

A Game Theoretic Approach to Influence Limitation Problems

A PROJECT REPORT
SUBMITTED IN PARTIAL FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
Master of Engineering
IN
FACULTY OF ENGINEERING

by

Sourav Medya



Computer Science and Automation
Indian Institute of Science
Bangalore – 560 012

July 2012

Dedicated

To

My Parents

Acknowledgements

I am extremely grateful to my guide, Prof. Y. Narahari, for his valuable guidance and constant encouragement during the course of my project work. I have been extremely fortunate to have an advisor who gave me the freedom to explore on my own, and at the same time the guidance to recover when my steps faltered. I would also like to thank my faculty advisor Dr. Aditya Kanade for his support during my stay at IISc.

I thank the Department of Computer Science and Automation for providing excellent ambiance and facilities. I am grateful to all the faculty members of CSA department for enlightening me with their knowledge during the last two years. The Game Theory lab has provided me a very pleasant atmosphere for doing my project work. I wish to express my gratitude to my fellow lab-mates and friends Rohith D. Vallam, Swaprava Nath, Pankaj Dayama, Prem Raj H, Ramasuri Narayanam, Swapnil Dhamal, Satyanath Bhat, Shweta Jain, Chidanand Sridhar, Sk. Mohammadul Haque, Ratul Ray and Ashutosh Verma. I am thankful to them for reading my reports, commenting on my views and helping me understand and enrich my ideas.

Abstract

Influence limitation problem is an important problem on networks, especially on computer and social networks. Here we study two kinds of influence limitation problems: (1) Virus inoculation and (2) limitation of misinformation. Large communities are always susceptible toward the spread of epidemic outbreaks. The occurrence of a widespread disease transmission like the propagation of viruses through computer network raises the natural question of optimal inoculation strategies. Here we study the vaccination of graphs against the outbreak of infectious diseases in a social network under preventive and reactive settings. We also study the problem of limitation of misinformation in a social network. Algorithms for limiting misinformation are needed to save a social network from panic.

First, we study the inoculation problem under a probabilistic model in preventive setting. Here a social network is represented as a graph and probability on a directed edge represents the chance of target node getting infected given that the source node is already infected. We consider the optimization problem: given that the number of nodes to be vaccinated is limited to k , our objective is to minimize the overall expected number of infected nodes in the network. Due to NP-hardness of the problem, we propose a heuristic algorithm which is based on Shapley value. We also investigate the performance of the game theoretic centralities as opposed to conventional centralities in solving this problem.

Second, we propose a model of inoculation under reactive strategies. If we have a social network under the same probabilistic model, we look for the strategies when the graph is already infected. We consider a second optimization problem: given that a virus

infection starts propagating from a specified source and vaccination is to be used to limit the spread of infection, we are interested in selecting k nodes at which the vaccination is to be given so as to save maximum number of nodes from the infection. In this report, we propose two approaches for this infection limitation optimization problem.

Third, we investigate another kind of influence limitation problem where positive influence is used to limit the effect of negative influence. Given a negative campaign/misinformation which starts from a specified source and a positive/counter campaign to be initiated to limit the effect of the former, we are interested in finding the top k suitable nodes where the positive campaign is to be triggered. We propose an improved Shapley value based approach for this limitation of misinformation problem.

Contents

Acknowledgements	i
Abstract	ii
Keywords	viii
1 Introduction	1
1.1 Virus Inoculation Problem	1
1.2 Limitation of Misinformation Problem	2
2 Preliminaries	3
2.1 Cooperative Game Theory	3
2.2 Shapley Value	3
2.3 Submodularity	5
3 Relevant Work	6
3.1 Current State of The Art	6
3.1.1 Virus Inoculation Problem and Game	6
3.1.2 Influence Maximization Problem	8
3.1.3 Limitation of Misinformation Problem	9
3.2 Contributions	10
3.3 Experimental Setup and Organization of this manual	12
4 Virus Inoculation in Preventive Setting	16
4.1 Model	16
4.2 Degree Centrality vs Game Theoretic Approach	18
4.3 Experimental Results and Analysis	23
4.4 Greedy Algorithm vs Game Theoretic Approach	23
4.5 Experimental Results and Analysis	28
5 Virus Inoculation in Reactive Setting	29
5.1 Issues	29
5.2 Problem Formulation	30
5.3 Reactive Method based Algorithms	31
5.4 Experimental Results and Analysis for GCM	39

5.5	Another Model: Modified General Cascade Model	39
5.6	Experimental Results and Analysis for MGCM	41
6	Algorithms for Limitation of Misinformation	43
6.1	Issues	43
6.2	Problem Formulation and Algorithm	44
6.3	Experimental Results and Analysis	46
7	Conclusion and Future Work	50
	Bibliography	52

List of Tables

3.1	Data Sets used in the experiments	14
4.1	Notation Table for Section 4	19
4.2	Data Sets with q used in the experiments	28
5.1	Notation Table for Section 5	32

List of Figures

4.1	<i>This figure shows how $g(T)$ does not satisfy supermodular and submodular property</i>	17
4.2	<i>Bad nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Netscience co-authorship data set, (ii) Celegans data set, (iii) Football data set</i>	18
4.3	<i>Bad nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Football data set, (ii) Celegans data set</i>	23
4.4	<i>This figure shows how to compute the best node by SV</i>	26
5.1	<i>(i) To show how $g(.)$ does not satisfy submodular property; (ii) is the live graph for the infection influence propagation; (iii) shows the edges which are absent in live graph due to sampling</i>	36
5.2	<i>Under GCM: Bad nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Netscience co-authorship data set, (ii) Jazz data set, (iii) Football data set</i>	37
5.3	<i>Under GCM: Saved nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Smart Grid data set, (ii) Jazz data set, (iii) Football data set under GCM</i>	37
5.4	<i>Under MGCM: Bad nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Celegans data set, (ii) Jazz data set, (iii) Football data set</i>	41
5.5	<i>Under MGCM: Saved nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Celegans data set, (ii) Jazz data set, (iii) Football data set</i>	41
6.1	<i>Bad nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Football data set, (ii) Jazz data set, (iii) Celegans data set</i>	46
6.2	<i>Good nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Football data set, (ii) Jazz data set, (iii) Celegans data set</i>	47
6.3	<i>Not bad nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Football data set, (ii) Jazz data set, (iii) Celegans data set</i>	47
6.4	<i>Saved nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Football data set, (ii) Jazz data set, (iii) Celegans data set</i>	47

Keywords

Social networks, Virus inoculation, Shapley value, Submodular function, Supermodular function, Infection influence, Misinformation

Chapter 1

Introduction

We consider two kinds of influence limitation problems: one is virus inoculation and the other is limitation of misinformation. First we discuss the virus inoculation problem.

1.1 Virus Inoculation Problem

Large communities are always susceptible toward the spread of outbreaks, for example, Swine flu. The need of inoculation strategies is quite inherent from this kind of occurrence of infection. The problem of propagation of viruses through email and computer networks has a common structure with that of epidemics problem. Due to different constraints, it is not possible to vaccinate every individual or computer during a virus outbreak. So, we have to use the structure of network to identify the nodes or users whose inoculation would result in a lower number of infected nodes or people, respectively.

Most of the existing models [15], [13], [10], [7] assume a common structure, i.e., if a node is infected in the network it will infect all its neighbours with probability 1. As the assumption is not realistic, Abbasi and Hoda [2] propose a new model where an infected node can infect its neighbour with probability depending upon the type of relation between nodes, type of disease etc. To solve the problem of vaccination in this model is hard as the original problem itself is already NP-complete.

1.2 Limitation of Misinformation Problem

Social networks have been used as a medium of high speed information diffusion. Now a days, as the online social networks have large number of users, information can propagate faster than common media like news papers, news channels etc. But this power of diffusion can be misused easily. Panic caused due to large scale misinformation about swine flu [14] is one such example. It is clear that there is a possibility of spreading rumour through social network. We need effective ways to check the news on social network to save the society from panic.

Given a set of adversary nodes I_R who start misinformation campaign R (R for rival) and it is detected after delay d , after which a positive/counter campaign G (G for good) is initiated to limit the spread of misinformation campaign, we consider the problem of finding the most suitable nodes in the network where positive campaign can be initiated. This kind of positive influence is absent in virus inoculation problem.

Chapter 2

Preliminaries

2.1 Cooperative Game Theory

A cooperative game [20, 17] with transferable utility is defined as the pair (N, v) where $N = \{1, 2, \dots, n\}$ is the set of players and the mapping $v : 2^N \rightarrow R$ is a characteristic function, that assigns a value to each of the 2^N possible subsets of N , with $v(\emptyset) = 0$.

The value $v(S)$ for any subset S of N represents the total transferable utility that can be attained by the members in S without any help from the members in $N \setminus S$. It is also known as the worth of coalition S . The set of players N is called the grand coalition and $v(N)$ is called the value of the grand coalition.

2.2 Shapley Value

Many solution concepts such as the Shapley value, the core, and the nucleolus can be used to analyze a cooperative game. Each of them provides a method for dividing the total value of the game among individual players. A brief description of the Shapley value is given below.

Shapley value [22, 17] developed by Shapley, considers the relative importance of each player to the game in deciding the payoff to be allocated to the players and provides a unique expected payoff allocation for any cooperative game (N, v) . Shapley value is a

unique mapping that satisfies three properties namely linearity, symmetry, and carrier property. Some players may contribute more to the value of the coalition than others, thus it is important that gains are distributed fairly among the players, which is ensured by the above three properties.

The Shapley value of a cooperative game (N, v) denoted by $\Phi(N, v)$ is given as follows,

$$\Phi(N, v) = (\Phi_1(N, v), \Phi_2(N, v), \dots, \Phi_n(N, v))$$

where the Shapley value, $\Phi_i(N, v)$ is the expected payoff of a player i is given by,

$$\Phi_i(N, v) = \sum_{C \subseteq N \setminus i} \frac{|C|!(n - |C| - 1)!}{n!} \{v(C \cup \{i\}) - v(C)\}$$

Several equivalent formulations for the Shapley value exist in the literature. One such alternative formulations which uses the concept of marginal contribution of the node is described as follows, Given a node $i \in N$ and a subset $C \subseteq N$ such that $i \notin C$, the marginal contribution of node i to the coalition C is defined as,

$$v(C \cup \{i\}) - v(C), \forall C \subseteq N \setminus \{i\}$$

Let Ω denote the set of all $n!$ permutations on N . Let π be a permutation in Ω and $C_i(\pi)$ denote the set of nodes appearing before node i in the permutation π . The average marginal contribution of node i to the given cooperative game, which is also the Shapley value of player i is defined as,

$$\frac{1}{n!} \sum_{\pi \in \Omega} [v(C_i(\pi) \cup \{i\}) - v(C_i(\pi))]$$

To find the marginal contribution of a node we have work with $n!$ permutations. Thus, the computational complexity of the naive approach is $O((\frac{n}{e})^n)$ which is not tractable if n is large. To overcome this issue, many techniques have been proposed, such as [6] to approximately calculate the Shapley values. For our simulations we follow the method

proposed in [6]. Thus, instead of using all the $n!$ permutations, we consider only $t = O(n)$ permutations.

2.3 Submodularity

Consider a function $f(\cdot)$ as a mapping from a collection of sets to non-negative real numbers. We say that f is *submodular* [19, 12] if it satisfies a natural diminishing returns property, which implies that the marginal gain from adding an element to a set S is at least as high as the marginal gain from adding the same element to a superset of S . Formally, a submodular function satisfies

$$f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T)$$

for all elements v and all pairs of sets $S \subseteq T$.

