Source Localization of Graph Diffusion

Derek Wang

Sep 29th, 2022

ILLINOIS

Outline

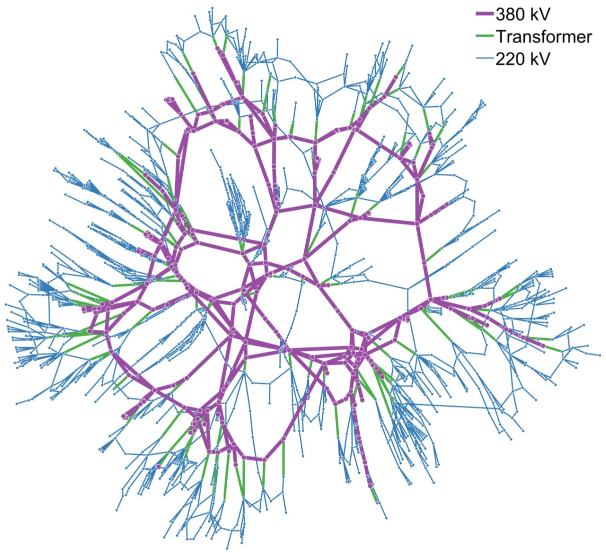
- Background
- Source Localization of Graph Diffusion:
  - Multiple Source Detection without Knowing the Underlying Propagation Model (AAAI 2017)
  - Multiple Rumor Source Detection with Graph Convolutional Networks (CIKM 2019)
  - Source Localization of Graph Diffusion via Variational Autoencoders for Graph Inverse Problems (KDD 2022)
- Summary

Outline

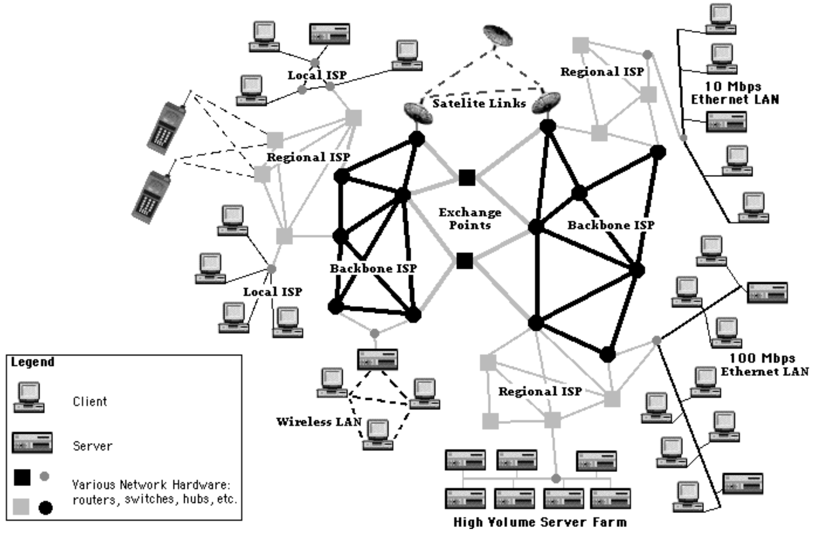
- Background
- Source Localization of Graph Diffusion:
  - Multiple Source Detection without Knowing the Underlying Propagation Model (AAAI 2017)
  - Multiple Rumor Source Detection with Graph Convolutional Networks (CIKM 2019)
  - Source Localization of Graph Diffusion via Variational Autoencoders for Graph Inverse Problems (KDD 2022)
- Summary

Diffusion Over Graphs

Social Networks



Power Networks



Internet

What to propagate?

Rumors

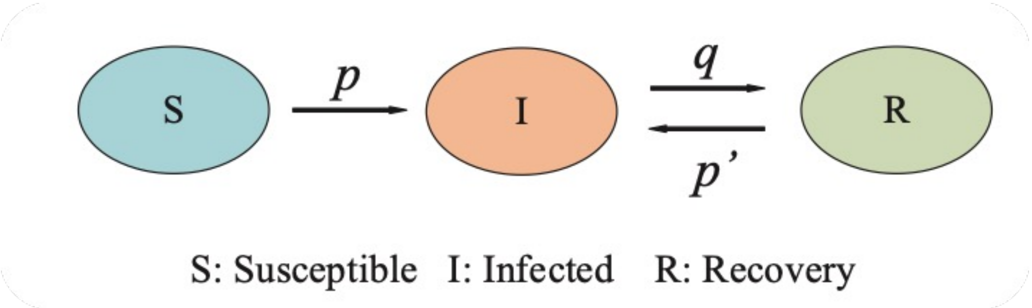
Disease

Power Failures

Computer Virus

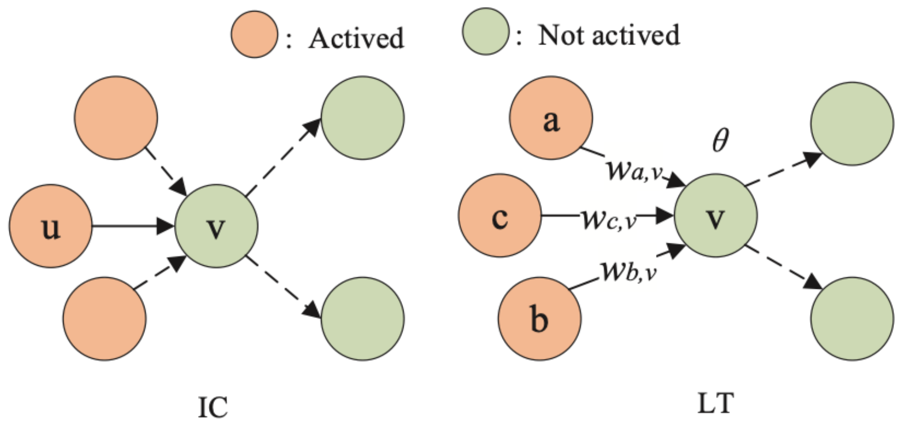
Propagation Models

- **Infection** Models:
  - Susceptible-Infected (SI): p to get infected. Infected nodes remain infected forever.
  - Susceptible-Infected-Recovered (SIR): p to get infected, and q to recover.
  - Susceptible-Infected-Susceptible (SIS): p to get infected, q to recover and p' to get infected again.



- Anderson, Roy M., and Robert M. May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1992.
- Allen, Linda JS. "Some discrete-time SI, SIR, and SIS epidemic models." *Mathematical biosciences* 124, no. 1 (1994): 83-105.
- Dong, Ming, 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*, pp. 569-578. 2019.

