Sequential Attention Source Identification Based on Feature Representation

Dongpeng Hou[1,2] , **Zhen Wang**[1,2*] , **Chao Gao**[2*] and **Xuelong Li**[2]

[1]School of Mechanical Engineering, Northwestern Polytechnical University (NWPU)
[2]School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University (NWPU)
w-zhen@nwpu.edu.cn, cgao@nwpu.edu.cn

Abstract

Snapshot observation based source localization has been widely studied due to its accessibility and low cost. However, the interaction of users in existing methods does not be addressed in time-varying infection scenarios. So these methods have a decreased accuracy in heterogeneous interaction scenarios. To solve this critical issue, this paper proposes a sequence-to-sequence based localization framework called **T**emporal-sequence based **G**raph **A**ttention **S**ource **I**dentification (TGASI) based on an inductive learning idea. More specifically, the encoder focuses on generating multiple features by estimating the influence probability between two users, and the decoder distinguishes the importance of prediction sources in different timestamps by a designed temporal attention mechanism. It's worth mentioning that the inductive learning idea ensures that TGASI can detect the sources in new scenarios without knowing other prior knowledge, which proves the scalability of TGASI. Comprehensive experiments with the SOTA methods demonstrate the higher detection performance and scalability in different scenarios of TGASI.

1 Introduction

Numerous studies have demonstrated that rumors spread quickly in social networks [Qian *et al.*, 2018; Wang *et al.*, 2022b], which could be extremely harmful to society. The issues of automatically resisting rumors, i.e., content-based rumor identification and rumor source identification, have garnered much attention in recent years [Dong *et al.*, 2019; Jin *et al.*, 2017; Song *et al.*, 2021; Lao *et al.*, 2021; Yang *et al.*, 2021; Khoo *et al.*, 2020]. However, it is still difficult to effectively cut off and control the propagation of rumors in social networks if only to identify the fake news instead of locating the rumor sources. Therefore, more and more studies focus on the problem of rumor source identification in social networks based on the network structure [Jiang *et al.*, 2016; Jin and Wu, 2021].

The corresponding methods of rumor source identification are classified into two types, including infection-status-based methods and sensor-based methods [Dong *et al.*, 2019]. In
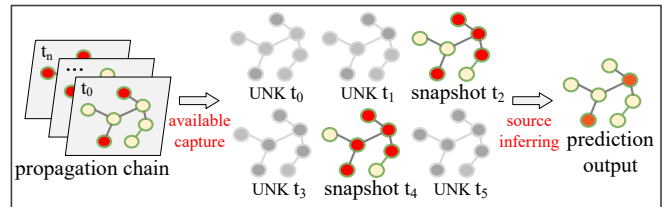

Figure 1: The illustration of the multiple rumor source detection problem under available snapshots with discrete timestamps. Due to the constraints of time and space consumption, the timestamps of snapshots we can capture are limited and discrete.

practice, the sensors sometimes cannot be presciently deployed in the break area in advance, and the sensor deployment process requires time and space overhead [Paluch *et al.*, 2020]. In contrast, the infection-status-based methods have broader application scenarios because the snapshots are easier to obtain in practice [Diepold *et al.*, 2022].

However, most of the infection-status-based methods, e.g., IVGD [Wang *et al.*, 2022a], SL_VAE [Ling *et al.*, 2022], assume that the attributes of all individuals (e.g., the infection probability or influence intensity) in a network is known in advance. Obviously, it is difficult to obtain such information in practice due to the huge labor and statistical costs. Although the SOTA methods, e.g., GCNSI [Dong *et al.*, 2019] and GCSSI [Dong *et al.*, 2022], try to identify the rumor sources without using the information of the underlying propagation model, they do not explicitly consider the user's interactive behavior when several time-varying snapshots are conveniently available. Therefore, the lack of essential behavior representation results in the limited performance of current localization methods in the complicated social network. Besides, the scalability of the model reflects the extensive application and robustness in practice [Agarwal *et al.*, 2021]. But the source localization methods so far have not considered the transfer performance of the model in new scenarios (different propagation models or different social networks, denoted as inductive learning [Michalski, 1983]). In summary, there is no sequence based inductive localization model for time-varying snapshots to consider the impact of user's behavior features on propagation.

In this paper, we study the multiple rumor source detection (MRSD) problem and focus on addressing the above

issues. Our goal is to design a generic source localization framework with transferability, especially can analyze the behavior diversity in time-varying infection scenarios. In practice, there should be a potentially heterogeneous influence between any two users in a social network no matter what the complex propagation models are. And such the corresponding influence can be evaluated from a collection of available snapshots with time-varying infection characteristics in discontinuous timestamps. So we can infer the sources by using these available captured snapshots. Such a problem is presented in Fig. 1. To this end, considering the excellent embedding ability of Graph Neural Network (GNN) [Welling and Kipf, 2016] for node attributes (or node features) and the outstanding modeling ability of the Gate Recurrent Unit (GRU) [Chung *et al.*, 2014] for time series data, we propose a novel sequence-to-sequence model called Temporal-sequence based Graph Attention Source Identification (TGASI). The input is a set of infection snapshots at available timestamps in a social network, and the output is a set of predicted sources. The TGASI model includes a GNN-based encoder to generate and aggregate the low-dimensional feature embedding of each individual, and a GRU-based decoder to infer the source with a temporal attention mechanism. The major contributions of TGASI are as follows.

- A generic sequence-to-sequence framework with transferability is designed for source localization, enabling TGASI to transfer to new diffusion models and social networks for source prediction without knowing other prior knowledge. And a unique loss function is designed to further enhance the performance of TGASI in the source localization task. Code is available at https://github.com/cgao-comp/TGASI.

- We design a deep module based on the stationary distribution of a Markov chain to learn the user's probability transition matrix in order to differentiate the user behavior diversity and consider the influence of heterogenous behavior on propagation. Then a high quality representation of the interactive behavior embedding is obtained.

