Influence Dynamics on Social Networks

A Thesis

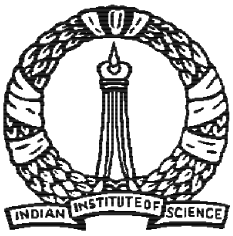
Submitted for the Degree of

Doctor of Philosophy

in the Faculty of Engineering

by

Srinivasan Venkatramanan



Electrical Communication Engineering
Indian Institute of Science, Bangalore
Bangalore – 560 012 (INDIA)

April 2014

To
my Parents and Sisters

[1] *Not all pages are intentionally left blank. Some, like our minds, get unintentionally filled.*

Acknowledgements

Quite often, the most (or perhaps the only) interesting part of a Ph.D thesis turns out to be the acknowledgements section. It is in this section, that we get a glimpse at the scaffolds of the cathedral, the palletes and brushes that made the painting possible. It is these short snippets that give hope to a faltering graduate student in his mid-Ph.D crisis, and make him realize that, in Henry Ford's words, "the greatest discovery a man makes is to find he can do what he was afraid he couldn't do." It is also in writing these words, the author realizes that with the beginning approaching its end, the journey has just begun.

I could not have hoped for a better place to begin this journey, than the Department of ECE, Indian Institute of Science, and under Prof. Anurag Kumar, my advisor. He is an unending repository of knowledge and enthusiasm, and I am also sincerely grateful for his patience and motivation. Often times, he has been the beacon and/or the tugboat during my graduate research. I will also carry along in my research journey the valuable lessons I learnt under him, in organizing one's thoughts and resources. I also thank him for providing wonderful lab resources, and funds for my conference travels.

I sincerely believe that, what a man creates is just a result of combinatorial exploration and exploitation of what he consumes. I would like to thank all the faculty members of the departments of ECE, Math and CSA at IISc., through whose courses I could set up the necessary foundations for my research. I would especially like to thank Prof. Rajesh Sundaresan (ECE) and Prof. Narahari (CSA), with whom I have had the closest interactions. I also thank Prof. Srikanth Iyer (Math), who, along with Prof. Rajesh and Prof. Narahari, was part of my comprehensive examination committee. I extend my gratitude to Prof. Vinod Sharma and Prof. Vijay Kumar who were the successive

department chairmen during my Ph.D. I am also thankful to Prof. Eitan Altman, one of my collaborators from INRIA (France), with whom I have had several insightful discussions. I would also like to thank the staff at Network Labs Office and the ECE Office, particularly Mrs. Chandrika, Ms. Priyanka, Mr. Mahesh, and Mr. Srinivasa Murthy for their friendly help throughout my Ph.D.

During my stay at the Institute, I got by a little, with a lot of help from friends [2]. I really enjoyed all the technical and philosophical discussions with the past and present Ph.D. students of Network Engineering Lab (NEL) and all my friends from ECE. In particular, I would like to acknowledge Venkatesh, Prem, Manoj, Chandramani, Naveen, Vineeth, Arpan, Abhijit, and Avinash, with whom I have shared not just the lab space, but also many candid discussions about research and life in general.

I am lucky to have had a wide social experience outside research, thanks to my friends at A mess, Rhythmica (IISc music band), and Quiz club. My life at IISc. would have been incomplete, but for those precious moments with my "mess gang". I am indebted to their friendship, for keeping me sane during my PhD (with a necessary doze of insanity). I would like to particularly thank Raman, Lakshminarasimhan, Prasad, Harish, Chandru, Janakiraman, Deepika and Arun who have been like my extended family at IISc. There is no better way to make music than with friends, and rarely a better way to make friends than through music, and Rhythmica provided me a chance to do both, the memories of which I will cherish forever.

I cannot begin to thank my parents and sisters, for being the one true reason for all my endeavors, for being there always with their unconditional love and support. This thesis is dedicated to them. I thank the Omnipresent Consciousness for bringing all these *influences* into my life, and I pray that I am reminded to be forever grateful for all the small and big things that have and will come my way.

According to the Hofstadter's Law, "it always takes longer than you expect, even when you take into account Hofstadter's Law", and it is nowhere more true than in graduate school. I salute the quintessential spirit of the graduate student, and along the lines of Hofstadter, I would like to acknowledge all those graduate students who have not acknowledged themselves in their thesis.

[2]This line was written with a "little" help from *The Beatles.*

Abstract

With online social networks such as Facebook and Twitter becoming globally popular, there is renewed interest in understanding the structural and dynamical properties of social networks. In this thesis we study several stochastic models arising in the context of the spread of influence or information in social networks. Our objective is to provide compact and accurate quantitative descriptions of the spread processes, to understand the effects of various system parameters, and to design policies for the control of such diffusions.

One of the well established models for influence spread in social networks is the threshold model. An individual's threshold indicates the minimum level of "influence" that must be exerted, by other members of the population engaged in some activity, before the individual will join the activity. We begin with the well-known Linear Threshold (LT) model introduced by Kempe et al. [1]. We analytically characterize the expected influence for a given initial set under the LT model, and provide an equivalent interpretation in terms of acyclic path probabilities in a Markov chain. We derive explicit optimal initial sets for some simple networks and also study the effectiveness of the Pagerank [2] algorithm for the problem of influence maximization. Using insights from our analytical characterization, we then propose a computationally efficient G1-sieving algorithm for influence maximization and show that it performs on par with the greedy algorithm, through experiments on a coauthorship dataset.

The Markov chain characterisation gives only limited insights into the *dynamics* of influence spread and the effects of the various parameters. We next provide such insights in a restricted setting, namely that of a homogeneous version of the LT model but with a

general threshold distribution, by taking the fluid limit of a probabilistically scaled version of the spread Markov process. We observe that the threshold distribution features in the fluid limit via its *hazard function*. We study the effect of various threshold distributions and show that the influence evolution can exhibit qualitatively different behaviors, depending on the threshold distribution, even in a homogeneous setting. We show that under the exponential threshold distribution, the LT model becomes equivalent to the *SIR* (Susceptible-Infected-Recovered) epidemic model [3]. We also show how our approach is easily amenable to networks with heterogeneous community structures.

Hundreds of millions of people today interact with social networks via their mobile devices. If the peer-to-peer radios on such devices are used, then influence spread and information spread can take place opportunistically when pairs of such devices come in proximity. In this context, we develop a framework for content delivery in mobile opportunistic networks with joint evolution of content popularity and availability. We model the evolution of influence and content spread using a multi-layer controlled epidemic model, and, using the monotonicity properties of the o.d.e.s, prove that a time-threshold policy for copying to relay nodes is delay-cost optimal.

Information spread occurs seldom in isolation on online social networks. Several contents might spread simultaneously, competing for the common resource of user attention. Hence, we turn our attention to the study of competition between content creators for a common population, across multiple social networks, as a non-cooperative game. We characterize the best response function, and observe that it has a threshold structure. We obtain the Nash equilibria and study the effect of cost parameters on the equilibrium budget allocation by the content creators. Another key aspect to capturing competition between contents, is to understand how a single end-user receives and processes content. Most social networks' interface involves a *timeline*, a reverse chronological list of contents displayed to the user, similar to an email inbox. We study competition between content creators for visibility on a social network user's timeline. We study a non-cooperative game among content creators over timelines of fixed size, show that the equilibrium rate of operation under a symmetric setting, exhibits a non-monotonic behavior with increasing number of players. We then consider timelines of infinite size, along with a behavioral model for user's scanning behavior, while also accounting for variability in quality (influence weight) among content creators. We obtain integral equations, that capture the evolution of average influence of competing contents on a social network user's timeline, and study various content competition formulations involving quality and quantity.

Contents

CHAPTER 1

Introduction

The origin of study of networks is generally traced back to 1736, when Leonhard Euler addressed the problem of circumnavigating the bridges of Königsberg, thus leading to the field now known as graph theory. Ever since, networks (or graphs, their mathematical abstractions) have been used to model a collection of interlinked entities and their relationships. Such systems could be natural, as in the case of ecological networks (e.g., foodwebs), and biological networks (e.g., protein interaction networks, metabolic networks), or man-made, as in the case of transportation networks, power grids, or the Internet.

One such network that permeates our everyday life is our social network. In a social network, each node represents an individual and if two individuals are related, there is a link between the corresponding nodes in the social network. Such social ties can be either personal (friends, family, acquaintances, etc.) or professional (co-workers, business partners, etc.). The links could also be directed, as in the case of parent-child relationships or employer-employee relationships. Such networks play an explicit role in information spread among its members (for instance, about job opportunities [4]), spread of disease [5], adoption of innovations [6]). They also implicitly dictate which products we buy [7], how we vote [8] and even our emotional state [9].

Though the idea of modeling a society as a network of individuals seems straightforward, it was not until the late 1960s that sociologists began modeling human relationships

in the form of a social network, to study human behavior in groups. Through his now famous experiment [10], Stanley Milgram demonstrated the surprisingly *small-world* nature (small average path length) of social networks, thus piquing the curiosity of several sociologists to investigate the impact of network topology on social interactions. A little later, Granovetter [4] identified that most people get crucial information on job opportunities through their acquaintances (as opposed to their close friends circle), thus expounding the *strength of weak ties* in social relationships.

With the advent of the Internet, digital interactions such as emails, newsgroups etc. paved the way for a richer and more complicated global social network of interactions. This also enabled a more concrete empirical documentation of social interactions, shifting the emphasis in social science from a qualitative to a quantitative approach. Around the same time, thanks to the physics community, researchers began noticing over-arching similarities in real-world network data arising from various domains. Until then real-world networks were generally modeled using random networks [11] or regular networks, but the evidence provided by data indicated that their characteristics were vastly different from either abstractions. For instance, real world social networks seemed to exhibit both small average path length and high clustering coefficient, unlike random networks (low clustering), or regular networks (high average path length). Watts and Strogatz [12] noted that the small-world property was not restricted to social networks, but also showed up in power grids, neuronal networks, etc. Meanwhile, Barabasi et al. [13] observed that most real-world networks seem to exhibit a scale-free property, i.e., where some nodes, called hubs, have many more connections than others, with the degree distribution of the network exhibiting a power-law behavior. The discovery of small-world and scale-free properties of real-world networks spawned a decade of intense activity in various research communities cutting across disciplines, leading to the body of literature now collectively referred to as network science [14]. Such investigations are aimed at providing a richer mathematical foundation for understanding the structure and dynamics of complex networks.

While the study of dynamics *of* complex networks pertains to the study of evolution of network structure, the study of dynamics *on* complex networks [15] focuses on processes that take place on the networks, such as diffusion. In the context of social networks, such diffusion processes can help explain the spread of influence [1], collective behavior [16], epidemics [3], adoption of innovation [6, 17], etc. With the proliferation of social media and e-commerce websites, there is renewed interest in such models, since characterizing

influence spread could help identify influential users who can be targeted for viral marketing [18]. Since the same process can also explain how content sharing occurs on the Internet [19] (i.e., information spread), they can help predict evolution of content popularity, identifying which contents may go viral [20, 21]. The understanding and prediction of popularity evolution of such content is crucial to the content provider for (i) choosing appropriate content caching strategies for quick delivery (ii) deploying better advertisement mechanisms for increased monetization. Content-hosting platforms also use this popularity information to re-rank the content, thus making highly popular content more visible, leading to an analogue of the preferential attachment [22] phenomena on complex networks. This also explains why a significant fraction of content gets very limited views, while a few go "viral" [23].

Over the past decade, online social networks such as Facebook (1.1 billion) and Twitter (350 million), have amassed a significant fraction of the Internet population within the span of a decade, and have had a considerable impact on how human society exchanges information. They have profoundly impacted the dynamics of the news cycle [24], and have enabled quick dissemination of time-critical information, for instance in the case of natural disasters [25], [26]. Large-scale social uprisings such as the Arab Spring [27], Occupy Wall Street [28], have also been organized primarily via the social media. While the above are some instances of how the consumers use social media, their daily online interactions on the social media websites leave behind a *digital exhaust* [29], helping the social network sites better understand the user preferences. Since the Internet economy is predominantly based on advertising, the big data generated via user profiling provides invaluable information to viral marketers who can then carry out targeted, personalized campaigns [1].

Thesis Summary: Since the renewed interest in the area is predominantly due to the availability of *big data* from online social networks such as Facebook, Twitter, etc., there is a vast body of empirical research pertaining to the structure and dynamics of social networks [30], [31]. However, in this thesis, we take an analytical approach to studying models of influence dynamics that arise in social networks. We believe that the problems addressed in this thesis and the analytical solutions developed herein, contribute to the theoretical understanding of the information diffusion process and pave new directions for both analytical and empirical study in the area of influence dynamics on social networks [1]. The first two chapters will analytically study one of the well known models for influence

[1]Since the underlying dynamics are similar, unless when explicitly required, we will use the terms

spread, known as the threshold model [16]. In a social network, an individual's threshold indicates the minimum level of "influence" that must be exerted, by other members of the population engaged in some activity, before the individual will join the activity. Chapter 2 focuses on the Linear Threshold (LT) model introduced by Kempe et al. in [1], and analytically characterizes the spread of influence for a given social network, using which we propose a computationally efficient algorithm for the problem of influence maximization. Chapter 3 provides the link between the LT model in [1] and the threshold model introduced by [16] to study collective behavior. We do so through the use of fluid limits, obtaining a system of o.d.e.s characterizing the influence evolution, and also show the equivalence between the LT model and the (Susceptible-Infected-Recovered) SIR model from epidemiology [3].

While Chapters 2 and 3 focus on the spread of a single diffusion process, In Chapter 4, we look at a possible application scenario where influence spread models are used in conjunction with content spread in a mobile opportunistic network setting. The influence spread model is used to characterize the evolution of interest in the content. In this scenario, we identify optimal content forwarding mechanisms, which optimize for a combination of delivery delay to the interested nodes and wasteful copying to the nodes not interested in the content. In Chapter 5, we look at competition between contents in the context of online social networks. While models of competitive influence spread have been studied in the past [32], [33], this chapter focuses on two aspects of competition heretofore not addressed in existing literature. A common theme in Chapter 5 is that of limited user attention. Limited user attention can manifest itself either as (a) the amount of time a user spends in a given social network [34] or (b) the way a user receives information while present in the social network [35]. The first part of Chapter 5, studies competition between contents, when users choose probabilistically among multiple social networks. The second part of Chapter 5 studies the effect of user attention when he scans the various contents presented to him in the form of a *timeline*.

Since a significant portion of this thesis focuses on influence spread, in the following section, we begin with a brief literature survey of epidemic processes on networks. Though the process has analogues in various disciplines ranging from telecommunication, biology, economics, etc., we will restrict ourselves to the literature from the fields of epidemiology and social networks. We then briefly provide an overview of each of our subproblems, discussing some existing literature. We finally conclude this chapter by summarizing the

information spread and influence spread interchangeably

key contributions of this thesis. Subsequent chapters discuss each of the problems in detail, beginning with the respective system models.

1.1 Literature Survey

1.1.1 Epidemics on Networks

Mathematical modeling of epidemics date back to 1927, with the Kermack-McKendrick birth-death model [36], now widely known as the SIR (Susceptible-Infected-Recovered) epidemic model. They identified the epidemiological threshold, and showed conditions under which the epidemic outbreak occurs. Further extensions have been done to the SIR model [37] in order to accommodate a variety of disease specific assumptions. Most epidemic models assume homogeneous mixing, i.e. the agents are uniformly distributed over a geographical region, with the disease spread parameters uniform across all the agents. Though such models are analytically tractable, they might be unrealistic for most practical scenarios. This lead to the development of network epidemiology [38], which studies the spread of a contagion through a network. Here, in addition to the disease parameters such as infection and recovery rates, the network topology also plays a significant role in the spread of epidemics.

It was identified that, in the case of sexually transmitted diseases, the variance in the node degree played a significant role in early stages of the epidemics [39]. Newman [40] studied the spread of disease on networks under the susceptible-infected-removed (SIR) model and showed how concepts from percolation theory can be used to study these models on a wide variety of networks. He also provided exact solutions for the SIR model on networks, even when the infectiveness times and transmission probabilities were non-uniform and correlated. Pastor-Satorras and Vespignani [41] studied the dynamics of infection spread on a scale-free network, and identified the absence of an epidemic threshold, and validated it using real-world data from computer virus infections. A detailed survey of the analytical, simulation and sampling methods employed in the area of network epidemiology is provided in [42].

Rumor spreading on social networks can also be classified based on who initiates the spread during pairwise contacts. In the *push-model*, the nodes that are aware of the rumor, contact their neighbours to spread them. Such models have been used to model spread of computer viruses, and also in studying data storage [43] and distributed computation on networks [44]. On the contrary, the hybrid *push-pull model* [45] aim to capture the effect

of gossiping on social networks, where nodes periodically poll their neighbours for any information update. By using the hybrid model, in [46], the authors establish the relation between the conductance of a graph and the time taken for rumor spread. Rumor spreading models have also been studied as random-walks on graphs [47], and it has been observed that the characteristics of spread are closely related to the spectral properties of the underlying graph. For instance, [48] demonstrated that the speed of virus spread is directly related to the spectral radius of the adjacency matrix of the graph. In [49], it was proposed that influence spread can be tractably represented as dynamics of networked Markov chains, with the state of each node evolving in time. They show that the interactions can be studied in a reduced-order influence matrix, whose eigenstructures reveal how the influence spreads through the full-order transition matrix. In Chapter 2, we provide interesting relations between self-avoiding paths in Markov chains and influence spread under the LT model, and show how measures like Pagerank [2] perform for the problem of influence maximization.

1.1.2 Influence Maximization

While network diffusion processes have been investigated in various domains, it was not until the early 2000s that it attracted the attention of computer scientists. With widespread adoption of Internet, humans (aided by computers) and computers were exchanging information at an unprecedented scale. While a community of researchers were interested in network diffusion from the perspective of containment of computer viruses [50], another community, motivated by the concept of viral marketing, were looking at the problem of influence maximization. Viral marketing hinges on the concept of a customer's network value through his neighbours, apart from his intrinsic value. For services whose utility increases with the number of adopters (such as Hotmail, MSN chat messenger), they observed that piggybacking such viral marketing ads on the actual messages led to exponential adoption curves [51].

Domingos and Richardson [18] were the first to study information diffusion under the viral marketing perspective, and they posed the *combinatorial optimization problem of choosing the initial set of customers to maximize the net profits*, and showed that choosing the right set of users for the marketing campaign could make a significant difference. Kempe et al. [1] studied the problem of choosing the most influential initial set using two different models of information propagation, namely the Independent Cascade model (IC model) and the Linear Threshold model (LT model). While in the IC model, each node

independently tries to influence each of its neighbours, under the LT model, a susceptible node gets influenced due to the collective influence of his neighbours. Under both models, they showed that the problem of influence maximization is NP-hard and the objective function is *submodular*, i.e., exhibiting diminishing returns. Using this insight, they proposed a *greedy hill-climbing algorithm* that was shown to achieve an approximation factor of $(1 - 1/e)$. They also provided generalizations of the two models, and showed the equivalence between the generalized versions of the LT and IC models.

Following Kempe et al.'s seminal work, there have been two primary directions of focus in the influence maximization literature: scalable heuristic replacements for greedy algorithm, efficient computation of the greedy algorithm. Guille et al. [31] provide a survey of the representative methods capturing the state of the art in this field. Leskovec et al. [52] studied the problem of cost-effective outbreak detection in the context of Internet news blogs. Outbreak detection can be modeled as selecting nodes in a network, in order to detect the cascade at the earliest. Simply put, this amounts to identifying the blogs a user should follow, to avoid missing important stories. They then showed that the problem of influence maximization is equivalent to that of cost-effective outbreak detection, and developed lazy forward-selection based heuristics (CELF), to efficiently solve the problem. Chen et al. [53] provided approaches to improve the running time of the greedy algorithm, as well as proposed degree discount heuristics for the independent cascade model that are easily computable, and almost matching the performance of the greedy algorithm. Later, they observed that influence computation in (directed acyclic graphs) DAGs [54] can be performed in linear time, and thus provided a scalable influence maximization algorithm for the linear threshold model. There have also been approaches based on cooperative game theory to obtain the set of most influential nodes [55]. Thus, there is need for scalable heuristics for the influence maximization problem, motivated by the analytical structure of influence dynamics, which we provide in Chapter 2.

1.1.3 Threshold models for Influence spread

Though the LT model was first formalized by Kempe et al. [1], threshold models are well established in sociology for modeling the evolution of popularity in human populations. Everett [6] explored the adoption of innovations, employing examples from rural sociology, and noted the diversity in people's propensity to adopt an innovation. He categorized them into various adopter groups (see Figure 1.1) in what is now known as the Everett's bell

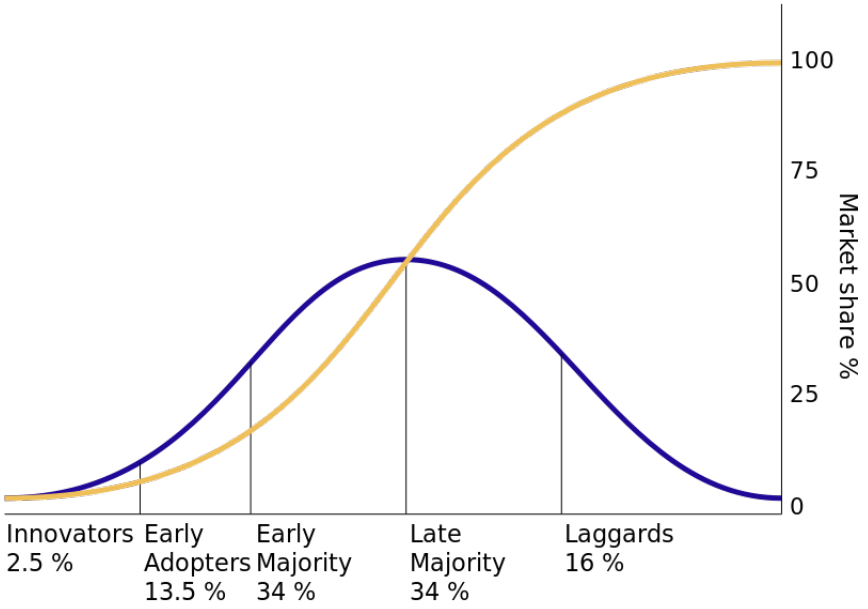


Figure 1.1: Everett's "bell curve". It is also known as the Technology Adoption Lifecycle since it represents the adoption of a new innovation over time. The bell shaped curve represents the incremental change in adoption over time, whereas the "s" shaped curve shows the cumulative adoption. Under the threshold interpretation, the early adopters (innovators) can be thought of as having the least threshold for adoption, and so on.(Source [6])

curve[2]. It is used to represent the process of adoption of a new product over time. Under the threshold interpretation, users in the leftmost class (innovators) can be interpreted as having the least threshold to adopt an innovation, and the ones in the rightmost class (laggards) as having the highest threshold (and hence least susceptible). The problem of interest is to understand the *evolution of the diffusion process, given the threshold distribution.*

Granovetter, in his seminal work [16] on collective behavior, aimed to use threshold distributions to model the spread of binary decisions among a group of rational agents, for instance during riots, voting, etc. Using his model, he calculated the equilibrium, i.e., steady state split of the population (between the binary decisions) given the threshold distribution, and also considered the stability of such equilibria. Valente [17] further refined this approach to threshold phenomena based on personal networks (local neighbourhood of an individual) as against whole social systems, and empirically studied datasets on the

[2]Everett's use of the term "bell curve", is not be taken to imply that the thresholds are normally distributed.

adoption of medical and rural innovation. Recently, in the context of user-generated content on the Internet, game theoretic analysis has shown that threshold based policies could emerge as equilibria when user's seek to maximize their utility (based on the perceived quality of the content) [56].

While in epidemiological models infection events are treated as independent of each other, in threshold models, the impact of additional influence depends on the total influence already acting on the node. In [57], Dodds and Watts showed that using an *influence response function*, one can bridge these two classes of models, and by looking at the influence response function, one can identify which epidemics will spread. Watts [58] further adapted Granovetter's model to networks, and observed that an epidemic could occur only if the underlying network had a "percolating vulnerable cluster," i.e., a relatively small population that once activated triggers a large spread of epidemic. He demonstrated that the existence of such a critical mass depended on heterogeneity of both the thresholds and the connectivity (degree) of the nodes. The above two efforts, however, focus on epidemic models where the influence dynamics is order-independent, i.e., vertices update states in a random, asynchronous fashion. Thus, they do not encompass models like the LT model proposed in [1]. Also, in [1], the threshold distribution is assumed to be uniform, and it does not provide any characterizations for a general threshold distribution. In Chapter 3, we provide a necessary link between the LT model and classic epidemiological models, obtaining fluid limits of the LT model under a general threshold distribution.

1.1.4 Content spread in MONETs

While the first two chapters of this thesis focus on the influence spread process, Chapter 3 studies one possible application of influence spread models in the context of a Mobile Opportunistic NETworks (MONETs). Due to the ubiquity of cellular networks, there has been a proliferation of hand-held mobile devices. The idea of MONETs is to exploit the mobility of users carrying such devices to transfer content in a device-to-device (D2D) [59] fashion during chance meetings. This is enabled by low-power radio interfaces on these devices (such as Bluetooth or WiFi Direct), and provides the opportunity for creating a multi-hopping communication network completely bypassing the cellular infrastructure. The main advantage of such a system, is to use the local connectivity (Bluetooth,802.11) to provide global connectivity for the mobile devices. Since such a scheme cannot meet delay guarantees, applications that utilize mobile opportunistic networking must be *delay tolerant*. On the other hand, such a peer-to-peer (P2P) content delivery is scalable [60],

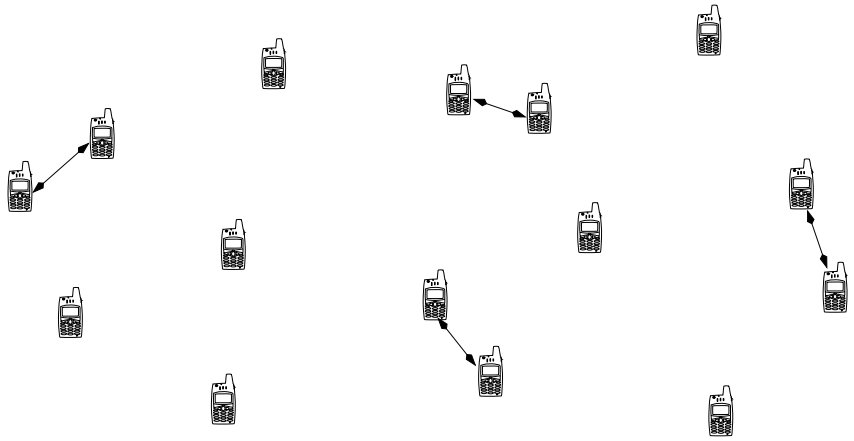


Figure 1.2: MONET over personal devices

as the rate of service scales in proportion to the number of peers in the system with little additional cost to the system planner.

Also termed as, Pocket-Switched Networks (PSNs) [61], research in the area of MONETs has been facilitated by the availability of large datasets [62] of network traces of bluetooth enabled motes. In [61] the authors discuss the major open problems in the context of PSNs, and stress on the importance of characterizing human mobility, to design better forwarding algorithms in such networks. [63], [64] propose mobility models that take into account social and community structure among the users in the population, inferred from the mobility traces. The effect of inter-contact time distributions on the content delivery properties of PSNs is studied in [65]. In [66], the authors propose BUBBLE Rap, a social-based forwarding algorithm which is shown to provide increased forwarding efficiency compared to social-oblivious algorithms. The algorithm uses knowledge about the hierarchical community structure and the centrality data from the underlying social network. They show how the algorithm can be implemented in a decentralized manner in pocket switched networks. The impact of social preferences in DTN routing is analyzed in [67], where they propose Social Selfishness Aware Routing (SSAR) to allow for user selfishness and obtain heuristics for better routing performance.

There has been considerable work in proposing improved architecture and protocols for an opportunistic network over mobile nodes [68], [69]. An interesting aspect, that is largely unexplored in prior literature of content delivery in MONETs, is the *evolution of interest in the content*. Models from epidemiology and influence spread literature can be used to capture this demand evolution. The content could be a video (news footage, sports highlights, a movie teaser, etc.) or an audio file (a recent song, a popular ringtone, etc.). For recently "released" content, the set of nodes interested in receiving the content, (also

known as *destinations*), need not be static. The interest in the content could grow, due to the influence of destinations already present in the population. Such influence could be mediated via a centralized server, which uses a low bit-rate channel to broadcast the current popularity of the content, or by interactions between the mobile nodes themselves. Thus, it may be necessary to deliver the content to destinations while keeping track of the demand evolution.

