Joint Source Localization in Different Platforms via Implicit Propagation Characteristics of Similar Topics

Zhen Wang[3,1,2] , **Dongpeng Hou**[1,2] , **Shu Yin**[4,2] , **Chao Gao**[2] and **Xianghua Li**[2*]
[1]School of Mechanical Engineering, Northwestern Polytechnical University
[2]School of Artificial Intelligence, OPtics and ElectroNics (iOPEN), Northwestern Polytechnical University
[3]School of Cybersecurity, Northwestern Polytechnical University
[4]School of Computer Science, Northwestern Polytechnical University
li_xianghua@nwpu.edu.cn

Abstract

Different social media are widely used in our daily lives. Inspired by the fact that similar topics have similar propagation characteristics, we mine the implicit knowledge of cascades with similar topics from different platforms to enhance the localization performance for scenarios where limited propagation data leads to the weak learning ability of existing localization models. In this work, we first construct a multiple platform propagation cascade dataset, aligning similar topics from both Twitter and Weibo, and enriching it with user profiles. Leveraging this dataset, we propose a Dual-channel Source Localization Framework (DSLF) for the joint cascades with similar topics. Specifically, a self-loop attention based graph convolutional network is designed to adaptively adjust the neighborhood aggregation scheme of different users with heterogeneous features in the message-passing process. Additionally, a dual-structure based Kullback-Leibler (KL) regularization module is proposed to constrain the latent distribution space of the source probabilities of similar characteristic-level users for a similar topic, enhancing the robustness of the model. Extensive experiments across Twitter and Weibo platforms demonstrate the superiority of the proposed DSLF over the SOTA methods. The code is available at https://github.com/cgao-comp/DSLF.

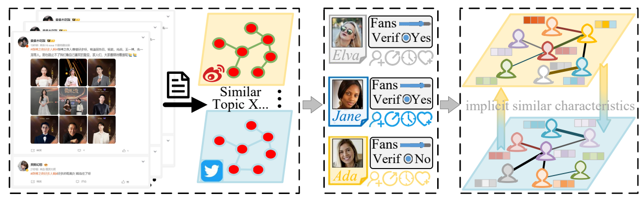

Figure 1: The illustration of multiple platform based source localization grounded in real-world propagation data. Considering the similar propagation cascade features between similar topics, the cascades are aligned across different platforms based on textual content. Further acknowledging the human-driven fact of propagation, we crawl and enrich the cascades with user profiles using the UIDs within each cascade. Lastly, considering the high cost of collecting inter-layer spreading rates, we rely solely on the propagation snapshots available within each platform for source inference.

1 Introduction

In today's digital age, the widespread usage of social media platforms like Weibo and Twitter has dramatically changed the spreading way of information [Ji *et al.*, 2017]. However, the appearance of malicious fake news or sensationalized topics on multiple platforms presents major challenges to the economy and society [Yin *et al.*, 2024; Islam *et al.*, 2020; Wang *et al.*, 2022b]. Focusing on their features and locating the sources are key steps in controlling abnormal spread.

Snapshot based and sensor based source identification methods are developed to locate the sources in the different scenarios [Jiang *et al.*, 2016; Jin and Wu, 2021]. Since the sensor deployment process requires time and space overhead [Paluch *et al.*, 2020a], more and more studies focus on the snapshot branch due to the accessibility and low cost of snapshot acquiring. The captured snapshot provides a set of users who participated in a topic. Then the inference strategy is proposed to identify the first participant in the snapshot.

Considering the complexity of propagation dynamics, most existing snapshot-based methods require abundant cascades to achieve an acceptable performance. A network with fewer than 200 nodes typically needs thousands of cascades to provide reasonable localization performance [Dong *et al.*, 2019]. In practical applications, however, the amount of data is usually insufficient. For example, only a few thousand observed cascades are captured in a million-scale Weibo network [Ma *et al.*, 2017]. Consequently, accurately locating the source using limited data in real-world scenarios becomes a significant challenge. Fortunately, we have identified that there exist potential similar features between cascades with similar topics in different platforms [Hou *et al.*, 2024b]. And the localization performance can be enhanced by leveraging this experience in source inference. However, most snapshot-based methods focus on a single-layer network localization [Hou *et al.*, 2023; Ling *et al.*, 2022], thereby limiting their applicability in joint localization scenarios. Although several localization methods for multiplex networks are proposed [Paluch *et al.*, 2020b], their required prior information, such as spreading rates between two layers, is challenging to obtain.

To address these issues, our work seeks to bridge the 'simulation-to-reality' gap by employing real-world data collected from Twitter and Weibo for joint source localization. As can be seen from Fig. 1, first, different real propagation cascade data of similar topics from Weibo and Twitter is aligned based on the comments, and the user-specific profiles are injected and aligned with the nodes within the cascades based on the unique user identification (UID) of each node in the datasets. Subsequently, a localization method tailored for dual-platform similar topics is developed. It's noteworthy that the propagation way of information is not confined to online platforms alone. The offline behaviors and external knowledge can also greatly affect intra-layer and inter-layer spreading rates. However, the prior information is costly to make the statistics. Therefore, to ensure the scalability of our proposed localization framework, we do not leverage external experience or prior knowledge and only use observed snapshot information in each platform to infer the source.

Based on two sets of snapshots at available timestamps observed from Weibo and Twitter, respectively, we propose a Dual-channel Source Localization Framework (DSLF) based on the regularized constraint for cascades with similar topics in different platforms. Specifically, eight essential propagation features is constructed for each user at every timestamp. These are further integrated with seven unique user profile features, such as tweet count and fan numbers, to create a comprehensive representation for each user. Then a self-loop attention based GCN is designed to dynamically assess and adapt to the varying importance of heterogeneous users in the information aggregation process of GCN instead of fixed coefficients. Moreover, a variational autoencoder is employed for each user at every timestamp to learn the latent distribution space of becoming a source, allowing us to develop a dual-structure based Kullback-Leibler (KL) regularized constraint module, which realizes the practical experience that the behavior and determination of users with analogous behavioral characteristics regarding similar topics are likely to stem from approximate distribution spaces. To further refine our model, a composite loss function is incorporated to enable DSLF to capture and learn the propagation dynamics in dual platforms more accurately. The major contributions of DSLF are as follows.

