M. Sc Engineering Thesis

Development of Graph Theoretic Models for Business Promotion on Social Networks

by

Safique Ahmed Faruque

Submitted to

Department of Computer Science and Engineering

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE & ENGINEERING

Department of Computer Science and Engineering

BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

Dhaka-1000, Bangladesh

December, 2013

The thesis titled "**Development of Graph Theoretic Models for Business Promotion on Social Networks**", submitted by Safique Ahmed Faruque, Roll No. 0409052041P, Session April 2009, to the Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, has been accepted as satisfactory in partial fulfillment of the requirements for the degree of Masters of Science in Computer Science and Engineering and approved as to its style and contents. Examination held on December 9, 2013.

Board of Examiners

1. ______________________________

Dr. Md. Saidur Rahman — Chairman
Professor — (Supervisor)
Department of CSE, BUET, Dhaka.

2. ______________________________

Dr. Abu Sayed Md. Latiful Hoque — Member
Professor & Head — (Ex-officio)
Department of CSE, BUET, Dhaka.

3. ______________________________

Dr. Mohammad Mahfuzul Islam — Member
Professor
Department of CSE, BUET, Dhaka.

4. ______________________________

Dr. Mahmuda Naznin — Member
Associate Professor
Department of CSE, BUET, Dhaka.

5. ______________________________

Dr. Md. Rezaul Karim — Member
Professor — (External)
Department of Computer Science and Engineering
University of Dhaka
Dhaka-1000, Bangladesh.

Candidate's Declaration

This is to certify that the work presented in this thesis entitled "**Development of Graph Theoretic Models for Business Promotion on Social Networks**" is the outcome of the investigation carried out by me under the supervision of Professor Dr. Md. Saidur Rahman in the Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology(BUET), Dhaka. It is also declared that neither this thesis nor any part thereof has been submitted or is being currently submitted anywhere else for the award of any degree or diploma.

Safique Ahmed Faruque
Candidate

Contents

Board of Examiners **i**

Candidate's Declaration **ii**

Acknowledgments **viii**

Abstract **ix**

1 Introduction **1**

1.1 An Example of Mapping 3

1.1.1 Problem Definition in Business Perspective 3

1.1.2 Problem Definition in Graph Theoretic Approach 4

1.2 Motivation 4

1.3 Previous Works 5

1.4 Scope of This Thesis 6

1.4.1 Business Problems to Model 6

1.4.2 Graph Algorithmic Solutions 7

1.4.3 Defining Graph Classes 8

1.5 Thesis Organization 9

2 Preliminaries **10**

2.1 Basic Terminology 10

2.1.1 Graphs 10

2.1.2 Simple Graphs and Multigraphs 11

2.1.3 Directed and Undirected Graphs 11

2.1.4 Paths and Cycles 13

2.1.5 Connectivity 13

2.1.6 Dominating Set 14

2.2 Complexity of Algorithms 15
2.2.1 The Notation $O(n)$ 15
2.2.2 Polynomial Algorithms 16
2.3 Social Network Analysis and Marketing 16
2.3.1 Social Networks 16
2.3.2 Social Networking Sites 16
2.3.3 Social Network Marketing 17
2.3.4 Direct Marketing 18
2.3.5 Viral Marketing 18
2.3.6 Seeding Strategy 19

3 Social Network Marketing Mappable Problems 20
3.1 Introduction 20
3.2 Direct Marketing Problem 20
3.2.1 Direct Marketing 21
3.2.2 Problem Definition in Business Perspective 21
3.2.3 Problem Definition in Graph Theoretic Approach 21
3.3 Seeding Strategy Problem 22
3.3.1 Problem Definition in Business Perspective 23
3.3.2 Problem Definition in Graph Theoretic Approach 23
3.4 Stop Propagating Problem 23
3.4.1 Problem Definition in Business Perspective 24
3.4.2 Problem Definition in Graph Theoretic Approach 24
3.5 Target Market Finding Problem 24
3.5.1 Problem Definition in Business Perspective 25
3.5.2 Problem Definition in Graph Theoretic Approach 26
3.6 Most Influential People Finding Problem 26
3.6.1 Problem Definition in Business Perspective 27
3.6.2 Problem Definition in Graph Theoretic Approach 27
3.7 Conclusion 28

4 Seeding Strategy Problem 29
4.1 Introduction 29
4.2 Selecting Promising Clients 32
4.2.1 Distance Based Scheme 32

4.2.2 Influence Based Scheme 32
4.2.3 Distance and Influence Based Scheme 33
4.3 Conclusion 35

5 Direct Marketing Problem **36**
5.1 Introduction 36
5.2 Problem Statement 37
5.3 Graph Classes 40
5.4 Properties for Identifying Graph Classes 41
5.5 Case Studies 42
5.5.1 Choosing from MDS 43
5.5.2 Graphs of Minimum Degree k 43
5.6 A Heuristic Algorithm for Covering Graph 45
5.6.1 Algorithm 45
5.6.2 Experiments 48
5.7 Conclusion 48

6 Stop Propagating Problem **51**
6.1 Introduction 51
6.2 Models of Spreading Information 52
6.3 Problem Statement 53
6.4 Algorithm for Stopping Propagation 55
6.5 Extension of the Problem 57
6.6 Conclusion 58

7 Conclusion **59**

References **61**

Index **65**

List of Figures

1.1 Representation of social network as a graph. 2

2.1 A graph with six vertices and eight edges. 11
2.2 Multigraphs. 12
2.3 Undirected and directed graphs. 12
2.4 Connected and disconnected graphs. 14
2.5 Dominating sets 14
2.6 Representation of a social network 17
2.7 Information propagation in viral marketing. 19
2.8 Seeds in a social network 19

3.1 Model of direct marketing problem 22
3.2 Model of seeding strategy problem 23
3.3 Model of stopping propagation problem 25
3.4 Model of target market finding problem 26

4.1 A social network 31
4.2 Illustration of distance based scheme 33
4.3 Illustration of influence based scheme 34

5.1 Illustration of direct marketing model 39
5.2 Example of a non-optimally covered network 40
5.3 Campaign friendly 40
5.4 Campaignable 41
5.5 Property of position 42
5.6 Property of cycle 42
5.7 Comparison of choosing initial set from MDS 43
5.8 Direct marketing campaign on a graph of minimum degree k . . 44

5.9 Pattern of having short of candidate vertices. 44
5.10 Cut vertices of a graph 46
5.11 Sample input graphs from rome library 49

6.1 Propagating environment 54
6.2 Blocker sets for stopping propagation 54
6.3 Illustration of Algorithm Filter_Adjacent 55
6.4 Illustration of extended problem 57

Acknowledgments

I would like to thank, first of all, my supervisor Professor Dr. Md. Saidur Rahman. He introduced me to this wonderful and fascinating world of *graph theory* and *graph drawing*. His wonderful classes in graduate level stimulated me to work in this field under his supervision. Whenever I found myself in stalemate condition, his helpful direction revived my interest and enthusiasm all the time. Without his constant supervision and words of encouragement, it would almost impossible for me to finish this work.

I am indebted to Dr. Mossa Anisa Khatun, who shed light on the study of marketing problems on social network. I also thank the members of my thesis committee: Prof. Dr. A. S. M. Latiful Hoque, Prof. Dr. Mohammad Mahfuzul Islam, Associate Prof. Dr. Mahmuda Naznin, and Prof. Dr. Md. Rezaul Karim for their valuable and helpful suggestions. This thesis work is done in the *Graph Drawing and Information Visualization Lab*, Dept. of CSE, BUET. It is my great pleasure to work in such a wonderful environment.

I would like to express my deep gratitude to all the members of our research group for their valuable suggestions and continual encouragements. My special thanks goes to Aftab Hussain and Iqbal Hossain. My parents also supported me and encouraged me to the best of their ability. My heart-felt gratitude goes to them.

Abstract

A social network is a social structure made up of a set of actors (such as individuals or organizations) and the dyadic ties between these actors. It is becoming increasingly an important part of marketing platform of any business. Its methodical analysis has gained significant amount of interest for researching not only at the field of modern sociology but also at communication studies, information science, network science etc. Different marketing managers adopt different strategy to market their product on social networks. They face a lot of problems while marketing their products on social networks.

On the other hand, a social network is nothing but a large graph, and hence a graph $G = (V, E)$ can be represented as a social network where each vertex $v \in V$ represents a user and there exists an edge $(u, v) \in E$ if the users corresponding to vertices u and v are acquaintance. Due to the ability of mapping a social network to a graph, it came into our mind that some of the social network marketing problems could be mapped to graph theoretic problems. In this thesis we list out social network marketing problems those are mappable to graph theoretic problems. We dig into some of the problems to formulate them in details and then we map those problems to graph theoretic problems. For executing a direct marketing campaign which is based on social network data, we examine different cases to cover the whole network with minimum round of campaign. We develop a graph theoretic model for selecting seeds at a viral marketing campaign on a social network where the existance of business rivals is considered. We devise an algorithm to stop propagation of harmful information on a social network. We also identify graph classes based on suitability of campaigning.

Chapter 1

Introduction

Social network analysis has become a buzzword as a research topic not only in the field of modern sociology but also in the field of communication studies, information science, network analysis, behavioral sciences, etc. Researchers of different sectors have started significant research on algorithmic analysis of social networks and it is evident from the many recent publications on social networks [6, 23, 8, 29]. Social networking sites are the major source of the data used in these research and analysis. These sites (e.g., Facebook, LinkedIn, Twitter etc.) are also one of the fastest growing arenas of the World Wide Web. These sites rely upon user-generated content to attract and retain visitors and they also accumulate user information those are valuable for targeted marketing in marketing sector. By considering these sites as marketing media, marketers can reduce marketing costs, make information propagation faster and influence customers easily to buy a product or recieve a service. Now-a-days media of different kinds of marketing like direct marketing, mass marketing are gradually being replaced from the traditional context to social network context. Viral marketing cannot be thought of now a days without social networking sites. But extracting proper information regarding the data of social networks from this media is not simple for marketers. Large chain of relationships and different kinds of attributes like mutual friendship, being small world, community membership make the social network a complex network for marketers.

Representation of a social network usually consists of the user and their connections, considering the users as vertices and connections as edges turns it into a simple huge graph. Figure 1.1 (a) shows an example of social network and Figure 1.1(b) shows its representation as a graph.

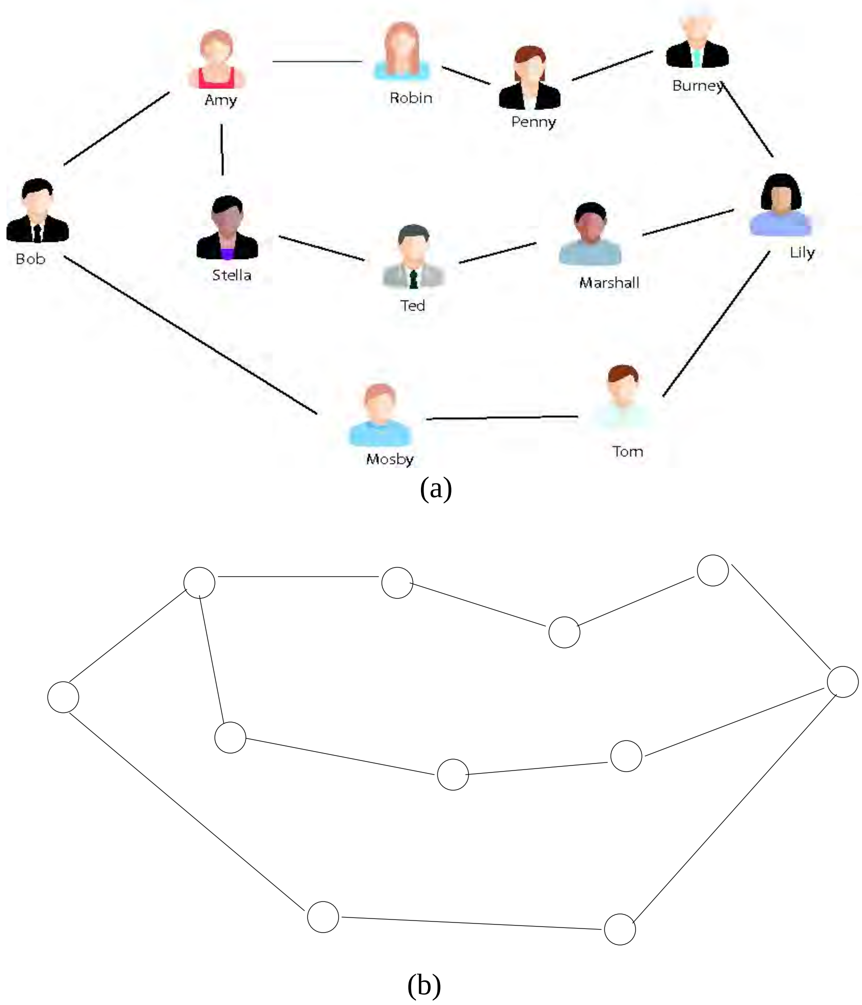

Figure 1.1: Representation of social network as a graph.

Fortunately, graph theory is rich in theoretical results as well as applications to real world problems. There has been significant number of research on complex networks, different states of networks, different activity on networks to solve graph theoretic problems. This enriched stock of algorithms and techniques can be applied to solve social network marketing problems as well. To solve such type of problems, primary activity is modeling, a process involving formulating a problem in such a way that it can be approached by algorithms and techniques in Graph Theory.

We identify some of the problems that are mappable to graph theoretic problems and map those marketing problems to graph theoretic problems. We develop a graph theoretic model of seeding strategy for a viral marketing campaign on a social network where the existence of business rivals is considered. We figure out some solutions of some of those business problems using graph theoretic algorithms. In this thesis, we also define graph classes base on suitability for business campaign by considering the structural properties of graphs.

In the rest of this chapter, we provide an example of mapping a marketing problem to a graph theoretic problem, necessary background and scope of this thesis. In Section 1.1, we present an example of a business problem and its modeling to graph theoretic problem. In Section 1.2, we describe the motivation behind this thesis work. Some research on business problem modeling are presented in Section 1.3. In Section 1.4, we depict the scope and and objectives of this thesis. Under this section we provide the list of the problems those we have mapped in this thesis and classification of graph base on suitability of direct marketing campaign. Finally the thesis organization is narrated in Section 1.5.

1.1 An Example of Mapping

In this section we present an example of mapping a social network marketing problem to a graph theoretic problem. Here, we will consider a marketing campaign where the information is spreaded through word of mouth communication. This type of marketing campaign is known as *Viral marketing*. Viral marketing is cheaper than any other marketing. It is now-a-days widely used to spread information on social network. But initialization of a viral marketing campaign is very crucial. Successful marketing campaign highly depends on properly choosing the initial persons. Company needs to find those active persons who are suitable to spread the product information. These persons can vary according to the nature of the product and the state of current usage of this type of product. Here we define a problem to find these persons in a social network where users have not started consuming the considering product but they consume similar products of other manufacturers.

1.1.1 Problem Definition in Business Perspective

Let company A wants to advertise its new product to prospective customers. But there are some people in the network who are already convinced by other business rivals. It might be difficult to convince the people adjacent to the already convinced people. The problem is to find out people who are more prospective customers to be convinced.

1.1.2 Problem Definition in Graph Theoretic Approach

Let $G = (V, E)$ be a simple graph representing the social network where each vertex $v \in V$ represents a customer in marketing perspective and there exists an edge $(u, v) \in E$ if the customers corresponding to vertices u and v are acquaintances. Let C be the set of vertices corresponding to the persons convinced by rivals. The problem is to find a set of k vertices $S \in V$ - C such that they are in longest distances from the vertices in C.

1.2 Motivation

Social network analysis has gained significant attention in recent years, largely due to the success of social networking sites and the consequent availability of social network data. In spite of the growing interest, however, there is little understanding of the applications of analyzing social networks on business problems. While there is a large body of research on different problems on social networks, there is a gap between the algorithms and techniques developed by the research community and their deployment in real-world business problems. Therefore the business problems on social networks are still largely unexplored.

Extensive researches on graph theory have already been taken place and huge amount of researches are still going on. This area is rich with many kinds of algorithms and techniques that solve graph theoretic and graph drawing problems. As representation of social networks by graphs is not only possible but also the most popular way, it is likely that business problems can also be solved by this rich stock of algorithms. Except some of the business problems, most of the business problems are not formulated to apply these algorithms. As a result these problems remain unresolved.

Furthermore, if some business problems are formulated for being applied graph theoretic algorithms, it actually opens the door of analysis and optimization using graph theory.