- We design the dynamic infection features and static topology features to enrich the representation of low-dimensional embedding. The propagation information based dynamic features can better fit the source localization task, and topological features can better incorporate the graph-level task into a sequence based framework.

- A one-timestamp based attention mechanism is introduced to further comprehensively consider the weight of the source prediction probability at different timestamps. Dynamically distinguishing the importance of different timestamps can solve the imbalance of infection information under temporal characteristics.

2 Related Work

TGASI is a sequence based framework to solve the source localization problem based on the infection status observations. Therefore, the infection-status-based localization and sequence-to-sequence learning are summarized.

2.1 Infection-status-based Source Localization

The aim of the source localization is to infer a set of sources in a network based on some observations. Due to the convenience and feasibility of snapshot acquisition, several recent surveys on this topic are available [Jiang *et al.*, 2016; Shelke and Attar, 2019]. A novelty method LPSI designs a source prominence based label propagation method to identify the source without using the prior knowledge [Wang *et al.*, 2017]. With the development of deep learning, Dong et al. inspired by LPSI design a GCN based source identification (GCNSI) model to solve the MRSD problem [Dong *et al.*, 2019]. However, they only simply consider the centrality features of individuals. Therefore, some studies focus on more features in the source localization task, such as the SIGN method extracting the individual's dynamic features [Li *et al.*, 2021], and the MCGNN method considering the edge features [Sun *et al.*, 2021]. Moreover, the dynamic features of the propagation are also considered before executing the source inferring process, like IVGD [Wang *et al.*, 2022a], SL_VAE [Ling *et al.*, 2022]. However, all the above methods do not consider and analyze the impact of time series characteristics on source localization. In summary, there is no generic sequence-to-sequence model for source localization to comprehensively consider the behavior features under available time-varying snapshots.

2.2 Sequence-to-Sequence Learning

With the rapid development of deep learning, sequence-to-sequence learning shows great performance on various tasks based on the characteristic of time series [Sutskever *et al.*, 2014]. For example, Chung et al. propose the gated recurrent unit (GRU), whose performance is comparable to LSTM [Hochreiter and Schmidhuber, 1997] with lightweight overhead [Chung *et al.*, 2014]. And GCSSI is a sequence-to-sequence framework using GRU as a recurrence unit to learn the time series features [Dong *et al.*, 2022]. Currently, the novel encoder-decoder architecture for supervised discrete time dynamic graph (DTDG) learning focuses on the various dynamics in the real world, which brings new ideas to the embedding of temporal information [Zhu *et al.*, 2022]. TGASI is based on the one of DTDG architectures, called the static graph encoder and sequential decoder, for the MRSD task. The dynamic intuition in TGASI refers to the dynamic infection features of each individual in different timestamps. We improve such a framework to analyze the behavioral features of individuals under time-varying characteristics based on the inductive learning idea, so that the trained TGASI can be applied to different networks and different propagations.

3 Methodology

3.1 Problem Definition

Given a collection of captured available snapshots $\{\mathcal{G}_{s_j} = (V, E, Y_{s_j}) \mid j = 1, 2, 3, ...\}$ with different timestamps $s_j \in T = \{t_1, t_2, t_3, ...\}$ based on an undirected network $G = (V, E)$, where T is the timestamp set of complete propagation chain (including unavailable snapshots, i.e., $t_1 \leq s_1 < s_{n'} \leq t_n$, and $n' < n$), $Y = \{Y_1, ..., Y_{|V|}\}$ is the set of the node state, and $Y_i = 1$ if the observed state of node v_i is infected, otherwise
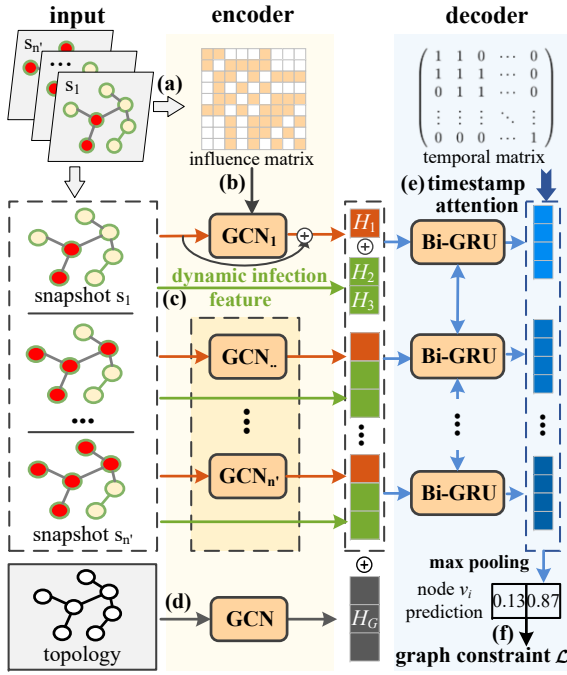
input **encoder** **decoder**

Figure 2: The illustration of TGASI based on the sequence-to-sequence framework. The features generation encoder includes four innovative modules. (a) influence probability transition matrix $\mathcal{W}$ is estimated by available captured snapshots. (b) coarse-grained source probability feature in one timestamp s is designed based on the influence matrix $\mathcal{W}$. (c) dynamic infected feature H_2 and uninfected feature H_3 is generated from each node's neighbors. (d) lower dimensional embedding H_G of the topology structure is obtained based on the topology graph. The GRU-based decoder with temporal attention uses Bi-GRU to decode the embedding, then (e) a one-timestamp based attention mechanism is designed for time series information in order to distinguish and weight the importance of decoder information at each different timestamp for source localization task. What's more, (f) a graph constraint based loss function is specially designed for the localization task to train TGASI.

Y_i=0. And we denote the original rumor sources set as $R \subset V$. The goal of our method in the MRSD problem is to predict a source set R^* which can maximize indicator like $\frac{R \cap R^*}{R \cup R^*}$.