Shakkottai and Johari [70] explored the aspect of demand evolution by characterizing it using an epidemic model and obtained a hybrid of P2P and client-server architecture to efficiently meet the demand. In such a scenario, the system objective could be to facilitate the spread of content to as many destinations as possible. In doing this, the nodes not yet interested in the content (also known as *relays*) can be used to cache the content, thus making the content more available in the population. This could cost the system planner, due to the need for incentivizing the relay nodes to participate in relaying the content. The overall objective would then be to deliver the content to as many destinations as possible, while minimizing the number of nodes that have the content but never became interested in it. This aspect of exploiting the relay nodes for content delivery was not explored in [70]. In the delay tolerant networking context, there is prior literature (see [71], [72] and the references therein) on the optimal opportunistic copying of content in order to optimize delivery delay and/or wasteful copying to relays, but neither of them consider evolution of interest in the content. In Chapter 4, we combine a model for interest evolution, with content delivery in mobile opportunistic networks, and demonstrate the delay-cost optimality of a time-threshold policy.

1.1.5 Content Competition on Social Networks

Innovations seldom spread in isolation on a social network. When several contents spread simultaneously on a social network, they interact with each other, either competitively or cooperatively. For instance, two news stories on a related topic could "help" spread each other (users who read one of them might be directed explicitly to the other). On the other hand, in the case of ad campaigns of competing companies, one spreads at the cost of the other. This effect was studied using a statistical model for cooperating and competing contagions in [73] and evaluated on contagions spreading on Twitter, where they observed a significant level of interaction between the contagions. Karrer and Newman [74] studied competition between two epidemics on a complex network, and demonstrated that there was a sharp transition between the dominance of one disease and the dominance of

the other, dependent on the relative epidemic parameters. They also demonstrated the existence of a coexistence regime, where both epidemics infect a significant fraction of the network. Beutel et al. [75] considered epidemics that are not necessarily exclusive, for instance user adoptions of browsers, where a user could install both Firefox and Chrome. Even in the scenario of non-exclusive epidemics, they established that there exists a similar phase transition, between *winner-takes-all* (dominance of one disease) and co-existence.

In [33], the authors extend the Independent Cascade (IC) model proposed in [1] to capture multiple competing influences. They show that the resulting model is an extension of the Hotelling's model of competition [76], and is related to Voronoi games for competitive facility location [77]. Borodin et al. [32] studied competitive versions of influence spread under the Linear Threshold model, and showed the straightforward generalizations of the model yields objective functions which are non-submodular. This implies greedy algorithm for each competing user will no longer be $(1 - \frac{1}{e})$ optimal. They then provide a modified version of the competition problem, which is amenable to the greedy approach.

While the earlier work in this domain have been predominantly empirical, there have also been attempts to study the competition between contents from a game-theoretic framework. Goyal et al. [78] study the competition between firms with budgets for initial seeding of the network. They observe that when the adoption dynamics exhibits diminishing returns, the Price of Anarchy (measuring the inefficiency of resource use at equilibrium) is uniformly bounded, and the adoption dynamics which are not submodular could have unbounded PoA. They also show that the imbalances in player budgets could be amplified at equilibrium. Past work in this area have focused on the competition between contents in social networks over popularity and visibility. In [79], [80], the authors study competition for popularity between content creators, over advertisement spaces in social networks and websites like YouTube. It has been noted that these problems are similar in nature to the problem of competition over shelf space, a problem well studied in Operations Research (see [81]).

One aspect that is largely unexplored in the literature of content competition, is characterizing competition across multiple social networks. We observe that there are several online social networks, such as Google+, Facebook, Twitter with considerable user-base derived from a common pool of consumers (i.e., the set of all Internet users). Each consumer spends varying amounts of time in each of these social networks, depending on the network's popularity and usefulness. Thus, the content creators need to simultaneously manage campaigns across multiple social networks. Recent literature has studied

the aspect of migration across such social networks [34], but modeling content competition across several such social networks has not received sufficient attention, and we study one such model in Chapter 5.

1.1.6 Timeline model for User Attention

The competition between contents usually arises due to limited user attention [82], well-documented in the literature under the topic of *attention economy*. While a lot of emphasis, as described in the previous section, (Section 1.1.5) has been given to modeling competition for *content spread* in the "macroscopic" level of the network, the actual influence competition between contents occurs at the "microscopic" level of a single user. Backstrom et al. [83] propose to study the personal network of a user on Facebook, based on how the user divides his attention among his neighbours. They note that there is significant variation among how people divide their attention, with some users distributing it widely among a large group of friends, while a few focus a large part of their attention on a smaller subset of friends. Bo et al. [84] study a social network where users allocate their *attention budget* to their neighbours, in order to maximize their "utility" (measured by the delay in receiving important stories for that particular user). They then study how selfish decisions by the users could affect the information flow efficiency of the network, and propose incentive mechanisms to increase the network's efficiency. Neither of the above papers focuses on how a single end user processes the incoming information. Unless there is a clear understanding of how a single user processes, and thereby gets influenced by, the various contents reaching him, the macroscopic models of influence spread will be insufficient. Thus, in the context of an online social network, modeling the competition at the level of a single user, requires capturing how information, and thus influence, reaches him.

The most prevalent manner in which the incoming contents are displayed to an end user, is in the form of a *timeline*. An interesting feature to observe is that, all three major online social networks (Facebook, Google+, Twitter) use a *timeline* based template in their homepages (see Figure (1.3)). A timeline is a reverse chronologically ordered list of content, displayed in the homepage of the social network. Though very similar to the traditional e-mail inbox view, the ability to place the content (multimedia/hyperlinks) directly on the timeline provides a more engaging experience for the end user. Since users are connected (subscribe/follow/friend) to other users and brands' accounts on these sites, content pushed out by these accounts reach the user timeline. Facebook and Google+

Figure 1.3: Homepage snapshots from Facebook, Twitter, Google+. The timelines are highlighted in bold black lines. The images have been blurred to protect private data.

employ algorithms to sort the timeline and prioritise feeds (according to the learned preferences from past user interactions), while Twitter still provides a reverse chronological view.

Due to the reverse-chronological nature, as new content arrives, the existing contents get pushed further down the timeline. Taking limited user attention, and the limited screen space into consideration, the temporally ordered timeline structure implies that entries in the timeline are seen with less probability the lower down they are in the timeline. Recent research using eye-tracking technology, and studies on user scrolling behavior have confirmed the fact that user attention is focused mostly on the first few entries. In summary, Web users spend 80% of their time looking at information above the page fold. Although users do scroll, they allocate only 20% of their attention below the fold [85] (see Figure (1.4)). Thus when there are several content creators, all aiming to grab the user's attention, it can be viewed as competition between contents for visibility in a user's timeline.

In [35], the authors studied such a competition between *memes* using a model that assumed two time-ordered list of posts (timeline), called the screen and (user's) memory. Their focus was primarily empirical, and aimed to address the prevalent heterogenity in

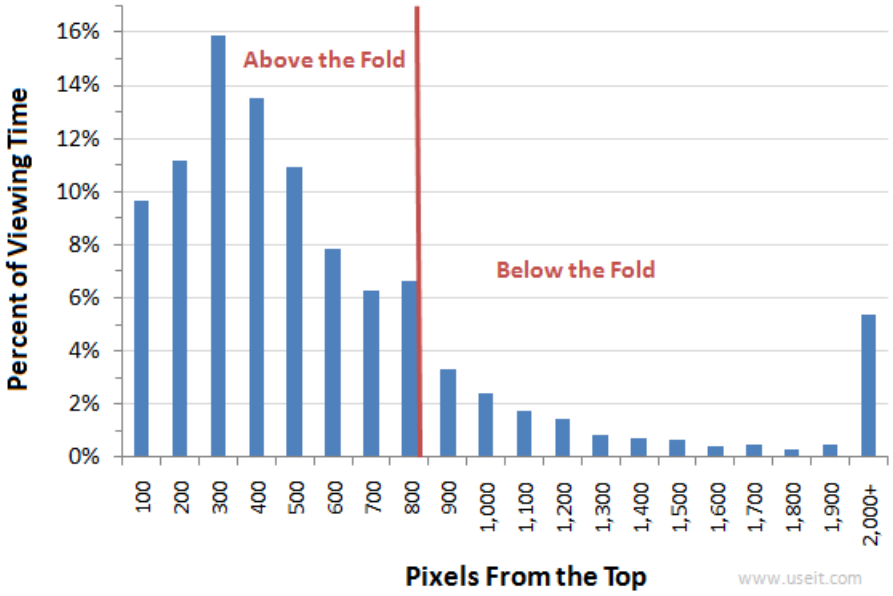


Figure 1.4: Distribution of user fixations on webpages (not restricted to social networks) along stripes that were 100 pixels tall. This justifies our assumption of considering a fixed timeline size. (Figure reproduced from [85])

the popularity (following power law) and persistence of memes on the Internet. Though such timeline based templates have been the norm in social media, very little effort has gone into understanding how the structure and dynamics of the timeline affects content consumption and the evolution of influence. One can view the timeline as a queue with a single server (user), and each of the competing contents as separate classes of customers. While in traditional queueing games literature [86], the customers aim to exit the queue at the earliest, in the context of timeline games, the contents wish to stay in the queue as long as possible (thus being more likely to influence the end user). In the second part of Chapter 5, we introduce the timeline model to study the evolution of influence in such scenarios, and provide some insights into content competition at the single-user level.

1.2 Our Contributions

In this thesis, we contribute towards the analytical understanding of various models of influence dynamics on social networks. The key results can be summarized as follows.

- ***Influence maximization under the LT model:*** We analytically characterize the influence spread under the Linear Threshold (LT) model [1], and provide an equivalent interpretation in terms of *acyclic path probabilities* in a Markov chain. We provide explicit characterization of optimal seed set for influence maximization for some toy networks, such as ring, star and node-degree based models. We also study

the performance of Pagerank [2] under various scenarios, and provide insights into its performance comparison with the greedy algorithm. Finally, using the insights from the analytical characterization, we propose the *G1-sieving* algorithm for the influence maximization problem, and show it performs on par with the greedy algorithm.

- ***Fluid limit of the LT model:*** We then proceed to derive the fluid limits of a homogeneous variant of the LT model, under a general threshold distribution. We note that the threshold distribution features in the fluid limit via its hazard function [87]. We then show that, under the exponential threshold distribution, the LT model becomes equivalent to the well known SIR model from epidemiology. We then proceed to consider the LT model under specific distributions, such as Weibull and loglogistic, and show that qualitatively different influence dynamics can arise due to different threshold distributions. We also provide the fluid limit for a heterogeneous LT model with communities, and provide numerical insights into a variant of the influence maximization problem in the context of communities.

- ***Coevolution of Influence and Content in MONETs:*** We model the evolution of popularity and the spread of content by models inspired from epidemiology. While the former is similar to an uncontrolled SIR (Susceptible-Infected-Recovered) process, the latter is modeled as a controlled SI (Susceptible-Infected) process, thus yielding what we call the *SIR-SI* model. The evolutions are modeled as coupled continuous time Markov chains, and using [88] we derive the fluid limit of the joint evolution process. We then obtain a time-threshold policy while copying to relays, that is delay-cost optimal for the problem of mobile content delivery.

- ***Competition over Multiple Social Networks:*** We model the competition between two content providers across multiple social networks as a non-cooperative game. We model the influence dynamics using epidemic models, and formulate the non-cooperative game in terms of the net budget used by each of the content creators. We obtain the best response functions for each of the users, and thus characterize the Nash equilibria, as the intersection of the best response functions. We also provide interesting remarks on the structure of the best response function and the threshold behavior. We also discuss a variant of the cost structure and discuss its similarities to competition models such as the Cournot duopoly model [89].

- ***Competition over Timelines in Social Networks:*** We study competition between content creators for visibility in a single social network user's timeline. We

begin with a timeline of fixed size, and formulate a non-cooperative game among N content creators, whose utility is the probability of finding their content on the timeline. We derive explicit equilibrium rate expressions under a symmetric cost scenario, and show that it exhibits non-monotonic behavior, with increase in number of players N. We then proceed to model influence evolution for a timeline of infinite size, with a behavioral model for user scanning the timeline. We obtain integral equations, capturing the evolution of influence for a single content creator in isolation, and in the presence of competing contents. We show that the system behaves like an $M/G/1$ queue *with immediate discount*. We then finally study non-cooperative games among the content creators, and provide some interesting insights.

CHAPTER 2

Influence Maximization Under the Linear Threshold Model

Recall (from Section 1.1.2) the influence maximization, where the objective is to identify the optimal initial set of K nodes to maximize the spread of influence in a social network. Though the combinatorial problem was first posed in [7], it was formalized in Kempe et al.'s seminal work [1]. They proposed two major models, namely the Linear Threshold (LT) model and the Independent Cascade (IC) model to capture the underlying influence dynamics (also known as the *activation process*). In the population, nodes that have not yet been influenced are termed as *inactive*. Nodes that *just* got influenced in the previous time step are termed as *active and infectious*, and they remain infectious for one time step, after which they move into the set of *active and non-infectious* nodes. In the IC model, a node which is active and infectious tries to influence each of its inactive neighbours, and succeeds independently in each attempt probabilistically. In the LT model, the activation of a susceptible node depends on the net influence it receives from all its active neighbours (including the non-infectious ones). So, while the IC model can be thought of as a *push* model, the LT model follows a *pull* based approach [45]. In this chapter, we develop upon the Linear Threshold model studied by [1]. Our major contributions are as follows:

- We derive recursive expressions for the expected influence of a given initial set (in Section 2.2), provide an equivalent interpretation via acyclic path probabilities in Markov chains, and provide an alternate proof of submodularity of the objective function (in Sections 2.3 and 2.4).

- In Section 2.5, we use the analytical expressions to derive the optimal initial set in some toy networks such as star, ring, degree-based models etc. We also provide some sample cases where the PageRank algorithm is provably optimal or performs very well and subsequently discuss the limitations of PageRank in more general cases.
- We finally propose the G1-Sieving algorithm to find the most influential set, based on the insights derived from the recursive expression (in Section 2.6) and demonstrate its performance comparison with the greedy algorithm on a coauthorship network dataset (in Section 2.7).

2.1 The Social Network Model

In this section, we briefly introduce the social network model, and the underlying activation model, with the associated notations. We then formally state the optimization problem of influence maximization, and describe the greedy hill-climbing approach discussed in [1].

Glossary of Notation

$\mathcal{N}$ - weighted directed graph of the entire social network
$\mathcal{N}\backslash\mathcal{A}$ - graph obtained by removing nodes in $\mathcal{A} \subseteq \mathcal{N}$ and all links to or from these nodes
$\mathbf{W}$ - influence matrix with $w_{i,j}$ as entries, gives the edge weights of $\mathcal{N}$
Θ_j - $U[0,1]$ random threshold chosen by node j
$b_j(A) = \sum_{i \in A} w_{i,j}$, total influence into node j from set A
$\mathcal{A}_0$ - Initial active set
A_k - Set of all active nodes at time step k, $A_0 \subset A_1 \subset A_2 \dots$
D_k - Set of nodes which were activated at time step k, $D_k = A_k \backslash A_{k-1}$
U - Random time at which the activation process stops, $U = \min_k\{A_k = A_{k-1}\}$
$g_j^{(\mathcal{N},\mathcal{A})}(k) = \mathbb{P}^{(\mathcal{N},\mathcal{A})}(j \in D_k) = \mathbb{P}^{(\mathcal{N})}(j \in D_k \big| \mathcal{A}_0 = \mathcal{A})$
$g_j^{(\mathcal{N},\mathcal{A})} = \mathbb{P}^{(\mathcal{N},\mathcal{A})}(j \in A_U)$
$\sigma^{(\mathcal{N},\mathcal{A})} = \mathbb{E}^{(\mathcal{N},\mathcal{A})}[|A_U|]$

Social Network Description Consider a social network is a weighted directed graph $\mathcal{N} = (V, E)$, where the edge weights $w_{i,j}$ give a measure of influence of node i on node j. There have been efforts [90] to learn these influence probabilities from the observed contagion traces on the social network. In this thesis, we assume that these weights $w_{i,j}$s

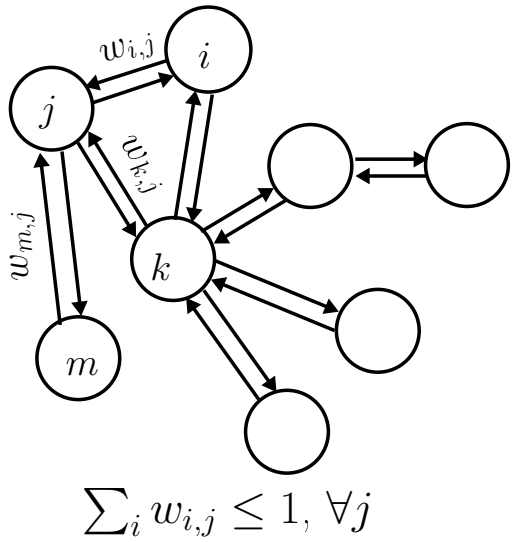


Figure 2.1: An influence graph of a social network under the Linear Threshold model.

are known. The activation process begins with an initial set of active nodes $\mathcal{A}_0$, and at each step k the set of active nodes A_k keeps increasing, due to the influence of the already active nodes. This goes on until a *terminal set* A_U is reached, from where the activation process cannot proceed further. We shall focus only on the *progressive case*, where nodes once activated, will never switch back to the inactive state.

Activation Models There are two widely used activation models, namely, *Linear Threshold model* and *Independent Cascade model*. According to the original formulation of the Linear Threshold model [1], the incoming edge weights to a node are normalized such that $\sum_{i \neq j} w_{i,j} \leq 1$. In this model, each node j randomly chooses a threshold Θ_j uniformly from [0,1] *at the beginning*. The threshold denotes the susceptance level of a node, i.e., the net amount of influence that is necessary to activate the node. At step k, a node j gets activated if, it had been inactive until step $k-1$ and

$$\sum_{i \in A_{k-1}} w_{i,j} \geq \Theta_j$$

In the *Independent Cascade model*, we start with an initial active set $\mathcal{A}_0$ and the activation proceeds according to the following randomized rule. Whenever a node i becomes active at step k, it is given one attempt at activating each of its inactive neighbours j, succeeding with probability $w_{i,j}$. If i succeeds, j becomes part of A_{k+1}, but whether or not i succeeds, it cannot make any more attempts at activating its neighbours in subsequent rounds. Again, the activation process continues until no more activations are possible. Kempe et

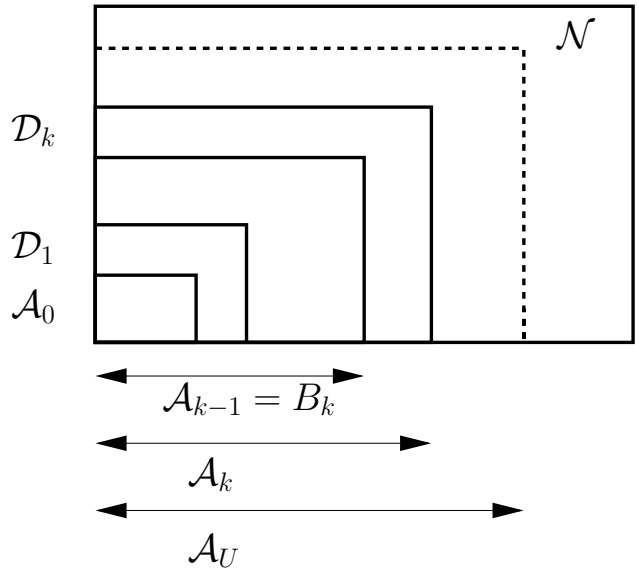


Figure 2.2: Evolution of the set of influenced nodes under the Linear Threshold model.

al. also provide generalizations of the above two models in [1], and show how the two generalized models are essentially equivalent. In the remaining sections of this thesis, we shall be discussing only the Linear Threshold model.

Problem Statement Given the initial set $\mathcal{A}_0$, the activation process evolves in discrete time steps according to the Linear Threshold model. Let A_k denote the set of all active nodes at time k. Since we are dealing with the progressive case, $A_0 \subset A_1 \subset \ldots \subseteq \mathcal{N}$. Let D_k denote the set of nodes which were activated at time k, i.e., $D_k = A_k \backslash A_{k-1}$ and $D_0 = \mathcal{A}_0$. Let U denote the random stopping time at which the activation process stops, i.e., $U = min_k\{A_k = A_{k-1}\}$. Define $\sigma^{(\mathcal{N},\mathcal{A}_0)} = \mathbb{E}^{(\mathcal{N},\mathcal{A}_0)}[|A_U|]$ to be the expected size of the *terminal set* A_U, starting with $\mathcal{A}_0$ as the initial set in the network $\mathcal{N}$. The influence maximization problem can then be formulated as follows:

$$\max \sigma^{(\mathcal{N},\mathcal{A}_0)} \tag{2.1}$$

$$\text{s.t. } \mathcal{A}_0 \subset \mathcal{N}$$

$$|\mathcal{A}_0| = K$$

Greedy Algorithm The greedy hill climbing solution for the influence maximization problem is shown in Algorithm 1. In the algorithm, K is the size of the required initial set, and the set X obtained after the K iterations is the greedy solution. It is noted in [1] that this achieves an approximation factor of $(1 - 1/e)$, and the proof involves the

submodularity (i.e., diminishing returns) and monotonicity of $\sigma^{(\mathcal{N},\mathcal{A})}$.

$X \leftarrow \emptyset$;
for $i = 1$ *to* K **do**
  Choose v_i such that $v_i = \arg\max_v \sigma^{(\mathcal{N},X\cup v)}$;
  $X \leftarrow X \cup v_i$;
end

Algorithm 1: Greedy Algorithm

2.2 Recursive Expression for $\sigma^{(\mathcal{N},\mathcal{A}_0)}$

As far as we know, there is no work on mathematically characterising the value of $\sigma^{(\mathcal{N},\mathcal{A}_0)}$ for the models introduced in [1]. Moreover, $\sigma^{(\mathcal{N},\mathcal{A}_0)}$ is generally obtained by simulating the activation process several times on the social network, and taking the average value. In this section, we derive an expression for $\sigma^{(\mathcal{N},i)}$ in recursive form, and hence give a general expression for $\sigma^{(\mathcal{N},\mathcal{A}_0)}$. We use this expression later to provide insights into various existing heuristics, and also for proposing an efficient algorithm that matches the greedy solution. Let us begin with the definition of $\sigma^{(\mathcal{N},\mathcal{A}_0)}$.

$$\sigma^{(\mathcal{N},\mathcal{A}_0)} = \mathbb{E}^{(\mathcal{N},\mathcal{A}_0)}[|A_U|]$$

Note that, since D_k's are disjoint, and $\bigcup_{k=0}^{\infty} D_k = A_U$, we can write,

$$\begin{aligned}
\sigma^{(\mathcal{N},\mathcal{A}_0)} &= \sum_{k=0}^{|\mathcal{N}|} \mathbb{E}^{(\mathcal{N},\mathcal{A}_0)}[|D_k|] \\
&= \sum_{k=0}^{|\mathcal{N}|} \mathbb{E}^{(\mathcal{N},\mathcal{A}_0)}[\sum_{j\in\mathcal{N}} I_{\{j\in D_k\}}] \\
&= \sum_{k=0}^{|\mathcal{N}|} \sum_{j\in\mathcal{N}} \mathbb{E}^{(\mathcal{N},\mathcal{A}_0)}[I_{\{j\in D_k\}}] \\
&= \sum_{k=0}^{|\mathcal{N}|} \sum_{j\in\mathcal{N}} g_j^{(\mathcal{N},\mathcal{A}_0)}(k)
\end{aligned}$$

In the above expressions, $I_{\{E\}}$ denotes the indicator variable for the event E, and we also use the fact that the total number of time steps of the activation process is bounded above by the number of nodes in the network, $|\mathcal{N}|$. Recall from the glossary of notations in

Section 2.1, $g_j^{(\mathcal{N},\mathcal{A}_0)}(k)$ gives the probability that node j is activated at the time step k, given that we start with $\mathcal{A}_0$ as the initial set in the network $\mathcal{N}$. We wish to state the following lemma, which will help us determine $g_j^{(\mathcal{N},\mathcal{A}_0)}(k)$.

Lemma 2.2.1. 1. $j \in \mathcal{A}_0$,

(a) $g_j^{(\mathcal{N},\mathcal{A}_0)}(0) = 1$

(b) $g_j^{(\mathcal{N},\mathcal{A}_0)}(k) = 0$, for all $k > 0$

2. $j \notin \mathcal{A}_0$,

(a) $g_j^{(\mathcal{N},\mathcal{A}_0)}(0) = 0$

(b) $g_j^{(\mathcal{N},\mathcal{A}_0)}(k) = \sum_{l \in \mathcal{N}\setminus\{j\}} g_l^{(\mathcal{N}\setminus\{j\},\mathcal{A}_0)}(k-1)\ w_{l,j}$, for all $k > 0$

Proof. Note that 1(a) and 2(a) are obvious, since $D_0 = \mathcal{A}_0$, chosen deterministically. 1(b) follows from 1(a) and the observation that $\sum_{k=0}^{\infty} g_j^{(\mathcal{N},\mathcal{A}_0)}(k) \leq 1$ by definition. For 2(b), since $\mathcal{A}_0 \subset \mathcal{N}\setminus\{j\}$,

$$g_j^{(\mathcal{N},\mathcal{A}_0)}(k) = \mathbb{P}^{(\mathcal{N}\setminus\{j\},\mathcal{A}_0)}\Big(b_j(A_{k-2}) < \Theta_j \leq b_j(A_{k-1})\Big)$$

Since $D_{k-1} = A_{k-1}\setminus A_{k-2}$, and Θ_j is chosen uniformly from [0,1], we can write,

$$\begin{aligned} g_j^{(\mathcal{N},\mathcal{A}_0)}(k) &= \mathbb{E}^{(\mathcal{N}\setminus\{j\},\mathcal{A}_0)}[b_j(D_{k-1})] \\ &= \mathbb{E}^{(\mathcal{N}\setminus\{j\},\mathcal{A}_0)}\Big[\sum_{l \in \mathcal{N}\setminus\{j\}} I_{\{l\in D_{k-1}\}} w_{l,j}\Big] \\ &= \sum_{l \in \mathcal{N}\setminus\{j\}} g_l^{(\mathcal{N}\setminus\{j\},\mathcal{A}_0)}(k-1)\ w_{l,j} \end{aligned}$$

□

2.2.1 Singleton Initial Set

Now, by Lemma 2.2.1, we can write, For $j \neq i$,

$$g_j^{(\mathcal{N},i)}(1) = w_{i,j}$$

$$g_j^{(\mathcal{N},i)}(2) = \sum_{k_1 \in \mathcal{N}\setminus\{j\}} g_{k_1}^{(\mathcal{N}\setminus\{j\},i)}(1) w_{k_1,j}$$

$$= \sum_{k_1 \in \mathcal{N}\setminus\{i,j\}} w_{i,k_1} w_{k_1,j}$$

$$\begin{aligned}
& g_j^{(\mathcal{N},i)}(3) \\
&= \sum_{k_1 \in \mathcal{N}\setminus\{j\}} g_{k_1}^{(\mathcal{N}\setminus\{j\},i)}(2) w_{k_1,j} \\
&= \sum_{k_1 \in \mathcal{N}\setminus\{i,j\}} g_{k_1}^{(\mathcal{N}\setminus\{j\},i)}(2) w_{k_1,j} \\
&= \sum_{k_1 \in \mathcal{N}\setminus\{i,j\}} \sum_{k_2 \in \mathcal{N}\setminus\{j,k_1\}} g_{k_2}^{(\mathcal{N}\setminus\{j,k_1\},i)}(1) w_{k_2,k_1} w_{k_1,j} \\
&= \sum_{k_1 \in \mathcal{N}\setminus\{i,j\}} \sum_{k_2 \in \mathcal{N}\setminus\{i,j,k_1\}} w_{i,k_2} w_{k_2,k_1} w_{k_1,j} \\
&= \sum_{k_1 \in \mathcal{N}\setminus\{i,j\}} \sum_{k_2 \in \mathcal{N}\setminus\{i,j,k_1\}} w_{i,k_1} w_{k_1,k_2} w_{k_2,j}
\end{aligned}$$

In the each of the steps, we substitute the expression for $g_j^{(\mathcal{N},i)}(k)$ and noting that $g_i^{(\mathcal{N},i)}(k) = 0$ for $k > 0$. The last step is obtained by suitably rearranging the terms.