1.3 Previous Works

Modeling business problems in graph theoretic approach has also gained interest to research. In most of the research works in this area, researchers first choose one business problem and model it as a graph theoretic problem. After that researchers analyze, do experiments and develop algorithms to solve the problems. In this section we briefly present some research works related to business problem modeling.

D. Kempe *et al.* [21] have considered seeding strategy problem to maximize the influence through a social network. As the optimization problem of selecting the most influential nodes is NP-hard, they have provided the first provable approximation guarantees for efficient algorithms. Using an analysis framework based on submodular functions, they have shown that a natural greedy strategy obtains a solution that is provably within 63% of optimal for several classes of models. Our work on seeding strategy problem is slightly different than this problem. On top of the problem, we have considered the existance of business rivals on social networks. Highly influential nodes, found by their algorithm, may not be highly influential in our scenario.

N. Mishra *et al.* [28] have worked on social network clustering problems. They have introduced a new criterion that overcomes some limitations of existing graph clustering by combining internal density with external sparsity for social networks. An algorithm has been given for provably finding the clusters, provided there is a sufficiently large gap between internal density and external sparsity. They also did some experiments on real social networks that illustrates the effectiveness of the algorithm. T. Qian et al. [33] also worked on refining graph partitioning for social network clustering. We have also included this problem in our list of mappable problems.

Besides these, significant research works have been done for different kinds of analysis on social networks like visual analysis [20], diffusion analysis [27], Marketing strategy analysis [16] etc. In this thesis, we list out some of the business problems and model them in such a way that one can analyze them using graph theoretic approach.

1.4 Scope of This Thesis

In this section we first present the list of business problems those are modeled in this thesis. After that we will mention the overview of graph algorithmic solutions for some of the business problems. Then we define graph classes based on suitability of direct marketing campaigning.

1.4.1 Business Problems to Model

In this section we prepare a list of mappable social network porblems. These problems will be modeled and graph theoretic algorithms will be applied on them.

Seeding strategy problem in presence of business rivals : Let company A wants to advertise its new product to prospective customers. But there are some people in the network who are already convinced by other business rivals. It might be difficult to convince the people adjacent to the already convinced people. So the problem is to find out people who are more prospective customers to be convinced.

Direct marketing problem : Assume that company A wants to advertise it's product to prospective customers. In the 1st round the company sends its representatives to some initially accessible customers. From the second round the company can only send their representative to the customers who have some acquaintance already reached by the representative of the company in some previous round. The company wishes to advertise on a social network of n people. If the company has $k < n$ representatives then how many rounds are necessary to cover the whole network? How to chose customers in each round to minimize the number of rounds?

Stop propagating problem : Let a product p of manufacturer A has a deficiency d. The harmful information i: p contains deficiency d has already been started to propagate through social network. Now A has removed the deficiency d from product p and decided to stop the propagation of i by sending another information $anti - i$: p has been overcome from the deficiency d. The problem is to find out the list of minimum persons to whom A will send $anti - i$ to stop propagation of i.

Target market finding problem : Let a company A has made a discount offer to consumer groups. Marketing manager of A has decided to spread this message through social network. But there are significant number of people exists who do not belongs to any group. On the other hand there also exists many people who belongs to multiple group. The problem is to find out the closely connected groups that can be overlapped and to exclude the people belongs to no group.

Most influential people finding problem : Let a company A wants to advertise its new product through social network. A needs to find a set of k people S called seed. A will directly market the product to them. Then, assuming they adopt the product, A relies on their influence to generate a large cascade of adoptions. A does not rely on any further direct promotion of the product.

The problem is how A choose the set S that ensures maximum cascade of adoption?

1.4.2 Graph Algorithmic Solutions

We first model the business problems in such a way that we can apply graph theoretic algorithms on it to solve the problems. Then we develop algorithms to solve the problems. In this thesis, we derived graph theoretic algorithms for Seeding strategy problem in presence of business rivals, Direct marketing problem and Stop propagating problem.

For seeding strategy problem in presence of business rivals, we consider two popular models of spreading informtaion. One is Linear Threshold Model and another is Independent Cascade Model. Based on these diffusion models we provide three schemes of seeding strategy in presence of business rivals.

In distance based scheme, we give an algorithm to rank promising clients using BFS labeling. We first mark each vertex that is convinced by business rival as a source vertex and execute BFS labeling. We consider the minimum label of each vertex as the rank of the vertex, and finally list the vertices in descending order of their rank.

In influence based scheme, we first delete the occupied vertices and then label them according to their degrees. Here also we list the vertices in descending order of their label.

In distance and influence based Scheme, we combine the distance based scheme and the influence based scheme. We compute weights w_1 and w_2 considering the distance based scheme and the influence based scheme as follows:

$w_1 = \frac{l_u}{max(L)}$, where l_u is the BFS label of vertex u and $max(L)$ is the largest BFS label among all the vertices.

$w_2 = \frac{d(u)}{\Delta}$, where $d(u)$ is the degree of u and Δ is the largest degree among all the vertices.

We define a function for ranking a vertex $f(r) = C_1 w_1 + C_2 w_2$. Finally we list the vertices in descending order of their rank.

For Direct marketing problem, we derived an algorithm to choose vertices in such a way that entire graph can be covered in optimum number of rounds. In this algorithm, first step is to find out all cut vertices of the considering graph as initial vertices to cover. Then from the next step an iteration will go on to choose the vertices of having maximum degrees of non-covered vertices.

For Stop propagating problem, we derived an algorithm to find out such minimum number of persons that convincing them can stop the propagation. According to this algorithm, we take an extra vertex and connect all adjacent vertices of infected vertices to it. Then we remove all the infected vertices. From all adjacent vertices of A in new temporary graph, we mark those vertices which have atleast one connection with $V - A$. These marked vertices are the set of minimum vertices to stop the propagation.

1.4.3 Defining Graph Classes

Based on our model of direct marketing over social network, every graph is not equally suitable for direct marketing campaign. Number of campaigning round will not be optimal, if in any particular round of campaign, any number of representatives face lack of enough customers to approach. Considering the suitability of campaigning we define three classes of graphs. These classes differ from each other on the structure of the graphs.

ii. Campaign friendly: These are the graphs on which any arbitrary way to campaign of vertices are optimal. As an example, every complete graph is campaign friendly.

iii. Campaignable: These are the graphs on which it is possible to complete campaigning in optimum round if the candidate vertices are chosen appropriately.

i. Campaign hard: These are the graphs on which it is impossible to complete campaigning in optimum round.

We describe some properties that help to identify a particular graph class. For example, if the graph is a cycle and number of representative (k) is 2, it is always optimally campaignable. Beside this, we develop an algorithm to choose appropriate vertices on a campaignable graph to campaign optimally.

1.5 Thesis Organization

The rest of this thesis is organized as follows. In Chapter 2, we give some basic terminology of graph theory and algorithmic theory. Chapter 3 contains a list of the mappable social network marketing problems and the corresponding graph theoretic problems. In Chapter 4, we describe our developed model for seeding strategy in presence of business rival. In Chapter 5, we elaborate the description of the problem regarding direct marketing based on social network data. This chapter also contains the classes of graphs based on the suitability for business campaign. We give an algorithm for stop propagation problem on social network in Chapter 6. Finally Chapter 7 concludes the thesis with a summary of the thesis with some future works.

Chapter 2

Preliminaries

In this chapter we define some basic terminology of graph theory, social network analysis and marketing that we will use throughout the rest of this thesis. In Section 2.1, we cover some definitions of standard graph-theoretical terms. Finally we introduce the notion of social network analysis and marketing in Section 2.3.

2.1 Basic Terminology

In this section we give some definitions of standard graph-theoretical terms used throughout this thesis. For readers interested in more details of graph theory we refer to [30, 31, 39].

2.1.1 Graphs

A *graph* G is a tuple (V, E) which consists of a finite set V of *vertices* and a finite set E of *edges*; each edge being an unordered pair of vertices. Figure 2.1 depicts a graph $G = (V, E)$ where each vertex in $V = \{v_1, v_2, \ldots, v_6\}$ is drawn as a small circle and each edge in $E = \{e_1, e_2, \ldots, e_8\}$ is drawn by a line segment.

We denote an edge joining two vertices u and v of the graph $G = (V, E)$ by (u, v) or simply by uv. If $uv \in E$ then the two vertices u and v of the graph G are said to be *adjacent*; the edge uv is then said to be *incident* to the vertices u and v; also the vertex u is said to be a neighbor of the vertex v (and vice versa). The *degree* of a vertex v in G, denoted by $d(v)$ or $deg(v)$, is the number of edges incident to v in G. In the graph shown in Figure 2.1 vertices v_1 and v_2

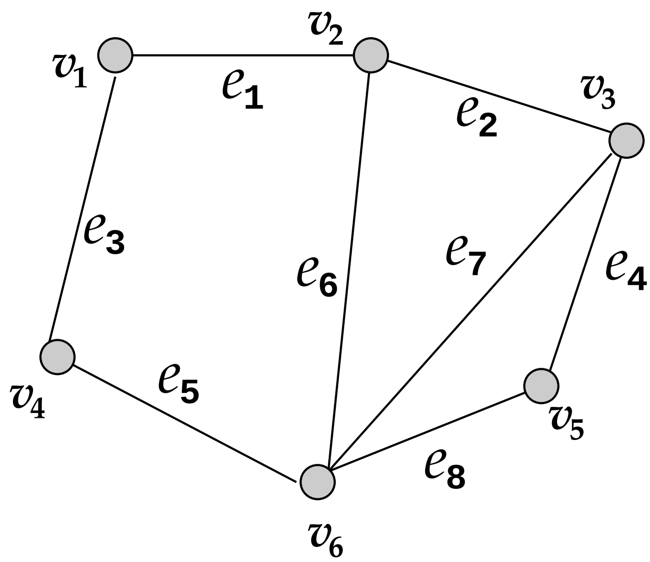

Figure 2.1: A graph with six vertices and eight edges.

are adjacent, and $d(v_6) = 4$, since four of the edges, namely e_5, e_6, e_7 and e_8 are incident to v_6.

2.1.2 Simple Graphs and Multigraphs

If a graph G has no "multiple edges" or "loops", then G is said to be a *simple graph*. *Multiple edges* join the same pair of vertices, while a *loop* joins a vertex with itself. The graph in Figure 2.1 is a simple graph.

A graph in which loops and multiple edges are allowed is called a *multigraph*. Multigraphs can arise from various applications. One example is the "call graph" that represents the telephone call history of a network. The graph in Figure 2.2(a) is a call graph that represents the call history among six subscribers. Note that there is no loop in this graph. Figure 2.2(b) illustrates another multigraph with multiple edges and loops.

Often it is clear from the context that the graph is simple. In such cases, a simple graph is called a *graph*. In the remainder of thesis we will only be concerned about simple graphs.

2.1.3 Directed and Undirected Graphs

In a *directed graph*, the edges do have a direction but in an *undirected graph*, the edges are undirected. Strictly speaking, each edge in a directed graph should be

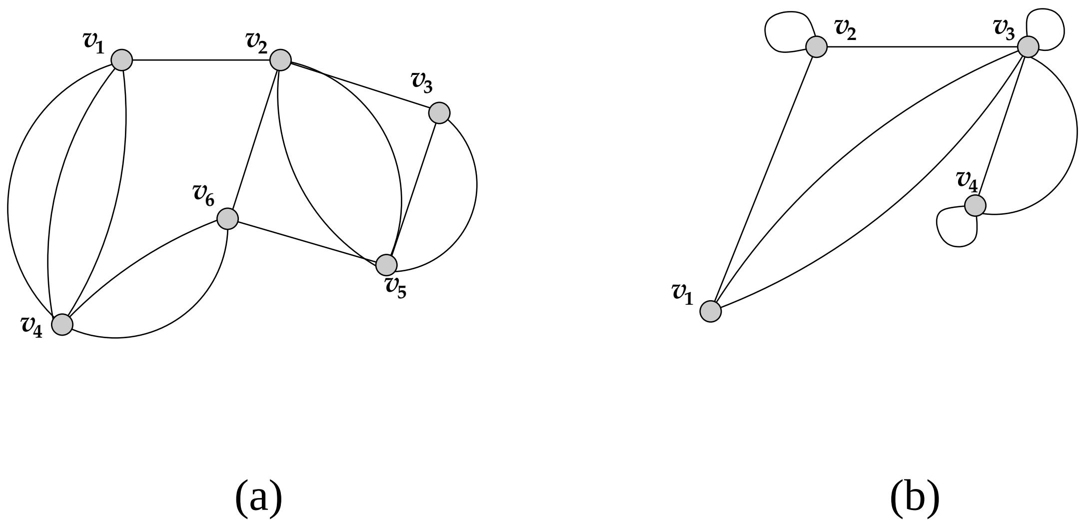

Figure 2.2: Multigraphs.

represented by a 2-tuple while for an undirected graph it should be represented by a 2-member subset of the vertex set. In Figure 2.3(a) and (b), we show an undirected and a directed graphs respectively. In this thesis, we will mean an undirected graph when we say "a graph" unless mentioned otherwise.

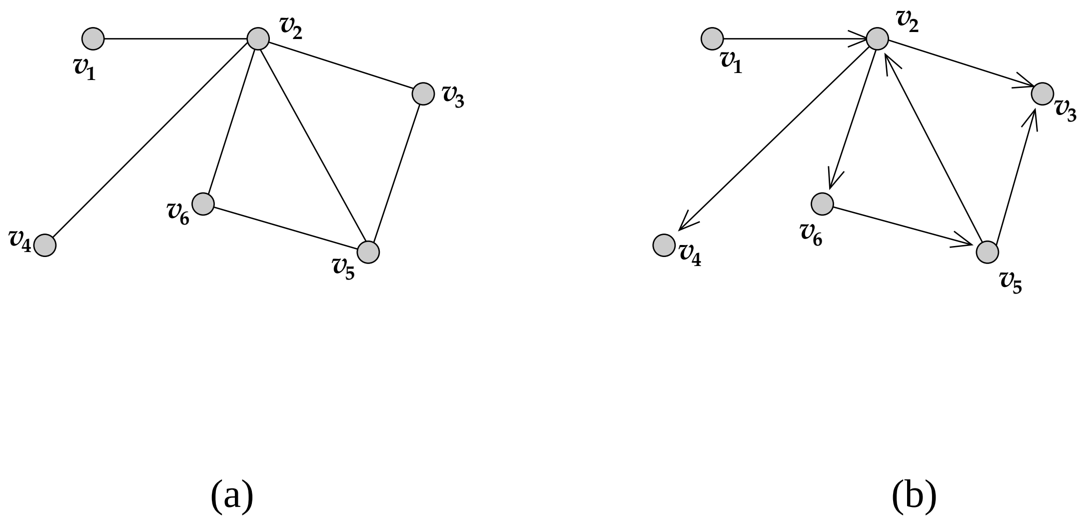

Figure 2.3: Undirected and directed graphs.

2.1.4 Paths and Cycles

A *walk*, $w = v_0, e_1, v_1, \ldots, v_{l-1}, e_l, v_l$, in a graph G is an alternating sequence of vertices and edges of G, beginning and ending with a vertex, in which each edge is incident to the two vertices immediately preceding and following it. The vertices v_0 and v_l are said to be the end-vertices of the walk w.

If the vertices $v_0, v_1, \ldots, v_l$ are distinct (except possibly v_0 and v_l), then the walk is called a *path* and usually denoted either by the sequence of vertices $v_0, v_1, \ldots, v_l$ or by the sequence of edges $e_1, e_2, \ldots, e_l$. The length of the path is l, one less than the number of vertices on the path. For any two vertices u and v of G, a u, v-path in G is a path whose end-vertices are u and v.

A walk or path w is *closed* if the end-vertices of w are the same. A closed path containing at least one edge is called a *cycle*. A cycle is called *Hamiltonian* if it consists of all the vertices of a graph G.

2.1.5 Connectivity

A graph G is *connected* if for any two distinct vertices u and v of G, there is a path between u and v. A graph which is not connected is called a *disconnected graph*. A *(connected) component* of a graph is a maximal connected subgraph. The graph in Figure 2.4(a) is a connected graph since there is a path between every pair of distinct vertices of the graph. On the other hand, the graph in Figure 2.4(b) is a disconnected graph since there is no path between, say, v_1 and v_5. The graph in Figure 2.4(b) has two connected components as indicated by the dotted lines. Note that every connected graph has only one component; the graph itself.

