Study of an Epidemic Multiple Behavior Diffusion Model in a Resource Constrained Social Network

by

Anindita Dey

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Master of Science

Approved November 2013 by the
Graduate Supervisory Committee:

Hari Sundaram, Chair
Pavan Turaga
Hasan Davulcu

ARIZONA STATE UNIVERSITY

December 2013

ABSTRACT

In contemporary society, sustainability and public well-being have been pressing challenges. Some of the important questions are: *how can sustainable practices, such as reducing carbon emission, be encouraged? , How can a healthy lifestyle be maintained?* Even though individuals are interested, they are unable to adopt these behaviors due to resource constraints. Developing a framework to enable cooperative behavior adoption and to sustain it for a long period of time is a major challenge. As a part of developing this framework, I am focusing on methods to understand behavior diffusion over time.

Facilitating behavior diffusion with resource constraints in a large population is qualitatively different from promoting cooperation in small groups. Previous work in social sciences has derived conditions for sustainable cooperative behavior in small homogeneous groups. However, how groups of individuals having resource constraint co-operate over extended periods of time is not well understood, and is the focus of my thesis.

I develop models to analyze behavior diffusion over time through the lens of epidemic models with the condition that individuals have resource constraint. I introduce an epidemic model SVRS (Susceptible-Volatile-Recovered-Susceptible) to accommodate multiple behavior adoption. I investigate the longitudinal effects of behavior diffusion by varying different properties of an individual such as resources,threshold and cost of behavior adoption. I also consider how behavior adoption of an individual varies with her knowledge of global adoption.

I evaluate my models on several synthetic topologies like complete regular graph, preferential attachment and small-world and make some interesting observations. Periodic injection of early adopters can help in boosting the spread of behaviors and sustain it for a longer period of time. Also, behavior propagation for the classical epidemic model SIRS (Susceptible-Infected-Recovered-Susceptible) does not continue for an infinite period of time as per conventional wisdom.

One interesting future direction is to investigate how behavior adoption is affected when number of individuals in a network changes. The affects on behavior adoption when availability of behavior changes with time can also be examined.

To my Family.

It has been an exciting and enriching journey towards my thesis. I am indebted to several people who have contributed directly and indirectly to my thesis. I would like to express my deep gratitude to my thesis advisor, Dr. Hari Sundaram for his patient guidance, inspiring thoughts and useful critiques during my master's study. His enthusiastic attitude towards research always inspired me throughout this journey. His approach to problem formulation and presentation of the same is a quality, which I try to emulate and has totally transformed my attitude towards research. Besides my advisor, I would like to thank Dr. Hasan Davulcu and Dr. Pavan Turaga for agreeing to be a part of my thesis committee and providing insightful comments on my thesis.

I would like to offer my special thanks to Kaushik Sarkar, for providing stimulating discussions and helpful hints whenever I was stuck on a problem.

Last but not the least, this thesis would not have been possible without the support of my entire family. I wish to thank my grandparents, my parents and my sister for their endless love and unconditional support. No words are enough to express my gratitude for all that they had to sacrifice and endure so that I may have this opportunity.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Chapter 1

INTRODUCTION

In this highly competitive and fast paced world , all of us want to accomplished a lot but are inhibited by several constraints on our resources. Even if an individual adopts a behavior, sustaining it over a long period of time is challenging. The question , why do even highly motivated individuals lose interest over time? intrigued me to pursue with my research in a quest to understand the reason. It is a challenging question, and one reason may be the core idea that people may not have resources to perform it. I am interested in understanding what happens when we incentivize a group of individuals in a community to adopt a behavior, assuming that not all individuals have enough resources to perform it.

Even ten years ago large scale social networking sites were thought to be a novel and emerging field. Now, most teenagers wake up and "Facebook" or "Tweet" their thoughts or whereabouts. Everything changes and so has the way we communicate or think. Posts of a friend riding a cool bike can influence others to take up biking. At the same time, this motivation can die down over a period of time. This leads to the next question of my research - is motivation cyclical? This has parallels in epidemiology, and through my research I am trying to explore that connection.

In Section 1.1, we introduce the idea of behavior diffusion in social network with resource constraint and it's parallels with epidemic models. In the following section, we present the context of our problem. In Section 1.3, we discuss the main research questions.

1.1 Epidemic Models of Behavior Diffusion in Social Networks with Resource Constraints

There has been a growing interest in social network analysis [26] , [51] , [54] , [60], [53] , [73], [31] fueled by the explosive growth of the Internet and online communities like Facebook, Twitter as well as blog , collaboration and email networks. A social network is essentially a graph of relationships and interactions among social entities such as individuals, groups of individuals, and organizations. It plays a critical role as a medium for the propagation of information, ideas, and influence among its members. Use of cell phones among college students or the rise of a political movement in an unstable society, as it did in the case of Egyptian population in the wake of gag order from the dictatorial government on local print media, provides a good example.

Social media can also be a powerful catalyst for environmental / social sustainability. Consider the case, when the habit of eco-friendly transportation alternatives needs to be spread among daily commuters or when the problem of obesity in population needs to be addressed through a profusion of awareness of group activities among individuals. These kinds of problems are termed as collective action problems [74] in the social sciences. It is noteworthy that, information or shared idea diffusion can either die out quickly or make significant inroads into the general population. It is exciting that with the advent of social networks, communication and trust formation becomes easy and less costly. It's possible that social network can increase the chances of sustainable behavior adoption.

We are motivated by collective action problems. To answer questions such as how does a person's limited resource, including time and money affect how she participates in real-world activities? How can change of network dynamics (like reducing the adoption cost or increasing the number of participants) affect her participation over time? A person's interest to adopt a new behavior, like riding a bike instead of driving a car to work, can be restricted due to lack of resources. In the real world, we are bombarded with choices and may like to adopt multiple behaviors. However, adopting every behavior has some cost involved with it. Current models of behavior adoption lack the idea that individuals may have significant resource constraints that preclude them from successfully adopting behaviors in which they are interested. Resource constraints not only limit individual participation, but also shape how behaviors spread in a network. Some recent works explored the resource constraint paradigm and presented seed selection mechanisms to maximize the behavior diffusion. However, they do not incorporate the long term effect of limited resources on behavior diffusion. Indeed, the change in network dynamics over time and resource constraints can have significant effects on how behavior diffusion in a social network shapes up.

The idea of sustained behaviors has parallels in epidemiology. Epidemic models namely Susceptible-Infected-Recovery (SIR) and Susceptible-Infected-Recovery-Susceptible (SIRS) [73] model the spread of an entity through a networked population .For example, in the spread of a disease through a population, contact between an infectious and a susceptible individual can lead to the transmission of infection. In a similar way, individuals or groups adopting a behavior can motivate other individuals or groups to adopt the behavior given they have the necessary resources. In SIR model, an infectious individual gets recovered after a period of time, similarly an individual can let go of an adopted behavior after a period of time.

1.2 The Problem Context

In this thesis, we present a model for multiple behavior diffusion that captures the complex dynamics of multiple behavior adoption in resource constrained networks over time. Our model has associated costs, utilities and adoption duration that are independent of the individual. Mindful of the work by Aral, Muchnik, and Sundararajan [4] and

Shalizi and Thomas [63], individuals in our model evaluate a utility function for each behavior combining intrinsic interest and social signals. An individual adopts behavior when she receives a social signal which is higher than her adoption threshold and when she has the resources to do so. In this study, we use three metrics: resource utilization, total adoption in the network and unique number of participants.

In our work on diffusion of behavior, the process of behavior adoption by an individual plays a central role. Adoption of behavior has been studied extensively in social psychology. Specifically, our model of behavior adoption by an individual is motivated by the Theory of Planned Behavior (TPB) postulated in [2]. According to this theory, attitude towards a behavior, subjective norm (or social pressure) and perceived behavioral control constitute the intention of adopting a behavior. Empirical studies show that not only intention is a good predictor of the actual behavior, but also it plays a direct causal role, i.e. with sufficiently high intention and actual control over behavior, an individual is expected to engage in the behavior whenever opportunity arises. Attitude towards a behavior from TPB motivates the concept of global utility for each behavior in our model. Similarly, the local influence in our diffusion model can be seen as playing the part of subjective norm (where the only referent is the social acquaintances). The notion of individual resource constraint in our behavior diffusion model mirrors the notion of perceived behavior control in TPB. However our work is markedly different from the related works in social psychology in two major aspects. In most of the works in social psychology, the main focus is on the individual. Although individuals play an important role in our model, we also consider the underlying social network connecting the individuals. Another subtle difference is that our model presents a dynamic view of the situation where the subjective norm or social pressure changes over time, and thereby enabling us to study the longitudinal shift in the pattern of behavior adoption. In contrast, the social psychological studies are snapshot studies providing at best a measure of correlation between before and after the snapshot point. Nevertheless, it can be said that the role of TPB in relation to our model is similar to that of a rational agent in relation to modern micro-economic theory.

1.3 Main Research Questions

Assume that you have a network of individuals and each individuals are connected to their neighbors in certain topologies. There are N individuals with defined resources, and are connected together with m edges. In the beginning, through an advertising campaign for example,k individuals are selected to be the seeds and the behavior will diffuse through the population over time. In my research, I am examining the following four questions related to the above scenario.

1.3.1 How varying resource availability of an individual can shape the behavior diffusion over time?

We model weekly resource variation, i.e individuals have access to different resources each day of the week. We analyze two variants of resource availability. In the first variant, all individuals have the majority of the resources available at the same time. For example, you want to do charity work and the majority of individuals have time during weekends. In the second variant, individuals vary in availability of their different resources. People answering to questions in an online forum can serve as a good example for this. Some individuals may want to set 30 mins every day of the week and some may set 4 hrs on a Saturday for participating in such forums.

1.3.2 What happens when adoption threshold of each individual vary over time?

In the Linear Threshold model, individuals are assigned a fixed threshold which never varies over time. However in the real world, the intent to adopt a behavior can change over time. For example, if a product becomes extremely popular, a person may be more inclined to adopt it as her threshold will drop.On the other hand, if you are an early adopter and you don't adopt the behavior in the beginning, you may become reluctant to adopt the behavior once it gets popular.

1.3.3 Can change in the behavior adoption cost affect behavior distribution in the long term?

In classical behavior adoption models, the cost of adopting a behavior does not change with time.In our model, we incorporate variation in behavior adoption cost over time. As an example, consider an individual riding a bike. In the beginning, the adoption of the behavior is costly as the individual spends time learning the skill, but as she gets better it becomes easier to perform. The cost can become lower when a skill gets acquired through performing the behavior repeatedly. Even after acquiring a skill, a base cost i.e. the base effort to perform the behavior, still remains.

1.3.4 How does the resource utilization change with varying global influence over time?

This is a variant of the Linear Threshold model which assumes that each node of the network is aware of the behavior adopted only by its neighbors. In our model, we are appending the influence of global adoption to the basic LT model. It has been seen in a prior experiment that the guests in a hotel reacts to the global adoption of behavior. Half of the guests received a note telling them to save water by reusing their towels, and the other half received one saying that more than 80 % of the guests previously have helped in conserving water by reusing the towels. It was inferred that the guests receiving the latter note adopted more to reusing the towels than their counterparts. Thus, showing that with global influence over time, an individual can be impacted.

1.4 Contribution of the thesis

The following are the main contribution of this thesis.

- We introduced an epidemic model SVRS (Susceptible-Volatile-Recovered-Susceptible) which is suitable for multiple behavior adoption.

- We looked into all variation of adoption parameters over time , like resource availability , behavior adoption costs, behavior adoption threshold and varying global influence .

- We also extended our model to investigate the scenarios where combinations of parameter variations took place .

1.5 Organization of the thesis

The next chapter discusses the related works on Information Diffusion and Epidemic models. Then, Chapter 3 on page 27 introduces the detailed problem description and describes the multiple behavior diffusion model. It also talks about the performance metrics used for evaluation of effectiveness. A description of influencing parameters and architecture to vary those parameters over time are mentioned in Chapter 4 on page 48. It also elaborates the simulation experiments starting from the network topologies to the results obtained from simulations. Finally, conclusive discussions with possible extensions and open issues are presented in Chapter 5 on page 69.

Chapter 2

RELATED WORKS

We review all related literature to our work in this chapter. The literature can be categorized into two parts: Information Cascade and Epidemic Diffusion.

2.1 *Information Cascades*

" When people are connected by a network, it becomes possible for them to influence each other's behavior and decisions." [24]

Kleinberg described this basic principle results in a number of social processes where networks tend to accumulate individual behavior to produce population wide collective outcomes. What products people buy, the opinions they hold, the different activities they perform, the technologies that define their lives and various other things put together can make nearly a limitless set of circumstances where a person can be influenced by another. The desire here is to understand why these influences occur and to rationally realize why one should sometimes imitate other's choice though his own information might suggest otherwise.

Kempe et al [41] discussed in their paper which set of individuals to target to trigger bigger cascades in a social network. Using submodular function based analysis framework Kempe et al [42] have studied two basic diffusion models namely - Independent Cascade Model and Linear Threshold Model. Their result shows 63 % performance improvement guarantee as compared to the previous node selection heuristics based on the degree and distance centrality. Calculating $\sigma(A)$, the influence function is tricky, and authors have used large-scale simulations to define the influence function. It is almost impossible to evaluate the influence function for real life applications by running simulations. Thus in our case, we used resource utilization and total adoption in the network as the metric for measuring the effectiveness.

Watt's journal [74] starts with the idea of influential nodes, which is used extensively in marketing and diffusion research. Influential(s) refers to a minority of individuals in a group whose influence is important to the formation of public opinion. By using different models and computer simulations, he figured out that social change takes place, not because of a few influentials, but by a group of easily influenced individuals, which he referred as a critical mass. He also showed that attributes of influentials are mostly accidents of timing and location than any special characteristics.Also, social changes are highly dependent on group structure, only if the right global combination of conditions exists. In our model, we used the concept of influentials in order to perform seed selection.

Leskovec et al [33] in their paper investigated the problem of tracing paths of information diffusion and influence. During information diffusion, it is often possible to directly observe when nodes become infected, but observing exactly who infects whom or who influences whom is a very difficult task. In many applications, network topology for propagations is unknown. Leskovec et al [44] addressed these challenges by developing a method for tracing paths of diffusion and influence through networks and inferring the networks over which contagions propagate. The efficient approximation algorithm they proposed scales to large datasets and gives near-optimal performance. Kleinberg et al focused on maximizing spread of behavior through a social network with a known topology while Leskovec et al attempted to define structure of the underlying network given information spreading is taking place.

Watts n Strogatz in their work on small world dynamics [72] have shown how a regular graph can be transformed into a "small world " graph by random rewiring where rewiring probability 0 implies completely regular topology and 1 imply completely chaotic topology. They studied the scenario when the probability lies between 0 and 1. They further probed the functional significance of small-world connectivity for dynamical systems. In our work, we have chosen small world as one of the representative network topology under multiple behaviors diffusion condition with resource constraint and epidemic scenario.

The origin of large but rare cascades that are triggered by small initial shocks is a phenomenon that manifests itself as diversely as cultural fads, collective action, the diffusion of norms and innovations, and cascading failures in infrastructure and organizational networks. Watts in the paper [73] presents a possible explanation of this phenomenon in terms of a sparse, random network of interacting agents whose decisions are determined by the actions of their neighbors according to a simple threshold rule. Two regimes are identified in which the network is susceptible to very large cascades - herein called global cascades - that occur very rarely. When cascade propagation is limited by the connectivity of the network, a power law distribution of cascade sizes is observed, analogous to the cluster size distribution in standard percolation theory and avalanches in self-organized criticality. But when the network is highly connected, cascade propagation is limited instead by the local stability of the nodes themselves, and the size distribution of cascades is bimodal, implying a more extreme kind of instability that is correspondingly harder to anticipate. In the first regime, where the distribution of network neighbors is highly skewed, it is found that the most connected nodes are far more likely than average nodes to trigger cascades, but not in the second regime. Finally, it is shown that heterogeneity plays an ambiguous role in determining a system's stability: increasingly heterogeneous thresholds make the system more vulnerable to global cascades; but an increasingly heterogeneous degree distribution makes it less vulnerable.

Bakshy et. al. [6] studied and modeled social influence based on the change in adoption rate due to the actions of one's friend using the social game Second Life [1] as a test bed. They found out that the adoption rate quickens as the number of friends adopting increases and this effect varies with the connectivity of a particular user. They further found that sharing among friends occurs more rapidly than sharing among strangers. Lastly, they examine the role of individuals, finding that some play a more active role in distributing content than others but that these influencers are distinct from the early adopters.

Bakshy et. al. didn't factor in the effect of fees on the transfer of assets, which we have considered as cost of adopting behavior. It is worth exploring, how an individual's behavior to accept asset changes when assets are costly to acquire i.e. when every behavior comes with a cost of adoption. Another interesting dimension they didn't explore is the long-term epidemic effect. Long-term epidemic effect of behavioral adoption may inhibit the spread of assets, favoring the spread of those where users adopts behavior and does not let go off their behaviors. The notion of temporal behavior adoption is shaped into our model using different epidemic modeling schemes.

Chen et. al. [18] , [19] studied the influence maximization problem from complementary directions - to improve the original greedy algorithm espoused by Kempe et al [42] and to propose new degree discount heuristics that improve influence spread. Their analysis is based on single behavior diffusion and doesn't consider other aspects of behavior propagation - namely, cost of behavior or time span of the behavior adoption.

Kleinberg [22] studied sequential influence models in social networks. The spread of influence among individuals in a social network can be naturally modeled in a probabilistic framework, but it is challenging to reason about differences between various models as well as to relate these models to actual social network data. In the paper they considered two of the most fundamental definitions of influence, one based on a small set of "snapshot"' observations of a social network and the other based on detailed temporal dynamics. The former is particularly useful because large-scale social network data sets are often available only in snapshots or crawls. The latter however provides a more detailed process model of how influence spreads. They studied the relationship between these two ways of measuring influence, in particular establishing how to infer the more detailed temporal measure from the more readily observable snapshot measure. They validated their analysis using the history of social interactions on Wikipedia; the result is the first large-scale study to exhibit a direct relationship between snapshot and temporal models of social influence. In our research we exhibits a dynamic study of behavior adoption over time.

Banerjee [8] , in his work showed an example, a person is in an unfamiliar town and she chooses Restaurant A through her own research. Upon reaching the place she finds it deserted but sees restaurant B is full. Now, if she takes into account that the diners in restaurant B have similar taste to hers and that they have their own set of information about the place then the logical choice for her is restaurant B. To analyze this further it can be assumed that each diner had imperfect but independent information about which restaurant is better. So, if restaurant B has more diners then it means that the collective information is more powerful than the person's private one and making this choice logical. In this case, we say that herding, or an information cascade, has occurred. This concept was also developed in other work around the same time by Bikhchandani, Hirshleifer, and Welch [12] , [75].

Thus, it can be said that an information cascade can take place when people make decisions sequentially. That is when people infer the result by observing their former counterparts and their actions, as is the case with the restaurant example. Here, people are not taking decisions on basis of being blind followers but informant ones, which is creating a information cascade. This is thus informed imitation though in many cases imitation does occur due to the desire to conform. Consider for example the following experiment performed by Milgram, Bickman, and Berkowitz in the 1960s [66]. In this experiment, people were asked to look up at the sky and the impact of this action on passerby was noted. It was seen when one person was standing looking up, almost no one notices. Then, when around 5 people were looking up a few noticed but not a very significant number. But, when around 15 people stood and looked up almost 45

It was observed that the force to conform becomes greater if the group involved in conforming becomes larger. Another explanation for this can also be information cascade. Initially with fewer numbers, passersby did not see a rational need to look up, but with growing number it was but logical to see if the information was there at all. Ultimately, information cascades may be at least part of the explanation for many types of imitation in social settings. This can be seen in everyday life where people do imitate others believing their information to be well thought out.

2.1.1 Diffusion in Networks

We now connect these two approaches by exploring some of the decision-making principles that can be used to model individual decision-making in a social network, leading people to align their behaviors with those of their network neighbors. The Diffusion of Innovations. We will consider specifically how new behaviors, practices, opinions, conventions, and technologies spread from person to person through a social network, as people influence their friends to adopt new ideas. Our understanding of how this process works is built on a long history of empirical work in sociology known as the diffusion of innovations [21], [59] , [67]. A number of now-classic studies done in the middle of the 20th century established a basic research strategy for studying the spread of a new technology or idea through a group of people, and analyzing the factors that facilitated or impeded its progress. Some of these early studies focused on cases in which the person-to-person influence was due primarily to informational effects: as people observed the decisions of their network neighbors, it provided indirect information that led them to try the innovation as well. Two of the most influential early pieces of research to capture such informational effects were

Ryan and Gross's study of the adoption of hybrid seed corn among farmers in Iowa [61] and Coleman, Katz, and Menzel's study of the adoption of tetracycline by physicians in the United States [21] . In Ryan and Gross's study, they interviewed farmers to determine how and when they decided to begin using hybrid seed corn; they found that while most of the farmers in their study first learned about hybrid seed corn from salesmen, most were first convinced to try using it based on the experience of neighbors in their community. Coleman, Katz, and Menzel went further when they studied the adoption of a new drug by doctors, mapping out the social connections among the doctors making decisions about adoption.

While these two studies clearly concerned very different communities and very different innovations, they - like other important studies of that period - shared a number of basic ingredients. In both cases, the novelty and initial lack of understanding of the innovation made it risky to adopt, but it was ultimately highly beneficial; in both cases, the early adopters had certain general characteristics, including higher socio-economic status and a tendency to travel more widely; and in both cases, decisions about adoption were made in the context of a social structure where people could observe what their neighbors, friends, and colleagues were doing.

Other important studies in the diffusion of innovations focused on settings in which decisions about adoption were driven primarily by direct-benefit effects rather than informational ones. A long line of diffusion research on communication technologies has explored such direct-benefit effects; the spread of technologies such as the telephone, the fax machine, and e-mail has depended on the incentives people have to communicate with friends who have already adopted the technology [64], [47]. As studies of this type began proliferating, researchers started to identify some of the common principles that applied across many different domains. In his influential book on the diffusion of innovations, Everett Rogers gathered together and articulated a number of these principles [59], including a set of recurring reasons why an innovation can fail to spread through a population, even when it is has significant relative advantage compared to existing practices. In particular, the success of an innovation also depends on its complexity for people to understand and implement; its observability, so that people can become aware that others are using it; its trial ability, so that people can mitigate its risks by adopting it gradually and incrementally; and perhaps most crucially, its overall compatibility with the social system that it is entering. Related to this, the principle of homophily can sometimes act as a barrier to diffusion: since people tend to interact with others who are like themselves, while new innovations tend to arrive from "outside" the system, it can be difficult for these innovations to make their way into a tightly-knit social community.

Shalizi [63] considered processes on social networks that can potentially involve three factors: homophily, or the formation of social ties due to matching individual traits; social contagion, also known as social influence; and the causal effect of an individual's covariates on their behavior or other measurable responses. They showed that, generically, all of these are confounded with each other. Distinguishing them from one another requires strong assumptions on the parametrization of the social process or on the adequacy of the covariates used (or both). In particular they demonstrated, with simple examples, that asymmetries in regression coefficients cannot identify causal effects, and that very simple models of imitation (a form of social contagion) can produce substantial correlations between an individual's enduring traits and their choices, even when there is no intrinsic affinity between them. They also suggested some possible constructive responses to these results.

2.1.2 Modeling Diffusion through a Network

We build our model for the diffusion of a new behavior in terms of a more basic, underlying model of individual decision-making: as individuals make decisions based on the choices of their neighbors, a particular pattern of behavior can begin to spread across the links of the network. To formulate such an individual-level model, it is possible to start either from informational effects [27], [7], [32] or direct-benefit effects [13], [25], [52] , [76]. In this chapter, we will focus on the latter, beginning with a natural model of direct-benefit effects in networks due to Stephen Morris [52].Network models based on direct-benefit effects involve the following underlying consideration: you have certain social network neighbors | friends, acquaintances, or colleagues and the benefits to you of adopting a new behavior increase as more and more of these neighbors adopt it. In such a case, simple self-interest will dictate that you should adopt the new behavior once a sufficient proportion of your neighbors have done so. For example, you may find it easier to collaborate with co-workers if you are using compatible technologies; similarly, you may find it easier to engage in social interaction all else being equal with people whose beliefs and opinions are similar to yours.

One of the fundamental things we learn from studying diffusion is that there is a crucial difference between learning about a new idea and actually deciding to adopt it. This contrast was already important in the early days of diffusion research. For example, Figure 19.10 comes from the original Ryan-Gross study of hybrid seed corn [61]; it shows a clear wave of awareness of this innovation that significantly precedes the wave of adoptions. Our models also illustrate this contrast. If we imagine that people first hear about an innovation when any of their neighbors first adopts, then we see for example in Figure 19.5 that nodes 4 and 9 are aware of A as a new behavior right away, but it takes further time for them to actually adopt it. In an even stronger direction, nodes 2 and 11-14 eventually become aware of A but never adopt it.

Centola and Macy [29] and Siegel [65] make the interesting observation that threshold models for diffusion thus highlight an interesting subtlety in the strength-of-weak-ties theory. Recall that the strength of weak ties is rooted in the idea that weak social connections, to people we see infrequently, often form local bridges in a social network. They therefore provide access to sources of information things like new job opportunities | that reside in parts of the network we otherwise wouldn't have access to. The trade-offs inherent in this picture have been used to motivate some of the reasons why many social movements tend to build support locally and relatively slowly. Although a world-spanning system of weak ties in the global friendship network is able to spread awareness of a joke or an on-line video with remarkable speed, political mobilization moves more sluggishly, needing to gain momentum within neighborhoods and small communities.