Note that, the above terms can be understood, as the influence of node i reaching node j through a path (without loops) of k hops. We can use this to derive the recursive equation for $\sigma^{(\mathcal{N},i)}$.

Theorem 2.2.2. Given a social network $\mathcal{N}$, with influence matrix $\mathbf{W}$, the total influence of any node i in the network under the LT model is given by

$$\sigma^{(\mathcal{N},i)} = 1 + \sum_{j \in \mathcal{N}\setminus\{i\}} w_{i,j} \sigma^{(\mathcal{N}\setminus\{i\},j)} \tag{2.2}$$

*Remark:*According to the above theorem, under LT model, the total influence of any node i in the network, is one (for the node i itself) plus the weighted sum of the influences of its neighbours in the network without i. It is interesting to see that (Figure 2.3), this allows us to decompose the problem into those involving i's neighbours. It is to be noted that the uniform distribution of the activation threshold is critical in allowing us to add up all influences exerted by i through its different neighbours.

Proof. We have,

$$\sigma^{(\mathcal{N},i)}$$

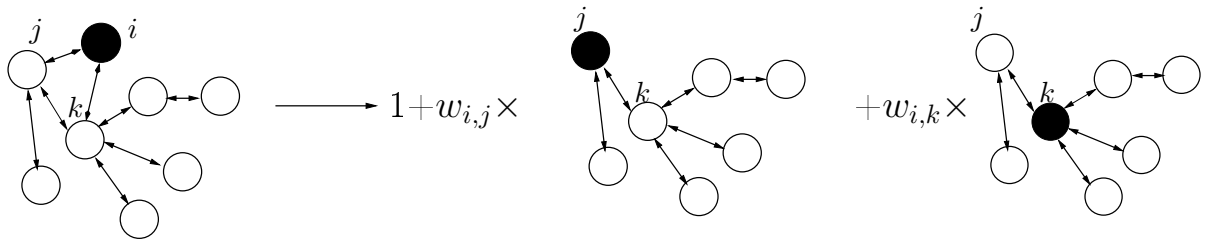


Figure 2.3: Influence of a single node evaluated through its neighbours (Theorem 2.2.2)

$$
\begin{aligned}
&= \sum_{k=0}^{\infty} \sum_{j \in \mathcal{N}} g_j^{(\mathcal{N},i)}(k) \\
&= 1 + \sum_{j \in \mathcal{N} \setminus \{i\}} g_j^{(\mathcal{N},i)}(1) + \sum_{j \in \mathcal{N} \setminus \{i\}} g_j^{(\mathcal{N},i)}(2) + \cdots \\
&= 1 + \sum_{j \in \mathcal{N} \setminus \{i\}} w_{i,j} + \sum_{j \in \mathcal{N} \setminus \{i\}} \sum_{k_1 \in \mathcal{N} \setminus \{i,j\}} w_{i,k_1} w_{k_1,j} + \cdots
\end{aligned}
$$

By changing variables and rearranging summations, this is equivalent to

$$
\sigma^{(\mathcal{N},i)} = 1 + \sum_{k_1 \in \mathcal{N} \setminus \{i\}} w_{i,k_1} \Bigg[1 + \sum_{k_2 \in \mathcal{N} \setminus \{i,k_1\}} w_{k_1,k_2} \Bigg[\cdots
$$

Note that this equation is recursive in nature, and hence we get the above theorem. □

2.2.2 Initial Set $\mathcal{A}_0$

A similar derivation can be done for any $\mathcal{A}_0$. In this case, again using Lemma 2.2.1, we get

$$
\begin{aligned}
g_j^{(\mathcal{N},\mathcal{A}_0)}(1) &= \sum_{i \in \mathcal{A}_0} w_{i,j} \\
g_j^{(\mathcal{N},\mathcal{A}_0)}(2) &= \sum_{i \in \mathcal{A}_0} \sum_{k_1 \neq j \;\; k_1 \notin \mathcal{A}_0} w_{i,k_1} w_{k_1,j} \\
g_j^{(\mathcal{N},\mathcal{A}_0)}(3) &= \sum_{i \in \mathcal{A}_0} \sum_{k_1 \neq j \;\; k_1 \notin \mathcal{A}_0} \sum_{k_2 \neq j, k_1 \;\; k_2 \notin \mathcal{A}_0} w_{i,k_1} w_{k_1,k_2} w_{k_2,j}
\end{aligned}
$$

and so on. Note that these terms can be understood, as the influence of nodes $i \in \mathcal{A}_0$ reaching node j through a path (without loops) of k steps, *without passing through any other node in* $\mathcal{A}_0$.

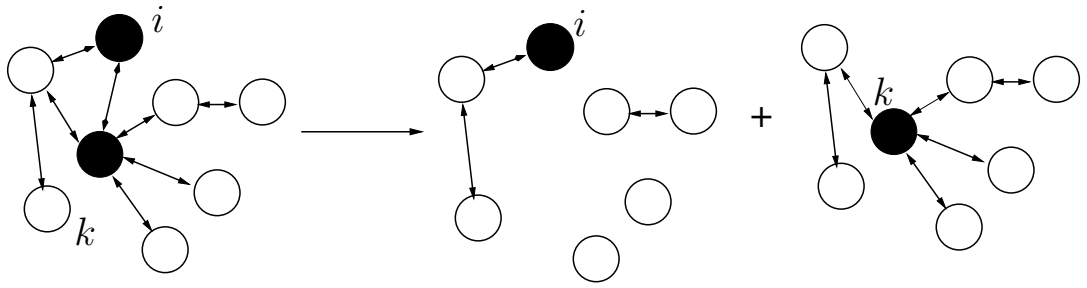


Figure 2.4: Evaluating set influences through individual influences (Theorem 2.2.3)

Also, having chosen $\mathcal{A}_0$, the edge weights $\{w_{i,j}, j \in \mathcal{A}_0\}$ do not have any effect on $\sigma^{(\mathcal{N},\mathcal{A}_0)}$. By the above two observations, we can thus divide the problem of finding $\sigma^{(\mathcal{N},\mathcal{A}_0)}$ into K subproblems, where $K = |\mathcal{A}_0|$. Define sub-networks $\mathcal{N}_i^{\mathcal{A}_0}$, for all $i \in \mathcal{A}_0$, such that,

$$\mathcal{N}_i^{\mathcal{A}_0} = \{\mathcal{N} \backslash \mathcal{A}_0\} \cup \{i\}$$

Then we can see that,

$$g_j^{(\mathcal{N},\mathcal{A}_0)}(k) = \sum_{i \in \mathcal{A}_0} g_j^{(\mathcal{N}_i^{\mathcal{A}_0}, i)}(k)$$

Now we can state the theorem as follows.

Theorem 2.2.3. Given a social network $\mathcal{N}$ with influence matrix $\mathbf{W}$ and an initial set $\mathcal{A}_0$, define sub-networks $\mathcal{N}_i^{\mathcal{A}_0}$, for all $i \in \mathcal{A}_0$, such that,

$$\mathcal{N}_i^{\mathcal{A}_0} = \{\mathcal{N} \backslash \mathcal{A}_0\} \cup \{i\}$$

Then the influence of the initial set $\mathcal{A}_0$ is given by,

$$\sigma^{(\mathcal{N},\mathcal{A}_0)} = \sum_{i \in \mathcal{A}_0} \sigma^{(\mathcal{N}_i^{\mathcal{A}_0}, i)} \tag{2.3}$$

Remark: It is interesting to see that some nodes might be left with a very small part of the network to directly influence. In Figure 2.4, by including k in the initial set, we have essentially restricted i's spread of influence to only two other nodes. To reach any other node, the influence has to pass through k, which is already included in the initial set.

Proof. The proof follows by substituting the $g_j^{(\mathcal{N},\mathcal{A}_0)}(k)$ expressions and noting that the edge weights $\{w_{i,j}, j \in \mathcal{A}_0\}$ do not have any effect on $\sigma^{(\mathcal{N},\mathcal{A}_0)}$. This allows us to split the problem of evaluating influences into K sub-problems each involving only one node from $\mathcal{A}_0$. □

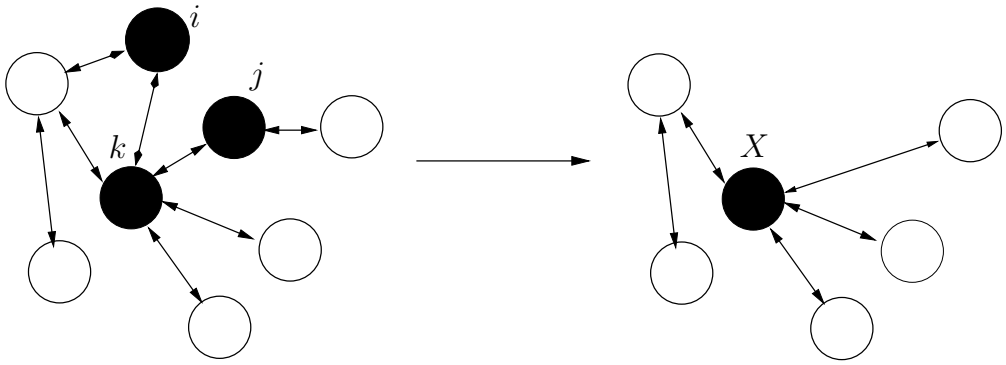


Figure 2.5: Replacing the initial set with a supernode (Theorem 2.2.4)

2.2.3 Replacing an initial set with a supernode

Theorem 2.2.4. Given a social network $\mathcal{N}$, with influence matrix $\mathbf{W}$, a given initial set $\mathcal{A}_0$ can be replaced by a supernode X where,

$$w_{X,v} = \sum_{i \in X} w_{i,v}$$

$$w_{v,X} = \sum_{i \in X} w_{v,i}$$

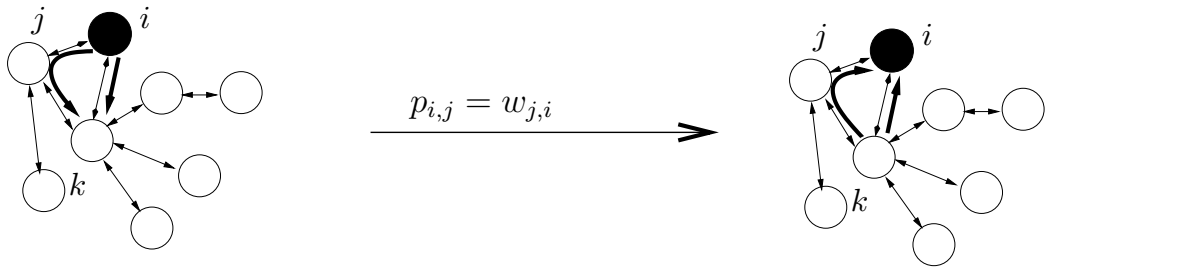
and, $\sigma^{(\mathcal{N},\mathcal{A}_0)}$ is equal to the influence of X in the modified network, with X being counted as $|X|$ nodes instead of 1.

Remark: The theorem allows us to reduce the initial problem of evaluating influence of a set into a problem of evaluating a node's influence(see Figure 2.5). This is possible because of the uniform distribution of the activation threshold and the fact that influence spreads in an acyclic manner branching out from the initial set.

Proof. The proof follows directly by using Theorem 2.2.2 for each sub-network in Theorem 2.2.3 and combining the common terms. □

2.3 Interpretation via Acyclic Path Probabilities in a DTMC

To begin with, since $\sum_{i \neq j} w_{i,j} \leq 1$, $\mathbf{W}$ in general need not be a stochastic matrix. In order to interpret the expressions for $g_j^{(\mathcal{N},\mathcal{A}_0)}$ in the Discrete Time Markov chains(DTMC) framework, we require $\mathbf{P} = \mathbf{W}^T$ to be row stochastic. Hence, we can set $w_{j,j} = 1 - \sum_{i \neq j} w_{i,j}$. Note that this does not affect the theory developed till now, since terms of



$g_k^{(\mathcal{N},i)} = w_{i,k} + w_{i,j} \times w_{j,k}$ $\qquad$ $c(k \to i) = p_{k,i} + p_{k,j} \times p_{j,i}$

Social Network $\qquad$ Equivalent Markov Chain

Figure 2.6: Equivalence between the social network under the LT model and Markov chains

the form $w_{i,i}$ do not feature in Equations 2.2 and 2.3. We shall also adopt a similar approach when we use the PageRank algorithm, where we will be calculating the stationary probability of a DTMC. From Lemma 2.2.1, for all $j \neq i$,

$$g_j^{(\mathcal{N},i)}(k) = \sum_{l_1 \neq i,j} \sum_{l_2 \neq i,j,l_1} \cdots \sum_{l_{k-1} \neq i,j,l_1,l_2,\ldots l_{k-2}} w_{i,l_1} w_{l_1,l_2} \ldots w_{l_{k-1},j}$$

Writing $\mathbf{P} = \mathbf{W}^T$ and interpreting $\mathbf{P}$ as a transition probability matrix for the DTMC $\{X_k\}$, obtained by reversing the edges in the social network, we have,

$$\begin{aligned} g_j^{(\mathcal{N},i)}(k) &= \sum_{l_1 \neq i,j} \sum_{l_2 \neq i,j,l_1} \cdots \sum_{l_{k-1} \neq i,j,l_1,l_2,\ldots l_{k-2}} p_{j,l_1} p_{l_1,l_2} \ldots p_{l_{k-1},i} \\ &= \mathbb{P}(\{X_m \in \mathcal{N} \backslash \{i\}, 0 \leq m < k, X_k = i, X_0 \neq X_1 \neq \cdots \neq X_k\} \big| X_0 = j) \\ &=: c_k(j \to i) \end{aligned}$$

and similarly,

$$\begin{aligned} g_j^{(\mathcal{N},\mathcal{A}_0)}(k) &= \mathbb{P}(\{X_m \in \mathcal{N} \backslash \mathcal{A}_0, 0 \leq m < k, X_k \in \mathcal{A}_0, X_0 \neq X_1 \neq \cdots \neq X_k\} \big| X_0 = j) \\ &=: c_k(j \to \mathcal{A}_0) \end{aligned}$$

Define,

$$c(j \to i) = \sum_{k=0}^{\infty} c_k(j \to i)$$

$$c(j \to \mathcal{A}_0) = \sum_{k=0}^{\infty} c_k(j \to \mathcal{A}_0)$$

In the DTMC represented by $\mathbb{P}$, given we start from state j, $c(j \to i)$ denotes the probability of hitting state i for the first time through an acyclic path from j, and $c(j \to \mathcal{A}_0)$ denotes the probability of hitting the set $\mathcal{A}_0$ for the first time through an acyclic path from j. Since $g_j^{(\mathcal{N},\mathcal{A}_0)} = c(j \to \mathcal{A}_0)$ we have,

$$\sigma^{(\mathcal{N},\mathcal{A}_0)} = |\mathcal{A}_0| + \sum_{j \notin \mathcal{A}_0} c(j \to \mathcal{A}_0) \tag{2.4}$$

Now the influence maximization problem can be restated as follows:
Let $\mathbf{W}$ *denote the influence matrix for the given problem. For the DTMC over the finite state space* $\mathcal{N}$*, with transition probability matrix* $\mathbf{P} = \mathbf{W}^T$*, choose* $\mathcal{A} \subset \mathcal{N}, |\mathcal{A}| = K$*, such that* $\sum_{j \in \mathcal{A}^c} c(j \to \mathcal{A})$ *is maximized. The set* $\mathcal{A}$ *thus obtained is the solution to the original influence maximization problem.*

Remark: Observe that computation of influence thus requires enumeration of all acyclic paths from the initial set to the remaining nodes. While powers of the adjacency matrix can be used to compute *all* paths (similar to how the Pagerank algorithm [91] works), there is no known efficient technique to compute acyclic paths. However, the above expression will help us arrive at efficient heuristics (as discussed in Section 2.6) to solve the influence maximization problem.

2.4 Submodularity of $\sigma^{(\mathcal{N},\mathcal{A}_0)}$

In [1], the authors prove that the objective function for influence maximization is submodular and monotone, and use it to establish the $(1 - \frac{1}{e})$ approximation guarantee of the greedy hill-climbing algorithm. In the previous sections, we have demonstrated the equivalence between computing acyclic path probabilities in Markov chains and the problem of influence maximization. In Section 2.9.1, we list and prove certain properties of acyclic path probabilities. Using them, in this section we shall prove the submodularity and monotonicity of acyclic path probabilities, and use it to provide an alternative proof for monotonicity and submodularity of $\sigma^{(\mathcal{N},\mathcal{A}_0)}$.

Lemma 2.4.1. $c(j \to \mathcal{A})$ is monotonically increasing in $\mathcal{A}$. i.e., For $\mathcal{A} \subseteq \mathcal{B}$, $c(j \to \mathcal{A}) \leq c(j \to \mathcal{B})$

Proof. We need to check,

$$c(j \to \mathcal{A} \cup \{v\}) - c(j \to \mathcal{A}) \stackrel{?}{\geq} 0$$

Substituting for $c(j \to \mathcal{A} \cup \{v\})$ and $c(j \to \mathcal{A})$ from Lemmas 2.9.2 and 2.9.3, we have

$$\begin{aligned} c(j \to \mathcal{A} \cup \{v\}) - c(j \to \mathcal{A}) &= c^{\mathcal{N}\setminus\mathcal{A}}(j \to v) - c(j \xrightarrow{v} \mathcal{A}) \\ &\geq 0 \end{aligned}$$

The last inequality results from Lemma 2.9.4. □

Lemma 2.4.2. $c(j \to \mathcal{A})$ is submodular in $\mathcal{A}$, i.e., $c(j \to \mathcal{A} \cup \{v\}) - c(j \to \mathcal{A})$ decreases, as $\mathcal{A}$ increases.

Proof.

$$\begin{aligned} c(j \to \mathcal{A} \cup \{v\}) - c(j \to \mathcal{A}) &= c^{\mathcal{N}\setminus\mathcal{A}}(j \to v) - c(j \xrightarrow{v} \mathcal{A}) \\ &= \sum_{L' \subseteq \mathcal{N}-(\mathcal{A}\cup\{j,v\})} \sum_{L \in \Pi(L')} p_{j,v}(L) - \\ &\quad \sum_{L' \subseteq \mathcal{N}-(\mathcal{A}\cup\{j,v\})} \sum_{L \in \Pi(L')} p_{j,v}(L) c^{\mathcal{N}\setminus(L'\cup\{j\})}(v \to \mathcal{A}) \\ &= \sum_{L' \subseteq \mathcal{N}-(\mathcal{A}\cup\{j,v\})} \sum_{L \in \Pi(L')} p_{j,v}(L) \Big[1 - c^{\mathcal{N}\setminus(L'\cup\{j\})}(v \to \mathcal{A})\Big] \end{aligned}$$

For a fixed L', as $\mathcal{A}$ increases, $c^{\mathcal{N}\setminus(L'\cup\{j\})}(v \to \mathcal{A})$ increases by Lemma 2.4.1, and hence the entire term inside the summation decreases, as $\mathcal{A}$ increases. Also, as $\mathcal{A}$ increases, the number of L' that satisfy the constraint in the first summation also decrease. Hence, $c(j \to \mathcal{A} \cup \{v\}) - c(j \to \mathcal{A})$ decreases, as $\mathcal{A}$ increases. Thus $c(j \to \mathcal{A})$ is submodular. □

2.4.1 Monotonicity and Submodularity of $\sigma^{(\mathcal{N},\mathcal{A}_0)}$

Recalling Equation 2.4,

$$\sigma^{(\mathcal{N},\mathcal{A}_0)} = |\mathcal{A}_0| + \sum_{j \notin \mathcal{A}_0} c(j \to \mathcal{A}_0)$$

Since $c(j \to \mathcal{A}) = 1$, for all $j \in \mathcal{A}$ by Lemma 2.9.1 we can write

$$\sigma^{(\mathcal{N},\mathcal{A}_0)} = \sum_{j \in \mathcal{N}} c(j \to \mathcal{A}_0)$$

$c(j \to \mathcal{A}_0)$ is monotonically increasing and submodular in A_0 by Lemmas 2.4.1 and 2.4.2 and since $\sigma^{(\mathcal{N},\mathcal{A}_0)}$ is a non-negative linear combination of such $c(j \to \mathcal{A}_0)$, this automatically proves the monotonicity and submodularity of $\sigma^{(\mathcal{N},\mathcal{A}_0)}$.

2.5 Examples

In this section, we will use the $\sigma^{(\mathcal{N},\mathcal{A}_0)}$ expressions on some simple networks, and obtain insights into the spread of influence and the optimal initial set in such networks. We also use the expression to obtain the optimal initial set for some uniform versions of the LT models. We also show that PageRank matches with the optimal solution obtained from the analytical expression in two of those cases. We provide simulation results for those cases in which PageRank is not optimal, but provides a very good approximation of the greedy solution.

2.5.1 Ring Topology

Consider a ring topology $\mathcal{N}$ with N homogeneous nodes(Figure 2.8). Such a topology could arise in proximity based social networks. Let $\alpha \leq 0.5$ denote the influence of any node on each of its two neighbours. In such a network, let $A(K)$ be a set of K nodes chosen, and denote the indices of the nodes in the ring by $(a_1, a_2, \ldots a_K)$. Let $(l_1, l_2, \ldots, l_K)$ be the number of nodes in between the chosen nodes in the ring, i.e. $l_1 = a_2 - a_1 - 1$, $l_2 = a_3 - a_2 - 1$ and so on, with $\sum_{i=1}^{K} l_i = N - K$. Then we can write the influence of $A(K)$ as,

$$\sigma^{(\mathcal{N},A(K))} = K + 2\frac{\alpha}{1-\alpha}\big(K - \sum_{i=1}^{K} \alpha^{l_i}\big)$$

Note that this expression is maximized only if all l_i's are equal. For simplicity if we assume that K divides N, then this means that in the optimal set, K nodes are equally separated along the ring. Also note that with $\alpha = 0.5$, for large N and small K, the influence of $A(K)$ grows as $3K$. This result is not evident from the way the network is constructed and arises directly from the analytical expression. This can be explained in a more general case, as we show in the Theorem 2.5.1.

2.5.2 Star Topology

Let us consider the star topology $\mathcal{N}$ with N nodes including the hub (see Figure 2.7). Let α be the influence exerted by the hub node on each of the peripheral nodes and let β be the influence of any one peripheral node on the hub. By LT model, $\alpha \leq 1$ and $\beta \leq \frac{1}{N-1}$. Such a setting is possible in authorization based social networks, where any data transfer between two nodes, has to pass through a central authority. Given K as the size of initial set, we can either have one hub node and $K-1$ peripheral nodes, or K peripheral nodes.

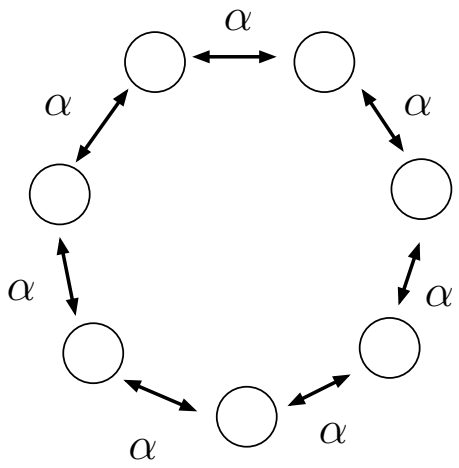


Figure 2.7: Ring topology

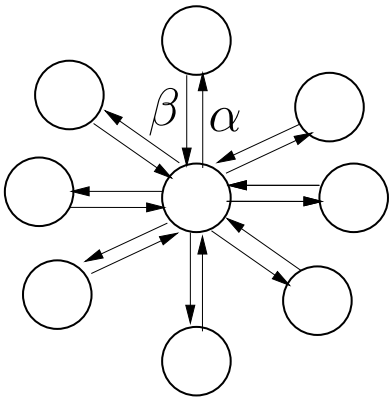


Figure 2.8: Star topology

Call the former set $H(K)$ and the latter $\tilde{H}(K)$. The influence functions can be readily written as follows using Theorem 2.2.4.

$$\sigma^{(\mathcal{N},H(K))} = K + \alpha(N-K)$$

$$\sigma^{(\mathcal{N},\tilde{H}(K))} = K + K\beta\big(1+\alpha(N-K-1)\big)$$

Given K and the values of α and β, one might be interested in knowing which of $H(K)$ or $\tilde{H}(K)$ is optimal, i.e. whether the hub node belongs to the optimal initial set. We see that there exists $\alpha^\star$, such that $\tilde{H}(K)$ is more *influential* than $H(K)$ for $\alpha < \alpha^\star$.

$$\alpha^\star = \frac{K}{(N-K)\frac{1}{\beta} - K(N-K-1)}$$

It is interesting to see that, there are cases where it would be advantageous to leave out the hub node from the initial set, in order to maximize the spread. This might seem counter-intuitive since any message has to pass through the hub node. But, to see why α has to be sufficiently large, consider $\beta = \frac{1}{N-1}$ and $K = N-1$. Then by picking K peripheral nodes we get influence of exactly N, whereas if $\alpha < 1$,then picking $K-1$ peripheral nodes and the hub node will give us an influence strictly less than N.

2.5.3 Node Degree based Model

We now look at a specific set of models, where the edge weight depends on the degree of the destination node. In such a case, each node weighs all its neighbours equally, and

thus the threshold is compared against the *local fraction* of active nodes, i.e., fraction of neighbours who are active. In this class of models, we start with an undirected graph without self-loops, whose adjacency matrix is given by A. Define W as follows.

$$w_{i,j} = a_{i,j}/d_j \tag{2.5}$$

where $d_j = \sum_i a_{i,j}$ is the degree of the node j. Let us restrict our attention to *acyclic graphs*. We then have the following theorem.

Theorem 2.5.1. Consider an acyclic undirected graph $\mathcal{N}$ represented by the adjacency matrix A. Let the influence matrix be generated by Equation (2.5). Then, for any node $i \in \mathcal{N}$,

$$\sigma^{(\mathcal{N},i)} = d_i + 1 \tag{2.6}$$

Proof. Given the acyclic graph $\mathcal{N}$ and the node i, view the graph as a tree $\mathcal{T}$ of depth D, with node i as the root. For any node j in the tree $\mathcal{T}$, let $P(j)$ be the parent of node j in $\mathcal{T}$, and $C(j)$ be its immediate child nodes. Define,

$$L_0 = \{j \in \mathcal{T} : C(j) = \emptyset\}\Big(\neq \emptyset \Big)$$

and for $0 < k \leq D$,

$$L_k = \{j \in \mathcal{T} : C(j) \subseteq \bigcup_{t=0}^{k-1} L_t, C(j) \cap L_{k-1} \neq \emptyset\}$$

Hence by definition of depth D we have, $L_D = \{i\}$ and it is easy to see that L_k's partition nodes in $\mathcal{N}$ into sets of nodes having the same depth. By Equation 2.2 we have,

$$\sigma^{(\mathcal{N},i)} = 1 + \sum_{j \in C(i)} \frac{1}{|C(j)| + 1} \sigma^{(\mathcal{N} \setminus P(j), j)} \tag{2.7}$$

where $P(j) = i$ and $j \in L_0 \cup \cdots \cup L_{D-1}$. We shall prove inductively that, $\forall j$,

$$\sigma^{(\mathcal{N} \setminus P(j), j)} = |C(j)| + 1 \tag{2.8}$$

We know that if $j \in L_0$, then it is true, since in that case $\sigma^{(\mathcal{N} \setminus P(j), j)} = 1$ and $C(j) = 0$.

