Robust Influence Maximization Algorithm Design for Online Social Network

Yingtong Dou

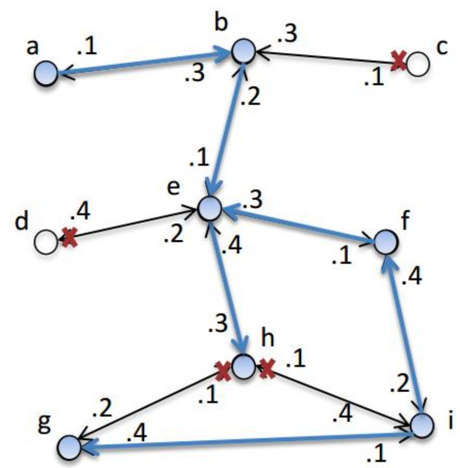
Outline

- 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.



[1] D. Kempe, J. M. Kleinberg, and ƒ. Tardos. Maximizing the spread of influence through a social network. In Proceedings of the 9th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 137–146, 2003.

Background

- Robust Optimization Objective 1[1][2]

$$\rho(S) = \min_{\sigma \in \Sigma} \frac{\sigma(S)}{\sigma(S_\sigma^*)}$$

- Robust Optimization Objective 2[3]

$$\delta(S, \boldsymbol{p}) = \max_{S' \in S} \sigma_P(S') - \sigma_P(S) \quad \delta^{MR}(S, \mathcal{P}) = \max_{p \in \mathcal{P}} \delta(S, \boldsymbol{p})$$

$$\delta^{MMR}(\mathcal{P}) = \max_{S \in \mathbb{S}} \delta^{MR}(S, \mathcal{P})$$

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

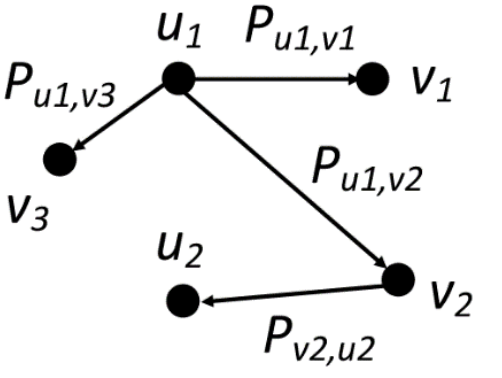
[1] He, X., & D. K. (2016). Robust Influence Maximization. Kdd, 885–894. http://doi.org/10.1145/2939672.2939760
[2] Chen, W., Lin, T., Tan, Z., Zhao, M., & Zhou, X. (2016). Robust Influence Maximization. Kdd, 795–804.
[3] Lowalekar, M., Varakantham, P., & Kumar, A. (2016). Robust Influence Maximization - (Extended Abstract). Aamas.

Model and Algorithms

- Centrality Independent Cascading (CIC) Model

$$P_{u,v} = \lambda \frac{C_u}{C_u + C_v}, \text{for } C_u + C_v \neq 0$$

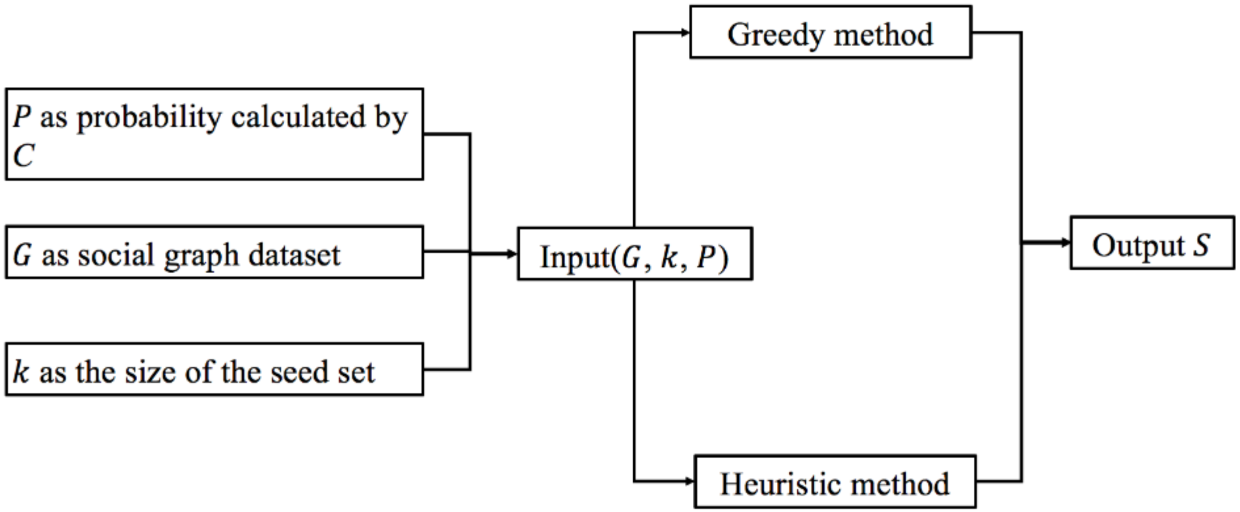
where $P_{u,v}$ is the edge activation probability from node u to node v, C_u is the centrality of node u, λ is modification coefficient.