3.2 Temporal-sequence Based Graph Attention Source Identification

In this part, an inductive framework based on the graph-level task for source localization is proposed, namely temporal-sequence based graph attention source identification. As shown in Fig. 2, the framework of TGASI is an encoder-decoder essentially. More specifically, a GNN-based encoder generates and embeds the node's features, which include coarse-grained-based source probability, dynamic infection features, and static topology features. And a GRU-based decoder with a temporal attention mechanism infers the source by decoding the embedding from the encoder.

Features Generation Encoder
Considering the complicated behavior diversity in social networks in practice, and such the potential relationship of

any two users can be characterized as a heterogenous influence. Then a deep heterogeneous influence estimation method is proposed to automatically learn the influence matrix $\mathcal{W} \in \mathbb{R}^{|V| \times |V|}$, where $\mathcal{W}_{ij}$ represents the probability that an infected-state node v_i successfully influences its uninfected neighbor v_j. Note that $\mathcal{W}_{ij} \neq \mathcal{W}_{ji}$ because people are influenced by celebrities to buy endorsement products but celebrities do not be influenced in reverse. More specifically, a propagation mechanism $f_\theta(\mathcal{G}_{s_j})$ is used to predict the infection state of the snapshot of the next timestamp $\mathcal{G}_{s_{j+1}}$. Here the f_θ is a nonlinear mapping function to characterize a propagation scheme based on the learnable matrix $\mathcal{W}$. Based on the mapping, we can construct the predicted snapshots matrix $\mathcal{G}'$, i.e.,

$$\mathcal{G}' = [f_\theta(\mathcal{G}_{s_1}), f_\theta(\mathcal{G}_{s_2}), f_\theta(\mathcal{G}_{s_3}), ...] \quad (1)$$

Note that the selection of the propagation mapping function f_θ is flexible, and it can be designed autonomously based on deep models, e.g., GCNs [Welling and Kipf, 2016] or attention-based graph models [Velickovic et al., 2018]. As for the detailed underlying design, some research has demonstrated that users in social networks have the ability to identify and prevent fraud [Baesens et al., 2015; Ventola, 2014]. So we use the IC strategy that only has one opportunity to activate the neighbors instead of the common widely used random walk that can be infected multiple times. What's more, the stationary distribution of a Markov chain revises the distribution bias of influence parameters (i.e., $\mathcal{W}$) in the diffusion process [Salamat et al., 2020]. Motivated by the stationary distribution of influence diffusion models [Xia et al., 2021], an approximate stationary distribution of a Markov chain based on the IC diffusion model is adopted for f_θ, as defined in Eq. (2).

$$f_\theta(Y_{v_i}^{s+1}) = 1 - \prod_{v_j \in \mathcal{N}(v_i)} \left(1 - \mathcal{W}_{ji}\left(Y_{v_j}^s - Y_{v_j}^{s-1}\right)\right) \quad (2)$$

where $\mathcal{N}(v_i)$ is the neighbors set of node v_i, and $Y^s \in \mathbb{R}^{|V|}$ is an available snapshot in timestamp s. The equation means that the state of node v_i in the next timestamp is approximately influenced by the heterogenous behavior of its neighbors in the timestamp s. In order to further improve the efficiency of the learning process of this propagation parameter on the graph-level task, we manually design two rules that conform to propagation. First, we set $\mathcal{W}_{ij}=0$ and remove the corresponding learning gradient if there is no edge between v_i and v_j. Second, the propagation parameter theoretically should be greater than 0, and we take the absolute value when the learning parameter $\mathcal{W}_{ij}<0$ in order to increase the penalty of MSE loss. Then we generalize the nonlinear mapping function f_θ from one node to network G, and use $f_\theta(\mathcal{G})$ to guide an MLP to evaluate the influence matrix $\mathcal{W}$.

Although there are many classic sequence-to-sequence models such as AGC-Seq2Seq [Zhang et al., 2019] and i-Revnet [Jacobsen et al., 2018; Chang et al., 2018] to directly identify the source based on the data of the snapshots, the sparse features (we only have two available features whether the state is infected or not) and lack of parameters in hidden layers restrict the ability to learn the spatial-temporal information in the localization task. So we construct some necessary feature embedding based on user behavior diversity $\mathcal{W}$.

First, We consider the coarse-grained source probability based on the behavioral interaction under each independent snapshot s, which is denoted as H_1^s. GNN can efficiently and conveniently embed the user's influence information in $\mathcal{W}$ into the message-passing process. Inspired by the invertible graph residual network, H_1^s is designed by using the following graph convolution operation.

$$
\begin{aligned}
H_1^s &= \varepsilon_{\text{GRN}}^s(Y^s, \boldsymbol{A} \mid \mathcal{W}) = \varepsilon_{\text{GCN}}^s(Y^s, \boldsymbol{A} \mid \mathcal{W}) + \alpha Y^s \\
&= \sigma\left(\widehat{\boldsymbol{D}}^{-\frac{1}{2}} \widehat{\boldsymbol{A}} \widehat{\boldsymbol{D}}^{-\frac{1}{2}} \boldsymbol{X} \boldsymbol{W}\right) + Y^s, s.t.\ L_\varepsilon < 1, \\
\widehat{\boldsymbol{A}} &= \boldsymbol{A} + \boldsymbol{I}_n + \mathcal{W}, X_i = \begin{cases} [1,0], & Y_{v_i}^s = 0 \\ [0,1], & Y_{v_i}^s = 1 \end{cases}
\end{aligned}
\tag{3}
$$

where $\boldsymbol{W} \in \mathbb{R}^{2 \times h_{one}}$ is a learnable weight matrix in the GCN module, 2 is the number of observed infection features which imply the infected state and uninfected state (denoted as X_i here), h_{one} is the hidden size of GCN in one timestamp s, $\sigma(\cdot)$ is the activation function, $H_1^s \in \mathbb{R}^{|v|}$ is the output feature of all nodes in a timestamp s, α is the residual weight, and L_ε is the Lipschitz constants to initialize $\boldsymbol{W}$. Here, we convert the observed state feature Y of snapshots into a one-hot encoding, characterizing that the infected state and uninfected state are categorical variables rather than numeric variables. The operation can enhance the learning ability and interpretability of the deep module. What's more, the learnable weight matrix $\boldsymbol{W}$ is normalized via the power iteration method to achieve the Lipschitz constant $L_\varepsilon < 1$ in order to guarantee the stronger robustness and accelerate convergence speed [Wang *et al.*, 2022a]. Then, we consider the infection characteristics of all nodes in each available timestamp s according to the source localization task, and design two infection features of node v_i in timestamp s, which is shown in Eq. (4) and Eq. (5).