Thresholds provide a possible reason: social movements tend to be inherently risky undertakings, and hence individuals tend to have higher thresholds for participating; under such conditions, local bridges that connect very different parts of the network are less useful. Such considerations provide a perspective on other well-known observations about social movements in the diffusion literature, such as Hedstrom's findings that such movements often spread geographically [38] , and McAdam's conclusion that strong ties, rather than weak ties, played the more significant role in recruitment to student activism during Freedom Summer in the 1960s [48] , [49].

2.1.3 Knowledge, Thresholds, and Collective Action

We now switch our discussion to a related topic that integrates network effects at both the population level and the local network level. We consider situations where coordination across a large segment of the population is important, and the underlying social network is serving to transmit information about people's willingness to participate. Collective Action and Pluralistic Ignorance. A useful motivating example is the problem of organizing a protest, uprising, or revolt under a repressive regime [20], [68], [34]. Imagine that you are living in such a society, and are aware of a public demonstration against the government that is planned for tomorrow. If an enormous number of people show up, then the government will be seriously weakened, and everyone in society including the demonstrators will benefit. But if only a few hundred show up, then the demonstrators will simply all be arrested (or worse), and it would have been better had everyone stayed home. In such circumstances, what should you do? This is an example of a collective action problem, where an activity produces benefits only if enough people participate. In this way, it is reminiscent of our analysis in Chapter 3 of population-level network effects: as with joining a large-scale demonstration, you only want to buy a fax machine if enough other people do. The starker setting of the present example highlights a few points, however. In the case of a fax machine, you can watch the experience of early adopters; you can read reviews and advertisements; you can canvass a wide array of friends and colleagues to see what they plan to do. Due to the much stronger negative payoffs associated with opposing a repressive government, many of these options are closed to you - you can talk about the idea with a small number of close friends whom you trust, but beyond this your decision about whether to show up for the demonstration is made difficult by a lack of knowledge of other people's willingness to participate, or of their criteria for deciding whether to participate.

These considerations illustrate some of the reasons why repressive governments work so hard to limit communication among their citizens. It is possible, for example, that a large fraction of the population is strong enough in its opposition to be willing to take extreme measures, but that most of these people believe they're in a small minority and hence view opposition as too risky. In this way, a government could survive long after there is enough strong opposition in principle to get rid of it.

This phenomenon is known as pluralistic ignorance [55], in which people have wildly erroneous estimates about the prevalence of certain opinions in the population at large. It is a principle that applies widely, not just in settings where a central authority is actively working to restrict information. For example, a survey conducted in the U.S. in 1970 (and replicated several times in the surrounding years with similar results) showed that while only a minority of white Americans at that point personally favored racial segregation, significantly more than 50

A Model for the Effect of Knowledge on Collective Action. Let's consider how the structure of the underlying social network can affect the way people make decisions about collective action, following a model and a set of illustrative examples proposed by Michael Chwe [20], [68]. Suppose that each person in a social network knows about a potential upcoming protest against the government, and she has a personal threshold which encodes her willingness to participate. A threshold of k means, "I will show up for the protest if I am sure that at least k people in total (including myself) will show up."

2.1.4 Applications of Social Network Analyzes

We now explore related application of social network analyzes in this section. Kleinberg , in his recent paper [5] analyzed romantic partnerships and the dispersion of social ties on Facebook. A crucial task in the analysis of on-line social-networking systems is to identify important people — those linked by strong social ties — within an individual's network neighborhood. In the paper, they investigate this question for a particular category of strong ties, those involving spouses or romantic partners. They organize their analysis around a basic question: given all the connections among a person's friends, can one recognize his or her romantic partner from the network structure alone? Using data from a large sample of Facebook users, they found that this task can be accomplished with high accuracy, but doing so requires the development of a new measure of tie strength that we term 'dispersion' — the extent to which two people's mutual friends are not themselves well-connected. The results offer methods for identifying types of structurally significant people in on-line applications, and suggest a potential expansion of existing theories of tie strength. This issue has also been dealt by [35].

Richardson [23] describe one of the major applications of data mining is in helping companies determine which potential customers to market to. If the expected profit from a customer is greater than the cost of marketing to her, the marketing action for that customer is executed. So far, work in this area has considered only the intrinsic value of the customer (i.e, the expected profit from sales to her). They propose to model also the customer's network value: the expected profit from sales to other customers she may influence to buy, the customers those may influence, and so on recursively. Instead of viewing a market as a set of independent entities, they view it as a social network and model it as a Markov random field. They show the advantages of this approach using a social network mined from a collaborative filtering database. Marketing that exploits the network value of customers—also known as viral marketing—can be extremely effective, but is still a black art. Their work can be viewed as a step towards providing a more solid foundation for it, taking advantage of the availability of large relevant databases.

Gruhl [37] in his paper studied the dynamics of information propagation in environments of low-overhead personal publishing, using a large collection of weblogs over time as our example domain. They characterized and model this collection at two levels. First, they present a macroscopic characterization of topic propagation through our corpus, formalizing the notion of long-running "chatter" topics consisting recursively of "spike" topics generated by outside world events, or more rarely, by resonances within the community. Second, they present a microscopic characterization of propagation from individual to individual, drawing on the theory of infectious diseases to model the flow. They proposed, validated, and employed an algorithm to induce the underlying propagation network from a sequence of posts, and report on the results.

Wasserman [70] focused on relationships among social entities, is used widely in the social and behavioral sciences, as well as in economics, marketing, and industrial engineering. Social Network Analysis: Methods and Applications reviewed and discussed methods for the analysis of social networks with a focus on applications of these methods to many substantive examples.

2.2 Epidemic Diffusion

The idea of epidemic diffusion is fluently pointed out by author Gladwell in his book "The Tipping Point" [30]. The idea is simplistic, and the best way to understand the emergence of fashion trends, the ebb and flow of crime waves. The book argued that the transformation of unknown books into bestsellers or the rise in risk of teenage smoking is to think of them as epidemics. Gladwell pointed out that ideas, products, behavior and messages, all spread just like viruses do.As mentioned in the book by Malcolm Gladwell [30], few examples of epidemics in actions are rise of Hush Puppies and the fall of New York's crime rate. On the basis of all these spread lies underlying patterns.

Leskovec in his paper [45] presents an analysis of a person-to-person recommendation network, consisting of 4 million people who made 16 million recommendations on half a million products. They observe the propagation of recommendations and the cascade sizes, which they explain by a simple stochastic model. They analyze how user behavior varies within user communities defined by a recommendation network. Product purchases follow a 'long tail' where a significant share of purchases belongs to rarely sold items. They establish how the recommendation network grows over time and how effective it is from the viewpoint of the sender and receiver of the recommendations. While on average recommendations are not very effective at inducing purchases and do not spread very far, they present a model that successfully identifies communities, product, and pricing categories for which viral marketing seems to be very effective.

Leskovec [46], discuss about cost-effective outbreak detection in networks.- Given a water distribution network, where should we place sensors to quickly detect contaminants? Or, which blogs should we read to avoid missing important stories? These seemingly different problems share common structure: Outbreak detection can be modeled as selecting nodes (sensor locations, blogs) in a network, in order to detect the spreading of a virus or information as quickly as possible. They present a general methodology for near optimal sensor placement in these and related problems. They demonstrate that many realistic outbreak detection objectives (e.g., detection likelihood, population affected) exhibit the property of "submodularity". They exploit submodularity to develop an efficient algorithm that scales to large problems, achieving near optimal placements, while being 700 times faster than a simple greedy algorithm. They also derive online bounds on the quality of the placements obtained by any algorithm. Their algorithms and bounds also handle cases where nodes (sensor locations, blogs) have different costs. They evaluate their approach on several large real-world problems,including a model of a water distribution network from the EPA, andreal blog data. The obtained sensor placements are provably near optimal, providing a constant fraction of the optimal solution. They show that the approach scales, achieving speedups and savings in storage of several orders of magnitude. They also show how the approach leads to deeper insights in both applications, answering multi-criteria trade-off, cost-sensitivity and generalization questions.

Grassberger [36] studied on the critical behavior of the general epidemic process and dynamical percolation. Scaling laws are formulated for the behavior of a space-dependent fluctuating general epidemic process near the critical point. Restricted to stationary properties, these laws describe also the critical behavior of random percolation. Monte Carlo calculations are used to estimate the critical exponents and the universal shape of the propagating wave, in the case of 2-dimensional space.

Canonical texts like [3], [39] combines mathematical models with extensive use of epidemiological and other data. The most widely studied epidemiological models include the so-called homogeneous models [50], which assume that every individual has equal contact to others in the population and that the rate of infection is determined by the density of the infected population. Kephart and White [43] were among the first to propose epidemiology-based models (the KW model) to analyze the propagation of computer viruses on homogeneous networks. However, Prakash el.at. [57] presents with an overwhelming evidence that real networks including social networks and routers etc, follow a power law structure instead. Pastor-Satorras and Vespignani [56] studied viral propagation for random power-law networks, and showed low or nonexistent epidemic thresholds, meaning that even an agent with extremely low infectivity could propagate and persist in the network. They use the "mean-field" approach where all graphs with a given degree distribution are considered equal. There is no particular reason why all such graphs should behave similarly in terms of viral propagation. In a recent work, Castellano and Pastor-Satorras [16] empirically argue that some special family of random power-law graphs has a non-vanishing threshold under the SIR model in the limit of infinite size, but provide no theoretical justification.

The population dynamics underlying the diffusion of ideas hold many qualitative similarities to those involved in the spread of infections. In spite of much suggestive evidence this analogy is hardly ever quantified in-useful ways. The standard benefit of modeling epidemics is the ability to estimate quantitatively population average parameters, such as interpersonal contact rates, incubation times, duration of infectious periods, etc. In most cases such quantities generalize naturally to the spread of ideas and provide a simple means of quantifying sociological and behavioral patterns. In the paper [10], Bettencourt apply several paradigmatic models of epidemics to empirical data on the advent and spread of Feynman diagrams through the theoretical physics communities of the USA, Japan, and the USSR in the period immediately after World War II. This test case has the advantage of having been studied historically in great detail, which allows validation of our results. They estimate the effectiveness of adoption of the idea in the three communities and find values for parameters reflecting both intentional social organization and long lifetimes for the idea. These features are probably general characteristics of the spread of ideas, but not of common epidemics.

Viral marketing has been one of the favorite strategies for marketers to achieve deeper market penetration. As such, viral marketing like recommendation network based marketing depends on the dynamics of the social influential interaction. The dynamics of the recommendations in social networks and their impact on the desired outcome in the form of purchase decisions can be studied as per the theory of local interaction games. In the paper [9], Banerjee tries to explore the effects of various parameters on such outcomes and proposes a model for studying these interactions incorporating the game theory based models and the fuzzy logic.

Viral marketing takes advantage of networks of influence among customers to inexpensively achieve large changes in behavior. Richardson's research [58] seeks to put it on a firmer footing by mining these networks from data, building probabilistic models of them, and using these models to choose the best viral marketing plan. Knowledge-sharing sites, where customers review products and advise each other, are a fertile source for this type of data mining. Paper extends their previous techniques, achieving a large reduction in computational cost, and apply them to data from a knowledge-sharing site. They optimize the amount of marketing funds spent on each customer, rather than just making a binary decision on whether to market to him. They take into account the fact that knowledge of the network is partial, and that gathering that knowledge can itself have a cost. Their results show the robustness and utility of their approach.

Ugander et.al [69] presented that the concept of contagion has steadily expanded from its original grounding in epidemic disease to describe a vast array of processes that spread across networks, notably social phenomena such as fads, political opinions, the adoption of new technologies, and financial decisions. Traditional models of social contagion have been based on physical analogies with biological contagion, in which the probability that an individual is affected by the contagion grows monotonically with the size of his or her contact neighborhoodâĂŤthe number of affected individuals with whom he or she is in contact. Whereas this contact neighborhood hypothesis has formed the underpinning of essentially all current models, it has been challenging to evaluate it due to the difficulty in obtaining detailed data on individual network neighborhoods during the course of a large-scale contagion process. Here, they study this question by analyzing the growth of Facebook, a rare example of a social process with genuinely global adoption. They find that the probability of contagion is tightly controlled by the number of connected components in an individual's contact neighborhood, rather than by the actual size of the neighborhood. Surprisingly, once this "structural diversity" is controlled for, the size of the contact neighborhood is,in fact,generally a negative predictor of contagion. More broadly, their analysis shows how data at the size and resolution of the Facebook network make possible the identification of subtle structural signals that go undetected at smaller scales yet hold pivotal predictive roles for the outcomes of social processes.

Newman [54] mapped the SIR model to a percolation problem on a network and studied thresholds for multiple competing viruses on special random graphs. Finally, Chakrabarti et.al. [17] and Ganesh et.al. [28] gave the threshold for the SIS model on arbitrary undirected networks. However, none of the earlier works focuses on long term effect of epidemic models for multiple behavior adoption by networks with resource constraint.

Our work is mindful of this literature but is different in several aspects. Most of the previous literature focus on diffusion of a single behavior and have not considered long-term effects of behavior diffusion. No other work considered the concept of user resource constraints and thereby does not apply directly to our problem of long term adoption of behaviors. Works of [11] and [15] discuss the problem of multiple competing influences, but they also do not include the resource constraints or the effects of individual parameter variation over time. To our knowledge, the present work is the first investigation of longitudinal effects of multiple behavior diffusion in a resource constrained social network.

Chapter 3

DESCRIPTIONS OF EPIDEMIC MODELS OF BEHAVIOR ADOPTION UNDER RESOURCE CONSTRAINT

3.1 Introduction

In this chapter, we introduce our behavior adoption models. Table 3.1 and Table 3.2 lists epidemic models and common terminology. In Section 3.2 , we describe our model of multiple behavior diffusion in a resource constrained network in the most general form. In Section 3.3, we point out all the model assumptions. Then, we introduce metrics for performance measurement in Section 3.4 followed by detailed descriptions of epidemic models in Section 3.5 and experimental results from simulations .

Table 3.1: Epidemic behavior models

Shortforms	**Descriptions**
SVS	Susceptible-Volatile-Susceptible
SIR	Susceptible-Infected-Recovered
$SIRS$	Susceptible-Infected-Recovered-Susceptible
$SVRS$	Susceptible-Volatile-Recovered-Susceptible

Table 3.2: Common Terminologies

Symbols	Descriptions
$G=(V,E)$	an undirected graph, each node $v \in V$ of the graph G represents an individual and an edge $e \in E$ between two nodes indicate a social relationship between the two individuals.
e	an edge
k	number of behaviors
i	each behavior
c^i	cost of behavior i and $0 \le c^i$
u^i	utility of behavior i and $u^i \le 1$
r_v	resource of node v and $0 \le r_v \le 1$
N_v	denote the set of neighbors of v in a network.
θ_v^i	fixed threshold for behavior i of each individual node $v \in V$
l_v^i	local network utility defined as the sum of influence weights–the social signal–exerted on v by its neighbors who have adopted behavior i
p_v^i	payoff for a behavior i is defined as the weighted sum of the intrinsic utility u^i and the local network utility l_v^i. That is, $p_v^i = wu^i + (1-w)l_v^i$. Where, w denotes the relative weight of the intrinsic utility.
d_I	infected duration for behavior i
d_R	recovered duration for behavior i
KS	adopted by knapsack
$\overline{KS}$	dropped by knapsack

3.2 A Model of Multiple Behavior Diffusion in a Resource Constrained Social Network

We now describe the model for each user, the properties of each behavior and the behavior adoption process. Conceptually our behavior adoption model can be described as follows - an individual adopts a new behaviors if the behavior has some value to him i.e. he has some interest in the behavior (intent), many of her friends have adopted the behavior (social signal), and she has enough available resource to pursue it (resource).

We represent the social network by an undirected graph $G = (V, E)$. Each node $v \in V$ of the graph G represents an individual and an edge $e \in E$ between two nodes indicate a social relationship between the two individuals.

We wish to spread k behaviors in the social network. Each behavior i has an associated cost c_i and a utility u_i. The cost refers to the cost of adoption and the utility refers to the intrinsic utility gained by an individual by adopting this behavior. In a simplification, we assume that both the cost c_i and the utility u_i of behavior i are intrinsic to the behavior and independent of the individual who adopts the behavior. Without loss of generality, we assume that $0 \leq c_i, u_i \leq 1$.

Individuals are resource constrained: an individual may have limited time, money or may not possess other material resources to adopt a behavior. Therefore, we assign a *fixed* resource $r(v)$ for each individual $v \in V$ towards adopting behaviors. The resource satisfies $0 \leq r(v) \leq 1$. For example, if we assume that individuals' resources are independent and identically distributed then the resource value $r(v)$ can be assumed to be obtained from a uniformly distributed random variable $U(0, 1)$. Let $N(v)$ denote the set of neighbors of v in the network. Then we assume that a neighboring node u asserts a social influence on node v with weight $1/|N(v)|$.

An individual will adopt a behavior i when she receives a strong social signal, has the resources to do so and when there is sufficiently high payoff in adopting the behavior. A behavior is a likely candidate for adoption when the strength of social signal exceeds a threshold, and the individual has enough resource to adopt the behavior. Figure 3.1 depicts the situation where the candidate behaviors are those for which the high social signal and resource availability conditions are met. We assume that each individual v has a different, fixed, threshold $\theta_i(v)$ for each behavior, and that each threshold is obtained independently from a uniformly distributed random variable $U(0,1)$. The strength of social signal is measured by $l_i(v)$ which is defined as the sum of influence weights—the social signal—exerted on v by its neighbors who have adopted behavior i. The payoff $p_i(v)$ for a behavior i is defined as the weighted sum of the intrinsic utility u_i and the local network utility $l_i(v)$. That is, $p_i(v) = w u_i + (1-w) l_i(v)$. Where, w denotes the relative weight of the intrinsic utility. Figure 3.2 shows this situation where the payoff is determined by social signal and intent. An individual adopts only those candidate behaviors that have high payoff (shown as the intersection between Candidate and Payoff in Figure 3.2). If there are multiple candidate behaviors, then an individual adopts a subset of candidate behaviors that maximizes total payoff.

Let us examine the diffusion of behavior over time, to illuminate the key ideas. The process takes place over discrete epochs [1] . We assume each node is aware of the behaviors adopted by her neighbors. The individual v first identifies all candidate behaviors. A behavior j is a candidate to be adopted if two conditions hold. First the social signal strength for behavior j must exceed the threshold for that behavior at node v, i.e. $l_j(v) \geq \theta_j(v)$. Second, the individual v must have the resources to adopt the behavior, i.e. $r(v) \geq c_j$. The first condition is the familiar Linear Threshold (LT) model [41]. Since there are

[1]Notice that while actions in a network are asynchronous, we can choose an appropriate time granularity for analysis to assume synchronized decision making.

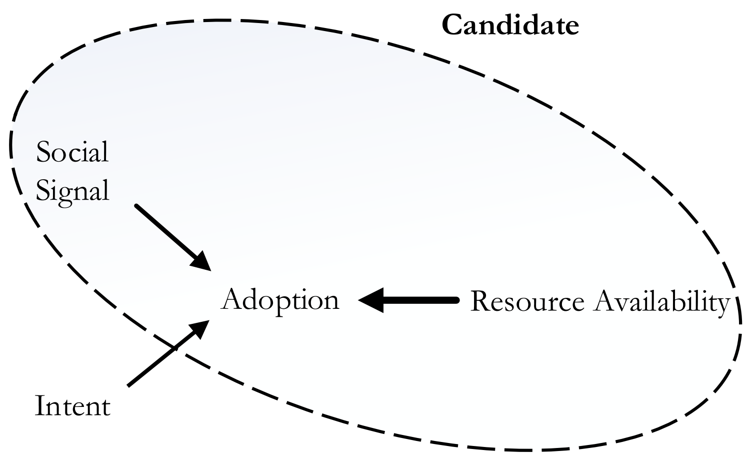

Figure 3.1: Sufficient social signal, intent and resource make an individual a candidate for adoption of a behavior.

multiple behaviors, the individual v chooses a subset of candidate behaviors that maximizes the total payoff subject to the condition that the sum of the adoption costs of the behaviors is less than the resource constraint. Let B_c be the set of candidate behaviors for an individual v. So v adopts a set of behaviors $B \subseteq B_c$ that maximizes $\sum_{i \in B} p_i(v)$ subject to the constraint that $\sum_{i \in B} c_i \leq r(v)$. At every epoch, the individual v evaluates all behaviors, including behaviors already adopted, to evaluate payoff. The behavior diffusion process continues until no additional adoption is possible.

In our diffusion model, we assume that the total resources available $r(v)$ at each node are known, while the threshold for adoption θ for any behavior is unknown. This assumption is reasonable if when people are willing to make public their available resources to participate in a set of behaviors. This can arise say in a private, mobile social network app focused on adoption of healthy behaviors including wellness, healthy eating and exercise, where individuals join the network to participate in healthy behaviors but each

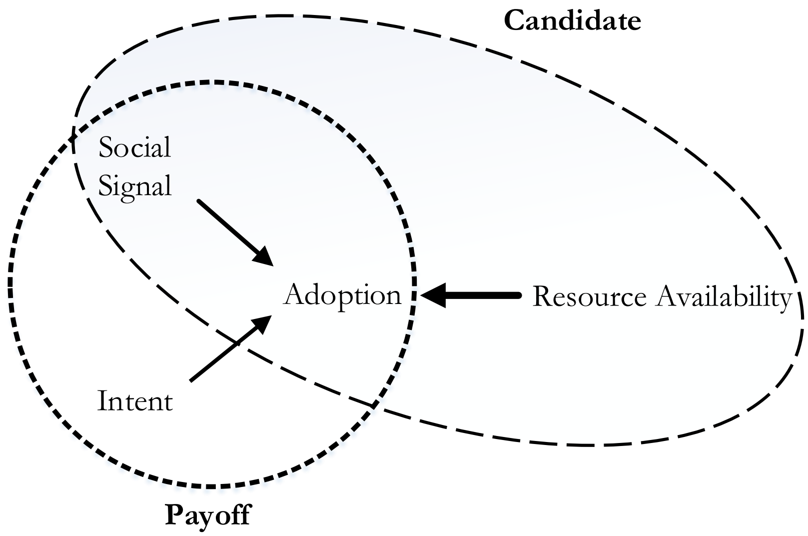

Figure 3.2: Payoff for adopting a behavior comes from an individual's intent and social signal.

individual is resource limited. An individual may declare that she has only one hour to spend on exercise each week, but would like to be nudged to participate in a health-related activity.

Figure 3.3 shows an illustration of the spread of behaviors with a four node network where three different behaviors - recycling, using public transport and eating locally grown food denoted by behaviors 1, 2 and 3 respectively. At time step 0 the state of the network is shown in 3.4. At this time step, for v, the social signal of eating locally grown food is weak. So v considers only recycling and using public transport for adoption. After maximizing payoff subject to the resource constraint, v adopts only recycling. Although public transport has strong social signal, v cannot adopt that behavior because it does not have enough resource. Notice that the payoff for recycling is higher than that of public transport, though the intrinsic utility of recycling was lower than that of public transport.

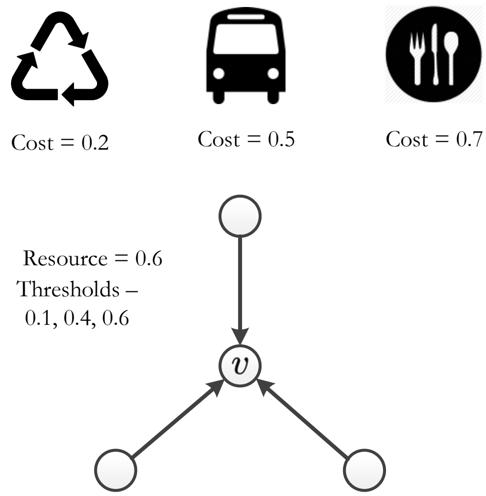

Figure 3.3: The three behaviors - (1) recycling, (2) using public transport, and (3) eating organic food with respective costs as well as the network is shown. The intrinsic utility of the behaviors are same as the cost. So $c_1 = u_1 = 0.2$, $c_2 = u_2 = 0.5$, $c_3 = u_3 = 0.7$. Resource of the node v, $r(v) = 0.6$, and the thresholds are - $\theta_1(v) = 0.1$, $\theta_2(v) = 0.4$, $\theta_3 = 0.6$.

3.3 Model Assumptions

We assume several conditions while performing all the experiments in this chapter.

- We assume number of nodes in the network to be constant over time.
- We assume individuals' resources availability , cost of performing a behavior , adoption threshold of a behavior and global influence are constant over the period of time.
- We assume the number of behavior/s is constant over time.
- While performing all the simulations , we assume that the time progresses with every ticks. Time tick is calculated by each discrete epochs.
- Every individuals are time bounded, either by the infected duration d_I or the recovery duration d_R.

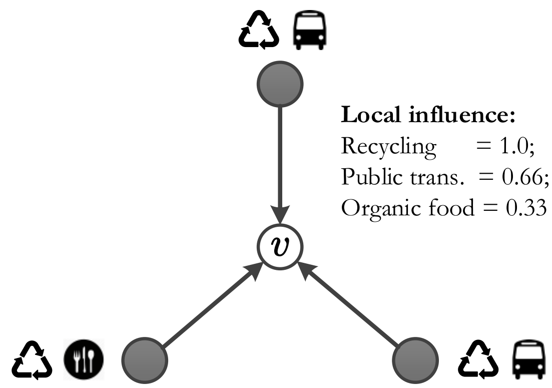

Figure 3.4: This is the network at time step 0. All three of the neighbors of v have adopted recycling, two of them have adopted public transport, and only one of them is eating organic food. v has not adopted any behavior yet. The local influences for the three behaviors are as follows - $l_1(v) = 1.0$, $l_2(v) = 0.66$, $l_3(v) = 0.33$.