- A topic-aligned propagation cascade dataset incorporating user profiles for Twitter and Weibo is developed to bridge the 'simulation-to-reality' gap in source localization research. And a dual-channel source localization framework (DSLF) for cascades with similar topics in different platforms is proposed, validating its practicality in real-world multiplex network scenarios. Extensive experiments further demonstrate that DSLF can be flexibly extendable to a single or more layer network.
- A self-loop attention-based GCN is designed to refine the neighborhood aggregation scheme, learning a higher-quality representation of graph structures and user profiles. which improves the source detection performance in real-world scenarios.
- A dual-structure based KL regularization is designed to focus on implicit joint features in the latent distribution space of the source probability between platforms. By realizing the alignment of behavioral distribution characteristics for users with similar profile levels about similar topics, the robustness of the model can be enhanced.

2 Related Work

2.1 Diffusion Models

In the last few years, many diffusion models have been proposed to capture the diffusion feature and simulate the propagation data which is applied for evaluating the performance of localization methods, such as the Susceptible-Infected (SI) model and the Susceptible-Infected-Recovered (SIR) model [Yang *et al.*, 2020; Paluch *et al.*, 2021; Zang *et al.*, 2015; Zhu and Ying, 2014; Tang *et al.*, 2018]. Further, recognizing that in reality, every individual in social networks has unique attributes, heterogeneous diffusion models like the heterogeneous SI (HSI) and heterogeneous SIR (HSIR) are proposed to consider varied infection and recovery rates respectively [Karrer and Newman, 2010; Ellison, 2020]. However, these models are still fundamentally based on the mean-field theory or assumptions of propagation dynamics. Therefore, it is necessary to utilize real-world propagation cascades to enhance the performance demonstration and expand the application scenarios of downstream localization methods.

2.2 Source Localization Methods

Due to the convenience and feasibility of snapshot acquisition, many works focus on snapshot based study. Dong et al. introduce a GCN based source identification model to tackle the multiple rumor source detection problem [Dong *et al.*, 2019]. Additionally, some methods construct the dynamic features of propagation before embarking on the source inference process, such as IVGD [Wang *et al.*, 2022a], MCGNN [Shu *et al.*, 2021] and SL_VAE [Ling *et al.*, 2022]. However, these methods do not consider the cross-platform nature of propagation. Although some works focus on source identification in multiplex networks [Paluch *et al.*, 2020b], some parameters of the underlying propagation model required as inputs in the inference are quite difficult to obtain in reality. What's more, these methods, grounded in simulated data for localization, do not consider the impact of real-world user profiles on propagation. In reality, individuals are the primary drivers of information diffusion. Therefore, distinguishing users while differentiating their characteristics in the source inference process can enhance the applicability in practical scenarios.

3 Method

3.1 Preliminary

Propagation Cascades

Focused on some specific topics or topics, we obtain K_1 number of available experienced historical propagation cascades $\mathcal{C}_k^{P_1}$=$(\mathcal{V}_k^{P_1}, \mathcal{E}_k^{P_1}, \mathcal{F}_k^{P_1})$ $(1 \leq k \leq K_1)$ from platform Twitter, and K_2 historical cascades $\mathcal{C}_k^{P_2}$=$(\mathcal{V}_k^{P_2}, \mathcal{E}_k^{P_2}, \mathcal{F}_k^{P_1})$ $(1 \leq k \leq K_2)$ from platform Weibo, where $\mathcal{V}_k^*$ is the participant user set with UID in a social media platform, $\mathcal{E}_k^*$ is the set of participant's directed propagation interaction (including comments or retweets from a user to another), and $\mathcal{F}_k^*$ is the feature set (i.e., profiles) associated with the users, including verification status (indicating authoritativeness), number of tweets (representing tweet activity), registration date (indicative of user seniority), number of fans (denoting popularity), number of followings (indicating information seeking behavior), ratio of fans to followings (reflecting credibility).

Historical Relationship Network

Drawing from K^* historical cascades $\mathcal{C}_k^*$=$(\mathcal{V}_k^*, \mathcal{E}_k^*, \mathcal{F}_k^*)$ in a social media platform $*$, we construct the historical relationship network $\boldsymbol{\mathcal{G}}^*$=$(\boldsymbol{\mathcal{V}}^*, \boldsymbol{\mathcal{E}}^*, \boldsymbol{\mathcal{F}}^*)$, which is a union graph by combining structural information of different cascades based on the same UIDs. Sincerely, we pick this idea from the field of diffusion inference [Ramezani *et al.*, 2023], where it is widely used as an intuitive yet effective approach when the underlying network is unknown. Specifically, if different cascades are interconnected by the presence of the same UID, it often implies that these cascades are not isolated incidents but indicate potential underlying similarities or connections driven by consistent shared interests or themes. Therefore, a connected historical relationship network represents a group of people (i.e., community) characterized by topic relevance, friendship ties, or similar attributes, even though these may not be explicitly known. Focusing on this distinct identified area, our investigative effort is to locate the sources from a new propagation within $\boldsymbol{\mathcal{G}}^*$.

Problem Definition

Having constructed the historical relationship network $\boldsymbol{\mathcal{G}}^{P_1}$=$(\boldsymbol{\mathcal{V}}^{P_1}, \boldsymbol{\mathcal{E}}^{P_1}, \boldsymbol{\mathcal{F}}^{P_1})$in Twitter and $\boldsymbol{\mathcal{G}}^{P_2}$=$(\boldsymbol{\mathcal{V}}^{P_2}, \boldsymbol{\mathcal{E}}^{P_2}, \boldsymbol{\mathcal{F}}^{P_2})$ in Weibo, as for a new propagation cascade of a focal topic or topic spreading in the concerned area $\boldsymbol{\mathcal{G}}^{P_1}$ and $\boldsymbol{\mathcal{G}}^{P_2}$, denoted as C^{P_1} and C^{P_2}, respectively, we only conveniently and randomly observe two collections of available snapshots $\{V_{t_1}^{P_1}, V_{t_2}^{P_1}, ..., V_{t_n}^{P_1}\}$ and $\{V_{t'_2}^{P_2}, V_{t'_2}^{P_2}, ..., V_{t'_n}^{P_2}\}$, where $V_{t_i}^*$ is the set of participants in area $\boldsymbol{\mathcal{G}}^*$ at a timestamp t_i. And we denote the original rumor sources set as $R^* \subset \boldsymbol{\mathcal{G}}^*$. The goal of our method is to predict a source set $\hat{R}$ which can maximize the indicator like $\frac{\hat{R} \cap R^*}{\hat{R} \cup R^*}$.