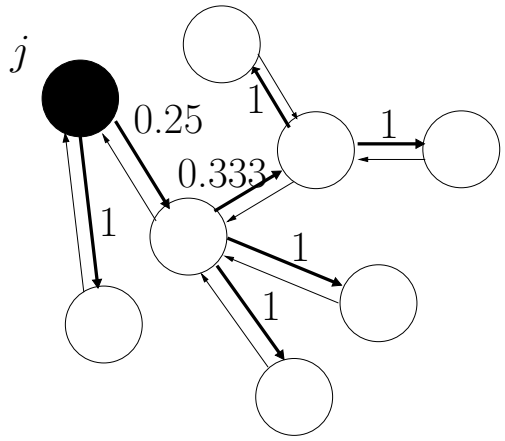


Figure 2.9: Node degree based model. Example network for Theorem 2.5.1. $\sigma^{(\mathcal{N},j)} = 1 + 1 + 0.25(1 + 1 + 1 + 0.333(1 + 1 + 1)) = 3 = d_j + 1$

Assume that the claim is valid for $j \in L_0, L_1, \cdots, L_k$. For $j \in L_{k+1}$,

$$\begin{aligned}
\sigma^{(\mathcal{N}\setminus P(j),j)} &= 1 + \sum_{l \in C(j)} \frac{1}{|C(l)| + 1} \sigma^{(\mathcal{N}\setminus P(l),l)} \\
&= 1 + \sum_{l \in C(j)} \frac{|C(l)| + 1}{|C(l)| + 1} \\
&= 1 + |C(j)|
\end{aligned}$$

The second equality in the above set of equations, is because Claim 2.8 is valid for $l \in \cup_{m=0}^{k} L_m$. Thus substituting the above in Equation 2.7, we have

$$\sigma^{(\mathcal{N},i)} = |C(i)| + 1 = d_i + 1$$

□

Thus we see that each edge yields exactly an expected influence of 1 irrespective of the network beyond that edge (see example in Figure 2.9). Note that, the local property (degree) of a node governs its global effect (the total influence of that node on the network). Thus, the most influential node is the node with the highest degree. In order to pick the second node for the greedy algorithm we need to maximize $\sigma^{(\mathcal{N},i\cup j)}$ where i is the node with the highest degree. By using Equations 2.2 and 2.3 we can write,

$$\sigma^{(\mathcal{N},i\cup j)} = \sigma^{(\mathcal{N}\setminus i,j)} + \sigma^{(\mathcal{N}\setminus j,i)}$$

$$\sigma^{(\mathcal{N},i)} = \sigma^{(\mathcal{N}\setminus j,i)} + w_{i,j}\sigma^{(\mathcal{N}\setminus i,j)}$$

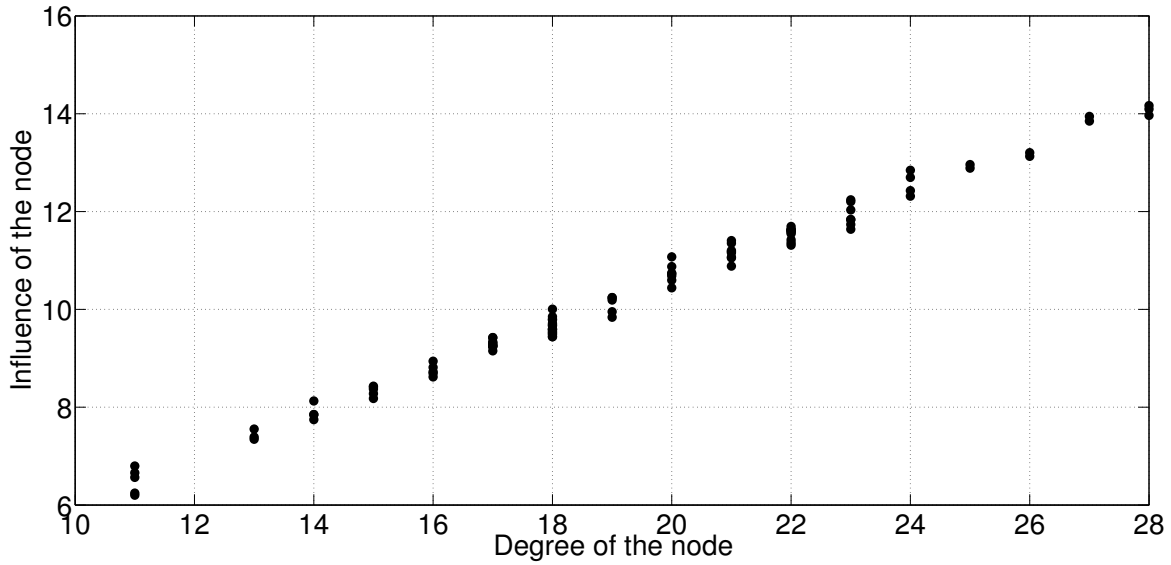


Figure 2.10: Scattergram between degree and influence of a node under the degree-based influence model in an Erdos-Renyi random graph of 100 nodes

$$\sigma^{(\mathcal{N},j)} = \sigma^{(\mathcal{N}\setminus i,j)} + w_{j,i}\sigma^{(\mathcal{N}\setminus j,i)}$$

From the above expressions, we find that if i and j are high degree nodes, their net influence can be approximated well by the sum of their individual influences, since $w_{i,j}$ and $w_{j,i}$ are small. From this, we can how picking the high degree nodes will give us a very good approximation of the greedy solution. By applying the PageRank algorithm using $P = \mathbf{W}^T$ as the transition probability matrix, where $\mathbf{W}$ is as defined in Equation 2.5 we get the stationary probability to be,

$$\pi_i = \frac{d_i}{\sum_j d_j}$$

and we find that PageRank algorithm matches with the heuristic of choosing the nodes according to degree and hence will give a good approximation of the greedy solution. We also tested PageRank algorithm on undirected graphs which may have cycles. It turns out that there is still a high correlation (see Figure 2.10) between the degree of a node and its individual influence. However, notice that the influence of a single node is less than its degree (as against the result in Theorem 2.5.1). However, even in this case picking the nodes in the decreasing order of degree yields a good estimate of the optimal initial set chosen by the greedy algorithm.

2.5.4 UISLT Model

We introduce a special case of the Linear Threshold model, called the Uniform Influence-Susceptance Linear Threshold model (UISLT) on a complete graph. In this model, we have two parameters α_i and β_i associated with the node i, a measure of the level of influence and susceptance of the node i. The social network is a complete graph with the matrix $\mathbf{W}$ defined as follows:
For all i, j with $i \neq j$,

$$w_{i,j} = \alpha_i \times \beta_j, \text{for all} j \neq i$$

and

$$w_{i,i} = 1 - \sum_{j \neq i} w_{j,i}$$

Note that α_i's and β_i's are chosen such that, $\sum_{j \neq i} w_{j,i} \leq 1$. This implies that, for all i,

$$\sum_{j \neq i} \alpha_j \leq \frac{1}{\beta_i}$$

Theorem 2.5.2. Let $\mathcal{N} = \mathcal{A}_0 \cup \{j_1, j_2, \cdots j_k\}$. where $K = |\mathcal{A}_0|$ and $k = |\mathcal{N}| - K$. Then the total influence of the initial set $\mathcal{A}_0$ in the UISLT model is given by

$$\sigma^{(\mathcal{N}, \mathcal{A}_0)} = |\mathcal{A}_0| + \alpha_{\mathcal{A}_0} \sum_{m=0}^{k-1} h^{(m)}(\alpha, \beta)$$

where $\alpha = \{\alpha_{j_1}, \alpha_{j_2}, \cdots \alpha_{j_k}\}$, $\beta = \{\beta_{j_1}, \beta_{j_2}, \cdots \beta_{j_k}\}$, $\alpha_{\mathcal{A}_0} = \sum_{i \in \mathcal{A}_0} \alpha_i$, and

$$h^m(\alpha, \beta) = \sum_{\{l_1, \ldots, l_{m+1}\} \subseteq \{1, \ldots, k\}} \beta_{j_{l_1}} \times \cdots \times \beta_{j_{l_{m+1}}} f^m(\alpha_{j_{l_1}}, \ldots, \alpha_{j_{l_{m+1}}})$$

where $f^0(x_1, \ldots, x_t) = 1$ and for all $m > 0$,

$$f^m(x_1, \ldots, x_t) = m! \sum_{\{y_1, \ldots, y_m\} \subseteq \{x_1, \ldots, x_t\}} y_1 \times \cdots \times y_m$$

Proof. The proof is by direct application of Equations 2.2 and 2.3. □

From the general UISLT model, we derive two special case, namely the Uniform Susceptance LT (USLT) model and the Uniform Influence LT (UILT) model. In both the UILT and the USLT model, we derive the optimal initial set and show that the Pagerank

also selects the same set of nodes.

USLT model For the Uniform Susceptance model, from the above general equation for UISLT, by setting $\alpha = (1, \ldots, 1)$, and noting that,

$$f^m(\alpha_{j_{l_1}}, \ldots, \alpha_{j_{l_{m+1}}}) = (m+1)!$$

$$\alpha_{\mathcal{A}_0} = |\mathcal{A}_0|$$

we get,

$$\sigma^{(\mathcal{N},\mathcal{A}_0)} = |\mathcal{A}_0| + |\mathcal{A}_0| \sum_{m=0}^{k-1} f^{m+1}(\beta)$$

where f^m is the same as defined earlier. Hence in this case $\sigma^{(\mathcal{N},\mathcal{A}_0)}$ is an increasing function of β. Thus to maximize $\sigma^{(\mathcal{N},\mathcal{A}_0)}$, we need to pick $\mathcal{A}_0$ such that, the nodes with maximum β_i are left out. Thus the optimal $\mathcal{A}_0$ in this case, is the set of K nodes with least β_i values. This also makes intuitive sense, since by picking this $\mathcal{A}_0$ (the least susceptible nodes), we ensure that the inactive nodes are the most susceptible ones, and hence maximizing expected cardinality of the terminal set. It turns out that when we apply the PageRank algorithm, we get the stationary probability as,

$$\pi_i = \frac{1/\beta_i}{\sum_j 1/\beta_j}$$

Thus choosing the nodes with top-k π_i yields us the same optimal set, i.e., the set of K nodes with least β_i values.

UILT model For the Uniform Influence model, from the above general equation for UISLT, by setting $\alpha = (1, \ldots, 1)$, we get,

$$h^m(\alpha) = \sum_{\{l_1,\ldots,l_{m+1}\} \subseteq \{1,\ldots,k\}} f^m(\alpha_{j_{l_1}}, \ldots, \alpha_{j_{l_{m+1}}})$$

Hence,

$$h^m(\alpha) = (k-m) f^m(\alpha_{j_1}, \ldots, \alpha_{j_k})$$

$$\sigma^{(\mathcal{N},\mathcal{A}_0)} = |\mathcal{A}_0| + k\alpha_{\mathcal{A}_0} + \alpha_{\mathcal{A}_0} \sum_{m=1}^{k-1} (k-m) f^m(\alpha_{j_1}, \ldots, \alpha_{j_k})$$

Hence in this case $\sigma^{(\mathcal{N},\mathcal{A}_0)}$ is an increasing function of $\alpha_{\mathcal{A}_0}$, fixing $|\mathcal{A}_0|$ to be K. Thus to maximize $\sigma^{(\mathcal{N},\mathcal{A}_0)}$, we need to pick $\mathcal{A}_0$ such that $\alpha_{\mathcal{A}_0}$ is maximized. Thus the optimal $\mathcal{A}_0$ in this case, is the set of K nodes with highest α_i values. This also makes intuitive sense, since α_i is a measure of influence, and the $\mathcal{A}_0$ thus obtained would be the set of most influential nodes. It turns out that when we apply the PageRank algorithm, we get the stationary probability as,

$$\pi_i = \frac{\alpha_i}{\sum_j \alpha_j}$$

Thus choosing the nodes with top-k π_i yields us the same optimal set, i.e., the set of K nodes with highest α_i values.

UISLT model and PageRank

Though for UILT and USLT models we could show that Pagerank is optimal, the PageRank algorithm need not be optimal for the UISLT case. For PageRank, in the UISLT case, the stationary probability is given by,

$$\pi_i = \frac{\alpha_i/\beta_i}{\sum_j \alpha_j/\beta_j}$$

Hence one might suspect that picking the nodes in increasing order of α_i/β_i could be optimal. But it could turn out to be suboptimal, since a node with β_i very close to zero could get chosen as the most influential node irrespective of its α_i. If we restrict the β_i, by not allowing it vary much, we see that the PageRank algorithm gives a very good approximation of the greedy solution. The following simulation was conducted on a complete graph with 50 nodes with the UISLT model. The α_i's were picked at random from a uniform distribution over $[0, 1]$ and β_i's were picked with a uniform distribution over $[\frac{0.5}{\sum_{j\neq i}\alpha_j}, \frac{1}{\sum_{j\neq i}\alpha_j}]$. It is found that PageRank performs on par with the greedy algorithm. Results are shown in Figure 2.11.

2.6 Improving the Greedy Algorithm

As shown by Kempe et al. the greedy hill-climbing solution achieves $(1-1/e)$ approximate solution for the influence maximization problem [1]. But finding the initial set using the greedy algorithm is computationally quite expensive. There have been several efforts in the literature to improve the execution speed of the greedy algorithm [52], [53]. In this section, based on the insights obtained from the analytical expressions, we provide two

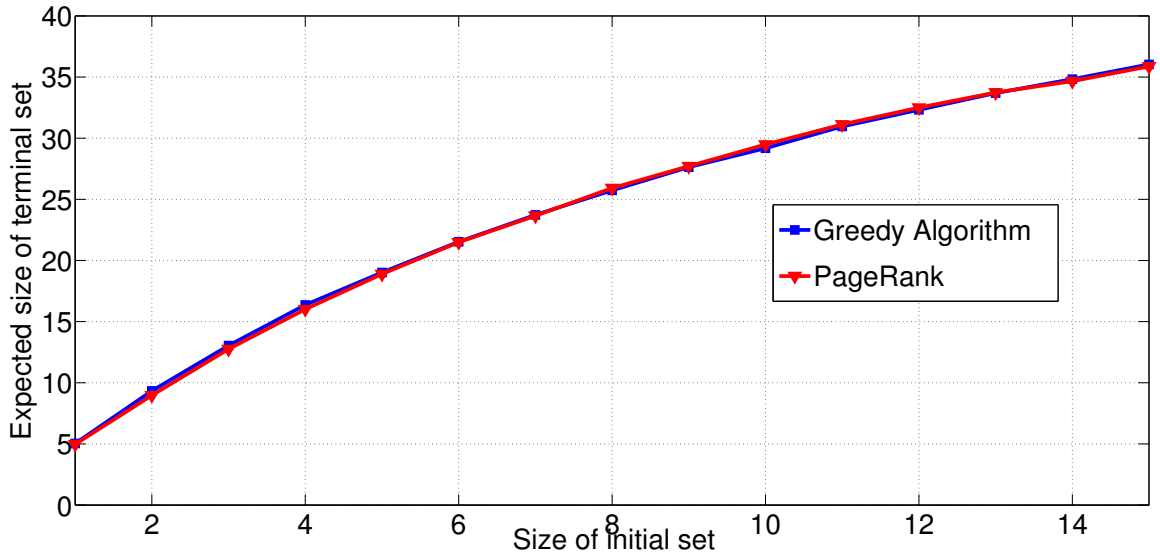


Figure 2.11: UISLT on a complete influence graph of 50 nodes, with α_i's and β_i's being picked as described in Section 2.5.4

techniques, namely thresholding and restriction, which achieve a very close approximation to the greedy solution.

During the greedy algorithm, the first stage involves computation of $\sigma^{(\mathcal{N},i)}$ for all nodes $i \in \mathcal{N}$. One can rank the nodes in decreasing order of individual influences to yield a rank list which we shall refer to as **G1** list. One possible solution for influence maximization is to pick the top K nodes to be the initial set. This solution is not as effective as the greedy solution, because it may contain *dummy* nodes, i.e., nodes which in spite of having a high individual influence, fail to provide high marginal contributions to the greedy solution set and hence get rejected by the greedy algorithm. We classify the *dummy* nodes into two categories, namely the *leechers* and *subordinates*. Let $\mathcal{X}$ denote the set of nodes in **G1** which are above i and have been picked to be part of the initial set. We define a node i to be an α-subordinate of set $\mathcal{X}$ if

$$g_i^{(\mathcal{N},\mathcal{X})} > \alpha$$

This means that node i gets activated at least α fraction of the time when we begin with $\mathcal{X}$ as the initial set. Thus, for high enough α, the marginal contribution of this node to the set $\mathcal{X}$ will be much smaller than its individual influence. A node i is a *leecher* to set $\mathcal{X}$, if

$$\sigma^{(\mathcal{N},i)} \approx \sum_{j \in \mathcal{X}} w_{i,j} \sigma^{(\mathcal{N} \setminus i,j)}$$

This means that the node i will have almost nil marginal contribution to a set which

already contains $\mathcal{X}$, since it primarily derives all its influence from those nodes. If one can identify and eliminate such subordinates and leechers from **G1** while picking the initial set, then a more effective initial set could be obtained. We use two techniques namely *thresholding* and *restriction* to filter out the subordinate nodes and leechers respectively. Thresholding involves comparing the $g_i^{(\mathcal{N},\mathcal{X})}$ with α and restriction involves evaluating $\sigma^{(\mathcal{N}\setminus\mathcal{X},i)}$ to pick the next best node.

2.6.1 G1-Sieving Algorithm

In this algorithm, we first evaluate $\sigma^{(\mathcal{N},i)}$ for $i \in \mathcal{N}$ and obtain the **G1** list. We start with $\mathcal{X} = \{i\}$. We remove the nodes which are α-subordinates of set $\mathcal{X}$ by evaluating $g_i^{(\mathcal{N},\mathcal{X})}$ for all nodes and comparing with the threshold α. Note that this involves just one influence computation, since $\sigma^{(\mathcal{N},\mathcal{X})}$ involves computing $g_i^{(\mathcal{N},\mathcal{X})}$. For the remaining nodes, if we choose to perform restriction, we compute $\sigma^{(\mathcal{N}\setminus\mathcal{X},i)}$ and pick the node that has the highest value. We add the picked node to $\mathcal{X}$ and repeat the procedure until we have a set of size K or we exhaust the entire list. The algorithm is shown in Algorithm 2. In most networks,

```
Evaluate σ^(N,i) for all i ∈ N;
Sort nodes in decreasing order of σ^(N,i) to get G1;
X = G1(1);
G1 = G1\X;
for c = 2 to K do
    if G1 = ∅ then
        break;
    end
    for i ∈ G1 do
        if g_i^(N,X) > α || σ^(N\X,i) < ε then
            G1 = G1\{i};
            continue;
        end
        Evaluate σ^(N\X,i);
    end
    v = arg max_{i∈G1} σ^(N\X,i);
    X = X ∪ v;
    G1 = G1\{v};
end
```

Algorithm 2: G1-Sieving Algorithm

we observe that G1-sieving with only thresholding yields a performance close to the greedy algorithm. Also, this requires only one influence computation per round (except the first

round for generating the **G1** list), as against N computations for the original greedy algorithm. Also, even if we perform restriction, it requires the computation of $\sigma^{(\mathcal{N}\setminus\mathcal{X},i)}$ for each i. Since, this is performed on a restricted network, and also with a singleton initial seed, techniques such as SCC (strongly connected component) [54] can be used to significantly reduce the computation. Thus, the G1-sieving algorithm requires significantly less computational effort in comparison to the greedy algorithm, and its performance is close to the greedy algorithm, as shown in Section 2.7.3.

2.7 Simulations

2.7.1 Coauthorship Networks

Newman [92] observed that scientific colloboration networks are excellent examples of social networks. In such networks, each node represents an author in the scientific community under consideration, and an edge exists between two nodes i and j if those two authors are listed as co-authors at least in one of the papers. Newman in [93] explains a method by which the strength of collaboration (symmetric) between two authors can be extracted. We use this data to obtain the Linear Threshold model parameters. The process is as follows. Let $\mathcal{R}$ denote the set of all papers under consideration in the scientific community, excluding the papers that only have a single author. For each $r \in \mathcal{R}$, let n_r represent the number of coauthors for paper r. Let $\mathcal{N}$ represent the union of all authors of the papers in $\mathcal{R}$. Define $\delta(i,r)$ to be 1 if author i was a co-author of paper r and zero otherwise. Then $\tilde{w}_{i,j}$ representing the strength of collaboration between authors i and j, for $i \neq j$, is given by,

$$\tilde{w}_{i,j} = \sum_{r\in\mathcal{R}} \frac{\delta(i,r)\delta(j,r)}{n_r - 1}$$

Using the strengths of collaboration $\tilde{w}_{i,j}$, we obtain the entries of influence matrix $\mathbf{W}$ by normalizing. i.e.,

$$w_{i,j} = \frac{\tilde{w}_{i,j}}{\sum_{k=1,k\neq j}^{|\mathcal{N}|} \tilde{w}_{i,j}}$$

The simulations in the following sections are carried out on a coauthorship network of NetScience community containing 1589 nodes.

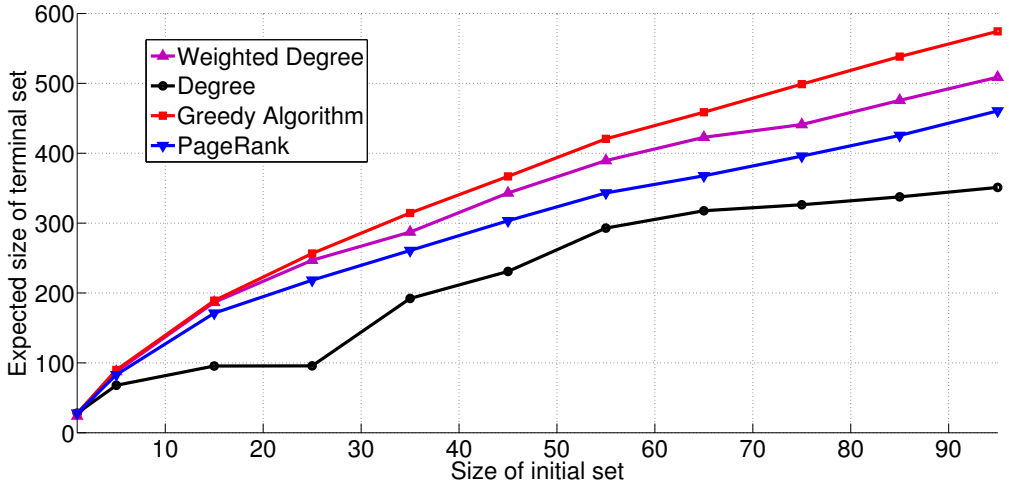


Figure 2.12: Comparison of PageRank and Greedy - undirected case

2.7.2 Comparing PageRank with the Greedy Algorithm

In Section 2.5 we examined various cases, where the PageRank algorithm was either optimal or performed very well. Here we report simulations comparing PageRank and Greedy algorithm on the Netscience dataset. They are also compared against heuristics such as the out-degree(obtained by counting the number of outgoing edges) and the weighted out-degree(summing up the weights on the outgoing edges). In the scenario where the $w_{i,j}$'s are obtained as above, it turned out that the PageRank algorithm's performance was much below that of the greedy algorithm. We ran another set of simulations where $\tilde{\mathbf{W}}$ was interpreted as directed, for experimental purposes. This means that the "strength of collaboration" $\tilde{w}_{i,j}$ between two authors i and j was completely assigned to the author with a higher index, i.e., for $i < j$, $\hat{w}_{j,i} = \tilde{w}_{i,j}$ and $\hat{w}_{i,j} = 0$. This would result in an influence graph which is directed, acyclic (DAG). It was found that PageRank algorithm performed on par with the greedy algorithm in networks that are directed, acyclic. The results are shown in figures 2.12 and 2.13.

We see that this is because, the PageRank algorithm essentially works with random walks on the given graph and finding the stationary probability. But as we have pointed out in Section 2.3 maximizing the spread of influence involves random walks that are "self-avoiding". Thus it turns out that in the directed case, where there are no cycles, the Pagerank computed for a given node correlates directly with its influence, whereas in the undirected case, it is not so, thus the Pagerank fails to perform on par with the greedy algorithm. Thus, even though the Pagerank can be calculated efficiently to give a rough estimate of the *network effect* of a node, in many cases it might perform poorly

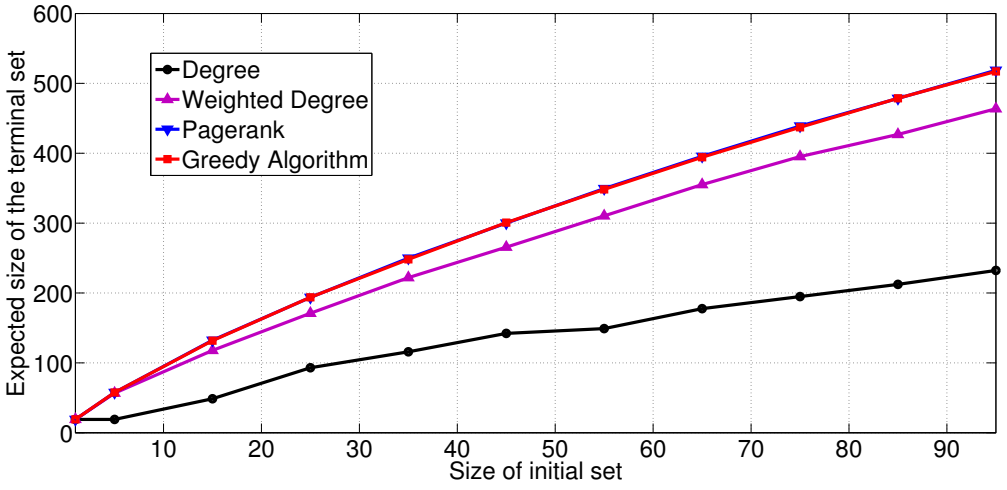


Figure 2.13: Comparison of PageRank and Greedy - directed case

compared to the greedy algorithm, depending on how far the influence graph is from a DAG (directed acyclic graph).

2.7.3 Comparison of G1-Sieving with Greedy Algorithm

We carried out two experiments with the NetScience dataset for the G1-sieving algorithm. The first involved using only the thresholding technique with various values for α. It is interesting to note the change in performance of the thresholding technique with variation of α as shown in Figure 2.14. For low values of α, the algorithm retains only the nodes whose influence domains are almost disjoint, and hence performs badly. Also, for high values of α, the algorithm does not remove many nodes, and the list almost resembles the **G1** list. It turned out that for this dataset, thresholding with $\alpha = 0.3$ provided a very good approximation of the greedy solution. In the second case, we implement the G1-Sieving algorithm with the threshold $\alpha = 0.3$ and with restriction. It is found that restriction provides a slight improvement over the solution with only thresholding. The results are shown in Figure 2.15. We find that G1-Sieving algorithm performs on par with the greedy algorithm.