The *connectivity* $\kappa(G)$ of a graph G is the minimum number of vertices whose removal results in a disconnected graph or a single-vertex graph K_1. We say that G is *k-connected* if $\kappa(G) \geq k$. 2-connected and 3- connected graphs are also called biconnected and triconnected graphs, respectively. A *block* is a maximal biconnected subgraph of G. We call a set of vertices in a connected graph G a *separator* or a *vertex cut* if the removal of the vertices in the set results in a disconnected or single-vertex graph. If a vertex-cut contains exactly one vertex then we call the vertex a *cut vertex*.

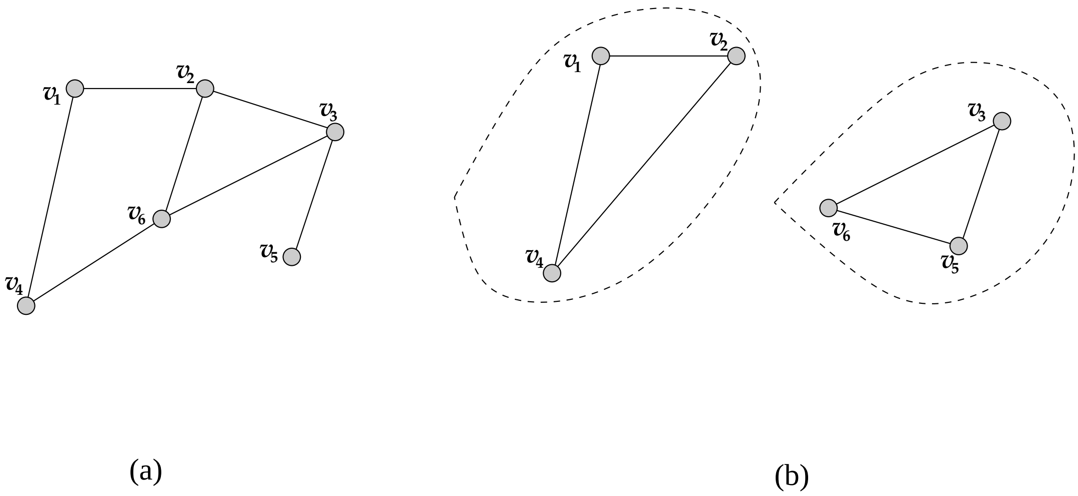

Figure 2.4: (a) A connected graph, (b) A disconnected graph with two connected components.

2.1.6 Dominating Set

A *dominating set* for a graph $G = (V, E)$ is a subset D of V such that every vertex not in D is adjacent to at least one member of D. Black marked vertices $D = \{a, e, f\}$ in Figure 2.5(i) is dominating set for the graph G where every vertex of the graph G is adjacent to atleast one of them. A *minimum dominating set* is the smallest dominating set for a graph. Multiple set can be minimum dominating set for a graph. In Figure 2.5(ii) minimum dominating set MDS for graph G is $\{b, e\}$ and the cardinality of this minimum dominating set $|MDS|$ is 2.

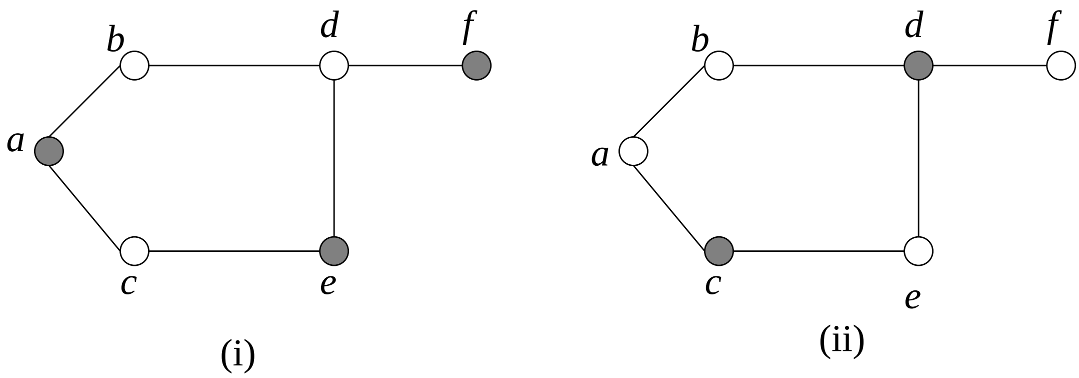

Figure 2.5: (i) A dominating set for a graph, (b) A minimum dominating set for a graph.

Finding a minimum dominating set is a fundamental math problem underlying routing that is highly applicable for communication networks like social networks. Member of dominating set constructs "cluster heads" in the network graph so one can rapidly, hierarchically, send message from one cluster to another and the cluster heads will oversee routing within and through the individual cluster. MDS is the smallest set of efficient routers or cluster heads. It is NP-hard in general, but efficient approximation algorithms do exists.

2.2 Complexity of Algorithms

In this section we briefly introduce some terminologies related to *complexity* of algorithms. For interested readers, we refer the book of Garey and Johnson [13].

The most widely accepted complexity measure for an algorithm is the *running time*, which is expressed by the number of operations it performs before producing the final answer. The number of operations required by an algorithm is not the same for all problem instances. Thus, we consider all inputs of a given size together, and we define the *complexity of the algorithm for that input size* to be the worst case behavior of the algorithm on any of these inputs. Then the running time is a function of size n of the input.

2.2.1 The Notation $O(n)$

In analyzing the complexity of an algorithm, we are often interested only in the "asymptotic behavior", that is, the behavior of the algorithm when applied to very large inputs. To deal with such a property of functions we shall use the following notations for asymptotic running time. Let $f(n)$ and $g(n)$ are the functions from the positive integers to the positive reals, then we write $f(n) = O(g(n))$ if there exists positive constants c_1 and c_2 such that $f(n) \leq c_1 g(n) + c_2$ for all n. Thus the running time of an algorithm may be bounded from above by phrasing like "takes time $O(n^2)$".

2.2.2 Polynomial Algorithms

An algorithm is said to be *polynomially bounded* (or simply *polynomial*) if its complexity is bounded by a polynomial of the size of a problem instance. Examples of such complexities are $O(n)$, $O(nlogn)$, $O(n^{100})$, etc. The remaining algorithms are usually referred as *exponential* or *non-polynomial.* Examples of such complexity are $O(2^n)$, $O(n!)$, etc. When the running time of an algorithm is bounded by $O(n)$, we call it a *linear-time* algorithm or simply a *linear* algorithm.

2.3 Social Network Analysis and Marketing

In this section we define some terminologies on social network analysis those used throughout the thesis. We also define some marketing terminologies. For interested readers, we refer to [38].

2.3.1 Social Networks

A Social Network is a social structure made up of a set of social actors (such as individuals or organizations) and a set of the dyadic ties between these actors [38].It is a theoretical construct useful to study relationships between individuals, groups, organizations, or even entire societies. The term is used to describe a social structure determined by interactions. This theoretical approach is, necessarily, relational. Because of having many different types of relations, singular or in combination, form these network configurations, network analytics are useful to a broad range of research enterprises. In social science, these fields of study include, but are not limited to anthropology, biology, communication studies, economics, geography, information science, organizational studies, social psychology, sociology, and sociolinguistics. Every individual of a social network is represented as a vertex in Figure 2.6 and the graph consists of all individuals and their relationships.

2.3.2 Social Networking Sites

Social network sites can be defined as web-based services that allow individuals to (1)construct a public or semi public profile in a bounded system, (2) articulate

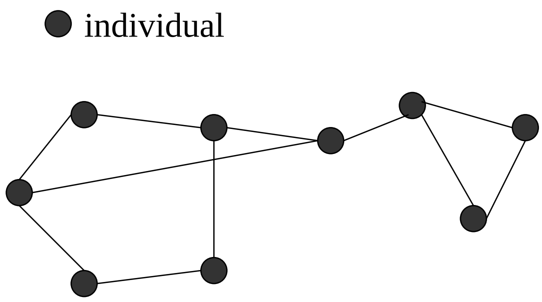

Figure 2.6: Representation of a social network

a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system [12]. Facebook, Twitter, Linked in, Pinterest, Myspace etc. are examples of social networking sites. A social networking site is a platform to build social networks or social relations among people who, for example, share interests, activities, backgrounds, or real-life connections. A social network site consists of a representation of each user (often a profile), his/her social links, and a variety of additional services.

2.3.3 Social Network Marketing

Social Network Marketing is a kind of marketing where the customers information are found from social networks. Irrespective of the media to communicate with customers, if potential customers are chosen from social networks in a marketing procedure, that marketing is known as *Social Network Marketing.* It usually refers to the process of gaining website traffic or attention through social network sites [36]. Social media like social networking sites are very much suitable for this type of marketing. Social network marketing programs usually center on efforts to create content that attracts attention and encourages readers to share it with their social networks. A corporate message spreads from user to user and presumably resonates because it appears to come from a trusted, third-party source, as opposed to the brand or company itself. Hence, this form of marketing is driven by word-of-mouth, meaning it results in earned media rather than paid media. Social media is a platform that is easily accessible to anyone with internet access. Increased communication for organizations fosters brand awareness and often, improved customer service. Additionally, social media serves as a relatively inexpensive platform for organizations to implement marketing campaigns.

2.3.4 Direct Marketing

In contrast to mass marketing, where a product is promoted indiscriminately to all potential customers, *direct marketing* attempts to first select the customers likely to be profitable and market only to those. It is a form of advertising that allows businesses to communicate straight to the customer, with advertising techniques that can include Cell Phone Text messaging, email, interactive consumer websites, online display ads, fliers, catalog distribution, promotional letters etc.

Direct marketing messages emphasize a focus on the customer, data and accountability [5]. A database of names often with certain other relevant information such as contact number/address, demographic information, purchase habits/history, company history, etc., is used to develop a list of targeted entities with some existing common interests, traits or characteristics. Generating such a database is often considered part of the Direct Marketing campaign. Marketing messages are addressed directly to this list of customer and/or prospects. Direct marketing relies on being able to address the members of a target market. Addressability comes in a variety of forms including email addresses, phone numbers, Web browser cookies, fax numbers and postal addresses [19]. Direct marketing emphasizes trackable, measurable responses, results and costs from prospects and/or customers regardless of medium.

2.3.5 Viral Marketing

Viral marketing refers to a marketing technique where marketing objectives are achieved through self-replicating viral process. It spreads information based on word-of-mouth technique. It is a method of product promotion that relies on getting customers to market an idea, product, or service on their own by telling their friends about it [40]. Figure 2.1 depicts how information propagates on social network in a viral marketing campaign.

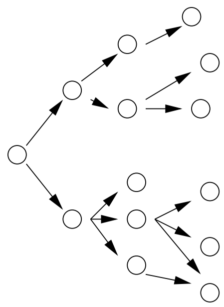

Figure 2.7: Information propagation in viral marketing.

2.3.6 Seeding Strategy

Seeding strategy is a strategy, which determines the initial set of targeted consumers chosen by the initiator of the viral marketing campaign. Each element of this initial set is known as *seed* of viral marketing campaign. Seeding strategy is one of the critical factors for success of viral marketing campaign. Seeding the right consumers yields up to eight times more referrals than seeding the wrong ones [17]. An example of a graph where seeds are identified by using a seeding strategy for a viral marketing is depicted on Figure 2.8.

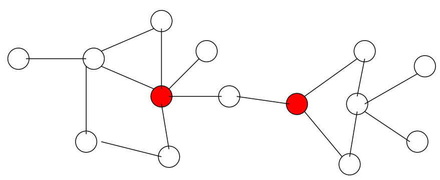

Figure 2.8: Seeds in a social network

Chapter 3

Social Network Marketing Mappable Problems

3.1 Introduction

Social network is widely used for marketing new as well as old products. Marketing based on social network has a large impact on sales and promotions of product. But it is more complex than a typical marketing as different factors like number of connectivity, interactivity etc influence the sales. So finding out the right person to market and execution of campaign in different state of network is not trivial problem. But the interesting fact that social network can be represented by simple graphs, so some of the problems of social network marketing can be mapped to the problems of graph theory. We list out this type of problems which are mappable. We present the problems in business perspective and then map them into graph theoretic problem. From Section 3.2 to Section 3.6 we describe the corresponding problems, define them in business perspective and define them in graph theoretic approach after mapping as well. Finally Section 3.7 concludes the chapter.

3.2 Direct Marketing Problem

In this section we define direct marketing and define the problem both in business perspective and graph theoretic approach.

3.2.1 Direct Marketing

In contrast to mass marketing, where a product is promoted indiscriminately to all potential customers, direct marketing attempts to first select the customers likely to be profitable and market only to those. Here marketing representatives go to the potential customers and try to convince them to buy the product. It is effective but costly. So choosing potential customers is crucial in this type of marketing. For this reason company takes different strategies to choose potential customers. Here we define the problem where company send their representative based on the social network data.

3.2.2 Problem Definition in Business Perspective

Assume that company A wants to advertise it's product to prospective customers. In the 1st round the company sends its representatives to some initially accessible customers. From the second round the company can only send their representative to the customers who have some acquaintance already reached by the representative of the company in some previous round. The company wishes to advertise on a social network of n people. If the company has $k < n$ representatives then how many rounds are necessary to cover the whole network? How to chose customers in each round to minimize the number of rounds?

3.2.3 Problem Definition in Graph Theoretic Approach

In graph theoretic approach we can reformulate the problem like this, let $G = (V, E)$ be a simple graph representing the social network where each vertex $v \in V$ represents a customer in marketing perspective and there exists an edge $(u, v) \in E$ if the customers corresponding to vertices u and v are acquaintances. Each round of campaign representatives will approach before at most k vertices that are not already approached earlier and adjacent to the already approached vertices. The problem is to find out how many rounds are needed to cover the whole network. And find out an algorithm to choose the vertices from candidate vertices in each round in such a way that minimizes the number of round. Figure 3.1 demonstrates the graph theoretic model of direct marketing problem. Here input is a graph of $n = 11$ vertices and number of representatives $k = 3$. Output is an algorithm to cover the entire graph in minimum steps.

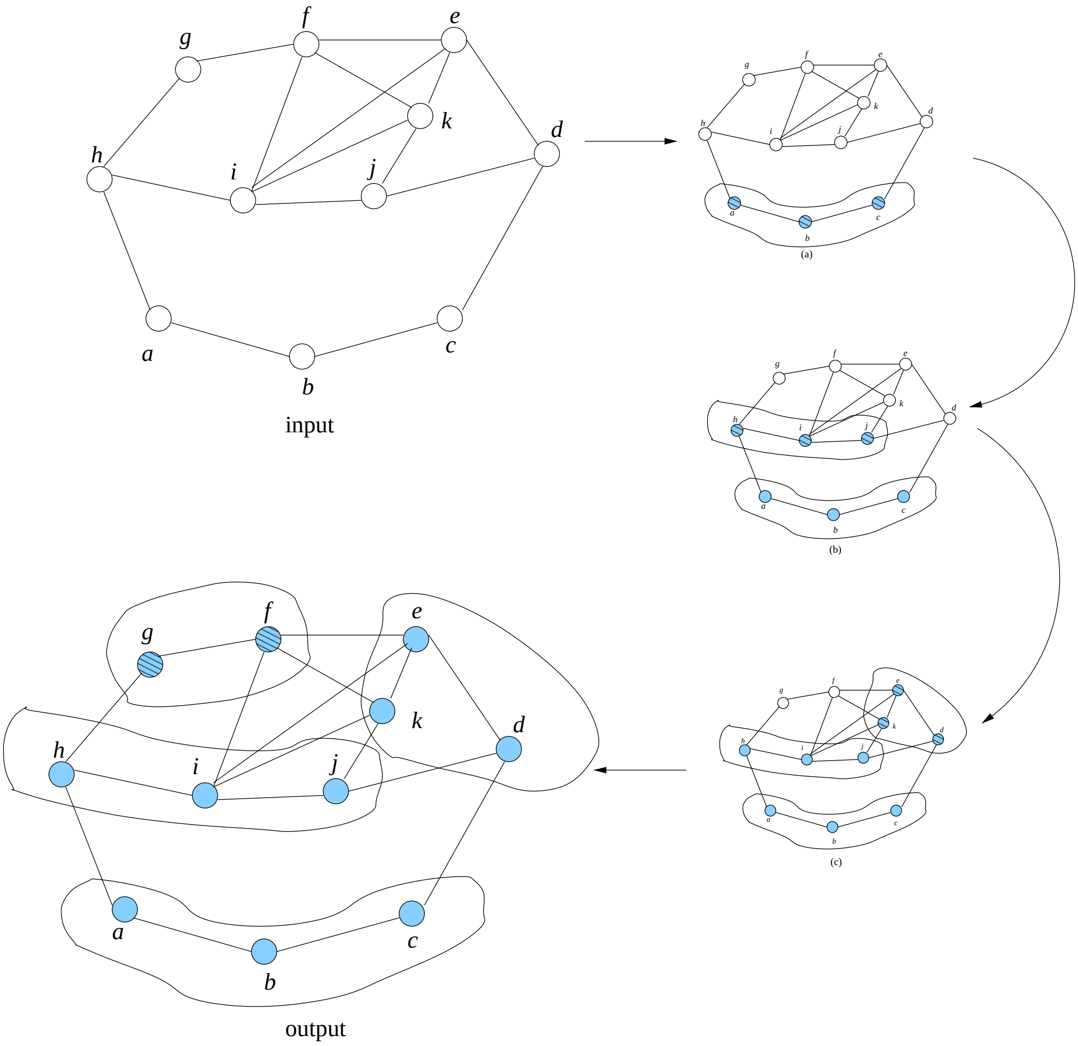

Figure 3.1: Model of direct marketing problem

3.3 Seeding Strategy Problem

Viral marketing is cheaper than any other marketing. It is now-a-days widely used to spread information in social network. But initialization of a viral marketing campaign is not easy. Company needs to find those active persons who are suitable to spread the product information. These persons can vary according to the nature of the product and the state of current usage of this type of product. Here we define a problem to find these persons in a social network where users already uses this type of product.

