A Game Theoretic Approach to Influence Limitation Problem

Final Presentation

Sourav Medya
Advisor: Prof Y. Narahari

Department of Computer Science & Automation
Indian Institute of Science, Bangalore
Bengaluru, India.

Two different problems

[1] Virus Inoculation problem

To identify the nodes in the graph with a constraint on the number of nodes such that vaccination of those nodes would result in a minimum number of infected nodes

[2] Limitation of Misinformation Problem

- Given:
  - Negative campaign/misinformation which starts from specified sources and detected after time delay d
  - Presence of positive/counter campaign to limit the former
- To Find: Top k suitable nodes where the positive campaign is to be triggered

Three Major Contributions

[1] Virus Inoculation: Preventive Methods (Before Infection)

- Compare game theoretic centralities with conventional centralities
- Propose a solution in non-submodular probabilistic setting

[2] Virus Inoculation: Reactive Methods (After Infection)

- Propose a novel reactive model for virus inoculation problem
- Propose two algorithms in the above setting
- Propose another model and proof for NP-hardness

[3] Limitation of Misinformation

- Propose a novel characteristic function
- Proposed heuristic outperforms existing heuristics

Virus Inoculation in Probabilistic Model

The Probabilistic Model:

- Model assumes probabilistic infection with vaccination cost (C) = infection cost (L)
- $f_s(T)$ be the the number of nodes that get infected by initial infected set S
- $f(T) = \sum_{S \subseteq V(G)} q(S) f_s(T)$ where $q(S)$ is the probability of set S to get infected initially and so, $q(S) = \prod_{i \in S} q_i \prod_{i \notin S} (1 - q_i)$
- Problem: Choose set T of k nodes to minimize $f(T)$

Definition (Submodular Function)

For every subset S, T of $V(G)$ where $S \subseteq T$ and any node v,
$f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T)$ (for supermodularity the inequality is reversed)

Lemma

$f(T)$ in the probabilistic model is not submodular.

Reference

Zeinab Abbasi and Hoda Heidari, “Toward Optimal Vaccination Strategies for Probabilistic Models,” in *Proceedings of the 20th International Conference on World wide web, WWW*; 2011, pp. 1-2.

Game Theoretic Approach

Key Points

- How does game theoretic centrality improve the solution?
- Game theoretic centrality vs conventional degree centrality

Characteristic Function $\nu_1 : 2^N \rightarrow \mathbb{R}$

$$\nu_1(C) = \begin{cases} 0 & \textit{if } C = \emptyset \\ \textit{size}(\textit{victim}(C)) & \textit{else} \end{cases}$$

victim(C)

$$\{v : v \in C \textit{ or } \exists u \in C \textit{ such that } (u, v) \in E(G)\}$$

Reference

JPo-An Chen, Mary David and David Kempe, “Better Vaccination Strategies for Better People,” in *Proceedings of the 11th ACM Symposium on Electronic Commerce, EC*; 2010, pp. 294–301.

Degree Centrality vs Game Theoretic Approach

Shapley Value

$$\phi(v_i) = \sum_{w_i \in \{v_i\} \cup N_{G_o}(v_i)} \frac{1}{1 + deg_{G_i}(w_i)}$$

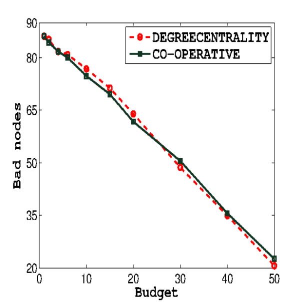
where

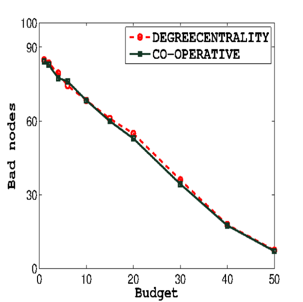
$N_{G_o}(v_i)$=$\{u \in V(G) : (v_i, u) \in E(G)\}$
$N_{G_i}(v_i)$=$\{u \in V(G') : (u, v_i) \in E(G)\}$
$deg_{G_o(v_i)} = |N_{G_o}|$
$deg_{G_i(v_i)} = |N_{G_i}|$

