1522
26. Wang Y, Li H, Zhang L, Zhao L, Li W (2024) Identifying influential nodes in social networks: centripetal centrality and seed exclusion approach. *Chaos* 163
27. Rashid Y, Bhat J (2023) Topological to deep learning era for identifying influencers in online social networks: a systematic review. *Multimed Tools Appl* 83
28. Zhou F, Lv L, Liu J, Mariani MS (2024) Beyond network centrality: individual-level behavioral traits for predicting information superspreaders in social media. *Natl Sci Rev* 11
29. Cha M, Haddadi H, Benevenuto F (2010) Measuring user influence in twitter: the million follower fallacy. In: *Proceedings Of the fourth international AAAI conference on weblogs and social media*
30. Song X, Chi Y, Hino K, Tseng B (2007) Identifying opinion leaders in the blogosphere. In: *Proceedings of the Sixteenth ACM conference on information and knowledge management*, pp 971–974
31. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. *Ann Rev Sociol* 27(1):415–444
32. Weng J, Lim E-P, Jiang J, He Q (2010) Twittrrank: finding topic-sensitive influential twitterers. In: *Proceedings of the third ACM international conference on web search and data mining*, pp 261–270
33. Dhali A, Gomasta SS, Anwar MM, Sarker IH (2020) Attribute-driven topical influential users detection in online social networks. In: *2020 IEEE Asia-pacific conference on computer science and data engineering (CSDE)*. IEEE, pp 1–5

34. Eliacik AB, Erdogan N (2018) Influential user weighted sentiment analysis on topic based microblogging community. *Expert Syst Appl* 92:403–418
35. Pal A, Counts S (2011) Identifying topical authorities in microblogs. In: *Proceedings of the fourth ACM international conference on web search and data mining*, pp 45–54
36. Zhou H, Zeng D, Zhang C (2009) Finding leaders from opinion networks. In: *2009 IEEE international conference on intelligence and security informatics*. IEEE, pp 266–268
37. Huang B, Yu G, Karimi HR (2014) The finding and dynamic detection of opinion leaders in social network. *Math Probl Eng* 2014
38. Dong G, Li B, Wei X, Qin T (2019) Mining key users of microblog topics based on trust model. *Int J Perform Eng* 15(11):3024
39. Chen Y, Wang X, Tang B, Xu R, Yuan B, Xiang X, Bu J (2014) Identifying opinion leaders from online comments. In: *Social media processing: third national conference, SMP 2014, Beijing, China, November 1-2, 2014*. *Proceedings*. Springer, pp 231–239
40. Jin B, Guo W (2022) Data driven modeling social media influence using differential equations. In: *2022 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, pp 504–507
41. Cercel D-C, Trausan-Matu S (2014) Opinion propagation in online social networks: a survey. In: *Proceedings of the 4th international conference on web intelligence, mining and semantics (WIMS14)*, pp 1–10
42. Kalman RE (1963) Mathematical description of linear dynamical systems. *J Soc Ind Appl Math Ser A: Control* 1(2):152–192
43. Lin C-T (1974) Structural controllability. *IEEE Trans Autom Control* 19(3):201–208
44. Liu Y-Y, Barabási A-L (2016) Control principles of complex systems. *Rev Mod Phys* 88(3):035006
45. Murota K (2010) *Matrices and matroids for systems analysis*. Springer
46. Liu Y-Y, Slotine J-J, Barabási A-L (2011) Controllability of complex networks. *Nature* 473(7346):167–173
47. Liu Y-Y, Slotine J-J, Barabási A-L (2012) Control centrality and hierarchical structure in complex networks
48. Hosoe S (1980) Determination of generic dimensions of controllable subspaces and its application. *IEEE Trans Autom Control* 25(6):1192–1196
49. Zou M, Guo W, Chu K-F (2024) Rewiring complex networks to achieve cluster synchronization using graph convolution networks with reinforcement learning. *IEEE Trans Netw Sci Eng* 11(5):4293–4304
50. Jing P, Su Y, Nie L, Bai X, Liu J, Wang M (2018) Low-rank multi-view embedding learning for micro-video popularity prediction. *IEEE Trans Knowl Data Eng* 30(8):1519–1532
51. Zuo X, Liu X, Yang B (2018) Coupled low rank approximation for collaborative filtering in social networks. *IEEE Access* 6:13326–13335
52. Donavalli A, Rege M, Liu X, Jafari-Khouzani K (2012) Low-rank matrix factorization and co-clustering algorithms for analyzing large data sets. In: *Data engineering and management: second international conference, ICDEM 2010, Tiruchirappalli, India, July 29-31, 2010*. *Revised Selected Papers*. Springer, pp 272–279
53. Candes E, Recht B (2012) Exact matrix completion via convex optimization. *Commun ACM* 55(6):111–119
54. Mahoney MW, Drineas P (2009) Cur matrix decompositions for improved data analysis. *Proc Natl Acad Sci* 106(3):697–702
55. Qiu K, Mao X, Shen X, Wang X, Li T, Gu Y (2017) Time-varying graph signal reconstruction. *IEEE J Sel Top Signal Process* 11(6):870–883
56. Wei Z, Li B, Sun C, Guo W (2020) Sampling and inference of networked dynamics using log-koopman nonlinear graph fourier transform. *IEEE Trans Signal Process* 68:6187–6197
57. Wei Z, Pagani A, Fu G, Guymer I, Chen W, McCann J, Guo W (2020) Optimal sampling of water distribution network dynamics using graph fourier transform. *IEEE Trans Netw Sci Eng* 7(3):1570–1582
58. Wei Z, Li B, Guo W (2019) Optimal sampling for dynamic complex networks with graph-banded initialization. *IEEE Access* 7:150294–150305
59. Wei Z, Pagani A, Guo W (2019) Monitoring networked infrastructure with minimum data via sequential graph fourier transforms. In: *2019 IEEE international smart cities conference (ISC2)*, pp 703–708
60. Dey S (2022) Dynamic mode decomposition and koopman theory. [arXiv:2211.07561](https://arxiv.org/abs/2211.07561)
61. Brunton SL (2019) Notes on koopman operator theory. Universität von Washington, Department of Mechanical Engineering, Zugriff 30
62. Valente TW, Pumpuang P (2007) Identifying opinion leaders to promote behavior change. *Health Educ Behav* 34(6):881–896
63. Bruns A, Stieglitz S (2014) Metrics for understanding communication on twitter. *Twitter and society [Digital Formations, Volume 89]*:69–82
64. Rogers EM, Singhal A, Quinlan MM (2014) *Diffusion of innovations*. Routledge, 432–448
65. Siahkali F, Samadi S, Kebriaei H (2024) Towards opinion shaping: A deep reinforcement learning approach in bot-user interactions. [arXiv:2409.11426](https://arxiv.org/abs/2409.11426)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.