A

Project Report

on

INFLUENCE MAXIMIZATION OF VIRAL MARKETING ON SOCIAL NETWORK

Submitted in Partial Fulfillment of

the Requirements for the Degree

of

Bachelor of Engineering

in

Computer Engineering

to

North Maharashtra University, Jalgaon

Submitted by

Hemangini Chaudhari
Juili Kulkarni
Payal Jagtap
Sanchita Wagh

Under the Guidance of

Mr. Nitin Y. Suryawanshi

DEPARTMENT OF COMPUTER ENGINEERING

SSBT's COLLEGE OF ENGINEERING AND TECHNOLOGY,

BAMBHORI, JALGAON - 425 001 (MS)

2016 - 2017

SSBT's COLLEGE OF ENGINEERING AND TECHNOLOGY, BAMBHORI, JALGAON - 425 001 (MS)
DEPARTMENT OF COMPUTER ENGINEERING

CERTIFICATE

This is to certify that the project entitled *Influence maximization of viral marketing on social network*, submitted by

Hemangini Chaudhari
Juili Kulkarni
Payal Jagtap
Sanchita Wagh

in partial fulfillment of the degree of *Bachelor of Engineering* in *Computer Engineering* has been satisfactorily carried out under my guidance as per the requirement of North Maharashtra University, Jalgaon.

Date: October 7, 2016
Place: Jalgaon

Mr. Nitin Y. Suryawanshi
Guide

Prof. Dr. Girish K. Patnaik   Prof. Dr. K. S. Wani
Head   **Principal**

Acknowledgements

We would like to express our deep gratitude and sincere thanks to all who helped us to complete this project work successfully. Our sincere thanks to our principal Prof Dr. K. S. Wani and Director of SSBT's COET, Prof Dr. S. P. Shikhawat for providing us facilities to complete our project work. Our gratitude goes to our Head of the Department, Prof. Dr. Girish K. Patnaik for granting us opportunity to conduct this project work. We sincerely thankful to our project guide, Mr. Nitin Y. Suryawanshi and project in charge Mr. Aakash Waghmare, for their valuable suggestions and guidance at the time of need. Also we would like to thanks our parents for always encouraging us.

<div align="right">

Hemangini Chaudhari

Juili Kulkarni

Payal Jagtap

Sanchita Wagh

</div>

Contents

List of Figures

Abstract

Influence maximization is introduced to maximize the profit of viral marketing in social networks. The weakness of influence maximization is that it does not distinguish specific users from others, even if some items can be only useful for the specific users. For such item, it is a better strategy to focus on maximizing the influence on the specific users. The mining problem of influence maximization is the following: given such a network with influence estimates,how should one select the set of initial users so that they eventually influence the largest number of users in the social network. While conducting experiments to evaluate the proposed method with real-life databases, and compare the results with those of existing methods that are adapted to the problem. From the experimental results, the proposed method is at least an order of magnitude faster than the existing methods in most cases while achieving high accuracy.

Chapter 1

Introduction

With the success of online social networks and microblogs such as Facebook, Flicker and Twitter, the phenomenon of influence exerted by users of such platforms on other users, and how it propagates in the network, has recently attracted the interest of computer scientists, information technologists, and marketing specialists. One of the key problems in this area is the identification of influential users, by targeting whom certain desirable marketing outcomes can be achieved. To utilize online interpersonal organizations as a promoting stage, there is loads of examination on the best way to utilize the proliferation of impact for viral advertising. Exploration issues is influence maximization (IMAX), which plans to discover k seed clients to amplify the spread of impact among clients in interpersonal organizations.This can be illustrated by one example that one small scale industry or company who developed small efficient online application for social network. But that company has some limited budget so that it only refers few no of initial users in the network. so company wants that these initial users would love these application and through these application others friends also start influencing through word-of-mouth effect and the problem is that to find out whom to select initial users which influence to the target users which are present in social network. And such problem is called as an Influence Maximization.The solution to this problem is based on the data mining perspective.Data mining is defined as extracting the information from the enormous set of data.The process of data mining includes integrating the data, selecting the useful data for data mining, cleaning the data to remove errors.

This chapter is elaborated in following sections: Section 1.1 describes Background of the project. The Motivation of the project is described in Section 1.2. Section 1.3 describes Problem Definition of project. The Scope of project is described in Section 1.4. Section 1.5 describes the Objective of the project. The Organization of the report is described in Section 1.6. Finally, section 1.7 contains Summary.

1.1 Background

Viral marketing is one of the key applications of influence maximization. In viral marketing, an item that a marketer wants to promote is diffused into social networks by word-of-mouth communication. From the perspective of marketing, influence maximization provides how to get the maximum profit from all the users in a social network through viral marketing. However, influence maximization is not always the most effective strategy for viral marketing, because there can be some items that are useful to only specific users. These specific users can be a few people with a common interest in a given item, some or all people in a community, or some or all users in a class. There is no limit for being specific users. For example, consider a marketer that is asked to promote a cosmetic product for women through viral marketing. For the cosmetic product, the specific users are female users who are likely to use it and male users who wish to purchase it as a gift for female users. In this case, the marketer does not need to be concerned about the other users because the cosmetic product is not useful to them. Instead, it is a better strategy to focus on maximizing the number of influenced specific users, but influence maximization has the weakness that it cannot distinguish them from the other users.