- Centrality Measurement Methods
  - Degree Centrality, PageRank Centrality, Eccentric Centrality, Closeness Centrality

Model and Algorithms

- Algorithm Flow Chart



- Problem Definition

$$\rho(S) = \min_{\sigma \in \Sigma} \frac{\sigma(S)}{\sigma(S_\sigma^*)} \quad \Rightarrow \quad S_{C,P}^* := \operatorname*{argmax}_{S \subseteq V, |S| = k} \min_{c \in C, p \in P} \frac{\sigma_{c,p}(S)}{\sigma_{c,p}\left(S_{c,p}^*\right)}$$

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

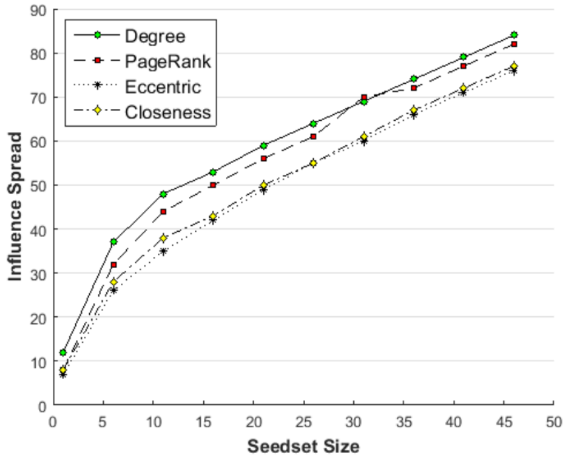
- Datasets

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

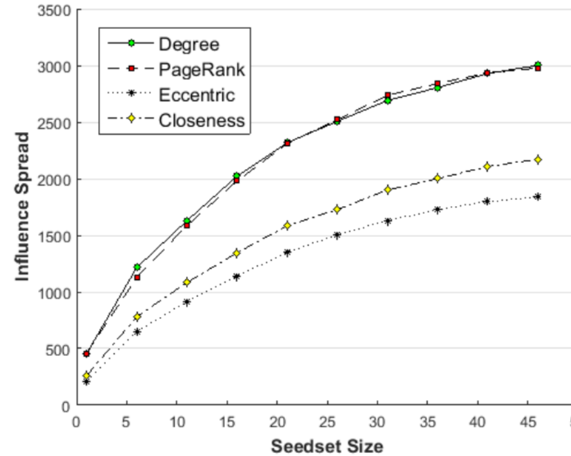
- Algorithms
  - NewDiscount, GreedyCIC, DegreeDiscount, NewGreedyIC
- Tools