Improvement

Shapley value is more for nodes with higher degree and nodes whose neighbours are weak

Results





Figure

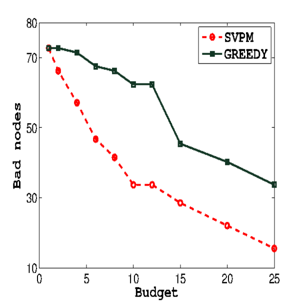
Bad (Infected) nodes (percentage of nodes) versus budget (percentage of nodes) using Football and Celegans data set

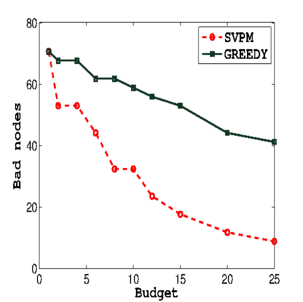
Proposed Preventive Method Based Algorithm: SVPM

Key Points

- Compare greedy algorithm vs SVPM
- SVPM: Greedy approach in game theoretic sense
- Solve an optimization problem using co-operative game theoretic tool
- Shapley value based concept which is agnostic to submodularity
- Characteristic Function: $\nu_2(C) = n - f(C)$
- Running Time: Similar

Results





Figure

Bad(Infected) nodes (percentage of nodes) versus budget (percentage of nodes) using Karate and Celegans data set

Model to Capture Reactive Methods

Model: General Cascade Model, GCM

- Take motivation from Independent Cascade Model(ICM)
- Infection starts at one node and spreads following ICM
- Detected at time delay *d*
- Some nodes get vaccinated and remove from graph
- Again infection progress following ICM
- Problem: to find nodes to give vaccination

Reference

D. Kempe, J. Kleinberg and E. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the 9th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD*; 2003, pp. 137-146.

Proposed Model to Capture Reactive Methods

Issues

- No models for virus inoculation have reactive strategy
- Assume $C = L$
- Introduce one kind of probability: Infection probability, p_{uv}

Problem Formulation:

- $g : S \rightarrow [0, n]$ where $S = \{W | W \subset V \text{ and } |W| \leq k\}$.
- The goal is to find a set W ($|W| \leq k$) of nodes to vaccinate after d to maximize $g(.)$

Lemma

Underlying function g in this model is not submodular.

Two Algorithms: SVRM and TRM

SVRM

- Based on Global Method
- Find most powerful nodes in the network based on infection probability
- Running Time: $O(qR(n+m)+nlogn)$ where q and R are number of permutations and iterations respectively

TRM

- Based on Local Method
- Find most likely nodes to be infected next
- Running Time: $O(k(n+m)+nlogn)$

Two Algorithms: SVRM and TRM

SVRM

$$\nu_3(C) = \sigma(C) \ \forall C \subseteq N$$

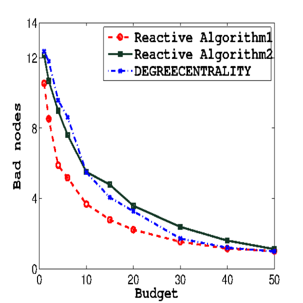
where $\sigma(C)$ gives the expected number of infected nodes if C is the initial infected set

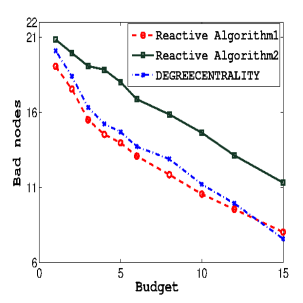
TRM

$$f_t(\Delta_{nxt}) = \sum_{T \in \Gamma(G')} \Pi_{(u,v)\in T} p_{uv}$$

where $|\Delta_{nxt}| = k$ and $\Gamma(G')$ is set of trees in G' with infected nodes and Δ_{nxt}