$$
H_2^s(v_i) = \frac{\sum_{v_j \in \mathcal{N}(v_i), Y_j^s = 1} \mathcal{W}_{ij}}{|\mathcal{N}(v_i)|}
\tag{4}
$$

$$
H_3^s(v_i) = \frac{\sum_{v_j \in \mathcal{N}(v_i), Y_j^s = 0} \mathcal{W}_{ij}}{|\mathcal{N}(v_i)|}
\tag{5}
$$

where $H_2^s \in \mathbb{R}^{|V|}$ and $H_2^s(v_i)$ is a infected neighbor's feature of v_i in an available timestamp s, and similarly H_3 is the uninfected neighbor's feature. The H_2 and H_3 are interpretable, and the proof is in the appendix due to the limited space.

Theorem 1. The ratio of the neighbor's infected state and uninfected state are special variants of infected feature H_2 and uninfected feature H_3, respectively.

The aim of designing these embedding (i.e., H_1-H_3) is to make the decoder efficiently learn the propagation features and predict the source. However, the common input of the time series decoder is sequence based data. It is necessary to additionally design the rules if the input includes graph information. There are two reasons why we do not directly process the decoder in the form of graph recurrent layer, such as $h_t = \sigma\left(\boldsymbol{A} H^s \boldsymbol{W}_{ih} + \boldsymbol{A} h_{t-1} \boldsymbol{W}_{hh} + b\right)$. First, the format constraints of the recurrent unit $\boldsymbol{W}_{ih} \in \mathbb{R}^{3 \times 2}$ and $\boldsymbol{W}_{hh} \in \mathbb{R}^{2 \times 2}$ would lead to too few learnable parameters and reduce the

model's learning capacity. The second reason is that designing topological features in the encoder expands the dimensionality of the embedding and increases the quality and interpretability of the embedding. So we use a one-layer GCN to design the topology feature embedding in a low-dimensional way to better solve the graph-level based localization task, as shown in Eq. (6).

$$
H_G = \varepsilon_{\text{GCN}_G}(\boldsymbol{I}_n, \boldsymbol{A}) = \sigma\left(\widehat{\boldsymbol{D}}^{-\frac{1}{2}} \widehat{\boldsymbol{A}} \widehat{\boldsymbol{D}}^{-\frac{1}{2}} \boldsymbol{I}_n \boldsymbol{W}\right)
\tag{6}
$$

where $\boldsymbol{W} \in \mathbb{R}^{|V| \times \sqrt{|V|}}$ is a learnable weight matrix, and $H_G \in \mathbb{R}^{|V| \times \sqrt{|V|}}$ is the output topology features of all nodes. Therefore, we implement the feature embedding from a high-dimensional graph to a low-dimensional vector.

We get the coarse-grained source probability H_1^s, the infection features H_2 and H_3 in the timestamp s, and the network topology H_G from the encoder. And the embedding H in the timestamp s can be obtained by concatenating these features.

$$
H^s = \text{CONCAT}(H_1^s, H_2^s, H_3^s, H_G)
\tag{7}
$$

GRU-based Decoder with Temporal Attention

For the available snapshots $\{\mathcal{G}_{s_j} = (V, E, Y_{s_j}) \mid j = 1, 2, 3, ...\}$ with different timestamps s_j, we can acquire the embedding H^{s_1}, H^{s_2}, H^{s_3}, etc., from the encoder. Next, a time series decoder needs to be designed to predict the source by using the embedding. To determine the output dimension, we identify the source localization problem as a node binary classification problem, then we use a target function that includes the loss of all nodes to constrain the training on graph-level G. A bidirectional lightweight time series module, i.e., Bi-GRU, is used as the decoder to predict the source. The update function is shown in Eq. (8).

$$
h_t = (1 - z_t) * n_t + z_t * h_{(t-1)}
\tag{8}
$$

where h_t and h_{t-1} are the hidden layer state of the model at the timestamp s_t and the previous timestamp, z_t and n_t are the update and new gates, respectively. We input the embedding of snapshot sequence $\{H^{s_1}, H^{s_2}, H^{s_3}, ...H^{s_\varsigma}\}$ into a Bi-GRU model, and the forward GRU infer the source based on the snapshots from the timestamp s_1 to s_ς, and we denote the output predicted binary classification for all nodes of the hidden layer in the timestamp s_j as $\overrightarrow{h}_j \in \mathbb{R}^{|V| \times 2}$. Similarly, the output of backward GRU is represented as $\overleftarrow{h}_j \in \mathbb{R}^{|V| \times 2}$. Finally, the two hidden layer states of the timestamp s_j are concatenated, i.e., $\hat{R}_j = \left[\overrightarrow{h}_j, \overleftarrow{h}_j\right] \in \mathbb{R}^{|V| \times 4}$.