1.2 Motivation

The idea behind viral marketing is that by targeting the most influential users in the network we can activate a chain reaction of influence driven by word-of-mouth, in such a way that with a very small marketing cost we can actually reach a very large portion of the network. Suppose we are given a social network, that is a graph whose nodes are users and links represent social relations among the users. Suppose we are also given the estimates of reciprocal influence between individuals connected in the network, and suppose that we want to push a new product in the market. The mining problem of influence maximization is the following: given such a network with influence estimates, how should one select the set of initial users so that they eventually influence the largest number of users in the social network. This problem has received a good deal of attention by the data mining.

1.3 Problem Definition

Influence maximization problem is query processing which are used to differ particular users from other users. So, for that web personalization is used. The aim of influence maximization is to increase the profit of viral promoting in social networks.But the influence maximization not differ particular user from other users, even if some of items is focused by influence

maximization strategy which is focus in this specific user.

Develop a web portal application for influence maximization of viral marketing.Users should be able to search for required information on the portal.Users can register with the portal to get further facilities.Multiple users visiting number of commercial sites these all data is collected in the databases,it can be in the form of logs,cookies,history etc...So,in the influence maximization process these all database logs are sorted and provide Ads,News,information etc. as per their behaviour and "Influence diffusion model" and target aware viral marketing.First,there are target users who have an interest in item.Second,there are non-target users who can be influence for the item to introduce it to their friends.Finally, there are non-target users who are immune to being influence for the item,because they do not want to introduce to their friends.

The proposed system is going to use data sets and whatever experiments is to be done that is evaluated into datasets and compare these experiment with existing method and from these experimental results.The proposed system is faster than the existing systems also have a high accuracy than the existing methods. Also, the existing methods,suppose even if some items are useful to that particular users still they does not differ that particular users from others.The proposed expectation model which is able to maximizing the influence on the specific users and also it can easily differ particular users from others.

It is easy to see that the influence diffusion model can handle the first case and the second case. However,the influence diffusion model cannot handle the third case, because it does not distinguish such immune nodes from the others.Nevertheless,we can easily modify the influence diffusion model to support the third case by adding one condition to it.Thus, for simplicity we stick to the original influence diffusion model to explain the proposed method.The proposed expectation model which is able to maximizing the influence on the specific users and also it can easily differ particular users from others.

1.4   Scope

The scope of this project is to overcome the weakness of influence maximization in viral marketing. It is a better strategy to focus on maximizing the number of influenced specific users, but influence maximization has the weakness that it cannot distinguish them from the other users.Online social networks are increasingly being recognized as an important source of information influencing the adoption and use of products and services. Viral marketing is the tactic of creating a process where interested people can market to each otheris therefore emerging as an important means to spread-the-word and stimulate the trial,adoption,and use of products and services. Consider the case of Hotmail, one of the earliest firms to tap the potential of viral marketing. Based predominantly on publicity from word-of-mouse, the

Web-based email service provider garnered one million registered subscribers in its first six months, hit two million subscribers two months later, and passed the eleven million mark in eighteen months.Wired magazine put this growth in perspective in its December 1998 issue: The Hotmail user base grew faster than [that of] any media company in historyfaster than CNN, faster than AOL, even faster than Seinfelds audience. By mid-2000, Hotmail had over 66 million users with 270,000 new accounts being established each day.

1.5 Objective

The aim of Influence maximization is to increase the profit of viral promoting in social networks. But Influence maximization not differ particular user from other users, even if some of the items are useful for that particular user. So these few types of item is focused by influence maximization strategy which is focused on these specific users. Influence maximization problem is a query processing which are used to differ particular users from other users.

1.6 Organization of Report

Chapter 1 describes the background of project with its scope and problem definition. Chapter 2 describes system analysis. Chapter 3 describes system requirements specification which includes software, hardware, functional and non functional requirements. Chapter 4 describes system design with the help of various unified modeling language diagrams. chapter 5 describes implementation details and flow of the system development. chapter 6 describes system testing, how to implement it and some of the test cases. Results and analysis are given in chapter 7. Finally, chapter 8 contains conclusion and summary.

1.7 Summary

This chapter covers the Introduction of the project, Problem definition, Scope and Objective of the project. In next chapter, the system analysis is described.

Chapter 2

System Analysis

System analysis is the process of gathering and interpreting facts, diagnosing problems and using the facts to improve the system. This chapter is elaborated in following sections: Section 2.1 describes the Literature Survey. Proposed System is discussed in Section 2.2. Section 2.3 describes Feasibility study. Project Scheduling is described in Section 2.4. Section 2.5 describes Effort Allocation. Finally, section 2.6 contains Summary.