3.2 Overall Framework

We have temporal data for each user at each timestamp in each platform. However, while the recurrent module typically processes sequence data, it is also expected to learn network topology information inductively. Hence in a single platform $*$, before inputting temporal data in a sequential model, it is required to incorporate graph information. Considering the excellent embedding ability of GNNs on graph topology, an intuitive and feasible way is to design a graph-based recurrent layer in the update process. A simplified, vectorized version of the graph-based recurrent layer can be specified as $h_{t_i} = \sigma\Big(AH^{*t_i}W_{i\text{-}h} + b_{i\text{-}h} + Ah_{t_{i-1}}W_{h\text{-}h} + b_{h\text{-}h}\Big)$, where σ is the activation function, h_{t_i} is the hidden layer output for all nodes at a timestamp t_i, A is a variant matrix of a certain topological structure. $W_{i\text{-}h}$ and $W_{h\text{-}h}$ are the learnable input-hidden and hidden-hidden weights, $b_{i\text{-}h}$ and $b_{h\text{-}h}$ are the corresponding bias. Based on this, we develop an encoder-decoder framework tailored for enhancing temporal data through graph-based feature aggregation, called dual-channel source localization framework (DSLF). As illustrated in Fig. 2, after constructing unique characteristics for each user in each snapshot, a self-loop attention based GCN encoder primarily aggregates topology information into temporal data. Further, a dual-structure based KL regularization decoder is designed to sequence and regularize the augmented temporal data from a multi-channel perspective. Finally, a joint loss is designed to optimize the overall performance by enhancing the single-channel specificity and ensuring robust multi-channel alignment.

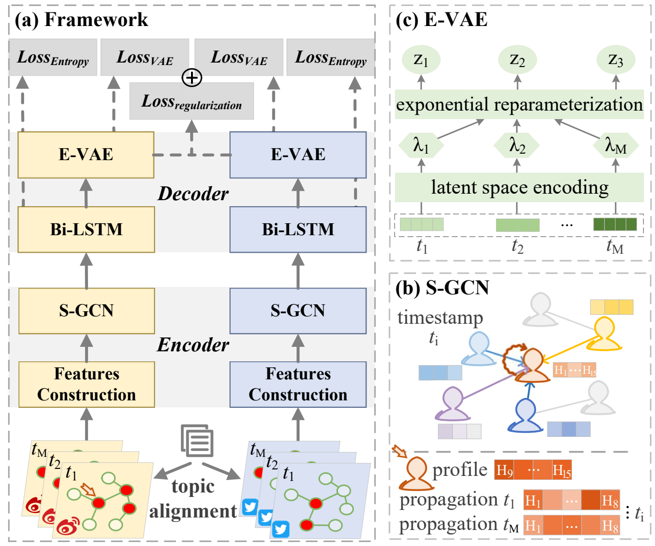

Figure 2: The overview of DSLF, where t_1-t_M are available snapshots with discrete timestamps, (a) is the structure of DSLF, (b) is the self-loop attention based GCN (S-GCN), (c) is the exponential variational autoencoder (E-VAE) sampling from an exponential distribution.

3.3 Features Construction

For a series of propagation snapshots $(V_{t_1}^*, V_{t_2}^*, ...)$ of a topic, historical relationships in $\boldsymbol{\mathcal{G}}^*$ is used to extract a knowledge based subgraph $G_{t_i}^*$ or adjacency $A_{t_i}^*$ in each timestamp. To ensure the structural consistency of each $A_{t_i}^*$ across different timestamps which is a preprocessing idea for the deep learning module, and to perceive the future potential participants from the experience in $\boldsymbol{\mathcal{G}}^*$, first, we unify the largest subgraph $\tilde{G}^*(\tilde{V}^*, \tilde{E}^*)$ or $\tilde{A}^*$ from the focused area $\boldsymbol{\mathcal{G}}^*$ by extracting additional one-hop relationships of $V_{max(t_i)}^*$.

$$\tilde{V}^* = \{u \mid v \in V_{max(t_i)}^*, u \in N^{\boldsymbol{\mathcal{G}}^*}(v)\} \cup V_{max(t_i)}^*, \quad (1)$$

$$\tilde{E}^* = \{(v_i, v_j) \mid v_i, v_j \in \tilde{V}^* \text{ and } (v_i, v_j) \in \boldsymbol{\mathcal{E}}^*\}, \quad (2)$$

where $N^{\boldsymbol{\mathcal{G}}^*}(v)$ is the neighbor set of user v in a single historical network $\boldsymbol{\mathcal{G}}^*$. Then, a single snapshot $V_{t_i}^*$ can be mapped onto $\tilde{G}^*$ and is denoted as $\hat{G}^*_{t_i}(\hat{V}^*_{t_i}, \hat{E}^*_{t_i}, \hat{Y}^*_{t_i})$. Here, $\hat{Y}^*_{t_i}(v_j) = 1$ indicates that a user v_j has participated our concerned topic in platform $*$ at timestamp t_i. Simultaneously, the state and characteristics of users in each timestamp are retained and the topological consistency of different timestamps is also guaranteed (i.e., $|\hat{V}^*_{t_1}|$=$|\hat{V}^*_{t_2}|$=$|\hat{V}^*_{t_i}|$=$|\tilde{V}^*|$).