3.3.1 Problem Definition in Business Perspective

Let company A wants to advertise its new product to prospective customers. But there are some people in the network who are already convinced by other business rivals. It might be difficult to convince the people adjacent to the already convinced people. The problem is to find out people who are more prospective customers to be convinced.

3.3.2 Problem Definition in Graph Theoretic Approach

Let $G = (V, E)$ be a simple graph representing the social network where each vertex $v \in V$ represents a customer in marketing perspective and there exists an edge $(u, v) \in E$ if the customers corresponding to vertices u and v are acquaintances. Let C be the set of vertices corresponding to the persons convinced by rivals. The problem is to find a set of k vertices $S \in V - C$ such that they are in longest distances from the vertices in C.

Figure 3.2 demonstrates the graph theoretic model of seeding strategy problem. Here input is a graph where a and i have already been convinced by the business rivals. Output is a set S of such vertices that are in longest distances from a and c. In this example $S = \{g, f\}$.

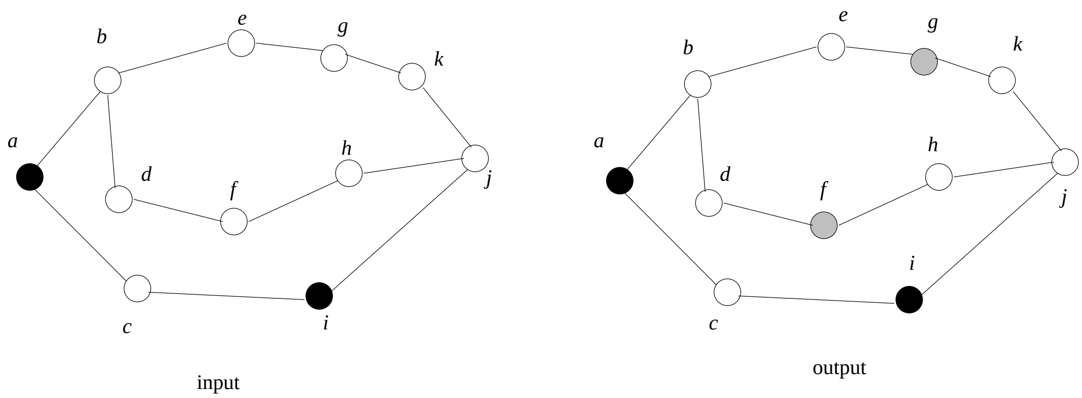

Figure 3.2: Model of seeding strategy problem

3.4 Stop Propagating Problem

In social network marketing, easier spreading of information helps to build reputation of organizations, but on the other hand, spreading of harmful information can ruin the achieved reputation within a short time. Even after overcoming from any deficiency of a product, if propagation of harmful information has not been stopped explicitely, it continues spreading. There is a practical importance to stop the propagation of harmful information which have already been started. Here we formulate the problem in business perspective as well as in graph theoretic approach.

3.4.1 Problem Definition in Business Perspective

Let a product p of manufacturer A has a deficiency d. The harmful information i: p contains deficiency d has already been started to propagate through social network. Now A has removed the deficiency d from product p and decided to stop the propagation of i by sending another information $anti - i$: p has been overcome from the deficiency d. The problem is to find out the list of minimum persons to whom A will send $anti - i$ to stop propagation of i.

3.4.2 Problem Definition in Graph Theoretic Approach

Let $G = (V, E)$ be a simple graph representing the social network where each vertex $v \in V$ represents a customer in marketing perspective and there exists an edge $(u, v) \in E$ if the customers corresponding to vertices u and v are acquaintances. Let the set of infected people H. The problem is to find out the set of vertices containing minimum elements that can stop propagating i that can save maximum no of unaffected vertices.

In Figure 3.3, we can see that input of this problem is a graph $G = (V, E)$ and a set of infected vertices $H = \{a, b, c, j, r\}$. As the problem is to find out the set of vertices containing minimum elements that can stop propagating, in this example output is a blocker set of vertices $B = \{g, q\}$. If B can be blocked, propagation will be stopped.

3.5 Target Market Finding Problem

The first step of marketing is to find out the niche market. Fnding out the niche market is known as target marketing. Formally a target market is a group of customers that the business has decided to aim its marketing efforts

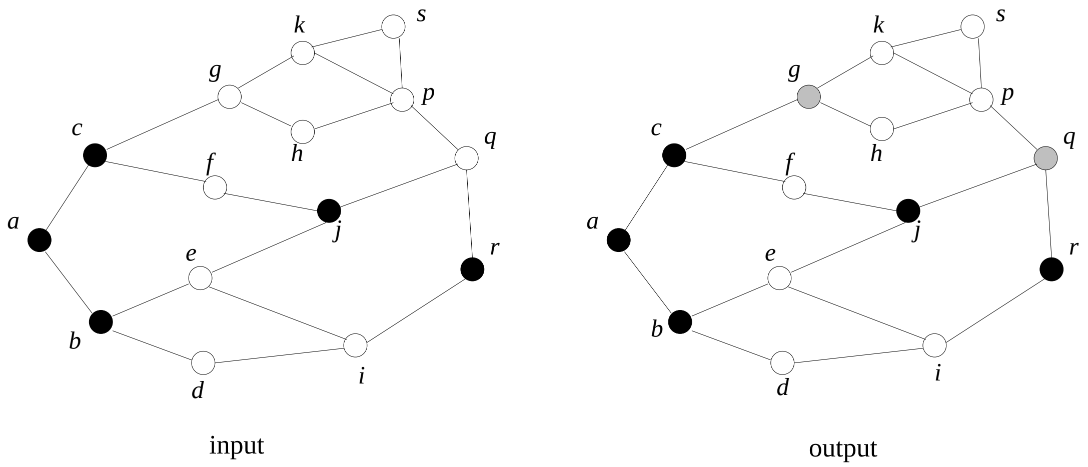

Figure 3.3: Model of stopping propagation problem

and ultimately its merchandise towards [24]. Target marketing offers a marketer many benefits to make the most of his marketing efforts and helps to ensure he will get results. The importance of target marketing in the online environment cannot be denied.

The first step of the target marketing process is to segment the market into groups of people with common characteristics and common needs, or market segments. While market segmentation can be performed on any large group of prospective buyers, such as the general market, the most logical approach is to identify the market segments that comprise the natural markets, which are those groups of people with whom marketers have a natural access due to similar values, lifestyles, experiences, attitudes, and so on.

Similarly, marketing based on the social network data also needs to segment the network which is called cluster. Now we know the discovery of close-knit clusters in these networks is of fundamental and practical interest. But existing clustering criteria are limited in that clusters typically do not overlap, all vertices are clustered and/or external sparsity is ignored. We need a new criterion that overcomes these limitations by combining internal density with external sparsity in a natural way.

3.5.1 Problem Definition in Business Perspective

Let a company A has made a discount offer to consumer groups. Marketing manager of A has decided to spread this message through social network. But there are significant number of people exists who do not belongs to any group. On the other hand there also exists many people who belongs to multiple group. The problem is to find out the closely connected groups that can be overlapped and to exclude the people belongs to no group.

3.5.2 Problem Definition in Graph Theoretic Approach

Let $G = (V, E)$ be a simple graph representing the social network where each vertex $v \in V$ represents a member in marketing perspective and there exists an edge $(u, v) \in E$ if the customers corresponding to vertices u and v are acquaintances. The problem is to identify clusters that are internally dense and externally sparse, that can be overlapped and scattered vertices are not be clustered.

Figure 3.4 demonstrates the graph theoretic model of target market finding problem. Here we can see that input of the problem is a graph. In this input graph $\{a, b, c, d\}$ and $\{d, e, f, g\}$ are two closely connected group whereas h and i are loosly connected with the groups. The problem is to find out closely connected groups and exclude the vertices those are not likely group members of any group. Output is two connected components $\{a, b, c, d\}$ and $\{d, e, f, g\}$. Output also excludes the loosly connected vertices h and i.

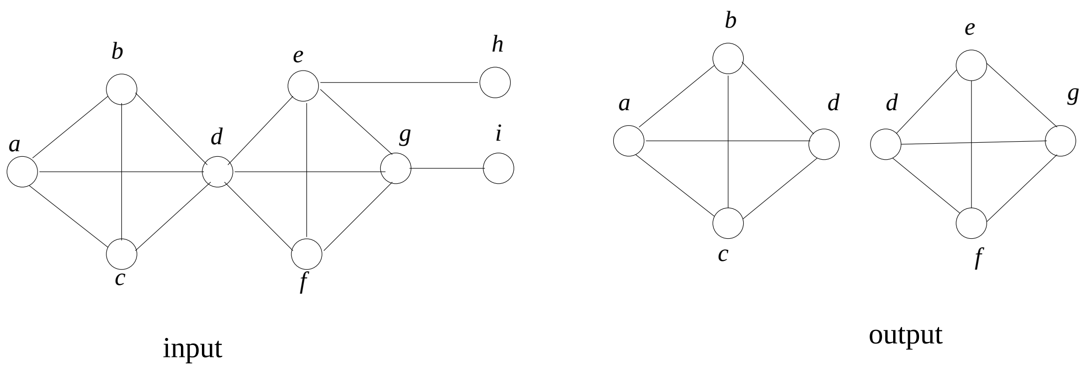

Figure 3.4: Model of target market finding problem

3.6 Most Influential People Finding Problem

Viral marketing includes marketing strategy that encourages individiuals to pass on a marketing message to others using the social network. This creates the potential of exponential growth of the message, as an individual passes on the message to his friends, those friends pass it on to their friends and so on. Viral marketing is popular because marketing campaign is relatively easy to implement, has a realatively low cost, and yields a high and rapid response rate.

There is a problem in executing a viral marketing campaign is finding a set of influential people from a group of people. More specifically, an advertiser may be unable to send its message to all the members of social network. So it is desirable for the advertiser to select a number of people in the social network who will maximize theumber of people who ultimaely recieve the advertising message.

3.6.1 Problem Definition in Business Perspective

Let a company A wants to advertise its new product through social network. A needs to find a set of k people S called seed. A will directly market the product to them. Then, assuming they adopt the product, A relies on their influence to generate a large cascade of adoptions. A does not rely on any further direct promotion of the product.

The problem is how A choose the set S that ensures maximum cascade of adoption?

3.6.2 Problem Definition in Graph Theoretic Approach

Let $G = (V, E)$ be a simple graph representing the social network where each vertex $v \in V$ represents a customer in marketing perspective and there exists an edge $(u, v) \in E$ if the customers corresponding to vertices u and v are acquaintances. Let there is a influence function $f(.)$ defined as follows: For a set S subset of V of nodes, $f(S)$ is the expected number of active nodes at the end of the process. Assuming that S is the set of nodes that are initially active. The problem is how large we can make $f(S)$ if we are allowed to choose a set S of size k?

3.7 Conclusion

This chapter is consists of the problems those are faced by the marketing people who works on social network. Every porblems of this list of problems are described in two perspectives. One is business perspective and another is graph theoretic approach. In each description, we first presented the general description of the problem then we presented the definition of the problem in business perspective and finally we presented the defintion of the problem in graph theoretic appraoch. Besides these list of problems there are still some marketing problems those can be mappable to graph theoretic problems.

Chapter 4

Seeding Strategy Problem

4.1 Introduction

Introducing a new product to the market is a challenging task for a business. Due to enourmous importance of launcing new products, researchers have concentrated their attention in formulating successful strategies for promoting new products to the market [4] [9] [18] [25]. In recent years many new product launch campaigns are adopting "word-of-mouth" strategy. In word-of-mouth strategy, for introducing a new product into a competitive market, managers consider a program in which an initial group of customer, which is called a *seed*, receives the product early on so that their word of mouth begins to drive sales [25]. In such a program, selection of appropriate seed is a crucial step for being successful in business [7] [26]. Domingos and Richardson [10] [11] posed a fundamental algorithmic problem for such systems as follows. Suppose that we have data on a social network, with estimates for the extent to which individuals influence one another we would like to market a new product that we hope to be adopted by a large fraction of the network. The premise of viral marketing is that by initially targeting a few "influential" members of the network, say, giving them free samples of the product, we can trigger a cascade of influence by which friends will recommend the product to other friends, and many individuals will ultimately try it. But how should we choose the initial individuals to use for seeding this process? In this chapter we focus on selection of seeds for word-of-mouth marketing on social networks. Furthermore, we consider a competitive market and assume that some of the customers in a social network are already convinced by the product owners of business rivals.

Social diffusion phenomena play significant role in selection of seeds for word-of-mouth marketing. In well studied models of social diffusion, some initial members of social network start diffusing ideas, innovations or information. According to the process of the models, information propagate through a social network. *Linear Threshold Model* [14] and *Independent Cascading Model* [15] [34] are two well studied diffusion models which are relevant to our work. According to *Linear Threshold Model*, at each step, an inactive node becomes active if the sum of the weights of the edges with active neighbors exceeds predefined threshold. In general we can say, at each step, an neutral customer becomes influential customer if his/her acquaintances collectively have enough influence factor to convince him/her [21]. In the cascade model, each individual has a single, probabilistic chance to activate each inactive node for which he is a neighbor after becoming active himself. A very simple example is the independent cascade model, in which the probability that an individual is activated by a newly active neighbor is independent of the set of neighbors who have attempted to activate him in the past [41] [21].

We consider a scenerio of having impact of business rivals in the network. We propose three schemes for the selection of seeds for word-of-mouth marketing motivated by the two diffusion models described above. Our first scheme is distance based, second scheme is influence based and the third one is a combination of them.

We now give a outline of our scenerio. Assume that company A wants to advertise it's new product to prospective customers. First Company A finds out the people already convinced by other business rivals (competitors). We will call them *occupied* throughout the chapter. According to *Linear Threshold Model* there is a high probability of person to be convinced for possessing any product that is already possessed by his/her acquaintances. Company A wants to find out the list of persons who have less connection to the occupied persons. Figure 4.1 illustrates a social network where Bob and Tom are convinced customers of the rivals of Company A. It is likely that Mosby will influenced by Bob or Tom. It will be difficult to convince Mosby. However, Ted and Penny are at a farthest distance from bob and tom, so it will be rather easy for Company A to convince them. So we consider Ted and Penny as prospective customers for

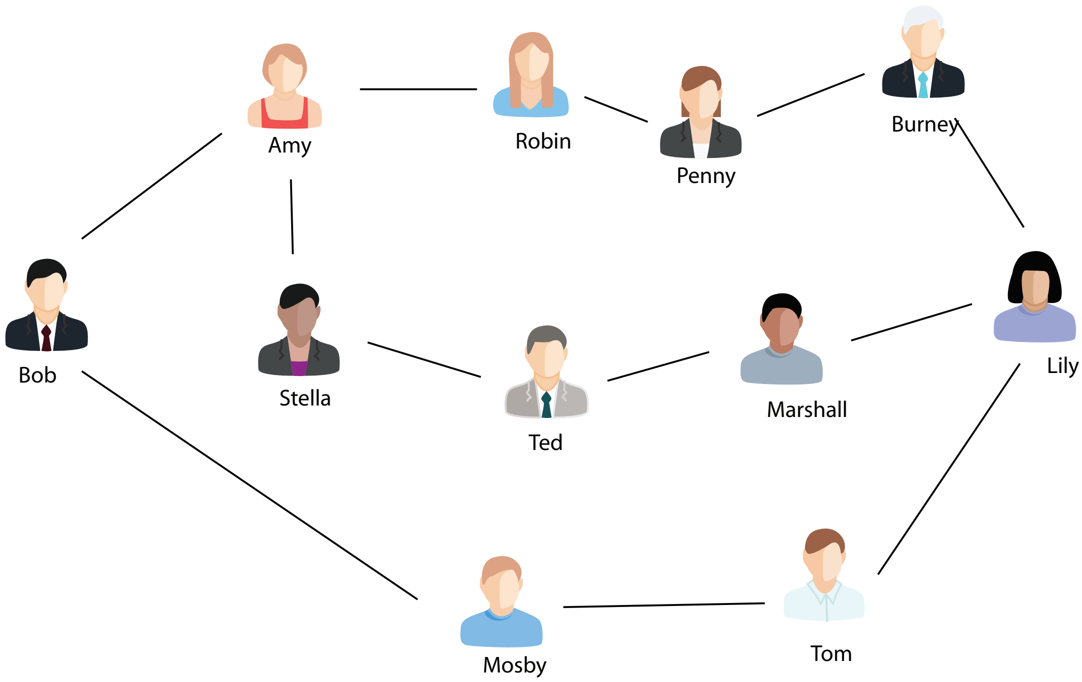

Figure 4.1: A social network

Company A. Our goal is to find such prospective candidates in a social network. We define a graph model for the problem as follows. Let $G = (V, E)$ be a simple graph representing the social network where each vertex $v \in V$ represents a customer and there exist an edge $(u, v) \in E$ if the customers corresponding to vertices u and v are acquainted with each other.