2.1 Literature Survey

We highlight two factors that play a key role in determining the nature of influence episodes in viral marketing. The first is the role of the influencerwhether the attempt to influence is passive or actively persuasive. The second is the level of network externalities the additional benefits accruing from broader usage of the product or service being recommended within a user community.There are four types of role of the influencer:

- Awareness Creation and Benefits Signaling (ACBS): The role of the influencer in persuasion is passive and the network externalities are minimal. Users emailing online greeting cards from Web sites such as Hallmark or BlueMountain to connected others represent typical instances of ACBS. When a user sends out a card from the site, the recipients get a personalized email message informing them of a greeting created by the sender available at the site and providing the URL to access it. The URL directs visitors to the card on site and once there, he or she is offered the choice to send a greeting to the original sender or to a connected other. In this process, recipients are made aware of the service offered by the site and are persuaded to use it. The role of the influencer is mainly to create awareness and signal benefits to others within their social network and can be particularly influential in encouraging trial and adoption of novel products and services.

- Targeted Recommendation: The influencer plays an active role in spreading the word

and the network externalities are minimal. For example, Honda Motor Europe unveiled the Honda Civic with a lifestyle-oriented campaign Live the Civic at the Paris Motor Show in September 2000. Honda used a viral marketing technique to sustain customer interest and gather data on potential buyers for the car.A set of amusing video clips playable over the Internet were distributed to 300 influencers to kick off a contest where participants chances of winning increased if they forwarded the videos to others. Reports suggest that this was responsible for nearly half a million individuals visiting the Honda Web site for the new car in the first three months with the video clips being forwarded almost 80,000 times.The viral marketing campaign is also credited with generating over 10,000 promising leads for the new car.

- Signaling Use, Group Membership (SGM): The influencers role is passive but there are significant externalities accruing to both the recipient and the influencer. Instances include the use of specific kinds of products, for example, file compression utilities such as winzip and animation software such as Flash. When a user sends the connected other a file compressed using winzip as an email attachment or makes a Flash animation available on a homepage, the recommenders role in spreading the word about the software is passive.In the initial stages of the lifecycle when a software package is not widely known and used, early users are generally viewed as being technically advanced.

- Motivated Evangelism (ME): where recommenders play an active role in influencing connected others and there are significant network externalities accruing to both influencers and recipients.ICQan instant messaging applicationand Dialpadan application to place telephone calls over the Internetare instances of motivated evangelism. In these instances, the influencer as well as the recipient need to use the product for either of them to benefit. The structure of benefits motivates early adopters to actively persuade connected others to also try the product so that they can both use the product.

2.2   Proposed System

The proposed system is the system which basically deals with the following assumption. The basic assumption is that when users see their social contacts performing an action they may decide to perform the action themselves. In truth, when users perform an action, they may have any one of a number of reasons for doing so: they may have heard of it outside of the online social network and may have decided it is worthwhile; the action may be very popular (e.g., buying an iPhone 4S may be such an action);or they may be genuinely influenced by seeing their social contacts perform that action.Generally, the focuss on the important

problem of distinguishing real social influence from homophily and other external factors. Homophily is a term coined by sociologists in the 1950s to explain the tendency of individuals to associate and bond with similar others. This is usually expressed by the famous adage birds of a feather flock together. Homophily assumes selection, i.e., the fact that it is the similarity between users to breed connections.

2.3   Feasibility Study

Feasibility study aim to uncover the strengths and weaknesses of the existing business or proposed venture opportunities and threats as presented by the environment. The results of feasibility study are used to make a decision whether to proceed with the project or whether to table it. There are various types of feasibilities available that depend on different factors like technical, operational. A feasibility study could be used to test a proposal for new system. First, the scale and scope of influence is considerably expandedcomputer mediation allows a much larger number of individuals to be connected by informational linkages than feasible through face-to-face contact or through conventional media such as the telephone. For instance, the effort to email a message to all contacts in the address book is only marginally more than the effort in sending the message to just one recipient. On average, the reach of individualsthe number of connected others1 that they can influence, increases considerably. Further, this reduction in the effort needed to reach out to others increases the number of occasions when individuals act on their natural impulse to share knowledge and pass along information they consider useful and timely to others in their social network. Together, this results in an enormous increase in the extent of influence through knowledge-sharing and information transfer in online networks.

2.3.1   Economical Feasibility

The project is economically feasible as its an open source project meaning most of the resources required for the development for free thus making its very much feasible in accordance to economy.

2.3.2   Operational Feasibility

This project aims at user friendly interface any non skilled person can handle it .For example, students are also included in non-skilled people but they can easily make their accounts and access the social network as they are the very important part of the influencer group, insert there key interest and behavioral pattern through visiting that blog again and again.Thus,

the activities performed by the student will be reflected in datasets. which will store data at back-end MySQL Server.The operational feasibility of the project is at greater extent.