Propagation Models



- **Influence** Models:
  - Independent Cascade (IC)
    - In each iteration, each node has a chance to active its neighbors.
    - If a neighbor v is not activated by the node u successfully, it will never be activated by u again, however, v can still be activated by other neighbors.
  - Linear Threshold (LT)
    - Each node has an activation threshold θ in $[0,1]$. Each edge e is initialized with a weight w_e.
    - If the sum of weights of edges from infected neighbors of node v is greater than the threshold, then node v gets infected.

- Goldenberg, Jacob, Barak Libai, and Eitan Muller. "Talk of the network: A complex systems look at the underlying process of word-of-mouth." *Marketing letters* 12, no. 3 (2001): 211-223.
- Kempe, David, Jon Kleinberg, and Éva Tardos. "Maximizing the spread of influence through a social network." In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 137-146. 2003.
- Dong, Ming, 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*, pp. 569-578. 2019.

Motivations & Solutions

- Why localize diffusion source?
  - Spread of misinformation causes huge damage.
  - It is crucial to detect diffusion source and cut off propagation paths in early stage.

- Solution directions:
  - Infection status based analysis: only infection status at the most recent time step is available.
  - Observer node deployment: partial observations of nodes' status are available along the diffusion path.

- Dong, Ming, 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*, pp. 569-578. 2019.
- Shah, Devavrat, and Tauhid Zaman. "Rumors in a network: Who's the culprit?." *IEEE Transactions on information theory*57, no. 8 (2011): 5163-5181.
- Zhu, Kai, Zhen Chen, and Lei Ying. "Catch'em all: Locating multiple diffusion sources in networks with partial observations." In *Thirty-First AAAI Conference on Artificial Intelligence*. 2017.

Outline

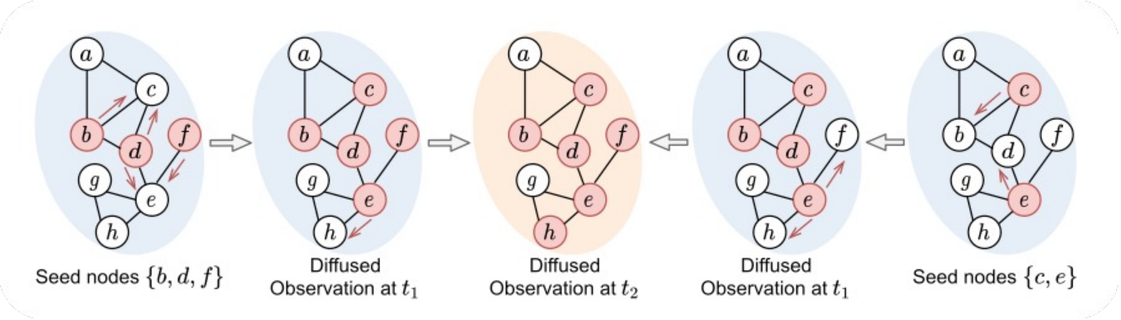
- Background
- Source Localization of Graph Diffusion:
  - Multiple Source Detection without Knowing the Underlying Propagation Model (AAAI 2017)
  - Multiple Rumor Source Detection with Graph Convolutional Networks (CIKM 2019)
  - Source Localization of Graph Diffusion via Variational Autoencoders for Graph Inverse Problems (KDD 2022)
- Summary

Problem Definition

- **Given**: a graph $G(V, E)$ where V is the set of nodes and E is the set of edges. Suppose there exists an indicator infection vector $\boldsymbol{y} \in \{0,1\}^{|V|}$ of all nodes in G, which describes the observed infection status of a subset of nodes in G: 1 for infected, 0 for uninfected.
- **Output**: the predicted diffusion source vector $\widetilde{\boldsymbol{x}} \in \{0,1\}^{|V|}$, such that $||\widetilde{\boldsymbol{x}} - \boldsymbol{x}||_2^2$ is minimized, where $\boldsymbol{x}$ is actual diffusion source vector.

Note: Inverse of graph diffusion.

Challenges



- Intrinsic patterns of diffusion source.
- Uncertainty of graph diffusion.
  - Different diffusion source might lead to same observations.
- Generalization to different propagation model.
  - Diffusion process is affected by various factors: network parameters, immunity power, transmission rate, etc.
- Difficulty to leverage knowledge in propagation models for modeling their inverse processes.

- Ling, Chen, 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*, pp. 1010-1020. 2022.
- Wang, Junxiang, Junji Jiang, and Liang Zhao. "An Invertible Graph Diffusion Neural Network for Source Localization." In *Proceedings of the ACM Web Conference 2022*, pp. 1058-1069. 2022.

Outline

- Background
- Source Localization of Graph Diffusion:
    - Multiple Source Detection without Knowing the Underlying Propagation Model (AAAI 2017)
    - Multiple Rumor Source Detection with Graph Convolutional Networks (CIKM 2019)
    - Source Localization of Graph Diffusion via Variational Autoencoders for Graph Inverse Problems (KDD 2022)
- Summary

Motivation & Assumption

- **Idea**: source prominence.
  - Nodes surrounded by larger proportions of infected nodes are more likely to be source ones.
  - At the margin of an infected region, nodes tend to have less infected neighbors.
  - Aligns with different propagation models: nodes close to source nodes have higher probability get infected.
- **Assumption**: no access to underlying propagation model.
  - Diffusion over graph in real world are usually more complex.
  - It is hard to acquire the real values of parameters of pre-select propagation model.

- Wang, Zheng, 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*, vol. 31, no. 1. 2017.

Label Propagation based Source Identification

- **Step 1**: Assign labels to nodes
  - Label vector $\mathcal{G}^0 = Y$: 1 for infected nodes, and -1 for uninfected nodes.
- **Step 2**: Label propagation on the network
  - $\mathcal{G}_i^{t+1} = \alpha \sum_{j:j\in\mathcal{N}(i)} S_{ij}\mathcal{G}_i^t + (1-\alpha)Y_i$
  - S_{ij} represents the label propagation probability from node j to node i.
  - α is the fraction of label information that node i gets from neighbors.
- **Step 3**: Sources identification
  - $\mathcal{G}^t$ converges to $\mathcal{G}^*$ at the end of propagation process.
  - Select source node if: 1) node i is an infected node initially, i.e., $Y_i = 1$; 2) its final label value $\mathcal{G}_i^*$ is larger than those of its neighbors.

- Wang, Zheng, 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*, vol. 31, no. 1. 2017.