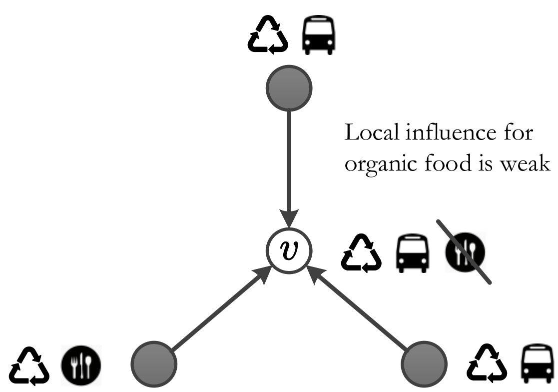

Figure 3.5: Local influence for organic food is less than the threshold, i.e $l_3(v) < \theta_3(v)$. So v will not consider organic food for adoption.

- We assume every individuals are in Susceptible (S) state from the very first time stamp.
- Another important assumption is that all resources r_v , behavior cost c^i , behavior utility u^i , infected duration d_I and recovered duration d_R are assigned in the very beginning and remains unchanged over time.

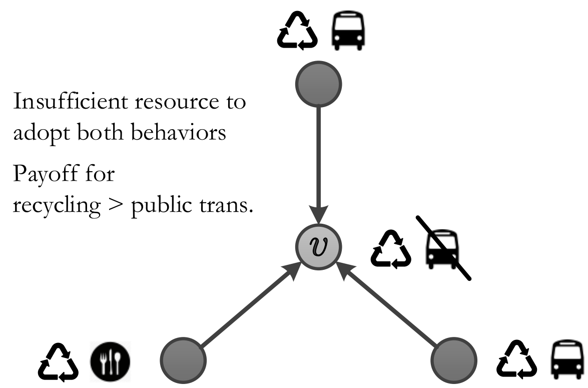

Figure 3.6: $c_1 + c_2 > r(v)$, so v's resource is insufficient for adopting both recycling and public transport. Payoff for recycling, $p_1(v) = 0.5 \times 0.2 + 0.5 \times 1.0 = 0.6$, and public transport, $p_2(v) = 0.5 \times 0.5 + 0.5 \times 0.66 = 0.58$ ($w = 0.5$). For v the payoff for recycling is higher than the payoff for using public transport though the intrinsic utility of public transport is higher than that of recycling. So v will adopt recycling at the end of time step 1.

3.4 Measurement of the Diffusion

We measure the effectiveness of the diffusion process with three metrics: total participation, total adoption and resource utilization. Since the behavior adoption is a stochastic process, we compute the expected value of each metric through simulation.

3.4.1 Total Participation

This metric counts the the expected number of individuals who have adopted at least one behavior (i.e. become active) during the process. For example, one goal for an advertiser of a product may be to maximize the total number of unique adoptees. Exact computation of this metric is shown to be #P-hard [19].

3.4.2 Total Adoption

In contrast to total participation, we need to keep track of the total number of adoptions of any behavior during the diffusion process. This metric counts the expected number of adoptions over all the behaviors. Notice that since an individual can adopt more than one behavior, total adoption cannot be less than the total participation. For the single behavior adoption problem, these two metrics will have the same value.

3.4.3 Resource Utilization

This metric captures the *efficiency* of the network to adopt costly behaviors. Not all resources available in a social network may be used for behavior adoption. This is because individuals have variable resources, and they may be unable to adopt the subset of behaviors that fully takes advantage of their desire to participate because of two reasons. First, they may have many more resources than needed to adopt the behavior. Second, if their friends have limited resources, then the social signals that they receive will be about adopting low-cost resources, and hence a particular individual may never see costly behaviors in their social circle that they could potentially adopt. Let us assume that a node v with resource r_v has adopted one or more behaviors. Let s_v be the amount of resource that v has used to adopt those behaviors, where $s \leq r_v$. Therefore, the individual has $r_v - s_v$ amount of his resource remaining unused. *Resource utilization* is the expected value of the ratio $\sum_{v \in V} s_v / \sum_{v \in V} r_v$ i.e. the ratio of total utilized resource to the total amount of available resource of all the individuals in the social network.

3.5 SVS Model of Multiple Behavior Diffusion in a Resource Constrained Network

For the SVS (Susceptible-Volatile-Susceptible) model , we shall assume that the adoption of behaviors is “Sticky”, that is, once a node adopts a behavior it never gets rid of that. This simplification is also known as *progressive* behavior adoption [41]. Notice that once a node adopts a behavior its resource to adopt other behaviors decreases. The “Sticky" model is named as SVS (Susceptible-Volatile-Susceptible) . Figure 3.5 shows the state-diagram of SVS model. Every individual remains in susceptible state (S) , where she is prone to adopt a behavior and remains until the local signal l_v^i of behavior i for node v is greater than the adoption threshold θ_v^i . An individual also remains in the Susceptible state (S) when she is not picked by knapsack $\overline{KS}$, indicating that the payoff p^i is low. Once the required conditions are met, node v adopts behavior i and moves to the volatile(V) state. The reason of naming the state to be volatile, because any competitive behavior with higher payoff can knock out the existing behavior . Note that the model is behavior dependent and individual independent. An individual adopts a behavior and continues to stay in V state when the behavior is chosen by knapsack.

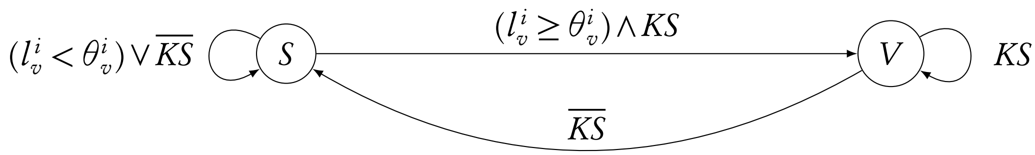

Figure 3.7: State-Diagram for SVS model: where S is the susceptible state where an individual is prone to adopt a behavior. An individual remains in the susceptible state if local utility l_v^i of behavior i for node v is lower than the threshold θ_v^i for behavior i, or dropped in knapsack $\overline{KS}$. Once the required conditions are met, node v adopts behavior i and moves to the volatile (V) state.

3.5.1 *Simulation Results and Discussion of SVS*

We perform the following simulation setup to execute all our experiments. There are fixed parameters and controlling parameter. Fixed parameters are the number of nodes as 100 , number of seeds as 10 . We are considering single behavior propagation with behavior cost as 0.5 and behavior utility as 0.5. We use “ Hill Climbing “ algorithm for the seed selection and perform uniform distribution of seeds across the network. Note that the resources of individuals, adoption threshold and adoption cost of the behavior remains constant across the network over time. For the experiment mention in Figure 3.8 , the adopted duration is set as 2. This experiment is a comparative study of SVS

epidemic model across time , over three different network topologies such as preferential attachment , small-world and complete regular graph.

Interestingly , in Figure 3.8 we observe a decrease in resource utilization in case of network topologies like smallworld and regular complete graph when compared with preferential attachment. Our explanation is that with increase of neighbors , an individual needs to have higher number of neighbors adopting the behavior to influence her.

3.6 SIR Model of Multiple Behavior Diffusion in a Resource Constrained Network

In this thesis, we examine the connection between epidemic models and the behavior adoption process. In classical SIR (Susceptible-Infected-Recovered) model, an infected individual recovers after a period of time. In our model , we represent the period of infection as d_I. This concept is similar to the idea presented by [17] . The recovered state (R) is synonymous to dropping the behavior for an indefinite time period. Ugander et. al. [69] in his paper also showed the concept of contagion has steadily expanded from its original grounding in epidemic disease to describe a vast array of processes that spread across networks. The key difference of our model with these previous works is the idea of multiple behavior adoption under limited resources. As shown in Figure 3.6 , an individual will remain in S state till her local social signal l_v^i is not greater than her behavior threshold θ_v^i or may not be selected in knapsack. As we are considering multiple behavior adoption, we also look into the payoff p_i factor to adopt a behavior. Only if the p_i of a behavior i is greater than threshold θ_v^i , then the individual adopts the behavior. Once a behavior is adopted , she is consider to be in infected state(I) and remains in the state for d_I period of time.

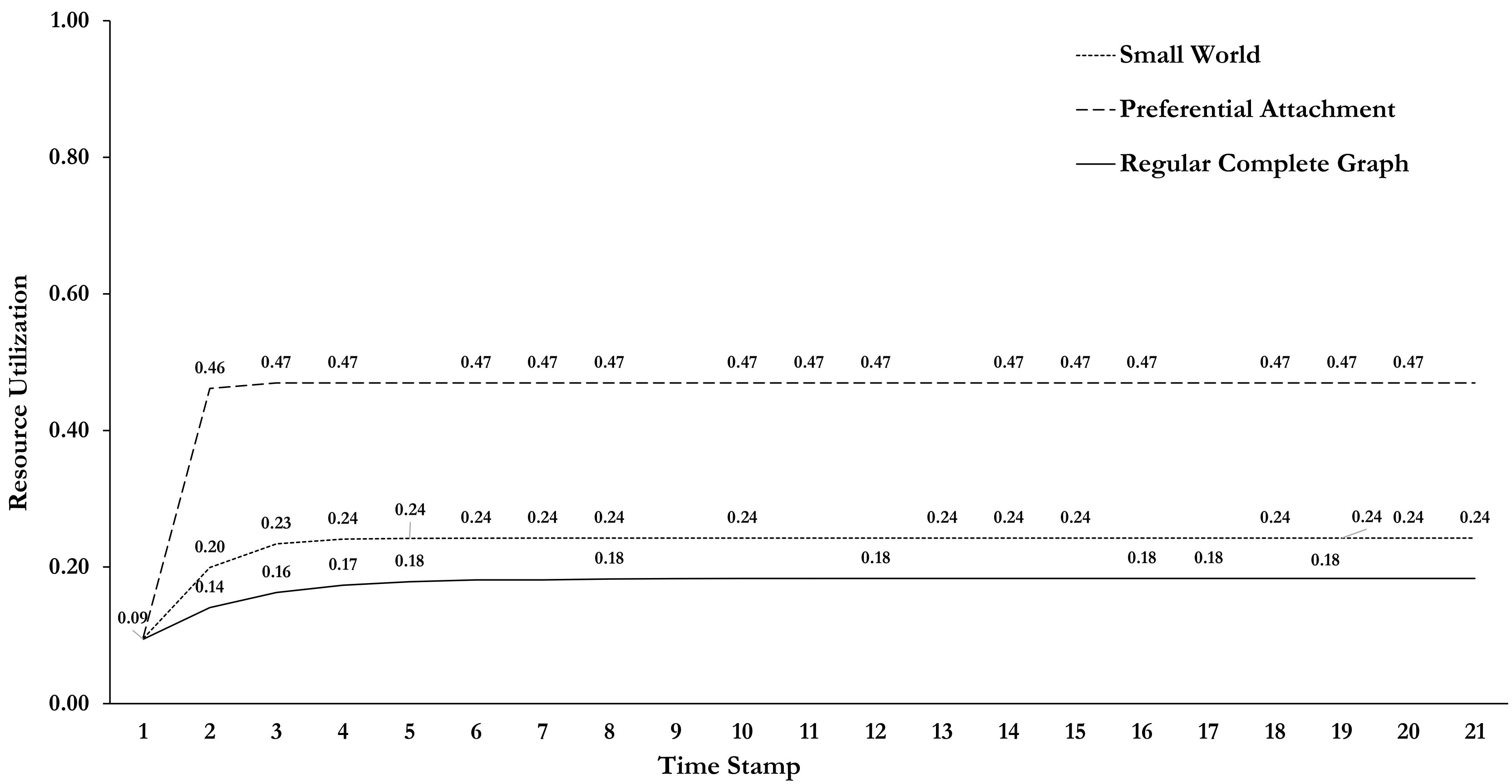

Figure 3.8: SVS model , key observation is the decrease in resource utilization as the number of neighbors increases. Fixed Parameters :Number of nodes = 100 , number of seeds = 10 , number of behavior = 1 , cost of behavior = 0.5 , utility = 0.5 , infected duration = 2 , recovered duration = 1 , seed selection algo = hill climbing , seed distribution = uniform. Varying Parameters : Epidemic Model used = SVS ,Network Topologies = [PA , smallworld , regular-complete-graph]

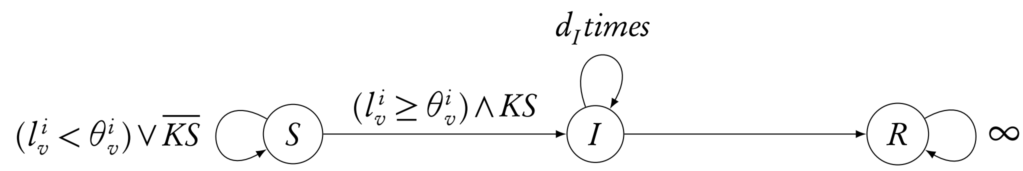

Figure 3.9: State-Diagram for SIR Model: where S is the susceptible state where an individual is prone to adopt a behavior. An individual adopts a behavior i and transforms to infected I state if local utility l_v^i of behavior i for node v is greater than the threshold θ_v^i for behavior i and get chosen in knapsack KS. Node v remains in I state for d_I (infected duration) times. After d_I period of time, node v moves to R (recovered) state and remains there for an infinite amount of time.

3.6.1 Simulation Results and Discussion of SIR

We setup the experiment in the following manner. There are fixed parameters and varying parameter. Fixed parameters are the number of nodes as 100 , number of seeds as 10 . We are considering single behavior propagation with behavior cost as 0.5 and behavior utility as 0.5. We use " Hill Climbing " algorithm for the seed selection and perform uniform distribution of seeds across the network. Note that the resources of individuals, adoption threshold and adoption cost of the behavior remains constant across the network over time. For the experiment mention in Figure 3.10 , the adopted duration is taken as 'two' time-stamp. This experiment is a comparative study of SIR epidemic model across time , over three different network topologies such as preferential attachment , small-world and complete regular graph.

In Figure 3.10 , we observe that the behavior propagation in the system dies down quickly when we perform the experiment on PA network. However, behavior propagation continues for a longer period of time in case of smallworld and complete regular graph. Dependency of behavior adoption on the number of neighbors of an individual provides a possible explanation. In smallworld and regular complete graph , behaviors propagate for a longer period due to higher regularity in number of neighbors of an indi-

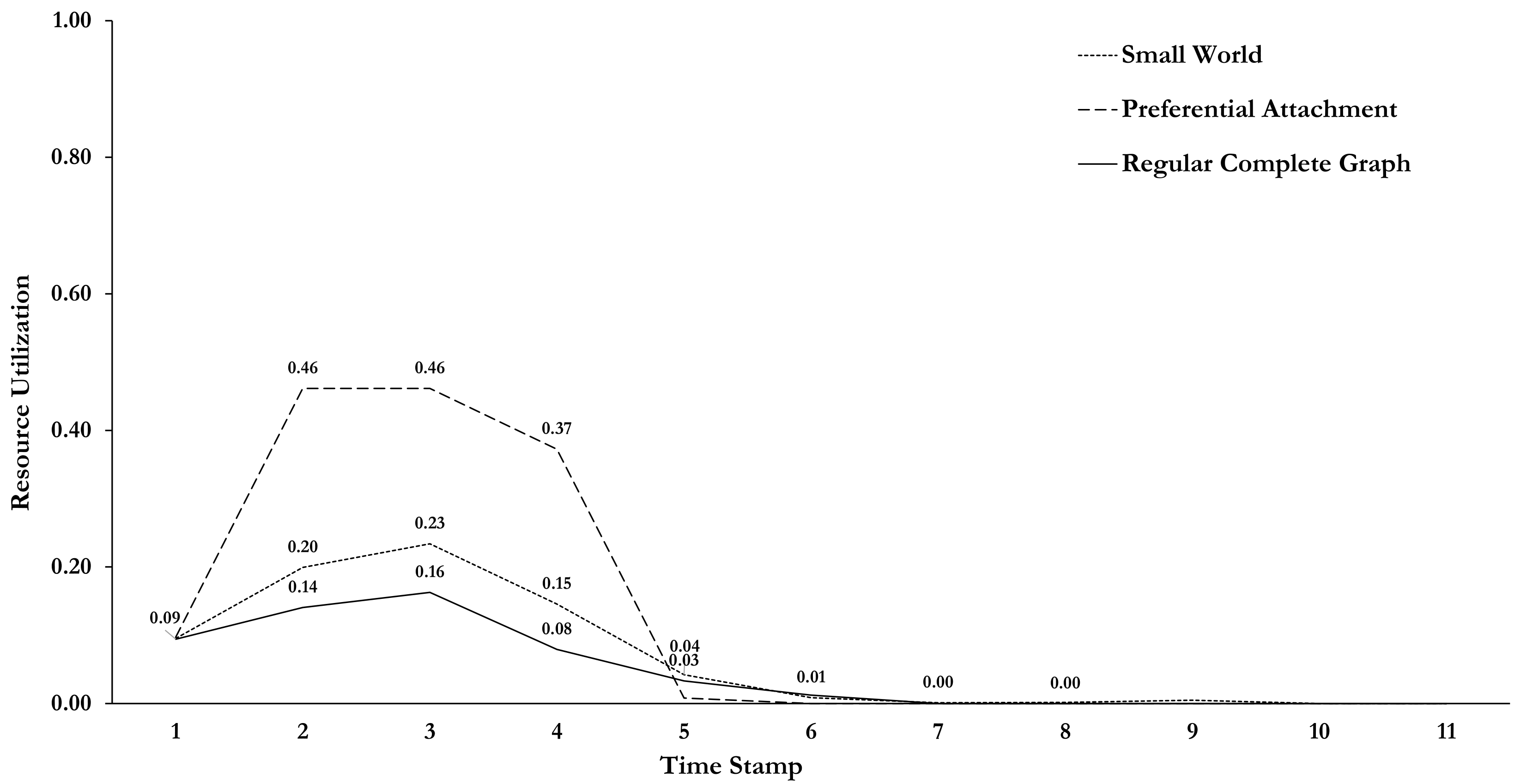

Figure 3.10: Experiment for *SIR* model over three different topologies and we observe the behavior to remain in the system for longer period of time in case of smallworld and complete regular graph comparing it with PA. Fixed parameter are : Number of nodes = 100 , number of seeds = 10 , number of behavior = 1 , cost of behavior = 0.5 , utility = 0.5 , infected duration = 2 , recovered duration = 1 , seed selection algo = hill climbing , seed distribution = uniform . Varying Parameters : Epidemic Model used = SIR, Network Topologies are PA, SW, CompleteRegularGraph]

vidual. In the case of PA network, we choose seeds which can impart the highest influence which eventually are some of the highest degree nodes. Hence, seed nodes propagates the behavior to peripheral nodes, and peripheral nodes are not powerful enough to impose influence on any other higher degree nodes which leading to the extinction of behavior propagation.

3.7 SIRS Model of Multiple Behavior Diffusion in a Resource Constrained Network

This model is an extension of the SIR model as we will see from its construction.

$$S \rightarrow I \rightarrow R \rightarrow S$$

The only difference is that it allows individuals in recovered state R to be free of infection and rejoin the susceptible state S after d_R recovered duration.

In Figure 3.7 , we illustrate $SIRS$ model using a state diagram. We assume every individuals are in susceptible S state till the social signal l_v^i is greater than the threshold θ_v^i and payoff p^i is high enough to get adopted by knapsack. Once an individual adopts a behavior, she moves to infected state I and continues till the infected duration d_I is over. Then an individual moves to recovered state R where she remains for d_R duration. This model can be explained with an example. An individual adopts jogging and performs it for two weeks , then she sprained her leg and failed to perform the behavior for the next week. At the end of her recovery time , she is again susceptible to continue with jogging if she has enough time and motivation to perform it.

3.7.1 Simulation Results and Discussion of SIRS

We setup the experiment in the following manner. There are fixed parameters and varying parameter. Fixed parameters are the number of nodes as 100 , number of seeds as 10 . We are considering single behavior propagation with behavior cost as 0.5 and behav-

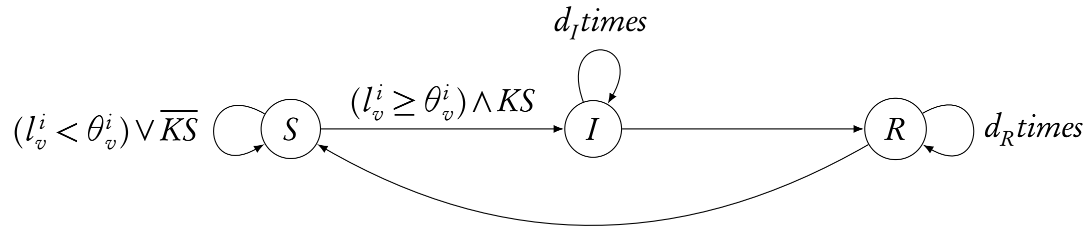

Figure 3.11: State-Diagram for $SIRS$ Model: where S is the susceptible state where an individual is prone to adopt behavior. An individual adopts a behavior i and transforms to infected I state if local utility l_v^i of behavior i for node v is greater than the threshold θ_v^i for behavior i and get chosen in knapsack KS. Node v remains in I state for d_I (infected duration) times. After d_I period of time, node v moves to R (recovered) state and remains there for d_R (recovered duration) period of time before converting to susceptible state.

ior utility as 0.5. We use " Hill Climbing " algorithm for the seed selection and perform uniform distribution of seeds across the network. Note that the resources of individuals, adoption threshold and adoption cost of the behavior remains constant across the network over time. For the experiment mention in Figure 3.12 , the adopted duration is taken as 'two' time-stamp. This experiment is a comparative study of $SIRS$ epidemic model across time , over three different network topologies such as preferential attachment , small-world and complete regular graph.

An interesting insight from this experiment is, in contrary of the classical $SIRS$ model where behavior propagation continues for infinite amount of time. Here we observe that propagation is network dependent. While performing the experiment on a more realistic network like preferential attachment, behavior adoption dies down after a period of time.

3.8 SVRS Model of Multiple Behavior Diffusion in a Resource Constrained Network

This $SVRS$ (Susceptible-Volatile-Recovered-Susceptible) model is a variant of $SIRS$ model.

$$S \rightarrow V \rightarrow R \rightarrow S$$

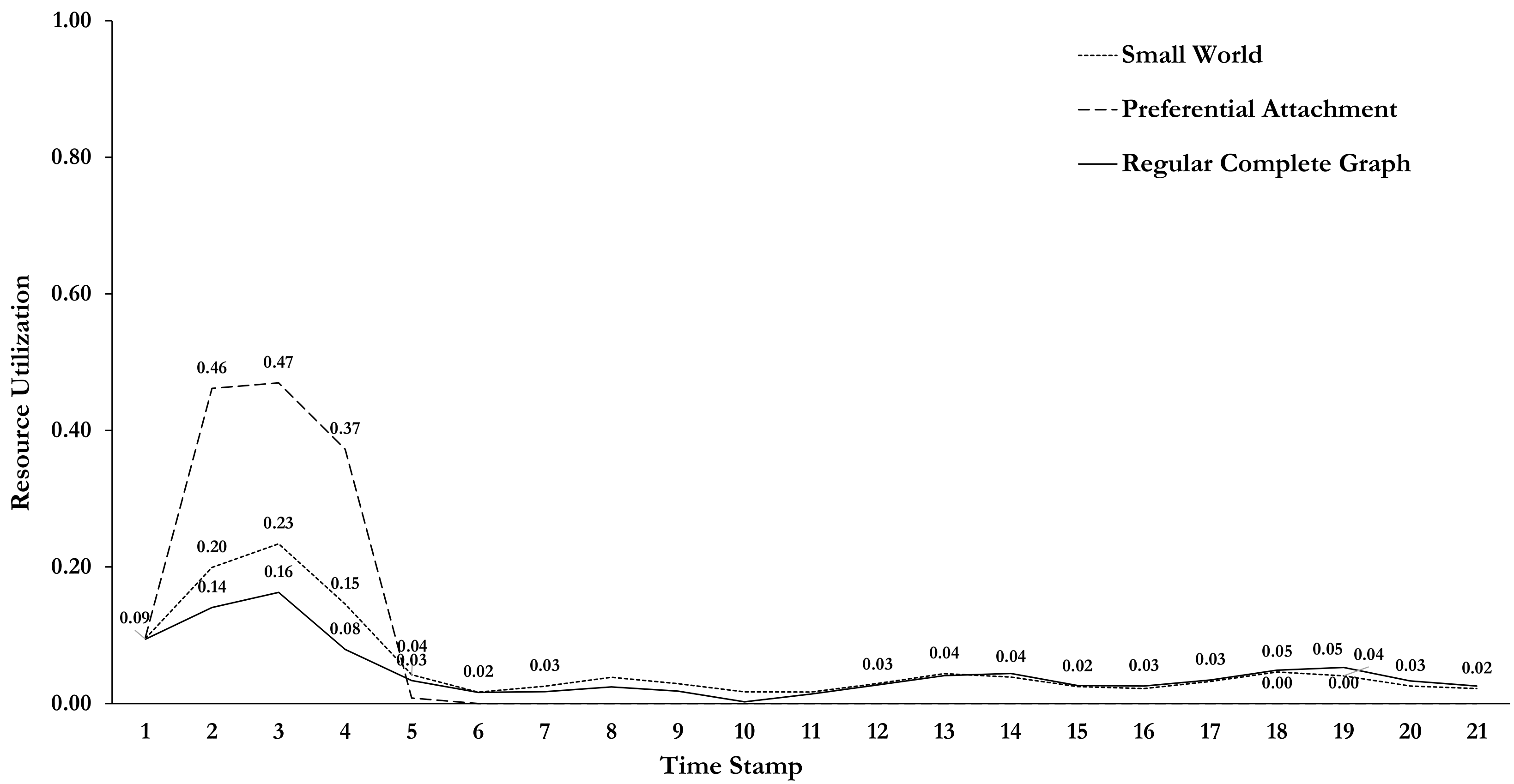

Figure 3.12: SIRS Model where interestingly we observe the behavior propagation is network dependent and dies out after a period of time in contrary of classical wisdom- Fixed Parameters :Number of nodes = 100 , number of seeds = 10 , number of behavior = 1 , cost of behavior = 0.5 , utility = 0.5 , infected duration = 2 , recovered duration = 1 , seed selection algo = hill climbing , seed distribution = uniform , resource = fixed , Threshold = fixed . Controlled Parameters : Epidemic Model used = SIRS , Network Topologies = PA, SW, CompleteRegularGraph

The major difference is, this model allows multiple behavior adoptions and behavior switches over time. For example, an individual starts jogging with her friends, but after a month she bought a bike and can start biking instead. In this example, there is a behavior switch taking place. In order to incorporate that individuals tend to adopt multiple behaviors and switching of behaviors , we introduced the volatile state V.