2.3.3 Technical Feasibility

The technical feasibility of this project is also fair enough.The technical person as well as non-technical person can play a role of influencer and can be one of the influenced intial users sets through which the viral marketing is done.The project is technically feasible and can make changes for further future extent.

2.4 Project Scheduling

Project Scheduling is all about scheduling the project i.e its each and every phase and managing time needed for each phase to complete. It gives the details of how much time is required to complete each and every phase of software engineering for this project.

2.5 Effort Allocation

The Effort Allocation deals with the involvement of each team member in the project. The Phases defined are Literature survey, Requirements Analysis, Planning, Analysis, Design, Implementation, testing.Thus, success of the project lies within the effort of each team member.

2.6 Summary

This chapter describes the overall phases of project development and also covers the analysis of the project such as it's Feasibility study, Literature Survey and also an overview of proposed system. The next chapter describes the Software Requirement Specification.

Chapter 3

System Requirement Specification

Software Requirement Specification is the official statement of what is required to the system developers. It should include both user requirements and a detailed specification of the system requirements. Requirement analysis is done in order to understand the problem the software system is to solve. This chapter is elaborated in following sections: Section 3.1 describes the Hardware Requirements. Software Requirements are explained in Section 3.2. Section 3.3 explains the Functional Requirements. Non Functional Requirements are described in Section 3.4. Section. Finally, section 3.5 contains Summary.

3.1 Hardware Requirements

The hardware requirements includes:

- 15gb of disk memory or disk partition.

- Ram-4 GB.

- AMD/Intel 32-bit or 64-bit Processor.

3.2 Software Requirements

The software requirements includes:

- JDK 1.7.0.

- OS Ubuntu 14.04.

- MySQL Server.

3.3   Functional Requirements

Functional requirements include both functional and non functional requirement aspect related to the project. Installed JDK 1.7.0 and MySql Server must provide development of the project. Thus a pre installed Linux operating system i.e Ubuntu version 14.04.

3.4   Non-Functional Requirements

Non functional requirement include various requirement but the most prominent one are access to database logs and prerequisite datasets available for data mining. Thus, the differentiation of special users from others by influence maximization phenomenon.

3.5   Summary

In this chapter, Hardware Requirements, Software Requirements, Functional and Non-Functional Requirements are explained. In next chapter, the System Design is described through various UML diagram.

Chapter 4

System Design

System design provides the understanding and procedural details necessary for implementing the system. This chapter is elaborated in following sections: Section 4.1 describes the System Architecture, Section 4.2 describes the Unified Modelling Diagrams and 4.3 contains Summary.

4.1 System Architecture

The system architecture provide details of how the components or modules are integrated and is described with the help of Unified Modelling Diagrams. In Unified Modelling Diagrams there are 6 diagrams they are Use Case Diagram, Class Diagram, Sequence Diagram, State Diagram, Component Diagram, Deployment Diagram.

4.2 UML Diagrams

4.2.1 Use Case Diagram

A Use Case diagram shows the interaction between the system and entities external to the system. These entities are called actors which have specific role in the system. The Figure 4.1 , shows the Use Case Diagram for proposed system.