When $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{N}$ and set sizes are very small compared to $|\mathcal{N}|$, due to the monotonicity of the influence function, evaluating $\sigma^{(\mathcal{N},\mathcal{A})}$ is faster than $\sigma^{(\mathcal{N},\mathcal{B})}$, since the latter involves more number of activations, whereas for set sizes closer to $|\mathcal{N}|$, evaluating $\sigma^{(\mathcal{N},\mathcal{B})}$ is faster than $\sigma^{(\mathcal{N},\mathcal{A})}$, since the former involves fewer nodes to be activated. Also for a given set size, the time taken for evaluating the influence of a set $\mathcal{A}$ increases with the

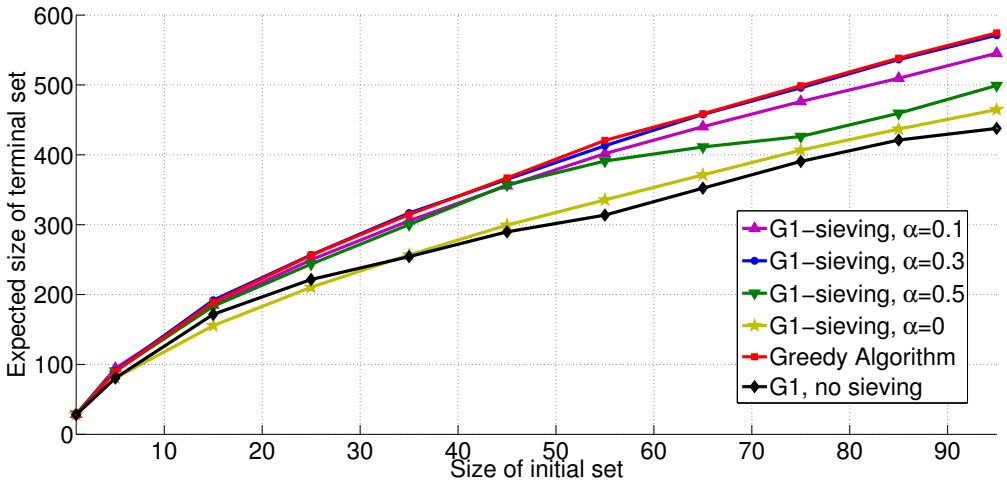


Figure 2.14: Comparison of various sieving thresholds with Greedy algorithm

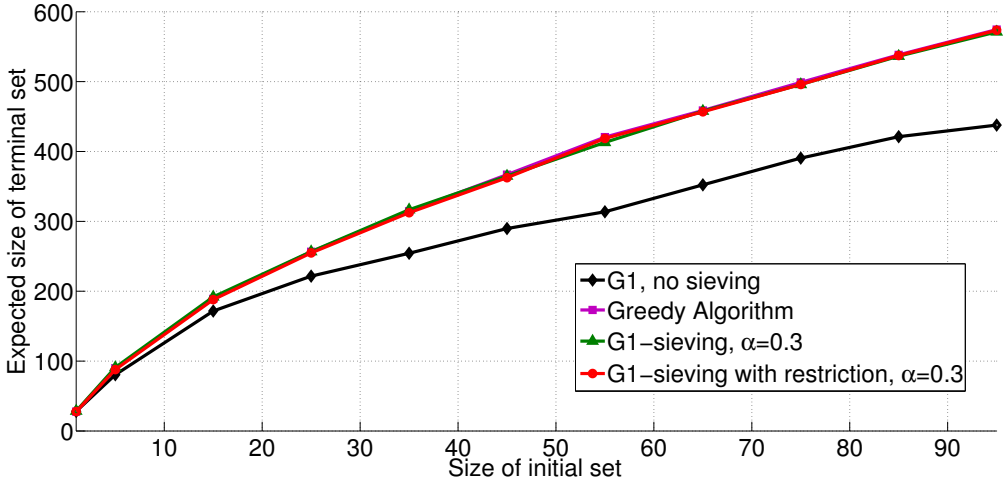


Figure 2.15: Comparison of G1-Sieving with Greedy algorithm

expected influence of the set. Keeping these in mind, we see that G1-Sieving algorithm runs much faster than greedy algorithm, since it evaluates influences for only sets with single node as the initial set, and as it proceeds it evaluates influences for fewer influential nodes in a restricted graph. Also, the technique of employing only thresholding, performs almost on par with the Greedy algorithm, given the fact that it involves evaluating $\sigma^{(\mathcal{N},i)}$ for all $i \in \mathcal{N}$, i.e., the first stage of greedy algorithm, and then subsequently evaluating influences (to get the activation probabilities $g_j^{(\mathcal{N},\mathcal{X})}$) only once per round. For the G1-Sieving algorithm, we obtained the optimal α using simulations. It would be interesting to look at ways to estimate α based on the graph structure. One can also use variable α, by having a higher threshold for initial few rounds.

Recall that, using the submodularity of the objective function, Kempe et al. [1] were able to provide a $(1 - \frac{1}{e})$ approximation guarantee for the greedy hill-climbing algorithm. Also, when the expectation computations are not exact, they obtain a $(1 - \frac{1}{e} - \epsilon)$ approximation for any $\epsilon > 0$. However, in our case, since G1-sieving is a *heuristic* based on the insights derived from the analytical expression, we were unable to provide any theoretical bounds (approximation guarantees) on its performance. Thus, the impact of inexact expectation computations is also unclear. We only demonstrate G1-sieving's effectiveness via numerical simulations in comparison with the greedy algorithm.

2.8 Summary

In this chapter, we derived an analytical expression for the influence of a given set in a social network under the Linear Threshold model. We used it to provide an equivalent interpretation of the influence maximization problem in terms of acyclic path probabilities in a Markov chain. We derived explicit optimal initial sets for some example networks, and showed scenarios where Pagerank was optimal. We used the insights from the analytical expressions to propose an efficient algorithm for influence maximization, which performs on par with the greedy algorithm.

An interesting implication of this work is the role played by self avoiding random walks in the analytical expression for the influence function. Finding an efficient way to compute these probabilities will speed up the influence computations. PageRank algorithm was found to be sub-optimal in graphs that were not acyclic, since it does not discount paths which have cycles. As an interesting aside, one can even model the walk of random surfer on the Web graph to be a "self-avoiding" random walk which might lead to better

Web-page ranking algorithms.

2.9 Appendix

2.9.1 Properties of Acyclic path probabilities

We shall use the interpretation in terms of acyclic path probabilities in the DTMC, to provide an alternate proof of submodularity of $\sigma^{(\mathcal{N},\mathcal{A}_0)}$. In this Appendix, we state and prove a few properties of such acyclic path probabilities. Let us first introduce a more elaborate notation.

$$\begin{aligned} c^{\mathcal{W}}(j \xrightarrow{v} \mathcal{D}) &= \mathbb{P}\Big(\bigcup_{k=0}^{\infty} \{X_m \in \mathcal{W}\backslash\mathcal{D}, 0 \leq m < k, X_k \in \mathcal{D}, \\ &\qquad X_0 \neq X_1 \neq \cdots \neq X_k, \exists \;\; 0 \leq u < k, s.t.\; X_u = v\} \big| X_0 = j \Big) \\ &= \sum_{k=0}^{\infty} \mathbb{P}\Big(\{X_m \in \mathcal{W}\backslash\mathcal{D}, 0 \leq m < k, X_k \in \mathcal{D}, \\ &\qquad X_0 \neq X_1 \neq \cdots \neq X_k, \exists \;\; 0 \leq u < k, s.t.\; X_u = v\} \big| X_0 = j \Big) \end{aligned} \tag{2.9}$$

Here $c^{\mathcal{W}}(j \xrightarrow{v} \mathcal{D})$ is the probability in the DTMC of reaching the set $\mathcal{D}$ for the first time, through an acyclic path via node v, using nodes only from $\mathcal{W}$ as intermediate nodes, given that we start from node j. If $\mathcal{W} = \mathcal{N}$, we do not explicitly mention it in the notation. Also, v is an optional argument, which when omitted, removes the constraint that $X_u = v$ for some $0 \leq u \leq k$. In all useful cases that we consider, we assume $v, j \notin \mathcal{D}$ and also $v, j \in \mathcal{W}$.

Some properties of Acyclic path probabilities are as follows:

Lemma 2.9.1. $c(j \to \mathcal{A}) = 1$, for all $j \in \mathcal{A}$

Proof.

$$\begin{aligned} c(j \to A) &= \sum_{k=0}^{\infty} \mathbb{P}\Big(\{X_m \in \mathcal{N}\backslash\mathcal{A}, 0 \leq m < k\}, X_k \in \mathcal{A}, X_0 \neq \cdots \neq X_k \big| X_0 = j \Big) \\ &= \mathbb{P}(X_0 \in \mathcal{A} \big| X_0 = j) + \\ &\sum_{k=1}^{\infty} \mathbb{P}\Big(X_0 \in \mathcal{N}\backslash\mathcal{A}, \{X_m \in \mathcal{N}\backslash\mathcal{A}, 0 < m < k\}, X_k \in \mathcal{A}, X_0 \neq X_1 \neq \cdots X_k \big| X_0 = j \Big) \end{aligned}$$

Since $j \in \mathcal{A}$, all the terms in the summation are zero, and hence, $c(j \to A) = \mathbb{P}(X_0 \in$

$\mathcal{A}\big|X_0 = j) = 1$ □

Lemma 2.9.2. $c(j \to \mathcal{A}) = c(j \xrightarrow{v} \mathcal{A}) + c^{\mathcal{N}\setminus\{v\}}(j \to \mathcal{A})$

This property states that the probability of reaching $\mathcal{A}$ starting from j through an acyclic path, can be split into the probability of those paths via node v and those that avoid node v.

Proof.

$$\begin{aligned}
c(j \to A) &= \sum_{k=0}^{\infty} \mathbb{P}\Big(\{X_m \in \mathcal{N}\setminus\mathcal{A}, 0 \le m < k, X_k \in \mathcal{A}, X_0 \neq X_1 \neq \cdots \neq X_k\}\big|X_0 = j \Big) \\
&= \sum_{k=0}^{\infty} \mathbb{P}\Big(\{X_m \in \mathcal{N}\setminus\mathcal{A}, 0 \le m < k, X_k \in \mathcal{A}, \\
&\qquad\qquad X_0 \neq X_1 \neq \cdots \neq X_k, \exists\ 0 \le u < k,\ s.t.\ X_u = v\}\big|X_0 = j \Big) + \\
&\quad \sum_{k=0}^{\infty} \mathbb{P}\Big(\{X_m \in \mathcal{N}\setminus\mathcal{A}, 0 \le m < k, \\
&\qquad\qquad X_k \in \mathcal{A}, X_0 \neq X_1 \neq \cdots \neq X_k, \nexists\ 0 \le u < k,\ s.t.\ X_u = v\}\big|X_0 = j \Big) \\
&= c(j \xrightarrow{v} \mathcal{A}) + \\
&\sum_{k=0}^{\infty} \mathbb{P}\Big(\{X_m \in (\mathcal{N}\setminus\{v\})\setminus\mathcal{A}, 0 \le m < k, X_k \in \mathcal{A}\setminus\{v\}, X_0 \neq X_1 \neq \cdots \neq X_k\}\big|X_0 = j \Big)
\end{aligned}$$

Since, $v \notin \mathcal{A}$, $\mathcal{A}\setminus\{v\} = \mathcal{A}$ and we get, $c(j \to \mathcal{A}) = c(j \xrightarrow{v} \mathcal{A}) + c^{\mathcal{N}\setminus\{v\}}(j \to \mathcal{A})$ □

Lemma 2.9.3. $c(j \to \mathcal{A} \cup \{v\}) = c^{\mathcal{N}\setminus\mathcal{A}}(j \to v) + c^{\mathcal{N}\setminus\{v\}}(j \to \mathcal{A})$, for all $v \notin \mathcal{A}$

This property means that the probability of reaching $\mathcal{A} \cup \{v\}$ through acyclic path, can be split into the probability of reaching $\mathcal{A}$ avoiding node v, and that of reaching node v, avoiding set $\mathcal{A}$.

Proof.

$$\begin{aligned}
c(j \to \mathcal{A} \cup \{v\}) &= \sum_{k=0}^{\infty} \mathbb{P}\Big(\{X_m \in \mathcal{N}\setminus(\mathcal{A} \cup \{v\}), 0 \le m < k, X_k \in \mathcal{A} \cup \{v\}, \\
&\qquad\qquad X_0 \neq X_1 \neq \cdots \neq X_k\}\big|X_0 = j \Big) \\
&= \sum_{k=0}^{\infty} \mathbb{P}\Big(\{X_m \in (\mathcal{N}\setminus\mathcal{A})\setminus\{v\}, 0 \le m < k, X_k = v,
\end{aligned}$$

$$
\begin{aligned}
& X_0 \neq X_1 \neq \cdots \neq X_k\} \big| X_0 = j \Big) + \\
\sum_{k=0}^{\infty} \mathbb{P}\Big(& \{X_m \in (\mathcal{N}\backslash\{v\})\backslash\mathcal{A}, 0 \leq m < k, X_k \in \mathcal{A}, \\
& X_0 \neq X_1 \neq \cdots \neq X_k\} \big| X_0 = j \Big) \\
= \; & c^{\mathcal{N}\backslash\mathcal{A}}(j \to v) + c^{\mathcal{N}\backslash\{v\}}(j \to \mathcal{A})
\end{aligned}
$$

The second equality results from the assumption that $v \notin \mathcal{A}$. Note that this property can be extended to $c(j \to \mathcal{A} \cup \mathcal{B})$, where $\mathcal{A}$ and $\mathcal{B}$ are any two disjoint sets. □

Before stating and proving the next property, we wish to prove the following two sublemmas.

Sublemma 2.9.1. $c(j \to v) = \sum_{L' \subseteq \mathcal{N}-\{j,v\}} \sum_{L \in \Pi(L')} p_{j,v}(L)$ where $\Pi(L')$ is the set of all permutations of L', and for $L = \{l_1, l_2, \cdots l_{k-1}\}$,

$$
p_{j,v}(L) = \begin{cases} p_{j,l_1} p_{l_1,l_2} \cdots p_{l_{k-1},v} & \text{if } L = \{l_1, \ldots l_{k-1}\} \\ p_{j,v} & \text{if } L = \emptyset \end{cases} \tag{2.10}
$$

Proof.

$$
\begin{aligned}
c(j \to v) \; = \; & \sum_{k=1}^{\infty} \mathbb{P}\Big(X_1 = l_1, X_2 = l_2, \ldots X_{k-1} = l_{k-1}, X_k = v, \\
& l_1 \neq l_2 \neq \ldots l_{k-1} \neq j \neq v \big| X_0 = j \Big) \\
= \; & \sum_{L' \subseteq \mathcal{N}-\{j,v\}} \sum_{L \in \Pi(L')} p_{j,v}(L)
\end{aligned}
$$

The second equality is by using the chain rule and the Markov property of X_k. □

Sublemma 2.9.2.

$$
c(j \xrightarrow{v} \mathcal{A}) = \sum_{L' \subseteq \mathcal{N}\backslash\mathcal{A}-\{j,v\}} \sum_{L \in \Pi(L')} p_{j,v}(L) c^{\mathcal{N}\backslash(L' \cup \{j\})}(v \to \mathcal{A})
$$

Proof.

$$
c(j \xrightarrow{v} \mathcal{A}) \; = \; \sum_{k=0}^{\infty} \mathbb{P}\Big(\{X_m \in \mathcal{N}\backslash\mathcal{A}, 0 \leq m < k, X_k \in \mathcal{A},
$$

$$
\begin{aligned}
&\qquad X_0 \neq X_1 \neq \cdots \neq X_k, \\
&\qquad \exists\ 0 \leq u < k,\ s.t.\ X_u = v\}\big|X_0 = j\Big) \\
&= \sum_{k=0}^{\infty} \sum_{L' \subseteq \mathcal{N}\setminus\mathcal{A}} \sum_{u=1}^{k-1} \mathbb{P}\Big(X_1 = l_1, \ldots, X_{u-1} = l_{u-1}, X_u = v, \\
&\qquad l_i \notin \mathcal{A}, \{X_m \in \mathcal{N}\setminus\mathcal{A}, u < m < k\}, X_k \in \mathcal{A}, \\
&\qquad X_0 \neq X_1 \neq \cdots \neq X_k\big|X_0 = j\Big) \\
&= \sum_{k=0}^{\infty} \sum_{L' \subseteq \mathcal{N}\setminus\mathcal{A}} \sum_{u=1}^{k-1} \mathbb{P}\Big(X_1 = l_1, \ldots, X_{u-1} = l_{u-1}, X_u = v, \\
&\quad l_1 \neq \ldots \neq l_{u-1}, l_i \neq v \neq j,\ l_i \notin \mathcal{A}, \\
&\quad \{X_m \in (\mathcal{N}\setminus\{L' \cup j\})\setminus\mathcal{A}, u \leq m < k\}, X_k \in \mathcal{A}\big|X_0 = j\Big) \\
&= \sum_{k=0}^{\infty} \sum_{L' \subseteq \mathcal{N}\setminus\mathcal{A}} \sum_{u=1}^{k-1} \mathbb{P}\Big(X_1 = l_1, \ldots, X_{u-1} = l_{u-1}, X_u = v, \\
&\qquad l_1 \neq \ldots \neq l_{u-1}, l_i \neq v \neq j, l_i \notin \mathcal{A}\big|X_0 = j\Big) \times \\
&\quad \mathbb{P}\Big(\{X_m \in (\mathcal{N}\setminus\{L' \cup j\})\setminus\mathcal{A}, u \leq m < k\}, \\
&\qquad\qquad X_k \in \mathcal{A}\big|X_u = v\Big) \\
&= \sum_{L' \subseteq \mathcal{N}-(\mathcal{A}\cup\{j,v\})} \sum_{L \in \Pi(L')} p_{j,v}(L) c^{\mathcal{N}\setminus\{L'\cup j\}}(v \to \mathcal{A})
\end{aligned}
$$

where $p_{j,v}(L)$ defined as above in Equation 2.10 and $L' = \{l_1, \ldots l_{u-1}\}$. The penultimate equality is by applying Markov property at $X_u = v$. □

Now we can state and prove the next property.

Lemma 2.9.4. $c(j \xrightarrow{v} \mathcal{A}) \leq c^{\mathcal{N}\setminus\mathcal{A}}(j \to v)$

Proof.

$$
\begin{aligned}
c(j \xrightarrow{v} \mathcal{A}) &= \sum_{L' \subseteq \mathcal{N}-(\mathcal{A}\cup\{j,v\})} \sum_{L \in \Pi(L')} p_{j,v}(L) c^{\mathcal{N}\setminus(L'\cup\{j\})}(v \to \mathcal{A}) \\
&\leq \sum_{L' \subseteq \mathcal{N}-(\mathcal{A}\cup\{j,v\})} \sum_{L \in \Pi(L')} p_{j,v}(L) = c^{\mathcal{N}\setminus\mathcal{A}}(j \to v)
\end{aligned}
$$

In the above proof the equalities are due to Sublemmas 2.9.1 and 2.9.2. The inequality is because $c^{\mathcal{N}\setminus(L'\cup\{j\})}(v \to \mathcal{A})$, being a probability term, is less than 1. □

Lemma 2.9.5. For $\mathcal{A} \subseteq \mathcal{B}$, $c^{\mathcal{N}\setminus\mathcal{A}}(j \to v) \geq c^{\mathcal{N}\setminus\mathcal{B}}(j \to v)$

Proof. Note that,

$$c^{\mathcal{N}\setminus\mathcal{A}}(j \to v) = \sum_{L' \subseteq \mathcal{N}-(\mathcal{A}\cup\{j,v\})} \sum_{L \in \Pi(L')} p_{j,v}(L) \tag{2.11}$$

In the expression on the right, as $\mathcal{A}$ increases, the number of possible L' decreases, hence $c^{\mathcal{N}\setminus\mathcal{A}}(j \to v)$ decreases. □

CHAPTER 3

Fluid Limit Characterization of the Linear Threshold Model

In the previous chapter, we considered a threshold based model for influence, where a node compared the net influence of its neighbours in the social network against a threshold, to decide whether to participate in a given activity, i.e., get influenced. However, in several Internet content platforms, users can obtain information on the global popularity of an item of content, for instance the number of views for a YouTube video. Global metrics such as viewcount provide a crude signal to the user about the quality of the content. Given the limited attention span and vast quantity of content available on the Internet, the tendency of a particular user to view a video or read an article increases with the number of people who have already viewed/read it. Hence a YouTube video with many views or a news article with many "Likes" on Facebook is more likely to be accessed than others. Such a behavior, where the users are oblivious of the network, can also be modeled using threshold models, as proposed in [16].

Our work in this chapter is motivated by two earlier efforts to employ a threshold model for the propensity of a user to be influenced by others in the population [1] [16]. For a detailed survey of threshold models in modeling influence dynamics, we refer the reader to Section 1.1.3. In this chapter, we begin with a homogeneous version of the LT model of [1] for a general threshold distribution (while the model in [1] assumes uniform distribution of threshold). We show that the influence evolution is a Markov process, and proceed to study its dynamics in the fluid limit [88].

Before deriving the fluid limit, we will now discuss certain modeling details in these

two important models of threshold based spread of influence, to motivate our work. Granovetter [16], rooted in sociology, considered a *general influence threshold distributions*, and characterized the spread of influence by a simple difference equation. For instance, if the threshold of individuals in the population is distributed with c.d.f. $F(\cdot)$, letting $r(t)$ denote the number of influenced individuals at time t, the evolution is described by the difference equation

$$r(t+1) = F(r(t)) \tag{3.1}$$

and thus, the equilibrium outcome is a fixed point of Equation 3.1 (see Figure 3.1). This approach provides an explicit dynamics of influence spread, and characterizes the fraction of the population that is eventually influenced. Though the analysis seems reasonable at first glance, careful inspection reveals that *the implied system dynamics will involve nodes resampling their thresholds at every timestep.* One should note that the threshold distribution is introduced to capture the variation in the unknown norms/preferences of the individuals in the population, but once sampled, they should remain unaltered during the process. Also, in Equation 3.1, there is no distinction between nodes that are already active and the nodes that are still susceptible to influence. This contradicts the assumption that the spread of influence is progressive (where nodes once influenced will remain active until the end of the process).

On the other hand, Kempe et al. [1] assumed a uniform distribution for the influence threshold, and focused on the algorithmic problem of selecting an initial seed set (of a given size) so as to maximize the final influenced set. Although, the model has progressive spread dynamics, the evolution itself was not a primary concern in [1]. It was explicitly noted that the thresholds need to be sampled only once *at the beginning of the process.*

Finally, for the uniform distribution of threshold (as assumed throughout in [1]), Equation 3.1 does not yield any useful insight, i.e., it does not predict a spread of influence.

Thus, although both Granovetter [16] and Kempe at al. [1] work with influence threshold distributions, the dynamics of the influence process are quite different. Further, although Kempe et al. work only with uniformly distributed threshold, Granovetter permits more general threshold distributions. The point of departure of our work is to adopt the idea of general threshold distributions from [16], while retaining the more natural model of sampling each individual's threshold just once in the beginning, and of the progressive spread of influence from [1].

Our Contributions: We adopt a fluid limit approach for analytically characterizing the

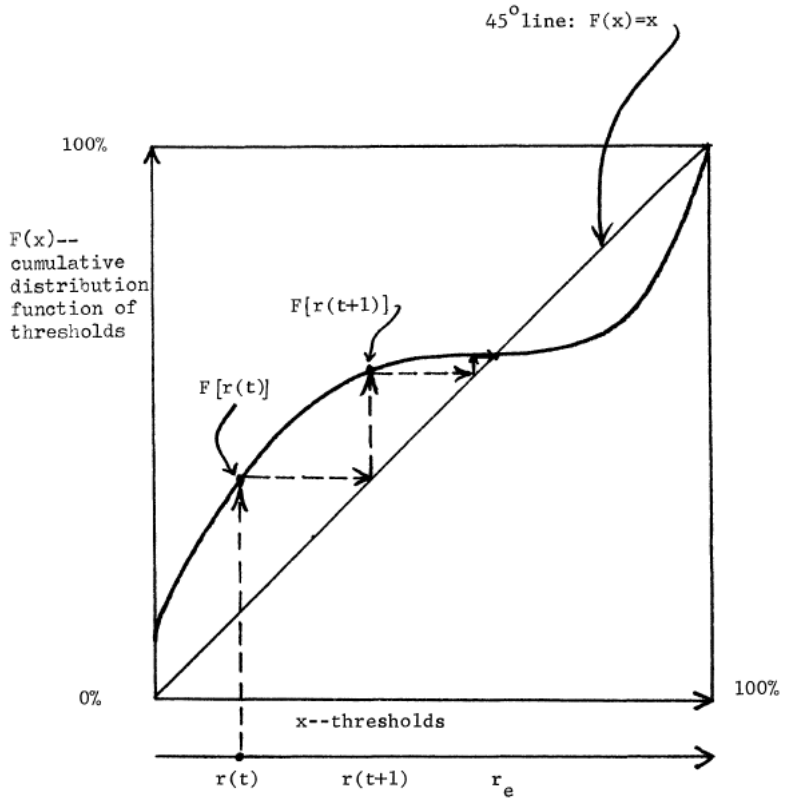


Figure 3.1: Granovetter's Fixed point dynamics. The intersection(s) of the cumulative distribution function with the 45° line, indicate the equilibrium points, i.e. solution to the fixed point equation 3.1. (Source [16])

dynamics of the spread of influence in the Linear Threshold model, with general threshold distributions. In order to do this, in Section 3.1 we propose a homogeneous version of the Linear Threshold model called the Homogeneous Influence Linear Threshold (HILT) model, with arbitrary threshold distribution. In Section 3.2 we characterize the evolution of influence in the HILT model as a Markov process, and using Kurtz's theorem [88], we derive a system of ordinary differential equations (o.d.e.). We provide simulation results that show that the o.d.e. approximates the original process fairly well for large values of N, the population size. In Section 3.3, we explicitly solve the o.d.e. for the uniform threshold distribution thus providing an explicit characterization of the evolution of influence in the model of Kempe et al. [1]. We also provide an analytical expression for the terminal spread of influence and use it to address some optimization problems (Section 5.2.7).

We note that the threshold distribution features in the o.d.e. via the *hazard function* [87], commonly used in survival/failure analysis. To the best of our knowledge, this is the first work that incorporates the hazard function to characterize variation among the individuals in an epidemic model. The variation of risk as an epidemic progresses has been

empirically observed in recent studies in veterinary medicine [94] which has highlighted the need to "consider possible differences in the risk of infection among subgroups in the population".

We then proceed to study the effect of the threshold distribution in Section 3.5. We also note that under the exponential threshold distribution, the fluid dynamics of the HILT model is equivalent to that of the classic SIR model from epidemiology [95], thus providing another interesting link between the influence spread and epidemics literature. Finally, we also show that the analysis can be extended to a heterogeneous system with communities (Section 3.6), and conclude with some possible future directions.

Comment on Network Topology: It has been noted that network topology plays a crucial role in the spread of influence on a social network [96]. In this work, however, we consider only completely connected graphs, while deriving the fluid limit equations. While it is necessary to study the impact of degree distribution on the fluid dynamics, our primary focus is to provide the missing link between Kempe's model [1] and Granovetter's model [16], and thus do not discuss the effect of network topology in this work.

3.1 Mathematical Model

Recall the network model described in Kempe et al. [1], used in Chapter 2. The social network was a weighted directed graph $\mathcal{N} = (V, E)$, where the edge weight $w_{i,j}$ gives a measure of influence of node i on node j. For clarity, we describe the activation process again, as follows. The activation process (see Figure 3.2) begins with an initial set of active, infectious nodes $\mathcal{A}_0 = \mathcal{D}_0$ and takes place in discrete time steps. Each active node spreads its influence to *each* of its inactive neighbours. By the activation process, some of the neighbours become activated to be part of $\mathcal{D}_1$, and can spread their influence in the next step. At the end of each step the population is partitioned into three sets of nodes: nodes that were just activated in that step $\mathcal{D}_k$ and have not yet had a chance to exert their influence (also referred to as *infectious* nodes), active nodes that have already exercised their influence $\mathcal{B}_k = \mathcal{A}_{k-1}$ and, hence, are no longer infectious, and the set of inactive nodes $(\mathcal{N} \backslash \mathcal{A}_k)$. Note that $D_k \subseteq A_k$. The activation process stops at a random time U when there are no more infectious nodes, i.e., $D_U = \emptyset$ and a *terminal set* A_U is reached, from where the activation process cannot proceed further. We also assume that once a node has become active, it cannot become inactive (*progressive case*).

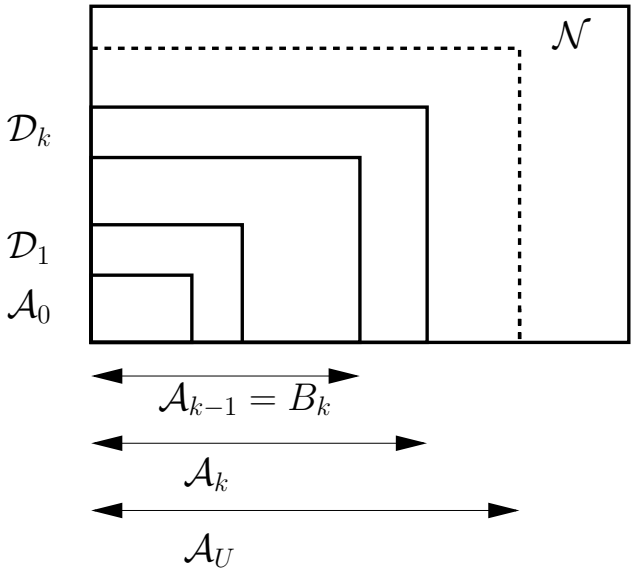


Figure 3.2: Evolution of the set of influenced nodes under the Linear Threshold model.