Consider a function f that is submodular, takes only nonnegative values, and is monotone non-decreasing i.e., $f(S \cup \{v\}) \geq f(S)$ for all elements v and sets S and we want to find a k -element set S for which $f(S)$ is maximized. This is a well known NP-hard optimization problem. The result of [19] shows that a greedy hill-climbing algorithm approximates the optimum to within a factor of $(1 - 1/e)$, where e is the base of the natural logarithm. According to the greedy algorithm, we start with an empty set and repeatedly add an element that gives the maximum marginal gain.

THEOREM: [19]: For a non-negative, monotone submodular function f , let S be a set of size k obtained by selecting elements one at a time, each time choosing an element that provides the largest marginal increase in the function value. Let S^* be a set that maximizes the value of f over all k -element sets. Then $f(S) \geq (1 - 1/e) \cdot f(S^*)$; in other words, S provides a $(1 - 1/e)$ approximation.

Chapter 3

Relevant Work

3.1 Current State of The Art

The study of epidemic disease has always been a vital topic and as a result, scientists from different fields: epidemiologists, economists, and computer scientists, have investigated the problem of vaccination. Different models have been proposed regarding the transmission process of viruses and also the type of impact of the infections on the whole society. Similarly, various algorithms have been proposed to minimize the effect of infection. In this section, we describe the relevant work related to virus inoculation strategies and some other problems like influence maximization and influence limitation, which have some similarities with the virus inoculation problem.

3.1.1 Virus Inoculation Problem and Game

The problem of virus inoculation in social networks is first introduced by Aspnes, Chang and Yampolskiy [3]. To model the containment of the spread of viruses in a network, the authors propose a game in which each node has to decide whether to install anti-virus software or to accept the risk of infection. Each node acts selfishly to minimize its own expected cost. They show that the price of anarchy is $O(n)$ in the worst case which shows that it is undesirable for selfish agents to use Nash equilibrium strategies. Though in some cases Nash Equilibrium can be computed in $O(n^3)$ time, they show

that it is NP-hard to compute a social optimum. They reduce the problem to *sum of squares partition problem*. They propose a greedy algorithm based on sparse cuts and approximate the problem within a factor of $O(\log^2 n)$.

Moscibroda, Schmid and Wattenhofer [15] allow users to be malicious/byzantine rather than only being selfish. Byzantine players are those players who want to minimize the performance of the system and they are not worried about their own profits. The authors look at the impact of malicious users on the system efficiency when compared to the social optimal outcome, or in the presence of only selfish users. To capture the decrease in efficiency they introduce the price of malice. Moreover, they have quantified and upper bounded the *Fear Factor*, which is the gain in system efficiency arising from the increased willingness of selfish individuals to cooperate in order to counter by the presence of malicious players. They examine this problem with respect to containing the spread of a virus in a software system.

Meier, Oswald, Schmid and Wattenhofer [13] propose a framework which allows to measure how much benefits the users of a social network get if they consider the welfare of their neighbors in the network instead of acting purely selfish. This notion is referred to as the windfall of friendship. They look at equilibria for the problem and also show that social welfare does not increase in relation with the extent with which players care for each other. But for few special graph the notion of friendship helps to remove some undesired equilibria. They show that the computation of best and worst *Friendship Nash Equilibrium* is NP-hard in this setting.

Daz, Mitsche, Rustagi and Saia [10] show interest to achieve windfall of malice in the absence of malicious users. They prove with the help of a mediator it is possible to achieve windfall of malice even without malicious users. For certain special classes of games they show that even with a mediator we cannot achieve windfall of malice. They propose new methods for designing mediators which can be used to improve social welfare. They also show the limits of their techniques by proving an impossibility result: for a large class of games, no mediator can improve the social welfare over the best Nash Equilibria.

The basic model of virus inoculation game assumes users to be selfish. But this assumption is found to be violated frequently in public goods and other experiments where participants act somewhat altruistically or spitefully. Chen, David and Kempe [7] consider the notion of altruism while addressing virus inoculation strategies. The authors improve the approximation guarantee for finding the best vaccination strategy existing in the literature, by rounding a natural linear program with region-growing techniques. They analyze the impact of autonomy on the part of the nodes. They also analyze the “Price of Opting Out” under partially altruistic behavior. They show that with positive altruism, Nash Equilibria may not exist, but the Price of Opting Out is at most $\frac{1}{\beta}$ (whereas the Price of Anarchy can remain at $\Theta(n)$).

The models those have been studied so far assume that once an individual is infected all its adjacent individuals would be infected with probability 1. As the assumption is not realistic, Abbasi and Hoda [2], describe a model where the neighbors of an infected node can be infected with some probability which depends upon contact between the users or the type of the disease. This problem is more challenging as the problem is NP-hard in earlier stricter models. Here the authors design several heuristic algorithms to handle it. Their algorithms have a common basic structure: for all of them, they define a measure (M) by means of which they calculate the vaccination priority of the nodes. Precisely, they iteratively calculate M for all remaining nodes of the input network and vaccinate the node whose M is extremum, until sufficient number of nodes get vaccinated. They call this template as Iterative Candidate Selection Algorithm (ICSA). They have used four techniques for this measure: high degree, high betweenness, greedy and local search.

3.1.2 Influence Maximization Problem

The algorithmic problem of influence maximization is very similar to the virus inoculation game problem. The problem of influence maximization in social networks is first proposed by Domingos and Richardson [8], [9] they follow the same thing. Their model considers the customer network value and how the influence of a customer can cascade sales of a product. Rather than viewing the market as a set of independent entities they

view the market as a social network.

Kempe, Kleinberg and Tardos [12] prove that for simple models of information diffusion in social network the problem of influence maximization is NP-hard. They show that for complex models of information diffusion it is not possible to have even constant factor approximation. But the good news is that if the influence function is submodular then one can give constant factor approximation algorithm. They use independent cascade model and linear threshold model under which the influence function is submodular. They propose a greedy algorithm which uses the assumption of having known influence probability to the algorithm.

In recent years, many people have proposed some algorithms to solve the influence maximization in social networks. Ramasuri and Narahari [18] propose a new way of solving the influence maximization problem using the concept of cooperative game theoretic tool named as Shapley value. They use the SPIN (ShaPley value based Influential Nodes) algorithm for solving the top-k nodes problem and the λ -coverage problem. The submodular property of the influence function is not used by their algorithm and hence is a more general approach. They show that the Shapley value based approach outperforms some of the existing algorithms in the literature.

3.1.3 Limitation of Misinformation Problem

Budak, Agrawal and Abbadi [4] study the problem of limiting misinformation for the first time in the literature. They use competing campaigns which may be positive and negative campaign in a social network and address the problem of influence limitation. In their setting a set of adversary nodes which start negative campaign and try to influence the users with misinformation. They aim to find a set of influential nodes in the network, which can be used as the seeds for the positive campaign which limits the effect of misinformation with some positive influence to the users. They show that the influence limitation problem is NP-hard and also show that influence limitation function in its most general setting is not submodular. They propose restricted versions of the problem and show that these restricted versions satisfy the sub-modular property and

hence provide good approximation guarantees.

Premmraj and Narahari [21] consider more generalized version of the problem described in [4]. In the generalized setting, as the problem is still NP-hard they propose a Shapley value based heuristic to solve this problem. In the generalized setting of influence limitation problem the underlying influence function is non submodular. But their approach works equally effectively for both the submodular and the non-submodular cases. They do experiments on standard real world social network datasets and by extensive experimental results they show that this approach outperforms the existing heuristics in [4]. They also addresses the the problem of λ -coverage for limiting of misinformation. They go beyond the negative and positive campaign and discuss the problem of influence limitation when multiple competing campaigns propagate simultaneously in a social network. In real world scenario there can be multiple campaign who use to diffuse information in the social network. The authors show interest in finding influential nodes among all competing campaigns and propose a extension of the previous Shapley value based algorithm to solve the problem.

3.2 Contributions

Current State of Art and Its Limitations

So far we have seen that the virus inoculation problem under probabilistic model [2] is NP-hard as well as not submodular. So, neither polynomial time algorithm nor any constant factor approximation algorithm can be proposed. Hence we propose heuristic which is agnostic towards submodularity. As far as we know, no work considers approach based on cooperative game theory for this problem. We show that this approach works well as compared to conventional centralities. Moreover, for virus inoculation problem, every method in literature considers strategies only in preventive setting. We consider virus inoculation problem in reactive setting and propose algorithms to solve this problem. Also, existing literature shows that limitation of misinformation problem is NP-hard and not submodular under Multi-Campaign Independent Cascade Model [4]. So, again,

neither polynomial time algorithm nor any constant factor approximation algorithm can be proposed. We propose heuristic which outperforms the existing ones.

In this work, we consider two kinds of influence limitation problems: virus inoculation and limitation of misinformation. In the first problem, no positive influence is used unlike in the second problem. In virus inoculation problem, we consider two settings. In the first setting, we work with the probabilistic model proposed in [2].

Virus Inoculation Problem under Preventive Strategies

We show for virus inoculation problem, game theoretic centralities are better than conventional centrality measures. The key idea is that we have to calculate a node's ability to protect the network from virus infection when it is combined together with other nodes. Here our assumption is that each node cooperates with the other nodes. We use cooperative game theory for proposing game theoretic centralities.

Game theoretic network centrality has been proposed in [11, 18] as a tool to solve similar problems. Moreover in [1], the authors calculate different centrality measures based on Shapley Value (SV) with different characteristic functions. We use SV based centrality as in those papers. The idea is to define a cooperative game on the given network assuming the nodes cooperate. The SV of a node is considered as the centrality of that node as SV calculates its average marginal contribution to all possible coalitions of agents.

As the problem is NP-hard, we propose a heuristic which uses the concept of SV from cooperative game theory. This algorithm outperforms all the algorithms existing in the literature. We call it SV-based heuristic for Probabilistic Model (SVPM). One cannot get a provable constant-factor approximation algorithm using a simple greedy algorithm as the objective function of this problem is neither submodular nor supermodular [21]. We prove that the underlying objective function is neither submodular nor supermodular using simple counter examples in Section 4.

Virus Inoculation Problem under Reactive Strategies

In the second part, we propose virus inoculation problem in reactive setting. All these approaches described in Section ?? only consider preventive strategies to solve the problem. We propose a new model which captures a more realistic situation than the models described above, including the model in [2], for the virus inoculation problem. We call it General Cascade Model for Virus Inoculation (GCM). All the approaches to solve the virus inoculation problem are based only on preventive methods, which means every approach solves the problem under the assumption that no infection has occurred before implementation of the vaccination decisions. Here, we propose a model which is able to capture reactive strategies, which can be taken after the infection has started in the graph. We also prove that the underlying function is not submodular. As we cannot get a good approximation, we propose a heuristic algorithm which is based on SV. We propose another algorithm based on finding trees in the graph. At last, we propose modified version of GCM (MGCM) and show that virus inoculation problem under this model is NP-hard.

Limitation of Misinformation

In the third part, we work with the same model as proposed in [21]. We use same SV based approach as used by the authors. In using SV based method, the important part is to choose characteristic function. We propose a novel characteristic function in our algorithm and show that our algorithm outperforms the existing algorithm in [21].

3.3 Experimental Setup and Organization of this manual

In this section, we describe the experimental setup for all the experiments. We consider different models for our experiments. For Section 4.3 and 4.5, we consider the probabilistic model in [2]. For Section 5.4 and 5.6, we consider GCM and MGCM as model of

propagating the infection respectively. In Section 6.3 we consider Multi-Campaign Independent Cascade Model (MCICM) [4] as the mode of positive and negative campaign. We use several real world social network dataset for our experiments. A summary of all the data sets described below is given in Table 3.1. The data sets are as follows:

Zacharys Karate club data set

This is a network of friendships between the 34 members of a karate club at a US university, as described by Wayne Zachary in 1977. This has 34 nodes.