python SNAP Networkx Gephi

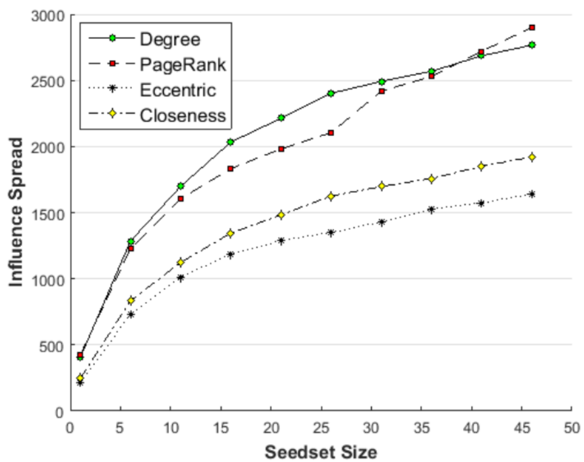
Experiments and Results



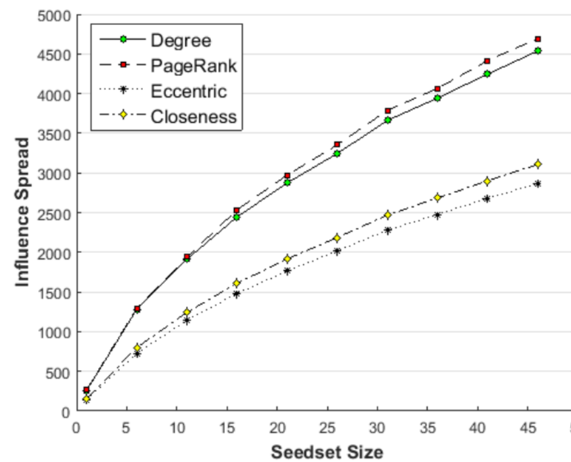
(a)



(b)

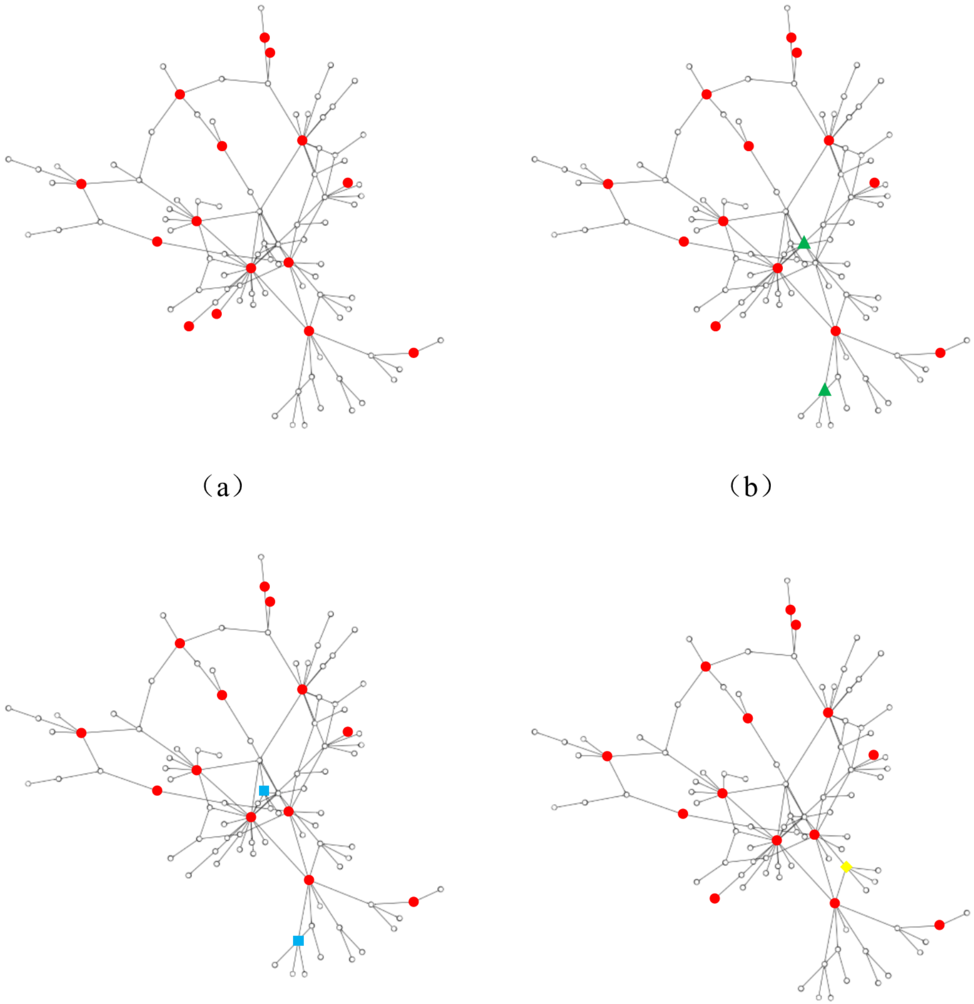


(c)