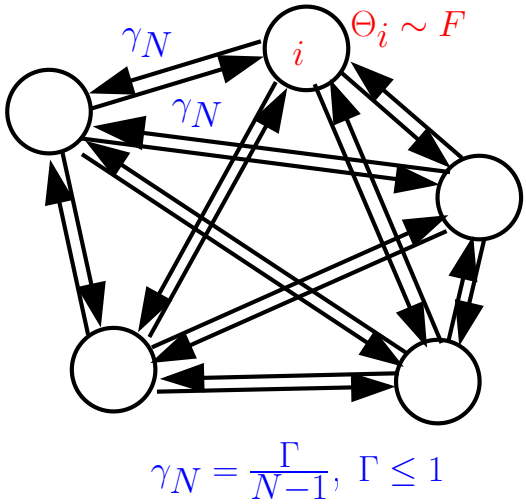


Figure 3.3: The HILT Network model

3.1.1 HILT Network model

In Chapter 2, we considered the Linear Threshold model on an arbitrary weighted directed graph. In this Chapter, in order to obtain a fluid limit, we propose a homogeneous version of the Linear Threshold model. Consider the population to be a social network $\mathcal{N}$ of N nodes where the graph is complete, and each edge carries the same weight. Also, let the thresholds Θ_j be chosen from an arbitrary threshold distribution with cumulative density function (c.d.f.) F (generalizing the uniform distribution assumption in Chapter 2). We call this the Homogeneous Influence Linear Threshold (HILT) network model, and the influence matrix $\mathbf{W}$ is such that for all $i \neq j$, $w_{i,j} = \gamma_N$.

Carrying over the assumption in [1]'s model, we will assume that $\gamma_N \leq \frac{1}{N-1}$ when

dealing with uniform threshold distribution. This is because, under uniform distribution, the maximum threshold is 1, and it does not make sense to consider influences greater than 1. In Section 3.5, when discussing various threshold distributions with unbounded support, we show that this restriction can be removed.

3.2 A Scaled Markov Chain and its Fluid Limit

Consider the HILT model on N nodes, and with edge weights $\gamma(N)$ such that $\lim_{N\to\infty} \gamma(N)N = \Gamma$, and the threshold distribution at the nodes given by F. In this section we will use Kurtz's theorem [88] to obtain a two dimensional o.d.e. that can serve as a fluid approximation for the evolution of the stochastic processes in the HILT model.

Let $A(k)$ and $D(k)$, respectively, be the *sizes* of the active and infectious sets at time k. In the HILT model, due to homogeneity, the precise membership of these sets is irrelevant and it is sufficient to keep track of set sizes. Instead of $A(k)$, we will work with $B(k) = A(k-1)$ to distinguish the active nodes that have exercised their influence, and the infectious nodes. Recall that, $B(k)$ is the size of the subset of active nodes that *have exercised their influence* by time k, whereas $D(k)$ is the size of the subset of active nodes at time k that *have not yet had a chance to exert their influence* on the inactive nodes.

A subtle but crucial point to note is that, in the original model, all nodes activated at time k ($j \in \mathcal{D}(k)$), will exert their influence at time $k+1$. Thus the two definitions of $\mathcal{D}(k)$ ("nodes activated at time k","subset of active nodes at time k who have not yet exerted their influence") are equivalent. However, as we shall see in Section 3.2.1, when nodes probabilistically delay the exertion of influence, these two definitions are no longer equivalent. Thus, from now on, $D(k)$ (or to be precise $D^N(k)$) represents the size of the subset of active nodes at time k that *have not yet exerted their influence*, irrespective of their times of activation.

3.2.1 A Scaled Markov Chain

It is easy to observe that $(B(k), D(k))$ is a discrete time Markov chain (DTMC) (see Appendix 3.8.1). In order to obtain an approximating o.d.e., we need to work with an appropriately scaled Markov process $(B^N(k), D^N(k))$, which can be thought of as evolving on a time scale N times faster than that of the original system. We can visualize this process as evolving over "minislots" of duration $1/N$, whereas the original process evolves at the epochs $0, 1, 2, \cdots$. Since this new process runs on a faster time scale, we need to slow

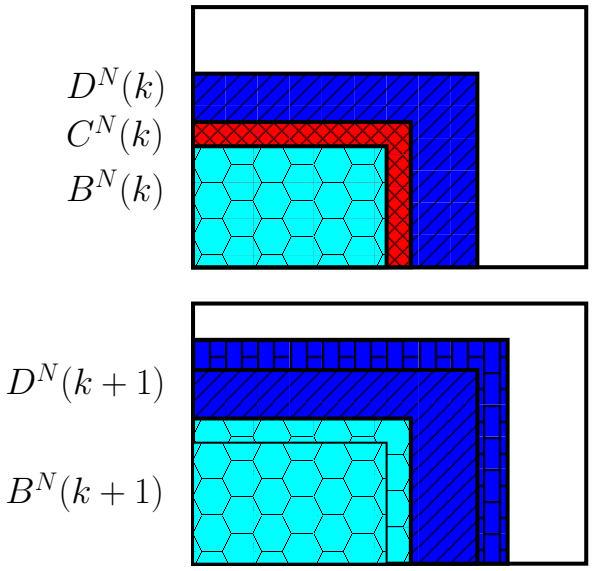


Figure 3.4: Evolution of the scaled process

down its dynamics. In each minislot, each node in $D^N(k)$ decides to spread its influence with probability $\frac{1}{N}$ or defer with probability $1 - \frac{1}{N}$. In the former case, it contributes its influence of γ and then moves to the set $B^N(k+1)$, else it stays in $D^N(k+1)$ set (see Figure 3.4). A similar scaling has been used in the context of the analysis of random multi-access algorithms by Bordenave et al. [97]. The reason for such a scaling is explained in Appendix 3.8.2, where we contrast it with the traditional amplitude and time scaling. The evolution of this process can be written as follows:

$$B^N(k+1) = B^N(k) + \frac{D^N(k)}{N} + Y^N(k+1)$$

$$\begin{aligned} D^N(k+1) &= \frac{N-1}{N} D^N(k) + Z^N(k+1) \\ &+ \frac{F(\gamma(B^N(k) + \frac{D^N(k)}{N})) - F(\gamma(B^N(k))}{1 - F(\gamma(B^N(k))} \times (N - B^N(k) - D^N(k)) \end{aligned}$$

where $Y^N(k+1)$ and $Z^N(k+1)$ are zero mean random variables. Dividing the evolution equations by N (the number of nodes in the network) and defining $\tilde{B}^N(k) = \frac{B^N(k)}{N}$, $\tilde{D}^N(k) = \frac{D^N(k)}{N}$, we can obtain the drifts for $\tilde{B}^N(k), \tilde{D}^N(k)$ for the fraction of nodes in each state.

$$\tilde{B}^N(k+1) = \tilde{B}^N(k) + \frac{\tilde{D}^N(k)}{N} + \tilde{Y}^N(k+1)$$

$$\begin{aligned}\tilde{D}^N(k+1) &= \frac{N-1}{N}\tilde{D}^N(k) + Z^N(k+1) \\ &+ \frac{F(\gamma(\tilde{B}^N(k) + \frac{\tilde{D}^N(k)}{N})) - F(\gamma(\tilde{B}^N(k))}{1 - F(\gamma(\tilde{B}^N(k))} \times (N - \tilde{B}^N(k) - \tilde{D}^N(k))\end{aligned}$$

Let $f_1^N(\tilde{B}^N(k), \tilde{D}^N(k))$ and $f_2^N(\tilde{B}^N(k), \tilde{D}^N(k))$ denote the mean drifts of $\tilde{B}^N(k), \tilde{D}^N(k)$. Consider the limiting drift function of $f_2^N(.)$ and observe that,

$$\begin{aligned}\lim_{N\to\infty} N\left(\frac{F(x+\frac{y}{N}) - F(x)}{1-F(x)}\right) &= \lim_{N\to\infty} \frac{y}{1-F(x)} \frac{F(x+\frac{y}{N}) - F(x)}{y/N} \\ &= \frac{yf(x)}{1-F(x)}\end{aligned}$$

Now consider $f_1(b,d) = d$ and $f_2(b,d) = \frac{f(\Gamma b)\Gamma d}{1-F(\Gamma b)}(1-b-d) - d$ and define,

$$f(b,d) := \Big(f_1(b,d), f_2(b,d)\Big)$$

Theorem 3.2.1. Given the Markov process $(\tilde{B}^N(k), \tilde{D}^N(k))$, we have for each $T > 0$ and each $\epsilon > 0$,

$$P\Big(\sup_{0\le t\le T} \big|\big|\big(\tilde{B}^N(\lfloor Nt\rfloor), \tilde{D}^N(\lfloor Nt\rfloor)\big) - \big(b(t), d(t)\big)\big|\big| > \epsilon\Big) \overset{N\to\infty}{\to} 0$$

where $(b(t), d(t))$ is the unique solution to the ODE,

$$\dot{b} = d$$

$$\dot{d} = \frac{f(\Gamma b)\Gamma d}{1-F(\Gamma b)}(1-b-d) - d$$

with initial conditions $(b(0) = 0, d(0) = a(0))$.

Proof. This is essentially an instance of Kurtz's theorem [88]; Also see [98]. In Appendix 3.8.3 we provide the statement of Kurtz's theorem for our context, and the verify the necessary conditions to guarantee the convergence of the Markov processes to the fluid limit o.d.e. □

We know that the hazard function corresponding to the c.d.f. $F(x)$ is given by

$$h_F(x) = \frac{f(x)}{1 - F(x)}$$

and hence the o.d.e. becomes,

$$\dot{b} = d \tag{3.2}$$

$$\dot{d} = h_F(\Gamma b)\Gamma d(1 - b - d) - d \tag{3.3}$$

Remark: It is necessary to re-emphasize the need for assuming homogeneity (i.e., identical nodes connected by a complete graph with identical edge weights). Under the homogeneous model, it can be seen that, it is sufficient to keep track of the size of each set $\mathcal{A}(k)$, $\mathcal{B}(k)$ and $\mathcal{D}(k)$, instead of the actual membership of the sets. This makes the analysis of influence dynamics tractable.

Remark: Though there have been several o.d.e. based models to study epidemics on networks [95], this is the first model which incorporates the threshold distribution (in the form of a hazard function) in the o.d.e. dynamics. This leads to a richer model, capable of producing a wide range of epidemic curves (as discussed in Section 3.5).

3.2.2 Accuracy of the o.d.e. approximation

Figure 3.5 shows the convergence of the scaled process to the o.d.e. with increasing network sizes $N = 50, 100, 500, 1000$ and for $\Gamma = 0.9$ and $d_0 = 0.2$. We observe that for $N = 1000$ the o.d.e. approximates the scaled process fairly well.

As noted in Appendix 3.8.2, the probabilistic scaling does not exactly replicate the original process. Hence, we have also compared the evolution of the original unscaled process for a fixed value of $N = 1000$, with the o.d.e approximation. The results are shown in Figure 3.6, where multiple sample paths of the original process (obtained by using different random number seeds) are plotted along with the (deterministic) o.d.e. solution. We find that the o.d.e. solution approximates the mean evolution of the original process well.

3.3 Uniform Threshold Distribution

In this section, we will consider the o.d.e. approximation of the HILT process, under the uniform distribution of threshold. The hazard function for uniform distribution is given

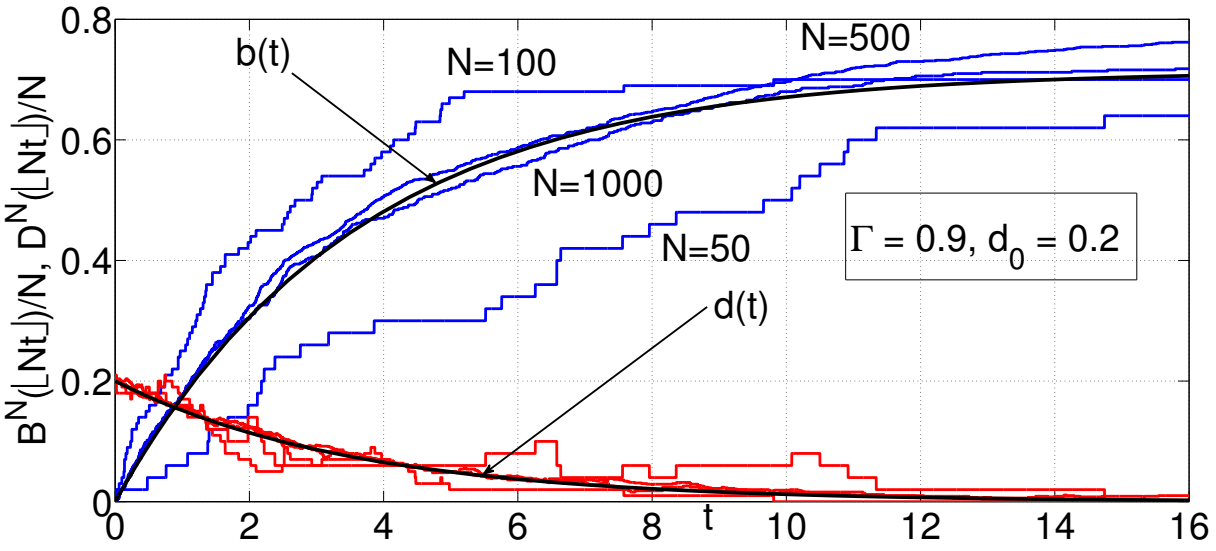


Figure 3.5: Trajectory of the fluid limit $(b(t), d(t))$ plotted along with samplepaths of the scaled process $\tilde{B}^N(\lfloor Nt \rfloor), \tilde{D}^N(\lfloor Nt \rfloor)$ for $N = 50, 100, 500, 1000$.

by $h_F(x) = \frac{1}{1-x}$ and thus the system of o.d.e. becomes,

$$\dot{b} = d$$

$$\dot{d} = -d + \frac{\Gamma d}{1 - \Gamma b}(1 - b - d)$$

It turns out that we can explicitly solve the above system, thus yielding closed form expressions for $(b(t), d(t))$. We will derive these closed form expressions and use these explicit expressions.

3.3.1 Solution to the o.d.e.

On solving the o.d.e. for uniform distribution, with initial conditions $b(0) = 0, d(0) = d_0$ and defining $r = 1 - \Gamma + \Gamma d_0$, we get

$$b(t) = \frac{d_0}{r} - \frac{d_0}{r} e^{-rt}$$

$$d(t) = d_0 e^{-rt}$$

In Appendix 3.8.4 we provide the steps involved in obtaining the solution. From the above equations we can state the following theorem:

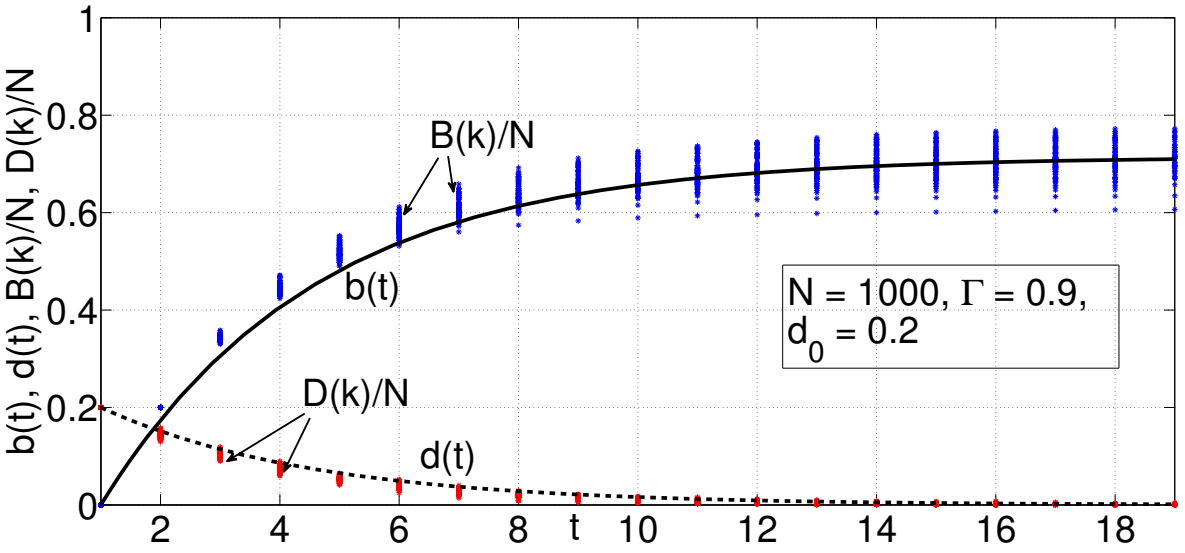


Figure 3.6: Trajectory of the fluid limit $(b(t), d(t))$ plotted along with multiple runs of the original process (normalized) $(\frac{B(k)}{N}, \frac{D(k)}{N})$ for $N = 1000$.

3.3.2 Terminal spread of influence

The following theorem results from a simple observation of the o.d.e.'s.

Theorem 3.3.1. Given that we start with d_0 fraction of nodes in the infectious set in an HILT network with parameter Γ, then the final fraction of activated nodes will be $\frac{d_0}{r}$ where $r = 1 - \Gamma + \Gamma d_0$.

Remarks:

- We might also be interested in the question of choosing the right d_0 which can give us the required b_∞, and we see that

$$d_0 = \frac{b_\infty(1-\Gamma)}{1 - b_\infty\Gamma}$$

- We observe that, for large N, as long as $\Gamma < 1$ we cannot influence the entire population (i.e., $b_\infty = 1$) unless we start off with the entire population active (i.e., $d_0 = 1$). But if $\Gamma = 1$ then $b_\infty = 1$ provided $d_0 > 0$.

Consider the discrete influence process (the Kempe model [1]), and let $\sigma^{(\mathcal{N},\mathcal{A}_0)} = \mathbb{E}^{(\mathcal{N},\mathcal{A}_0)}[|A_U|]$ be the expected size of the *terminal set* A_U, starting with $\mathcal{A}_0$ as the initial set in the network $\mathcal{N}$. Since all initial sets are equivalent in the HILT model, we will be interested in the influence of a set of size m. Define $h_\gamma^{(N)}(m) := \sigma^{(\mathcal{N},\mathcal{A})}$, for all $\mathcal{A}$ of size m.

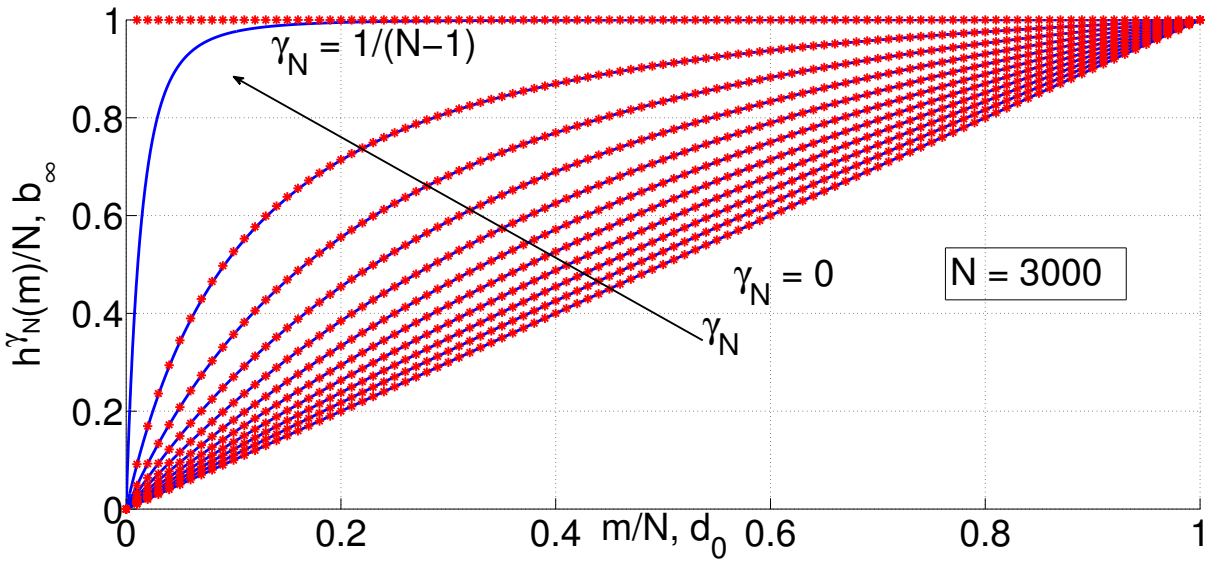


Figure 3.7: $h_\gamma^{(N)}(k)$ versus k for $N = 3000$ (shown by solid lines) and b_∞ versus d_0 (shown by asterisks) for various values of γ_N.

By using results from Chapter 2, we can show that,

$$h_\gamma^{(N)}(m) = m[1 + (N-m)\gamma[1 + (N-m-1)\gamma[1 + \cdots$$

The behaviour of $h_\gamma^{(N)}(m)$ as a function of γ_N and m can be seen in Figure 3.7 (depicted by solid lines), for a network of 3000 nodes. We also superimpose the behavior of b_∞ against d_0 (depicted by asterisks). We observe that there is an exact match, except for $\Gamma = 1$. For $\Gamma = 1$, as seen earlier, we know that $b_\infty = 1$ as long as $d_0 > 0$. This is however true only in the fluid limit, and hence the discrepancy for finite N.

Taking $\Gamma = N\gamma$ and $d_0 = \frac{m}{N}$, we can show that as $N \to \infty$, $\frac{h_\gamma^{(N)}(m)}{N} \to b_\infty$. See Appendix 3.8.5 for the proof. This provides another verification of the accuracy of the o.d.e. approximation for large N.

3.4 Time constrained optimization

While the analytical expression $h_\gamma^N(m)$ derived earlier for HILT gives only the expected size of the terminal set, the o.d.e. dynamics approximates the trajectory of influence evolution, for large N. This can be useful, especially in problem settings where the time taken by the process for the spread of influence is also considered, in addition to the size of the initial set.

Theorem 3.4.1. Given the initial fraction of infected nodes d_0 in an HILT network with parameter Γ, the time we have to wait to get at least α ($\alpha < \frac{d_0}{r}$) fraction of nodes active is given by,

$$T(\alpha, d_0, \Gamma) = \frac{1}{r} \ln \left(\frac{1-r}{1-\frac{\alpha}{d_0} r} \right)$$

where $r = 1 - \Gamma + \Gamma d_0$.

Proof. Firstly, note that since $a_\infty = b_\infty = \frac{d_0}{r}$, α must be less than $\frac{d_0}{r}$. Since we are observing the process at a finite time T, $d(T)$ is not zero. Hence, we should look at the value of $a(T) = b(T) + d(T)$ and set it to α. We get,

$$a(T) = d_0\big(\frac{1}{r} - (\frac{1}{r} - 1)e^{-rT}\big) = \alpha$$

Rearranging terms,we get the expression for $T(\alpha, d_0, \Gamma)$. □

A more interesting question would be to determine the d_0 to be chosen so that by time T we will have at least α fraction of the nodes activated, in the HILT network with parameter Γ. Unfortunately, we will not be able to get a closed form expression for this, and it can be solved numerically using the following fixed point equation.

$$e^{-rT} = \frac{1 - \frac{\alpha}{d_0} r}{1 - r}$$

We can use the *iterative bisection method* obtain the fixed point of the above equation. Let $F(d_0) = e^{-rT}$ and $G(d_0) = \frac{1-\frac{\alpha}{d_0} r}{1-r}$. We know that $d_0^\star$ that solves $F(d_0) = G(d_0)$ will lie in $[\frac{\alpha(1-\Gamma)}{1-\alpha\Gamma}, 1]$ and that the solution is unique, since $a(T)$ is a monotonic function in d_0. We also know that for $d_0 < d_0^\star$, $F(d_0) > G(d_0)$ and for $d_0 > d_0^\star$, $F(d_0) < G(d_0)$.

Under the above conditions, we find that the bisection method will converge to $d_0^\star$. This is shown as Algorithm 3. The method is illustrated in Figure 3.8 for parameters $\Gamma = 0.8$, $\alpha = 0.7$, $T = 15$.

The variation of $d_0^\star$ with respect to the parameters α, Γ and T can be seen in Figures 3.9,3.10,3.11.

In Figure 3.9, note that for $\Gamma = 0$, as expected, $d_0 = \alpha$, i.e., since there is no social interaction ($\Gamma = 0$), our terminal spread of influence will be equal to the initial seeding. Also, note that as the target time T is reduced, we require higher values of d_0 to achieve the same α (for $T = 0$, $d_0 = \alpha$). For $\Gamma = 1$, d_0 asymptotically approaches 0 for large T. From Figure 3.10, we see that as α increases, for a given T, the required d_0 monotonically increases.