4.2.2 Class Diagram

A Class diagram is used to represent the static view of the system. It mainly use Classes, interfaces and their relationships. The classes are connected with each other according to their relationship with each other. The class Diagram contents the attributes and operations of the required classes, which shows various properties of the classes as well a set of task performed by each class which is shown in Figure 4.2.
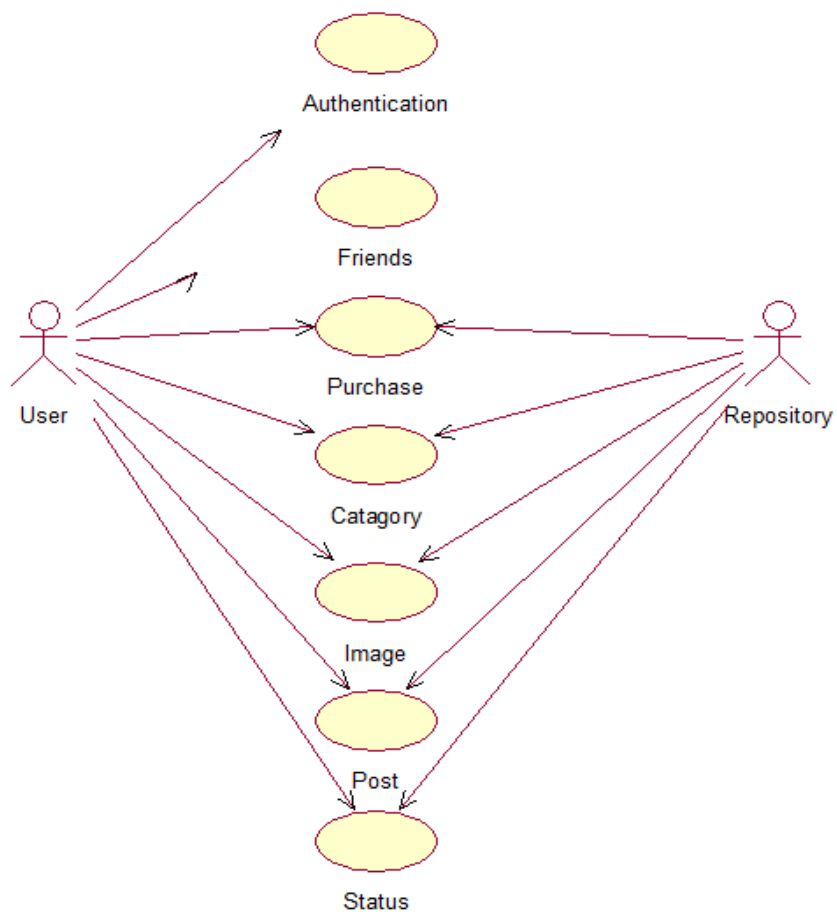
Fig. Use case diagram for Influence maximization of viral marketing on social network
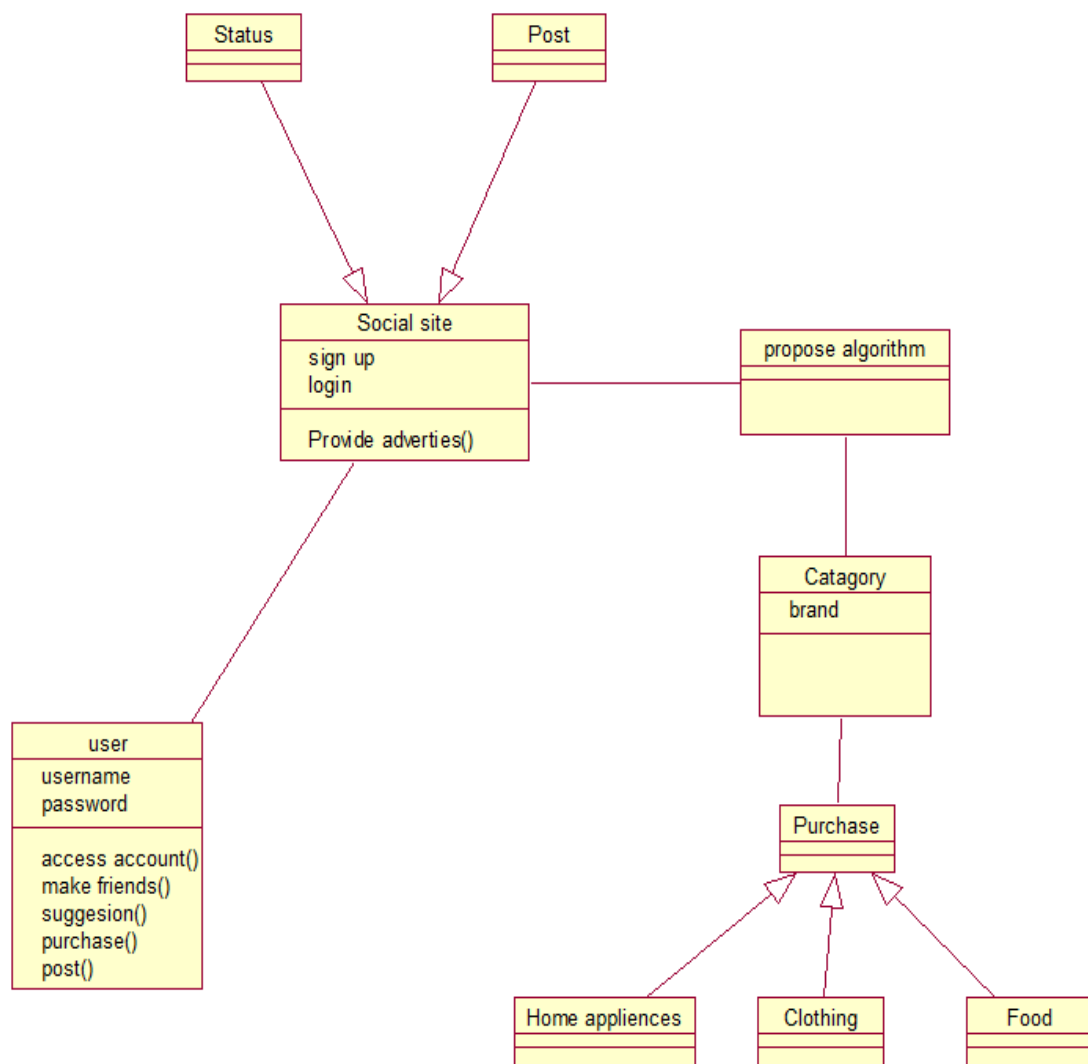
Figure 4.1: Use Case diagram.