In Figure 3.8 , we explained $SVRS$ model using a state diagram. We assume every individuals are in susceptible S state till the social signal l_v^i is greater than the threshold θ_v^i and payoff p^i is high enough to get adopted by knapsack. Once an individual adopts a behavior, she moves to volatile state V. In volatile state she can either continue performing the behavior for d_I period of time , else she can adopt a new behavior which gives her higher payoff p^i . If she adopts a behavior and performs it for d_I period of time, she finishes performing the behavior and moves to the recovered state. For example, if an individual wants to write blog on a topic and she spends two days to perform that. Then she gets done with bloging and becomes susceptible to perform other activities. Since , after completion of performing a behavior an individual drops the behavior, we kept a recovered state in our model. The recovery duration can negligible and one can get susceptible to other behavior/s immediately .The main motivation behind developing this model is to capture the periodic nature of human behavior adoption in everyday life.

3.8.1 Simulation Results and Discussion of SVRS

We setup the experiment in the following manner. There are fixed parameters and varying parameter. Fixed parameters are the number of nodes as 100 , number of seeds as 10 . We are considering single behavior propagation with behavior cost as 0.5 and behavior utility as 0.5. We use " Hill Climbing " algorithm for the seed selection and perform uniform distribution of seeds across the network. Note that the resources of individuals, adoption threshold and adoption cost of the behavior remains constant across the net-

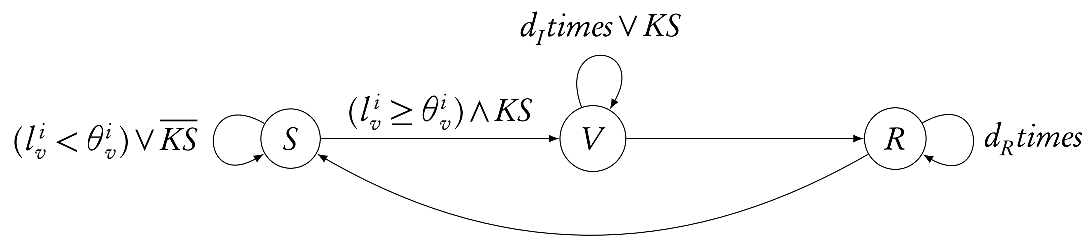

Figure 3.13: State-Diagram for $SVRS$ Model: where S is the susceptible state where an individual is prone to adopt behavior. An individual adopts a behavior i and transforms to infected I state if local utility V of behavior i for node v is greater than the threshold θ_v^i for behavior i and get chosen in knapsack KS. Node v remains in I state for d_I (infected duration) times. After d_I period of time, node v moves to R (recovered) state and remains there for d_R (recovered duration) period of time before converting to susceptible state.

work over time. For the experiment mention in Figure 3.14 , the adopted duration is taken as 'two' time-stamp. This experiment is a comparative study of $SVRS$ epidemic model across time , over three different network topologies such as preferential attachment [14] , small-world [71], [51] and complete regular graph.

Referring to Figure 3.12 and Figure 3.14, the difference between SIRS model and SVRS model is that there is a negligible recovery time . Hence even when the node is infected it is participating in knapsack. Due to negligible recovery time in the case of PA network, we inspected that the seed node is influencing peripheral nodes. Also,the behavior propagates back from the peripheral nodes to the seed nodes before the peripheral nodes recover. Therefore we find a nice periodic graph as a result.

3.9 Chapter Summary

In this chapter, we introduced all our epidemic models and discussed about the simulation results. We assume that the parameters like resource availability , adoption threshold , adoption cost and global influence are constant. Understanding the parameters variation over time is an interesting direction to explore. Chapter 5, we introduce those models.

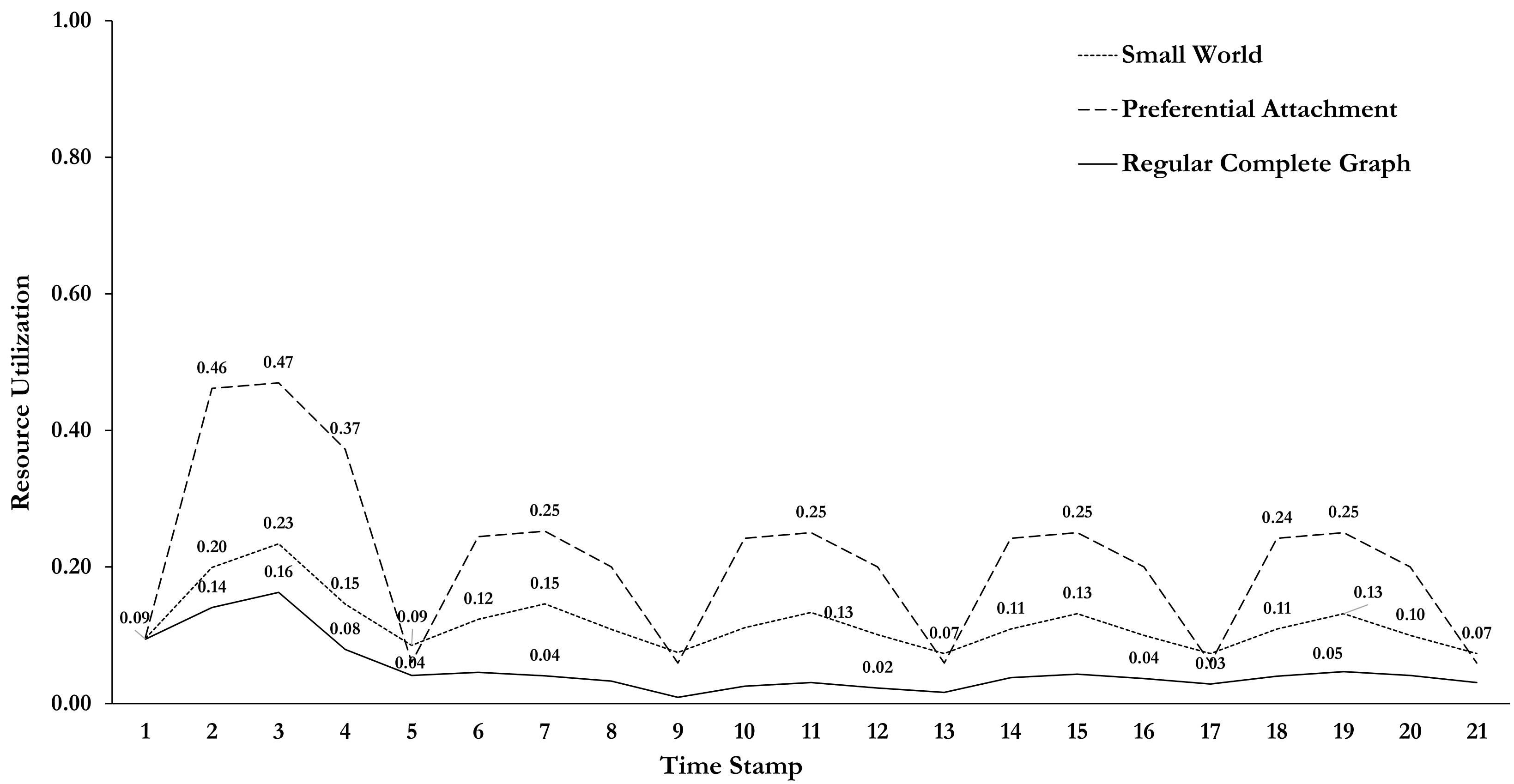

Figure 3.14: SVRS Model captures the periodic nature of an individual's behavior adoption. Interestingly the effect of seeds nodes is on the first period only and eventually resource utilization decreases with time - Fixed Parameters :Number of nodes = 100 , number of seeds = 10 , number of behavior = 1 , cost of behavior = 0.5 , utility = 0.5 , infected duration = 2 , recovered duration = 1 , seed selection algo = hill climbing , seed distribution = uniform , resource = fixed , Threshold = fixed . Varying Parameters : Epidemic Model Used = SVRS , Network Topologies = PA, SW and CompleteRegularGraph

Chapter 4

VARYING PARAMETERS OVER TIME

In this chapter, we formulate four different questions to better understand how behavior propagates in a resource constraint social network. In each section we model variation of different parameter over time and present results with discussions. In Section 4.1 , we discuss about resource availability variation over time. In Section 4.2 , we introduce adoption threshold variation over time. In Section 4.3, we model the variation of adoption cost over time. In Section 4.4 , we model the variation of global influence over time. In Section 4.5 , we compare all the parameter variations with combination of parameter variations over time.

4.1 *How varying resource availability of an individual can shape the behavior diffusion over time?*

In this section, we model weekly variation in availability of resources.Briefly, the idea is, an individual may not have equal resources available every day of a week. For example, an individual can have four hours for social service on the weekends but no time during the weekdays. We introduce two variants of resource availability in this chapter. In our model, we are representing the resource available with an individual v at time stamp t as $r_v(t)$.

4.1.0.1 Weekend Peaks

The main idea behind weekend peaks is that most of the population is having more resource during the same time. For example, most of the individuals are having more resources during their weekends than weekdays.

4.1.0.2 Random Peaks

In “Random Peaks" , we build the model in such a way that individuals can have higher resource on any two consecutive days of the week. For example, in a community, if one individual is having more time during weekends, another individual may have more time on Wednesdays and Thursdays.

4.1.1 Description of Model Resource Availability Variation

We want to model the variation of available resources for an individual over the different days of a week. Generally one can assume that more resource will be available for pursuing behaviors over the weekend than the week days. We want our model of resource variation to satisfy the following three conditions -

1. Experimenter can specify the fraction of average total weekly resource which will be available over the weekends. Let α be the fraction of average total weekly resource that is available on average over the weekends for the individuals. If we assign individual resources uniformly at random from $[0,1]$ for all the seven days of the week, then the value of α will be $2/7$, i.e. 28.57% of average total weekly resource will be available over the weekends.

2. The average total weekly resource should be 3.5. Note that this would be the average total weekly resource if we assign individual resource uniformly at random from $[0,1]$ for all the seven day (the daily average will be 0.5, so the weekly average will be $0.5 \times 7 = 3.5$). This will also be the seven day average for the single resource base case. We want to keep the average resource value fixed at 3.5, because we want to compare the resource variation scenario with the no variation model.

3. The average resource of the population for any day of the week must be at most 1.0.

4.1.1.1 *Naive Attempt:*

Let the average total weekly resource be x. So $x(1-\alpha) = 2.5$. Let $\beta \times 0.5$ be the average weekend resource. So $x\alpha = \beta$, i.e. $\beta = \frac{\alpha}{1-\alpha}.2.5$.

So we will first assign a number chosen uniformly at random from $[0,1]$ as the resource value for each of the seven days of the week. Then we will multiply the weekend resources by β. But this can make the weekend resource more than 1. So we will normalize the daily resources, i.e. divide the resource of each day by the sum of the total weekly resource. Although this approach satisfies condition 1 and 3, it does not satisfy condition 2 mentioned above. In fact the total weekly resource will be exactly 1.0 in this approach, which is not acceptable. The network will become much poorer than in the no variation model.

4.1.1.2 *A Better Approach:*

If we let α take any arbitrary value then it is not possible to satisfy condition 2. Precisely, α need to satisfy the following condition - $3.5\alpha \leq 2$, i.e. $\alpha \leq 0.57$. So we let the experimenter vary the value of α from 29% to 57%.

Let $\bar{w}$ be the average weekday resource. So we know that $5\bar{w} = 3.5(1-\alpha)$, i.e. $\bar{w} = \frac{3.5(1-\alpha)}{5}$.

For $0.29 \leq \alpha \leq 0.57$, $\bar{w} < 0.5$. So for the 5 weekdays we assign the resource value uniformly at random from $[0, 2\bar{w}]$. The average weekend resource will be $\frac{3.5\alpha}{2} > 0.5$. So we assign the two weekend resources uniformly at random from $[3.5\bar{w} - 1, 1]$.

4.1.2 *Fixing the values of α for resource availability variation*

As mentioned in the previous section, we are calculating the alpha value by performing exhaustive experiments and details of these experiments are given in Appendix

4.1.3 Simulation Setup for Varying Resource Availability Over Time

In this section, we are mentioning all the common simulation setups which we follow to obtain our results. We perform experiments and concluded combination of PA network and SVRS epidemic model is the best combination . Hence, all results follows with this combination. We also perform extensive experiments on other topologies and network , they are mentioned in the Appendix. There are few fixed parameters and few varying parameters. Fixed Parameters are number of nodes as 100 , number of seeds 10 , number of behavior 1 , cost of behavior 0.5 , utility 0.5 , infected duration 5 , recovered duration 2 , seed selection algo = hill climbing , seed distribution = uniform , Threshold average over 1000 runs . Varying Parameters : Epidemic Duration Model is SVRS, Network Topology as PA. We are comparing between "Weekend Peaks" and "Random Peaks".

4.1.4 Simulation Results and Discussions for Single Behavior

In this section, we present the simulation results from single behavior followed by discussions. Note that , all fixed parameters and controlled parameter used during each experiments are mentioned in details with each result.

Observations for Figure 4.1

- Main difference between "Weekend Peaks" and "Random Peaks" model is the maximum utilization. Maximum utilization when resources are synchronized is 0.46 whereas maximum utilization when resource variation is not synchronized is 0.23.

- Second, we observe a periodic nature for the "Weekend Peaks" resource variation. During weekends or during the higher resource timestamps, we observe a rise in resource utilization. Note that cost of behavior is 0.5 and during weekdays no node has resource as 0.5, hence during weekdays no adoption is taking place. We also

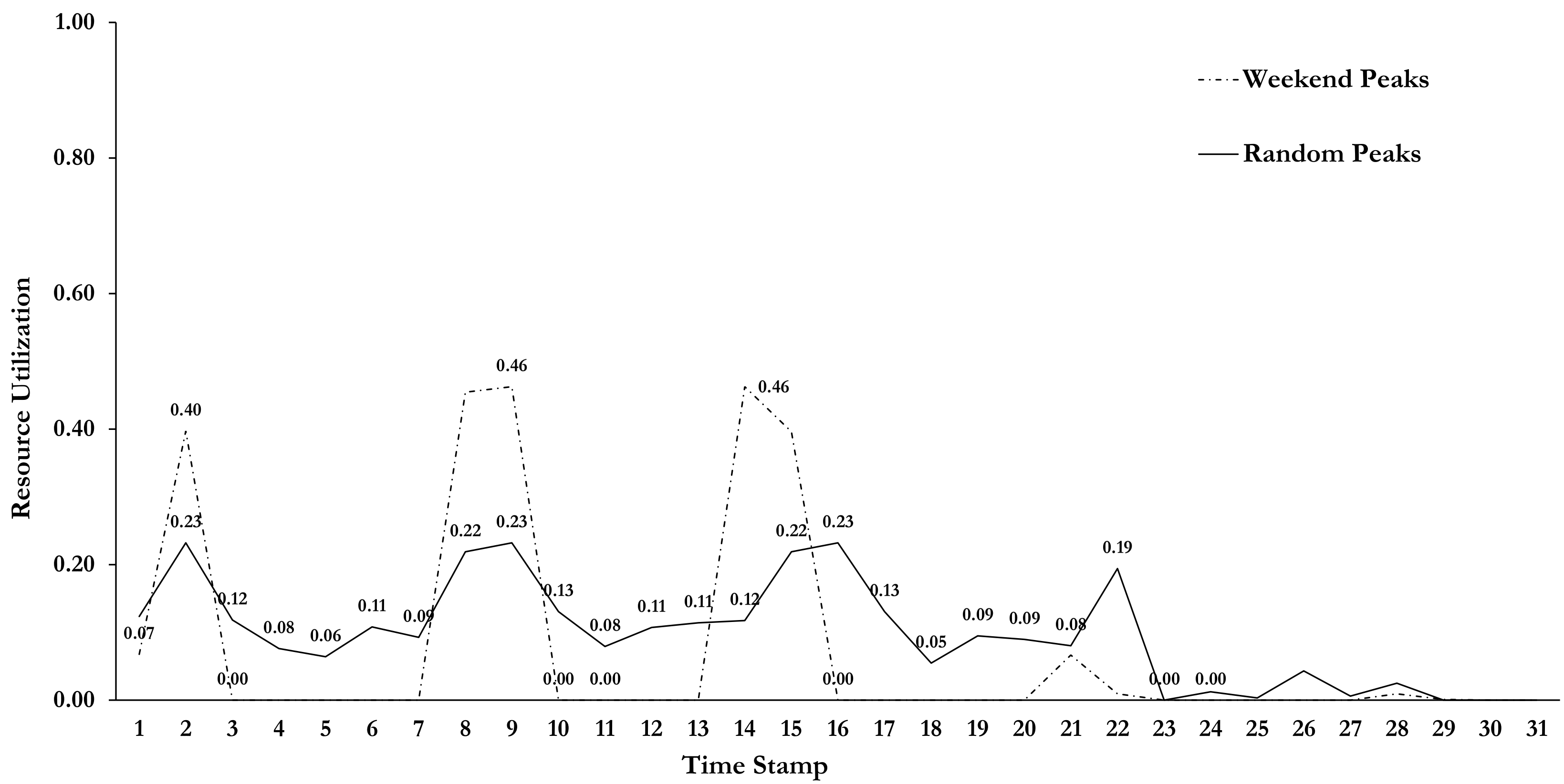

Figure 4.1: Resource Variation with time using SIR model on PA network - Fixed Parameters are number of nodes as 100 , number of seeds 10 , number of behavior 1 , cost of behavior 0.5 , utility 0.5 , infected duration 5 , recovered duration 2 , seed selection algo = hill climbing , seed distribution = uniform , Threshold average over 1000 runs . Varying Parameters : Epidemic Duration Model is SVRS, Network Topology as PA. We are comparing between “Weekend Peaks“ and “Random Peaks“.

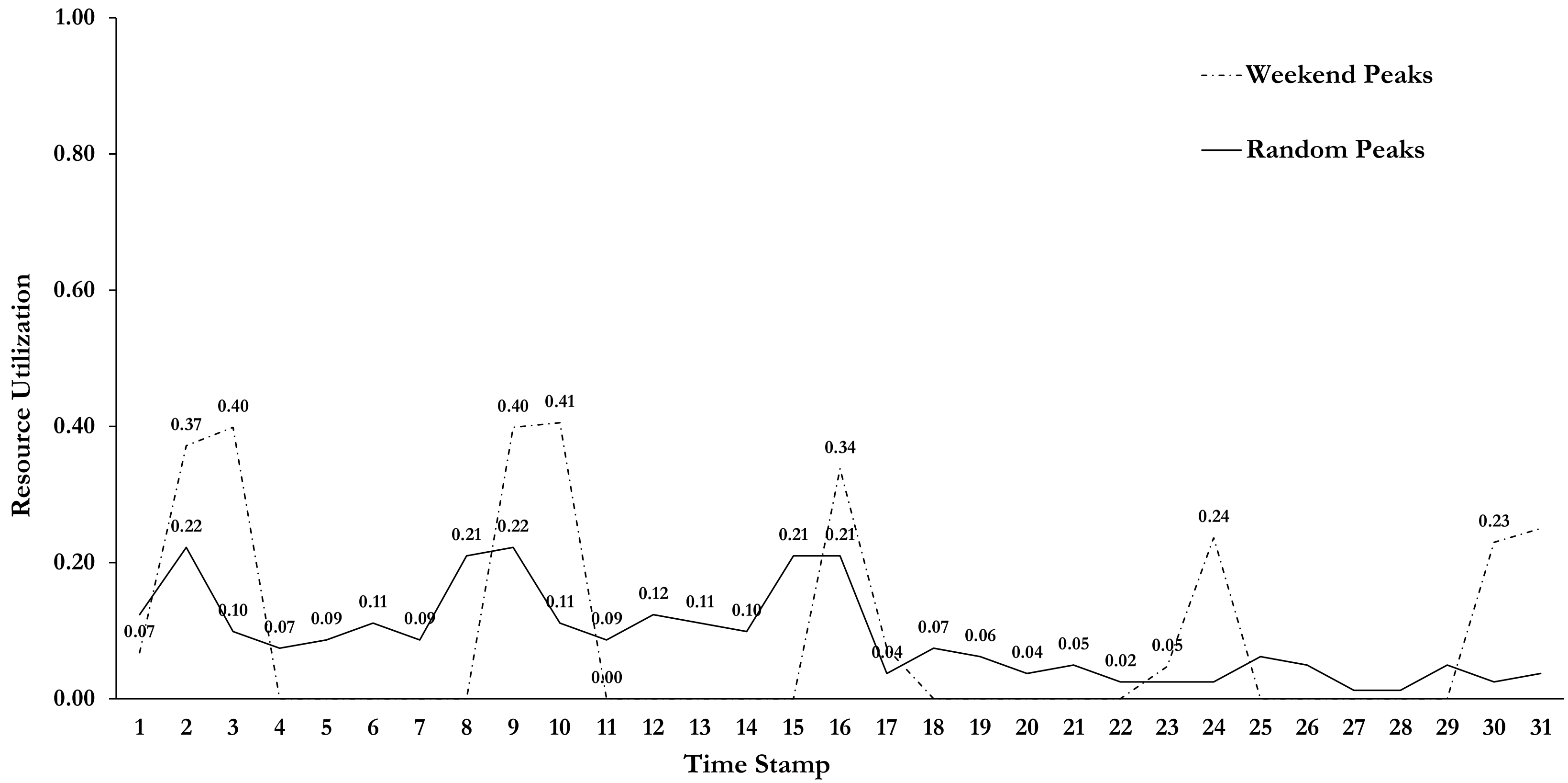

Figure 4.2: Resource Variation with time using SIRS model on PA network -Fixed Parameters are number of nodes as 100 , number of seeds 10 , number of behavior 1 , cost of behavior 0.5 , utility 0.5 , infected duration 4 , recovered duration 3 , seed selection algo = hill climbing , seed distribution uniform , Threshold average over 1000 runs . Varying Parameters : Epidemic Duration Model is SVRS, Network Topology as PA. We are comparing between "Weekend Peaks" and "Random Peaks".

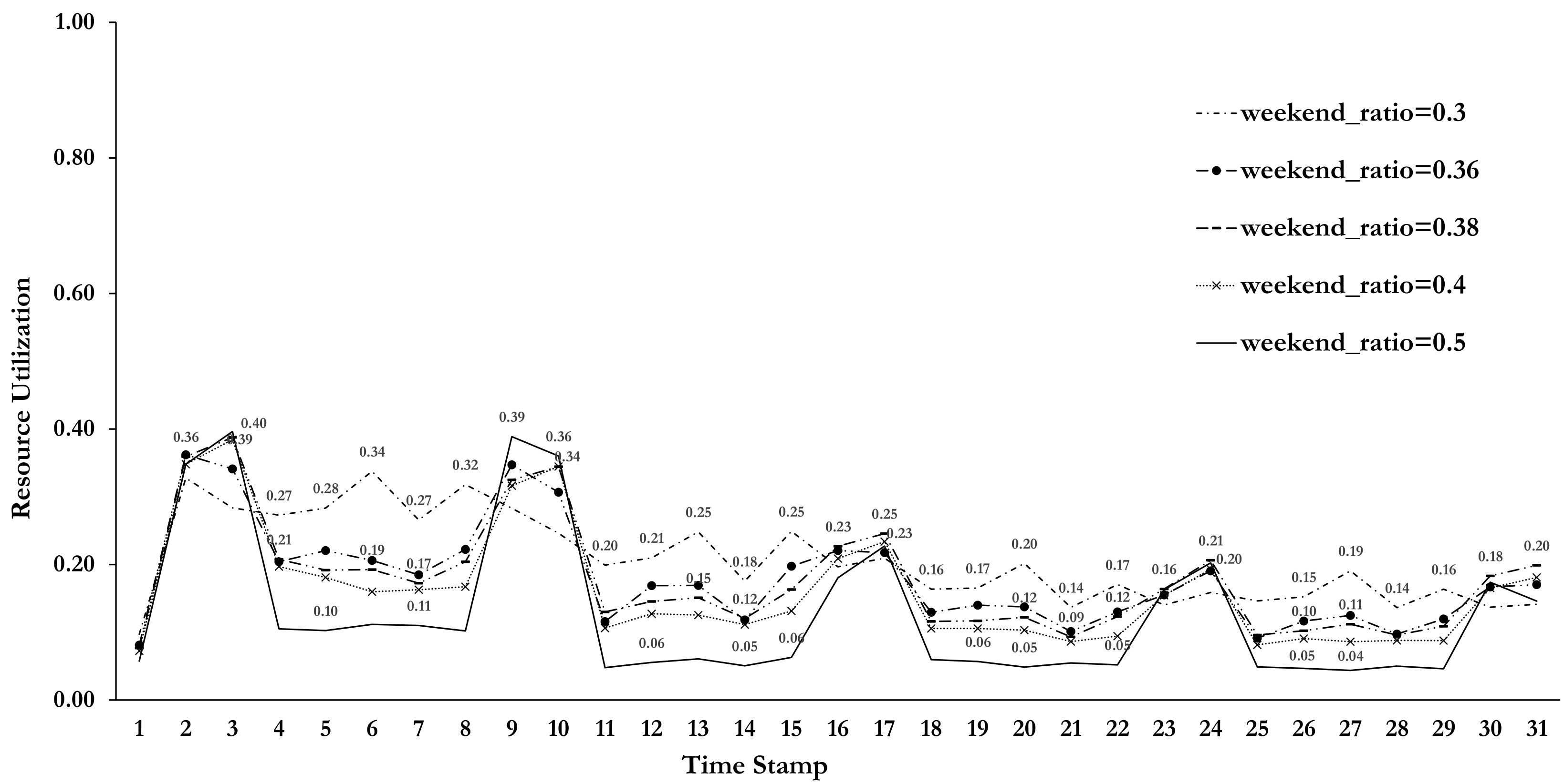

Figure 4.3: Resource Variation with time using SIRS model on PA network - Fixed Parameters are number of nodes as 100 , number of seeds 10 , number of behavior 1 , cost of behavior 0.5 , utility 0.5 , infected duration 4 , recovered duration 8 , seed selection algo = hill climbing , seed distribution = uniform , Threshold average over 1000 runs . Varying Parameters : Epidemic Duration Model is SVRS, Network Topology as PA. We are comparing between "Weekend Peaks" and "Random Peaks".]