Let $C \subset V$ be the set of vertices corresponding to the persons convinced by rivals. Now the problem is to find k vertices in $V - C$ such that they are in longest distances from the vertices in C. Similar problems have been studied on Euclidian plane as "Voronoi diagrams" and "largest empty circle" for finding suitable location for new facilities where the distances are Euclidian distance [32]. However, our problem is different; in our case, we are studying the problem on social networks, namely on graphs.

The rest of the chapter is organized as follows. In section 2 we present the graph theoretic terminologies. In Section 3 we present our three schemes for selecting seeds for a word of mouth program on a social network. Finally Section 4 is a conclusion. Results presented in this chapter have appeared in [22].

4.2 Selecting Promising Clients

In this section we present three schemes of selecting promising clients based on social network data and considering the presence of business rivals.

4.2.1 Distance Based Scheme

In this section we give an algorithm to rank promising clients using BFS labeling.

Let $G = (V, E)$ be a simple graph representing the social network where each vertex $v \in V$ represents a customer and there exist an edge $(u, v) \in E$ if the customers corresponding to vertices u and v are acquainted with each other.

Let $C \subset V$ be the set of vertices corresponding to the persons convinced by rivals. Now the problem is to find k vertices in $V - C$ such that they are having longest distances from the vertices in C.

We first compute BFS labeling of the vertices of G taking each of the vertices of C as the source vertices. For each vertex we then choose the label which is minimum among all the labels obtained by taking the vertices in C as source vertices. In the graph in Figure 4.2(a) the vertices in C are a, l and q. BFS labeling of graph G is obtained taking a as the source vertex in Figure 4.2(b). Similarly BFS labeling is obtained taking l and q as the source vertices in Figures 4.2(c) and (d) respectively. The minimum labels among the labels in Figures 4.2(b) (c) and (d) are shown in Figure 4.2(a). We consider the minimum label of each vertex as the rank of the vertex, and finally list the vertices in descending order of their rank.

4.2.2 Influence Based Scheme

According to *Independent Cascade* model, one active person u is given a single chance to activate each currently inactive neighbor w; it succeeds with a probability $p_{v,w}$. In general terms people having more *social influence* have more probability to succeed. *Social influence* is a phenomenon that the actions of a user can induce his/her friends to behave in a similar way [1]. Epidemiology studies indicate that vertex having more connection foster the spread of diseases [2], which suggests in parallel that vertex having more connection should be more attractive for seeding viral marketing campaigns. Backstrom et al. [3] examined the membership problem in an online community. They observed

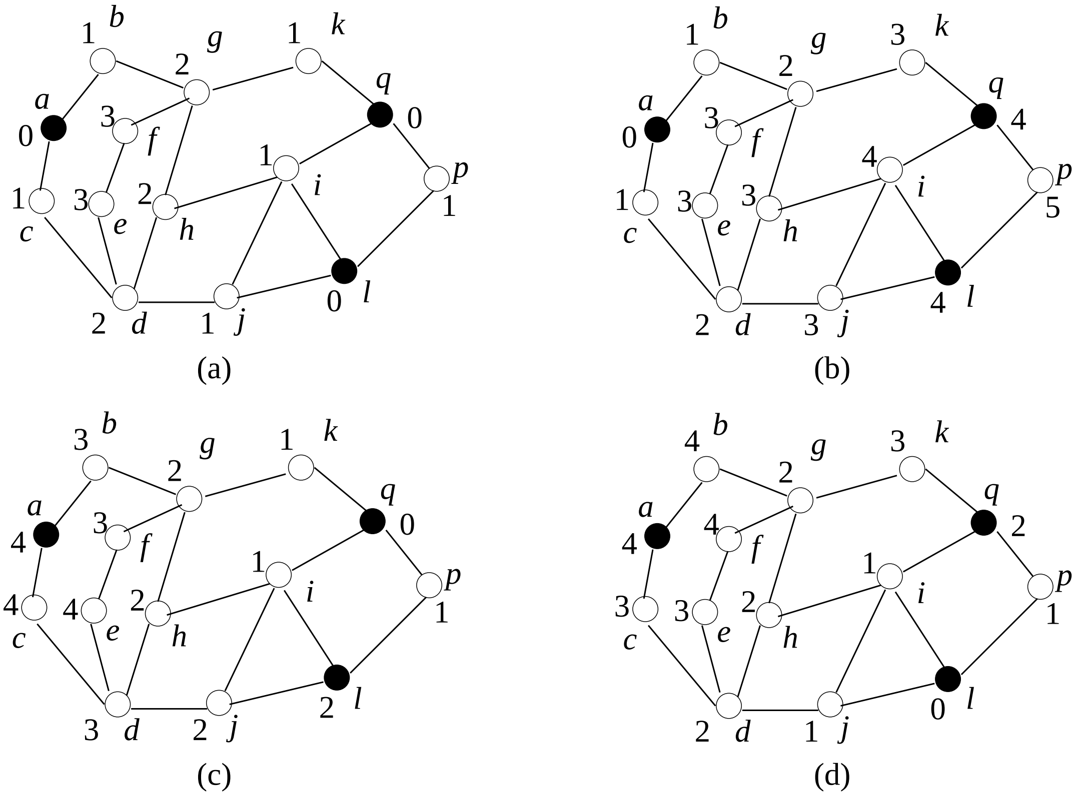

Figure 4.2: (a) A graph G with 3 occupied vertices a, l and q, (b) G with BFS labeling taking a as the source vertex (c) G with BFS labeling taking l as the source vertex, and (d) G with BFS labeling where q as the source vertex.

correlation between the action of a user joining an online community and the number of friends who are already members of that community. From reverse perspective, we can say that users having more affiliation have more influence in social network. Here comes the necessity of user ranking considering degrees of vertices. We propose a scheme where we first delete the occupied vertex and then label them according to their degrees $l_u = d(u)$ here $d(u)$ is degree of vertex u after deleting the occupied vertices in the graph. Influence based labeling is illustrated in Figure 4.3; we label the vertices by their degrees after deleting the occupied vertices.

4.2.3 Distance and Influence Based Scheme

In this section we give a scheme for ranking promising clients which is a combination of the two schemes above. We consider the following three cases.

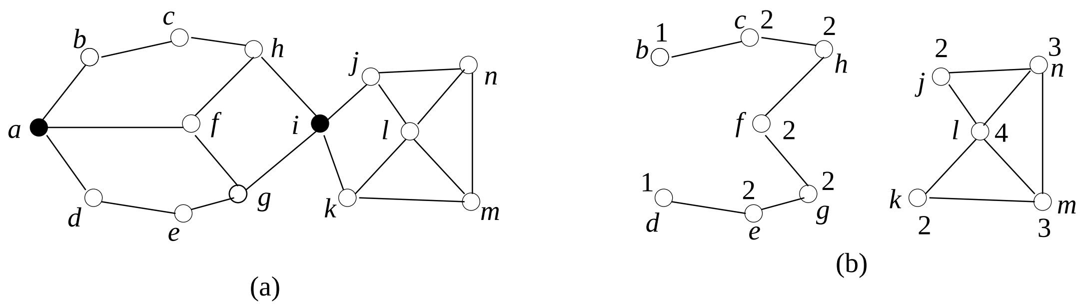

Figure 4.3: (a) A graph G with 2 occupied vertices and (b) influence based labeling after deleting the occupied vertices.

Case 1. *Some vertices are occupied by other business rivals.* We first compute a weight w_1 for a vertex u considering the distance based scheme as follows. Let L be the set of labels obtained by BFS labeling, let $max(L)$ be the maximum label in L and let the label of vertex u be l_u. Then we compute $w_1 = \frac{l_u}{max(L)}$ which can have value at most 1.

We now compute a weight w_2 considering influence based scheme. We first delete all the occupied vertices in the graph. Let Δ be the maximum degree of the graph after deleting the occupied vertices. We compute $w_2 = \frac{d(u)}{\Delta}$.

Taking weights w_1 and w_2 we finally formulate the function $f(r) = C_1w_1 + C_2w_2$ for ranking of vertices, where C_1 and C_2 are the variables satisfying the conditions $0 < C_1, C_2 \leq 1$ and $C_1 + C_2 = 1$. In different networks the values of C_1 and C_2 will vary. For a default case, we take $C_1 = C_2 = 0.5$

Case 2. *No vertex is occupied by other business rivals.* In this case there is no source to start BFS labeling. Only degree of vertices considered for ranking of vertices.

Case 3. *All the vertices are occupied by other business rivals.* In this case, after BFS labeling we will get all the labels 0. Here also degree of vertices are considered for ranking of vertices.

In case 2 and 3, we can choose a vertex of the maximum degrees as a seed. But if we want multiple seeds, we choose the vertices that are far from each other to minimize the overlapping of their influence domain.

4.3 Conclusion

In this chapter we have presented three schemes for selecting seeds for word of mouth marketing on social networks. Our distance based scheme works when there are some customers in the social network who are already convinced by the rival. The influence based scheme works fine even if no customer is convinced by rivals.

Launching a new product needs extensive marketing and to make it successful finding target market is mandatory. If the marketing media is social network which is widely used now a days, to find out the promising clients we can use this algorithm and start campaigning by making them convinced.

Chapter 5

Direct Marketing Problem

5.1 Introduction

Direct marketing is a form of advertising in which physical marketing materials are provided to consumers in order to communicate information about a product or service. Direct marketing does not involve advertisements placed on the internet, on television or over the radio. Here marketers communicate straight to the customers. Direct marketing is practiced by businesses of all sizes from the smallest startup to the leaders. A well executed direct advertising campaign can prove a positive return on investment (ROI) by showing how many potential customers responded to a clear call-to-action. Even well designed general advertisements rarely can prove their impact on the organizations bottom line.

Social network has a strong impact on the possibility of buying a product by customers. If friends of a customer become a consumer of a product, it becomes easier to convince the customer to buy the product. So in direct marketing it is better to choose the potential customers who have friends those are already convinced or at least approached earlier.

Marketing managers face problems to optimize the cost while executing campaign. They have to manage limited marketing representatives in a such way that marketing representatives will cover the network with minimum campaigning round. Each marketing representative is supposed to go to a potential customer in each round of campaign. They choose the potential customer from the list of candidate potential customers who have convinced or approached friends. If the number of candidate potential customers are less than the number of representatives in a particular round, some of the representatives will stay idle in that round. Marketings managers objective is to send marketing representatives in such a way that minimizes the number of campaigning round.

We refer this type of campaigning on social network as *direct marketing camapigning model* in this chapter. In this model every graph is not equally suitable for campaigning. Some networks require less number of rounds where as some networks require more number of rounds to cover the whole network. It is due to the structure of the networks. In this chapter we identify graph classes based on suitability of campaigning. Achieving optimum number of rounds to cover the whole network depends not only on the number of representatives but also on the structures of the graphs. Another important factor to achieve the optimum number of rounds is initially accessible customers. Some campaignable graphs can become campaign hard graphs if initially accessible customers are predifined and cannot be changed. In this chapter we describe the graph classes and some properties to identify graph classes. We also derive an algorithm to achieve optimum number of round on a campaignable graph.

In Section 5.2 we describe the problem, elaborate it with an example. In Section 5.3 we define the graph classes. Properties to identify the graph classes are described in Section 5.4. Section 5.5 presents two case study where first one shows whether we need to choose initially accessible customers from minimum dominating set or not and second one shows that graphs of minimum degree k are not always coverable optimally where k is the number of representatives. Algorithm to achieve optimum number of round on a campaignable graph is depicted in Section 5.6. Finally Section 5.7 concludes the chapter.

5.2 Problem Statement

In this section we formulate a social network marketing problem where company send their representatives to potential customers directly. But before sending them company select potential customers from social network data. Here we also present corresponding graph theoretic problem after being mapped from social network marketing problem.

Assume that company A wants to advertise it's product to prospective customers. They choose a network over which marketing campaign will be hold.

They have the social network data from which they know relationships between the persons of the network. In the 1st round the company sends its representatives to some initially accessible customers. From the second round the company can only send their representative to the customers who have some acquaintance already reached by the representative of the company in some previous round. We refer this campaigning models as *direct marketing campaigning model.* The company wishes to advertise on a social network of n people. Here comes two questions

1. If the company has $k < n$ representatives then how many rounds are necessary to cover the whole network?

2. How to choose customers in each round to minimize the number of rounds?

A lower bound $r_o = \lceil \frac{n}{k} \rceil$ on the number of rounds is trivial. Let r_m be the minimum number of rounds necessary to cover the network. Then $r_m \geq r_o = \lceil \frac{n}{k} \rceil$. However, it is not clear whether it is possible to achieve the lower bound or not.

The answer of the second question is not easy. In each round it needs to analyze the current state of candidate vertices and choose accordingly. We are going to map the problem into graph theoretic problem now.

Let $G = (V, E)$ be a simple graph representing the social network where each vertex $v \in V$ represents a customer in marketing perspective and there exists an edge $(u, v) \in E$ if the customers corresponding to vertices u and v are acquaintance. Here Figure 5.1 represents a social network of $n = 11$ people and number of marketing representatives is $k = 3$. Hence at least $r_o = \lceil \frac{n}{k} \rceil = \lceil \frac{11}{3} \rceil$ $= 4$ rounds are necessary.

Let a, b, c are the initially accessible persons. In first round $r = 1$, Company A will send its $k = 3$ marketing representatives to 3 initially accessible persons a, b and c. In second round $r = 2$, it can send its representatives to any 3 of h, i, j and d. We call this set of vertices as candidate vertices of a particular round, so set of candidate vertices in round 2 is $\{h,\ i,\ j,\ d\}$. Say it sends its representatives to h, i, j. Now the set of candidate vertices is $\{d,\ e,\ f,\ g,\ k\}$. From this set it can choose any 3 say d, e, k and sends representatives to them in third round $r = 3$. Finally in round four $r = 4$, company sends to g and f.