(d)

Experiments and Results



（a） （b） （c） （d）

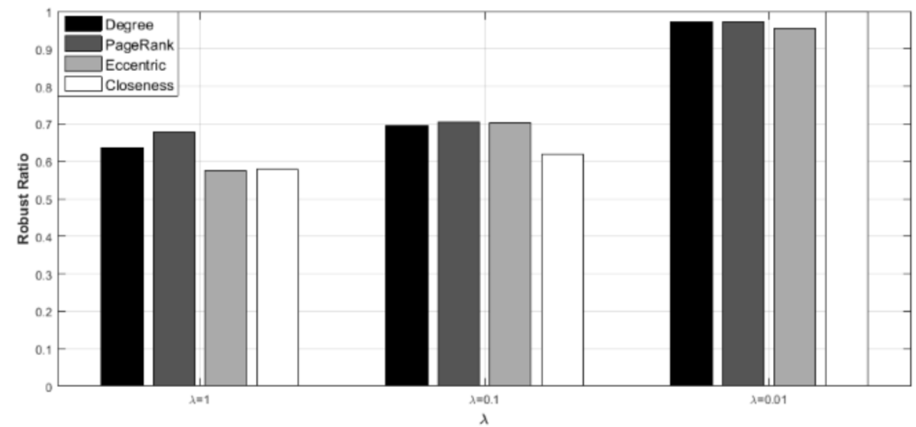
Experiments and Results

- Average Running Time of Algorithms under Different Experiment Sets

Algorithms	Methods	Datasets	ART(seconds)	Algorithms	Methods	Datasets	ART(seconds)
Degree	Heuristic	FBMIT	22.2	Degree	Greedy	FBMIT	268.7
		Epinions	30.1			Epinions	**839.7**
PageRank	Heuristic	FBMIT	24.1	PageRank	Greedy	FBMIT	203.0
		Epinions	29.9			Epinions	**842.0**
Eccentric	Heuristic	FBMIT	22.5	Eccentric	Greedy	FBMIT	202.6
		Epinions	36.9			Epinions	**880.2**
Closeness	Heuristic	FBMIT	24.6	Closeness	Greedy	FBMIT	198.4
		Epinions	42.4			Epinions	**881.2**
DegreeDiscount	Heuristic	FBMIT	22.9	NewGreedyIC	Greedy	FBMIT	194.7
		Epinions	27.0			Epinions	**905.1**

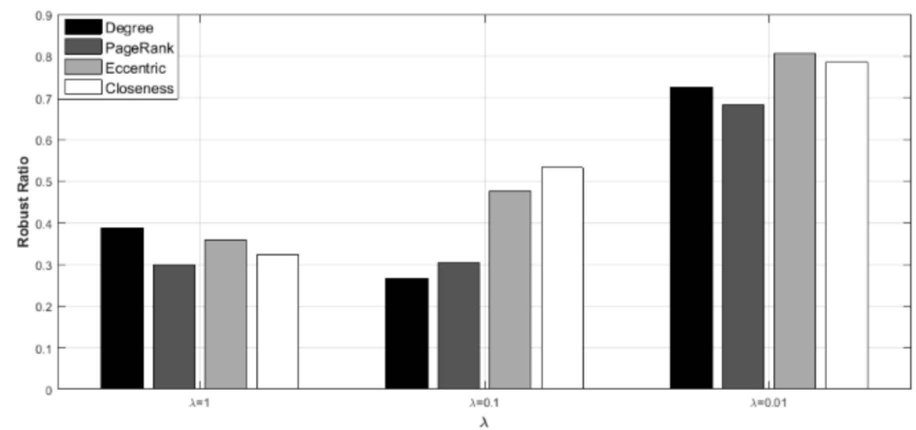
Experiments and Results

- Robust Performance of Algorithms under Different λ Value.



(a)

- (a) is under FBMIT
(b) is under Epinions



(b)

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]

[1] Chen, W., Wang, Y., & Yang, S. (2009). Efficient influence maximization in social networks. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28 - July*(Vol.199-208, pp.199-208). DBLP.
[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 coeeficient $\boldsymbol{\lambda}$**=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