American College Football data set

This is a network of American football games between division IA colleges during regular season fall 2000. There are 115 nodes in this network.

Jazz data set

This is a social network of jazz musicians. There are 198 nodes in this network.

C. elegans Neural network data set

This data set describes the neural network of the worm *Caenorhabditis elegans* (C. elegans). There are 297 nodes in this dataset.

Erdos971 data set

This is a collaboration network of Erdos in 1971. This has 472 nodes.

Roget data set

This has 1022 nodes.

Smart Grid data set

This has 1059 nodes.

Data Set	Number of nodes	Number of edges
Karate	34	78
Football	115	615
Jazz	198	5485
C. elegans-neural	297	2359
Erdos971	472	1314
Roget	1022	2748
SmartGrid	1059	4922
Netscience	1589	2742
Power	4941	6549

Table 3.1: Data Sets used in the experiments

Netscience data set

This is a co-authorship network of scientists working on network theory. There are 1589 nodes in this dataset.

Power grid data set

This is a network representing the topology of the western states power grid of US. There are 4941 nodes in this dataset.

All the experiments are executed on cluster of 8 nodes with (i) 4 cores and (ii) 64-bit Linux operating system. On each data set, every experimental result is taken as the average of 1000 – 10000 repetitions depending on the type of the data set. All the experiments are carried out using MATLAB. We use the following convention while plotting the performance curves for various algorithms: the X-axis represents the size of budget which is initial target set, and Y-axis represents the percentage of nodes with their respective states.

The outline of the rest of the report is as follows: we describe the models and problem formulations separately in Chapters 4 and 5 for virus inoculation problem. Our

proposed algorithms are also described in these chapters. We study the limitation of misinformation in Chapter 6 and give a new heuristic. We show the experimental results in individual chapter. We conclude with summary and future work in Chapter 7.

Chapter 4

Virus Inoculation in Preventive Setting

This section is organized as follows: we describe the probabilistic model in Section 4.1. We compare the algorithms regarding game theoretic approach and conventional centrality in Section 4.2. We give experimental results of the performances of these algorithms in Section 4.3. Our proposed algorithm, SVRM is described in Section 4.4. We show its performance in Section 4.5.

4.1 Model

In this section, we work with the model proposed in [2]. The model is as follows: the social network is represented by a graph $G = (V, E)$, where all the users are denoted by nodes. The nodes have edges between them whenever the probability of getting infected of one node given that the other is already infected, is non-zero. For example, if two friends are in same room and one gets some infection then there is a chance of the other person in the room getting infected; so as per our model there will be an edge between two friends who are roommates. The graph is directed because the probabilities are not symmetric. In the example, the probability of infection may vary depending upon cautiousness of the two roommates.

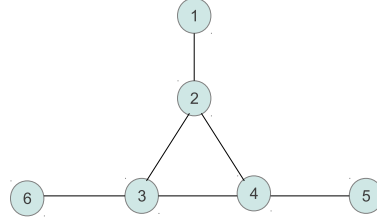


Figure 4.1: This figure shows how $g(T)$ does not satisfy supermodular and submodular property

In this graph, initially node i can get infected with certain probability q_i . For all $i, j \in V$, e_{ij} is the edge from node i to node j , p_{ij} is the probability of node j getting infected, given that node i is already infected. The problem is the same as in [2]: given probabilities q_i , p_{ij} and k , we have to choose k nodes which are to be vaccinated in order to minimize the total expected number of infected nodes in the given graph. Let $f(T)$ be the expected number of nodes that get infected after set T is inoculated, $f_s(T)$ be the the number of nodes that get infected by the propagation of virus according to the random process with initial infected set S . So, it can be observed that $f(T) = \sum_{S \subseteq V(G)} q(S) f_s(T)$ where $q(S)$ is the probability of set S getting infected initially and so, $q(S) = \prod_{i \in S} q_i \prod_{i \notin S} (1 - q_i)$.

LEMMA 1. $f(T)$ is neither submodular nor supermodular.

Proof. We prove the lemma by taking a new function $g(T)$ where $g(T) = n - f(T)$, n being the total number of nodes in the graph $G = (V, E)$. In this case $g(T)$ is the expected number of nodes which will not be infected in the process. The function $f(\cdot)$ is called sub-modular if it satisfies $f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T)$ for every node $v \in V(G)$ and for every subset S, T of $V(G)$ where $S \subseteq T$. In definition of supermodularity the inequality is reversed. Now substituting the value of $f(T)$ in terms of $g(T)$, it can be seen that $g(T)$ is submodular if and only if $f(T)$ is supermodular and $g(T)$ is supermodular if and only if $f(T)$ is submodular.

By simple counter example we prove that $g(T)$ is not supermodular and that will

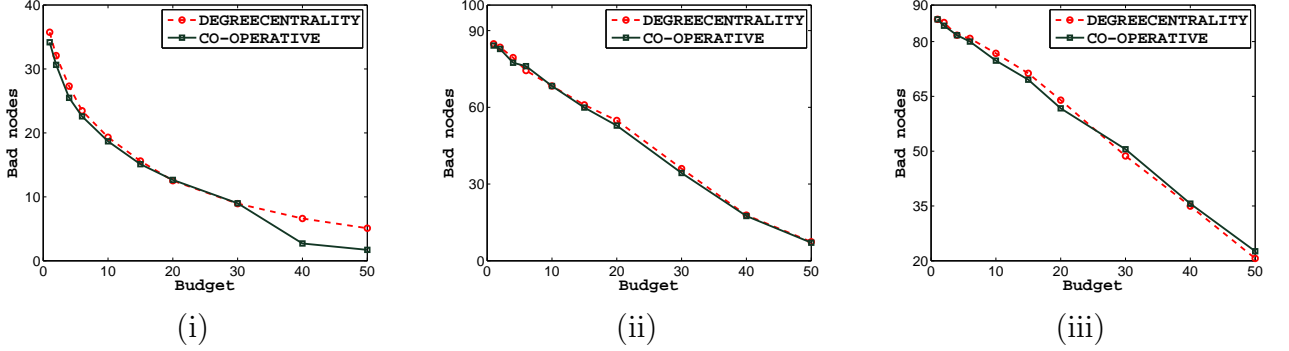


Figure 4.2: Bad nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Netscience co-authorship data set, (ii) Celegans data set, (iii) Football data set

prove $f(T)$ is not sub-modular. Similarly, we show $g(T)$ is not submodular and that will prove $f(T)$ is not supermodular.

Let us assume $q_i = q$ for every node i in V and $p = 1 - q$, we also assume that the graph in Figure 1 is our live graph for this random process. A live graph is a subgraph of the original graph G obtained after sampling based on edge probabilities. We take an undirected graph because we assume that $p_{ij} = p_{ji}$ for all i, j in V . Now take $T_2 = \{2\}$, $T_1 = \phi$ and v to be node 4. To calculate $g(T_2)$, remove 2 from the graph. It will result into two components; one consists of one node and the other of four nodes. First one will be saved by probability p and the other one will be saved with probability p^4 . By definition of $g(\cdot)$, $g(T_2)$ is $1p + 4p^4$. So, $g(T_2 \cup \{v\}) = 2p + 2p^2$, $g(T_2) = p + 4p^4$, $g(T_1 \cup \{v\}) = p + 4p^4$, $g(T_1) = 6p^6$. Now it is clear that if we take $p = 0.9$ then $g(T_1 \cup \{v\}) - g(T_1) \geq g(T_2 \cup \{v\}) - g(T_2)$. So we can conclude g is not supermodular and so f is not submodular. But if we take $p = 0.1$ then $g(T_2 \cup \{v\}) - g(T_2) \geq g(T_1 \cup \{v\}) - g(T_1)$. So we can conclude g is not submodular and so f is not supermodular.

□

4.2 Degree Centrality vs Game Theoretic Approach

The basic model of virus inoculation game assumes users to be selfish. But this assumption is found to be violated frequently in real life situations where participants act

Symbol	Meaning
$G = (V, E)$	Social network represented by a graph where V is the set of vertices and E is set of edges.
T	Set to be vaccinated,i.e. infection cannot propagate through the nodes in the set
$f(T)$	Expected number of infected nodes with vaccinated set T and $ T = k$
k	$ T $
$g(T)$	Expected number of saved nodes with vaccinated set T
q_i	Initial probability of node i getting infected
p_{ij}	The probability of node j getting infected, given that node i is already infected
$f_s(T)$	Expected number of infected nodes with initial infected set S and vaccinated set T
$g_1(N, \nu_1)$	A cooperative game with N as set of players and $\nu_1(.)$ as the characteristic function.
$\phi_{g_1}(v_i)$	SV of node v_i in game g_1
$g_2(N, \nu_2)$	Another co-operative game with N as set of players and $\nu_2(.)$ as the characteristic function
n	Number of nodes in the graph G , i.e., total number of players or nodes
m	Number of edges in the graph G

somewhat altruistically or spitefully. Chen, David and Kempe [7] introduce the notion of altruism while addressing virus inoculation strategies and improve the approximation guarantee for finding the best vaccination strategy. We use similar concepts here to solve the problem on probabilistic setting which has been modeled by Abbasi and Hoda [2].

We use game theoretic based centralities to solve the virus inoculation problem in this model. At first, we compare two algorithms: one is based on degree centrality which is a conventional network centrality and other is based on game theoretic centrality which uses SV with characteristic function based on degree of nodes. We show that for one live graph we get an exact closed-form expression of SV of each node.

We consider one live graph $G'(V, E)$, which is directed and unweighted. To deduce closed-form expression of SV of each node, we take similar approach as in [1]. We first define “victim” of a coalition of players(nodes). Here, “victim” of a subset $C \subseteq V(G')$ is the set $\{v \in V(G') : v \in C \text{ or } \exists u \in C \text{ such that } (u, v) \in E(G')\}$, i.e. the set of those nodes which are reachable at one hop distance from the nodes in the coalition C . Based on this set “victim”, we define a co-operative game $g_1(N, \nu_1)$ on this network $G'(V, E)$ where $N = V(G')$ is set of players and $\nu_1 : 2^{V(G')} \rightarrow \mathbb{R}$ the characteristic function defined as

$$\nu_1(C) = \begin{cases} 0 & \text{if } C = \emptyset \\ \text{size}(\text{victim}(C)) & \text{else} \end{cases} \quad (4.1)$$

This kind of game on a network is already investigated in [18]. Here we show that SV of each node can be used as a better centrality measure than degree centrality for virus inoculation problem. To evaluate SV of a particular node v_i , we have to calculate the marginal contribution of that node to the coalition S_i which occurs before this node in a random permutation of nodes. Let $N_{G'_o}(v_i)$ be the set $\{u \in V(G') : (v_i, u) \in E(G')\}$, $N_{G'_i}(v_i)$ as the set $\{u \in V(G') : (u, v_i) \in E(G')\}$ and $\deg_{G'_o} = |N_{G'_o}|$ and $\deg_{G'_i} = |N_{G'_i}|$.