After obtaining a series of snapshot based subgraphs with topological consistency $\hat{G}^*_{t_1}, \hat{G}^*_{t_2}, ...$, a sequence-to-sequence model is designed to locate the source user of a topic propagation in $\boldsymbol{\mathcal{G}}^*$. First in the encoder phase, to better solve the user-level-based source localization task, some unique propagation dynamic features are designed for each user combined with user profiles to differentiate each unique user. There are eight explicit dynamic indicators at a timestamp t_i (denoted as $H_1^{*t_i}$-$H_8^{*t_i}$) are constructed to characterize the time-varying based dynamic features of an individual at a timestamp t_i in a single platform $*$. Among them, the ratio of participated neighbors and non-participated neighbors of v_j at a timestamp t_i are shown in Eq. (3) and Eq. (4), respectively.

$$H_1^{*t_i}(v_j) = \frac{\sum_{v_k \in \mathcal{N}^{\hat{G}^*_{t_i}(v_j)}} Y^*_{t_i}(v_k)}{|\mathcal{N}^{\hat{G}^*_{t_i}}(v_j)|}, \tag{3}$$

$$H_2^{*t_i}(v_j) = 1 - H_1^{*t_i}(v_j). \tag{4}$$

What's more, we also consider the normalized number of participated and non-participated neighbors of v_j at a timestamp t_i are shown in Eq. (5) and Eq. (6), respectively.

$$H_3^{*t_i}(v_j) = \frac{\sum_{v_k \in \mathcal{N}^{\hat{G}^*_{t_i}(v_j)}} Y^*_{t_i}(v_k)}{\max_{u \in \hat{V}^*_{t_i}} \left(\mathcal{N}^{\hat{G}^*_{t_i}}(u)\right)}, \tag{5}$$

$$H_4^{*t_i}(v_j) = \frac{|\mathcal{N}^{\hat{G}^*_{t_i}}(v_j)| - \sum_{v_k \in \mathcal{N}^{\hat{G}^*_{t_i}(v_j)}} Y^*_{t_i}(v_k)}{\max_{u \in \hat{V}^*_{t_i}} \left(\mathcal{N}^{\hat{G}^*_{t_i}}(u)\right)}. \tag{6}$$

Here, features H_1-H_4 indicate that we are not solely focused on the dynamic ratio of neighbor users. Both the total number of participated and non-participated neighbors emphasize our concern for the precise count of neighbors' states, not just their proportions. For example, considering that a user only has one neighbor and the neighbor participates in the topic, then H_1 is a relatively large feature indicator. However, indicator H_3 of such a user is small. Therefore, both normalized numerical features and proportional features need to be considered. Moreover, the original state $Y^*_{t_i}(v_j)$ of each user in $\hat{G}^*_{t_i}$, whether participated (H_5) or non-participated (H_6), collectively represents the essential property of the individual. Furthermore, the normalized snapshot sequence index (H_7) is introduced to indicate the relative timing of each user's first appearance in the sequence of captured snapshots. Additionally, we also pay attention to the degree centrality (H_8) [Simmie *et al.*, 2013] which can reflect the celebrity effect in social networks. After these eight propagation dynamic features are obtained, the complete feature embedding $H^{*t_i}(v_j) \in \mathbb{R}^{15}$ at a timestamp t_i in a single platform $*$ can be obtained by concatenating the six normalized user profile features in $\boldsymbol{\mathcal{F}}^*(v_j)$ (verification status is a binary variable with two dimensions).

3.4 Self-loop Attention Based GCN

Since we have the features of users at different timestamps, we can input the temporal information into a recurrent model in a sequence-input-oriented way. However, direct input may result in the loss of prior information about the topological structure. Therefore, we can utilize the topology-based aggregation properties of GNNs. And in each platform, the original GCN at a timestamp t_i for feature aggregation of $H^{*t_i} \in \mathbb{R}^{|\hat{V}^*_{t_i}|*15}$ in the spectrum domain convolutions is defined as follows:

$$H^{*t_i} \leftarrow \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{*t_i}W\right), \tag{7}$$

where W are the learnable weights in the module. $\tilde{A} = A + I$, where A is the corresponding adjacency matrix of $\hat{G}^*_{t_i}$ and I is an identity matrix. $\tilde{D}$ is the corresponding degree matrix of $\tilde{A}$. However, in the aggregation process of single-layer GCN, we observe that nodes with higher degrees, despite having a larger number of interacting neighbors, often contribute with relatively lower feature weights in the aggregation, impacting both themselves and their neighboring nodes. More details can be seen from Fig. 3. This tendency highlights a nuanced challenge within the single-layer GCN module, where the influence of highly connected nodes might be diminished in the aggregation process. Although as the number of GCN layers increases, the influence of a node is proportional to its degree, the limited extent to which GCNs with an appropriate number of layers enhance the aggregation of features for high-degree nodes may not sufficiently address the required focus on the importance of celebrities. This conclusion is revealed from our comprehensive statistical analysis [Hou *et al.*, 2024b] for propagation cascades datasets in Twitter and Weibo, that is, high-influence users in social networks, such as celebrities with large fans, often engage in topics that attract a vast number of directed participants and have a great impact on these neighbors. In other words, more attention than that of the original GCN with several layers is demanded to enhance the impact of celebrities, and contrarily, the impact of fringe users can be weakened.

Manually setting the coefficient of each user through numerous experiments is a time-consuming process. Therefore, we propose a self-loop attention based GCN to revise the coefficient weight of propagation dynamic features and profiles $H^{*t_i}(v_j)$ of each user during the process of information aggregation. In this way, celebrities can weaken the average of feature influence by its neighbors, so as to better reflect the real-world level of influence of their characteristics during the aggregation process. As shown in Eq. (8), we add a learnable diagonal matrix to personalize the element value on the diagonal of the matrix $\Lambda \in \mathbb{R}^{|\hat{V}^*_{t_i}|*|\hat{V}^*_{t_i}|}$.

$$H^{*t_i} \leftarrow \sigma\left[(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}} + \Lambda)H^{*t_i}W\right]. \tag{8}$$

The diagonal elements of this matrix are controlled by a multi-head attention mechanism. And a single-layer BP neural network $\vec{a} \in \mathbb{R}^{15}$ is applied for each head of the attention

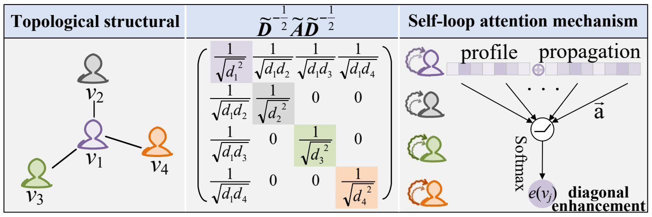

Figure 3: The message passing strategy in a single-layer GCN and the enhancement strategy based on a self-loop attention mechanism. For the normalized form of v_1, it can be seen that the value of the diagonal element is smaller than the value of its neighbor elements due to the larger degree of v_1 (e.g., $d_1^2 > d_1 d_2$). Therefore, v_1 allocates a diminished influence to each of its adjacent neighbors if only one graph convolutional layer is used. This observation is contrary to the established understanding in media scenarios where celebrities with higher fans and followers should naturally exert a more significant influence on others. To solve this, a self-loop attention mechanism $\vec{a}$ is proposed to evaluate user's influence through his profiles and dynamic propagation features in order to enhance the normalized matrix's diagonal elements of GCN.

