Robust Influence Maximization Algorithm Design for Online Social Network

Yingtong Dou
Outline

- Background: Robust Influence Maximization (RIM)
- Model and Algorithms
  - CIC Model with NewDiscount and GreedyCIC algorithms
- Experiments and Results
  - Experiments under various noise
- Problems and Solutions
  - Datasets & Algorithms
- Conclusion and Future Work
  - New algorithm has better performance
  - Get the optimal match for robust performance
Background

- **Influence Maximization**
  - Find a set with $k$ nodes in a specific social network graph as the initial influence propagation nodes that make the final number of influenced nodes the largest. [1]

- **Robust Influence Maximization**
  - Find the seed set which has stable performance under various uncertainty factors of the model.

Background

- Robust Optimization Objective 1\(^{[1][2]}\)
  \[ \rho(S) = \min_{\sigma \in \Sigma} \frac{\sigma(S)}{\sigma(S^*)} \]

- Robust Optimization Objective 2\(^{[3]}\)
  \[ \delta(S, p) = \max_{S' \in S} \sigma_p(S') - \sigma_p(S) \]
  \[ \delta^{MR}(S, \mathcal{P}) = \max_{p \in \mathcal{P}} \delta(S, p) \]
  \[ \delta^{MMR}(\mathcal{P}) = \max_{S \in \mathcal{S}} \delta^{MR}(S, \mathcal{P}) \]

- Motivation
  - Noise on edge probability
  - Centrality based probability evaluation

Model and Algorithms

- Centrality Independent Cascading (CIC) Model

\[ P_{u,v} = \lambda \frac{C_u}{C_u + C_v} \], for \( C_u + C_v \neq 0 \)

where \( P_{u,v} \) is the edge activation probability from node \( u \) to node \( v \), \( C_u \) is the centrality of node \( u \), \( \lambda \) is modification coefficient.

- Centrality Measurement Methods
  - Degree Centrality, PageRank Centrality, Eccentric Centrality, Closeness Centrality
Model and Algorithms

- Algorithm Flow Chart

- Problem Definition

\[ \rho(S) = \min_{\sigma \in \Sigma} \frac{\sigma(S)}{\sigma(S^*)} \implies S^*_{c,p} := \arg\max_{S \subseteq V, |S|=k} \min_{c \in C, p \in P} \frac{\sigma_{c,p}(S)}{\sigma_{c,p}(S^*_{c,p})} \]
Model and Algorithms

- **GreedyCIC**
  - Add edge probability space as input to NewGreedyIC and set iteration times to 200

- **NewDiscount**
  - Add edge probability space as input to DegreeDiscount and define new $dd_v$ of node $v$
Experiments and Results

Datasets

| Name    | |Nodes| |Edges| Density | Description                                                                 |
|---------|----------------|------|------|------|-----------------------------------------------------------------------------|
| Retweet | 96             | 117  | 0.0257 | Retweeting users network with #political and #copen in Twitter              |
| FBMIT   | 6.4K           | 251.2K | 0.0123 | User network in Facebook whose university is MIT                            |
| Epinions| 26.6K          | 100.1K | 0.003  | User network from a online product reviewing website                        |
| Douban  | 154.9K         | 327.2K | 2.73e-05 | All the links among users of Chinese social website called Douban           |

Algorithms

- NewDiscount, GreedyCIC, DegreeDiscount, NewGreedyIC

Tools

- Python
- SNAP
- Networkx
- Gephi
Experiments and Results

(a)  
(b)  
(c)  
(d)
Experiments and Results

(a)

(b)

(c)

(d)
## Experiments and Results

**Average Running Time of Algorithms under Different Experiment Sets**

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Methods</th>
<th>Datasets</th>
<th>ART(seconds)</th>
<th>Algorithms</th>
<th>Methods</th>
<th>Datasets</th>
<th>ART(seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
<td>Heuristic</td>
<td>FBMIT</td>
<td>22.2</td>
<td>Degree</td>
<td>Greedy</td>
<td>FBMIT</td>
<td>268.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Epinions</td>
<td>30.1</td>
<td></td>
<td></td>
<td>Epinions</td>
<td>839.7</td>
</tr>
<tr>
<td>PageRank</td>
<td>Heuristic</td>
<td>FBMIT</td>
<td>24.1</td>
<td>PageRank</td>
<td>Greedy</td>
<td>FBMIT</td>
<td>203.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Epinions</td>
<td>29.9</td>
<td></td>
<td></td>
<td>Epinions</td>
<td>842.0</td>
</tr>
<tr>
<td>Eccentric</td>
<td>Heuristic</td>
<td>FBMIT</td>
<td>22.5</td>
<td>Eccentric</td>
<td>Greedy</td>
<td>FBMIT</td>
<td>202.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Epinions</td>
<td>36.9</td>
<td></td>
<td></td>
<td>Epinions</td>
<td>880.2</td>
</tr>
<tr>
<td>Closeness</td>
<td>Heuristic</td>
<td>FBMIT</td>
<td>24.6</td>
<td>Closeness</td>
<td>Greedy</td>
<td>FBMIT</td>
<td>198.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Epinions</td>
<td>42.4</td>
<td></td>
<td></td>
<td>Epinions</td>
<td>881.2</td>
</tr>
<tr>
<td>DegreeDiscount</td>
<td>Heuristic</td>
<td>FBMIT</td>
<td>22.9</td>
<td>NewGreedyIC</td>
<td>Greedy</td>
<td>FBMIT</td>
<td>194.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Epinions</td>
<td>27.0</td>
<td></td>
<td></td>
<td>Epinions</td>
<td>905.1</td>
</tr>
</tbody>
</table>
Robust Performance of Algorithms under Different $\lambda$ Value.

(a) is under FBMIT
(b) is under Epinions
Problems and Solutions

- Sina Weibo dataset is hard to get
  - Discuss with the supervisor about the challenge
  - Use Douban dataset to replace

- Which algorithm to improve?
  - Talk with Dr. Zhou about my confusion
  - Take the work of Chen et al. as baseline\(^1\)[2]


\[^2\] Chen, W., Yuan, Y., & Zhang, L. (2010). Scalable Influence Maximization in Social Networks under the Linear Threshold Model. *ICDM 2010, the, IEEE International Conference on Data Mining, Sydney, Australia, 14-17 December* (pp.88-97). DBLP.
Conclusion and Future Work

- The influence spread is indeed influenced by various uncertainty

- PageRank Centrality and Degree Centrality based algorithms have similar performance

- PageRank Centrality based greedy algorithm has better influence spread and takes less time than NewGreedyIC in large dataset

- All the algorithms have best robust performance when modification coefficient $\lambda=0.01$
Conclusion and Future Work

For my new model
- More datasets and algorithms should be evaluated
- Evaluate more centrality measurement methods
- Design new methods evaluating the edge probability
- Compare the performance in directed and undirected network

For RIM problem
- Design new robust optimizing objective
- Evaluate other noise factors in the model
- Improve the efficiency to find robust results of the problem
- Do experiments on large datasets with distributed system
Questions & Answers

Yingtong Dou