Fig.Class diagram for Influence maximization of viral marketing on social network

Figure 4.2: Class Diagram.

4.2.3 Sequence Diagram

The Figure 4.3 shows the Sequence diagram. Sequence diagram shows the sequence of user and repository.

4.2.4 State Diagram

The Figure 4.4 shows the State diagram. In the project there are four states Register, Login, Shopping and Checkout between Start and End state.
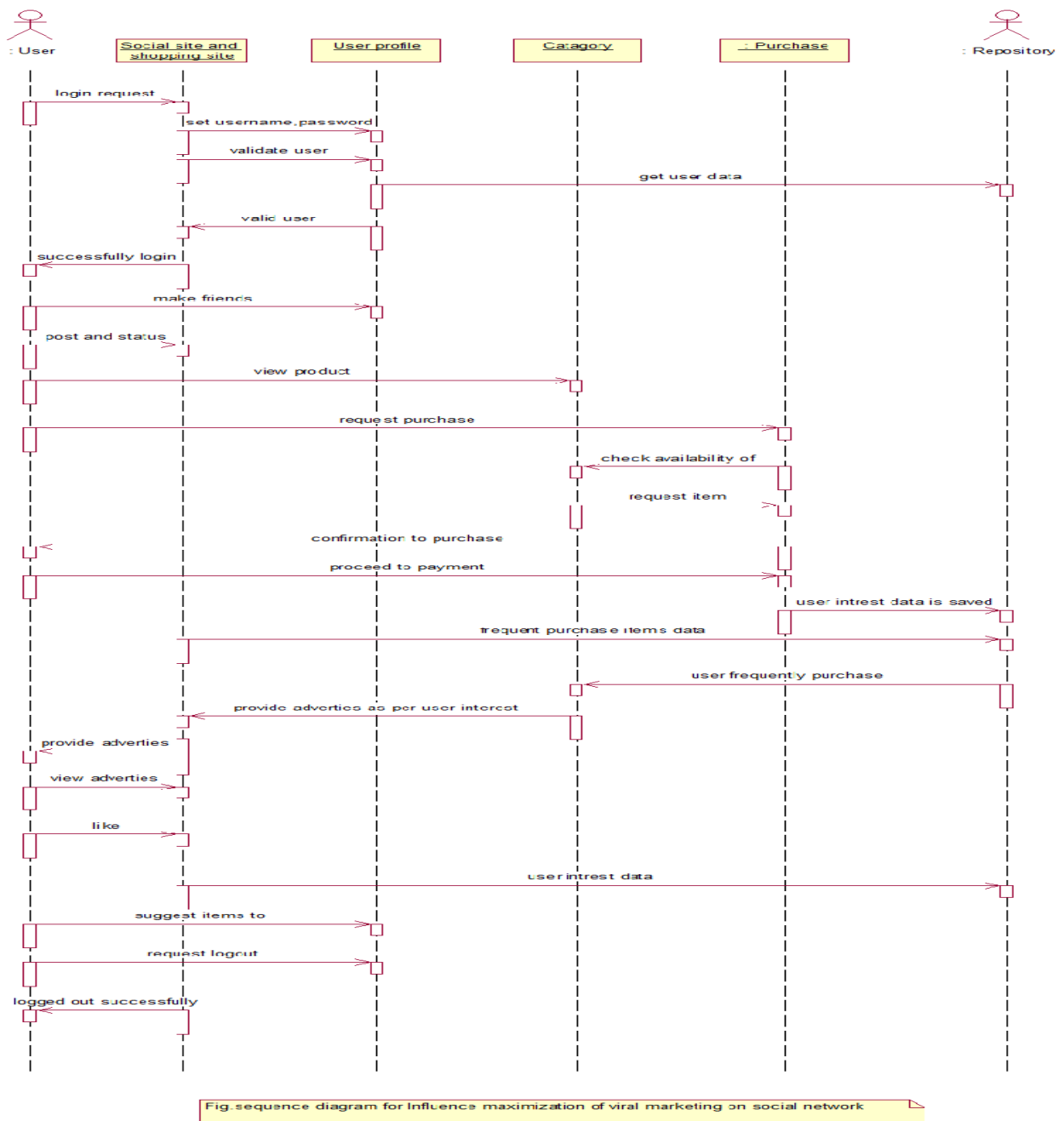
4.2.5 Component Diagram

A component diagram shows the organization and dependencies among set of components. These diagrams are used to model static view of the system. The Figure 4.5, shows the component diagram for the proposed system.