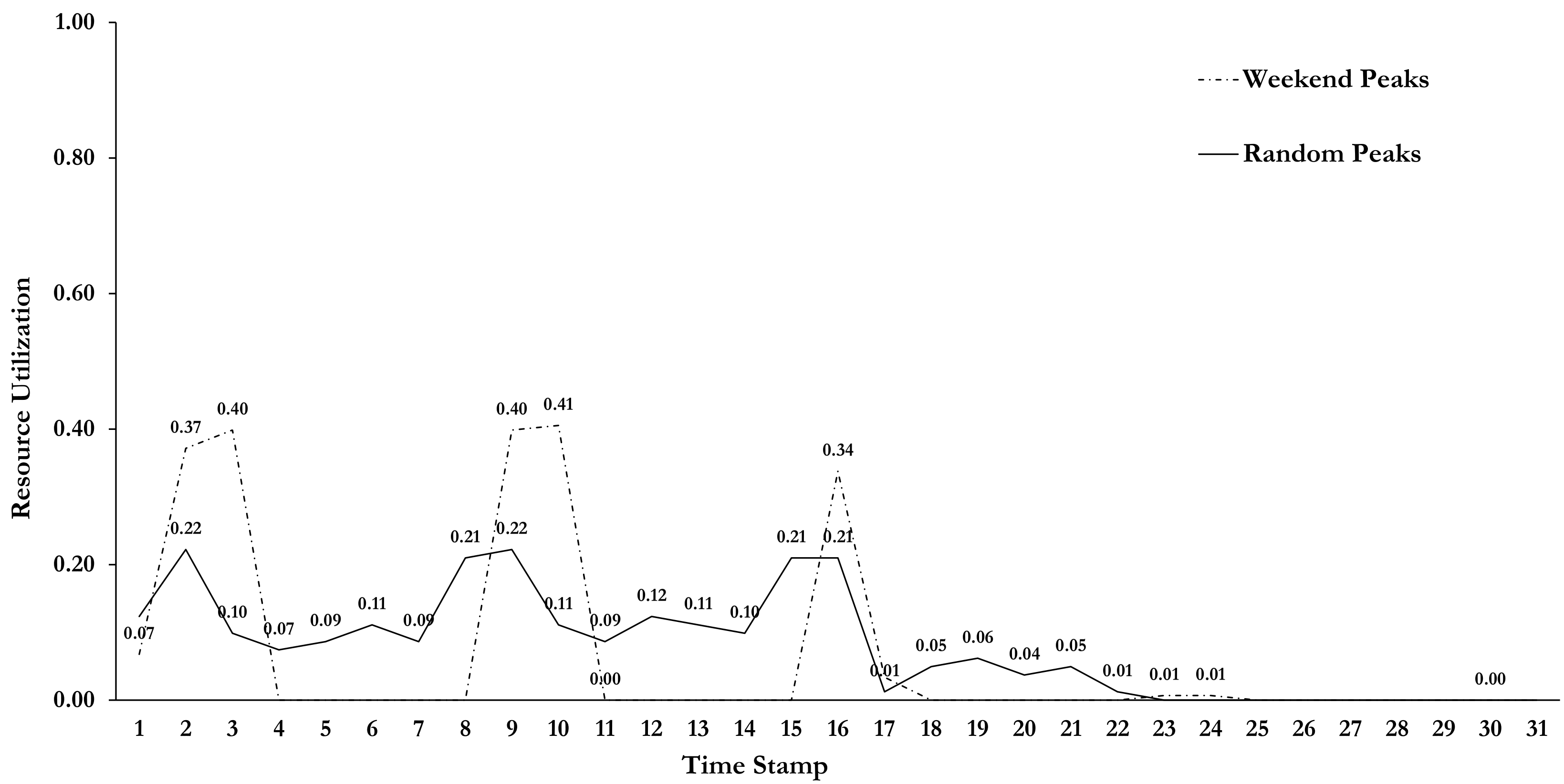

Figure 4.4: Resource Variation with time using SIRS model on PA network - Fixed Parameters are number of nodes as 100 , number of seeds 10 , number of behavior 1 , cost of behavior 0.5 , utility 0.5 , infected duration 4 , recovered duration 8 , seed selection algo = hill climbing , seed distribution = uniform , Threshold average over 1000 runs . Varying Parameters : Epidemic Duration Model is SVRS, Network Topology as PA. We are comparing between “Weekend Peaks“ and “Random Peaks“.]

observe for "Random Peaks" model utilization never reaches zero. The reason is, unlike "Weekend Peaks" there are nodes having 0.5 or more resources even during weekdays and hence individuals continues to perform the behavior.

4.2 *What happens when adoption threshold of each individual vary over time?*

In this section, we model variation of the individuals' intent over time. In case of classical *Linear Threshold* model the threshold remains constant through out the behavior adoption. In our model adoption threshold varies with time. According to *Linear Threshold* model, an individual will only adopt a behavior when the local social signal is high. For example, you want to learn swimming but the swimming club is off the route from home to office. After few months, your office got sifted near the swimming club location, then it becomes more convenient for you to enroll.

In the"varying adoption threshold" model , we represent the change in adoption threshold over time. The process takes place over discrete epochs [1] . We assume that each node is aware of the behaviors adopted in the network. The individual v first identifies all candidate behaviors. A behavior i is a candidate to be adopted if two conditions hold true. First, the individual v must have the resources to adopt the behavior, i.e $r_v \geq c^i$. Second, the social signal strength for behavior i must exceed the threshold for that behavior at node v, i.e $l_v^i \geq \theta_v^i$. Note that, all nodes are assigned a threshold in the very beginning $\theta_v^i(zero)$. This is similar to an individual initially having an adoption threshold who can get influenced over time. The second condition is similar to *Linear Threshold* (LT) model. Here, we are representing a variant of LT model, where adoption threshold for a behavior of a node θ_v^i changes over time. The change in adoption threshold depends on the knowledge of global adoption of behavior i in the network. We model the threshold variation

[1]Notice that while actions in a network are asynchronous, we can choose an appropriate time granularity for analysis to assume synchronized decision making.

such that an uniformly random fraction of the population will have a positive effect to global adoption of the behavior whereas the rest of the population will have a negative effect. For example, if an individual learns about the increasing global adoption of organic food, she can be affected positively and her adoption threshold may get reduced. On the other hand, an individual can have a negative effect, which leads to increase in her adoption threshold.

4.2.1 Mathematical Model for Varying Threshold Over Time

$$\theta_v^i(t) = \theta_0^i - x^i(t) * \alpha \tag{4.1}$$

$$\theta_v^i(t) = \theta_0^i + x^i(t) * \alpha \tag{4.2}$$

where $\theta_v^i(t)$ =final threshold for behavior i , node v at time t ,

θ_0^i =initial threshold at t_0 ,

α =maximum threshold drop or gain,

$x_i = n_i/N$.

Hence, when $x^i = 0$, then $\theta_v^i(t) = \theta_0^i$ and when $x^i = 1$, then $\theta_v^i(t) = \theta_0^i + \alpha$, which is the stable threshold θ_s^i.

As mentioned in the previous section, we are calculating the alpha and the ratio of individuals who are more susceptible by performing exhaustive experiments which are mentioned in Appendix A on page 84.

4.2.2 Simulation Setup for adoption threshold variation over time

We perform all our experiments under the following conditions. There are two parameters, we fixed few parameters like Number of nodes as 100 , number of seeds as 10 , number of behavior as 1 , cost of behavior as 0.5 , behavior utility 0.5 , infected duration as 2 , recovered duration 1 , seed selection algo as $hillclimbing$, seed distribution as $uniform$. We perform all experiments by averaging 1000 threshold runs. The varying parameters are "Threshold variation " and "No Threshold Variation". Also we assign 70% of the population to have positive effect on threshold variation and 30% population will have negative effect. We assign threshold drop or gain to 0.2. Both the threshold drop/gain and the percentage of population having positive and negative effects are assigned after extensive experimental study and careful observation. All results are given in the Appendix.

4.2.3 Simulation Results and Discussions for Single Behavior

In this section, we present the simulation results from single behavior followed by discussion on the results. Note , all fixed parameters and controlled parameter used during each experiments are mentioned in details with every result.

For our first experiment, we randomly select half of the nodes to have positive effect to threshold variation for behavior i and the rest will have negative effects.This experiment is comparing resource utilization of "Threshold Variation on" i.e the threshold variation over time and "Threshold Variation off" i.e having a fixed threshold for a single run. Figure 4.5, gives us an interesting insight and we observe with "Threshold Variation on", we can have 7 - 60 % raise in resource utilization. On careful observation, we see the resource utilization increase is due the combination of both positive and negative effect. In the first epoch, we are selecting high degree nodes as seeds, and with progression of time, behavior i is propagating to peripheral nodes. In case of "Threshold Variation off", peripheral nodes are unable to influence the seed node leading to decrease in behavior adoption whereas interestingly for "Threshold Variation on" variant, there are few seed nodes having a negative effect and adoption threshold is getting decreased. Due to a de-

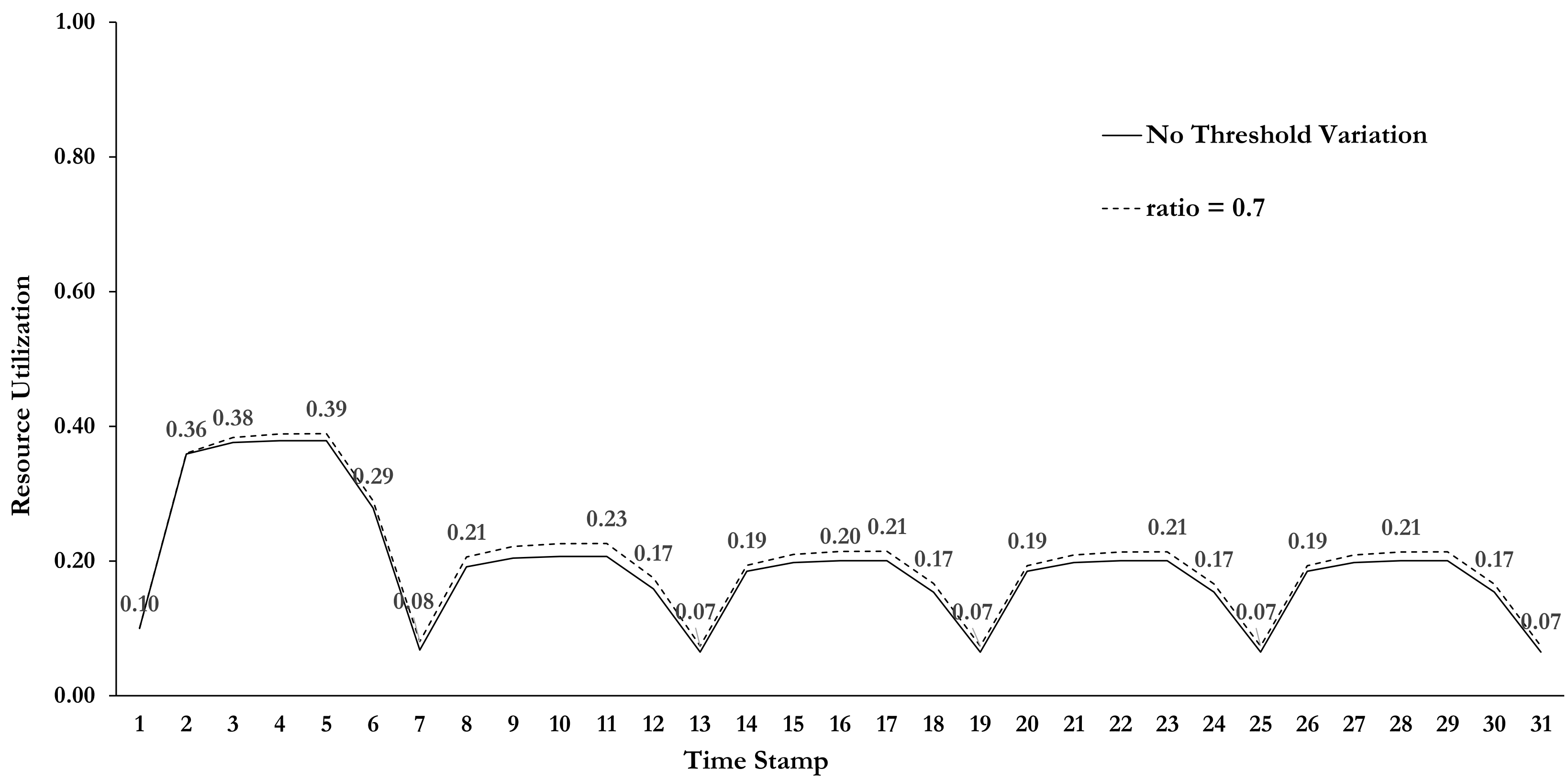

Figure 4.5: $SVRS$- PA Threshold Ratio Variation ,$\alpha = 0.2$-There are two parameters, we fixed few parameters like Number of nodes as 100 , number of seeds as 10 , number of behavior as 1 , cost of behavior as 0.5 , behavior utility 0.5 , infected duration as 2 , recovered duration 1 , seed selection algo as $hillclimbing$, seed distribution as $uniform$. We perform all experiments by averaging 1000 threshold runs. The varying parameters are "Threshold variation " and "No Threshold Variation". Also we assign 70% of the population to have positive effect on threshold variation and 30% population will have negative effect. We assign threshold drop or gain to 0.2. Both the threshold drop/gain and the percentage of population having positive and negative effects are assigned after extensive experimental study and careful observation. All results are given in the Appendix

crease in threshold , peripheral nodes are able to influence the seed nodes and bring an increase in resource utilization.

4.3 *Can change in the behavior adoption cost affects behavior distribution in the long term?*

This Section starts with the mathematical model and graphical representation on how adoption cost varies with time. Followed by initial experimental results and discussions. Our main motivation is to examine *What happens when adoption cost varies over time?*. In order to model the cost variation, we take an intuitive approach. We are considering every behavior i is associated with a cost c^i at time $zero$. For example, an individual starts with biking, initially there is a learning curve and time to learn how to bike is the cost in our model. As time progresses, if an individual carries on with biking her proficiency increases and eventually cost decreases. We model this decrease as an exponential decrease and it happens when the node is in the infected state.

4.3.1 *Mathematical Model for Varying Adoption Cost Over Time*

$$c^i(t) = c^i + \alpha * e^{-ct} \tag{4.3}$$

where c^i = adoption cost of of behavior i , exponentially decreasing ,
c^i = behavior dependent constant ,
α =max cost drop ,
ct = cumulative infected time.

4.3.2 *Simulation setup for varying adoption cost over time*

We perform all our experiments under the following conditions. There are two parameters, we fixed few parameters like Number of nodes as 100 , number of seeds as 10 , number of behavior as 1 , cost of behavior as 0.5 , behavior utility 0.5 , infected duration as 2 , recovered duration 1 , seed selection algo as $hillclimbing$, seed distribution as $uniform$. We perform all experiments by averaging 1000 threshold runs. The varying parameters are "Cost variation " and "No Cost Variation". We assign constant as 10 and the cost variation drop as 0.2. Both the adoption cost drop and the constant are assigned after extensive experimental study and careful observation. All results are given in the Appendix.

4.3.3 *Simulation Results and Discussions for Single Behavior*

From figure 4.6, we observe the following insights.

- *SVS Duration Model* :In case of Regular complete graph, we saw the highest increase in adoption 25.16% with respect to 11% in case of PA and 10.76% for small world. It shows that reduction of adoption cost can give more boosts to highly connected individuals.
- *SIRS Duration Model*: An interesting conclusion is when the recovery duration comes into play, in both PA and complete regular graph we observed a 12% increase in adoption.
- *SVRS Duration Model*: Similar to SIRS model, there is an overall 12% increase in adoption in SVRS epidemic model.

4.4 *How does the resource utilization change with varying global influence over time?*

In this thesis, in all other models, we have taken local influence into account. However, global knowledge plays a big role in behavior adoption that we learn from history of newspapers. Hence , we are interested to examine the effects of global influence variation over time on behavior adoption.

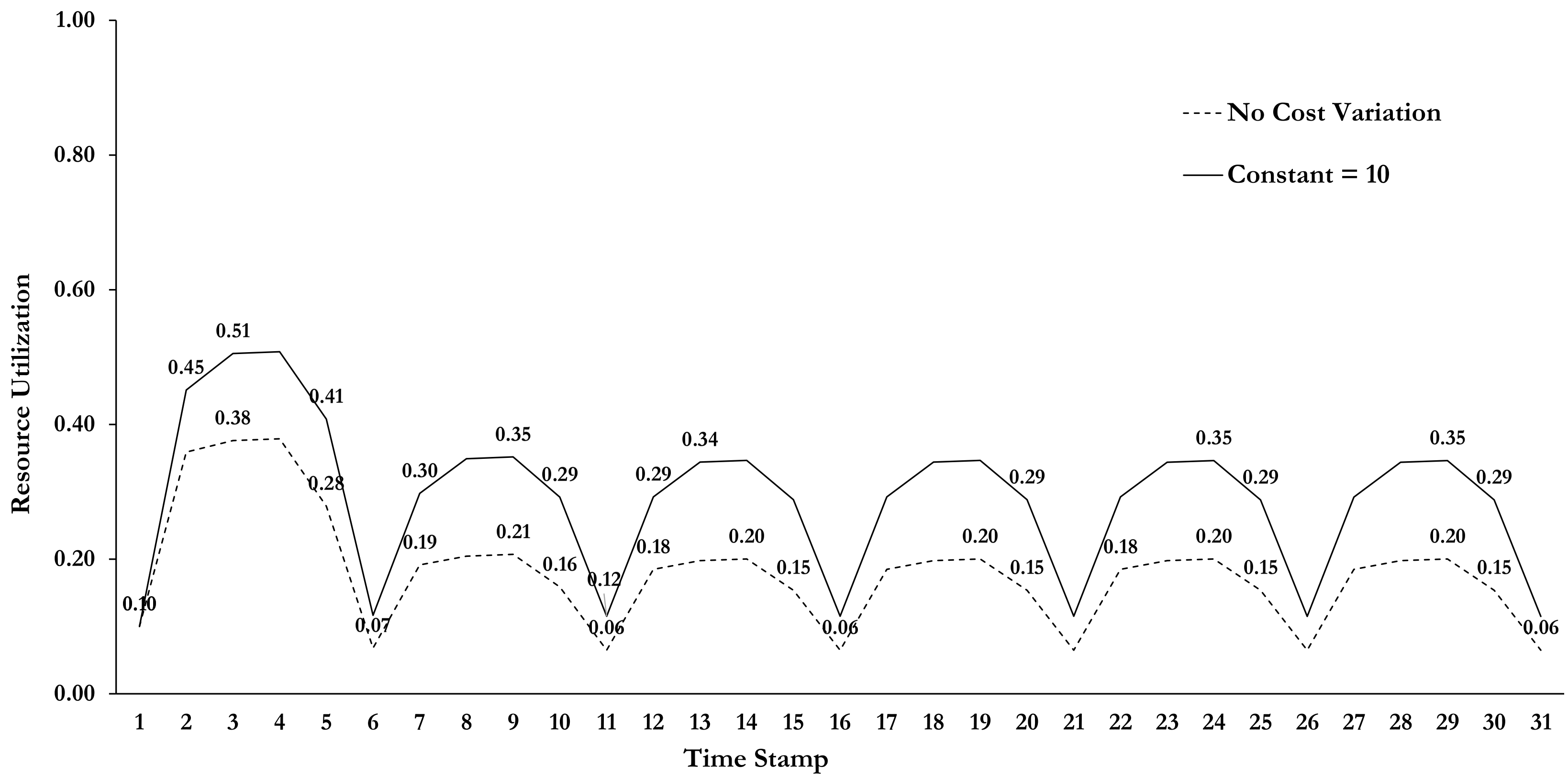

Figure 4.6: $SVRS$- PA Cost Variation , $\alpha = 0.2$ -We perform all our experiments under the following conditions. There are two parameters, we fixed few parameters like Number of nodes as 100 , number of seeds as 10 , number of behavior as 1 , cost of behavior as 0.5 , behavior utility 0.5 , infected duration as 2 , recovered duration 1 , seed selection algo as $hillclimbing$, seed distribution as $uniform$. We perform all experiments by averaging 1000 threshold runs. The varying parameters are "Cost variation " and "No Cost Variation". We assign constant as 10 and the cost variation drop as 0.2. Both the adoption cost drop and the constant are assigned after extensive experimental study and careful observation. All results are given in the Appendix

4.4.1 Mathematical Model of Varying Global Influence

$$\omega^i = \min(1, l_v^i + x_i) \tag{4.4}$$

$$\omega^i = \max(0, l_v^i - x_i) \tag{4.5}$$

where ω^i = total influence ,

l_v^i =local influence ,

$x_i = n_i/N$, global influence .

4.4.2 Simulation setup for global influence variation over time

We perform all our experiments under the following conditions. There are two parameters, we fixed few parameters like Number of nodes as 100 , number of seeds as 10 , number of behavior as 1 , cost of behavior as 0.5 , behavior utility 0.5 , infected duration as 2 , recovered duration 1 , seed selection algo as $hillclimbing$, seed distribution as $uniform$. We perform all experiments by averaging 1000 threshold runs. The varying parameters are "Global Influence Variation " and "No Global Influence Variation". We assign ratio of population as 70 : 30, i.e 70 % of the population will have positive effect to the variation of global influence over time whereas , 30 % of the population will have negative effect of the global influence. This ratio of population is assigned after extensive experimental study and careful observations. All results are given in the Appendix.

4.4.3 Simulation Results and Discussions for Single Behavior

In this section, we present the simulation results for single behavior adoption followed by discussion on the results. Note that all fixed parameters and controlled parameters used during each experiments are mentioned in detail with each result.