Results

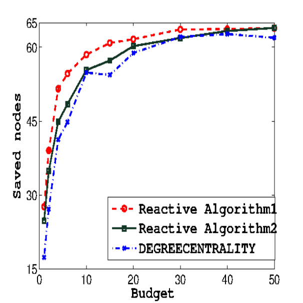


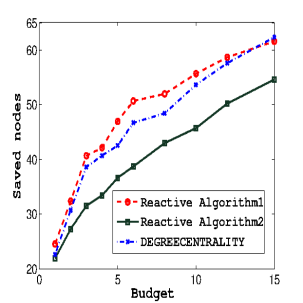


Figure

Bad nodes (percentage of nodes) versus budget (percentage of nodes) using Football and Jazz data set

Results





Figure

Saved nodes (percentage of nodes) versus budget (percentage of nodes) using Football and Jazz data set

Model to Capture Reactive Methods

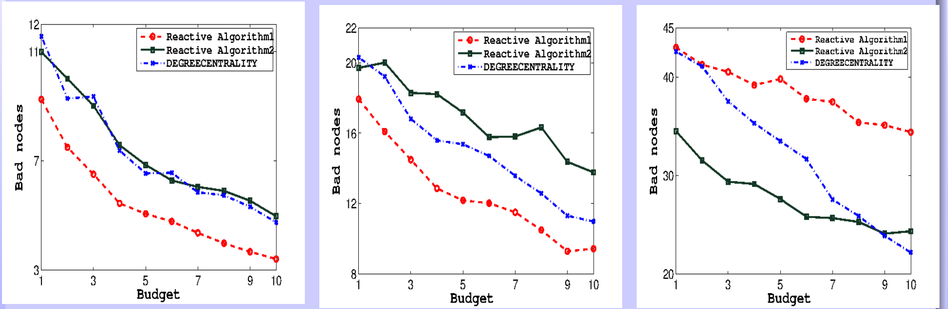
Model: Modified General Cascade Model, MGCM

- Assumption on GCM
- Neighbours of vaccinated nodes are also saved
- Problem: to find nodes to give vaccination

Lemma

Virus inoculation problem under MGCM is NP-hard

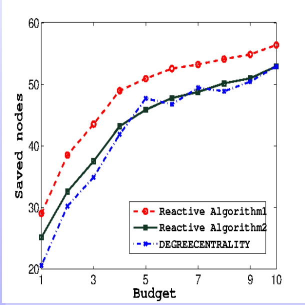
Results

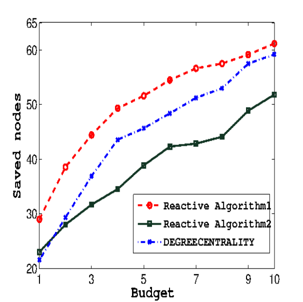


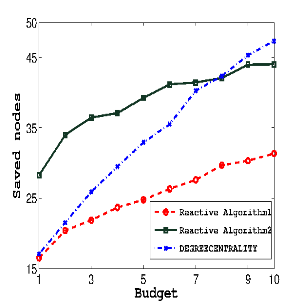
Figure

Bad nodes (percentage of nodes) versus budget (percentage of nodes) using Football, Jazz and elegans data set

Results







Figure

Saved nodes (percentage of nodes) versus budget (percentage of nodes) using Football, Jazz and elegans data set

Model

Limiting Misinformation Problem

- The problem is first proposed by Budak, Agarwal and Abbadi
- More generalized version is proposed and solved by Premm Raj

Limiting Misinformation vs Virus Inoculation

- Limiting rumour does not involve costs
- No question of preventive methods, only involves reactive methods
- Unlike virus inoculation problem, positive influence is used to limit misinformation

Reference

[1] Ceren Budak and Divyakant Agrawal and Amr El Abbadi, "Limiting the Spread of Misinformation in Social Networks," in *Proceedings of the 20th International conference on World Wide Web, WWW*; 2011, pp. 665–674.