$d_0^{min} = \frac{\alpha(1-\Gamma)}{1-\alpha\Gamma}$;
$d_0^{max} = 1$;
while *1* **do**
  $x = (d_0^{min} + d_0^{max})/2$;
  if $F(x) - G(x) > 0$ **then**
    $d_0^{min} = x$;
  else
    $d_0^{max} = x$;
  end
  if $|F(x) - G(x)| < \epsilon$ **then**
    break;
  end
end
$d_0^\star = x$;

Algorithm 3: Iterative Bisection method

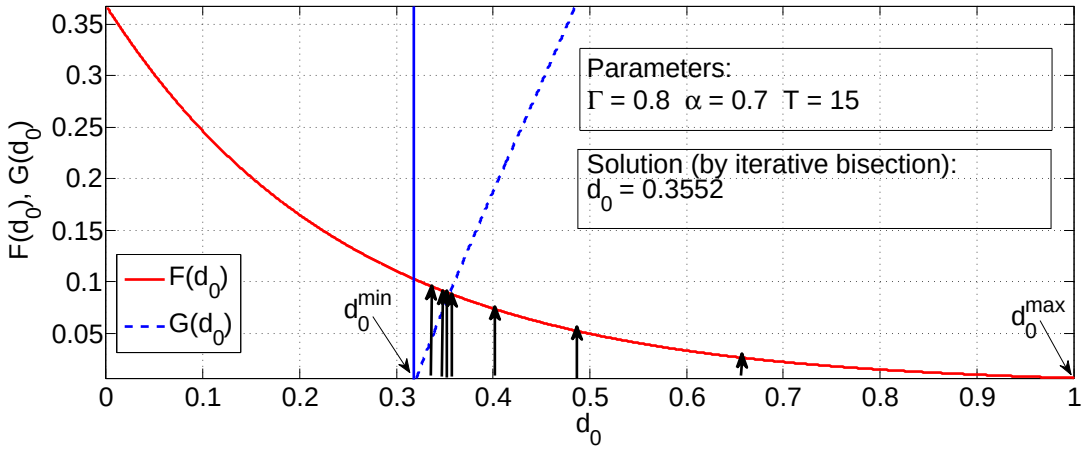


Figure 3.8: Evaluating d_0 by Iterative Bisection Method

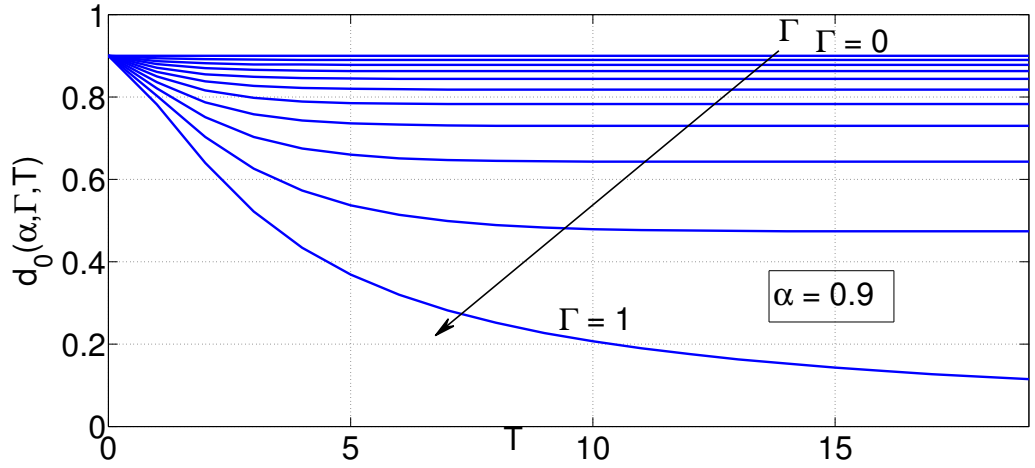


Figure 3.9: Variation of $d_0^\star$ across T for various values of Γ with $\alpha = 0.9$

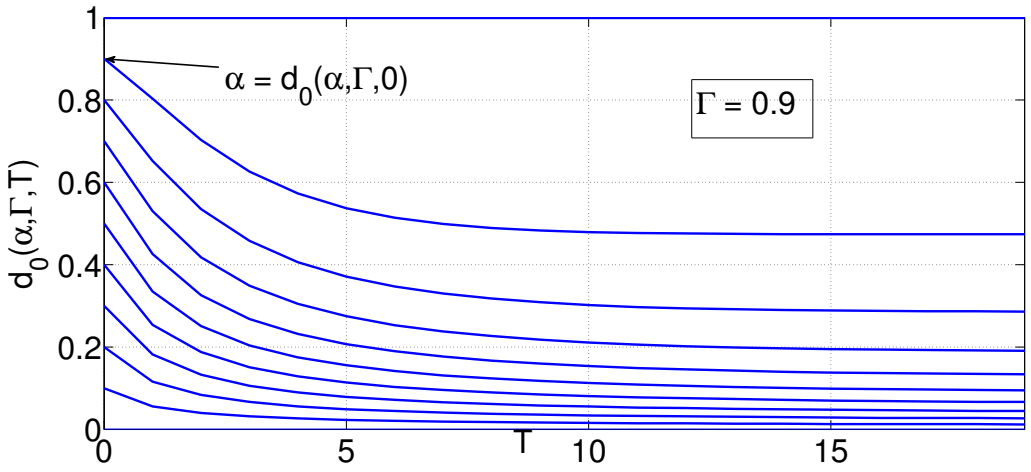


Figure 3.10: Variation of $d_0^\star$ across T for various values of α with $\Gamma = 0.9$

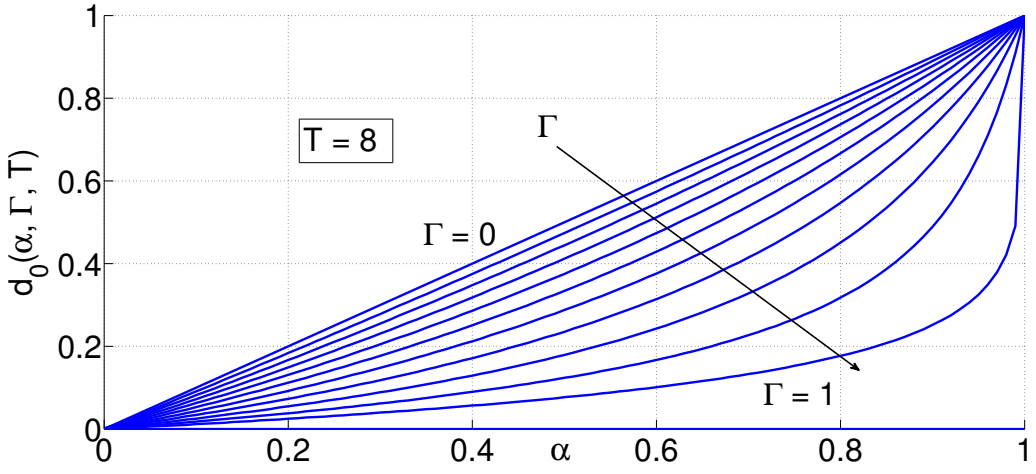


Figure 3.11: Variation of $d_0^\star$ across α for various values of Γ at $T = 8$

Finally, Figure 3.11, shows that the d_0 `vs` α behavior for $T = 8$ is qualitatively similar to the one in Figure 3.7 depicting d_0 `vs` b_∞.

3.5 Effect of the Threshold distribution

Recall that the evolution of influence is given by:

$$\dot{b} = d$$

$$\dot{d} = h_F(\Gamma b)\Gamma d(1 - b - d) - d$$

Note that the evolution depends on the distribution of threshold via its hazard function,

$$h(x) = \frac{f(x)}{1 - F(x)}$$

where $f(\cdot)$ and $F(\cdot)$ are the probability density and cumulative distribution functions of the threshold distribution, respectively. Hazard functions are widely used in failure/survival analysis. In this section, we will consider threshold distributions with different hazard function characteristics, and study the spread of influence.

As indicated earlier, the o.d.e. derived is valid for any $\Gamma > 0$, and in this Section, we will also consider cases when $\Gamma > 1$, while discussing threshold distributions with unbounded support. However, for uniform threshold distribution, we will restrict $\Gamma \leq 1$, since under this case $h_F(x) = \frac{1}{1-x}$, valid only for $x \in [0, 1]$.

3.5.1 Exponential distribution

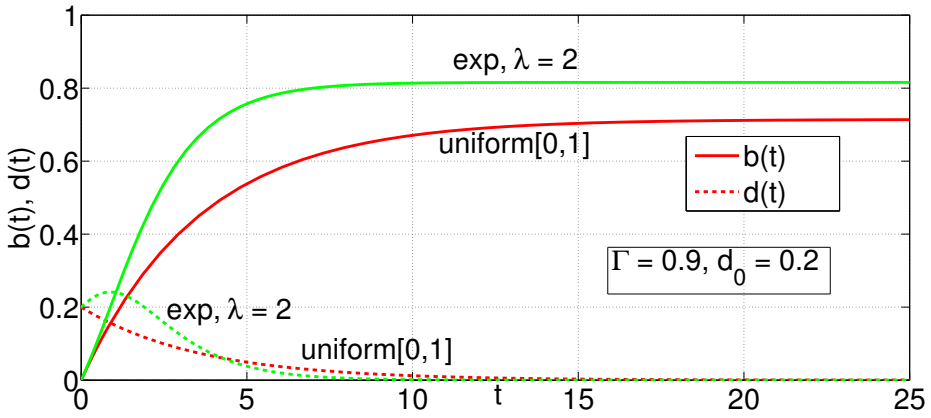
Exponential distribution is widely used in scenarios where there is need for a constant hazard rate. This is also due to the fact that exponential distribution is the only memoryless continuous distribution. Consider the threshold θ_i distributed as exponential with parameter λ. We have

$$f(x; \lambda) = \lambda e^{-\lambda x},\ x \geq 0$$

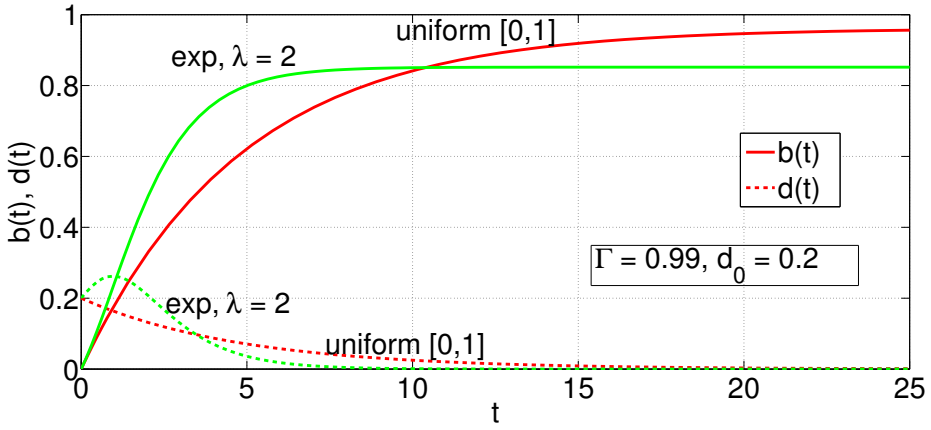
$$F(x; \lambda) = 1 - e^{-\lambda x},\ x \geq 0$$

Thus we get $h_F(x) = \lambda$. Plugging this in the o.d.e. expression we get,

$$\dot{b} = d$$



(a) small Γ regime



(b) large Γ regime

Figure 3.12: Comparison of influence spread between Uniform threshold distribution and Exponential threshold distribution with the same mean. We still use $\Gamma < 1$, since we are dealing with the uniform distribution

$$\dot{d} = -d + \lambda\Gamma d(1 - b - d)$$

Observe that the above system of o.d.e. is equivalent to the dynamics of an SIR *(Susceptible-Infective-Recovered)* epidemic, with infection rate $\lambda\Gamma$ and recovery rate 1 [95]. The $b(t)$ and $d(t)$ processes respectively are equivalent to the *Recovered* and *Infective* processes of the SIR epidemic model. Thus we see that the under exponential distribution of threshold, the Linear Threshold model, in its fluid limit, is equivalent to a special case of the SIR model. This equivalence provides a hitherto undocumented link between influence spread models from viral marketing literature (Linear Threshold model) and a traditional epidemic model (SIR model).

Figures 3.5.1 and 3.5.1 compare the influence evolution under uniform and exponential distribution of threshold. Note that for the same mean threshold ($\mathbf{E}\theta = 0.5$) and smaller value of Γ (Figure 3.5.1), exponential case yields a larger terminal influence spread. This is because, under the exponential distribution, there are more nodes with threshold close to zero. This also explains the steeper increase of $b(t)$ for exponential distribution compared to the uniform distribution case. In fact, from the respective o.d.e.s it is clear that $\dot{a}(0) = \dot{b}(0) + \dot{d}(0)$ for uniform distribution, is half that of exponential distribution with the same mean.

But, for larger values of Γ (Figure 3.5.1), uniform distribution yields a larger terminal influence spread. This is because, in the uniform case, the thresholds are bounded above by 1, while in the exponential case, the support set for thresholds is unbounded. Thus, under the uniform distribution, as Γ approaches 1, the terminal spread of influence approaches 1 (as noted in Section 3.3.2).

3.5.2 Weibull distribution

Another distribution which is widely used in survival analysis is the Weibull distribution. The probability density function of a Weibull random variable is given by,

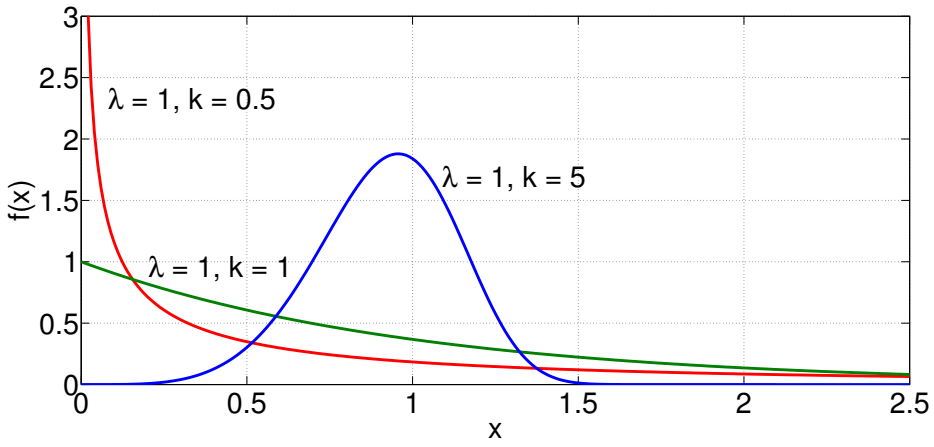
$$f(x : \lambda, k) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{(k-1)} e^{-(\frac{x}{\lambda})^k},\ x \geq 0$$

If the random variable X is the time to failure, then under the Weibull distribution, the failure rate is proportional to a power of time. The hazard function is given by,

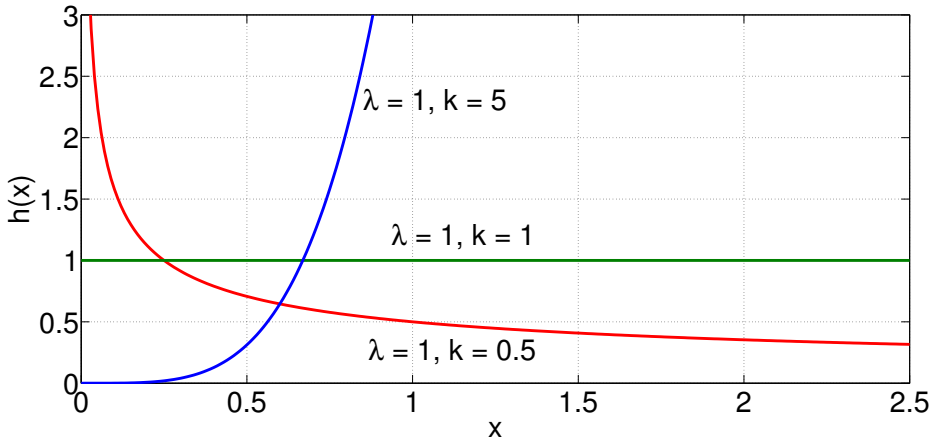
$$h(x; \lambda, k) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{(k-1)},\ x \geq 0$$

In the above expression λ is often referred to as the *scale* parameter and k is referred to as the *shape* parameter. Figure 3.5.2 shows the probability density function of Weibull distribution for various values of k. Note that for $k > 1$, there are significantly high number of users with higher values of threshold, i.e., less susceptible to the spread of influence. The hazard rate for Weibull distribution can be increasing, constant or decreasing depending on the value of k. This is demonstrated in the Figure 3.5.2.

- $k < 1$ leads to decreasing hazard rate. This implies that nodes are less likely to become activated by an instantaneous influence, as the existing influence (which



(a) Probability density function



(b) Hazard function

Figure 3.13: Weibull distribution for different values of k

failed to activate the node) on them increases.

- $k = 1$ yields constant hazard rate, and in that case Weibull distribution is just the exponential distribution.

- $k > 1$ yields an increasing hazard rate, which implies nodes are more likely to become activated by an instantaneous influence, as the existing influence on them increases.

The HILT o.d.e. under Weibull distribution of threshold can be written as follows:

$$\dot{b} = d$$

$$\dot{d} = -d + \Gamma d \frac{k}{\lambda} \left(\frac{\Gamma b}{\lambda} \right)^{k-1} (1 - b - d)$$

Figures 3.5.2 and 3.5.2 demonstrate the evolution of the o.d.e under the Weibull distribution of threshold, for different values of k in the small and large regimes for Γ. For smaller Γ (Figure 3.5.2), we observe that as k increases, the spread of influence decreases. This is expected, since from Figure 3.5.2 it is clear that, for larger values of k, Weibull distribution puts more mass on larger values of threshold, i.e., nodes are less susceptible to influence. Further, for $k = 5$, Figure 3.5.2 shows that the total spread of influence is 0.2, equal to the initial seeding $d_0 = 0.2$. This implies the influence does not spread at all, since the node thresholds are much higher, compared to the net influence generated by d_0 (due to smaller Γ).

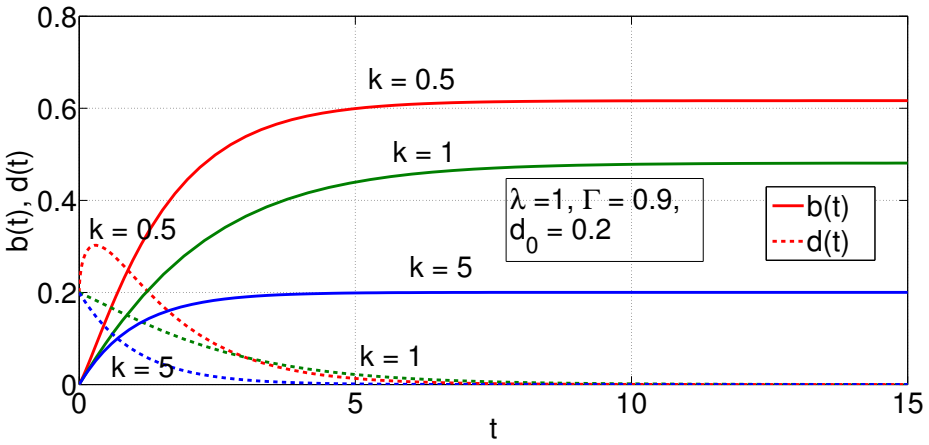
For larger Γ (Figure 3.5.2), we see that the trend is reversed, i.e., as k increases, the spread of influence increases. It is to be noted that the $\dot{b} = d$ near 0 is larger for smaller k, similar to the small Γ regime. However, from Figures 3.5.2 and 3.5.2 we see that smaller values of k have heavier tails (and lower hazard rates), thus leading to stagnation of influence after the initial surge.

Another interesting feature to note is that, unlike the small Γ regime, for $k = 5$, we get a much higher influence spread. Also, unlike other values of k, here $\dot{b} = d$ exhibits a non-monotonic behavior even after it begins to decrease, i.e., $d(t)$ is not unimodal. Such behavior has not been observed until now in the classic epidemiology framework, especially in a homogeneous setting. In traditional epidemic models like *SIR*, the I process (equivalent to $\dot{b}$) might exhibit an initial increase, but once it begins decreasing, continues to steadily decrease to zero. But, in our dynamics, the presence of hazard rate (increasing, in this case) leads to such non-unimodal characteristics of $d(t)$.

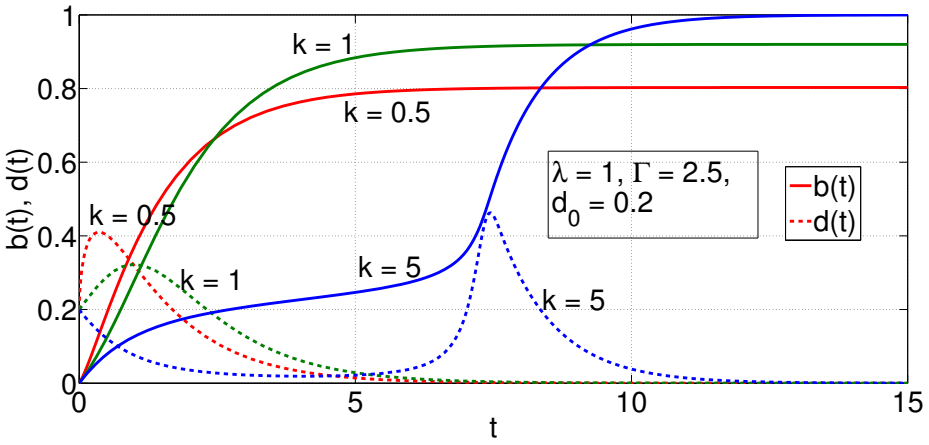
3.5.3 Loglogistic distribution

Loglogistic distribution is the probability distribution of a random variable who logarithm follows the logistic distribution. It has similar shape characteristics to log-normal distribution, but has heavier tails. The probability density function and the hazard function are given by,

$$f(x; \alpha, \beta) = \frac{(\frac{\beta}{\alpha})(\frac{x}{\alpha})^{\beta-1}}{(1 + (\frac{x}{\alpha})^{\beta})^2}, \; x \geq 0$$



(a) small Γ regime



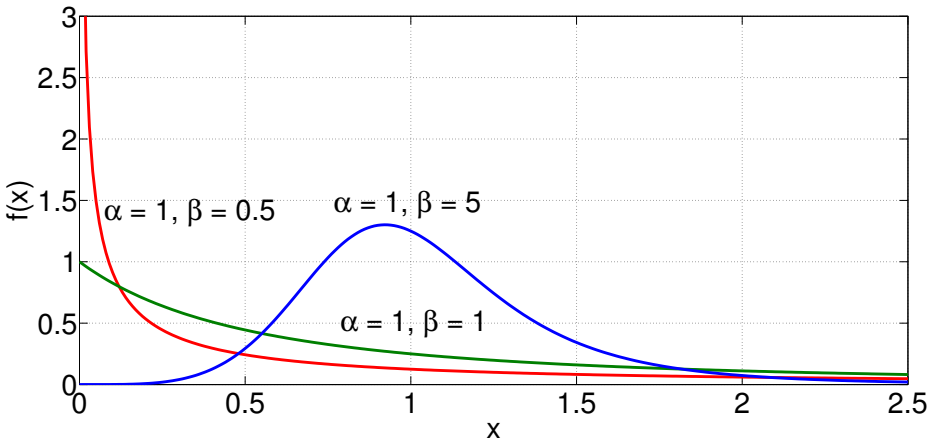
(b) large Γ regime

Figure 3.14: Comparison of influence spread between Weibull threshold distributions with different values of k

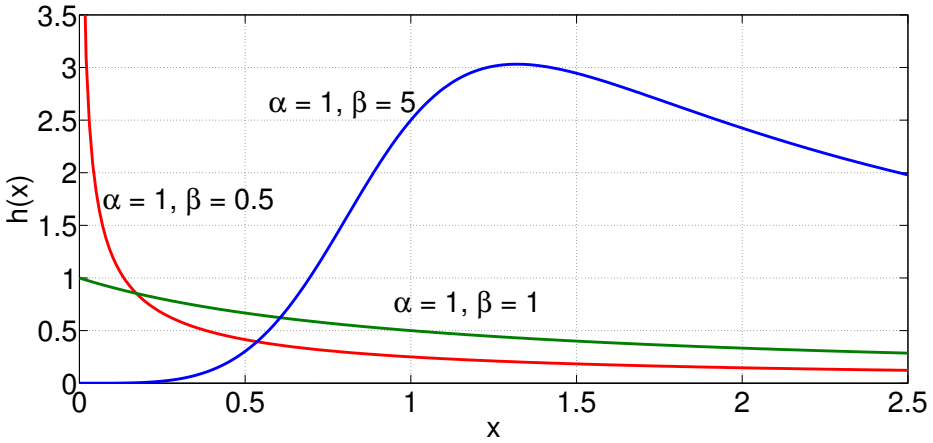
$$h(x;\alpha,\beta) = \frac{\beta}{\alpha}\left[\frac{(\frac{x}{\alpha})^{\beta-1}}{1+(\frac{x}{\alpha})^{\beta}}\right],\ x \geq 0$$

The parameter α functions as the scale parameter and β is referred to as the shape parameter. Also for $\beta > 1$, the distribution is unimodal, and is more concentrated as β increases (see Figure 3.5.3).

Similar to the Weibull distribution, one can obtain different failure characteristics by tuning the β parameter. For $\beta \leq 1$, the hazard rate decreases monotonically. But unlike the Weibull distribution, for $\beta > 1$, the hazard function exhibits non-monotonic behavior (see Figure 3.5.3).



(a) Probability density function



(b) Hazard function

Figure 3.15: loglogistic distribution for different values of k

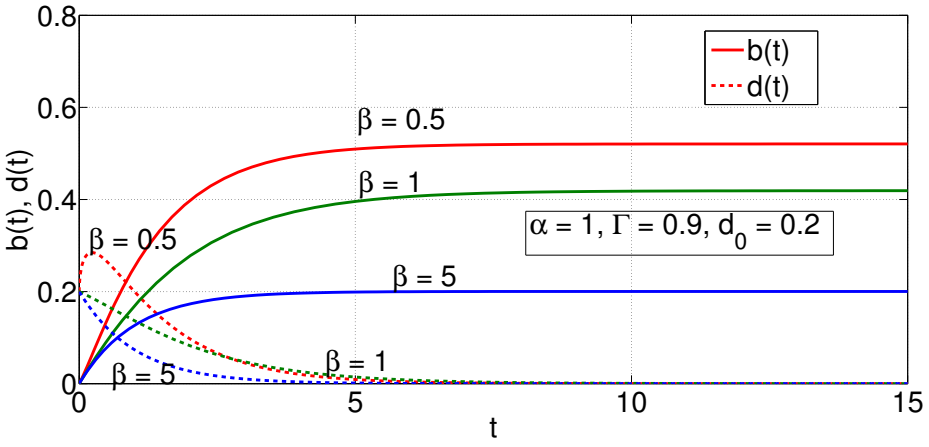
The HILT o.d.e. under loglogistic distribution of threshold can be written as follows:

$$\dot{b} = d$$

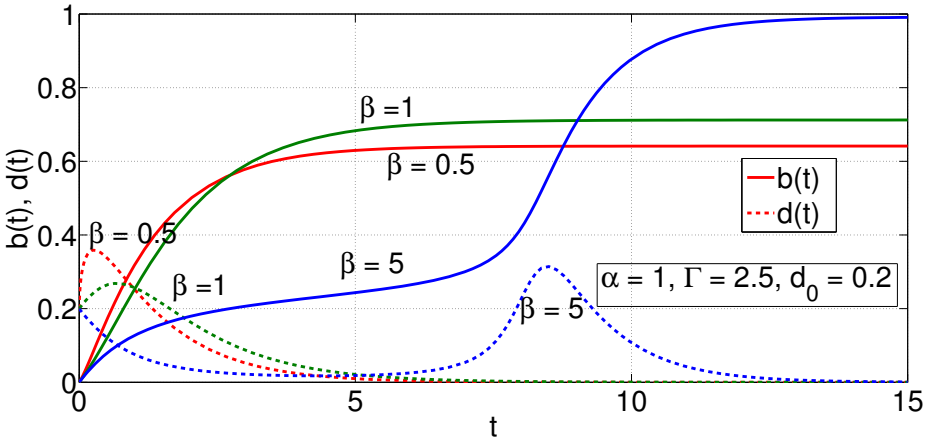
$$\dot{d} = -d + \Gamma d \frac{\beta}{\alpha} \left[\frac{(\frac{\Gamma b}{\alpha})^{\beta-1}}{1 + (\frac{\Gamma b}{\alpha})^{\beta}} \right] (1 - b - d)$$

Figures 3.5.3 and 3.5.3 demonstrate the evolution of the influence spread o.d.e. under the loglogistic distribution of threshold, for different values of k in the small and large regimes for Γ. We note that for small Γ, the evolution of influence is qualitatively similar, but under the loglogistic distribution, we get a smaller influence spread, due to heavier

tails. Also, in the large Γ regime, we note that for $\beta = 5$, we again get a non-unimodal behavior for $d(t)$. But the second peak is less pronounced in the loglogistic distribution than the Weibull distribution, since the loglogistic distribution exhibits a non-monotonic hazard rate.



(a) small Γ regime



(b) large Γ regime

Figure 3.16: Comparison of influence spread between loglogistic threshold distributions with different values of k

Thus we see that, the incorporation of hazard rate into the o.d.e. (resulting from a fluid limit characterization of the LT model) yields qualitatively different characteristics compared to the standard epidemic models. To the best of our knowledge, this is the first work that analytically characterizes the evolution of influence under different threshold distributions. This is also the first work to incorporate hazard functions into the epidemic models, thus providing a way to capture heterogeneity in the population. Further, due

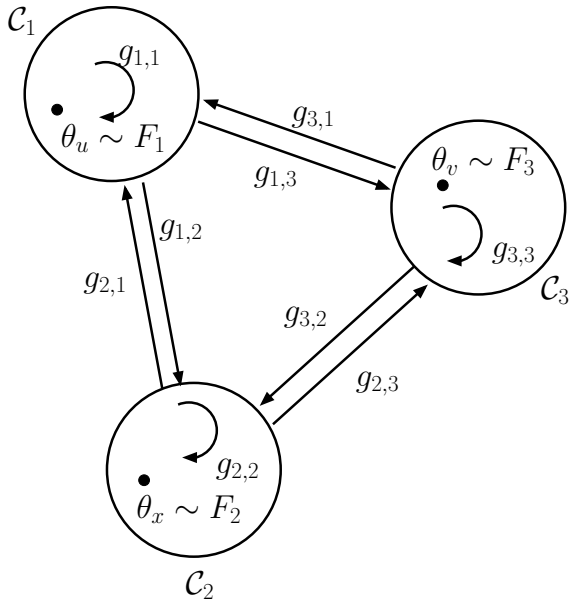


Figure 3.17: A heterogeneous network with three communities, shown with entries of the influence matrix $\mathcal{G}$. Nodes u, x and v belong to communities $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$, and have their thresholds distributed according to $F_1(\cdot)$, $F_2(\cdot)$ and $F_3(\cdot)$ respectively.

to the one-one correspondence between a given hazard function and its corresponding cumulative distribution [87], one can begin with the hazard function in the o.d.e. (obtained by curve-fitting to existing epidemic data) and ascertain the threshold distribution of the population.

3.6 Multiclass HILT model

A natural extension to the HILT model would be to consider the evolution of information spread in an heterogeneous network. Such a scenario might arise in a network with communities, where the interactions within a community might be stronger than the interaction across communities. These have been traditionally studied under the term *stratified epidemics* [99]. Consider a network with M communities $(\mathcal{C}_i)_{i=1}^M$ and let $(N_i)_{i=1}^M$ denote the number of nodes in each community. Let $\mathcal{G}$ be the influence matrix, whose entries $g_{i,j}$ indicates the strength of influence from community i to community j (see Figure 3.17).