Now we can see that, for this example and if the company chooses candidate

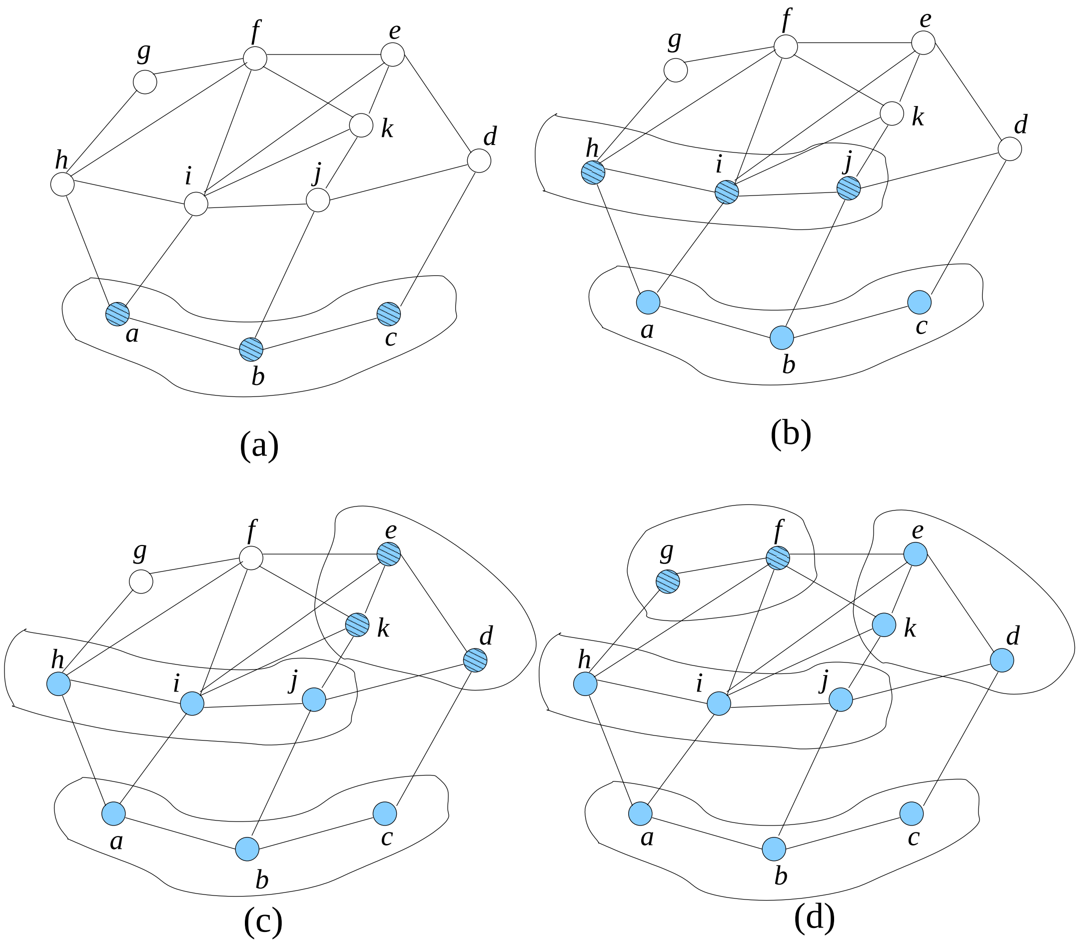

Figure 5.1: a) Round 1: approach to initially accessible customers b) Round 2: Approach to h, i and j c) Round 3: Approach to d, e and k d) Round 4: Approach to g and f

vertices in this way it can cover the whole network within 4 rounds which is optimal.

But if the network expands more and if the company chooses candidate vertices in this way it cannot always cover the whole network in optimal round. In Figure 5.2 we extends the previous network and now $n = 15$. So the optimal number of round is $r_o = \lceil \frac{n}{k} \rceil = \lceil \frac{15}{3} \rceil = 5$. But if the company choose the candidate vertices in similar way, company needs to send its representative 6 times to cover the whole network. Fifth and sixth rounds are depicted in Figure 5.2(a) and (b) respectively.

Here company A wants to cover the network in 5 rounds rather than 6. So the problem is to develop an algorithm to choose customers in each round to

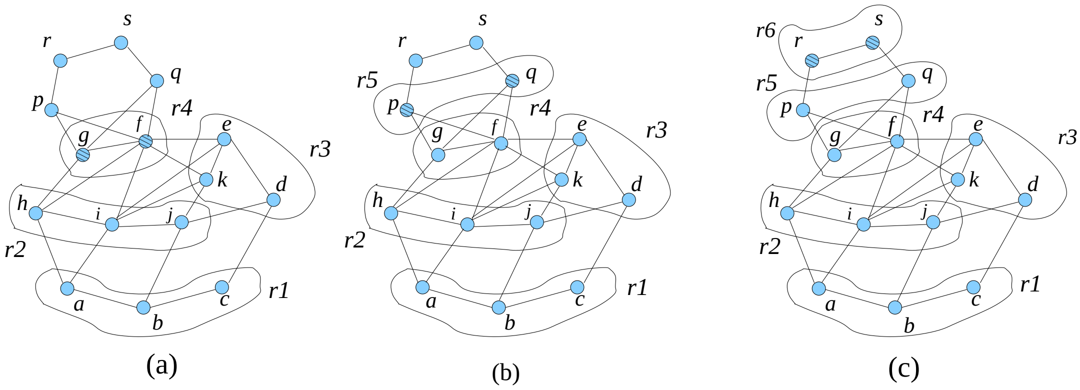

Figure 5.2: a) Round 4: Extended example of the network Figure 5.1 b) Round 5: Approach to p and q c) Round 6: Approach to r and s

minimize the number of rounds.

5.3 Graph Classes

In this section we define graph classes base on suitability of direct marketing campaigning. Here we classify graphs in three classes.

- *Campaign friendly*
- *Campaignable*
- *Campaign hard*

i. Campaign friendly: These are the graphs on which any arbitrary way to campaign is optimal. Example: Every complete graph is *campaign friendly*. In Figure 5.3 for any initially accessible vertices the whole network can be covered in optimum number of rounds.

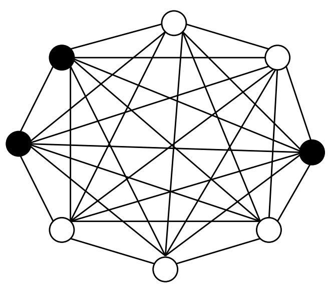

Figure 5.3: Campaign friendly

ii. Campaignable: These are the graphs on which it is possible to campaign in optimum round if the candidate vertices are chosen appropriately. Graph of Figure 5.4(a) is a campaignable graph because it is possible to campaign in optimum number of rounds if the candidate vertices are chosen appropriately as shown in Figure 5.4(b).

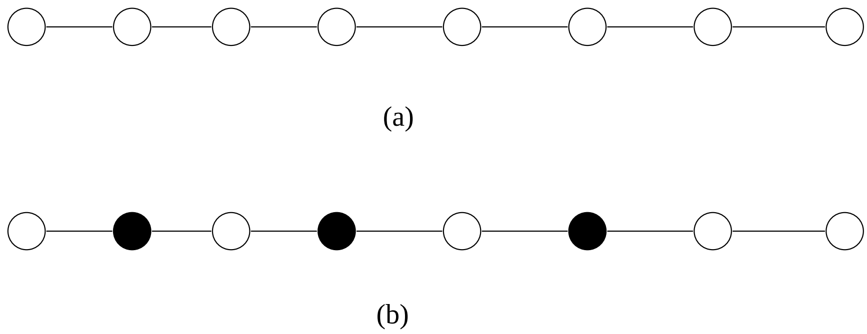

Figure 5.4: Campaignable

iii. Campaign hard: These are the graphs on which it is impossible to campaign in optimum number of rounds.

5.4 Properties for Identifying Graph Classes

In this section we present some properties of graphs from which one can identify the class of a graph. Identifying the class of a graph is an interesting topic. Not only the structure but also the number of representatives is the key factor of the identification of a graph class. We will find out an algorithm to identify the class of a graph in fututre research. But there are some cases where one come up with decision that the considering graph belongs to which graph. Followings are the cases, where the properties proclaim the graph is optimally campaignable or not.

- If $k = 1$, it is always optimally campaignable for every connected graph.
- If G is a path and initially accessible vertices are in every n/k th position, it is always optimally campaignable. Here n refers to the number of persons in network and k refers to number of representatives. Figure 5.5

depicts this case where $n = 9$ and $k = 3$. If the initially accessible vertices are in 3rd, 6th and 9th postion, it is always campaignable.

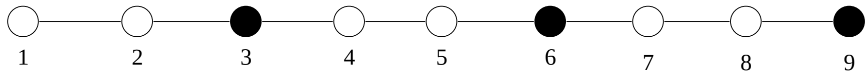

Figure 5.5: Property of position

- A *complete graph* is always optimally campaignable. In every round there will be enough candidate vertices to approach. So whatever vertices are chosen, it will always be optimally campaignable.
- If G is a cycle and number of representatives $k = 2$, it is always optimally campaignable. In Figure 5.6, we can see this case where there will not be a short of candidate vertices in any round.

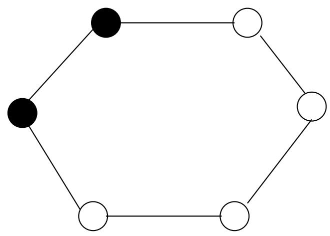

Figure 5.6: Property of cycle

- Graphs of minimum degree 3 are not always campaign friendly if the number of representatives $k = 3$. As every vertex of these type of graph have atleast 3 neighbors, it biases the mind that there will not be short of candidate vertices in any round. But actually there are examples of short of candidate vertices in this case. We depicted this case in Figure 5.8 where $n = 15$, $k = 3$ and number of round $r = 6$ which is larger than optimal number of round $r_o = 5$.

5.5 Case Studies

In this section we present two experiments for finding some properties of graphs base on suitability of campaigning. In first experiment we choose initial vertices from minimum dominating set of a graph to check whether it gives optimum solution or not. In second experiment we consider a graph in which minimum number of degrees of every vertex is 3. We experimented to verify that if number of representatives is $k = 3$, the whole network can be covered in minimum number of rounds. But we found an example showing that any arbitrary way to campaign is not optimal.

5.5.1 Choosing from MDS

In this section we will present an example showing that, choosing initially accessible persons from MDS does not give the optimum solution always.

In general perspective, as all the vertices are connected with the vertices from MDS, it biases the thinking of choosing from MDS. We have seen from Section 5.2, in different cases choosing initially accessible vertices from MDS results optimum number of rounds. But for case 4 it does not give always the optimum result. Figure 5.7 shows an example of choosing from MDS and other than MDS where filled vertices are in set of MDS.

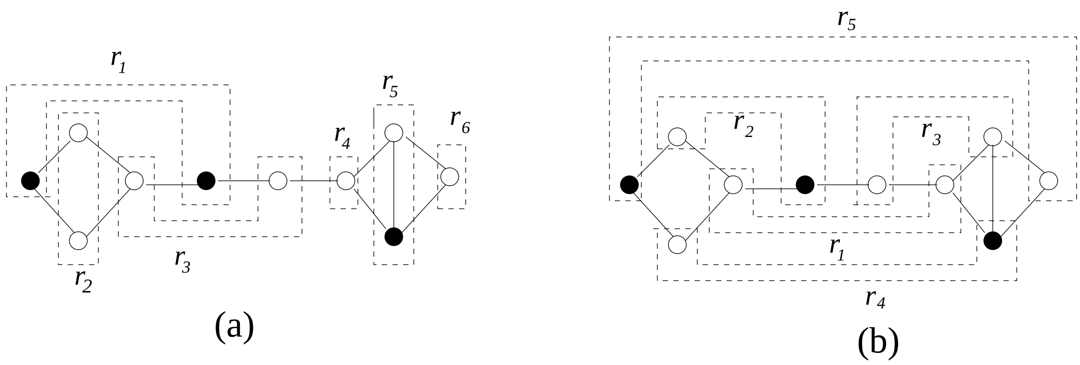

Figure 5.7: a) Choosing the seeds from MDS and spread thereby b) Choosing the seeds from other than MDS and spread thereby

From Figure 5.7 (a) needs 6 rounds to cover all the vertices whereas (b) needs 5 rounds to cover all the vertices. So we can say that choosing seeds from MDS may not give the optimum result.

5.5.2 Graphs of Minimum Degree k

In this section we present another example to show that, a graph is not a campaign friendly graph even if all the vertices of the graph have minimum k degrees.

From Figure 5.7(a) we can see that it is taking more than minimum number of rounds to cover all the vertices. We are giving another example of it because we will establish an algorithm to choose the potential clients in such a way that we can cover all the vertices in minimum number of rounds of campaign.

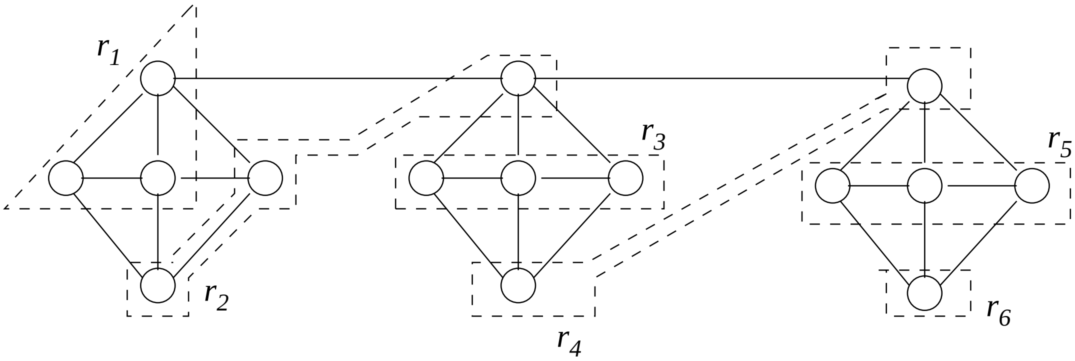

Figure 5.8: Rounds of campaign where minimum degree of the graph is greater equal than maximum limit of approaching.

Here, number of degree $n = 15$ and maximum limit of approaching before $k = 3$. So minimum number of round is $\frac{n}{k} = \frac{15}{3} = 5$. But approaching to wrong client has made the number of round from 5 to 6.

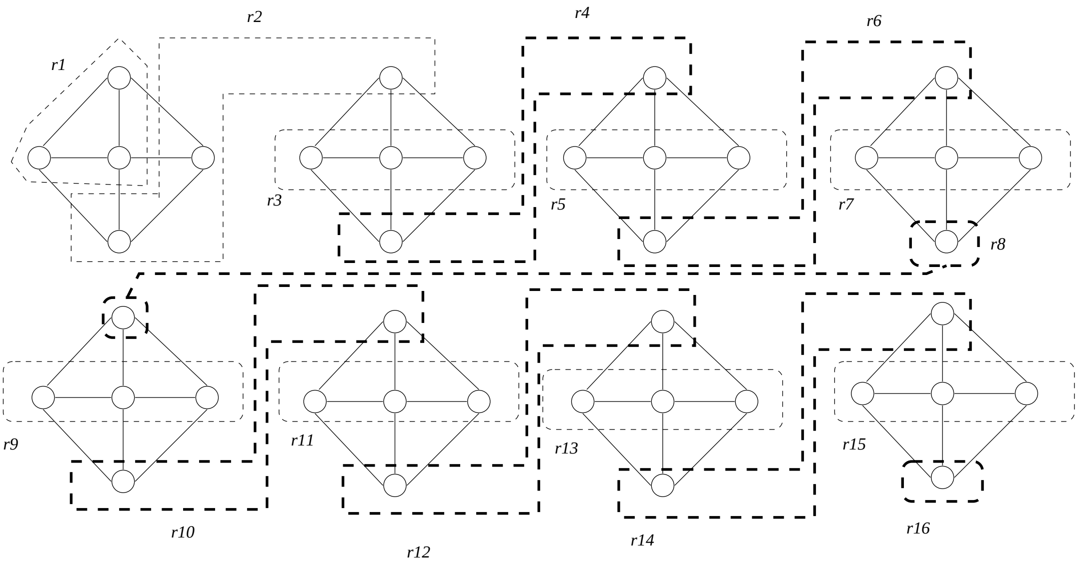

Figure 5.9: Pattern of having short of candidate vertices.

If we extend the graph with similar pattern, there will arise a short of candidate vertices in every next rounds and number of rounds will be increased by 1 for every 20 new vertices.

5.6 A Heuristic Algorithm for Covering Graph

In this section we present a heuristic algorithm which we have developed to choose vertices in such a way that entire graph can be covered in minimum number of rounds. We also show an example how this algorithm works. We coded this algorithm in php and comparisons of minimum number of round and number of round obtained from the algorithm for several input graphs are shown in tabular format. We provide the complexity of the algortihm at the end of this section.

5.6.1 Algorithm

According to this algorithm we need to find all cut vertices of the considering graph, say G. Let the set of cut vertices is C^G for graph G. For initial step we choose vertices from this set C^G. If number of cut vertices is smaller than number of representatives, we choose rest of the vertices according to number of degree. From the next steps we choose those vertices which have maximum degrees of untraversed/open vertices until the whole network has been covered. We call this algorithm k**-Cover** algorithm.

Algorithm k-Cover
Input: A graph G and number of representatives k.
Output: Collection of best chosen vertices set S in every step to cover the whole network in minimum round.
begin
Find all cut vertices of G, say C^G is the set of cut vertices of graph G;
if $|C^G| >= k$ **then**
    Choose initial k vertices from any element of C^G;
else
    Choose initial $|C^G|$ vertices from any element of C^G;
    Sort $V - C^G$ according to number of degree in descendent order;
    Choose first $k - |C^G|$ vertices from sorted $V - C^G$;
while !covered all vertices, **do**
  begin
    increment number of rounds r.