It is to be visualized that a node v_i marginally contributes a node $w_i \in N_{G'_o}(v_i)$ to the coalition $\text{victim}(S_i)$ if and only if neither w_i nor any node from $N_{G'_i}(w_i)$ is present in S_i . For computing the SV, if the permutations are chosen uniformly at random, by

combinatorial arguments, one can show that the above condition is satisfied with probability $\frac{1}{1+deg_{G'_i}(w_i)}$. If we denote B_{v_i, w_i} the Bernoulli random variable that v_i marginally contributes w_i to $victim(S_i)$. Thus:

$$E[B_{v_i, w_i}] = Pr[(N_{G'_i}(w_i) \cup \{w_i\}) \cap S_i = \emptyset] = \frac{1}{1 + deg_{G'_i}(w_i)} \quad (4.2)$$

So, the SV $\phi_{g1}(v_i)$, which is the expected marginal contribution of v_i , is given by

$$\phi_{g1}(v_i) = \sum_{w_i \in \{v_i\} \cup N_{G'_o}(v_i)} E[B_{v_i, w_i}] \quad (4.3)$$

$$\sum_{w_i \in \{v_i\} \cup N_{G'_o}(v_i)} E[B_{v_i, w_i}] = \sum_{w_i \in \{v_i\} \cup N_{G'_o}(v_i)} \frac{1}{1 + deg_{G'_i}(w_i)} \quad (4.4)$$

The above form is clearly an exact closed-form expression for computing the SV of each node in the network for one live graph. Now, in the virus inoculation problem, the given graph has probabilities on its edges. So, to compute degree centrality or to compute SV with this closed form described above, one can simply add the probabilities on the edges or sampling can be done from a certain number of simulations. In Monte-Carlo simulations technique, for each round, both the algorithms take $O(|V| + |E|)$ operations. Algorithm 1 describes the procedure to compute SV of each node and to choose those nodes, which have higher centrality, to be inoculated.

We have already argued the running time of calculation for SV is $O(|V| + |E|)$. For sorting it takes $O(|V| \log |V|)$. So total running time is $O(R(|V| + |E|) + |V| \log |V|)$. We show in Section 4.3 that centrality based on the above formula works better than conventional degree centrality. Some intuition can be derived from the above formula. It can be seen that an arbitrary node v has more SV based centrality when it has more out-degree, i.e., $N_{G_o}(v)$ is more, and if its neighbours in $N_{G_o}(v)$ have lower in-degree. So removal of such a node by vaccination from the graph decreases the probability of spreading the infection as it disconnects the weak nodes in terms of in-degree.

Algorithm 1 Calculating the vaccination set based on SVs for $g_1(N, \nu_1)$

Require: Social network graph $G(V, E)$, the probabilities q_i, p_{ij} , Number of simulations or rounds R , Number of nodes k which has to be vaccinated

```

1: if  $k < n$  then
2:   for  $i = 1$  to  $R$  do
3:      $G' =$  Live graph for the random process of infection
4:     for  $v \in V(G')$  do
5:        $\text{ShapleyValue}[v] = \frac{1}{1 + \deg_{G'_i}(v)}$ 
6:       for  $w \in N_{G'_o}$  do
7:          $\text{ShapleyValue}[v] + = \frac{1}{1 + \deg_{G'_i}(w)}$ 
8:       end for
9:     end for
10:  end for
11:   $\text{ShapleyValue}[v] = \frac{\text{ShapleyValue}[v]}{R}$ 
12:  Sort the array ShapleyValue and Choose the top  $k$  nodes with highest SVs from
    ShapleyValue
13: end if
14: return Vaccination set of size  $k$ 

```

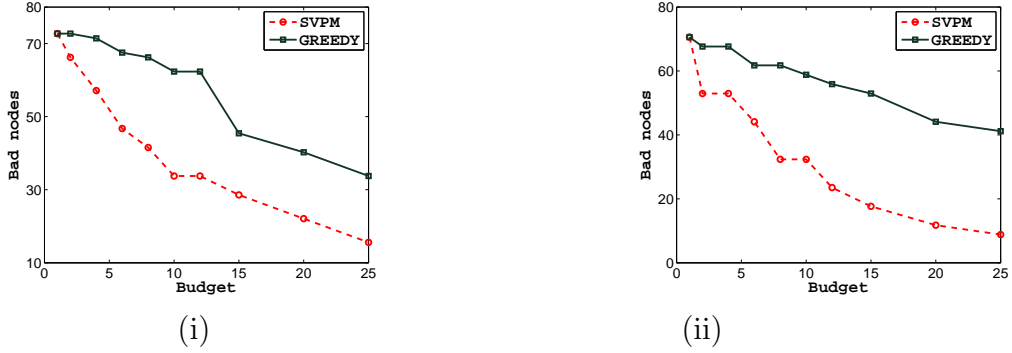


Figure 4.3: Bad nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Football data set, (ii) Celegans data set

4.3 Experimental Results and Analysis

Figure 4.2 shows the performance of Algorithm 1 vis-a-vis the algorithm based on conventional degree centrality on real world social network data sets consisting of various number of edges. In particular, Figure 4.2(i) corresponds to Netscience co-authorship data set, Figure 4.2(ii) corresponds to Celegans data set and Figure 4.2(iii) corresponds to Football data set. It can be observed that the efficacy of Algorithm 1 is better than that of the algorithm based on conventional degree centrality for most of the cases.

4.4 Greedy Algorithm vs Game Theoretic Approach

Abassi and Hoda [2], the authors propose four heuristics for virus inoculation as the problem is NP-hard. They show by experiments that the greedy algorithm is the best algorithm to be used. In Algorithm 2, we present the greedy algorithm described in [2].

The algorithm iteratively computes strength of each node to spread the infection. Then it chooses the node inoculation to minimize the infection. In step 4, the algorithm takes $O(R(n + m))$ time where n represents number of nodes, m represents number of edges to compute $f(\cdot)$ (Table 4.1) via live graphs. As the total running time is concerned, the greedy algorithm takes $O(kRn(n + m))$ time because of two loops in step 2 and 3.

Algorithm 2 Greedy Algorithm

Require: Social network graph $G(V, E)$, the probabilities q_i, p_{ij} , Number of simulations or rounds R , Number of nodes k which has to be vaccinated

```

1: Initialize  $T = \phi$ 
2: for  $i = 1$  to  $k$  do
3:   for  $v \in V(G)$  and  $v \notin T$  do
4:     compute  $f(T \cup \{v\})$  by  $R$  number of rounds
5:   end for
6:   Let  $u$  be the node with minimum  $f(T \cup \{u\})$ 
7:    $T = T \cup \{u\}$ 
8:    $V(G) = V(G) - \{u\}$ 
9: end for
10: return Vaccination set  $T$  of size  $k$ 

```

Myerson value and SV based approach

Now we propose another heuristic based on some game theoretic tools. We take an approach which is similar to the greedy algorithm but here our main assumption is that nodes or players are rational and they cooperate with each other. Our concern is about strength of a node in saving other nodes of the graph when the node is combined with other nodes. To use SV choosing the characteristic function is important. Based on that function, SV of each node represents the node's marginal ability contributed to a coalition. Myerson value [16] takes co-graph into account. Co-graph describes the cooperation structure amongst the players. If we assume every player can cooperate with any other player, then the co-graph with all the players will be a complete graph. In that case, Myerson value and SV are same. Now, to save one network, the practical assumption is to use a complete co-graph as we take preventive strategies here. So, to save the network, each node in the network cooperates with every other node to save itself from infection. Concluding, we use Myerson value assuming a complete co-graph structure, which turns out to be same as using SV here.

Whenever the underlying function is non-submodular, the SV approach works well [21], [18]. In the context of infection limitation, SV approach is quite natural because this concept of cooperative game theory captures the marginal contribution which a user brings into the system. For limiting the number of infected user, the number of additional nodes saved by a user is his marginal contribution to the coalition.

Given a social network as a graph $G(V, E)$, we define the following cooperative game $g_2(N, \nu_2)$: Let N be the set of players, here $N = V(G)$ and $n = |N|$. We define the characteristic function $\nu_2 : 2^N \rightarrow \mathbb{R}$ used for computing the SVs of nodes as follows:

$$\nu_2(C) = n - f(C) \quad \forall C \subseteq N \quad (4.5)$$

i.e., given a set of users C , our characteristic function $\nu_2(C)$ is same as $g(C)$ which is the difference between n and the expected number of nodes affected by the infection. Clearly $\nu_2(\phi) = 0$ as $g(\phi) = 0$. Given the graph and corresponding probabilities, we are interested in finding T , the set of users where $|T| = k$, such that after vaccinating this set of users, the number of infected node in the whole graph, under this probabilistic model, is minimum.

According to the SV based heuristic (SVPM) described in Algorithm 3, we are given the social network graph, the probabilities q_i, p_{ij} for all nodes i, j in V , the number of simulation for which the heuristic is to be run and the number of nodes to be vaccinated, k . In each round, we compute a live graph for the random process. Since computation of SV requires all possible permutations making it computationally hard, we compute approximate SV by a sampling based approach. Consider a set \mathbf{P} which contains random permutations; we choose $q = |\mathbf{P}|$ such that $q = O(n)$. Using the permutations in \mathbf{P} , we compute approximate SV of each node. In this algorithm:

- $MC[v]$ represents the marginal contribution of node v .
- In permutation π_j , $B_v(\pi_j)$ represents the set of node occurring before node v .
- $ShapleyValue[v]$ represents SV of node v .

Once the number of simulations are completed, we can choose k nodes with the highest SVs as the top k nodes to minimize the infection in the given social network.

Computational Complexity

Here we generate the live graph R times and compute $\nu_2(C)$ which can be seen as $g(C)$. Computing this characteristic function value takes $O(n + m)$, as it requires traversing over all the nodes and edges of a live graph to compute number of nodes in the connected components. We compute these values for q number of permutations. In the last part of the algorithm, we sort the array which take $O(n \log(n))$ time. Thus, the overall running time is $O(qR(n + m) + n \log(n))$. It is to be noted that q is polynomial in n . The greedy algorithm also takes similar time to execute. We show in Section 4.5 that SVPM(Algorithm 3) works better than Greedy algorithm(Algorithm 2).

An Example

Let us assume $q_i = q = 0.5$ for all nodes i in V , $p = 1 - q$ and consider $k = 1$, for a particular graph. We also assume that the graph in Figure 4 is our live graph for this random process in one round of a simulation. We consider an undirected graph because of the assumption that $p_{ij} = p_{ji} = 0.5$ for all i, j in V . The live graph has been constructed based on these probabilities. Now by computing the SVs for each node we get $\phi_1 = 3 - (1 + 1 + 0.5)$, $\phi_2 = 3 - (0.5 + 0.5 + 0.5)$, $\phi_3 = 3 - (1 + 1 + 0.5)$ for nodes 1, 2 and 3 respectively. So node 2 is the resulting node to be vaccinated by the algorithm. A careful look at the graph also tells that node 2 is the best node to be vaccinated.