For any timestamp s_j, each node $\hat{R}_j(v)$ has 4 features output by the Bi-GRU. However, it is difficult to determine exactly at which timestamp the source probability can be evaluated. Because the infection information is imbalanced in different timestamps. That is, the information in the earlier propagation stage is lacking. And the information in the later propagation stage is highly coupled and overlapping. Thus, we design a temporal attention mechanism to dynamically adjust the weight of predicted source probability in different timestamps. A one-timestamp adjacency matrix ϖ is designed to enhance the attention ability for the infection timestamps, which is defined as follows.

$$\varpi = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \qquad (9)$$

And the self-attention focusing on the weight of different timestamps is implemented by a shared attentional mechanism a: $\text{CONCAT}(\mathbb{R}^{|\boldsymbol{W}_A^T \hat{R}(v)|}, \mathbb{R}^{|\boldsymbol{W}_A^T \hat{R}(v)|}) \to \mathbb{R}$ to compute the attention coefficients e_{ij} corresponding to the timestamp s_i and timestamp s_j of a node v, which indicates the importance of timestamp j's features to timestamp i.

$$e_{ij}(v) = a(\boldsymbol{W}_A^T \hat{R}_i(v), \boldsymbol{W}_A^T \hat{R}_j(v)) \quad s.t. \quad \text{abs}(i-j) \le 1 \tag{10}$$

where $\boldsymbol{W}_A \in \mathbb{R}^{4 \times 2}$, here the input feature 4 is the double-times binary classification. And we apply a single-layer BP neural network for the attention mechanism a, which is a weight vector with $\vec{a} \in \mathbb{R}^{2|\boldsymbol{W}_A^T \hat{R}(v)|}$. Then to make coefficients easily comparable across different timestamps, we normalize the attentional mechanism.

$$\phi_{ij}(v) = \text{softmax}_j\left(e_{ij}(v)\right) = \frac{\exp\left(e_{ij}(v)\right)}{\sum_{k \in \varpi_{ik}} \exp\left(e_{ik}(v)\right)}$$

$$= \frac{\exp\left(\text{ReLU}\left(\vec{\mathbf{a}}^T \left[\boldsymbol{W}_A^T \hat{R}_i(v) \| \boldsymbol{W}_A^T \hat{R}_j(v)\right]\right)\right)}{\sum_{k \in \varpi_{ik}} \exp\left(\text{ReLU}\left(\vec{\mathbf{a}}^T \left[\boldsymbol{W}_A^T \hat{R}_i(v) \| \boldsymbol{W}_A^T \hat{R}_j(v)\right]\right)\right)} \tag{11}$$

The normalized attention coefficients ϕ_{ij} are used to compute a linear combination of the four features in one timestamp corresponding to them, to serve as the final source prediction for a node v in timestamp s_i with $\mathcal{K}$-head attention, which is shown in Eq. (12).

$$\hat{R}'_i(v) = \text{softmax}\left(\frac{1}{\mathcal{K}} \sum_1^{\mathcal{K}} \sigma\left(\sum_{j \in \varpi_{ij}} \phi_{ij} \boldsymbol{W}_A^T \hat{R}_i(v)\right)\right) \tag{12}$$

After the final prediction $\hat{R}'_i(v)$ for each node v in each timestamp s_i finishes, we pick the needed number (denoted as Z) of predicted nodes by ranking the highest probability as the predicted source, such a source set R^* is shown in Eq. (13).

$$R^* = \{\underset{v \in V}{\text{argmax}}\left(\hat{R}'_i(v)\right) \mid v \notin R^*, |R^*| = Z\} \tag{13}$$

3.3 Loss Function

Further, we design a unique loss function for TGASI to better adapt to the task of source localization.

$$\mathcal{L} = \mathcal{L}_{Entropy}(R_{\text{hot}}, R^*_{\text{hot}}) + \mathcal{L}_{MSE}(R_{\text{hot}}, H_1^{s_i}) + \mathcal{L}_G(R_{\text{hot}}, R^*_{\text{hot}}) \tag{14}$$

Among them, $\mathcal{L}_{Entropy}$ using binary cross entropy is the main loss to train the complete process of TGASI. $\mathcal{L}_{MSE}$ using mean square error is an encoder loss to constrain the learning process of source probability prediction for a single timestamp s_i. And $\mathcal{L}_G$ is a designed auxiliary loss to enhance the learning ability of the TGASI on the localization task.

$$\mathcal{L}_G(R_{\text{hot}}, R^*_{\text{hot}}) = \frac{\sum R_{\text{hot}}}{|V|} \sum_{\substack{v_i \in R_{\text{hot}}(v_i)=0}} (R^*_{\text{hot}}[v_i, 0])$$

$$+ (1 - \frac{\sum R_{\text{hot}}}{|V|}) \sum_{\substack{v_i \in R_{\text{hot}}(v_i)=1}} (R^*_{\text{hot}}[v_i, 1]) \tag{15}$$

4 Experiments

4.1 Datasets and Baselines

Six social networks are selected to evaluate the performance of all localization methods [1].

	Network	$|V|$	$|E|$	$\langle k \rangle$
G_1	Karate	34	78	4.6
G_2	Jazz	198	2742	27.7
G_3	Facebook	4039	88234	43.69
G_4	Twitch-ES	4648	59382	25.55
G_5	Wiki-Vote	7115	103689	29.15
G_6	Page-Large	22470	171002	15.22

Table 1: Information of the real-world datasets.

To demonstrate the validity and novelty of the current work, we compare the TGASI method with the SOTA localization methods, including IVGD [Wang *et al.*, 2022a], SL_VAE [Ling *et al.*, 2022], GCSSI [Dong *et al.*, 2022], SIGN [Li *et al.*, 2021], MCGNN [Sun *et al.*, 2021], Res-GCN [Shah *et al.*, 2020], GCNSI [Dong *et al.*, 2019].