Choose the set of vertices S_r of having maximum degrees of open vertices.

Push this set S_r to S.

end

end.

We now present an example where we will follow these steps and it will eventually cover the graph in minimum rounds. In Figure 5.4 we can see that $n = 22$ and say $k = 3$. Optimum number of rounds $r_o = \lceil \frac{22}{3} \rceil = 8$. Here set of cut vertices is $C^G = \{f, g, l, r\}$.

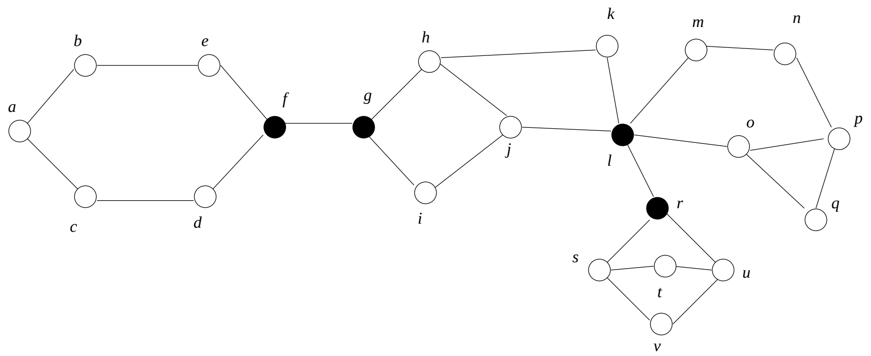

Figure 5.10: Cut vertices of a graph

Let the set of candidate vertices of a round r is denoted as C_r and set of chosen vertices of a round r is denoted as R_r.

Round 1:

$C_1 = C^G = \{f, g, l, r\}$

$R_1 = \{f, g, l\}$

Round 2:

$C_2 = \{d, e, h, i, k, m, o, r\}$.

It can be represented like this after adding the number of open vertices as superscript on the name of vertices.

$C_2 = \{d^1, e^1, h^2, i^1, j^2, k^1, m^1, o^2, r^2\}$

After sorting descendently by number of open vertices

$C_2 = \{h, o, r, d, e, i, k, m\}$

$R_2 = \{h, o, r\}$

Round 3:

$C_3 = \{d^1, e^1, i^1, j^1, k^1, m^1, p^2, q^1, s^2, u^2\}$ After sorting descendently by number of open vertices

$C_3 = \{p, s, u, d, e, i, j, k, m, q\}$

$R_3 = \{p, s, u\}$

Round 4:

$C_4 = \{d^1, e^1, i^1, j^1, k^1, m^1, n^1, q^1, t^1, v^1\}$

As all element of C_4 have same number of open vertices, we can choose any k of them $R_4 = \{d, e, i\}$

Round 5:

$C_5 = \{b^1, c^1, j^1, k^1, m^1, n^1, q^1, t^1, v^1\}$

Similar as round 4

$R_5 = \{b, c, j\}$

Round 6:

$C_6 = \{a^1, k^1, m^1, n^1, q^1, t^1, v^1\}$

$R_6 = \{a, k, m\}$

Round 7:

$C_7 = \{n^1, q^1, t^1, v^1\}$

$R_7 = \{n, q, t\}$

Round 8:

$C_8 = \{v^1\}$

$R_8 = \{v\}$

Thus according to the algorithm in this example we can cover whole network in minimum number of rounds.

We now present time complexity of the algorithm. We have the following theorem.

Theorem 5.6.1 *Algorithm k-**Cover** takes $O(n^2 logn)$ time.*

Proof. Let G be a graph and k be the number of representatives. Finding cut vertices in step 1 of the algorithm can be achived in linear time. If we use tarjan's algorithm [35], time complexity of finding cut vertices will be $O(n+m)$ where n and m are number of vertices and number of edges of G respectively. In step 2, choosing initial k vertices from cut vertices can be achieved in linear time as complexity is $O(k)$. Time complexity of sorting in step 3 is $O(nlogn)$ using marge sort. Upto this, all steps can be achieved in linear time but step 5 to 7 will iterate n times and step 6 requires a sorting. The time complexity of step 4 to 7 is $O(n^2logn)$ which is polynomial. Therefore, we can say that whole network can be covered in polynomial time. $\mathcal{Q.E.D}$.

5.6.2 Experiments

We now present a experimental comparison of optimal number of rounds and number of rounds obtained from Algorithm k-**Cover**. We use Rome library (only undirected) and a sub graph of facebook network from stanford large network dataset collection as standard graph drawing test library. We present here randomly chosen 5 graphs from rome library and 3 graphs from facebook network as input and execute our code for Algorithm k-**Cover**. Table 5.1 depicts the results after execution the code. Figure 5.11 depicts the input graphs in consideration.

Let number of vertices of the input graph be n, number of representatives be k, optimal number of rounds be $r_o = \lceil \frac{n}{k} \rceil$, number of rounds obtained from the algorithm be r and number of rounds obtained after randomly chosen be r_r.

5.7 Conclusion

This chapter consists of the problem description regarding direct marketing based on social network data. First we have elaborated the problem and define a model of direct marketing campaign. Then we have defined three graph classes base on the suitability of direct marketing. We described general properties of graphs to identify the graph classes. After that we have presented an example where choosing initially accessible persons from MDS is not optimal and we

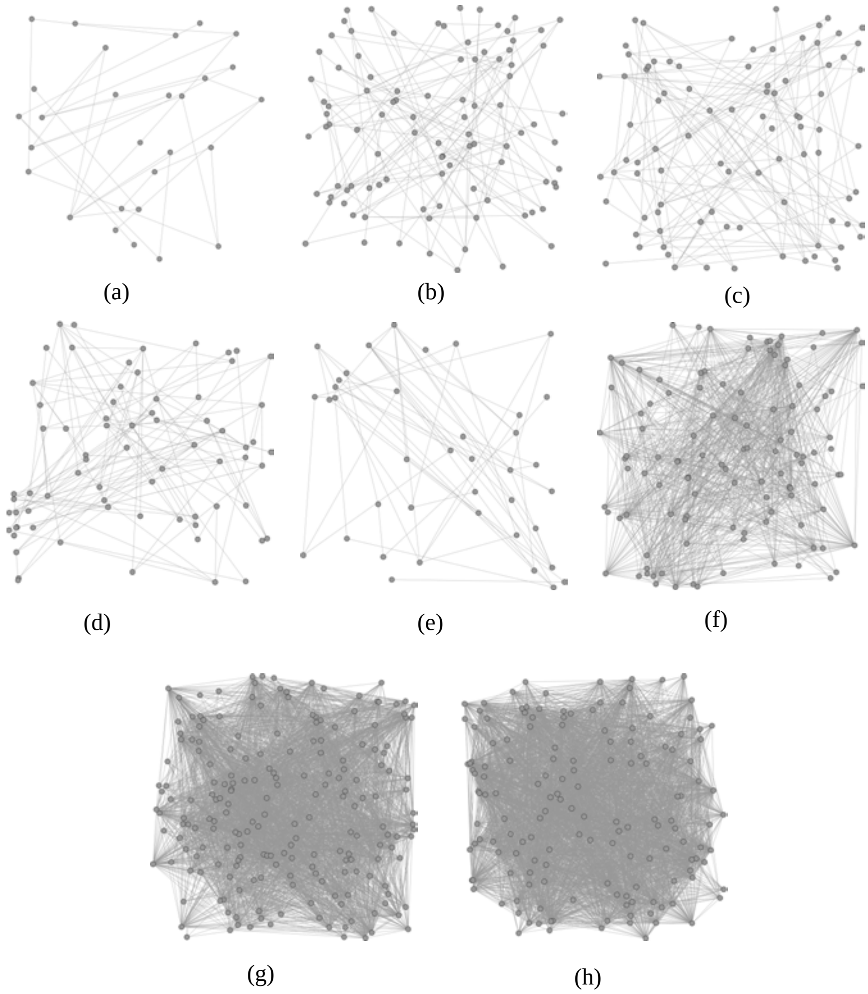

Figure 5.11: Input graphs (a) input graph #1 (b) input graph #2 (c) input graph #3 (d) input graph #4 (e) input graph #5 (f) input graph #6 (g) input graph #7 (h) input graph #8

have also investigated and found that graph of minimum degree k is not always optimal where number of repsentatives is k. We describe a pattern of minimum degree-k network structure where number of round is directly propotional to number of this pattern. With the increase of this pattern, number of rounds increases. We have developed an algorithm to choose vertices in a way that makes the number of rounds optimal.

Input	Source	n	k	r_o	r	r_r
#1	Rome library	28	3	10	10	10
#2	Rome library	97	3	33	33	33
#3	Rome library	89	3	30	30	31
#4	Rome library	72	3	24	24	25
#5	Rome library	39	3	13	13	15
#6	Facebook network	113	3	38	38	39
#7	Facebook network	203	3	67	67	67
#8	Facebook network	153	3	51	51	51

Table 5.1: Comparison table

Chapter 6

Stop Propagating Problem

6.1 Introduction

Harmful information often cause very serious damage on state of reputation of a brand or a product. Companies always try to be aware about preventing disclosure of harmful information. Inspite of their awareness there are cases where harmful information starts propagating. If any company cannot prevent such propagation, it has to face a large amount of loss. They need to find out the persons who know the harmful information and convince them to stop propagating the information. As resources are limited, it is hard to convince all the persons who know the information. Company wants to minimize the list of persons to be convinced. Here comes the interest of study on minimization problem. In this chapter we develop an algorithm to find out such minimum number of persons that convincing them can stop the propagation. Before that we will formulate the problem as a graph theoretical problem.

In Section 6.2 we present some models of spreading information as well as spreading harmful information. Problem statement and elaboration of the problem with example are presented in Section 6.3. Our algorithm is presented in Section 6.4. Extension of the problem and limitation of the algorithm is described in Section 6.5. And finally conclusion is mentioned in Section 6.6

6.2 Models of Spreading Information

Stopping the propagation of harmful information is highly dependent on the model of spreading information. Models for diffusion of information in a network have been studied extensively in various disciplines, including computer science, sociology, physics and epidemiology. In this chapter, we consider two different types of models: general information propagation models and harmful information like gossip, negative advertising propagation model.
Two popular models for general information propagation are Linear threshold model and the independent cascade model, considered by Kempe et al [21]. According to *Linear Threshold Model*, at each step, an inactive node becomes active if the sum of the weights of the edges with active neighbors exceeds predefined threshold. In general we can say, at each step, an neutral customer becomes influential customer if his/her acquaintances collectively have enough influence factor to convince him/her [21].
According to the Independent cascade model, each individual has a single, probabilistic chance to activate each inactive node for which he is a neighbor after becoming active himself. A very simple example is the independent cascade model, in which the probability that an individual is activated by a newly active neighbor is independent of the set of neighbors who have attempted to activate him in the past [42] [21].

For the propagation of harmful information like gossip or negative rumour, well known models are SIR (Susceptible-Infected-Removed) model and SIS (Susceptible-Infected-Susceptible) model for epidemic spread [37]. According to the *SIR model*, a vertex may be in any of the following three states: Susceptible, in the case it does not have the harmful message/virus, but it can become infected if exposed to it; Infected, in the case that it has the harmful message and can pass it on; Removed (or Recovered), in the case that it used to have the virus, but it recovered (or died), and now it is permanently immunized and it no longer participates in the virus propagation process. According to the *SIS model*, we assume that a node may be cured from the virus, but it is not immunized, and thus it can become infected again. Therefore, a node alternates between the susceptible and immunized states.

Here we only consider the SIR model for propagating harmful information also known as virus on the network. This model describes the states of nodes in different time. No person will send harmful information after being immunized. But for propagating the information we consider here the Independent Cascade Model where the probability is always 1.

6.3 Problem Statement

Now we will formulate the problem in business perspective:

Let a product p, say a laptop of manufacturer XYZ has a deficiency d say has a "shorter battery life". The harmful information i say "p has a shorter battery life problem". This information i has already been started spreading over the social network. The media of propagating the information is social networking sites and people on the network are propagating the information by sharing to their friends. According to *independent cascade model* each individual has a probabilistic chance to activate others. If the probability is maximum, each sharing of information can make all adjacent individuals convinced to share this information. In this case, we assume this probability is maximum.

Now, in this state of the network, XYZ has removed the deficiency d that means p has a longer battery life now. XYZ decided to stop the propagation of i by sending another message "p has been overcome from the deficiency d", we can call it *anti-i*. But there are two models of propagating harmful informations. Considering *SIS* model, only sending anti harmful information *anti-i* in single time is not enough. According to this model, individuals who have received *anti-i* may also propagate harmful information i, if he/she again receives i. But this model is not appropriate for this case as individuals does not send harmful information i second time after getting the information *anti-i*. *SIR* model is perfectly applicable in this case where any individual need not to be convinced twice to stop propagation of harmful information. Considering *SIR* model, the problem is to find out the list of minimum persons to whom XYZ will send this *anti-i* information to stop propagation of i.

In graph theoretic approach Let $G = (V, E)$ be a simple graph representing the social network where each vertex $v \in V$ represents a customer and there exist an edge $(u, v) \in E$ if the customers corresponding to vertices u and v are acquaintances. Let the set of activated/infected people H by information i. In Figure 6.1, $H = \{a, b, c, j, r\}$.

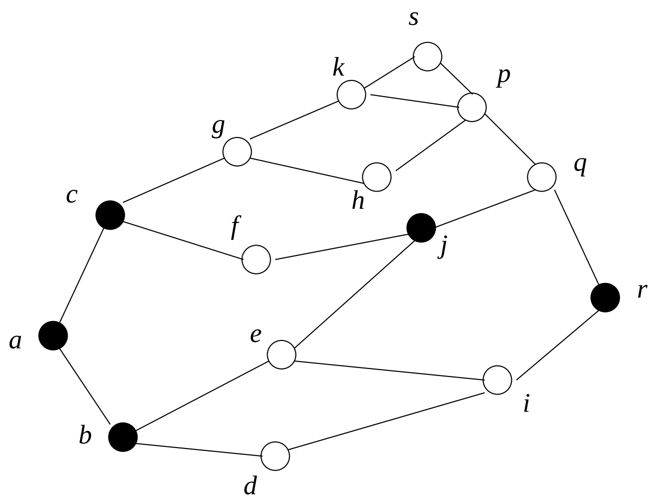

Figure 6.1: Propagating environment

If XYZ immunize the network by sending the *anti-i* information to all the adjacent vertices of H, the information i will not be able to be spreaded. Say A is the set of this adjacent vertices. In Figure 6.2(a), blocking entire $A = \{d, e, f, g, i, q\}$ will stop the propagation of harmful message. Now the problem is to find out a set of vertices containing minimum elements that can stop propagating i. If XYZ sends the *anti-i* information to only g and q like in the Figure 6.2(b), that also stop the propagation of the harmful information.

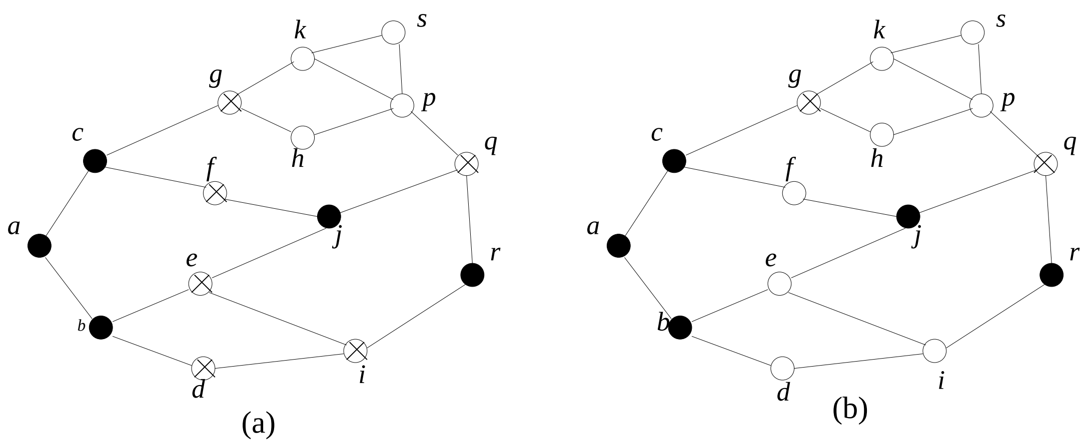

Figure 6.2: Stopped propagation by blocking (a) all adjacent vertices (b) only two vertices

6.4 Algorithm for Stopping Propagation

In this section we present an algorithm for finding minimum blocker set to stop propagation. First we provide the formal steps of the algorithm and we call our algorithm **Filter_Adjacent** for the rest of this section. Based on this algorithm we present a theorem and its correctness proof.