Figure 4.4: This figure shows how to compute the best node by SV

Algorithm 3 SV based heuristic for Probabilistic Model (SVPM)

Require: Social network graph $G(V, E)$, the probabilities q_i, p_{ij} , Number of simulations or rounds R , Number of nodes k which have to be vaccinated

```

1: Let  $\pi_j$  be the  $j^{th}$  permutation in  $\mathbf{P}$ 
2: for  $v= 1$  to  $n$  do
3:    $MC[v] \leftarrow 0$ 
4: end for
5: for  $j= 1$  to  $q$  do
6:   for  $i= 1$  to  $n$  do
7:      $arr[i] \leftarrow 0$ 
8:   end for
9:   for  $r= 1$  to  $R$  do
10:    Generate a live graph based on the random edge probabilities
11:    for  $v= 1$  to  $n$  do
12:       $arr[v] \leftarrow arr[v] + \nu_2(B_v(\pi_j) \cup \{v\}) - \nu_2(B_v(\pi_j))$ 
13:    end for
14:  end for
15:  for  $v= 1$  to  $n$  do
16:     $MC[v] \leftarrow arr[v]/R$ 
17:  end for
18: end for
19: for  $v= 1$  to  $n$  do
20:    $ShapleyValue[v] \leftarrow \frac{MC[v]}{q}$ 
21: end for
22: Sort the array ShapleyValue and Choose the top  $k$  nodes with highest SVs from
    ShapleyValue
23: return Vaccination set of size  $k$ 

```

Data Set	Number of nodes (n)	q
Karate	34	100
Football	115	300
Jazz	198	200
C. elegans-neural	297	300
Erdos971	472	200
Roget	1022	150
SmartGrid	1059	150
Netscience	1589	150
Power	4941	200

Table 4.2: Data Sets with q used in the experiments

4.5 Experimental Results and Analysis

Figure 4.3 shows the performance of SVPM vis-a-vis the greedy algorithm (Algorithm 2) on real world social network data sets consisting of various number of edges. In particular, Figure 4.3(i) corresponds to Football data set, Figure 4.3(ii) corresponds to Celegans data set and Figure 4.3(iii) corresponds to Football data set. It is clear that the efficacy of SVPM is good as that of the greedy algorithm.

Recall that SVPM algorithm works with q number of permutations in the sampled set \mathbf{P} . So, the running time depends on value of q . Table 4.2 shows the values of q for various data sets.

Chapter 5

Virus Inoculation in Reactive Setting

This section is organized as follows: we describe the issues regarding reactive setting in Section 5.1. We formulate the virus inoculation problem under General Cascade Model (GCM) in Section 5.2. We propose two algorithms in Section 5.3. We give experimental results of the performances of these algorithms in Section 5.4. We propose another model, Modified General Cascade Model (MGCM) in 5.5. We show the performances of the same algorithms under MGCM in Section 5.6.

5.1 Issues

It is clear from the relevant work in Section ?? that all approaches to solve the virus inoculation problem are based only on preventive methods, that is, every approach solves the problem under the assumption that no infection has occurred before the implementation of vaccination decisions. These types of strategies are called preventive strategies or methods. To the best of our knowledge, no paper uses any of the reactive methods to solve the vaccination problem. By reactive method or strategies, we mean that any strategy that has been taken after the start of infection by a virus in a graph. Here, we propose a model which is able to capture the reactive strategies in the context of virus

inoculation problem. We describe the model in the following subsection.

5.2 Problem Formulation

We formally define the model along with the optimization problem for reactive setting that we address in this report.

Model

We define our model as follows: a social network can be represented by a directed graph $G = (V, E)$ consisting of nodes V and edges E . A node j is a neighbor of node i if there is an edge $e_{i,j} \in E$ from node i to node j in G .

We take motivation from the Independent Cascade Model (ICM) [12] to formulate our model. We call our model General Cascade Model for Virus Inoculation (GCM). In GCM, infection starts from one node and spreads following ICM. After some time the infection is detected and to minimize the infection some nodes are vaccinated, i.e., deleted from the graph. Then, again the infection spreads following ICM in the rest of the graph. Like in the preventive setting, we assign infection probability p_{ij} to every edge $e_{i,j} \in E$. We call it infection influence: if node i is infected, it will try to infect node j with probability p_{ij} . The concept of infection influence is same as the probabilities of edges under probabilistic model described in [2]. The process of spreading infection and inoculation occurs in discrete steps. The infection starts from one node. In the infection process, if a node i gets infected by virus in step t then it has only one chance to infect its unaffected neighbor j with probability p_{ij} in step $t + 1$. If it succeeds then j becomes infected in step $t + 1$; but whether or not i succeeds, it cannot make any further attempts in subsequent steps.

We also refer to p_{ij} as the probability of the edge $e_{i,j}$ being live. If there are two or more nodes trying to activate (infect) j at the same time, at most one of them will be successful. If node j has several newly activated (infected) neighbours, those attempts can be sequenced arbitrarily.

Problem Definition

We define the vaccination of top k nodes problem as follows: the spread of infection by virus starts from one node r_1 and is detected with delay d , at which point of time, vaccination is given to the selected k nodes. Hence, the problem is to find k nodes to be vaccinated. Thus we will remove these nodes from the graph as infection will not propagate through them so as to save the maximum number of nodes from the infection. This strategy of minimizing the infection in a social network is one kind of reactive strategy, that has not been studied yet. We define a function $g : S \rightarrow [0, n]$ where $S = \{W | W \subset V \text{ and } |W| \leq k\}$. Let W be the set of initial vaccinated nodes and $g(W)$ be expected number of nodes that can be saved from infection in the given random process. The goal is to find a set W ($|W| \leq k$) of nodes to vaccinate in order to maximize $g(W)$.

5.3 Reactive Method based Algorithms

Here, we give two algorithms for solving this problem. Both these methods are based on negative approaches, that is, our method is to find which nodes are the most harmful ones. It is desirable to vaccinate them to save the network in worst case situation. In these two algorithms, we use different method to guess the harmful nodes. The first approach is based on finding the local harmful nodes. The approach in the first algorithm is similar to the approach taken in [4] for prediction algorithms. The next is based on finding the global harmful nodes.

Algorithm based on local method

In the model, we assume that the infection starts from one node. The infection in our model, GCM, follows a tree structure as it starts from one node and follow the infection process of ICM. The first algorithm is based on an important question: Which k nodes are most likely to get infected in the process immediately after the infection is detected? We want to vaccinate these nodes to minimize the overall infection in the graph.

Symbol	Meaning
$G = (V, E)$	Social network represented by a graph
p_{ij}	The probability of node j getting infected, given that the node i is already infected
W	Set to be vaccinated after the delay
$g(W)$	Expected number of saved nodes with vaccinated set W in GCM
k	$ W $
$\Psi(W)$	Expected number of saved nodes with vaccinated set W in MGCM
$g_3(N, \nu_3)$	A co-operative game with N as set of players and $\nu_3(.)$ as the characteristic function
$\sigma(C)$	Influence function that gives expected number of infected nodes with initial infected set C
d	Delay between starting of infection and detection of infection
n	Number of nodes in the graph G , i.e., total number of players or nodes
m	Number of edges in the graph G

Table 5.1: Notation Table for Section 5

We are given graph, the set of nodes Δ_{given} which are already infected by the infection process. Our aim is to choose k nodes which are most likely to get activated next. This can be viewed as an optimization problem, finding a set Δ_{next}^* of k nodes that could potentially maximize the possibility of spread of infection.

We construct a graph $G' = (V', E')$ from the original graph $G = (V, E)$ where $V' = \Delta_{given} \cup \Delta_{next}$ and $E' = \{(u, v) | (u, v) \in E \text{ and } u \in V' \text{ and } v \in V'\}$. As the infection process takes tree structure, we are interested in finding the number of spanning trees in G'_1 where G'_1 is a connected component of G' for all possible Δ_{next} which are in the same component as that of Δ_{given} . Now we are in a position to describe the underlying function. We consider a new function $f_t(\cdot)$ that takes Δ_{next} as an argument and computes the chance of set Δ_{next} getting infected next by counting the spanning trees it can produce combined with the strength of each spanning tree. By strength of the spanning tree, we mean the multiplication of the probabilities on the edges and if the edge is live then we consider probability as 1. We define the function as follows:

$$f_t(\Delta_{next}) = \sum_{T \in \Gamma(G'_1)} \prod_{(u,v) \in T} p_{uv} \quad (5.1)$$

where $|\Delta_{next}| = k$, $\Gamma(G'_1)$ is the set of possible spanning trees in G'_1 and $p_{uv} = 1$ if $u, v \in \Delta_{given}$ i.e. u and v are already infected.

So, from the function definition, it can be viewed that nodes, in the set Δ_{next} which maximizes $f_t(\cdot)$, are the most likely nodes which are going to be infected next.

It is to be noted that the number of spanning trees of a graph can be exponential, making it computationally hard. So, we propose a heuristic to solve this optimization problem. One good heuristic should have the following properties: 1) Δ_{given} and Δ_{next} should form a spanning tree, 2) The heuristic should produce as many spanning trees as possible, so that it gets close to the optimal. Algorithm 4 provides the details of the heuristic. The already infected set is given by Δ_{given} . As the next step, we choose the nodes which will induce more number of trees in the structure. So, we choose the nodes that are connected to as many nodes in the set S as possible. Clearly, choosing nodes

by this technique will increase the number of parents in the already infected set of nodes and its chance of getting infected will be more. It is to be noted that this algorithm is based on a local method. Locally we want to vaccinate those nodes which are more likely to get infected. It may or may not be good for the whole graph to minimize the overall infection. This algorithm takes $O(n + m)$ in step 5. The total running time of the algorithm is $O(k(n + m))$. Next, we present an algorithm which is based on global approach to find harmful nodes. We show evaluation of this algorithm in Section 5.4.

Algorithm 4 Tree based heuristic for Reactive Model (TRM)

Require: Social network graph $G(V, E)$, the probabilities p_{ij} , the set Δ_{given} , Number of nodes k which has to be vaccinated

```

1:  $\Delta_{next}^* = \phi$ 
2:  $V' = V - \Delta_{given}$ 
3:  $S = \Delta_{given}$ 
4: while  $|\Delta_{next}^*| \leq k$  do
5:    $n^* = argmax_{n \in V'} \{deg(n)_S\}$ 
6:    $\Delta_{next}^* = \Delta_{next}^* \cup \{n^*\}$ 
7:    $V' = V' - \{n^*\}$ 
8:    $S = S \cup \Delta_{next}^*$ 
9: end while
10: return Vaccination set  $\Delta_{next}^*$  of size  $k$ 

```

Algorithm based on global method

This algorithm is based on global method. We are interested in finding the most powerful nodes in terms of spreading the infection in a network. It is also a negative approach as in Algorithm 4. We know the already infected nodes in the graph, to be precise, the infected tree is known to us as infection starts from one node and spreads following the GCM model. We find the most powerful nodes that could spread the infection in the same component.

We first prove that the function $g : S \rightarrow [0, n]$ (Table 5.1) where $S = \{W | W \subset V \text{ and } |W| \leq k\}$, is not submodular. Then we describe the algorithm to solve the virus inoculation problem under the GCM model. We show in Section 5.4 that these algorithm outperforms both Algorithm 4 and algorithm based on conventional degree centrality.

LEMMA 2. *The function $g : S \rightarrow [0, n]$ where $S = \{W | W \subset V \text{ and } |W| \leq k\}$ is not submodular.*

Proof. Let set $S = \phi$, $T = \{2\}$, and node x be 3 in Figure 5. From the definition of submodularity, if we show that $g(T \cup \{x\}) - g(T) \geq g(S \cup \{x\}) - g(S)$ then $g(\cdot)$ will be non-submodular as $S \subseteq T$. Now, in the live graph of infection influence propagation as in Figure 5, consider the infection probabilities as $p_{12} = 0.5$, $p_{24} = 0.5$, $p_{13} = 0.5$, $p_{34} = 0.5$. In the live graph of vaccination influence propagation, consider the probabilities as $p_{52} = 0.4$, $p_{63} = 0.4$. Consider the live graph generated as in Figure 5(ii). Now, $g(T \cup \{x\}) = 3$, $g(T) = 1$, $g(S \cup \{x\}) = 1$, $g(S) = 0$. So, $g(T \cup \{x\}) - g(T) \geq g(S \cup \{x\}) - g(S)$ and so $g(\cdot)$ is not submodular. \square

We use game theoretic approach to find the desirable k nodes. We are interested to find the ability of a node in infecting the network when it acts together with other nodes. We use SV here, as it captures marginal contribution of a node to a coalition, in spreading infection. As mentioned earlier and also in [21], [18], also even when the underlying function is non-submodular, the SV works well.