Label Propagation based Source Identification

Algorithm 1 Label Propagation based Source Identification (LPSI)

Input: The infected network G=(V, E), parameter α ;
The initial infection node vector Y.
Output: The source set $\mathcal{S}$.

Form the weight matrix W defined by $W_{ij} = 1$ if there exists an edge connecting nodes i and j;
Construct the matrix $S = D^{-1/2} W D^{-1/2}$, where D is a diagonal matrix with its (i,i)-element equal to the sum of the i-th row of W;
$\mathcal{G}^{t=0} \leftarrow Y$;
while $\mathcal{G}^t$ does not reach the convergence $\mathcal{G}^*$ **do**
  for each node i **do**
    $\mathcal{G}_i^{t+1} = \alpha \sum_{j:j \in \mathcal{N}(i)} S_{ij} \mathcal{G}_j^t + (1-\alpha) Y_i$;
  end for
  $t = t + 1$;
end while
$\mathcal{S} = \{\}$;
for each original infected node i **do**
  if $\mathcal{G}_i^* >$ all i's neighbors' $\mathcal{G}^*$ value **then**
    $\mathcal{S} = \mathcal{S} \cup \{i\}$;
  end if
end for
return $\mathcal{S}$;

Step 1: Assign Labels

Step 2: Label Propagation

Step 3: Source Identification

- Wang, Zheng, 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*, vol. 31, no. 1. 2017.

Convergence Analysis

- Label Propagation: $\mathcal{G}^{t+1} = \alpha S \mathcal{G}^{t} + (1-\alpha)Y$
- With initial condition $\mathcal{G}^{0} = Y$, we have:

$$\mathcal{G}^{t} = (\alpha S)^{t} Y + (1-\alpha)\sum_{i=0}^{t-1} (\alpha S)^{i}\, Y$$

- With parameter $0 < \alpha < 1$ and normalized matrix S will have:
  - $\lim_{t\to\infty} (\alpha S)^{t} = 0$ and $\lim_{t\to\infty} \sum_{i=0}^{t-1} (\alpha S)^{i} = (I - \alpha S)^{-1}$
  - I is the identity matrix.
- Thus, the iteration converges to $\mathcal{G}^{*} = (1-\alpha)(I - \alpha S)^{-1} Y$.

- Zhou, Dengyong, Olivier Bousquet, Thomas Lal, Jason Weston, and Bernhard Schölkopf. "Learning with local and global consistency." *Advances in neural information processing systems* 16 (2003).
- Wang, Fei, and Changshui Zhang. "Label propagation through linear neighborhoods." *IEEE Transactions on Knowledge and Data Engineering* 20, no. 1 (2007): 55-67.

Experiment

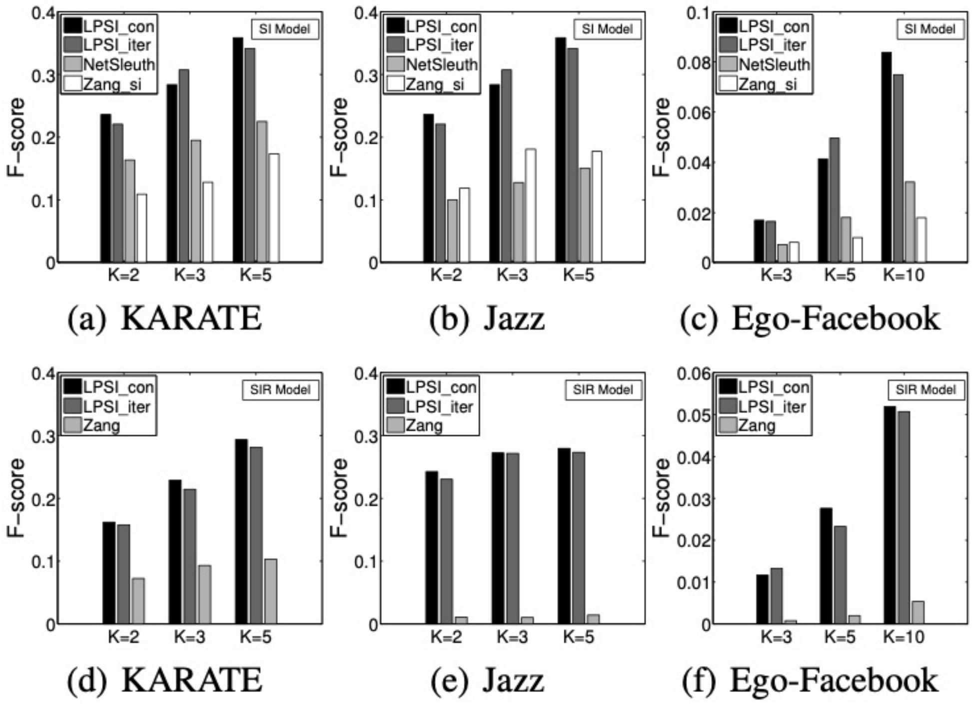


Figure 2: Source detection accuracies under infection models, i.e., SI model (row 1) and SIR model (row 2).

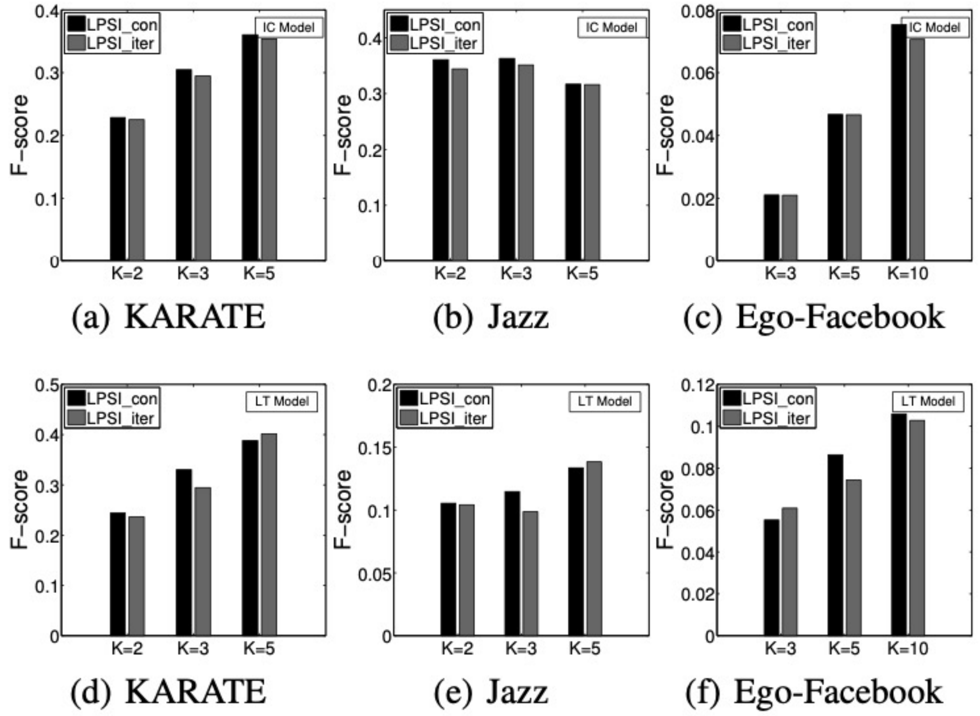


Figure 3: Source detection accuracies under influence models, i.e., IC model (row 1) and the LT model (row 2).