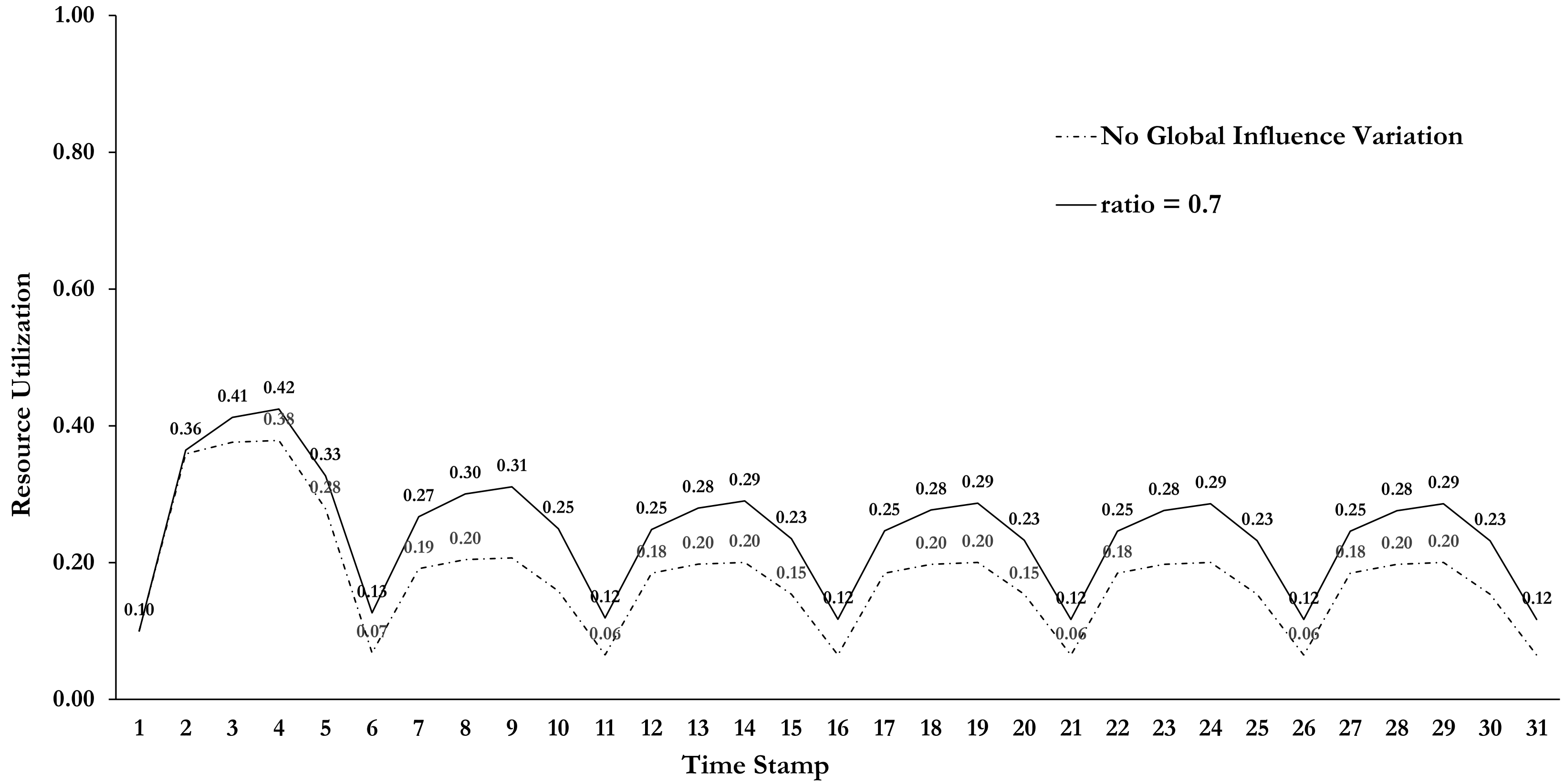

Figure 4.7: We perform all our experiments under the following conditions. There are two parameters, we fixed few parameters like Number of nodes as 100 , number of seeds as 10 , number of behavior as 1 , cost of behavior as 0.5 , behavior utility 0.5 , infected duration as 2 , recovered duration 1 , seed selection algo as $hillclimbing$, seed distribution as $uniform$. We perform all experiments by averaging 1000 threshold runs. The varying parameters are "Global Influence Variation " and "No Global Influence Variation". We assign ratio of population as 70 : 30, i.e 70 % of the population will have positive effect to the variation of global influence over time whereas , 30 % of the population will have negative effect of the global influence. This ratio of population is assigned after extensive experimental study and careful observations. All results are given in the Appendix

Figure 4.7 , shows the experimental results for varying global influence over time. In all the experimental results, we observe a drop in resource utilization. Reason for the drop in utilization is the effect of knowledge about adoption in global population. Note that the global influence is behavior dependent. The additional knowledge of the global behavior adoption is diluting the weight of social signal exerted by neighbors. For example , 50 % of neighbors of node v adopted behavior j, then local influence is 0.5 whereas only 20 % of the global population has adopted the behavior j. If we take the threshold θ_v^j of behavior j of node v is 0.5, with knowledge of only local adoption v will end up adopting behavior j. After addition of the global influence knowledge, the combined weight is less than the threshold of v. Therefore as a result of this experiment, we can see global influence is acting as a diluting factor and reducing resource utilization. We find there is a 21 % - 88% increase in behavior lifespan in the network. This is caused by participation of peripheral nodes having no or negligible local influence. We found that there are nodes having low threshold and enough resource for behavior adoption but do not have any knowledge of the presence of the behavior j in the network. By introducing the idea of global awareness, these nodes are getting activated, and behavior j remains in the network for a longer period of time.

4.5 *What happens when combinations of parameters varies over time?*

In this section , we examine the scenarios where multiple assumption parameters are varying over time. This not only is more realistic , but also provides us with insights which are comparable.

In this section we perform several parameter combination variation. We provide the experimental sequence and all parameter combination results in Appendix.In figure A.11 , we present the comparative study of these combinations. We use short forms to represent the combinations due to lack of graphical space. There "BM" is the base model where no variation of parameters are taking place. Starting from right hand side, we look into "T" , where adoption threshold is varying over time. Then "C" is the adoption cost variation . We represent variation of resource availability by "R". "GI" is the short form of global influence variation over time. Then combinations like "T + C " is the case where both threshold and cost are varying with time. "C + R " is another scenario where both adoption cost and the resource availability varies. We present variation of cost , resource availability and threshold by "C+R+T". In the above mentioned experiments we only consider local social signal on an individual , now we present the additive effect of global influence variation in the case of " C+R+T+GI". Note that all the constant are already fixed in previous sections. Here we are concentrating on combinational effects.

4.5.1 Simulation Results and Discussions

In figure 4.8 , we represent all the combination variations using a boxplot data visualization [40] , [62]. In descriptive statistics, a box plot or boxplot is a convenient way of graphically depicting groups of numerical data through their quartiles. Outliers are plotted as individual points. Box plots display differences between populations without making any assumptions of the underlying statistical distribution: they are non-parametric. The spacings between the different parts of the box help indicate the degree of dispersion (spread) and skewness in the data, and identify outliers. In addition to the points themselves, they allow one to visually estimate various L-estimators, notably the interquartile range, midhinge, range, mid-range, and trimean. We are representing our longitudinal distributions and in Figure 4.8 , there are few interesting intuitive observations.

- We observe adoption cost variation gives us a statistically significant increase in resource utilization.
- Similarly combination of threshold and cost variation of a behavior adoption gives

us a statistically significant increase in resource utilization.

- Interestingly , when resource availability variation comes into play , we observe an increase in range of resources among individuals but the maximum resource utilization is not getting benefited.

- While comparing to no variation model , which we are representing as "BM" , additive effect of global influence variation gives us an increase in resource utilization as well as extends the range of resource variation among individuals.

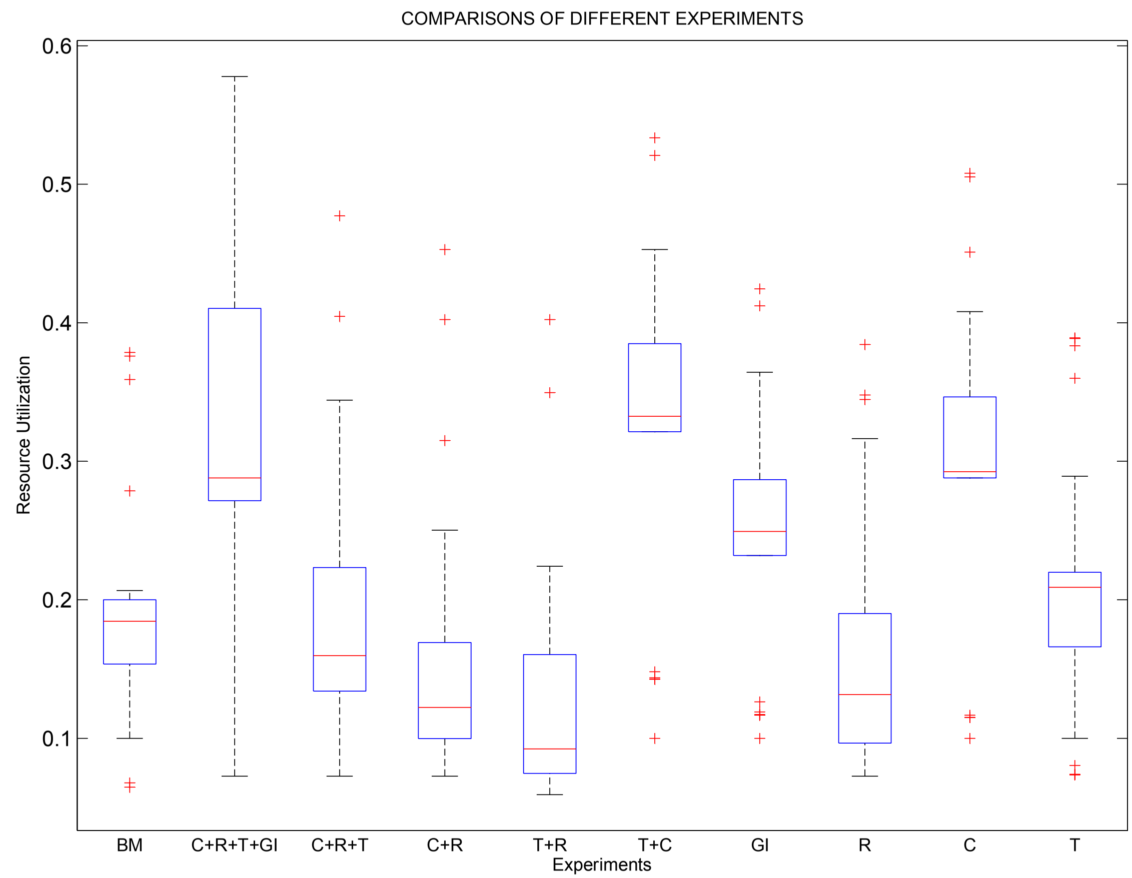

Figure 4.8: This is comparative study of parameter combinations. We use short forms to represent the combinations. There "BM" is the base model where no variation of parameters are taking place. Starting from right hand side, we look into "T" , where adoption threshold is varying over time. Then "C" is the adoption cost variation . We represent variation of resource availability by "R". "GI" is the short form of global influence variation over time. Then combinations like "T + C " is the case where both threshold and cost are varying with time. "C + R " is another scenario where both adoption cost and the resource availability varies. We present variation of cost , resource availability and threshold by "C+R+T". In the above mentioned experiments we only consider local social signal on an individual , now we present the additive effect of global influence variation in the case of " C+R+T+GI". Note that all the constant are already fixed in previous sections. Here we are concentrating on combinational effects.

Chapter 5

CONCLUSION AND FUTURE WORK

5.1 Introduction

In conclusion, we shall first summarize the work presented in this thesis, along the lines of the five questions that have been tackled here. They are behavior duration models, varying adoption cost, varying adoption threshold, varying resource availability and varying global influence. Then, in section 5.3, we shall present some possible improvements to the models used in this work and in section 5.4, we shall present a few potential areas of future research.

5.2 Research Summary

We now present a synopsis of the work done in this thesis by first summarizing our approach to epidemic models.

5.2.1 Behavior Duration Models

The thesis developed a novel framework for including epidemic models and costly behavior diffusion in a social network. We focused on social network having limited resource and focused our investigation on four epidemic models. The first two models are *SIR (Susceptible - Infected - Recovered)* and *SIRS (Susceptible - Infected - Recovered - Susceptible)* and the last two are *SVS (Susceptible - Volatile - Susceptible)* and *SVRS (Susceptible - Volatile - Recovered - Susceptible).* SVRS is an extension to SIRS . In SVRS , we introduced a new state called "Volatile" . The reason behind the new state is to make our model generalized for both single behavior and multiple behavior diffusion. In "multiple behavior" diffusion, we incorporated the idea of choosing the behavior/s which, will give an individual the optimized result. Hence, as she is open to a new behavior, she is "Volatile" in nature. In chapter 3 on page 27 , we presented our approach and showed how to calculate three parameters of measuring effectiveness. Three parameters we are using are ,

- Resource Utilization - cumulative amount of resource used in a system by all the nodes,
- Total Adoption - a summation of all the nodes' adopted behavior or behaviors. This is the total sum, which includes all available behavior in the system.
- Unique Adoption - This is summarization of nodes' adoption of each behavior in the system. This parameter is important as there can be few nodes in the system adopting more than one behavior and only "Total Adoption" will not reveal the real story.

In chapter 4 on page 48 , we presented our experimental results from the epidemic models and three key observations are summarized below. We run all the duration models on three synthetic network topologies (PA, SW , CG). Hence , while summarizing , we are going to mention the exact duration model and network topology .

- *SVS Duration Model* : In case of SW and Regular complete graph, we saw that with increase of graph regularity, utilization drops as more adopted neighbors are required to perform the adoption.
- *SIRS Duration Model*: Here an interesting observation is that there is a drop of the behavior in PA case. The reason being during seed selection, we are choosing those nodes, which can impart highest influence and eventually have the highest degrees. Now once those nodes are getting recovered the adoption is getting moved to peripheral nodes, but these peripheral nodes are not powerful enough to impose back

the adoption to the seed nodes. Whereas, in case of Smallworld or Complete regular graph, there remains a group of nodes, which can induce adoption to each other and the behavior propagates. From the result we can also see that the propagation is stabler when network is more regular.

- *SVRS Duration Model*: The difference between SIRS model and this model is that there is no recovery time and even when the node is infected it is in competition with other behaviors. As there is no recovery time, even in case of PA, we can see that the seed node is influencing the peripheral node and the peripheral node are in turn influencing the seed node and the behavior propagates.

5.2.2 *Varying Adoption Cost*

In chapter 4 on page 48 , we described how every behavior is associated with a cost and an individual is able to adopt that behavior only when the social signals from neighbors are high and individual has resource. A summarized observation in listed below.

- *SVS Duration Model* :In case of Regular complete graph, we saw the highest increase in adoption 25.16% with respect to 11% in case of PA and 10.76% for small world. It shows that reduction of adoption cost can give more boosts to highly connected individuals.

- *SIRS Duration Model*: Interesting conclusion is when the recovery duration comes into play, in both PA and CG we observed a 12% increase in adoption.

- *SVRS Duration Model*: Similar to SIRS model, there was an overall 12% increase in adoption.

5.2.3 Varying Adoption Threshold

In this thesis, we have used Linear Threshold model to design our behavior diffusion. According to Linear Threshold model every individual has a threshold of adopting a behavior. In chapter 4 on page 48, one question we targeted is “what happens when the adoption threshold varies with time?" Note that, we had two assumptions while performing these experiments. First, each node's can be more susceptible, threshold can either decrease as time progresses and we are taking randomly allotted 70% of the nodes. Second, threshold of each node can increase with time and we are taking 30% of the nodes. Below we cumulate all the compelling results and conclusions attained from of our experiments.

- *SVS Duration Model* : We observed a 4
- *SIRS Duration Model*: Interestingly, behavior sustained for a longer period of time in the system with change in threshold.
- *SVRS Duration Model*: This model was inferred to be the most effective model for the PA network. First, we observed a 5% increase in resource utilization. Second, after the first time period, we noted an almost 60% increase in resource utilization. The reason, being the combination of nodes having both increases and decreases of thresholds. Hence, we can conclude that diversity in a community will help for better behavior adoption and sustaining it for a longer period of time.

5.2.4 Varying Resource Availability

In this thesis, one of the questions we investigated is “how varying resource availability over time can effect behavior/s adoption ? ” In chapter 4 on page 48, we presented the complete model. Briefly, the idea is, an individual may not have equal resource available every day of a week. For example, an individual can have 4 hours for social service on the weekends but no time during the weekdays. We introduced two variants of resource availability in this thesis.

5.2.4.1 Weekend Peaks

The main idea behind weekend peaks is that most of the population is having more resource during the same time. For example, most of the individuals are having more resources during their weekends than weekdays.

5.2.4.2 Random Peaks

In “Random Peaks" , we built the model in such a way that individuals can have higher resource on any two consecutive days of the week. For example, in a community, if one individual is having 4 hours during weekends, another individual may have more time on Wednesdays and Thursdays.

Taking all variants of resource availability distinctions into account, we conclude all the experimental results below.

- *SVRS Duration Model*: Comparing the above-mentioned variants, we observed initial adoption is higher in case of “Weekend Peaks". As a conclusion, we can say there is more adoption if most of the individuals have higher resources at the same period of time. On the other hand, average resource utilization is higher in case of “Random Peaks". We, thus, concluded that if we target for higher average resource utilization over a period of time then “Random Peaks" is a better variant.

5.2.5 Varying Global Influence

Till now, in all of the above sections we have been using only local social signals for behavior diffusion. Therefore, intuitively in this thesis the next question we investigated is “what happens when the global influence becomes part of the adoption decision and varies over time?" For example, an individual now will not only have knowledge about how many of her neighbors adopted recycling but also about the number of people in her county who adopted recycling.

We found an additive effect of “varying global influence" with all the other models used in this thesis.

5.3 Improvements to the frameworks

The task of building a framework that answers questions to “collective human behavior" is a challenging one and this thesis is an effort in presenting a complete framework to the understanding of “how behavior diffusion changes over time both in case of single and multiple behaviors".

Below we suggest a few improvements in inference to this thesis.

5.3.1 Network Topologies

We performed all our experiments taking three synthetic topologies (PA, Small-world and Completed Regular Graph). However, it would be interesting to expand the model for other real-life networks. For example, it will be intriguing if we can try all the experiments on Obesity Dataset. We can then observe how individuals adopt and drop healthy behaviors during a course of time.

5.3.2 Representation of Time

In this thesis, we have incorporated several important ideas from varying adoption cost over time to the importance of varying resource availability over time. We have constraint time to discrete epochs. However, representation of time can be more intuitive by thinking it as a 24 hours span.

5.4 Future directions

We now highlight a few possible future research directions.

5.4.1 Noisy Networks

In a network, we receive social signals from our friends, but there is noise because we miss messages and or we check them late. In modeling the behavior adoption problem, we have ignored the role of constraints in how they affect the production and consumption of messages from peers. Explicit consideration of the cost of social signaling would not only make the model more realistic but also provide better bounds on the maximal resource utilization of the networks resources.

5.4.2 Change in population

In this thesis, we considered our network to be a closed network. For example, no node dies or there is no increase in population. However, it can be a possible extension to this model. It will not only make the network more dynamic but also will provide better understanding of behavior extinction from a network in the long term.

5.4.3 Change in behavior

In this thesis, one assumption is the number of behavior remains constant throughout the propagation of behaviors. Whereas, extending the model to “varying behavior availability" can provide us with interesting insights.

REFERENCES

[1] Second life http://secondlife.com/.

[2] Icek Ajzen. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211, December 1991.

[3] Roy M. Anderson and Robert M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, USA, 9 1992.

[4] Sinan Aral, Lev Muchnik, and Arun Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51):21544–21549, 2009.

[5] L. Backstrom and J. Kleinberg. Romantic Partnerships and the Dispersion of Social Ties: A Network Analysis of Relationship Status on Facebook. *ArXiv e-prints*, October 2013.

[6] Eytan Bakshy, Brian Karrer, and Lada A. Adamic. Social influence and the diffusion of user-created content. In John Chuang, Lance Fortnow, and Pearl Pu, editors, *ACM Conference on Electronic Commerce*, pages 325–334. ACM, 2009.

[7] V. Bala and S. Goyal. Learning from neighbours. *Review of Economic Studies*, 65:595–621, 1998.

[8] Abhijit V. Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817, 1992.

[9] S. Banerjee, H. Al-Qaheri, and A. ella Hassanien. Mining social networks for viral marketing using fuzzy logic. In *Mathematical/Analytical Modelling and Computer Simulation (AMS), 2010 Fourth Asia International Conference on*, pages 24–28, 2010.

[10] LuÃČÂŋs M.A. Bettencourt, Ariel CintrÃČÂşn-Arias, David I. Kaiser, and Carlos Castillo-ChÃČÂąvez. The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. *Physica A: Statistical Mechanics and its Applications*, 364(0):513 – 536, 2006.

[11] Shishir Bharathi, David Kempe, and Mahyar Salek. Competitive influence maximization in social networks. In *In WINE*, pages 306–311, 2007.

[12] S. Bikhchandani, D. Hirshleifer, and I. Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026, 1992.

[13] L. Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5:387–424, 1993.

[14] Duncan S Callaway, MEJ Newman, Steven H Strogatz, and Duncan J Watts. Network robustness and fragility: Percolation on random graphs. 85(25):5468–5471, December 2000.

[15] Tim Carnes, Rashekhar Nagarajan, Stefan M. Wild, and Anke Van Zuylen. Maximizing influence in a competitive social network: a follower's perspective. In *In ICEC ?07: Proceedings of the ninth international conference on Electronic commerce*, pages 351–360. ACM, 2007.

[16] Claudio Castellano and Romualdo Pastor-Satorras. Thresholds for epidemic spreading in networks. *CoRR*, abs/1010.1646, 2010.

[17] Deepayan Chakrabarti, Yang Wang, Chenxi Wang, Jurij Leskovec, and Christos Faloutsos. Epidemic thresholds in real networks.

[18] Wei Chen, Yajun Wang, and Siyu Yang. Efficient influence maximization in social networks. In *KDD*, pages 199–208, 2009.

[19] Wei Chen, Yifei Yuan, and Li Zhang. Scalable influence maximization in social networks under the linear threshold model.

[20] Michael Suk-Young Chwe. Structure and strategy in collective action. volume 105, pages 128–156, July 1999.

[21] James S. Coleman, Elihu Katz, and Herbert Menzel. *Medical innovation: A diffusion study*. Bobbs-Merrill, New York, 1966.

[22] Dan Cosley, Daniel P. Huttenlocher, Jon M. Kleinberg, Xiangyang Lan, and Siddharth Suri. Sequential influence models in social networks. In William W. Cohen and Samuel Gosling, editors, *ICWSM*. The AAAI Press, 2010.

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

[24] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

[25] Glenn Ellison. Learning, local interaction, and coordination. *Econometrica*, 61(5):1047–1071, 1993.

[26] Eyal Even-Dar and Asaf Shapira. A note on maximizing the spread of influence in social networks. In Xiaotie Deng and FanChung Graham, editors, *Internet and Network Economics*, volume 4858 of *Lecture Notes in Computer Science*, pages 281–286. Springer Berlin Heidelberg, 2007.

[27] Douglas Gale and Shachar Kariv. Bayesian learning in social networks. *Games and Economic Behavior*, 45(2):329–346, 2003.

[28] Ayalvadi J. Ganesh, Laurent MassouliÃľ, and Donald F. Towsley. The effect of network topology on the spread of epidemics. In *INFOCOM*, pages 1455–1466. IEEE, 2005.

[29] Golnaz Ghasemiesfeh, Roozbeh Ebrahimi, and Jie Gao. Complex contagion and the weakness of long ties in social networks: Revisited. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, EC '13, pages 507–524, New York, NY, USA, 2013. ACM.

[30] Malcolm Gladwell. *The Tipping Point. How Little Things Can Make A Big Difference*. Abacus, 2006.

[31] Jacob Goldenberg, Barak Libai, and Eitan Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 2001.

[32] Scott A. Golder, Dennis M. Wilkinson, and Bernardo A. Huberman. Rhythms of social interaction: messaging within a massive online network. *CoRR*, abs/cs/0611137, 2006.

[33] Manuel Gomez Rodriguez, Jure Leskovec, and Andreas Krause. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '10, pages 1019–1028, New York, NY, USA, 2010. ACM.

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

[35] M.S. Granovetter. The strength of weak ties. *American Journal of Sociology*, (78):1360–1380, 1973.

[36] P. Grassberger. On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences*, 63(2):157 – 172, 1983.

[37] Daniel Gruhl, R. Guha, David Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *WWW'03*, 2003.

[38] Peter Hedstrom. *Contagious collectivities: On the spatial diffusion of Swedish trade unions.*, volume 99. American Journal of Sociology, 1994.

[39] Herbert W. Hethcote. The mathematics of infectious diseases. *SIAM REVIEW*, 42(4):599–653, 2000.

[40] Peter Kampstra. Beanplot: A boxplot alternative for visual comparison of distributions. *Journal of Statistical Software, Code Snippets*, 28(1):1–9, 11 2008.

[41] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '03, pages 137–146, New York, NY, USA, 2003. ACM.

[42] David Kempe, Jon M. Kleinberg, and Éva Tardos. Influential nodes in a diffusion model for social networks. In *32nd International Colloquium on Automata, Languages and Programming (ICALP 2005)*, 2005.

[43] J.O. Kephart and S.R. White. Measuring and modeling computer virus prevalence. In *Research in Security and Privacy, 1993. Proceedings., 1993 IEEE Computer Society Symposium on*, pages 2–15, 1993.

[44] Masahiro Kimura and Kazumi Saito. Tractable models for information diffusion in social networks. In Johannes FÃijrnkranz, Tobias Scheffer, and Myra Spiliopoulou, editors, *PKDD*, volume 4213 of *Lecture Notes in Computer Science*, pages 259–271. Springer, 2006.

[45] Jure Leskovec, Lada A. Adamic, and Bernardo A. Huberman. The dynamics of viral marketing. 2007.

[46] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne M. VanBriesen, and Natalie S. Glance. Cost-effective outbreak detection in networks. In Pavel Berkhin, Rich Caruana, and Xindong Wu, editors, *KDD*, pages 420–429. ACM, 2007.

[47] M. LYNNE MARKUS. Toward a âĂIJcritical massâĂİ theory of interactive media: Universal access, interdependence and diffusion. *Communication Research*, 14(5):491–511, 1987.

[48] Doug McAdam. *Recruitment to high-risk activism: The case of Freedom Summer.*, volume 92. American Journal of Sociology, 1986.

[49] Doug McAdam. *Freedom Summer*. Oxford University Press, New York, 1988.

[50] A. G. M'Kendrick. Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130, 2 1925.

[51] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.*, 7:295, 1998.

[52] Stephen Morris. Contagion. *The Review of Economic Studies*, 67(1):57–78, 2000.

[53] M. E. J. Newman. The structure of scientific collaboration networks. *PNAS*, 98(2):404–409, 2001.

[54] M. E. J. Newman. The spread of epidemic disease on networks. *Physical Review*, E 66, 2002.

[55] Hubert J. O'Gorman. The discovery of pluralistic ignorance: An ironic lesson. *Journal of the History of the Behavioral Sciences*, 22(4):333–347, 1986.

[56] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physics Review Letters*, 86(14), april 2001.