We are given a social network as a graph $G(V, E)$. We define the following cooperative game $g_3(N, \nu_3)$ as follows: Let $N = V(G)$ be the set of players and $n = |N|$. We define the characteristic function $\nu_3 : 2^N \rightarrow \mathbb{R}$ for computing the SVs of nodes as follows:

$$\nu_3(C) = \sigma(C) \quad \forall C \subseteq N \quad (5.2)$$

i.e., given a set of users C , our characteristic function $\nu_3(C)$ is same as the expected number of infected nodes if C is the initial infected set. Clearly $\nu_3(\phi) = 0$ as $\sigma(\phi) = 0$.

Given a set of users, we are interested in finding the set of users W , where $|W| = k$, such that after vaccinating this set of users, the number of infected nodes in the whole

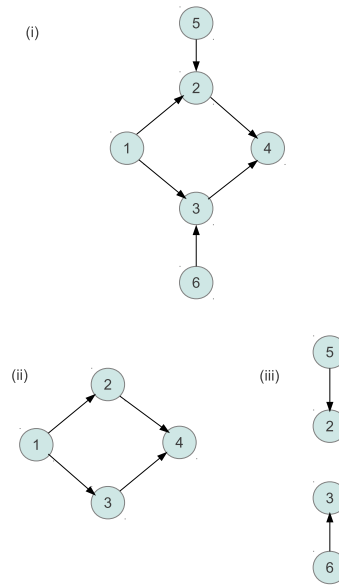


Figure 5.1: (i) To show how $g(\cdot)$ does not satisfy submodular property; (ii) is the live graph for the infection influence propagation; (iii) shows the edges which are absent in live graph due to sampling

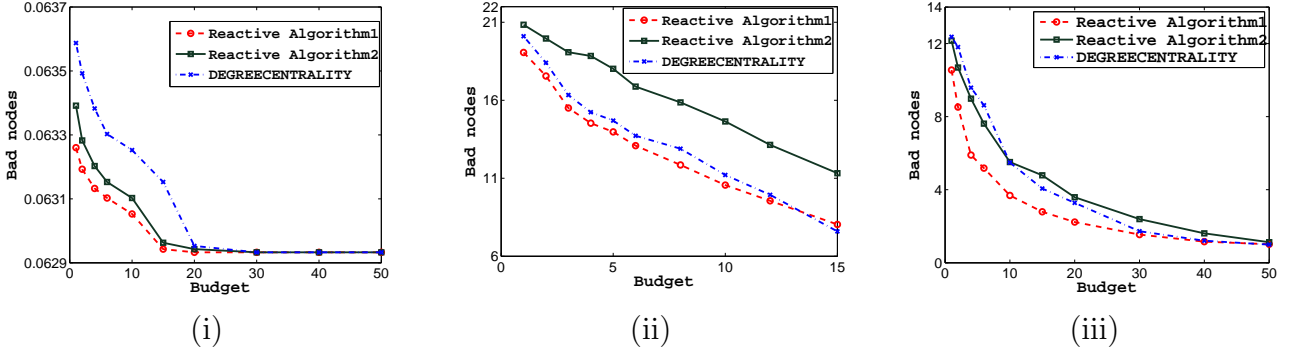


Figure 5.2: Under GCM: Bad nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Netscience co-authorship data set, (ii) Jazz data set, (iii) Football data set

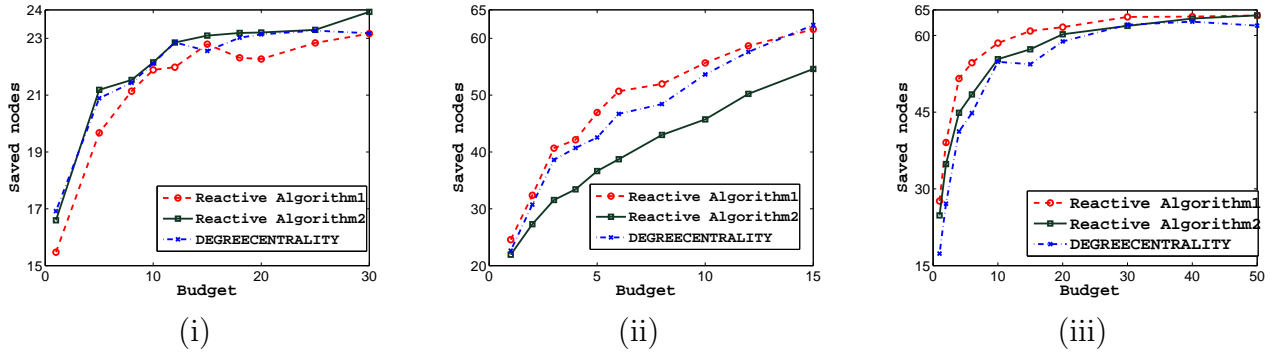


Figure 5.3: Under GCM: Saved nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Smart Grid data set, (ii) Jazz data set, (iii) Football data set under GCM

graph under GCM is minimized.

According to the SV based heuristic (SVRM, SV based heuristic for Reactive Model) described in Algorithm 5, we are given the social network graph, the probabilities p_{ij} for all nodes i, j in V , the number of rounds for which the heuristic is to be run, and the number of nodes, k , to be vaccinated. In each round, we compute a live graph for the infection influence propagation. Next, we construct a set (PB-set, Probably Best set depending on SV) as the set of users in this live graph. Now, for each node in PB-set, we compute its SV using $\nu_3(C)$ as the characteristic function. Once the number of rounds reaches R , we can choose k nodes with the highest SVs as the top k nodes to minimize the infection in the given social network. The time complexity will be same as that of

SVPM.

Algorithm 5 SV based heuristic for Reactive Model (SVRM)

Require: Social network graph $G(V, E)$, the probabilities p_{ij} , Number of simulations or rounds R , Delay d , Number of nodes k to be vaccinated

```

1: if  $k < n$  then
2:   for  $i = 1$  to  $R$  do
3:      $R_1$  = Live graph for the random process of infection influence propagation process
4:     Based on delay  $d$ , from  $R_1$ , remove the nodes which are already infected by the
       above stated random process and thus construct  $G_1$ 
5:     PB-set = The nodes in  $G_1$ 
6:     for each node in PB-set do
7:       Compute the SV with the characteristic function  $\nu_3(.)$  by following same
       steps(1 – 20) as in Algorithm 3.
8:     end for
9:   end for
10:  Choose the top  $k$  nodes with highest SVs to be vaccinated
11: end if

```

There can exist four types of nodes in a social network for the virus inoculation problem:

- Nodes which are not reachable by the infection influence propagation process, and hence are not going to be infected in the whole process; those nodes will not be in the live graph of that process, i.e., they will be absent in R_1
- As per the choice of characteristic function, SV will be less for those nodes which have neighbours with high SV. It is thus consistent with not choosing such a node for practical vaccination.
- Nodes which are definitely going to be infected due to delay associated with the start of vaccination and so they are removed in step 4 of Algorithm 5.

- Nodes which are going to be infected by the infection process but cannot be saved by the vaccination process.

5.4 Experimental Results and Analysis for GCM

In Figures 5.2 and 5.3, we call SVRM (Algorithm 5) as Reactive Algorithm1 and TRM (Algorithm 4) as Reactive Algorithm2. Under GCM, Figure 5.2 and 5.3 show the performance of TRM and SVRM vis-a-vis the algorithm based on conventional degree centrality on real world social network data sets with different densities. In particular, Figure 5.2(i) corresponds to Netscience co-authorship data set, Figure 5.3(ii) corresponds to Smart Grid data set, Figures 5.2(ii) and 5.3(ii) correspond to Jazz data set, Figures 5.2(iii) and 5.3(iii) correspond to Football data set. Jazz data set is the most dense data set. Smart Grid and Netscience co-authorship data set are less dense. It is clear that the efficacy of SVRM is better than that of the algorithm based on conventional degree centrality. But performance of SVRM degrades if density of the graph is less. Because if the graph is more dense the local method, TRM will not be enough to stop the spreading. Then vaccination to the global influential nodes will be better in worst case situation. So, TRM works well when the graph is sparse. Algorithm based on degree centrality works better than TRM on Jazz data set. For Football data set, it performs at par with degree centrality. But for other data sets, TRM works better than algorithm based on degree centrality.

5.5 Another Model: Modified General Cascade Model

Following GCM, we propose another model with some more assumptions to capture reactive strategies. We call this model Modified General Cascade Model for Virus Inoculation (MGCM), which is almost same as GCM model. Here also social network is represented by a graph and all the edges in the graph have infection probability, i.e. infection influence. The only additional assumption we make is that, if we choose any

node to be inoculated then this node will save its neighbours with probability 1 if the neighbours are not already affected by the infection. In GCM, such influences on neighbours from inoculated nodes are absent. Now the problem is to find the suitable k nodes to be vaccinated. Thus we will remove these nodes and their uninfected neighbours from graph as infection will not propagate through them, and then get the maximum number of nodes saved from the infection. We define a function $\Psi : S \rightarrow [0, n]$, where $S = \{W | W \subset V \text{ and } |W| \leq k\}$. Let W be the set of initial vaccinated nodes and $\Psi(W)$ be expected number of nodes that can be saved from infection in the given random process. The goal is to find a set W ($|W| \leq k$) of nodes to be vaccinated in order to maximize $\Psi(W)$. We show that virus inoculation problem under MGCM is NP-hard. We show the experimental results under MGCM in Section 5.6.

LEMMA 3. *Virus inoculation problem under MGCM is NP-hard*

Proof. Consider an instance of the NP-complete Set Cover problem, defined by a collection of subsets S_1, S_2, \dots, S_m for a universe set $U = \{u_1, u_2, \dots, u_n\}$; the problem is to decide whether there exists k subsets whose union is equal to U . We show that this problem instance can be viewed as a special case of virus inoculation problem under MGCM. Given an arbitrary instance of the Set Cover problem, we define a corresponding directed bipartite graph with $n + m + 1$ nodes: there is a node i corresponding to each set S_i , a node j corresponding to each element u_j , and a directed edge (i, j) whenever $u_j \in S_i$. We add one more adversary node n_a and a directed edge (n_a, j) for all u_j with infection probability $p_{r, n_a, j} = 1$. The Set Cover problem is equivalent to deciding if there is a set W of k nodes in this graph with $\Psi(k) \geq n + k$ when we detect the infection at time step 0 (when n_a itself is infected but has not contacted any of its neighbors yet). As per our definition for the instance, infection is a deterministic process, as all probabilities for adversary node to infect its neighbors are 0 or 1. Initially choosing the k nodes corresponding to sets in a Set Cover solution results in saving all n nodes corresponding to the ground set U as per the special assumption in MGCM. The Set Cover problem is solvable if any set W of k nodes has $\Psi(k) \geq n + k$. \square