- Wang, Zheng, 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*, vol. 31, no. 1. 2017.

Experiment

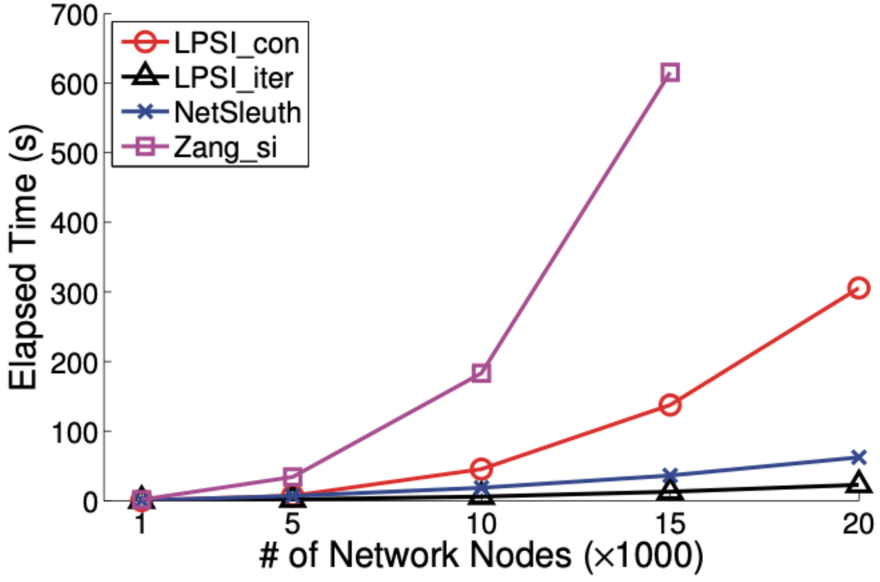


Figure 4: Scalability on synthetic data.

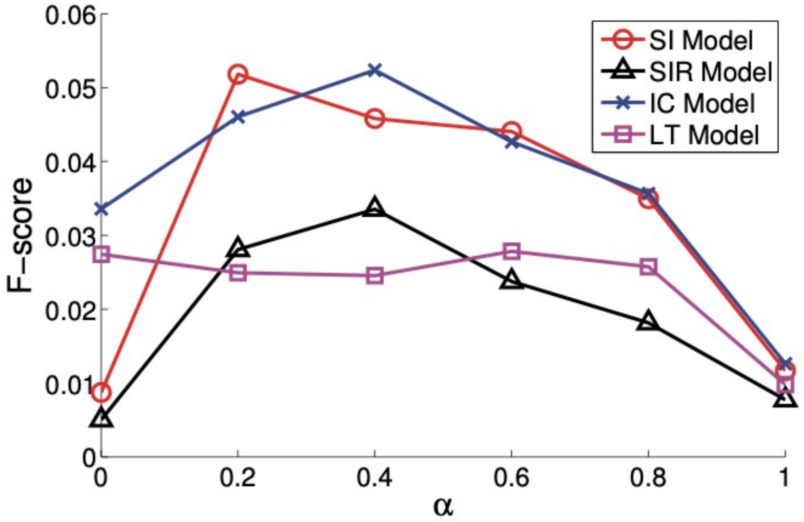


Figure 5: Parameter α in LPSI_con on Ego-Facebook with K=5.

- Wang, Zheng, 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*, vol. 31, no. 1. 2017.

Outline

- Background
- Source Localization of Graph Diffusion:
  - Multiple Source Detection without Knowing the Underlying Propagation Model (AAAI 2017)
  - Multiple Rumor Source Detection with Graph Convolutional Networks (CIKM 2019)
  - Source Localization of Graph Diffusion via Variational Autoencoders for Graph Inverse Problems (KDD 2022)
- Summary

Motivations

- Most existing works assume the underlying propagation model is known.
  - Impractical on real data since it is usually difficult to acquire actual and accurate propagation model.
- The LPSI attempts to solve this, but suffers that node label is simply an integer, which restricts prediction precision.
- First work that applies GCN for source detection problem.

- Dong, Ming, 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*, pp. 569-578. 2019.

Input Generation

- d_1: infection state value, where infected node to be 1 and uninfected to be -1.
- d_2: feature of source prominence, which is the output of LPSI.
- d_3: features for infected nodes.
- d_4: features for uninfected nodes.

Algorithm 1: Input Generation Algorithm

Input: The infected network $\mathcal{G} = (V, E, Y)$ where $Y = (Y_1, ..., Y_{|v|})^T$; parameter α;

Output: Infection state matrix $Y' = [d_1, d_2, d_3, d_4]$

Form the weight matrix W defined by $W_{ij} = 1$ if there exists an edge connecting nodes i and j;

Construct the matrix $S = D^{-1/2}WD^{-1/2}$, where D is a diagonal matrix with its (i,i)-element equal to the sum of the i-th row of W;

Initialize the $V_3 = Y$, $V_4 = Y$;

for $i < len(Y)$ **do**

  if Y_i == *-1* **then**

    $V_{3i} = 0$;

  else

    $V_{4i} = 0$;

  end

end

$d_1 = Y$;

$d_2 = (1 - \alpha)(I - \alpha S)^{-1} Y$;

$d_3 = (1 - \alpha)(I - \alpha S)^{-1} V_3$;

$d_4 = (1 - \alpha)(I - \alpha S)^{-1} V_4$;

$Y' = concatenate(d_1, d_2, d_3, d_4)$

- Dong, Ming, 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*, pp. 569-578. 2019.

Modified GCN and Loss Function

- Original GCN: semi-supervised learning.
- Multiple rumor source detection: supervised learning task.