4.2 Evaluation Metrics

By comprehensively considering the evaluation metrics of the baseline methods, we use two metrics, i.e., the standard F1-score [Sokolova *et al.*, 2006] (F1) and average error distance (AED) [Dong *et al.*, 2022].

$$F1\text{-}score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{16}$$

$$\Delta_{\text{AED}} = \min_{r^* \in \text{ permutation } (R^*)} \sum_{i=1}^{Q} \frac{d\left(r_i^*, r_i\right)}{Q} \tag{17}$$

4.3 Settings and Optimizations

By following existing methods in the field of information propagation [Xia *et al.*, 2021] and rumor source localization [Wang *et al.*, 2022a; Ling *et al.*, 2022], we use the independent cascade (IC) model as the underlying propagation model. Also in reference to their study, we pick 10% nodes as the ground-truth sources and simulate the IC propagation process based on G_1 to G_6 networks. And we set the low infection rates in the experiments due to the propagation characteristics of rumors in social media [Kou and Gray, 2017; Wang *et al.*, 2022b]. We independently generate 1000 sets

---

[1] The six datasets are available at: http://snap.stanford.edu

Network	G_1		G_2		G_3		G_4		G_5		G_6	
Algorithm	F1	AED	F1	AED	F1	AED	F1	AED	F1	AED	F1	AED
GCNSI	0.117	1.78	0.05	1.93	0.003	2.21	0.004	2.23	0.001	2.41	0	2.89
GCSSI	0.213	1.58	0.071	1.85	0.007	2.17	0.009	2.18	0.002	2.38	0	3.11
SIGN	0.421	1.09	0.377	1.21	0.065	1.88	0.055	1.91	0.013	2.11	0	2.86
ResGCN	0.410	1.11	0.371	1.17	0.117	1.69	0.130	1.61	0.025	1.89	0.007	2.17
MCGNN	0.312	1.36	0.297	1.42	0.110	1.87	0.123	1.88	0.031	1.98	0	2.90
IVGD	**0.537**	0.91	0.517	0.93	0.371	1.17	0.375	1.19	0.257	1.44	0.103	2.04
SL_VAE	0.377	1.18	0.353	1.22	0.291	1.39	0.277	1.43	0.196	1.73	0.012	2.35
TGASI	0.514	**0.89**	**0.596**	**0.72**	**0.727**	**0.58**	**0.672**	**0.64**	**0.661**	**0.62**	**0.493**	**1.49**

Table 2: Source identification performance on the test dataset of six social networks. The bold values represent the best results.

of propagation. Each set includes several available snapshots with different timestamps. And we use a 10-fold cross-validation strategy to divide the training dataset and the test dataset, then the final result is output by taking the average prediction in the test dataset from each fold. Moreover, an early stopping mechanism is designed in order to avoid overfitting in the training process.

4.4 Overall Experimental Results

The source detection performance based on the F1-score and AED metrics in G_1 to G_6 is shown in Tab. 2. The higher F1-score and the lower AED reveal better performances. In general, the models considering the propagation process generally have a better prediction performance than other methods, i.e., IVGD, and SL_VAE. And the most challenging baseline is IVGD. The main reasons are that GCNSI, SIGN, and ResGCN only design different training models from the perspective of neural network structure, GCCSI only considers the time series features, and MCGCN only solves the single source detection rather than the MRSD task. Although IVGD considers the individual's comprehensive interaction based on the influence, it does not consider the impact of time series characteristics in the sequence data on source detection.

Compared with the optimal baseline IVGD, TGASI is comparable in G_1, and TGASI based on the F1-score metric demonstrates an improvement of approximately 15% in G_2, 54% in G_3, 88% in G_4, etc. Similarly, the comparison superiority of TGASI on the AED metric also increases with the network scale. In conclusion, TGASI outperforms all the SOTA methods on the basis of rigorous metrics in all datasets. There are three reasons: (1) The influence matrix is embedded through the graph residual network to imply the user interaction at each timestamp. (2) The dynamic infection features and static topology features are designed to fit the source localization task. (3) The attention mechanism distinguishes the weight and importance of decoder information at each different timestamp for source localization.

4.5 Comparison of Different Loss Functions

We specially design the loss function in the sequence-to-sequence framework to fit the source localization task. In order to prove the necessity of each component of the loss function $\mathcal{L}$ we designed, we rigorously demonstrate the detection performance of TGASI by removing or replacing each item

in the loss function. Some localization studies only use MSE as the loss function, so we first use the MSE loss to replace the main loss item $\mathcal{L}_{Entropy}$, which is denoted as $\mathcal{L}_{Entropy \to MSE}$ ($\mathcal{L}_{E \to M}$). It can be seen from Tab. 3 that the performance of entropy loss on the TGASI model is better than that of MSE. Because we define the source localization task as a node binary classification task based on the graph-level constraint. What's more, the other two loss items, i.e., $\mathcal{L}_{MSE}$ and $\mathcal{L}_G$ in $\mathcal{L}$, also play a positive role in the learning process of the TGASI model. So we further keep the main loss item $\mathcal{L}_{Entropy}$ unchanged, and remove the loss item $\mathcal{L}_{MSE}$ or loss item $\mathcal{L}_G$, respectively. As can be seen from the last two rows of Tab. 3, we can conclude that both the loss item $\mathcal{L}_{MSE}$ and the loss item $\mathcal{L}_{MSE}$ positively guide the source detection of TGASI.

4.6 Ablation Study

We further study the influence of designed components of TGASI on the source detection performance to prove their contributions. The critical modules of the TGASI model include the coarse-grained feature based on the estimated influence matrix $\mathcal{W}$, the dynamic infection features H_2 and H_3, the topology features of a social network H_G, and the self-attention mechanism on different timestamps. So four variant models of TGASI are developed as follows.

- TGASI_W− uses a zero vector $[0]^{|v| \times 1}$ to replace the coarse-grained feature H_1 shown in Eq. (3).

- TGASI_D− uses two zero vectors $[0]^{|v| \times 1}$ to replace the dynamic features H_2 shown in Eq. (4) and H_3 shown in Eq. (5), respectively.

- TGASI_S− uses a zero matrix $[0]^{|v| \times \sqrt{|V|}}$ to replace the static topology features H_G shown in Eq. (6).