4.2.6 Deployment Diagram

Deployment diagram represents the configuration of run time processing nodes and the components that live on them which is shown in Figure 4.6.
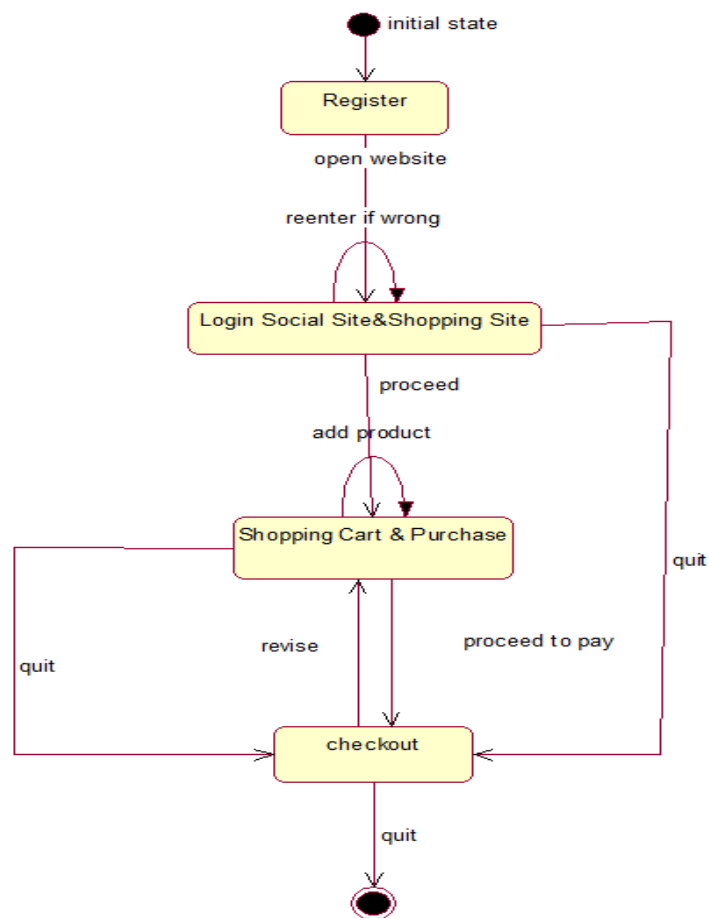
4.3 Summary

This chapter includes various UML diagrams which shows the overall design of project. The next chapter describes the conclusion.

Fig.sequence diagram for Influence maximization of viral marketing on social network

(a)

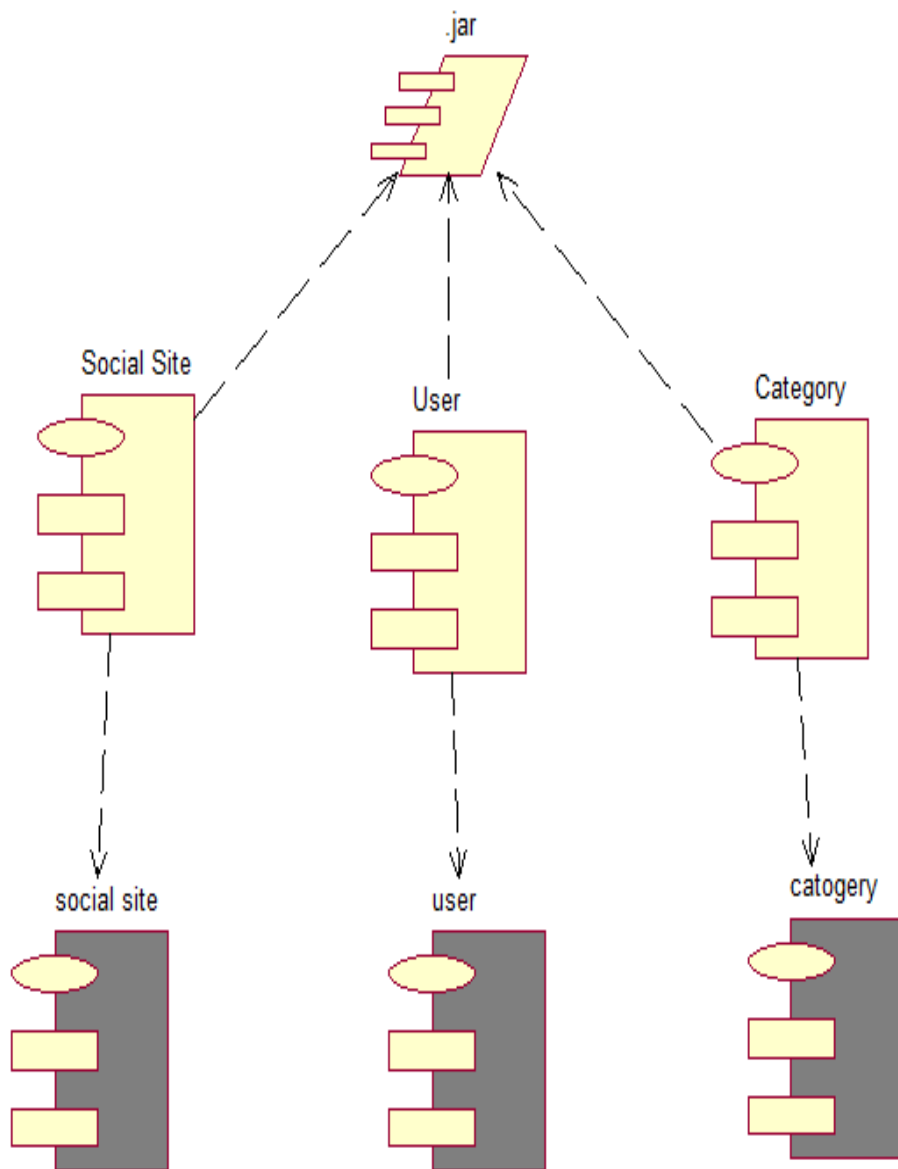Figure 4.3: Sequence Diagram

Figure 4.4: State diagram.