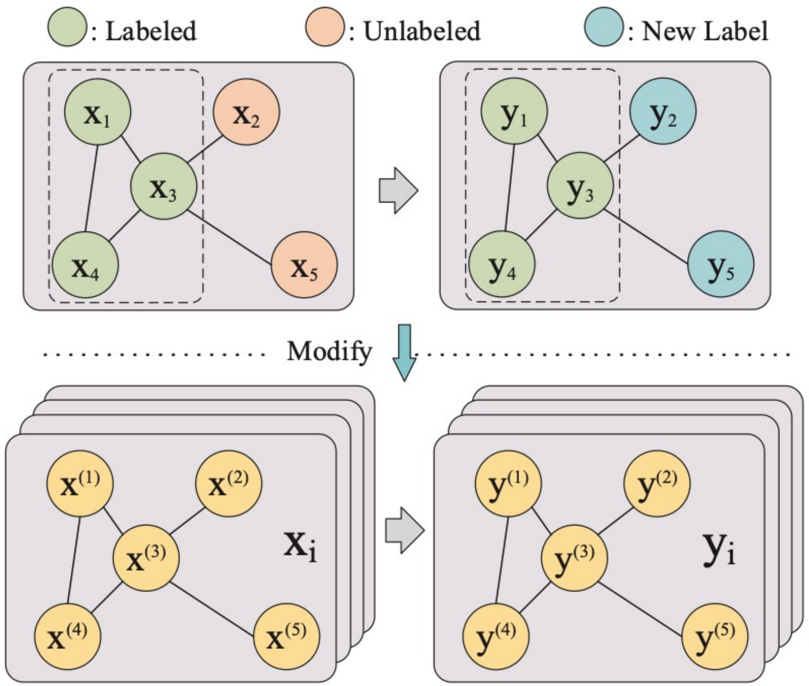


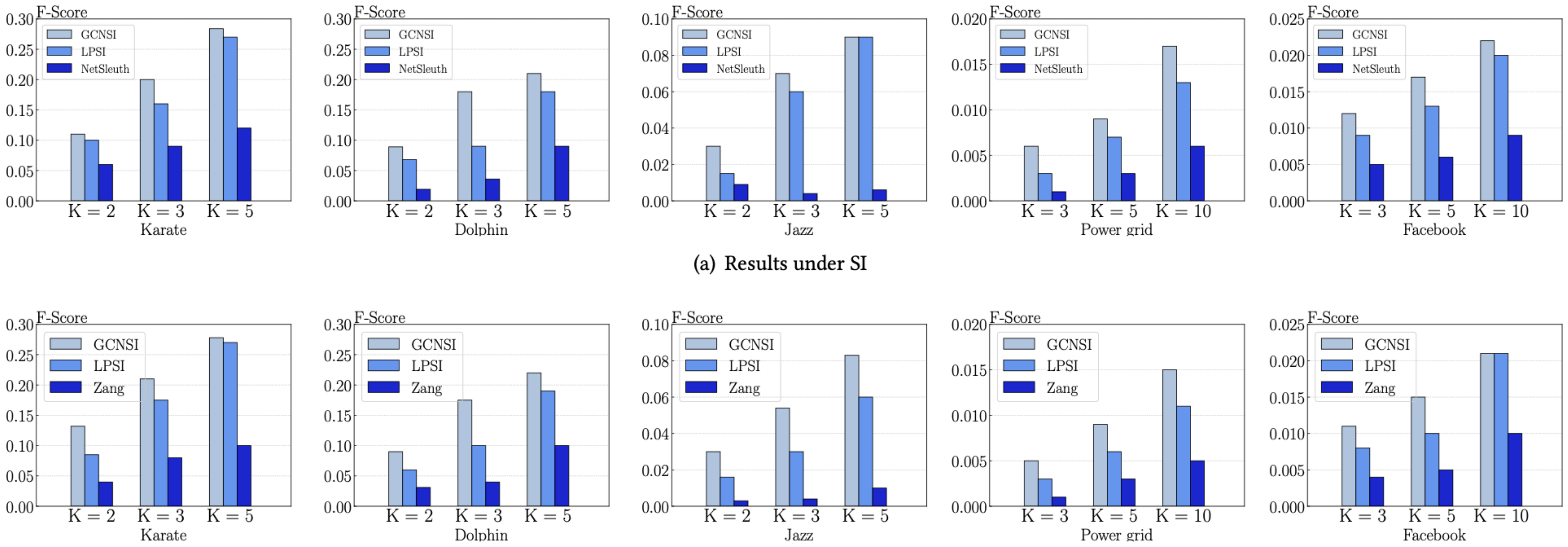
Figure 4: Traditional and modified GCN layer

- Loss Function:

$$L(y, y') = -\log \sigma(y') \times y - \log\big(1 - \sigma(y')\big) \times (1 - y) + \lambda ||w||_2$$

- Kipf, Thomas N., and Max Welling. "Semi-supervised classification with graph convolutional networks." *arXiv preprint arXiv:1609.02907* (2016).
- Dong, Ming, 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*, pp. 569-578. 2019.

Experiment



(a) Results under SI

(b) Results under SIR

- Dong, Ming, 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*, pp. 569-578. 2019.