[57] B. Aditya Prakash, Hanghang Tong, Nicholas Valler, Michalis Faloutsos, and Christos Faloutsos. Virus propagation on time-varying networks: Theory and immunization algorithms. In JosÃľ L. BalcÃązar, Francesco Bonchi, Aristides Gionis, and MichÃĺle Sebag, editors, *ECML/PKDD (3)*, volume 6323 of *Lecture Notes in Computer Science*, pages 99–114. Springer, 2010.

[58] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing, 2002.

[59] Everett M. Rogers. *Diffusion of innovations*. Free Press, New York, NY [u.a.], 5th edition, 08 2003.

[60] Everett M. Rogers and Everett Rogers. *Diffusion of Innovations, 5th Edition*. Free Press, 5th edition, August 2003.

[61] B. Ryan and N. Gross. The diffusion of hybrid seed corn in two iowa communities. *Rural Sociology*, 8(1):15–24, 1943.

[62] Kazumi Saito, Masahiro Kimura, and Hiroshi Motoda. Effective visualization of information diffusion process over complex networks. In Walter Daelemans, Bart Goethals, and Katharina Morik, editors, *ECML/PKDD (2)*, volume 5212 of *Lecture Notes in Computer Science*, pages 326–341. Springer, 2008.

[63] Cosma Rohilla Shalizi and Andrew C. Thomas. Homophily and contagion are generically confounded in observational social network studies, 2010. cite arxiv:1004.4704Comment: 27 pages, 9 figures. V2: Revised in response to referees. V3: Ditto.

[64] David B. Sicilia. America calling: A social history of the telephone to 1940. by claude s. fischer. berkeley: University of california press, 1992. pp. xv, 424. *The Journal of Economic History*, 53(02):451–452, June 1993.

[65] David A. Siegel. *Social networks and collective action.*, volume 53(1). American Journal of Political Science,, 2009.

[66] Leonard Bickman Stanley Milgram and Lawrence Berkowitz. *Note on the drawing power of crowds of different size. Journal of Personality and Social Psychology,*, volume 13(2). October 1969.

[67] David Strang and Sarah Soule. Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual Review of Sociology,*, 24:265–290, 1998.

[68] Michael Suk-Young Chwe. Communication and coordination in social networks. *The Review of Economic Studies*, 67(1), 2000.

[69] Johan Ugander, Lars Backstrom, Cameron Marlow, and Jon Kleinberg. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences*, 109(16):5962–5966, 2012.

[70] S. Wasserman and K. Faust. *Social network analysis: Methods and applications*. Cambridge Univ Pr, 1994.

[71] D.J. Watts. *Small worlds : the dynamics of networks between order and randomness*. Princeton Studies in Complexity. Princeton University Press, 2004.

[72] Duncan J. Watts. Networks, dynamics and the small world phenomenon. *AJS*, 1999.

[73] Duncan J. Watts. A simple model of global cascades on random networks. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 99, pages 5766–5771, April 30 2002. http://www.jstor.org/view/00278424/sp020038/02x3936j/0.

[74] Duncan J. Watts and Peter S. Dodds. Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research*, 34(4):441–458, December 2007.

[75] Ivo Welch. Sequential sales, learning, and cascades. *Journal of Finance*, 47:695–732, 1992.

[76] Peyton Young. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, 1998.

APPENDIX A

EXPERIMENTAL DATA

Table A.1: Design of Experimental Sequence

Exp No.	Varying Parameters	Fixed Parameters
1	θ_α	$\theta_{ratio}, c^i, R^i, GI_{ratio}$
2	θ_{ratio}	$\theta_\alpha, c^i, R^i, GI_{ratio}$
	Fix $\theta_{\alpha,ratio}$	
3	c_α	$c_C, \theta^i, R^i, GI_{ratio}$
4	c_C	$c_\alpha, \theta^i, R^i, GI_{ratio}$
	Fix $c_{\alpha,C}$	
5	R_α	$c^i, \theta^i, GI_{ratio}$
	Fix R_α	
6	GI_{ratio}	c^i, θ^i, R^i
	Fix GI_{ratio}	
7	$\theta_{\alpha,ratio}$ X $c_{\alpha,C}$	R_α, GI_{ratio}
8	$\theta_{\alpha,ratio}$ X R_α	$c_{\alpha,C}, GI_{ratio}$
9	$c_{\alpha,C}$ X R_α	θ_p, GI_{ratio}
10	$\theta_{\alpha,ratio}$ X $c_{\alpha,C}$ X R_α	GI_{ratio}
11	$\theta_{\alpha,ratio}$ X $c_{\alpha,C}$ X R_α X GI_{ratio}	N/A

Perform the above mentioned 11 experiments for “base model" , “SIR" , “SVRS" , “SIRS" on three networks “PA" , “small-world" and “complete graph".

Experiment to fix threshold variation parameters

We perform an extensive study on threshold variation to identify the threshold drop/gain and the ratio of population who will have positive or negative impact. By positive impact of threshold, we mean that the intention to perform a behavior of an individual will increase. Similarly , negative effect of threshold means that the intention to perform a behavior will decrease. In Figure A.6 , we perform alpha variation keeping population ration as 70 : 30 , i.e 70 % of the population will have positive effect and 30 % of the population will have negative effect over long time periods. We identify alpha as 0.2 . In Figure A.4 , we perform population ratio variation keeping alpha as 0.2.

Table A.2: Terminologies used in experimental sequence design

Symbols	Descriptions
θ_C	Threshold with changing constant.
θ_α	Threshold with changing α values
θ_{ratio}	Ratios are the ratio between nodes with increasing threshold and nodes with decreasing effect over time.
θ_p	Fixing all the parameter values for varying threshold θ
c_α	cost variation
R_α	Resource Variation
GI	Global Influence Variation

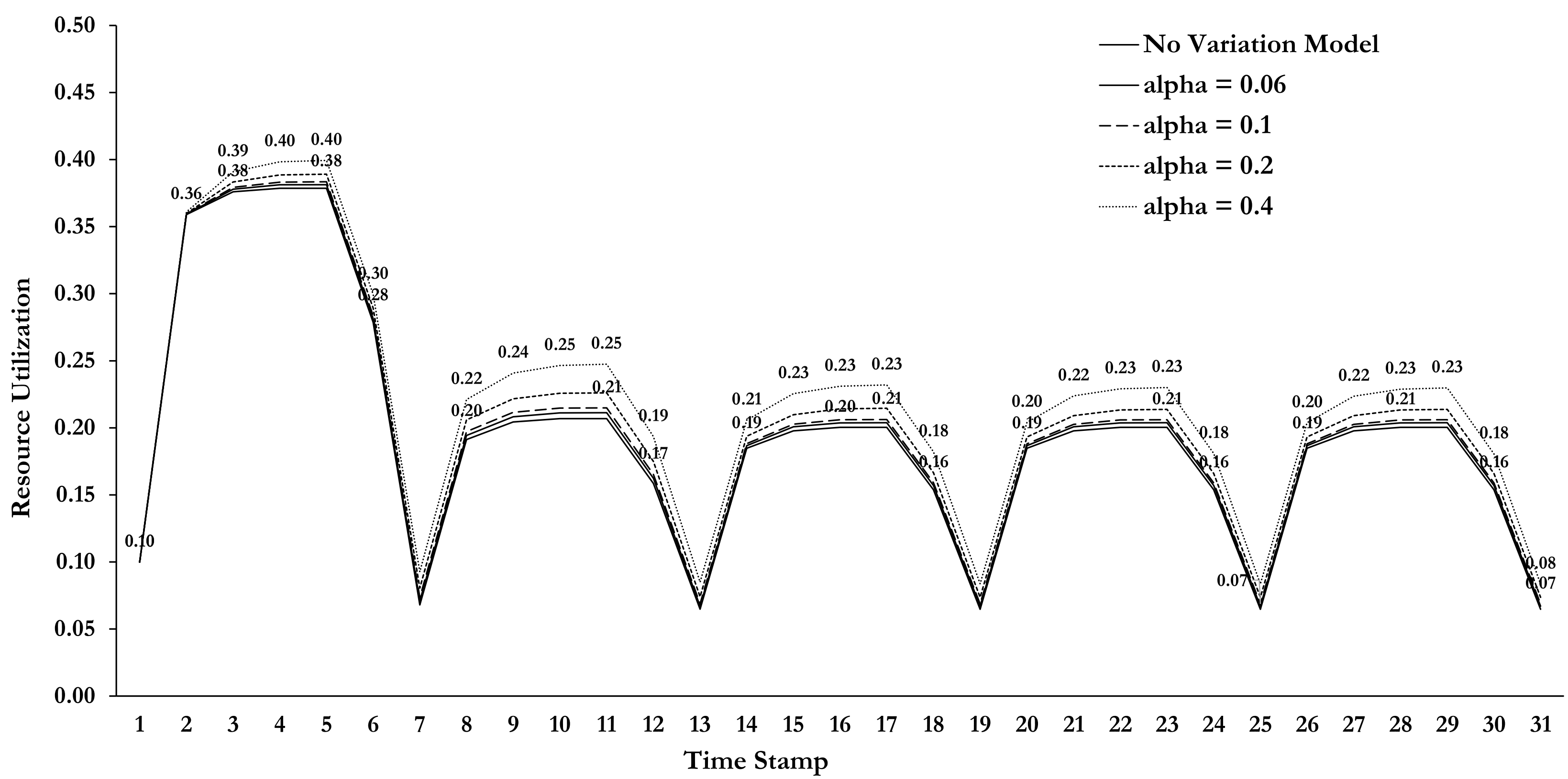

Figure A.1: $SVRS$- PA Threshold Alpha Variation ,$ratio = 0.7$ - This is an extensive study on threshold variation to identify the threshold drop/gain keeping population ration as 70 : 30 , i.e 70 % of the population will have positive effect and 30 % of the population will have negative effect over long time periods. We identify alpha as 0.2 .

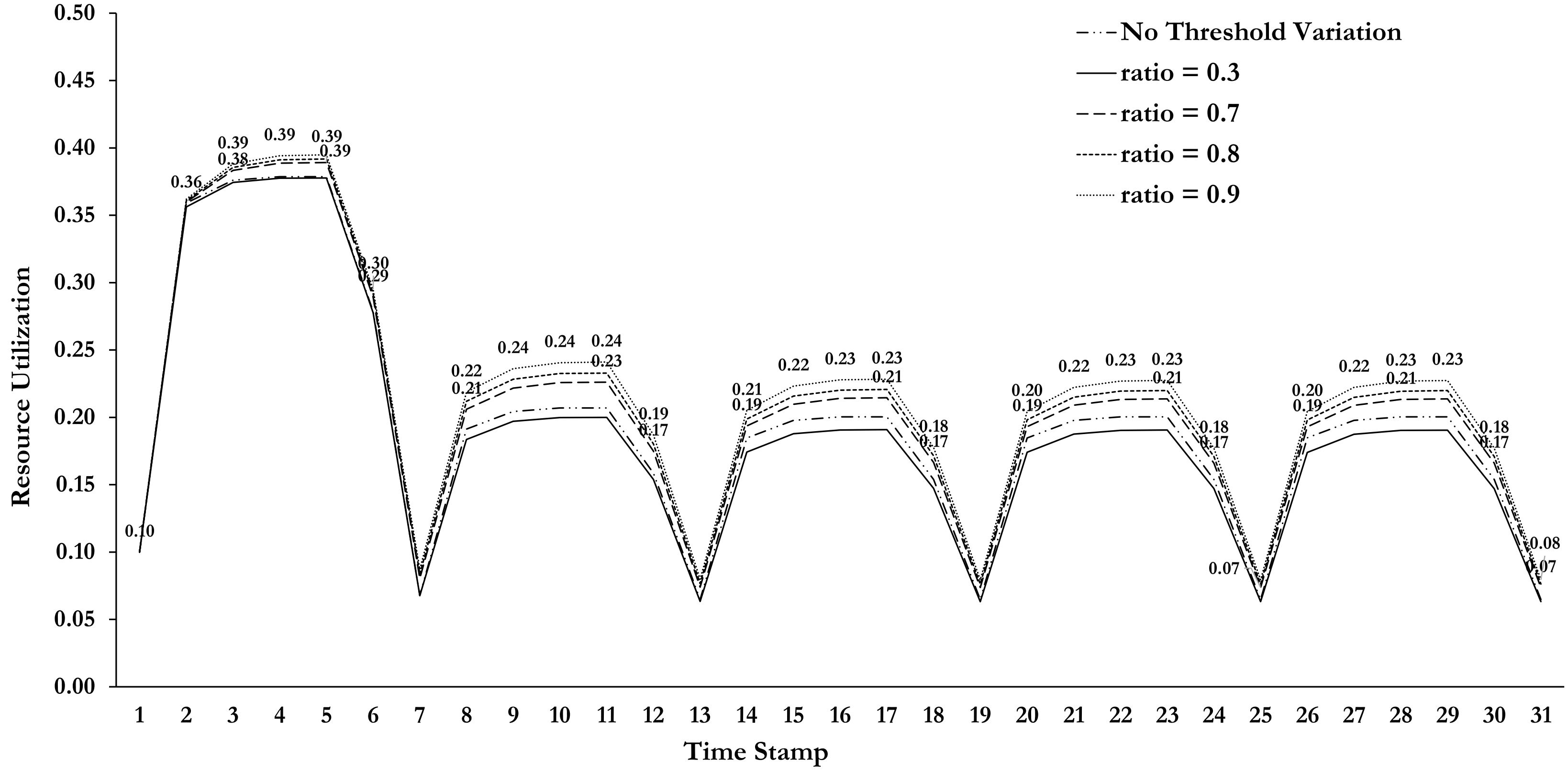

Figure A.2: $SVRS$- PA Threshold Ratio Variation ,$\alpha = 0.2$ - This is an extensive study on threshold variation to identify the ratio of population who will have positive or negative impact. By positive impact of threshold, we mean that the intention to perform a behavior of an individual will increase. Similarly , negative effect of threshold means that the intention to perform a behavior will decrease. We perform population ratio variation keeping alpha as 0.2 and the ratio 70 : 30 in the most interesting combination.

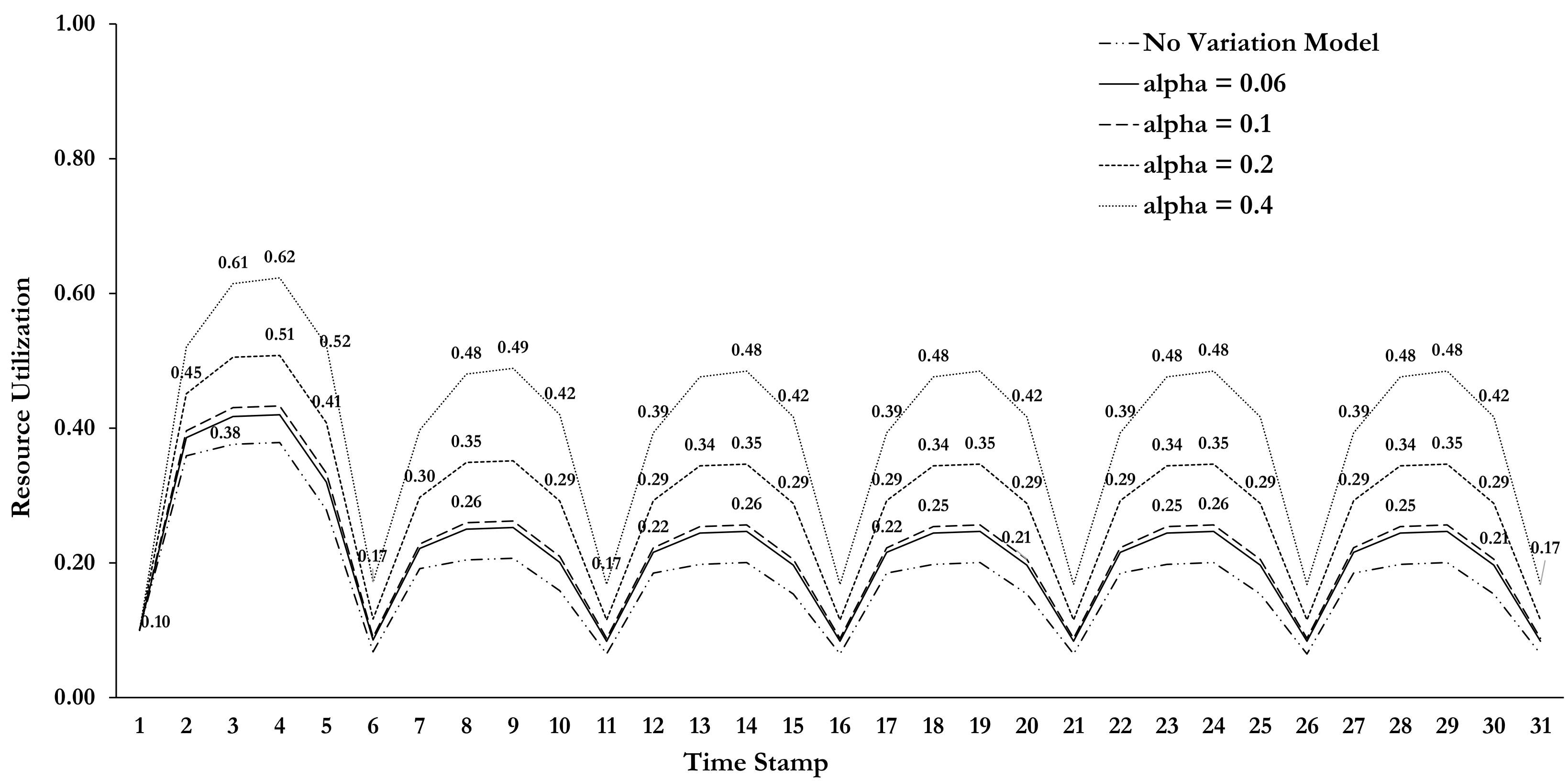

Figure A.3: $SVRS$- PA Cost α Variation , $Constant = 10$ - This is an extensive study on cost variation to identify the drop in adoption cost i.e the alpha over time. In this experiment, we kept constant as 10 and varied alpha to find the most appropriate alpha value. We observe that alpha as 0.2 is the most interesting value.

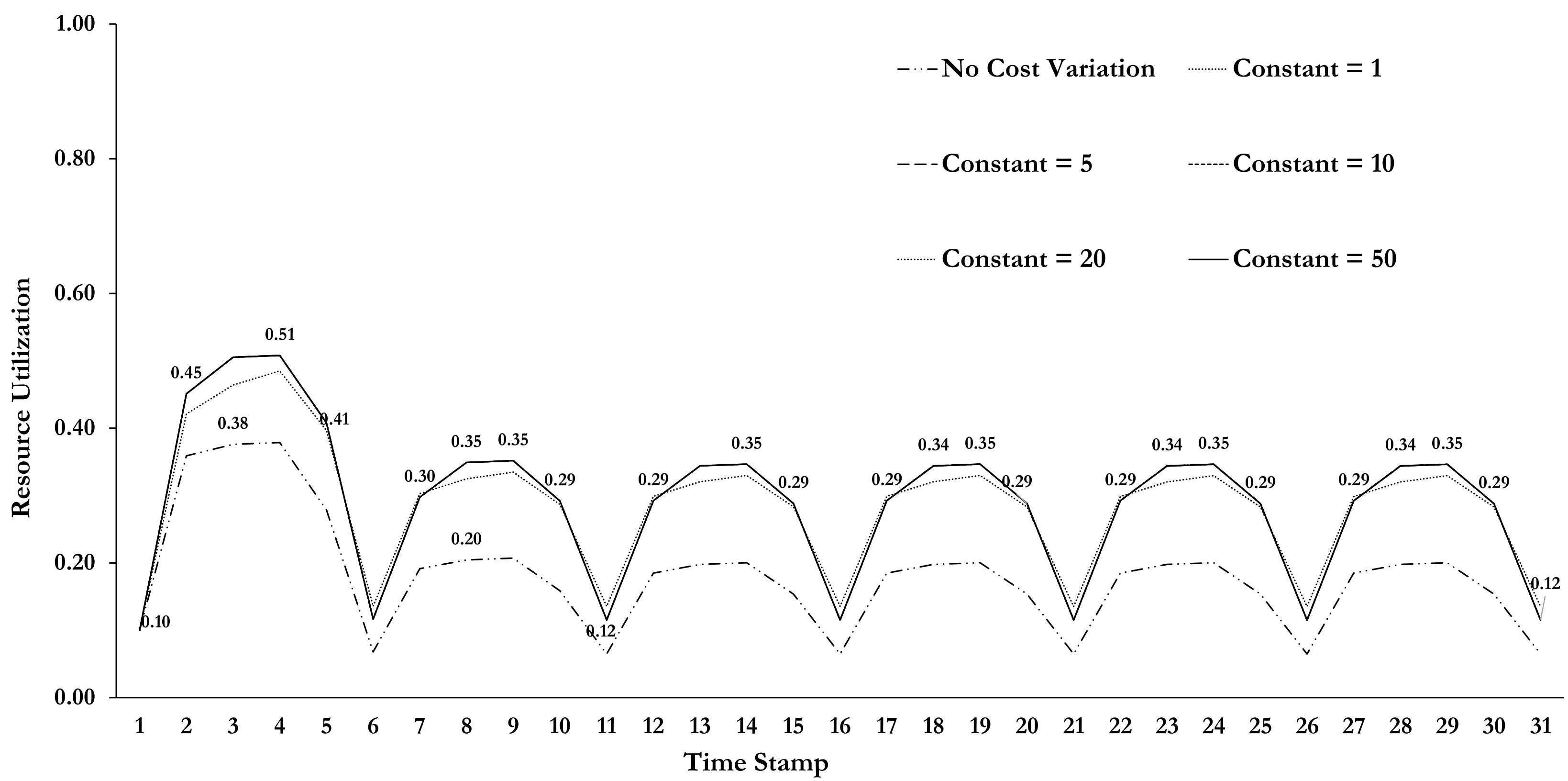

Figure A.4: $SVRS$- PA Cost Constant Variation , $\alpha = 0.2$ - This is an extensive study on cost variation to identify the constant in adoption cost over time. In this experiment, we kept α as 0.2 which we obtained form the previous experiment and varied constant over five values. After careful observation we find constant value as 10 to be the most appropriate one.

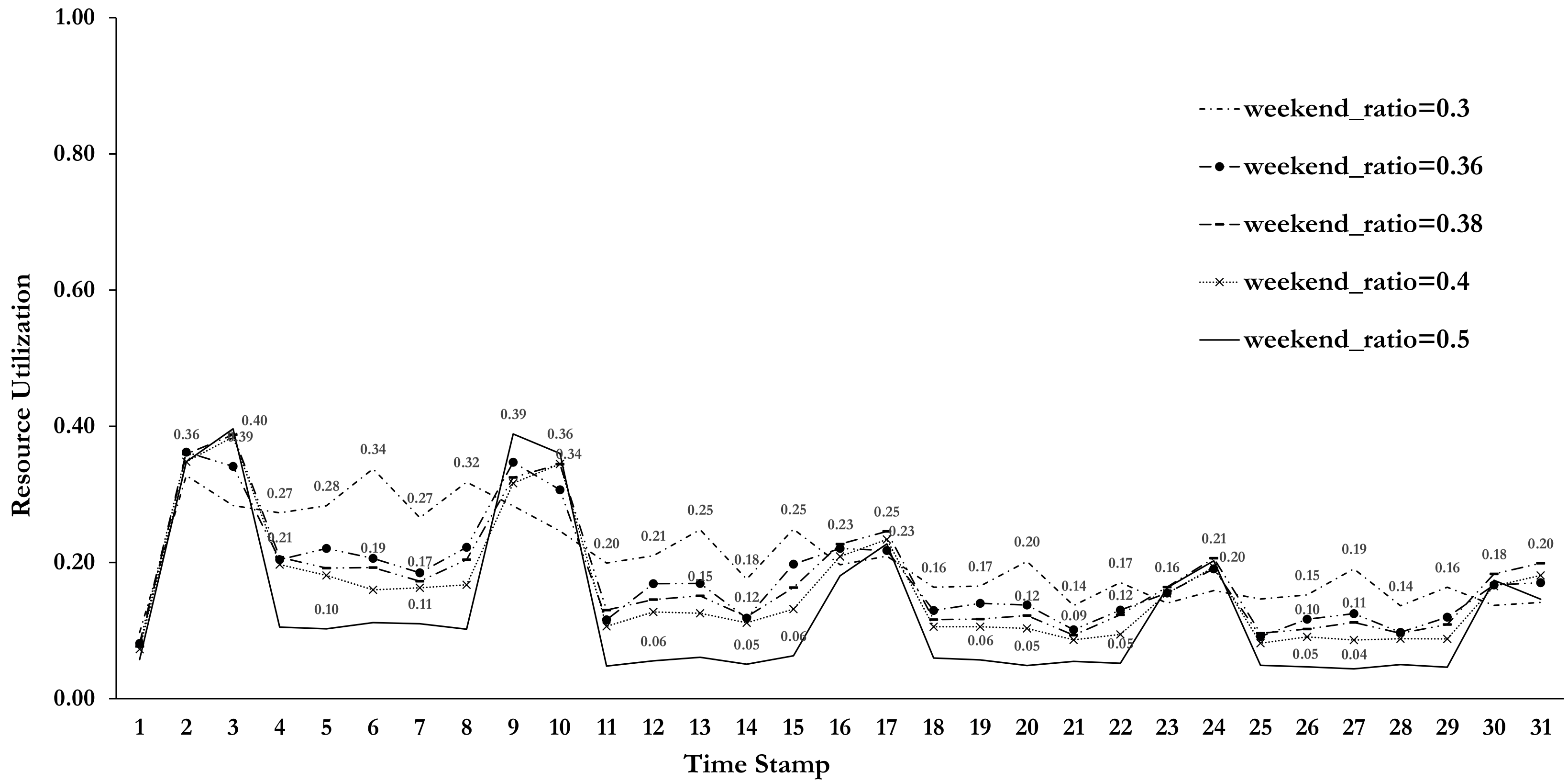

Figure A.5: *SVRS*- PA Resource variation with Weekend Peaks - In this experiment , we vary the average weekend resource availability. We observe if the average population weekend resource is 50% of their full resource then there are distinct peaks during weekends. As we are targeting population to have significantly higher resources during weekends , hence we are taking value as 50% for rest of the experiments.