To solve the problem we take an extra vertex say w represents the infected vertices. First we connect all members of A to the extra vertex w. Then we remove all the infected vertices. Now there is no infected vertices, only w is representing all of them. Let G' be the resulting graph. From all vertices of A in G', mark those vertices of $v \in A$ which have atleast one neighbor in $V - A$. These marked vertices are the set of minimum vertices to stop the propagation. This algorithm is denoted as Algorithm **Filter_Adjacent**.

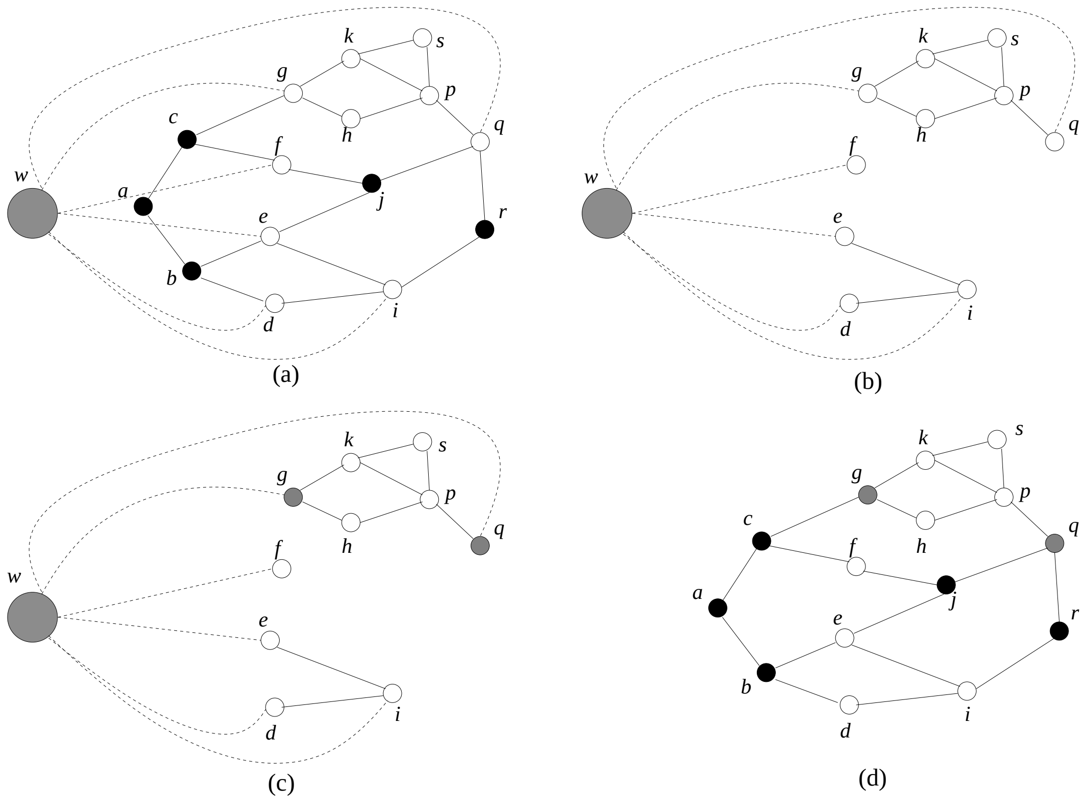

Figure 6.3: Steps of the algorithm (a) connect all adjacent vertices of infected vertices to w (b) delete all the infected vertices (c) mark the vertices that can block the propagation according to the algorithm (d) state of the graph after executing the algorithm

Let the given graph is $G = (V, E)$ Let the set of infected vertices is H and set of adjacent vertices of H is A

Algorithm Filter_Adjacent

Input: A graph G and a set of infected vertices H.

Output: A set of minimum number of vertices B that can block the propagation of harmful information.

begin

1 Let A be the set of neighbors of vertices in H. Connect each $u \in A$ to a new temporary vertex w;

2 Delete all vertices in H. Let G' be the resulting graph.

3 Take in B the vertices of A which has a neighbor in $V - A$ in G'.

end.

Minimum blocker set for stopping propagation can be found by Algorithm **Filter_Adjacent** in linear time. We have the following theorem.

Theorem 6.4.1 *Let G be a graph and H be a set of infected vertices. Then a minimum blocker set B can be obtained by Algorithm **Filter_Adjacent** in linear time.*

Proof. We first prove that B is a blocker set. Since only vertices in blocker set B is connected with the safe vertices $V - A - H$ amongst vertices in A, propagation must go through vertices in B. If each vertex in B is blocked, no propagation can reach to $V - A - H$ from H. Therefore B is a blocker set.

We next prove that B is the minimum blocker set. Assume that there is a blocker set B' such that $|B'| < |B|$. From the problem definition obviously $B' \subseteq A$ since taking a vertex which is not in A does not lead to a minimum solution. Then there is a vertex x such that $x \in B$ and $x \notin B'$. Since every vertex in B has a neighbor to $V - A - H$, information will be propagated through x and hence B' cannot be a blocker set. Hence B is the minimum blocker set.

We now prove the time complexity. Let G be a graph and H be the set of infected vertices in G. Maximum number of adjacent vertices of H can be $n - |H|$ where n is the number of vertices in graph G. If H is minimum, it can be $n - 1$. Worst case time complexity of step 1 is $O(n - 1)$ which can be simplified as $O(n)$. In step 2 we remove vertices as well as edges incident to them. Time complexity of step 2 is $O(n + m)$ where m is the number of edges. Similar to step 1, time complexity of step 3 is $O(n)$. $O(n + m)$ is the time complexity of total steps of the algorithm. Thus we can obtain a minimum blocker set in linear time. $\mathcal{Q.E.D}$.

6.5 Extension of the Problem

Up to this, manufacturer XYZ does not want to send *anti-i* to the people who have not received the information i. This means we have to choose only those people for blocking who are adjacent to infected people. But now we will see what happen when XYZ allows otherwise. Propagation blocker set $B = \{g, q\}$ of Figure 6.1 can be different set like $\{g, p\}$ also be the minimum.

Even there could be such set of elements that does not contain any element of A. If each element of B is not adjacent to any element of H, some nodes will be compromised. Let the compromised set C and safe set S. In Figure 6.4, we can see that set of harmful information propagators $H = \{a, b, c\}$ and minimum blocker set $B = \{j\}$. Hence compromised set $C = \{g, h, i\}$ and safe set $S = \{k, l, o, p, q, r, s, t\}$.

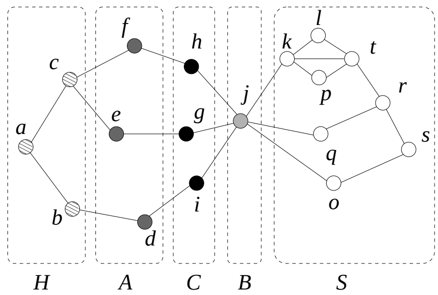

Figure 6.4: Illustration of extended problem

Natural relation between S and C could be
$|S| \times \tau > |C|$, where $0 < \tau \leq 1$.

Problem can be formulated like this, Let $G = (V, E)$ be a simple graph representing the social network where each vertex $v \in V$ represents an individual in social network and there exists an edge $(u, v) \in E$ if the individuals corresponding to vertices u and v are acquaintances.

Let H be the set of vertices that represents the people who the have already been activated to propagate harmful information i.

Let S be the set of vertices that represents safe people who will not receive i or *anti-i*

Let C be the set of vertices that represents compromised people who have been ignored to stop spreading i

Let τ be the ratio of $|S|$ and $|C|$, that defines how much compromise can be tolerated for how much saving.

If τ is given, the problem is to find out the set of vertices B containing minimum elements that can stop propagating of i.

6.6 Conclusion

In this chapter, we formulate the business problem of stopping a propagation of harmful information. Considering independent cascade model and SIR model, we map this problem into graph theoretic problem. Then we have developed an algorithm to find out the minimal blocker set from the adjacent vertices of infected vertices. We have also presented one extension of this problem which can be researched further. Every infected vertices does not have same probability to spread harmful information. Here we have considered only the maximum probability. With the decrease of probability, cardinality of minimal blocker set could be lower. This can be an interesting topic to research in future.

Chapter 7

Conclusion

In this thesis we have addressed different problems of social network marketing. We have formulated them in such a way that graph theoretic algorithms and techniques can be applied on them.

In Chapter 3 we have listed out all the social network marketing problems to be mapped and presented corresponding graph theoretic models along with examples.

In Chapter 4 we have dug deep into seeding strategy problem considering the existence of business rivals on the network . We provided three schemes for selecting seeds for a word of mouth program on a social network.

In Chapter 5 we have developed a model for direct marketing based on social network data. Based on that model we have developed a heuristic algorithm and we have done some experiments with sample input graphs from Rome graph library. We have also presented comparison table showing that in all cases of experiments our algorithm gives optimum result. We have also defined a classification criteria of graphs based on suitability of marketing on social networks.

In Chapter 6 we have presented some existing models of spreading harmful information. Then a problem of stopping propagation of harmful information has been formulated and mapped into a graph theoretic problem. Our algorithm can find minimum blocker set to stop the propagation of harmful information.

Graph theoretic models of social network marketing problems and research done in this thesis lead some practical interesting future works. The following is a brief list of future works related to our results which have been presented in this thesis.

- We have developed a model for direct marketing based on social network data. We found that choosing initial vertices from minimum dominating set is not always optimal. Besides that we also provide an example showing that networks of minimum degree k always not optimal. We have also developed a heuristic algorithm for choosing vertices in a way that cover the whole network. We believe that our algorithm can cover the whole network optimally, but we could not proved it yet. Development of an optimal algorithm is an interesting topic for future research.

- We have presented an algorithm for stopping propagation of harmful information on social networks. We have presented an extension of the problem. Our algorithm does not fit for this extended problem. It will be interesting to develop an algorithm for this extended problem.

References

[1] A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and correlation in social networks. In *Proceedings of the 14th International Conference on Knowledge Discovery and Data Mining*, pages 7–15. ACM, 2008.

[2] R. M. Anderson, R. M. May, and B. Anderson. *Infectious Diseases of Humans: Dynamics and Control*, volume 28. Wiley Online Library, 1992.

[3] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th International Conference on Knowledge Discovery and Data Mining*, pages 44–54. ACM, 2006.

[4] C. A. Benedetto. Identifying the key success factors in new product launch. *Journal of Product Innovation Management*, 16(6):530–544, 1999.

[5] S. Bob. *Successful Direct Marketing Methods.* NTC Business Books, 1989.

[6] A. Borodin, Y. Filmus, and J. Oren. Threshold models for competitive influence in social networks. *Internet and Network Economics*, pages 539–550, 2010.

[7] J. J. Brown and P. H. Reingen. Social ties and word-of-mouth referral behavior. *Journal of Consumer Research*, pages 350–362, 1987.

[8] Z. Cao, X. Chen, and C. Wang. How to schedule the marketing of products with negative externalities. *Computing and Combinatorics*, pages 122–133, 2013.

[9] L. G. Cooper. Social ties and word-of-mouth referral behavior. *The Journal of Marketing*, pages 1–16, 2000.

[10] P. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of the 7th International Conference on Knowledge Discovery and Data Mining*, pages 57–66. ACM, 2001.

[11] P. Domingos and M. Richardson. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining*, pages 61–70. ACM, 2002.

[12] N. B. Ellison. Social network sites: Definition, history and scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 2007.

[13] M. R. Garey and D. S. Johnson. *Computers and Intractability: a Guide to the Theory of NP-completeness*. W. H. Freeman and Company, New York, USA, 1979.

[14] J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Journal of Marketing Letters*, 12(3):211–223, 2001.

[15] M. Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.

[16] J. Hartline, V. Mirrokni, and M. Shundararajan. Optimal marketing strategies over social networks. In *Proceedings of the 17th International Conference on World Wide Web*, pages 189–198. ACM, 2008.

[17] O. Hinz, B. Skiera, C. Barrot, and J. U. Becker. Social contagionan empirical comparison of seeding strategies for viral marketing. *Journal of Marketing Research*, 75(6):55–71, 2011.

[18] A. A. El Houssi, K. P. Morel, and E. J. Hultink. Effectively communicating new product benefits to consumers: the use of analogy versus literal similarity. *Advances in Consumer Research*, 2005.

[19] S. Jim. *Customer Service on the Internet: Customer Support, Direct Marketing and Up-Selling*. NTC Business Books, 1996.

[20] H. Kang, L. Getoor, and L. Singh. Visual analysis of dinamic group membership in temporal social networks. *ACM SIGKDD Explorations Newsletter*, 9(2):13–21, 2007.

[21] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2003.

[22] A. Khatun, S. A. Faruque, and S. Rahman. Selecting promising customers on social networks for promoting new products in presence of rivals. *Journal of Marketing and Operations Management Research*, 1(4):305–312, 2013.

[23] G. Kossinets. Effects of missing data in social networks. *Social Networks*, 28(3):247–268, 2006.

[24] D. Kurtz. Contemporary marketing mason. *OH: South-Western Cengage Learning*, 2010.

[25] B. Libai, E. Muller, and R. Peres. Sources of social value in word of mouth programs. *Marketing Science Institute working paper*, pages 10–103, 2010.

[26] C. Long and R. C. W. Wong. Minimizing seed set for viral marketing. In *Proceedings of IEEE 11th International Conference on Data Mining*, pages 427–436. IEEE, 2011.

[27] G. R. Manuel, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1019–1028. ACM, 2010.

[28] N. Mishra, R. Schreiber, I. Stanton, and R. E. Tarjan. Finding strongly knit clusters in social networks. *Internet Mathematics*, 5(1-2):155–174, 2008.

[29] A. Mislove, Gummadi, Krishna P, Druschel, Bhattacharjee, and Bobby. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, pages 29–42. ACM, 2007.

[30] T. Nishizeki and N. Chiba. *Planar Graphs: Theory and Algorithms*. North-Holland, Amsterdam, 1988.

[31] T. Nishizeki and M. S. Rahman. *Planar Graph Drawing*. World Scientific, Singapore, 2004.

[32] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu. Spatial tessellations: concepts and applications of voronoi diagrams. *Concepts and Applications of Voronoi Diagram*, 1992.

[33] T. Qian, Y. Yang, and S. Wang. Refining graph partitioning for social network clustering. In *Web Information System Engineering–WISE 2010*, pages 77–90. Springer, 2010.

[34] T. C. Schelling. *Micromotives and Macrobehavior.* WW Norton & Company, 2006.

[35] R. Tarjan. Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2):146–160, 1972.

[36] C. Trattner and F. Kappe. Social stream marketing on facebook: a case study. *International Journal of Social and Humanistic Computing*, 2(1):86–103, 2013.

[37] H. Trottier and P. Philippe. Deterministic modeling of infectious diseases: theory and methods. *The Internet Journal of Infectious Diseases*, 1(2):3, 2001.

[38] S. Wasserman. *Social network analysis: Methods and Applications*, volume 8. Cambridge University Press, 1994.

[39] D. B. West. *Introduction to Graph Theory.* Prentice-Hall, Upper Saddle River, New Jersey, USA, 2001.

[40] R. F. Wilson. The six simple principles of viral marketing. *Web Marketing Today*, 70(1):232, 2000.

[41] R. J. Wilson. *Introduction to Graph Theory.* Pearson Education India, 1985.

[42] J. Wortman. Viral marketing and the diffusion of trends on social networks. *Technical Report MS-CIS-08-19, Department of Computer & Information Science*, 2008.

Index

$O(n)$ notation, 15
$\kappa(G)$, 13
k-connected, 13

adjacent, 10
asymptotic behavior, 15

block, 13

call graph, 11
Campaign friendly, 40
Campaign hard, 40
Campaignable, 40
complete graph, 42
complexity, 15
component, 13
connected graph, 13
connectivity, 13
cut vertex, 13
cycle, 13

degree, 10
direct marketing, 18
direct marketing camapigning model, 37
direct marketing campaigning model, 38
directed graph, 11
disconnected graph, 13
dominating set, 14

edges, 10
exponential, 16

graph, 10

Hamiltonian, 13

incident, 10

Linear Threshold Model, 51
linear-time, 16
loop, 11

minimum dominating set, 14
multigraph, 11

non-polynomial, 16

path, 13
polynomial, 16
polynomially bounded, 16

running time, 15

seed, 19
Seeding strategy, 19
separator, 13
simple graph, 11
SIR model, 51
SIS model, 51
Social Network Marketing, 17
Social network sites, 16

undirected graph, 11

vertex cut, 13
vertices, 10
Viral marketing, 3, 18

walk, 13