mechanism. To make coefficients easily comparable across all users, we normalize the self-loop attention mechanism.

$$\begin{aligned}\phi(v_j) &= \operatorname{softmax}_{v_k \in \hat{V}^*_{t_i}}\left(e(v_k)\right) = \frac{\exp\left(e(v_j)\right)}{\sum_{v_k \in \hat{V}^*_{t_i}} \exp\left(e(v_k)\right)} \\ &= \frac{\exp\left(\operatorname{ReLU}\left(\vec{a}^T\left[{W_A}^T H^{*t_i}(v_j)\right]\right)\right)}{\sum_{v_k \in \hat{V}^*_{t_i}} \exp\left(\operatorname{ReLU}\left(\vec{a}^T\left[{W_A}^T H^{*t_i}(v_k)\right]\right)\right)},\end{aligned} \tag{9}$$

where W_A is the learnable matrix in the attention module, and $e(v_j)$ is the self-loop attention coefficient of v_j. Further, the multi-head attention can be executed for a user and the average operator is selected as the pooling process. Furthermore, by diagonalizing the normalized self-loop attention coefficients $\phi(v_j)$ for each user, the matrix Λ can be obtained. Λ allows for the automated and dynamic adjustment of influence coefficients for each node, catering to its unique role in the network. The essence of this matrix is to "personalize" the influence of each node by enhancing or diminishing its diagonal element value, which is learned from data. This personalized approach is not just about optimizing influence but ensuring the model's adaptability and scalability.

3.5 Bi-LSTM for Sequential Propagation Features

After the topology aggregation based features H^{*t_i} at each timestamp t_i in a single platform $*$ is updated from the self-loop attention based GCN, a sequential decoder is developed to evaluate the source probability of each user based on the discrete temporal embedding. Propagation information from earlier stages is the prior knowledge for later source localization, and the inverse process of forward diffusion is the localization task. So we use a bidirectional lightweight time series module, Bi-LSTM, for source prediction. We input the augmented embedding of snapshot sequence $\{H^{*t_1}, H^{*t_2}, H^{*t_3}, \ldots\}$ in a single platform $*$ into a Bi-GRU model, and the forward LSTM infer the source based on the snapshots from the timestamp t_1 to $\max(t_i)$, and we denote the output predicted binary classification for all nodes of the hidden layer at the timestamp t_i as $\overrightarrow{h}_{t_i} \in \mathbb{R}^{|\hat{V}^*_{t_i}| \times 2}$. Similarly, the backward output is represented as $\overleftarrow{h}_{t_i} \in \mathbb{R}^{|\hat{V}^*_{t_i}| \times 2}$. Finally, the two hidden layer states of the timestamp t_i are concatenated, i.e., $\hat{O}^*_{t_i} = \left[\overrightarrow{h}_{t_i}, \overleftarrow{h}_{t_i}\right] \in \mathbb{R}^{|\hat{V}^*_{t_i}| \times 4}$.

3.6 E-VAE for User's Latent Distribution Space

Having identified the source features on a single platform from $\hat{O}^*_{t_i}$ in a coarse-grained way, it is important to note that even when appearing in different platforms, certain characteristics of a similar topic, such as the behaviors of participants involved and their attitudes towards the topic, tend to remain potentially consistent. These invariant features provide valuable insights into improving the detection performance and robustness of source localization. Therefore, based on the above considerations, we start from the perspective of Weibo and Twitter data and propose a dual-structure based KL regularization to capture the consistent implicit features of the cascades with similar topics in different platforms, which not only identifies the unique factors of each platform but also bridges all channels to reveal a comprehensive alignment of user behavior and decision-making outcomes.

In practice, users with equivalent behavioral characteristics, influence, or profiles, are likely to exhibit decision-making and actions stemming from approximate distribution spaces when dealing with the same topics. Therefore, the source probability of similar users across different platforms, as a reflection of their decision outcomes, is also highly likely to originate from similar distribution spaces. Drawing insights from the phenomenon analysis of the propagation cascades [Hou *et al.*, 2024b; Hou *et al.*, 2024a], we have developed an exponential variational autoencoder (E-VAE) instead of normal distribution sampling.

$$\log q_{\boldsymbol{\phi}}\left(\mathbf{z} \mid \hat{O}^*_{t_i}(v_j)\right) = \log \operatorname{Exp}\left(\mathbf{z}; \boldsymbol{\lambda}^{(v_j)}\right), \tag{10}$$

where $q_{\boldsymbol{\phi}}(\mathbf{z} \mid \hat{O}^*_{t_i}(v_j))$ represents the approximate posterior distribution of the latent variable $\mathbf{z}$ parameterized by $\boldsymbol{\phi}$, conditioned on the source probability evaluation of a user v_j at a timestamp t_i in a single platform $*$. Eq. (10) implies that the distribution of the latent variable $\mathbf{z}$ is inferred from the data output by the sequence model. The sequence LSTM model encapsulates temporal dependencies, and then the coarse-grained source features predicted by LSTM are leveraged to guide the VAE framework in learning the latent distribution spaces of user's decision-making outcomes. In this case, samples are drawn from an exponential distribution. Our goal is to model the latent variables $\mathbf{z}$ by using an exponential distribution when an observed data point user v_j is given. Therefore, $\log \operatorname{Exp}(\mathbf{z}; \boldsymbol{\lambda}^{(v_j)})$ represents the natural logarithm of the probability density function of an exponential distribution with a rate parameter $\boldsymbol{\lambda}^{(v_j)}$. Moreover, to ensure efficient backpropagation of gradients and avoid the potential issue of exploding exponentials within E-VAE, we use a combination of specific transformations and the reparameterization trick[1].

[1]The interpretability of E-VAE and the corresponding KL divergence are proven in the Appendix.