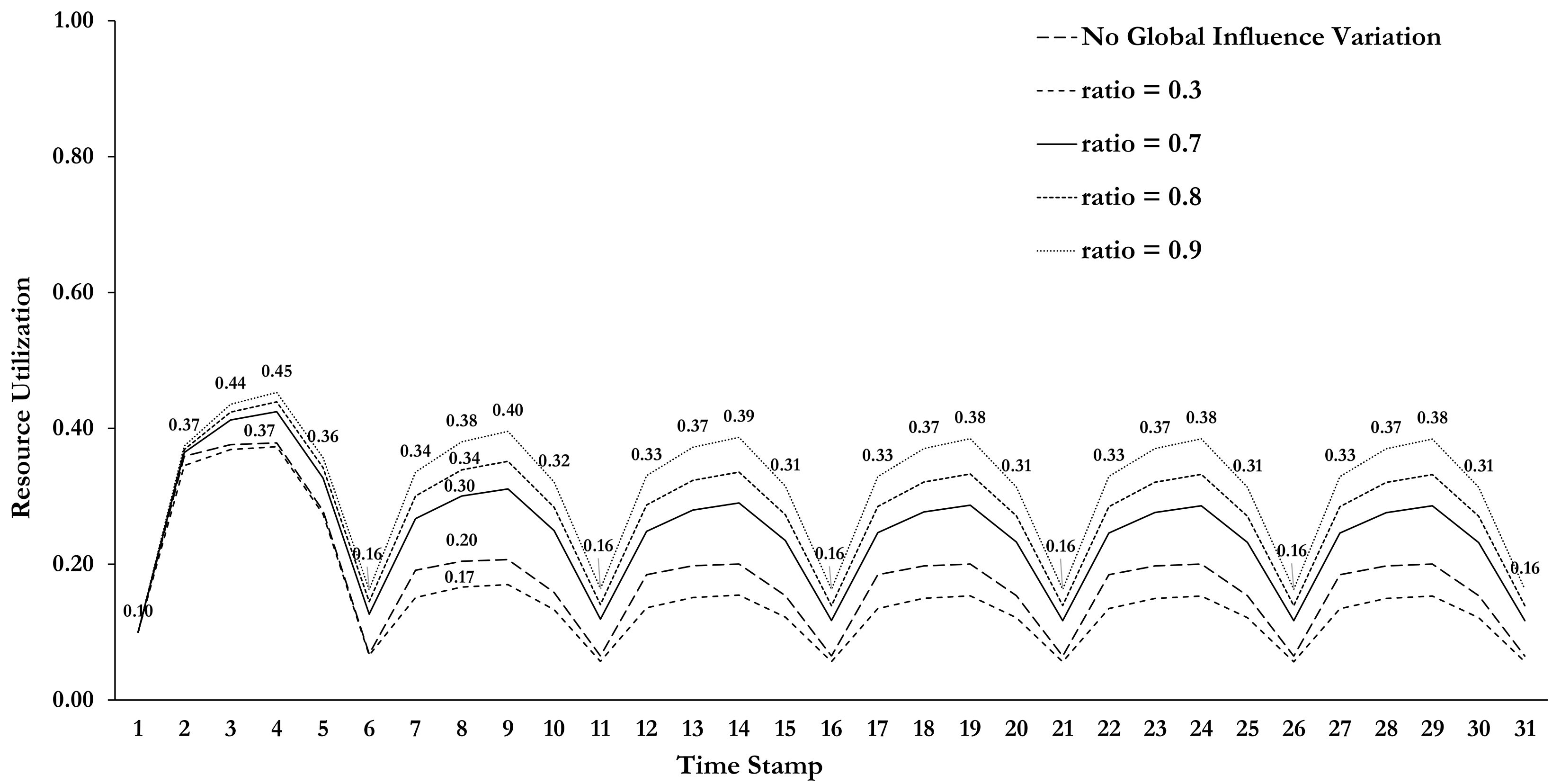

Figure A.6: *SVRS*- PA Global Influence Ratio Variation - In this experiment , we are varying the ratio of population having positive effects and negative effect over global influence variation. By positive effect means , an individual will be more keen to adopt a behavior if the global influence increases. And by negative effect we mean , an individual will be reluctant to adopt a behavior with increase in global influence. We perform a study with various population ratio and we observe that 70 : 30 ratio i.e 70% population having positive effect and 30 % population having negative effect is the most significant.

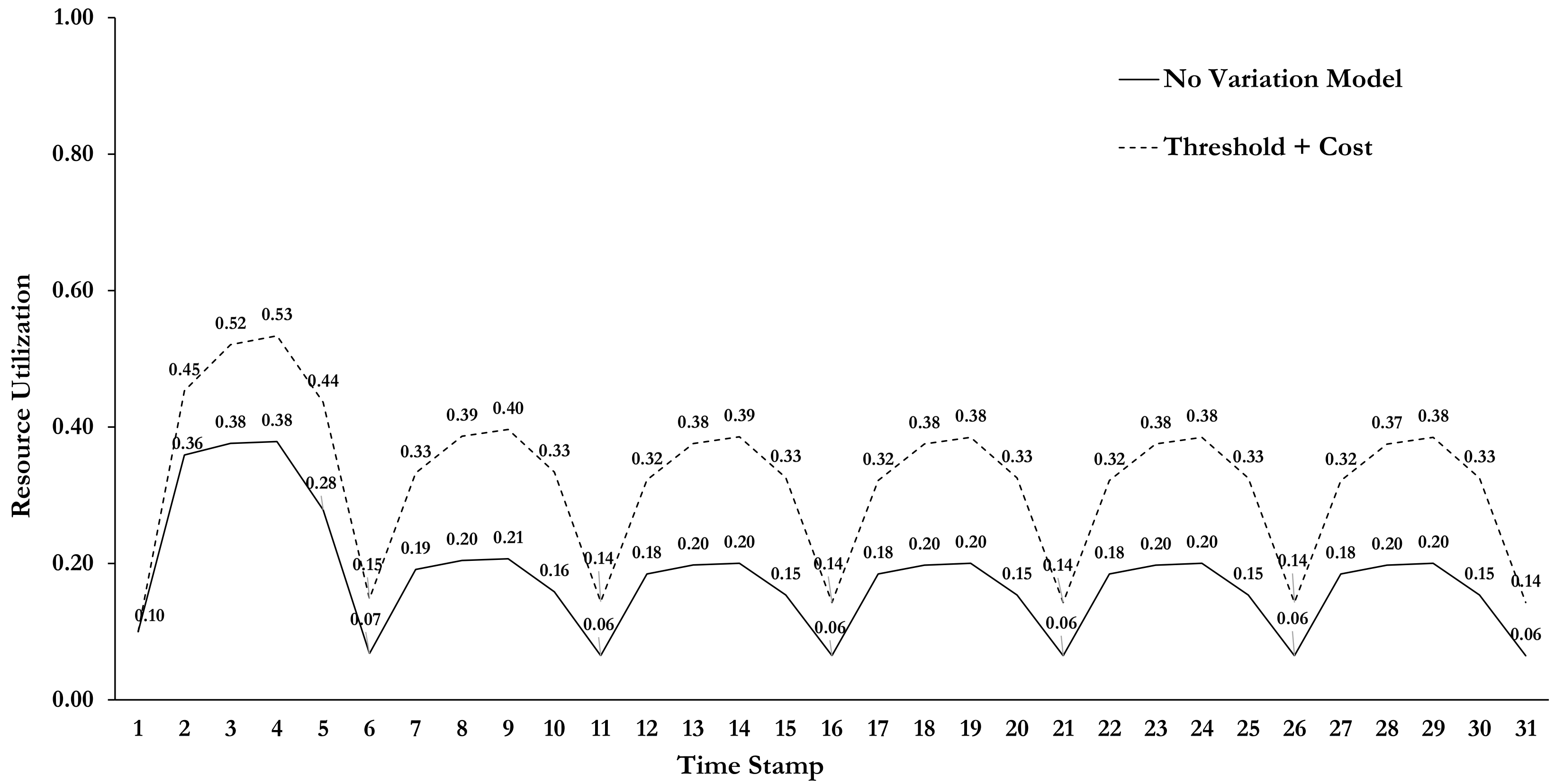

Figure A.7: $SVRS$- PA Threshold and Cost Variation $\alpha = 0.2$ and $constant = 10$ - In this experiment, we are varying both behavior adoption threshold and behavior cost over time. Note that we are fixing all the constants from previous experiments. Interestingly, we observe a statistically significant increase in resource utilization when both threshold and cost are varying together.

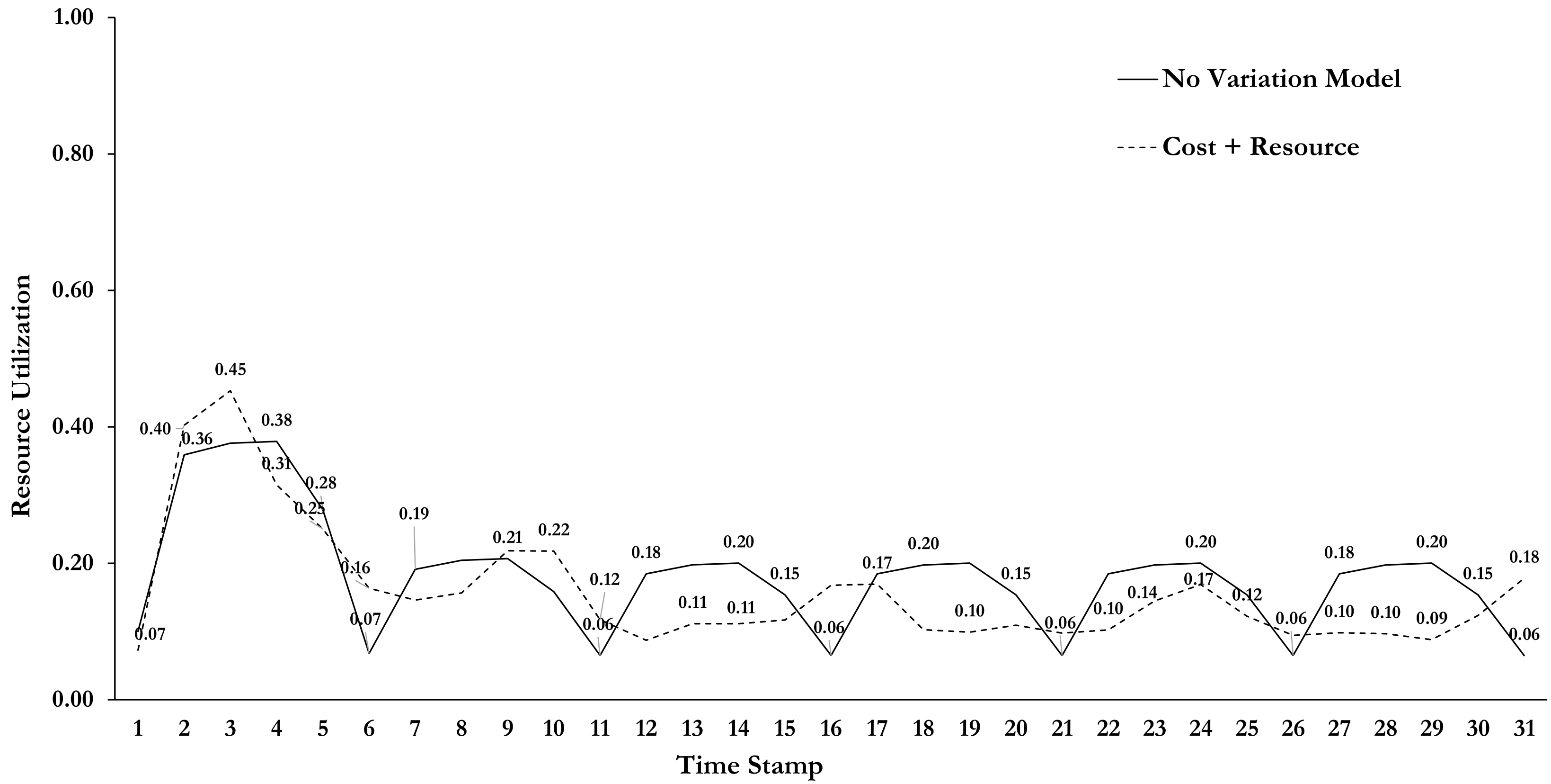

Figure A.8: $SVRS$- PA Cost and Resource Variation - In this experiment, we are varying both behavior cost and resource availability over time. Note that we are fixing all the constants from previous experiments by setting $weekend - ratio = 0.5$ and $\alpha = 0.2$. We observe that there is no significant increase in resource utilization but the resource variation among the population increases.

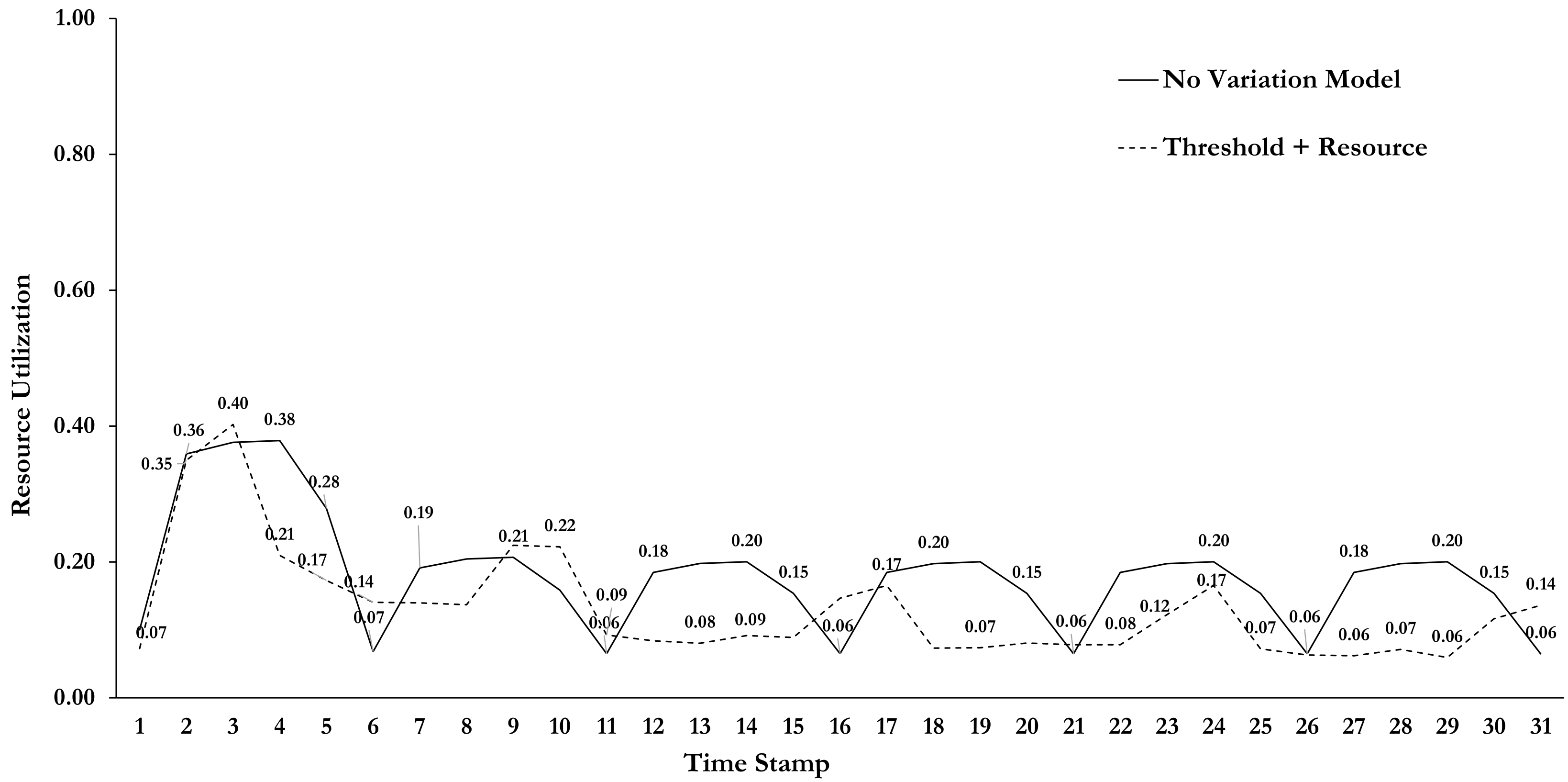

Figure A.9: $SVRS$- PA Threshold and Resource Variation - In this experiment, we are varying both behavior adoption threshold and resource availability over time. Note that we are fixing all the constants from previous experiments by setting $\alpha = 0.2$ and $weekend_ratio = 0.5$. We observe that there is no significant increase in resource utilization but the resource variation among the population increases.

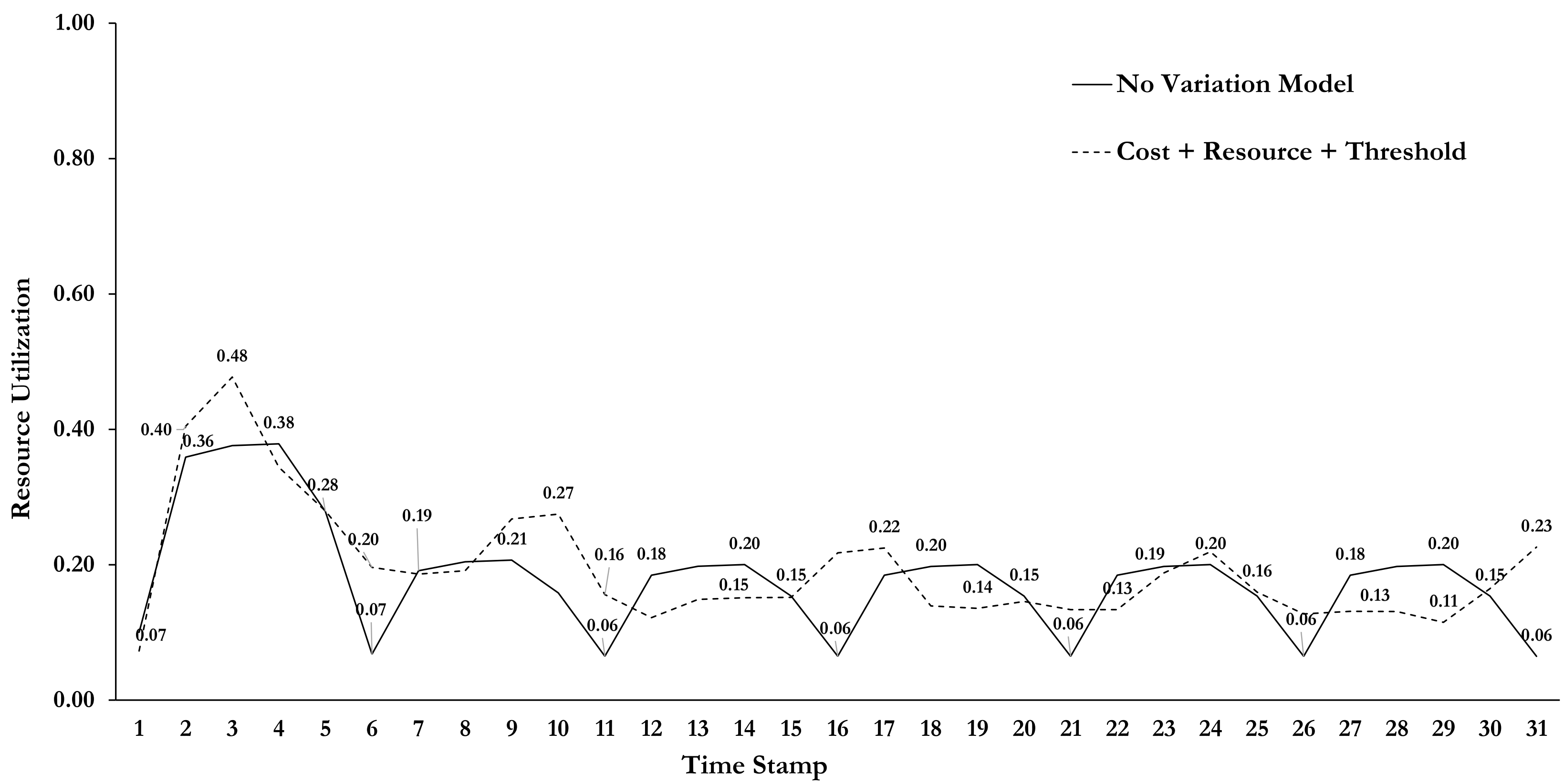

Figure A.10: $SVRS$- PA Threshold , Cost and Resource Variation - In this experiment, we are varying both behavior adoption threshold , behavior cost and resource availability over time. Note that we are fixing all the constants from previous experiments by setting $weekend_ratio = 0.5$, $\alpha = 0.2$ and $positive_ratio = 0.7$. Interestingly, we observe that there is no significant increase in resource utilization but the resource variation among the population increases.

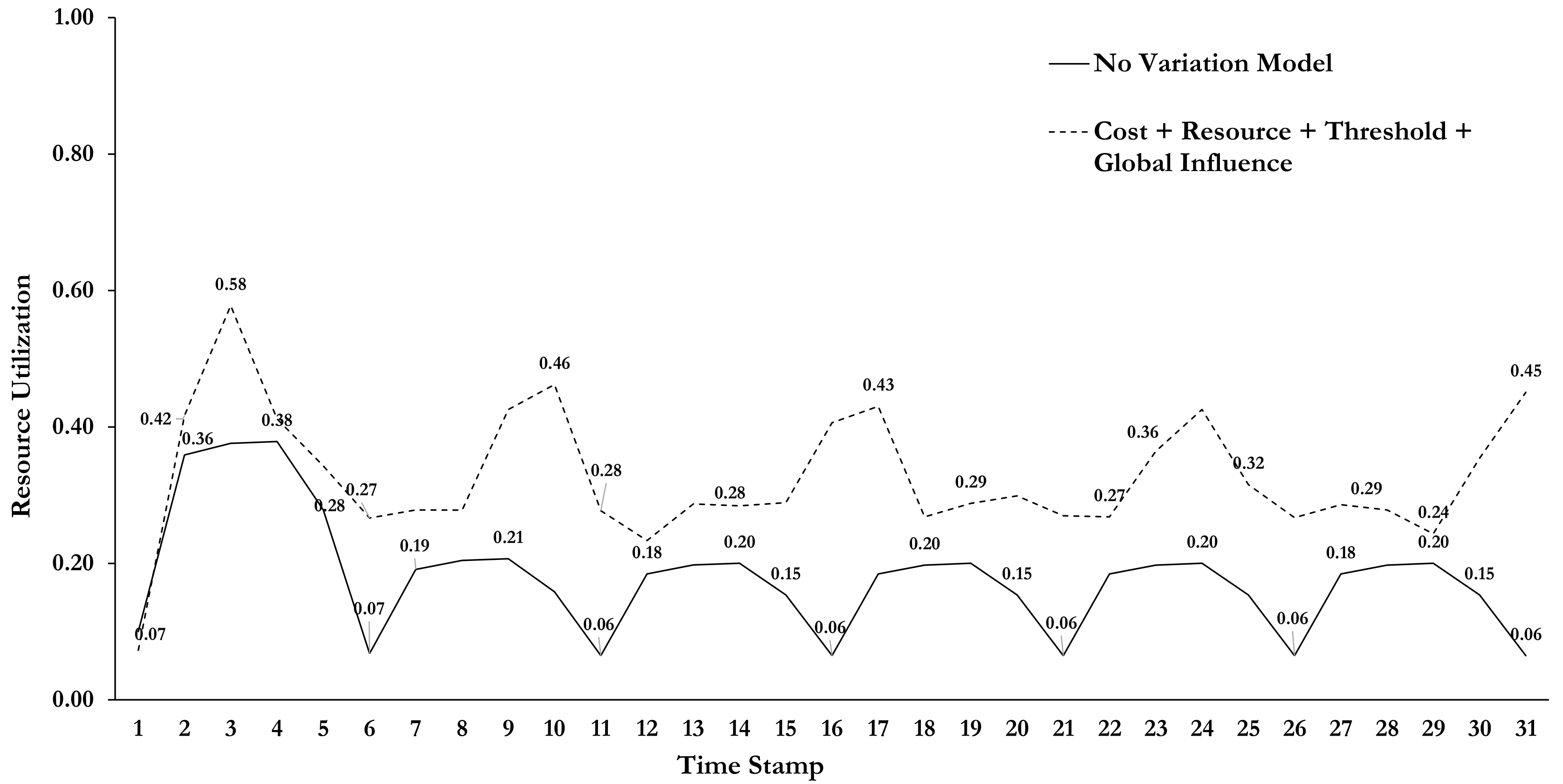

Figure A.11: *SVRS*- PA Threshold , Cost ,Resource and Global Influence Variation - In this experiment, we are varying both behavior adoption threshold , behavior cost , resource availability and global influence over time. Note that we are fixing all the constants from previous experiments. Interestingly, we observe a statistically significant increase in resource utilization when all the parameters are varying together and there is an increase in the range of resource availability among the population.