Fig. Component diagram for Influence maximization of viral marketing on social network
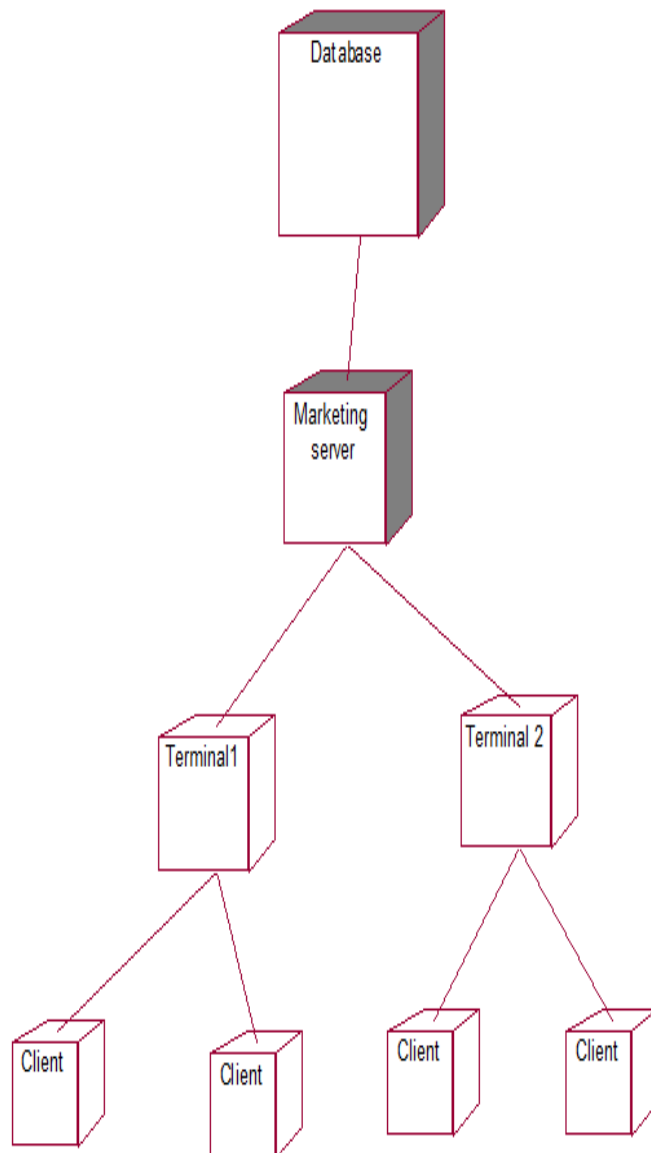
(a)

Figure 4.5: Component Diagram.

Fig.Deployment diagram for Influence maximization of viral marketing on social site

Figure 4.6: Deployment Diagram.

Conclusion and Future Scope

Viral marketing is a complex phenomenon and the proposed framework draws attention to a variety of factors important in determining outcomes: user influence and recipient behavior. Attributes of the product or service determine the quadrant and the dominant character of the user influence. The quadrants are associated with different mechanisms of compliance; the framework highlights contexts where the characteristics of the influencer and the recipient are key to explaining outcomes. The framework can also facilitate decisions regarding the nature of ancillary resources required.

Viral marketing is a powerful means for both marketers and recipients to benefit from the innate helpfulness of individuals in social networks. However, success things upon the recognition of the strong need for influencers to be viewed as knowledgeable helpers in the social network rather than as agents of the marketer. Schemes that make overt attempts to co-opt users to promote products and services are likely to upset the balance and reduce the effectiveness of the approach to the detriment of both the marketer and users who may have benefited from the knowledge-sharing acts of influencers. Firms would do well to reflect on this very carefully in planning viral marketing efforts.

Bibliography

[1] December 2011, Vol.12. Francesco Bonchi, "Influence Propagation in Social Network: A Data Mining Perspective".

[2] Pedro Domingos, "Mining Social Network for Viral Marketing".

[3] David Kempe and John Kleinberg, "Maximizing the Spread of Influence through a Social Network".

[4] Shishir Bharathi and David Kempe, "Competitive Influence Maximization in Social Networks"

[5] December 2003, Vol.46. Mani R. Subramani and Blagi Rajagopalan, "Knowledge-Sharing and Influence in Online in Social Networks via Viral Marketing".