Experiment

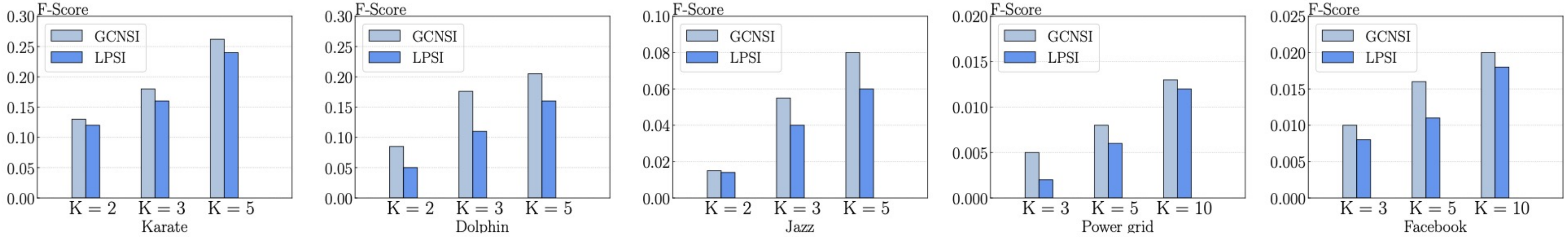


Figure 6: Results under Influence Model IC influence model

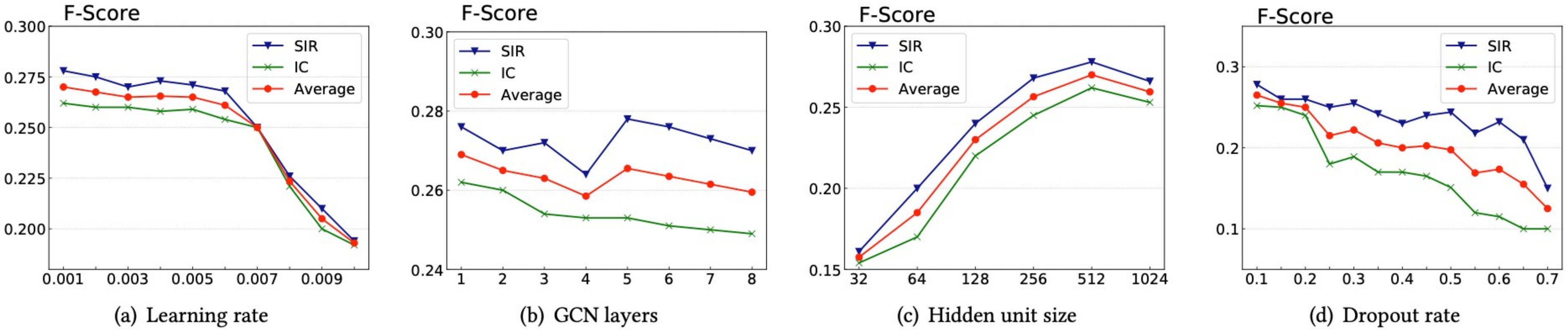


Figure 7: Impact of parameters Karate Dataset

- Dong, Ming, 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*, pp. 569-578. 2019.

Outline

- Background
- Source Localization of Graph Diffusion:
  - Multiple Source Detection without Knowing the Underlying Propagation Model (AAAI 2017)
  - Multiple Rumor Source Detection with Graph Convolutional Networks (CIKM 2019)
  - Source Localization of Graph Diffusion via Variational Autoencoders for Graph Inverse Problems (KDD 2022)
- Summary

Motivation

- Challenges of source localization of graph diffusion:
  - The difficulty of quantifying the **uncertainty** in source localization of graph diffusion.
  - The difficulty of characterizing the **intrinsic patterns** of the diffusion sources.
  - The difficulty of imposing the **generalization** under any underlying diffusion patterns.
- **Uncertainty**: deep generative model.
- **Intrinsic patterns**: learn from the observed source-observation pairs to encode the prior knowledge of the diffusion sources.
- **Generalization**: derive a unified objective function that makes sure the source localization is fully aware of any diffusion patterns.

- Ling, Chen, 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*, pp. 1010-1020. 2022.

Source Localization Variational Autoenocder

- Evidence Lower Bound (ELBO) of VAE:

$$ELBO = \mathbb{E}[\log p_\theta(x|z)] - \mathbb{KL}[q_\varphi(z|x)||p(z)] \leq \log p\,(x)$$

- ELBO for SL-VAE:

$$ELBO = \mathbb{E}\Big[\underbrace{\boxed{\log p_\psi(y|x,G)}}_{\text{forward loss}} + \underbrace{\boxed{\log p_\theta(x|z)}}_{\text{reconstruction loss}}\Big]$$
$$-\mathbb{KL}[q_\varphi(z|x)||p(z)] - \underbrace{\lambda||\max\left(0, y^{(j)} - y^{(i)}\right)||_2^2}_{\text{constraint on information diffusion}}, \forall x^{(i)} \supseteq x^{(j)}$$

- Monotonicity constraint on information diffusion: if one diffusion source set $x^{(i)}$ is the superset of another $x^{(j)}$, then the probability of each node being infected in $y^{(i)}$(estimated from $x^{(i)}$) should be $\geq y^{(j)}$.

- Ling, Chen, 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*, pp. 1010-1020. 2022.

SL-VAE Inference Framework

- Optimization problem for finding optimal diffusion source:

$$\max_x [p_\psi(y|x,G) \cdot \sum_z p_\theta(x|z) \cdot p(z)]$$

requires sampling

- Initialization of x ($\mathcal{L}_{init}$):

$$\min_x [-\log p_\psi(y|x,G) - \log p_\theta(x|\bar{z})]$$

prior knowledge

$\bar{z} = \frac{1}{N} q_\varphi(z|\hat{x})$ is the mean of all latent variables obtained from N diffusion sources $\hat{x}$ in the training set.

Algorithm 1: SL-VAE Inference Framework

Require: $p_\psi(y|x,G)$; $p_\theta(x|z)$; $q_\phi(z|x)$; $\mathcal{L}_{pred}$; $\mathcal{L}_{init}$; Numbers of iterations: $n_{init} < n_{opt}$; Success probability τ; Threshold δ; Learning rate α

```
1:  x ~ Bin(|V|, τ) // Sampling an initial x̃_init.
    // Optimize x with L_init for initialization.
2:  for n = 0, ..., n_init do
3:      x ← x − α · ∇L_init
4:      x ← trim(x, 0, 1)
5:      x ← threshold(x, δ)
6:  end for
    // Optimize x with L_pred(x) for final prediction.
7:  for n = 0, ..., n_opt do
8:      x ← x − α · ∇L_pred(x)
9:      x ← trim(x, 0, 1)
10:     x ← threshold(x, δ)
11: end for
12: x̃ ← x // Output the predicted diffusion source.
```


- Ling, Chen, 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*, pp. 1010-1020. 2022.

SL-VAE Inference Framework

- Objective for Prediction ($\mathcal{L}_{pred}$):

$$\min_x [-\log p_\psi(y|x,G) - \log[\sum_z \sum_{\hat{x}} p_\theta(x|z) \cdot q_\varphi(z|\hat{x})]]$$

prior knowledge

$$= \min_x [||y - p_\psi(y|x,G)||_2^2 - \log[\sum_j^N \prod_i^{|V|} f_\theta(z_i^j)^{x_i} (1 - f_\theta(z_i^j))^{1-x_i}]$$

- $\mathcal{L}_{init}$ help reduce the search space.