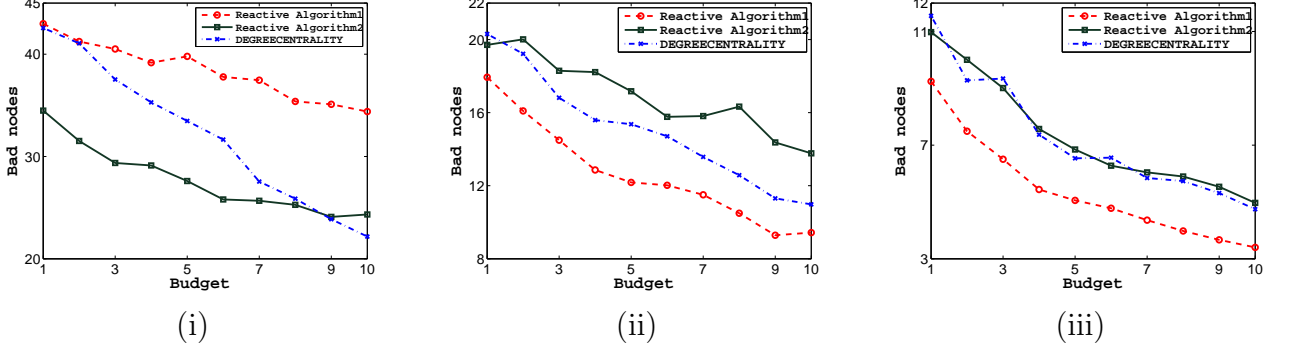


Figure 5.4: Under MGCM: Bad nodes (percentage of nodes) versus budget (percentage of nodes) using (i) *Celegans* data set, (ii) *Jazz* data set, (iii) *Football* data set

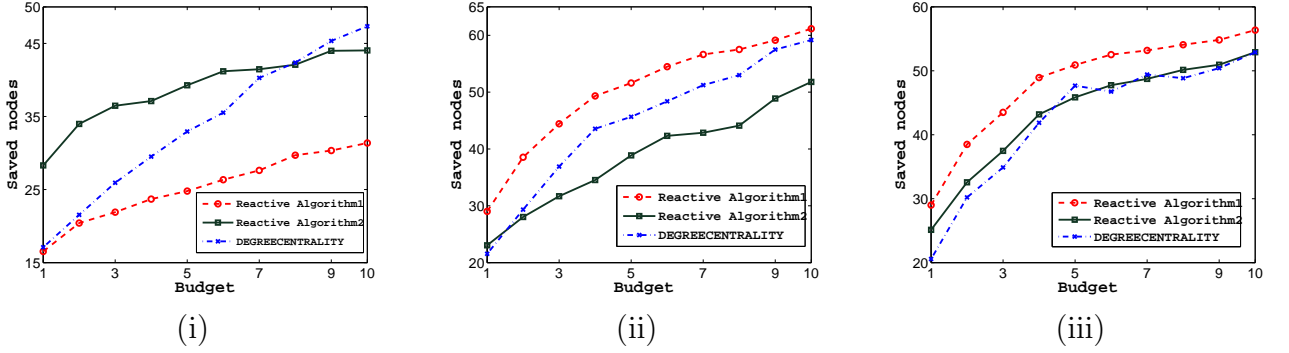


Figure 5.5: Under MGCM: Saved nodes (percentage of nodes) versus budget (percentage of nodes) using (i) *Celegans* data set, (ii) *Jazz* data set, (iii) *Football* data set

5.6 Experimental Results and Analysis for MGCM

In Figures 5.4 and 5.5, we call SVRM (Algorithm 5) as Reactive Algorithm1 and TRM (Algorithm 4) as Reactive Algorithm2. Under MGCM, Figures 5.4 and 5.5 show the performance of TRM and SVRM vis-a-vis the algorithm based on conventional degree centrality on real world social network data sets with different densities. In particular, Figures 5.4(i) and 5.5(i) corresponds to *Celegans* data set, Figures 5.4(ii) and 5.5(ii) correspond to *Jazz* data set, Figures 5.4(iii) and 5.5(iii) correspond to *Football* data set. *Jazz* data set is the most dense data set. *Celegans* data set is the least dense. It is clear that the efficacy of SVRM is better than that of the algorithm based on conventional

degree centrality. But performance of SVRM degrades if density of the graph is less. TRM works well when the graph is sparse. Algorithm based on degree centrality works better than TRM on Jazz data set. On Celegans data set, TRM works better than algorithm based on degree centrality. For Football data set, it performs at par with the degree centrality based algorithm.

Chapter 6

Algorithms for Limitation of Misinformation

This section is organized as follows: we describe the issues regarding the limitation of misinformation problem in Section 6.1. We formulate the problem and propose our algorithm in Section 6.2. We give experimental results of the performance of this algorithm in Section 6.3.

6.1 Issues

The problem of limitation of misinformation is already studied in [4] and [21]. The first one considers restricted version of the problem which is unable to capture real world scenarios. The second one uses SV based approach to solve a more generalized version of the problem. In the generalized scenario, the underlying function is not submodular. In this context, we also propose a SV based heuristic which considers a different characteristic function. We show in Section 6.3 that our algorithm outperforms the existing algorithm in [21].

6.2 Problem Formulation and Algorithm

We define the model along with the optimization problem for limitation of misinformation.

Model

We define our model as follows: a social network can be represented by a directed graph $G = (V, E)$ consisting of nodes V and edges E . A node j is a neighbor of node i if there is an edge $e_{i,j} \in E$ from node i to node j in G .

We work with MCICM described in [4]. There are two kind of probabilities on the edges of this graph. One is negative influence, denoted by $p_{R,i,j}$ and the other is positive influence $p_{W,i,j}$, for each edge $e_{i,j} \in E$. Negative campaign (R) of misinformation propagates based on negative influences and positive/counter campaign (W) propagates based on positive influences. Both these campaigns follow MCICM model to propagate.

Let the initial set of active nodes for cascade G (R) is denoted by I_G (I_R). When one node i is activated in any campaign at step t , it has only one chance to activate its inactive neighbour j . It succeeds with probability $p_{G,i,j}$ ($p_{R,i,j}$) in campaign G (R), given that no other neighbour of j tries to activate j in the same campaign in same time step. If there are two or more nodes try to activate same node at same time, at most one of them can succeed. Once a node becomes active, it remains active for the rest of the process. If campaign G and R both reaches to a node at the same time, campaign G takes effect.

Problem Definition

We define limitation of misinformation problem as follows: the spread of misinformation starts from one node r_1 and is detected with delay d , at which point of time, positive campaign is triggered at the selected k nodes. Hence, the problem is to find k nodes to initiate the positive campaign. We define a function $g : S \rightarrow [0, n]$ where $S = \{V' | V' \subset V \text{ and } |V'| \leq k\}$. Let T be the set of nodes where positive campaign has been triggered

and $g(T)$ be the expected number of nodes that can be saved from misinformation in the given random process. The goal is to find a set T ($|T| \leq k$) of nodes to be vaccinated in order to maximize $g(T)$.

Algorithm

Some important theorems about limitation of misinformation problem are as follows:

THEOREM 6.1. [4] : *The function $g : S \rightarrow [0, n]$ where $S = \{V' | V' \subset V \text{ and } |V'| \leq k\}$ is not submodular.*

THEOREM 6.2. [4] : *Limitation of misinformation problem is NP-hard even with the high effectiveness property.*

According to Theorem 2, limitation of misinformation problem is NP-hard in its generalized version. In the context of this kind of misinformation limitation, SV approach is a natural one because it captures the marginal contribution that a user brings into the system. For limiting misinformation, the number of nodes saved from misinformation by a user, is his marginal contribution to a coalition.

We define two functions as follows: $\sigma_R : 2^V \rightarrow [0, n]$ and $\sigma_G : 2^V \rightarrow [0, n]$. Here σ_R is the negative influence function which gives the expected number of misinformed nodes and σ_G is the positive influence function which gives the expected number of nodes influenced in the positive campaign. We define the characteristic function used for computing the SVs of nodes as follows:

$$\nu_4(C) = \sigma_R(C) + \sigma_G(C) \quad (6.1)$$

for all $C \subseteq V$, i.e., given a set of users C , our characteristic function $\nu_4(C)$ is same as the summation of two influence functions. The intuition behind the characteristic function is to select the nodes which have both more positive and negative influences, because we have to influence the graph by positive influence as much as possible and have to minimize the misinformation.

Given a set of users, we are interested in finding the set of users T , where $|T| = k$, such

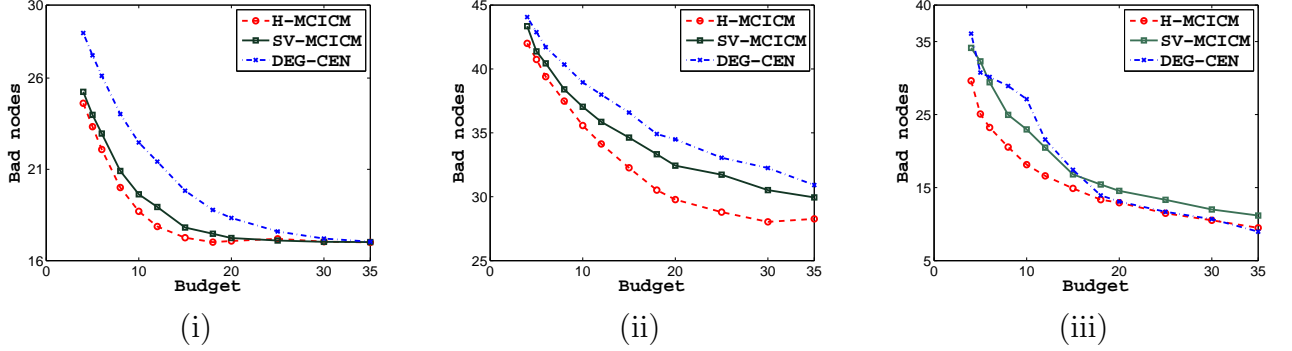


Figure 6.1: *Bad nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Football data set, (ii) Jazz data set, (iii) Celegans data set*

that after triggering positive influence to this set of users, the number of misinformed nodes in the whole graph, under MCICM, is minimized.

According to the SV based heuristic (we call it H-MCICM, Heuristic for MCICM) described in Algorithm 6, we are given the social network graph, the probabilities $p_{G,i,j}$, $p_{R,i,j}$ for all nodes i, j in V , the number of rounds for which the heuristic is to be run, and the number of nodes, k , to be chosen for triggering positive influence. In each round, we calculate a live graph for the positive and negative influence propagation. Next we construct a set (PB-set, Probably Best set depending on SV) as the set of users in that live graph. Now for each node in PB-set, we calculate its SV using $\nu_4(C)$ as the characteristic function. Once the number of rounds reaches S , we can choose k nodes with the highest SVs, as the top k nodes to minimize the misinformation in the given social network. We show in Section 6.3 that H-MCICM outperforms the algorithms proposed in [21].