Network	G_2		G_3		G_4	
Loss	F1	Epoch	F1	Epoch	F1	Epoch
$\mathcal{L}$	**0.5959**	5	**0.7266**	6	**0.6716**	6
$\mathcal{L}_{E \to M}$	0.5730	5	0.7013	6	0.6411	6
-$\mathcal{L}_{MSE}$	0.5811	7	0.7132	6	0.6607	8
-$\mathcal{L}_G$	0.5621	11	0.6933	10	0.6373	11

Table 3: The performance evaluation of variant loss in G_2 to G_4.
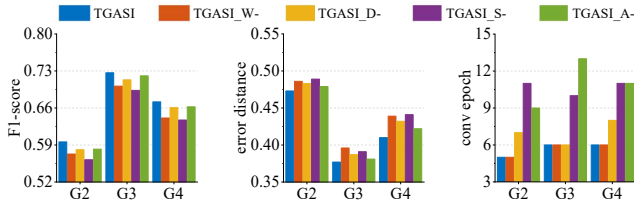
Figure 3: The performance evaluation of variant models from TGASI in G_2 to G_4.

- TGASI_A$-$ replaces the attention mechanism in Eqs. (9)-(12) with $\text{softmax}(\frac{1}{2}(\overrightarrow{h}_j + \overleftarrow{h}_j))$.

We set the parameters of four variant models as the same as those in Sec. 4.3 to guarantee the rigor of the experiments. Due to the limited space, we only present the experiment results in G_2-G_4. As can be seen from Fig. 3 that it will lead to a performance decrease or a delay in the convergence speed of model training no matter removing any critical modules.

4.7 Scalability

The scalability of the model can evaluate whether the model can be applied in a variety of scenarios, which can reflect the strength of its transferability and robustness. The decoder we designed mainly includes Bi-GRU and attention mechanism, which can be used in the inductive learning task. We demonstrate the inductive learning capabilities of TGASI from two perspectives, including social networks and propagation models. More specifically, when we have trained the TGASI model using the snapshot set $\{\mathcal{G}_{s_j} = (V, E, Y_{s_j}) \mid j = 1, 2, 3, ...\}$ based on a social network $G_1=(V,E)$, we further apply the trained TGASI to the other snapshot set $\{\mathcal{G'}_{s_{j'}} = (V', E', Y_{s_{j'}}) \mid j' = 1, 2, 3, ...\}$ based on the other social network $G_2=(V',E')$, where $V \neq V'$ and $E \neq E'$, and $Y_{s_{j'}}$ and Y_{s_j} are generated by different propagation models. Due to the better performance in G_3, we use the trained TGASI model in G_3 to conduct inductive learning.

In the inductive learning of social networks, we replace the original test set of G_3 with another batch of snapshot sets generated by G_1, G_2, G_4, G_5, and G_6. One issue that requires improvement in the future is the variation in the dimensionality of H_G caused by network scale, which only is addressed by using the original topological embedding here. In order to rigorously evaluate TGASI's performance in inductive learning, we also compare the inductive results with the native results of SOTA methods and TGASI on native datasets or propagation models. From figs. 4(a)-(b), the inductive learning ability of TGASI is similar to or better than the original detection performance of the optimal SOTA method. What's more, it can be seen from Fig. 4(c) that the inductive performance in G_4 or G_5 is stronger than that in G_1 or G_6. We conclude that the decreased extent of inductive performance is proportional to the deviation from the network scale corresponding to the trained baseline model.

In the inductive learning of propagation models, we use various propagation models, including the heterogeneous SI with infection rate $I \sim U(0.05, 0.15)$, the heterogeneous SIR with recovery rate $R \sim U(0, 0.05)$, the homogeneous SI with
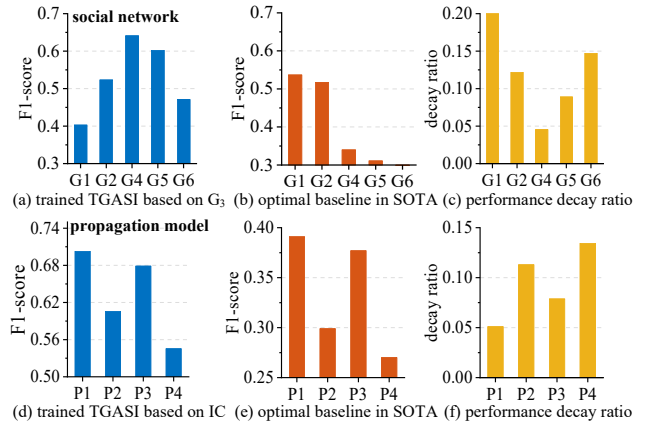

Figure 4: The performance evaluation of TGASI on the inductive learning task. (a)-(c) is the inductive learning on different social networks based on the IC model, and (d)-(f) is on different propagation models in G_3. $P1$, $P2$, $P3$, and $P4$ are denoted as homogeneous SI, homogeneous SIR, heterogeneous SI, and heterogeneous SIR.

$I = 0.1$, and the homogeneous SIR with $R = 0.02$, to evaluate the source detection performance of TGASI in different propagation models. Figs. 4(d)-(e) show that the inductive learning ability of TGASI based on the propagation model is overall higher than 0.5 on the F-score metric, and is better than the original performance of all SOTA methods. Moreover, the performance of TGASI on the homogeneous model is better than that of the heterogeneous model because it is easier to accurately learn the influence information $\mathcal{W}$ in the homogeneous models. And the performance of TGASI on the SI model is better than that of the SIR model because the interactive behavior and the temporal characteristic in the SIR model are more complicated.

5 Conclusion

In this paper, we pay attention to the source localization problem by considering the heterogeneous behavior diversity in time-varying infection scenarios, and we design a sequence-to-sequence framework TGASI with the transferability to share trained TGASI to different propagation models and social networks. Comprehensive experiments demonstrate the necessity of designed modules, i.e., a coarse-grained-based source probability feature, dynamic infection features, static topology features, a temporal attention mechanism, and a unique loss function for the localization task in TGASI. It's worth mentioning that TGASI is the first source localization framework that considers scalability based inductive learning. So a trained TGASI model is suitable for other social networks and propagation models. And model transferability experiments demonstrate that the inductive detection performance of TGASI is overall superior to the original detection performance of SOTA methods. However, we also analyze that the source detection performance of TGASI could deteriorate when it transfers to some scenarios. So we are motivated to propose a novel localization framework in the future that can efficiently handle scalability under various scenarios.