Algorithm 1: SL-VAE Inference Framework

Require: $p_\psi(y|x,G)$; $p_\theta(x|z)$; $q_\phi(z|x)$; $\mathcal{L}_{pred}$; $\mathcal{L}_{init}$; Numbers of iterations: $n_{init} < n_{opt}$; Success probability τ; Threshold δ; Learning rate α

```
x ~ Bin(|V|, τ) // Sampling an initial x̃_init.
// Optimize x with L_init for initialization.
for n = 0, ..., n_init do
    x ← x − α · ∇L_init
    x ← trim(x, 0, 1)
    x ← threshold(x, δ)
end for
// Optimize x with L_pred(x) for final prediction.
for n = 0, ..., n_opt do
    x ← x − α · ∇L_pred(x)
    x ← trim(x, 0, 1)
    x ← threshold(x, δ)
end for
x̃ ← x // Output the predicted diffusion source.
```


- Ling, Chen, 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*, pp. 1010-1020. 2022.

Experiment

	Jazz				Cora-ML				Power Grid				Karate				Network Science			
Methods	RE	PR	F1	AUC	RE	PR	F1	AUC	RE	PR	F1	AUC	RE	PR	F1	AUC	RE	PR	F1	AUC
LPSI	0.4789	0.1054	0.1716	0.4841	0.5954	0.1556	0.2466	0.6675	0.4953	0.4546	0.4737	0.9337	0.4667	0.1861	0.2855	0.6344	0.6044	0.4231	0.4978	0.8378
GCNSI	0.4368	0.1589	0.2329	0.6428	0.3619	0.1182	0.1780	0.5383	0.3477	0.1413	0.2099	0.5044	0.4333	0.1999	0.2613	0.6022	0.2247	0.1375	0.1706	0.4759
OJC	0.1798	0.1005	0.1289	0.5045	0.2239	0.2036	0.2133	0.5633	0.2871	0.1044	0.1531	0.5011	0.3611	0.2708	0.3095	0.6335	0.1233	0.3708	0.1851	0.5331
Netsleuth	0.1315	0.1087	0.1191	0.5432	0.2647	0.2647	0.2647	0.4688	0.5972	0.4975	0.5428	0.7651	0.3333	0.3333	0.3333	0.4355	0.3948	0.3283	0.3585	0.6528
SL-VAE	**0.9474**	**0.7193**	**0.8182**	**0.9777**	**0.9466**	**0.6717**	**0.7858**	**0.9582**	**0.9636**	**0.6648**	**0.7868**	**0.9636**	**0.6667**	**0.6667**	**0.6667**	**0.8172**	**0.9937**	**0.6738**	**0.8031**	**0.9705**

Table 1: Performance over comparison methods under SI diffusion pattern. (Best is highlighted with bold.)

	Jazz				Cora-ML				Power Grid				Karate				Network Science			
Methods	PR	RE	F1	AUC	PR	RE	F1	AUC	PR	RE	F1	AUC	PR	RE	F1	AUC	PR	RE	F1	AUC
LPSI	0.1153	0.3632	0.1698	0.5005	0.1072	0.4779	0.1752	0.4986	0.4865	0.4721	0.4784	0.5821	0.1284	0.4167	0.1936	0.5144	0.1362	0.4326	0.2072	0.5614
GCNSI	0.1419	0.3737	0.2055	0.6411	0.1158	0.3381	0.1725	0.5321	0.1133	0.2371	0.1533	0.5038	0.0721	0.1167	0.0879	0.5036	0.1047	0.3511	0.1613	0.5433
OJC	0.1543	0.2201	0.1814	0.5012	0.1414	0.1679	0.1535	0.5110	0.1414	0.1679	0.1535	0.5009	0.3750	0.1944	0.2500	0.5771	0.3977	0.1231	0.1800	0.5097
SL-VAE	**0.6667**	**0.8421**	**0.7442**	**0.9749**	**0.6695**	**0.5623**	**0.6112**	**0.9686**	**0.6846**	**0.6458**	**0.6646**	**0.9689**	**0.6250**	**0.8333**	**0.7143**	**0.8289**	**0.6541**	**0.5713**	**0.6099**	**0.9711**

Table 2: Performance over comparison methods under SIR diffusion pattern. (Best is highlighted with bold.)

	Digg-7556				Digg				Memetracker-7884				Memetracker			
Methods	PR	RE	F1	AUC	PR	RE	F1	AUC	PR	RE	F1	AUC	PR	RE	F1	AUC
LPSI	0.0026	0.0123	0.0043	0.4432	0.0079	0.2727	0.0155	0.5618	0.0132	0.3184	0.0253	0.5112	0.0087	0.2913	0.0169	0.5377
GCNSI	0.0114	0.3700	0.0221	0.4450	0.0123	0.2100	0.0232	0.4129	0.0211	0.3219	0.0396	0.4357	0.0197	0.2342	0.0363	0.4103
OJC	0.0118	0.0107	0.0112	0.5023	0.0635	0.0696	0.0664	0.5142	0.0542	0.0433	0.0481	0.4812	0.0331	0.0207	0.0255	0.5077
SL-VAE	**0.4131**	**0.6217**	**0.4655**	**0.5541**	**0.4297**	**0.5421**	**0.4792**	**0.6213**	**0.5113**	**0.6214**	**0.5610**	**0.5954**	**0.4612**	**0.5181**	**0.4880**	**0.6245**

Table 3: Performance evaluation over comparison methods under the real-world diffusion. (Best is highlighted with bold.)

- Ling, Chen, 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*, pp. 1010-1020. 2022.