3.7 Loss Function

Specifically, given users from both platforms, their latent distribution space of decision-making outcomes from the perspective of source probability can be obtained in Sec. 3.6. We now introduce the KL divergence constraints for approximating the distribution space of similar users across two platforms. Utilizing the chi-squared test from statistical analysis [Hou *et al.*, 2024b], we evaluate the influence of each user based on user profiles of $\mathcal{F}_k^{P_1}$ and $\mathcal{F}_k^{P_2}$. Then, as for the cascades with similar topics in different platforms, we read out the influence of users, rank the users on each platform in descending order of their influence, and perform cross-platform matching and alignment of users. Further, following an interpretable derivation, we design a decision-making regularization loss function to minimize the KL divergence between the two distributions for users in different platforms but with similar influences[2]. Without loss of generality, we define $P_f(\hat{Y}) = q_{\boldsymbol{\phi}}(\mathbf{z}|\hat{O}_{t_i}^{P_1}(v_j))$ as the distribution space of decision-making outcomes (e.g., source probability) for a user v_j in the Twitter platform, and $P_g(\hat{Y}) = q_{\boldsymbol{\phi}}(\mathbf{z}|\hat{O}_{t_i'}^{P_2}(v_j'))$ as that for a user v_j' in the Weibo platform.

$$\begin{aligned}\mathcal{L}_c(P_f(\hat{Y}), P_g(\hat{Y})) =& \frac{1}{2}[KL(P_f(\hat{Y}) \parallel P_g(\hat{Y})) \\ &+ KL(P_g(\hat{Y}) \parallel P_f(\hat{Y}))].\end{aligned} \tag{11}$$

Further, a unique loss function for DSLF is designed to better adapt to the task of source localization:

$$\begin{aligned}\mathcal{L}oss = \mathcal{L}_{\textit{w-Entropy}}^{P_1}(\hat{R}^{P_1}, R^{P_1}) + \mathcal{L}_{\textit{w-Entropy}}^{P_2}(\hat{R}^{P_2}, R^{P_2}) + \\ \alpha * (\mathcal{L}_{\textit{E-VAE}}^{P_1}(\hat{O}^{P_1}) + \mathcal{L}_{\textit{E-VAE}}^{P_2}(\hat{O}^{P_2})) + \beta * \mathcal{L}_c(P_f(\hat{Y}), P_g(\hat{Y})).\end{aligned} \tag{12}$$

The $\mathcal{L}oss$ encompasses the weighted (focusing on a small number of source nodes) binary cross entropy loss of source prediction task for each platform, E-VAE loss (including reconstruction error and KL divergence between the latent representation and the standard exponential distribution) for each platform, and regularization loss across both platforms. Here, α and β are served as the coefficients to balance each item.

4 Experiments

4.1 Experimental Setup

We use three datasets collected from two real-world social media platforms, Weibo and Twitter, for locating the sources of cascades with similar topics in different platforms [Liu *et al.*, 2015; Ma *et al.*, 2016; Ma *et al.*, 2017].

Statistic	Twitter	Weibo
#users	677,640	2,856,741
#users in $\boldsymbol{\mathcal{G}}^*$	677,058	2,856,519
#relations in $\boldsymbol{\mathcal{G}}^*$	828,546	3,508,596
#cascades	2,308	4,664
#topic-aligned cascades	1,831	1,831

Table 1: Statistics and relevant information of the datasets.

[2]The derivation of Eq. (11) is demonstrated in the Appendix.

And we consider TGASI [Hou *et al.*, 2023], IVGD [Wang *et al.*, 2022a], SL_VAE [Ling *et al.*, 2022], GCSSI [Dong *et al.*, 2022], and MCGNN [Shu *et al.*, 2021] for comparison. And to demonstrate the source prediction performance of all methods rigorously, the widely used $F1\text{-}score = \frac{2*Precision*Recall}{Precision+Recall}$ is chosen as the evaluation metric [Wang *et al.*, 2023].

In our experiments, we employ a 10-fold cross-validation strategy to divide the training and test datasets. Further, DSLF utilizes the training dataset for learning, and then the final result is output by averaging the prediction across each fold in the test dataset. Moreover, an early stopping mechanism is designed to avoid over-fitting in the training process. For optimization, the Adam optimizer is used, configured with a learning rate of 0.0005 for all model parameters. In the loss function, both α and β are set to 0.5. More settings can be found in the Appendix of supplementary files.

It's noteworthy that to demonstrate the generalizability of DSLF, we employ widely recognized simulated synthetic datasets for source localization, generated based on the IC and HSI propagation models, as commonly used in traditional localization methods [Ling *et al.*, 2022; Wang *et al.*, 2022a]. Similar to the settings of those studies, we select 10% of the nodes as ground-truth sources and simulated the IC and HSI propagation processes on a network with 4,039 nodes and 88,234 edges. The infection rate of each node follows the uniform distribution of $U(0.05, 0.15)$ [Hou *et al.*, 2023]. And we independently generate 1,000 sets of propagation data, each including several snapshots with distinct timestamps. Since the experimental datasets from the SOTA methods we follow are single-layer networks, only one channel of DSLF is used and the cross-platform based regularization loss is removed.

4.2 Overall Experimental Results

The source detection performance based on the real-world dataset and the paradigm of the traditional localization methods is illustrated in Tab. 2. When benchmarked against the optimal baseline TGASI, DSLF exhibits an average improvement of 82.1% in real-world datasets and a 15.2% enhancement in simulated datasets from the propagation models. In conclusion, DSLF outperforms all the SOTA methods based on rigorous metrics in all datasets. There are three key reasons for the significant improvement in real-world datasets: (1) The dynamic user features and profiles are constructed to enhance individual uniqueness in user-driven-based real-world propagation scenarios. (2) The self-loop attention-based GCN module refines neighborhood aggregation, improving graph and profile representation quality. (3) The KL regularization module bolsters cross-platform robustness by aligning behavioral distributions of similar user groups.

Furthermore, in simulated datasets, the detection performance of DSLF, although superior to SOTA methods, shows a less pronounced improvement compared to real-world datasets, for two reasons: (1) The simulated data lacks user profile information. (2) The KL regularization module can not be used in single-layer scenarios. It's noteworthy that DSLF transfers effectively in single-layer networks, indicating better generalizability. In summary, its continued higher

F1-score in single-layer networks than SOTA methods highlights the superiority of its designed modules.