Acknowledgments

References

[Agarwal *et al.*, 2021] Nidhi Agarwal, Akanksha Sondhi, Khyati Chopra, and Ghanapriya Singh. Transfer learning: Survey and classification. *Smart innovations in communication and computational sciences*, pages 145–155, 2021.

[Baesens *et al.*, 2015] Bart Baesens, Veronique Van Vlasselaer, and Wouter Verbeke. *Fraud analytics using descriptive, predictive, and social network techniques: a guide to data science for fraud detection*. John Wiley & Sons, 2015.

[Chang *et al.*, 2018] Bo Chang, Lili Meng, Eldad Haber, Lars Ruthotto, David Begert, and Elliot Holtham. Reversible architectures for arbitrarily deep residual neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[Chung *et al.*, 2014] Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning*, 12 2014.

[Diepold *et al.*, 2022] Klaus Diepold, Sven Gronauer, Matthias Kissel, and Mohamed Ali Tnani. Data science for the environment. *Trend Reports: Seminar on Machine Intelligence*, 2022.

[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.

[Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jurgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[Jacobsen *et al.*, 2018] Jorn-Henrik Jacobsen, Arnold Smeulders, and Edouard Oyallon. i-revnet: Deep invertible networks. *arXiv preprint arXiv:1802.07088*, 2018.

[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.

[Jin *et al.*, 2017] Zhiwei Jin, Juan Cao, Han Guo, Yongdong Zhang, and Jiebo Luo. Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 795–816, 2017.

[Khoo *et al.*, 2020] Ling Min Serena Khoo, Hai Leong Chieu, Zhong Qian, and Jing Jiang. Interpretable rumor detection in microblogs by attending to user interactions. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 8783–8790, 2020.

[Kou and Gray, 2017] Yubo Kou and Colin M Gray. Supporting distributed critique through interpretation and sense-making in an online creative community. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–18, 2017.

[Lao *et al.*, 2021] An Lao, Chongyang Shi, and Yayi Yang. Rumor detection with field of linear and non-linear propagation. In *Proceedings of the Web Conference*, pages 3178–3187, 2021.

[Li *et al.*, 2021] Liang Li, Jianye Zhou, Yuewen Jiang, and Biqing Huang. Propagation source identification of infectious diseases with graph convolutional networks. *Journal of biomedical informatics*, 116:103720, 2021.

[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.

[Michalski, 1983] Ryszard S Michalski. A theory and methodology of inductive learning. In *Machine learning*, pages 83–134. Elsevier, 1983.

[Paluch *et al.*, 2020] 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.

[Qian *et al.*, 2018] Feng Qian, Chengyue Gong, Karishma Sharma, and Yan Liu. Neural user response generator: Fake news detection with collective user intelligence. In *Proceedings of the International Joint Conference on Artificial Intelligence*, volume 18, pages 3834–3840, 2018.

[Salamat *et al.*, 2020] Amirreza Salamat, Xiao Luo, and Ali Jafari. Balnode2vec: Balanced random walk based versatile feature learning for networks. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2020.

[Shah *et al.*, 2020] Chintan Shah, Nima Dehmamy, Nicola Perra, and Matteo Chinazzi. Finding patient zero: Learning contagion source with graph neural networks. *arXiv preprint arXiv:2006.11913*, 2020.

[Shelke and Attar, 2019] Sushila Shelke and Vahida Attar. Source detection of rumor in social network–a review. *Online Social Networks and Media*, 9:30–42, 2019.

[Sokolova *et al.*, 2006] Marina Sokolova, Nathalie Japkowicz, and Stan Szpakowicz. Beyond accuracy, f-score and roc: A family of discriminant measures for performance evaluation. In *Australasian Joint Conference on Artificial Intelligence*, pages 1015–1021, 2006.

[Song *et al.*, 2021] Chenguang Song, Nianwen Ning, Yunlei Zhang, and Bin Wu. A multimodal fake news detection model based on crossmodal attention residual and multi-channel convolutional neural networks. *Information Processing & Management*, 58(1):102437, 2021.

[Sun *et al.*, 2021] Kuangchi Sun, Zhenfeng Huang, Hanling Mao, Aisong Qin, Xinxin Li, Weili Tang, and Jianbin Xiong. Multi-scale cluster-graph convolution network with multi-channel residual network for intelligent fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 71:1–12, 2021.

[Sutskever *et al.*, 2014] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27, 2014.

[Velickovic *et al.*, 2018] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *International Conference on Learning Representations*, 2018.

[Ventola, 2014] C Lee Ventola. Social media and health care professionals: benefits, risks, and best practices. *Pharmacy and therapeutics*, 39(7):491, 2014.

[Wang *et al.*, 2017] Zheng Wang, Chaokun Wang, Jisheng Pei, and Xiaojun Ye. Multiple source detection without knowing the underlying propagation model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 217–223, San Francisco, CA, 2017. PALO ALTO, CA 94303 USA.

[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.

[Welling and Kipf, 2016] Max Welling and Thomas N Kipf. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR 2017)*, 2016.

[Xia *et al.*, 2021] Wenwen Xia, Yuchen Li, Jun Wu, and Shenghong Li. Deepis: Susceptibility estimation on social networks. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 761–769, 2021.

[Yang *et al.*, 2021] Xiaoyu Yang, Yuefei Lyu, Tian Tian, Yifei Liu, Yudong Liu, and Xi Zhang. Rumor detection on social media with graph structured adversarial learning. In *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence*, pages 1417–1423, 2021.

[Zhang *et al.*, 2019] Zhengchao Zhang, Meng Li, Xi Lin, Yinhai Wang, and Fang He. Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies. *Transportation research part C: emerging technologies*, 105:297–322, 2019.

[Zhu *et al.*, 2022] Yuecai Zhu, Fuyuan Lyu, Chengming Hu, Xi Chen, and Xue Liu. Learnable encoder-decoder architecture for dynamic graph: A survey. *arXiv preprint arXiv:2203.10480*, 2022.