Experiment

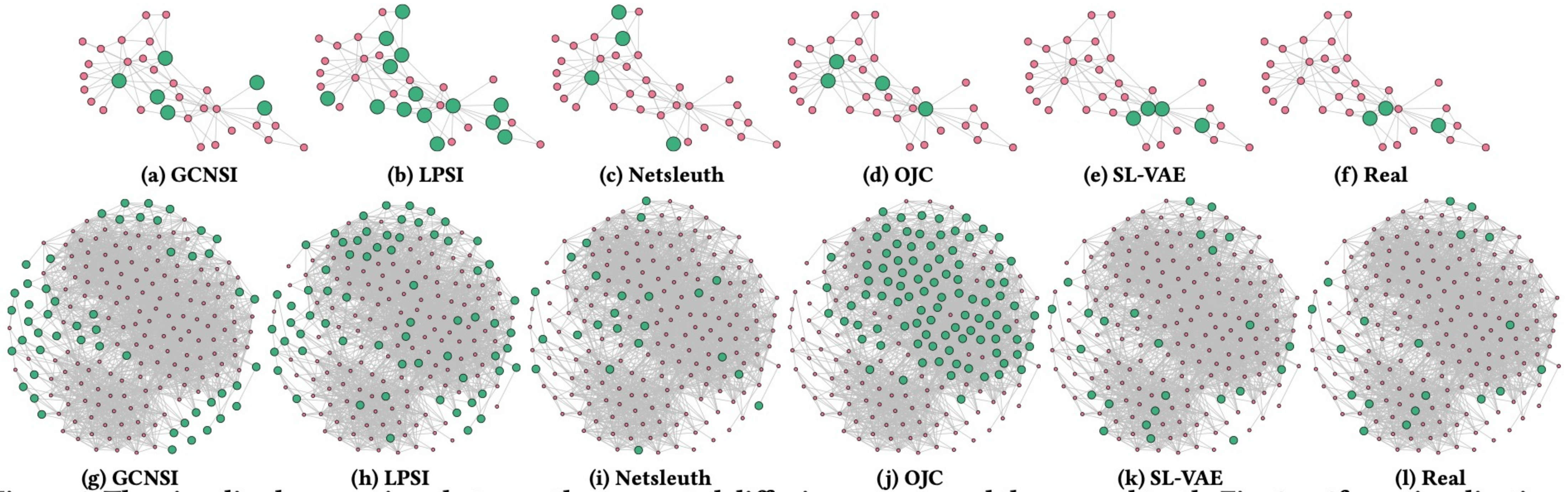
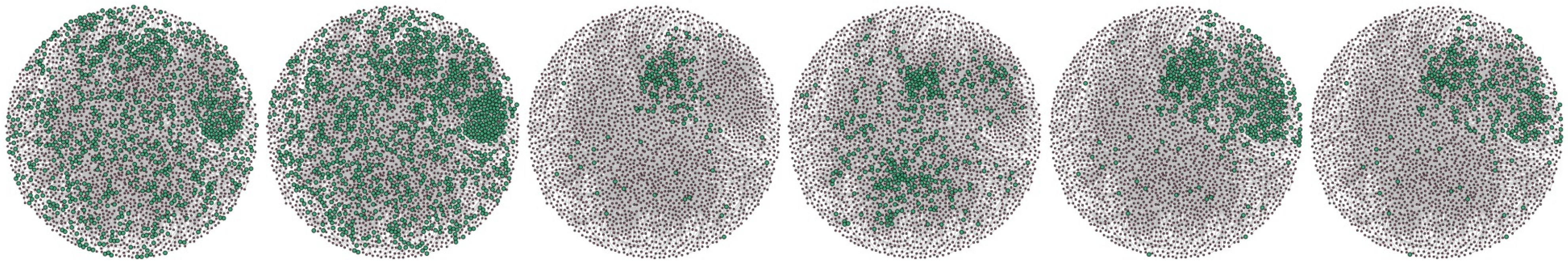


Figure 4: The visualized comparison between the generated diffusion sources and the ground truth. Fig. 4a - 4f are visualizations of Karate dataset, and Fig. 4g - 4l are visualizations of Jazz dataset. The predicted and ground truth diffusion sources are marked with green color while the rest of nodes are marked with red color.

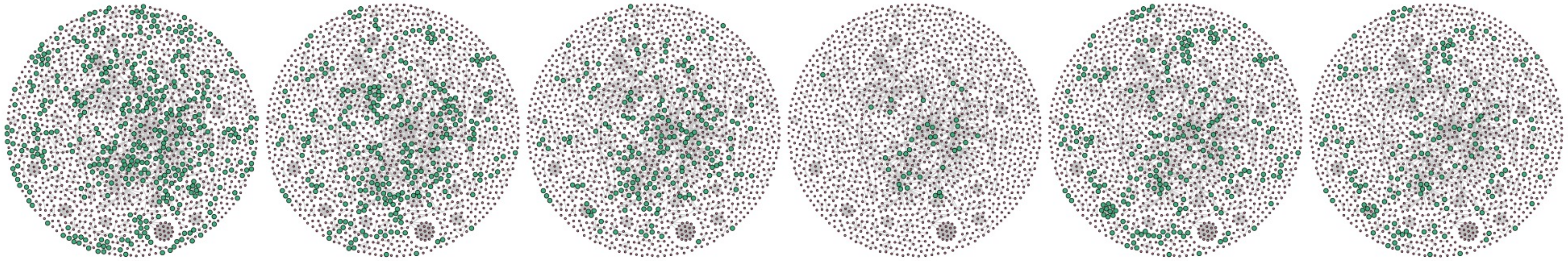
- Ling, Chen, Junji Jiang, Junxiang Wang, and Zhao Liang. "Source Localization of Graph Diffusion via Variational Autoencoders for Graph Inverse Problems." In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1010-1020. 2022.

Experiment



(a) GCNSI (b) LPSI (c) Netsleuth (d) OJC (e) SL-VAE (f) Real

Figure 6: The visualized comparison between the generated diffusion sources and the ground truth (Jazz).



(a) GCNSI (b) LPSI (c) Netsleuth (d) OJC (e) SL-VAE (f) Real

Figure 7: The visualized comparison between the generated diffusion sources and the ground truth (Network Science).

- Ling, Chen, 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*, pp. 1010-1020. 2022.

Outline

- Background
- Source Localization of Graph Diffusion:
  - Multiple Source Detection without Knowing the Underlying Propagation Model (AAAI 2017)
  - Multiple Rumor Source Detection with Graph Convolutional Networks (CIKM 2019)
  - Source Localization of Graph Diffusion via Variational Autoencoders for Graph Inverse Problems (KDD 2022)
- Summary

Summary

- **LPSI**:
  - First trial of localizing diffusion sources without assumption on underlying propagation model.
  - Use the idea of source prominence and label propagation on graphs.
- **GCNSI**:
  - Extend LPSI with GCN.
  - Enrich node features from different perspectives.
- **SL-VAE**:
  - Use VAE to quantify the uncertainty in source localization.
  - Use learned prior knowledge of diffusion source during inference.
  - Be able to generalize to different diffusion patterns.

Related References

- Wang, Junxiang, Junji Jiang, and Liang Zhao. "An Invertible Graph Diffusion Neural Network for Source Localization." In *Proceedings of the ACM Web Conference 2022*, pp. 1058-1069. 2022.
- Zhu, Kai, and Lei Ying. "Information source detection in the SIR model: A sample-path-based approach." *IEEE/ACM Transactions on Networking* 24, no. 1 (2014): 408-421.
- Nguyen, Hung T., Preetam Ghosh, Michael L. Mayo, and Thang N. Dinh. "Multiple infection sources identification with provable guarantees." In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pp. 1663-1672. 2016.
- Ying, Lei, and Kai Zhu. "Diffusion source localization in large networks." *Synthesis Lectures on Communication Networks*11, no. 1 (2018): 1-95.

Thanks for listening!
Q & A

ILLINOIS