Datasets	Real-world		Simulation	
Group	*Twitter*	*Weibo*	*HSI*	*IC*
TGASI [Hou *et al.*, 2023]	0.511	0.485	0.647	0.671
IVGD [Wang *et al.*, 2022a]	0.366	0.321	0.376	0.403
SL_VAE [Ling *et al.*, 2022]	0.352	0.340	0.318	0.364
GCSSI [Dong *et al.*, 2022]	0.225	0.265	0.164	0.185
MCGNN [Shu *et al.*, 2021]	0.271	0.188	0.121	0.155
DSLF (Ours)	**0.911**	**0.902**	**0.742**	**0.776**

Table 2: Source identification performance based on the real-world dataset and the simulated synthetic dataset based on F1-score metric.

4.3 Ablation Study for Modules in DSLF

We further study the influence of designed components of DSLF in the source detection performance to prove each part of the contribution. As shown in Tab. 3, it will lead to a performance decrease or a delay in the convergence speed of model training no matter removing or exchanging any critical modules. Due to the limited space, more discussion is demonstrated in the Appendix.

Modules	*Variants*	*Twitter*	*Weibo*	*Epoch*
Encoder	-Dynamic	0.841	0.833	5
	-Profiles	0.857	0.822	3
	-Attention	0.887	0.869	4
	-SGCN	0.836	0.801	5
Decoder	-EVAE	0.832	0.841	4
	EVAE→NVAE	0.903	0.875	3
	-KL	0.842	0.822	8
Loss	w-Entropy→MSE	0.799	0.754	5
	w-Entropy→Entropy	0.856	0.845	6
Origin	DSLF	**0.911**	**0.902**	**3**

Table 3: The performance evaluation of variant models from DSLF. - signifies that the corresponding modules are removed or masked with zeros. And the symbol → indicates that the original module is replaced by a new part.

4.4 Ablation Study for User Profiles

One highlight of our work is the consideration that propagation is user-driven. Therefore, individual characteristics may provide positive knowledge for source inference. Given this, we have collected profile information of nearly one million users based on actual UIDs in cascades. From Sec. 4.3, the source detection performance would decline if we do not consider the user profiles. Further, an essential study is determining which characteristics are more important and deserve more attention in the propagation dynamics of social media. Therefore, we conduct ablation experiments on all characteristics by keeping the other settings unchanged and masking only the selected characteristic. As can be seen from Tab. 4, these characteristics do not contribute equally to the localization performance. As the most important feature, a higher "ratio of fans to followings" suggests greater influence, while a lower ratio indicates a stronger capacity for information-seeking. Abundant numerical meaning ensures its highest importance. Furthermore, another notable conclusion is that the sum of each decline in performance resulting from individually masking each of the six features (15.29% summed decrease) exceeds that of all are masked simultaneously (a 6.30% decrease from the -Profile variant in Tab. 3). This suggests the presence of interaction effects among these features.

Variants	*Twitter*	*Weibo*
-Ratio	0.871 (↓4.59%)	0.855 (↓ 5.50%)
-Status	0.881 (↓ 3.41%)	0.859 (↓ 5.01%)
-Date	0.886 (↓ 2.82%)	0.867 (↓ 4.04%)
-Fans	0.890 (↓ 2.36%)	0.874 (↓ 3.20%)
-Followings	0.898 (↓ 1.45%)	0.886 (↓ 1.81%)
-Tweets	0.905 (↓ 0.66%)	0.891 (↓ 1.23%)
DSLF	**0.911**	**0.902**

Table 4: The performance evaluation of variant models by masking one of the user profiles.

4.5 Scalability

The developed dataset is based on dual-channel propagation and further DSLF is verified based on two layers. In fact, DSLF can be extended to any arbitrary layers of networks. This scalability is demonstrated by feature-driven learning processes adaptable to different scales of networks, universal applicability of self-loop attention for diagonal enhancement, flexible VAE module sampled from different distributions (we also provide another VAE sampling from the normal distribution in our source code), and flexible dual-channel KL constraints for any two layers in multiplex scenarios. More descriptions can be found in the Appendix. And to prove the feasibility, we randomly divided all 6,900 cascades from Weibo or Twitter datasets into more platforms to test the time complexity of DSLF.

Datasets	*F1-score*	*Epoch*	*second/iteration*
1 layers	0.933	2	4.1485
3 layers	0.923	3	4.6233
4 layers	0.894	3	5.1028
5 layers	0.901	3	5.5328

Table 5: The performance evaluation in different layers scenarios.

5 Conclusion

By developing a topic-aligned propagation cascade dataset enriched with user profiles, we bridge the gap between simulation and reality in source localization. Further, the DSLF method is developed which can infer the source in networks with any number of layers. Ablation experiments demonstrate the effectiveness of crawled user profiles, constructed propagation dynamic features, self-loop attention-based GCN, and dual-structure KL regularization modules of DSLF. In the future, we will focus on collecting multi-layered data to expand and refine our models, broadening the scope of our research and its applicability.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. U22B2036, 62271411, U22A2098, 11931015, 61976181), Fok Ying-Tong Education Foundation, China (No. 171105), the Fundamental Research Funds for the Central Universities (No. D5000230366), and the Tencent Foundation and XPLORER PRIZE. Dr. Xianghua Li is the corresponding author.

References

[Dong *et al.*, 2019] Ming Dong, Bolong Zheng, Nguyen Quoc Viet Hung, Han Su, and Guohui Li. Multiple rumor source detection with graph convolutional networks. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 569–578, 2019.

[Dong *et al.*, 2022] Ming Dong, Bolong Zheng, Guohui Li, Chenliang Li, Kai Zheng, and Xiaofang Zhou. Wavefront-based multiple rumor sources identification by multi-task learning. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2022.

[Ellison, 2020] Glenn Ellison. Implications of heterogeneous sir models for analyses of covid-19. Technical report, National Bureau of Economic Research, 2020.

[Hou *et al.*, 2023] Dongpeng Hou, Zhen Wang, Chao Gao, and Xuelong Li. Sequential attention source identification based on feature representation. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 4794–4802, 2023.

[Hou *et al.*, 2024a] Dongpeng Hou, Chao Gao, Xuelong Li, and Zhen Wang. Dag-aware variational autoencoder for social propagation graph generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 8508–8516, 2024.