[2] H Premm Raj and Y. Narahari, "Influence Limitation in Multi-Campaign Social Networks with Non-Submodular Influence Functions: A Shapley Value Based Approach," in *Proceedings of IEEE conference on Automation Science and Engineering (CASE)*; 2012

Our Algorithm

Key Points

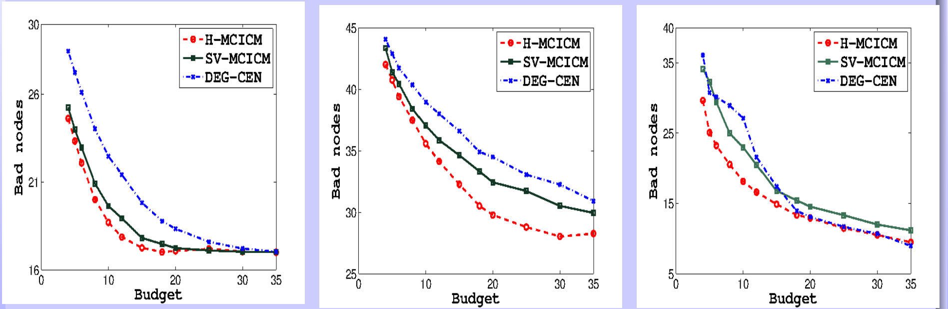
- Work with MCICM
- Algorithm uses Shapley value
- Novel characteristic function
- To stop the powerful nodes of negative campaign
- To trigger positive influence to the powerful nodes of positive campaign

H-MCICM

$$\nu_4(C) = \sigma_R(C) + \sigma_G(C) \;\; \forall C \subseteq N$$

where $\sigma_R(C)$ and $\sigma_G(C)$ give the expected number of nodes misinformed and influenced by positive campaign respectively

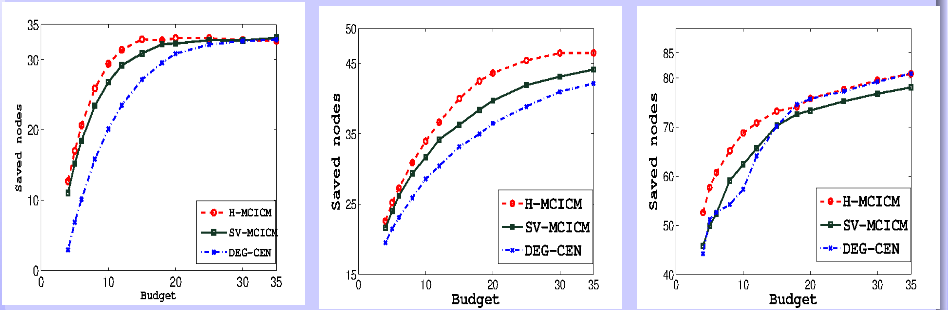
Results



Figure

Bad nodes (percentage of nodes) versus budget (percentage of nodes) using Football, Jazz and elegans data set

Results



Figure

Saved nodes (percentage of nodes) versus budget (percentage of nodes) using Football, Jazz and elegans data set

The Whole Picture

- Preventive Method for virus inoculation:
  - Work with probabilistic model
  - Minimization problem
  - Underlying function is not submodular
  - Any tool which does not depend on submodularity will work well

- Reactive Method for virus inoculation:
  - No model captures reactive strategies
  - Maximization problem
  - Underlying function is not submodular
  - Propose two methods to solve the problem

The Whole Picture

- Reactive Method for virus inoculation:
  - Modify the model with extra assumption
  - The problem under this model is NP-hard
- Modified Algorithm for Limiting Misinformation
  - Maximization problem
  - Underlying function is not submodular
  - Propose novel characteristic function

The Path Ahead: 1 of 2

Game Theoretic Aspects

- Incorporating Costs and Game: Does Nash Equilibrium exist?
    - If yes, how far is it from social optimum?
    - If no, what will be the social optimum?
- Network Formation: Developing a network of suitable characteristics which makes vaccination more effective.

The Path Ahead: 2 of 2

Algorithmic Aspects

- Special Structures:
  - Submodular functions can be well approximated by constant factor
  - In both cases underlying functions are non-submodular
  - Designing some algorithms which can well approximate certain kinds of network with some special graphical structures
- Agnostic Approximation Algorithm: Designing a good approximation algorithm which is agnostic towards underlying nature of the function

THANK YOU !!