6.3 Experimental Results and Analysis

Under MCICM, Figures 6.1, 6.2, 6.3 and 6.4 show the performance of H-MCICM (Algorithm 6) vis-a-vis SV-MCICM [21] and the algorithm based on conventional degree centrality on real world social network data sets with different number of edges and

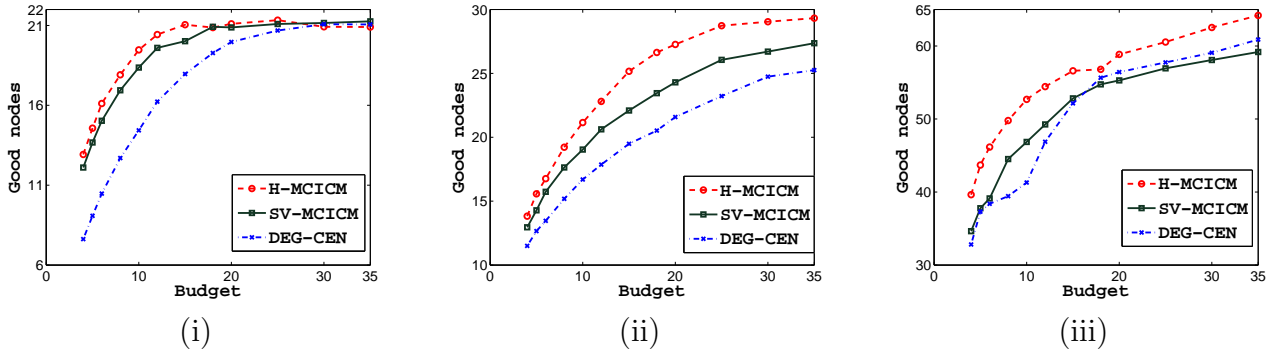


Figure 6.2: Good nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Football data set, (ii) Jazz data set, (iii) Celegans data set

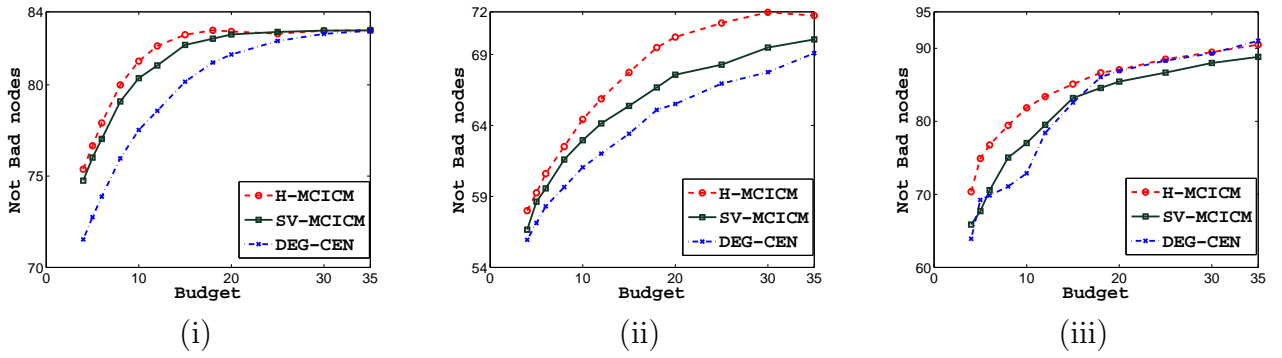


Figure 6.3: Not bad nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Football data set, (ii) Jazz data set, (iii) Celegans data set

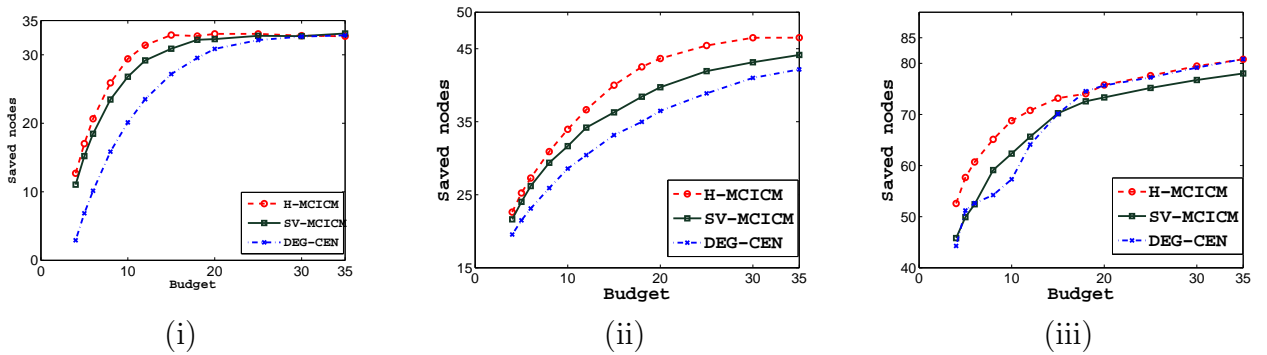


Figure 6.4: Saved nodes (percentage of nodes) versus budget (percentage of nodes) using (i) Football data set, (ii) Jazz data set, (iii) Celegans data set

Algorithm 6 Heuristic for MCICM (H-MCICM)

Require: Social network graph $G(V, E)$, the probabilities $p_{G,i,j}$, $p_{R,i,j}$, Number of simulations or rounds S , Delay d , Number of nodes k which has to be vaccinated

- 1: **if** $k < n$ **then**
 - 2: **for** $i = 1$ to S **do**
 - 3: R_1 = Live graph for the random process of negative influence/misinformation propagation process
 - 4: G_1 = Live graph for the random process of positive influence propagation process
 - 5: Based on delay d , from G remove the nodes which are already infected by the above stated random process and construct G'
 - 6: PB-set = The common nodes in G_1 , R_1 and G'
 - 7: **for** each node in PB-set **do**
 - 8: Compute the SV from G_1 and R_1 based on characteristic function $\nu_4(.)$
 - 9: **end for**
 - 10: **end for**
 - 11: Choose the top k nodes with highest SVs to be vaccinated
 - 12: **end if**
-

nodes. In particular, in these four figures, (i) corresponds to football data set, (ii) corresponds to Jazz data set and (iii) corresponds to Celegans data set. Here, “bad” nodes are those which get activated in negative campaign, “good” nodes are those which get activated in positive campaign. By “not bad” nodes, we mean those nodes which are not activated in negative campaign. “Saved” nodes are those which do not get activated by negative campaign because of the presence of positive campaign. It is clear that the efficacy of H-MCICM is better than that of SV-MCICM and the algorithm based on conventional degree centrality.

Chapter 7

Conclusion and Future Work

In this work, we have shown how game theoretic approach can help in solving virus inoculation problem under a probabilistic model. At first, we show cooperative game theoretic degree centrality is better than conventional degree centrality as a centrality measure. Secondly, we give a heuristic based on SV which outperforms existing heuristics for probabilistic models. As the underlying function is not submodular, SV performs well irrespective this fact. Assumption of submodularity in real world scenario is rare in context of vaccination inoculation problem. Third, we investigate the virus inoculation problem under reactive strategies. We design two algorithms to solve this problem under reactive setting. Both the algorithms perform better than conventional degree centrality. We propose one more model MGCM, a modified version of GCM model, to capture reactive settings and show that the virus inoculation problem under MGCM is NP-hard. We also propose a novel characteristic function to give an improved SV based approach to solve limitation of misinformation problem.

Some interesting directions for future work are as follows:

Incorporating Costs

We propose a heuristic to solve the virus inoculation problem under the model described by Abbasi and Hoda in [2]. We only consider the optimization problem and solve this problem using a game theoretic tool. To involve the cost of users for infection and

vaccination, the cost of the total budget explicitly and minimizing those costs can be considered as a good future direction to work under the proposed model, GCM.

Virus Inoculation Game

We consider virus inoculation problem in a general way. Strategic users in the network play a game called virus inoculation game between them and they want to minimize their own costs. Here by allowing the users to be strategic we aim to incorporate the actual virus inoculation game in the described model, GCM under the setting of reactive strategies and to design some proper mechanisms to solve this problem.

Special Structures with Non-submodularity

We have seen that if underlying function is not submodular, no good approximation can be found. So, one direction of future work can be to design some algorithms which can well approximate certain kinds of network with some special graphical structures in the context of underlying virus inoculation game.

Agnostic Approximation Algorithm

In general, to design a good approximation algorithm which will be agnostic towards underlying nature of the function (submodular, supermodular, both or none) can be seen as a good challenge in this area of work.

Bibliography

- [1] Karthik V. Aadithya, Balaraman Ravindran, Tomasz P. Michalak, and Nicholas R. Jennings. Efficient computation of the shapley value for centrality in networks. *Proceedings of the 6th International Conference on Internet and Network Economics, WINE*, pages 1–13, 2010.
- [2] Zeinab Abbasi and Hoda Heidari. Toward optimal vaccination strategies for probabilistic models. *Proceedings of the 20th International Conference on World wide web, WWW*, pages 1–2, 2011.
- [3] James Aspnes, Kevin Chang, and Aleksandr Yampolskiy. Inoculation strategies for victims of viruses and the sum-of-squares partition problem. *Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 43–52, 2005.
- [4] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. Limiting the spread of misinformation in social networks. *Proceedings of the 20th International conference on World Wide Web, WWW*, pages 665–674, 2011.
- [5] Tim Carnes, Chandrashekhar Nagarajan, Stefan M. Wild, and Anke van Zuylen. Maximizing influence in a competitive social network: a follower’s perspective. In *Proceedings of the 9th International conference on Electronic Commerce, EC ’07*, pages 351–360, 2007.
- [6] Javier Castro, Daniel Gómez, and Juan Tejada. Polynomial calculation of the shapley value based on sampling. *Computers and Operations Research*, 36:1726–1730, May 2009.

- [7] Po-An Chen, Mary David, and David Kempe. Better vaccination strategies for better people. *Proceedings of the 11th ACM Symposium on Electronic Commerce, EC*, pages 294–301, 2010.
- [8] Pedro Domingos and Matthew Richardson. Mining the network value of customers. *Proceedings of the 7th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD*, pages 47–56, 2001.
- [9] Pedro Domingos and Matthew Richardson. Mining knowledge-sharing sites for viral marketing. *Proceedings of the 8th ACM SIGKDD conference on Knowledge Discovery and Data Mining, KDD*, pages 61–70, 2002.
- [10] Josep Daz, Dieter Mitsche, Navin Rustagi, and Jared Saia. Fear in mediation: Exploiting the windfall of malice. *Proceedings of the 5th International Workshop on Internet and Network Economics, WINE*, pages 455–462, 2009.
- [11] D. Gomez, E. Gonzalez-Aranguena, C. Manuel, G. Owen, M. Del Pozo, and J. Tejada. Centrality and power in social networks: a game theoretic approach. *Mathematical Social Sciences*, 46:27–54, 2003.
- [12] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. *Proceedings of the 9th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD*, pages 137–146, 2003.
- [13] Dominic Meier, Yvonne Anne Oswald, and Stefan Schmid. On the windfall of friendship: Inoculation strategies on social networks. *Proceedings of the 9th ACM Symposium on Electronic Commerce, EC*, pages 294–301, 2008.
- [14] Evgeny Morozov. Swine flu: Twitters power to misinform. *Foriegn Policy*, 2009.
- [15] Thomas Moscibroda, Stefan Schmid, and Roger Wattenhofer. When selfish meets evil: Byzantine players in a virus inoculation game. *Proceedings of the 25th ACM Symposium on Principles of Distributed Computing*, pages 35–44, 2006.

- [16] Roger B. Myerson. Graphs and cooperation in games. *Mathematics of Operations Research*, 2:225–229, 1977.
- [17] Y. Narahari, Dinesh Garg, Ramasuri Narayanam, and Hastagiri Prakash. *Game Theoretic Problems in Network Economics and Mechanism Design Solutions*. Springer Book Series in Advanced Information and Knowledge Processing, Springer, London, 2009.
- [18] Ramasuri Narayanam and Y. Narahari. Determining top k nodes in social networks using the shapley value. *Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems, AAMAS*, 3:1509–1512, 2008.
- [19] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming*, 14:265–294, 1978.
- [20] Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, 2007.
- [21] H Prem Raj and Y. Narahari. Influence limitation in multi-campaign social networks with non-submodular influence functions: A shapley value based approach. In *(to appear) Proceedings of IEEE conference on Automation Science and Engineering (CASE)*, Seoul, Korea, August 2012.
- [22] Lloyd S. Shapley. A value for n-person games. In *Contributions to the Theory of Games, Volume I*, Princeton University Press, Princeton, New Jersey, USA, 1950.