[Hou *et al.*, 2024b] Dongpeng Hou, Shu Yin, Chao Gao, Xianghua Li, and Zhen Wang. Propagation dynamics of rumor vs. non-rumor across multiple social media platforms driven by user characteristics. *arXiv preprint arXiv:2401.17840*, 2024.

[Islam *et al.*, 2020] Md Saiful Islam, Tonmoy Sarkar, and Sazzad Hossain Khan. Covid-19–related infodemic and its impact on public health: A global social media analysis. *The American Journal of Tropical Medicine and Hygiene*, 103(4):1621, 2020.

[Ji *et al.*, 2017] Feng Ji, Wee Peng Tay, and Lav R Varshney. An algorithmic framework for estimating rumor sources with different start times. *IEEE Transactions on Signal Processing*, 65(10):2517–2530, 2017.

[Jiang *et al.*, 2016] Jiaojiao Jiang, Sheng Wen, Shui Yu, Yang Xiang, and Wanlei Zhou. Identifying propagation sources in networks: State-of-the-art and comparative studies. *IEEE Communications Surveys & Tutorials*, 19(1):465–481, 2016.

[Jin and Wu, 2021] Rong Jin and Weili Wu. Schemes of propagation models and source estimators for rumor source detection in online social networks: A short survey of a decade of research. *Discrete Mathematics, Algorithms and Applications*, 13(04):2130002, 2021.

[Karrer and Newman, 2010] Brian Karrer and Mark EJ Newman. Message passing approach for general epidemic models. *Physical Review E*, 82(1):016101, 2010.

[Ling *et al.*, 2022] Chen Ling, Junji Jiang, Junxiang Wang, and Zhao Liang. Source localization of graph diffusion via variational autoencoders for graph inverse problems. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1010–1020, 2022.

[Liu *et al.*, 2015] Xiaomo Liu, Armineh Nourbakhsh, Quanzhi Li, Rui Fang, and Sameena Shah. Real-time rumor debunking on twitter. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 1867–1870, 2015.

[Ma *et al.*, 2016] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, and Cha Meeyoung. Detecting rumors from microblogs with recurrent neural networks. In *The 25th International Joint Conference on Artificial Intelligence*, pages 3818–3824. AAAI, 2016.

[Ma *et al.*, 2017] Jing Ma, Wei Gao, and Kam-Fai Wong. Detect rumors in microblog posts using propagation structure via kernel learning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 708–717, 2017.

[Paluch *et al.*, 2020a] Robert Paluch, Lukasz G Gajewski, Janusz A Holyst, and Boleslaw K Szymanski. Optimizing sensors placement in complex networks for localization of hidden signal source: A review. *Future Generation Computer Systems*, 112:1070–1092, 2020.

[Paluch *et al.*, 2020b] Robert Paluch, Łukasz G Gajewski, Krzysztof Suchecki, and Janusz A Hołyst. Source location on multilayer networks. *arXiv preprint arXiv:2012.02023*, 2020.

[Paluch *et al.*, 2021] Robert Paluch, Łukasz Gajewski, Krzysztof Suchecki, and Bolesław Szymański. Enhancing maximum likelihood estimation of infection source localization. In *Simplicity of Complexity in Economic and Social Systems*, pages 21–41, 2021.

[Ramezani *et al.*, 2023] Maryam Ramezani, Aryan Ahadinia, Amirmohammad Ziaei Bideh, and Hamid R Rabiee. Joint inference of diffusion and structure in partially observed social networks using coupled matrix factorization. *ACM Transactions on Knowledge Discovery from Data*, 17(9):1–28, 2023.

[Shu *et al.*, 2021] Xincheng Shu, Bin Yu, Zhongyuan Ruan, Qingpeng Zhang, and Qi Xuan. Information source estimation with multi-channel graph neural network. *Graph Data Mining: Algorithm, Security and Application*, pages 1–27, 2021.

[Simmie *et al.*, 2013] D. Simmie, M.G. Vigliotti, and C. Hankin. Ranking twitter influence by combining network centrality and influence observables in an evolutionary model. In *2013 International Conference on Signal-Image Technology & Internet-Based Systems*, pages 491–498, 2013.

[Tang *et al.*, 2018] Wenchang Tang, Feng Ji, and Wee Peng Tay. Estimating infection sources in networks using partial timestamps. *IEEE Transactions on Information Forensics and Security*, 13(12):3035–3049, 2018.

[Wang *et al.*, 2022a] Junxiang Wang, Junji Jiang, and Liang Zhao. An invertible graph diffusion neural network for source localization. In *Proceedings of the ACM Web Conference*, pages 1058–1069, 2022.

[Wang *et al.*, 2022b] Zhen Wang, Dongpeng Hou, Chao Gao, Jiajin Huang, and Qi Xuan. A rapid source localization method in the early stage of large-scale network propagation. In *Proceedings of the ACM Web Conference*, pages 1372–1380, 2022.

[Wang *et al.*, 2023] Zhen Wang, Dongpeng Hou, Chao Gao, Xiaoyu Li, and Xuelong Li. Lightweight source localization for large-scale social networks. In *Proceedings of the ACM Web Conference 2023*, pages 286–294, 2023.

[Yang *et al.*, 2020] Fan Yang, Shuhong Yang, Yong Peng, Yabing Yao, Zhiwen Wang, Houjun Li, Jingxian Liu, Ruisheng Zhang, and Chungui Li. Locating the propagation source in complex networks with a direction-induced search based gaussian estimator. *Knowledge-Based Systems*, 195:105674, 2020.

[Yin *et al.*, 2024] Shu Yin, Peican Zhu, Lianwei Wu, Chao Gao, and Zhen Wang. Gamc: an unsupervised method for fake news detection using graph autoencoder with masking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 347–355, 2024.

[Zang *et al.*, 2015] Wenyu Zang, Chuan Zhou, Li Guo, and Peng Zhang. Topic-aware source locating in social networks. In *Proceedings of the 24th International Conference on World Wide Web*, pages 141–142, 2015.

[Zhu and Ying, 2014] Kai Zhu and Lei Ying. Information source detection in the sir model: A sample-path-based approach. *IEEE/ACM Transactions on Networking*, 24(1):408–421, 2014.