As earlier, we will appropriately normalize the edge weights, i.e., for $u \in \mathcal{C}_i$ and $v \in \mathcal{C}_j$, $w_{u,v} = \frac{g_{i,j}}{N}$, where $N = \sum_i N_i$ is the total population size. Let the nodes within community i have their thresholds distributed according to F_i, with hazard function h_{F_i}. We can then carry out an analysis similar to what was done for the HILT model in Section 3.2. We can show that the joint evolution is a Markov process, and we construct a scaled process

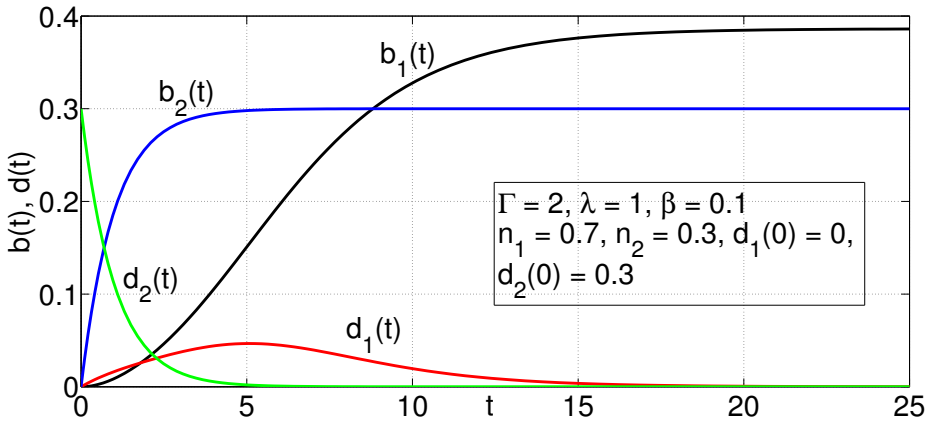


Figure 3.18: Evolution of influence in the two communities when the initial seeding is done in the smaller community (i.e., $d_1(0) = 0, d_2(0) = d_0 = 0.3$)

using the minislots approach and with appropriate probability scaling. Here the attempt probability of all infectious nodes during a given mini-slot scales as $\frac{1}{N}$, irrespective of which community they belong to. By applying Kurtz's theorem to the scaled process, we obtain the o.d.e.s representing the influence evolution. Let $(b_i(t), d_i(t))$ denote the non-infectious and infectious active nodes within community i. We can then describe their evolution by the following system of o.d.e.s similar to Equations 3.2 and 3.3 ($1 \leq i \leq M$):

$$\dot{b}_i = d_i$$

$$\dot{d}_i = -d_i + [\mathcal{G}^T \mathbf{d}]_i \; h_{F_i}([\mathcal{G}^T \mathbf{b}]_i)(n_i - b_i - d_i)$$

where $\mathbf{b} = (b_1, b_2, \cdots b_m)^T$, $\mathbf{d} = (d_1, d_2, \cdots d_m)^T$ and $n_i = \lim_{N\to\infty} \frac{N_i}{N}$. One possible objective function to maximize in this scenario would be the total spread of influence $\sum_i b_i(\infty)$, by suitably choosing the initial $\mathbf{d}(0)$ subject to the constraint $\sum_i d_i(0) = d_0$, for fixed system parameters, i.e., the threshold distributions F_i and the influence matrix $\mathcal{G}$. We were unable to obtain a universal analytical solution for this problem, but numerically demonstrate that depending the system parameters the results could be quite counter-intuitive.

For instance, consider a two community network with $N_1 = 0.7N$ and $N_2 = 0.3N$ as the relative community sizes. Let all the nodes in the population have their thresholds distributed according to an exponential distribution with parameter λ. Also assume that $g_{i,i} = \Gamma$ and $g_{i,j} = \beta$ for $i, j \in \{1, 2\}$. Figures 3.18 and 3.19 show the evolution of $(b_i(t), d_i(t))_i$ for $i = 1, 2$, for different initial conditions. While in Figure 3.18 (scenario 1) the entire initial seeding is done in the smaller community (i.e., $d_1(0) = 0, d_2(0) = d_0 = 0.3$

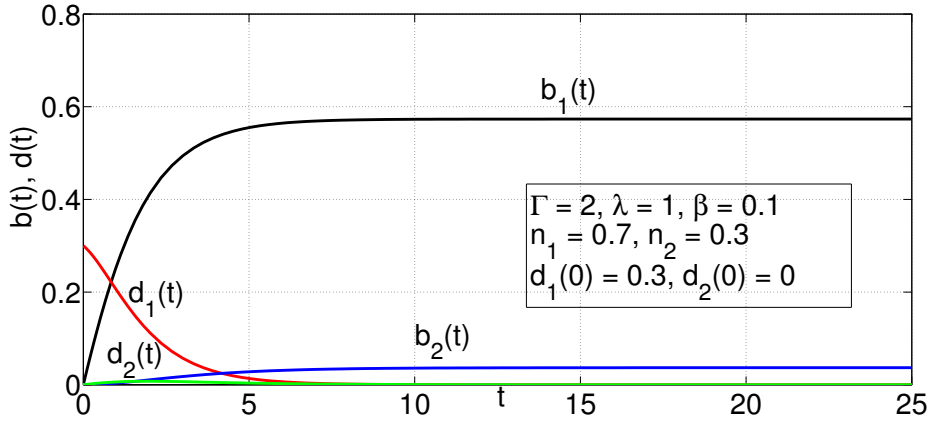


Figure 3.19: Evolution of influence in the two communities when the initial seeding is done in the larger community (i.e., $d_1(0) = d_0 = 0.3, d_2(0) = 0$)

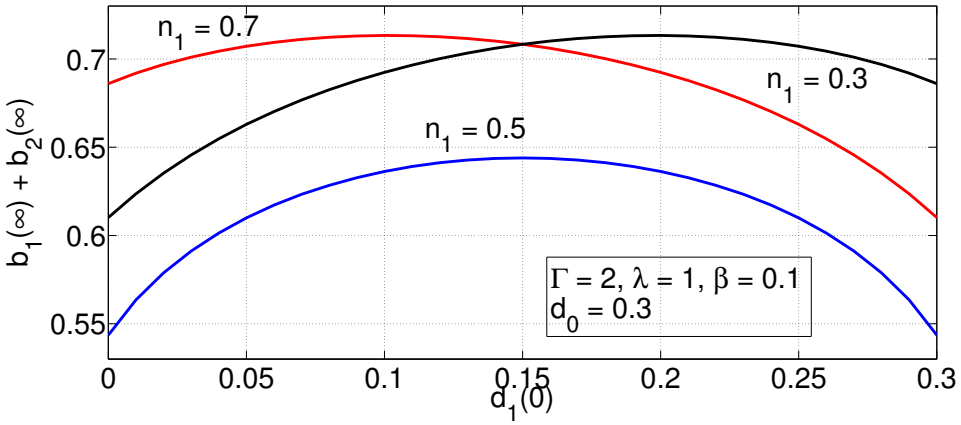


Figure 3.20: Total spread of influence $b_1(\infty) + b_2(\infty)$ for various allocations of initial seeding $(d_1(0), d_2(0))$ and different relative community sizes.

), in Figure 3.19 (scenario 2) the entire initial seeding is done in the larger community. We see that, the total spread of influence in scenario 1 is larger than in scenario 2. Further from Figure 3.20 it is clear that the optimal seeding for this setting is approximately $(d_1(0) = 0.1, d_2(0) = 0.2)$. It is surprising that we get a wider spread of influence by investing more in the smaller community. Thus we see that, even in a simple two community setting, the optimal seeding might be counter-intuitive. It would be an interesting future direction to analytically obtain the optimal seeding, given the influence matrix $\mathcal{G}$ and the threshold distributions $\{F_i\}_{1 \leq i \leq m}$.

3.7 Summary

In this chapter, we began with a homogeneous version of the Linear Threshold model proposed by Kempe et al. [1] in the context of viral marketing, and generalized it for arbitrary threshold distributions. We observed that the spread of influence evolves as a discrete time Markov chain. Under a certain scaling, we showed that the scaled Markov chain converges (in the sense of [88]) to a deterministic trajectory defined by an o.d.e.. The threshold distribution appears in terms of its hazard rate function in this o.d.e. We described how this approach complements the fixed point equation suggested by Granovetter [16], thus providing a link between two threads in the threshold model literature. Also, under the exponential distribution of threshold, we showed that the derived fluid dynamics are equivalent to the well-known SIR model in epidemiology. We also numerically demonstrated how incorporating the hazard function into the o.d.e. can provide qualitatively different characteristics compared to traditional epidemic models, even in a homogeneous setting.

3.8 Appendix

3.8.1 DTMC $(B(k), D(k))$

Let $\mathcal{F}_k$ denote the entire history of the processes up to time k, i.e., $\mathcal{F}_k = (B(l), D(l))_{l=0}^{l=k}$. To obtain the expected drift of $D(k)$, consider,

$$\begin{aligned}
&\mathbb{P}(D_{k+1} = l|\mathcal{F}_k) \\
&= \mathbb{P}(\sum_{j\notin B_k} I\{b_j(B_k) < \theta_j \leq b_j(B_{k+1})\} = l|\mathcal{F}_k) \\
&= \sum_{\substack{L\subseteq N\setminus B_{k+1} \\ |L|=l}} \prod_{j\in L} \mathbb{P}(I\{b_j(B_k) < \theta_j \leq b_j(B_{k+1})\} = 1|\mathcal{F}_k) \times \\
&\qquad \prod_{j\notin L} \mathbb{P}(I\{b_j(B_k) < \theta_j \leq b_j(B_{k+1})\} = 0|\mathcal{F}_k)
\end{aligned}$$

Let $|B_k| = b$ and $|D_k| = d$, then we have in the HILT model $b_j(B_k) = \Gamma b$ and $b_j(B_{k+1}) = \Gamma(b+d)$. Hence we can write,

$$
\begin{aligned}
&\mathbb{P}(D_{k+1} = l|\mathcal{F}_k) \\
&= \sum_{L\subseteq N\setminus B_{k+1}, |L|=l} \prod_{j\in L} \mathbb{P}(\theta_j \leq \Gamma(b+d)|\theta_j > \Gamma b) \times \\
&\qquad \prod_{j\notin L} \mathbb{P}(\theta_j > \Gamma(b+d)|\theta_j > \Gamma b) \\
&= \sum_{L\subseteq N\setminus B_{k+1}|L|=l} \left(\frac{F(\Gamma(b+d)) - F(\Gamma b)}{1-F(\Gamma b)}\right)^l \times \\
&\qquad \left(1 - \frac{F(\Gamma(b+d)) - F(\Gamma b)}{1-F(\Gamma b)}\right)^{(N-b-d-l)} \\
&= \binom{N-b-d}{l} \left(\frac{F(\Gamma(b+d)) - F(\Gamma b)}{1-F(\Gamma b)}\right)^l \times \\
&\qquad \left(1 - \frac{F(\Gamma(b+d)) - F(\Gamma b)}{1-F(\Gamma b)}\right)^{(N-b-d-l)} \\
&= P(D(k+1) = l|B(k) = b, D(k) = d)
\end{aligned}
$$

From the above equations we can clearly see that $(B(k), D(k))$ is a DTMC on the state space $[0, 1, \cdots, N] \times [0, 1, \cdots, N - B(k)]$.

3.8.2 Scaling the HILT Model

In this section, we will demonstrate the necessity for a probabilistic scaling (in addition to the amplitude and time scaling) to arrive at the mean drift expressions. Let $\mathcal{F}_k$ denote the entire history of the processes up to time k, i.e., $\mathcal{F}_k = (B(l), D(l))_{l=0}^{l=k}$. Begin with the drift equations for the unscaled process $(B(k), D(k))$.

$$
\mathbb{E}\Big[B(k+1) - B(k)|\mathcal{F}_k\Big] = D(k)
$$

$$
\begin{aligned}
&E\Big[D(k+1) - D(k)|\mathcal{F}_k\Big] \\
&= -D(k) + \\
&\quad \frac{F(\gamma(B(k)+D(k))) - F(\gamma B(k))}{1-F(\gamma B(k))} \times \\
&\qquad (N - D(k) - B(k))
\end{aligned}
$$

We shall now try scaling the process in the usual way, i.e.,by scaling down the amplitude by a factor of N, $\tilde{B}^N(k) = \frac{B(k)}{N}$, $\tilde{D}^N(k) = \frac{D(k)}{N}$. The evolution equations can then be written down as follows:

$$E\left[\tilde{B}^N(k+1) - \tilde{B}^N(k)|\mathcal{F}_k\right] = \tilde{D}^N(k)$$

$$\begin{aligned} & E\left[\tilde{D}^N(k+1) - \tilde{D}^N(k)|\mathcal{F}_k\right] \\ = \; & -\tilde{D}^N(k) + \\ & \frac{F(\gamma N(\tilde{B}^(k) + \tilde{D}^N(k))) - F(\gamma N\tilde{B}^N(k))}{1 - F(\gamma N\tilde{B}^N(k))} \times \\ & (1 - \tilde{D}^N(k) - \tilde{B}^N(k)) \end{aligned}$$

Using $d = \tilde{D}^N(k)$, $b = \tilde{B}^N(k)$ and $\Gamma = \gamma N$, we can write the drift function as,

$$f_N = \left(d, \frac{F(\Gamma(b+d) - F(\Gamma b)}{1 - F(\Gamma b)}(1 - b - d) - d\right)$$

where both d and b are fractions taking values from $[0, 1]$. It is clear that $\frac{f_N}{1/N}$ diverges with $N \to \infty$ but we want this quantity to converge to a function f (which is independent of N) so that we can apply Kurtz's theorem to obtain an approximating ODE. We can see that the problem in the above case is caused because the drift function in the original process scales with the state. Hence in this case, while scaling, we need to slow down the process by another factor of N. To this purpose, we use the probabilistic attempt model in our scaling. The same scaling has been used in the literature in the context of the analysis of Random Multi-Access Algorithms by Bordenave et al. [97]. Note that this modifies the dynamics of the original process. The o.d.e. will be the limit (in probability) of the stochastic process with modified dynamics as $N \to \infty$ but will be a heuristic approximation for the stochastic process with the original dynamics.

3.8.3 Proof of Theorem 3.2.1

Kurtz's theorem [88] provides us a way by which we can approximate the evolution of a pure jump Markov process by the solution of a derived ODE. In this paper we shall refer to [98] for an equivalent version of Kurtz's theorem, which is simpler to handle. It can be restated as follows to be directly used in our context.

Theorem 3.8.1. Given that,

(i) $f(b,d)$ is Lipschitz

(ii) $\sup_{(b,d)\in\Delta^{(N)}} \left| \frac{f^{(N)}(b,d)}{\frac{1}{N}} - f(b,d) \right| \overset{N\to\infty}{\to} 0$ where $\Delta^{(N)} = [0, \frac{1}{N}, \cdots, N] \times [0, \frac{1}{N}, \cdots, 1]$, with $b + d \leq 1$.

(iii) $E(|\tilde{Y}^{(N)}(k)|^2/\mathcal{F}_k) \leq \frac{C_0}{N^2}$ and $E(|\tilde{Z}^{(N)}(k)|^2/\mathcal{F}_k) \leq \frac{C_1}{N^2}$ where $\mathcal{F}_k = (B^{(N)}(0), D^{(N)}(0), \cdots, B^{(N)}(k), D^{(N)}(k))$ is the history of the process upto time k.

(iv) $\tilde{B}^{(N)}(0) \overset{p}{\to} b(0)$ and $\tilde{D}^{(N)}(k) \overset{p}{\to} d(0)$

then we have for each $T > 0$ and each $\epsilon > 0$,

$$P\left(\sup_{0\leq t\leq T} \left|\left| \left(\tilde{B}^N(\lfloor Nt \rfloor), \tilde{D}^N(\lfloor Nt \rfloor)\right) - \left(b(t), d(t)\right) \right|\right| > \epsilon \right) \overset{N\to\infty}{\to} 0$$

where $b(t)$ and $d(t)$ are defined as the solutions of the system of ODE,

$$\dot{b}(t) = f_1(b,d)$$

$$\dot{d}(t) = f_2(b,d)$$

with initial conditions $(b(0), d(0))$.

(i) **Lipschitz property**

Consider,

$$f_1(b,d) = d$$

$$f_2(b,d) = \frac{\Gamma d}{1-\Gamma b}(1-b-d) - d$$

$$\frac{\partial f_1}{\partial b} = 0; \frac{\partial f_1}{\partial d} = 1$$

$$\frac{\partial f_2}{\partial b} = \frac{\Gamma(d(\Gamma-1) - d^2\Gamma)}{(1-\Gamma b)^2}; \frac{\partial f_2}{\partial d} = \frac{\Gamma - 1}{1-\Gamma b} - \frac{2d\Gamma}{1-\Gamma b}$$

We see that each of the terms above is bounded when $(b,d) \in [0,1] \times [0, 1-b]$. Thus the norm of Jacobian $||Df(b,d)||$ is uniformly bounded, and it follows that $f(b,d) = (f_1(b,d), f_2(b,d))$ is Lipschitz.

(ii) **Uniform Convergence**

$$f^N(b,d) = \frac{1}{N}\left(d, \frac{N\gamma d}{1-N\gamma b}(1-b-d) - d\right)$$

$$f(b,d) = \left(d, \frac{\Gamma d}{1-\Gamma b}(1-b-d) - d\right)$$

By definition, $\Gamma = N\gamma$ and hence the uniform convergence of $\frac{f^N(b,d)}{1/N}$ to $f(b,d)$ in the domain $(b,d) \in [0, \frac{1}{N}, \cdots, N] \times [0, \frac{1}{N}, \cdots, b]$ is straightforward.

(iii) **Bounded Noise variance**

We can write the noise variances as follows:

$$\begin{aligned} E(|Y^N(k)|^2|\mathcal{F}_k) &= \frac{1}{N}(1-\frac{1}{N})D^N(k) \\ &\leq \tilde{D}^N(k) \leq 1 \end{aligned}$$

$$\begin{aligned} E(|Z^N(k)|^2|\mathcal{F}_k) &= \frac{1}{N}(1-\frac{1}{N})D^N(k) + \sum_{D^{\star N}(k)=0}^{D^N(k)} \left(\frac{D^{\star N}(k)\gamma}{1-\gamma B^N(k)}\right. \\ &\quad \times\big(1-\frac{D^{\star N}(k)\gamma}{1-\gamma B^N(k)}\big)(N-B^N(k)-D^N(k)) \\ &\quad \times\left.\binom{D^N(k)}{D^{\star N}(k)}\big(\frac{1}{N}\big)^{D^{\star N}(k)}\big(1-\frac{1}{N}\big)^{D^N(k)-D^{\star N}(k)}\right) \\ &\leq \tilde{D}^N(k) + \frac{\Gamma(1-\tilde{B}^N(k)-\tilde{D}^N(k))\tilde{D}^N(k)}{1-\Gamma\tilde{B}^N(k)} \end{aligned}$$

where $D^{\star N}(k)$ represents the number of nodes that succeed in contributing their influence, at the mini slot k. Since $\tilde{D}^N(k)$,$\tilde{B}^N(k)$ and Γ are less than or equal to 1, both the above terms can be bounded above by constants C_1 and C_2. Hence in the process involving fraction of nodes, we have

$$E(|\tilde{Y}^N(k)|^2|\mathcal{F}_k) = \frac{1}{N^2}E(|Y^N(k)|^2|\mathcal{F}_k) \leq \frac{C_1}{N^2}$$

$$E(|\tilde{Z}^N(k)|^2|\mathcal{F}_k) = \frac{1}{N^2}E(|Z^N(k)|^2|\mathcal{F}_k) \leq \frac{C_2}{N^2}$$

and we can see that the noise conditions are satisfied.

(iv) **Convergence of initial conditions**

By choice, we have $\tilde{B}^N(0) = b(0)$ and $\tilde{D}^N(0) = d(0)$.

Thus by Kurtz's theorem, we have for each $T > 0$ and each $\epsilon > 0$,

$$P\left(\sup_{0 \le t \le T} \left|\left| \left(\tilde{B}^N(\lfloor Nt \rfloor), \tilde{D}^N(\lfloor Nt \rfloor)\right) - \left(b(t), d(t)\right) \right|\right| > \epsilon \right) \overset{N \to \infty}{\to} 0$$

where $(b(t), d(t))$ is the unique solution of the o.d.e..

$$\dot{b}(t) = d(t)$$

$$\dot{d(t)} = -d(t) + \frac{\Gamma d(t)}{1 - \Gamma b(t)}(1 - b(t) - d(t))$$

with initial conditions $(b(0) = 0, d(0) = a(0))$.

■

3.8.4 Solution of the o.d.e.

Recall the system of o.d.e.s for the evolution of influence under the uniform distribution is given by,

$$\dot{b} = d$$

$$\dot{d} = -d + \frac{\Gamma d}{1 - \Gamma b}(1 - b - d)$$

Substituting for d in the second equation and simplifying, we get

$$\ddot{b} = \frac{\Gamma \dot{b} - \dot{b} - \Gamma \dot{b}^2}{1 - \Gamma b}$$

Note that,

$$\ddot{b} = \frac{\mathrm{d}d}{\mathrm{d}t} = \frac{\mathrm{d}d}{\mathrm{d}b} \times \frac{\mathrm{d}b}{\mathrm{d}t} = d\frac{\mathrm{d}d}{\mathrm{d}b}$$

Hence,

$$d\frac{\mathrm{d}d}{\mathrm{d}b} = \frac{\Gamma \dot{b} - \dot{b} - \Gamma \dot{b}^2}{1 - \Gamma b}$$

Separating the variables, integrating on both sides, and after taking anti-logarithm

$$\Gamma - 1 - \Gamma d = c_1(1 - \Gamma b)$$

Differentiating on both sides yields $\dot{d} = c_1 d$ and hence $d(t) = c_2 e^{c_1 t}$. Substituting in the above equation for $d(t)$ we get

$$b(t) = \frac{c_2}{c_1} e^{c_1 t} + \frac{1 + c_1 - \Gamma}{\Gamma c_1}$$

Solving for constants using the initial conditions $b(0) = 0$, $d(0) = d_0$ results in $c1 = -(1 + \Gamma d_0 - \Gamma) =: -r$ and $c_2 = d_0$. Hence we have,

$$b(t) = \frac{d_0}{r} - \frac{d_0}{r} e^{-rt}$$

$$d(t) = d_0 e^{-rt}$$

■

3.8.5 Convergence of $h_\gamma^{(N)}(k)$ to b_∞

The solution of the o.d.e. suggests that $\lim_{t\to\infty} b(t) = d_0/r$. This is consistent with the fact that $\lim_{t\to\infty} \frac{h_\gamma^{(N)}(k)}{N} \to \frac{d_0}{1-(1-d_0)\Gamma} = b_\infty$ as we now proceed to show.

$$h_\gamma^{(N)}(k) = k\Big[1 + (N-k)\gamma\Big[1 + (N-k-1)\gamma\Big[1 + \cdots$$

$$h_\gamma^{(N)}(k+1) = (k+1)\Big[1 + (N-k-1)\gamma\Big[1 + \cdots$$

Thus we can write,

$$h_\gamma^{(N)}(k) = k\left[1 + \gamma(N-k)\frac{h_\gamma^{(N)}(k+1)}{k+1}\right]$$

Now substituting for $k = d_0 N$ and noting that $\Gamma = \gamma N$ we have

$$\frac{h_\gamma^{(N)}(Nd_0)}{Nd_0} = 1 + \Gamma(1-d_0)\frac{h_\gamma^{(N)}(N(d_0 + \frac{1}{N}))}{N(d_0 + \frac{1}{N})}$$

Taking $N \to \infty$ and noting that $h_\gamma^{(N)}(k)$ is a continuous function, we have

$$\frac{1}{d_0}\frac{h_\gamma^{(N)}(Nd_0)}{N} = 1 + \frac{\Gamma(1-d_0)}{d_0}\frac{h_\gamma^{(N)}(Nd_0)}{N}$$

Take limits on both sides, and solving for the unknown,

$$\frac{h_\gamma^{(N)}(Nd_0)}{N} \to \frac{d_0}{1-(1-d_0)\Gamma} = b_\infty$$

■

CHAPTER 4

Coevolution of Content Popularity and Availability in MONETs

In the previous two chapters, we studied an influence evolution process in isolation. As motivated in Section 1.1.4, such influence dynamics can be used to understand the evolution of demand for content, in the context of mobile opportunistic networks. Understanding the demand evolution will allow content providers to design optimal content delivery mechanisms. To this end, in this chapter, we study a co-evolution process, where the interest in the content evolves in conjunction with content spread. We consider a population of N mobile nodes, and model their pair-wise meetings by independent Poisson point processes. The item of content is provided to a subset of the initially interested set of nodes. Epidemic models are used to model both the spread of popularity and the content. In this framework, we make the following contributions:

1. We model the evolution of popularity and the spread of content by models inspired from epidemiology. While the former is similar to an uncontrolled SIR (Susceptible-Infected-Recovered) process, the latter is modeled as a controlled SI (Susceptible-Infected) process, with the copying probability to a relay (σ) serving as a static control, thus yielding what we call the *SIR-SI* model. The evolutions are modeled as coupled continuous time Markov chains. We scale the population size N, and under certain assumptions on the scaling with N of the model parameters, we use Kurtz's theorem [88] to derive the fluid limit of the joint evolution process. Simulation results are provided to illustrate how large a value of N is required before the fluid limit is a good approximation to the finite population dynamics.

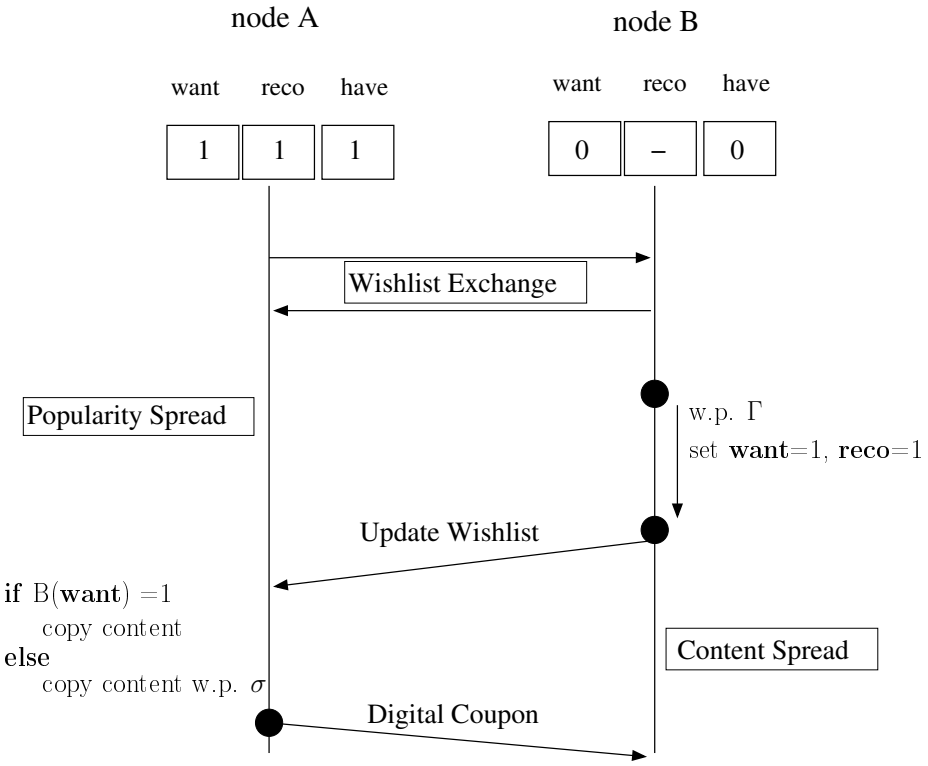


Figure 4.1: Sample exchange during a pairwise meeting of an active destination with content (node A) and a relay without content (node B).

2. Then, permitting the copying probability, σ, to vary with time, we obtain a system of controlled o.d.e.s for which we obtain the optimal control by direct arguments using certain monotonicity properties. This results in a time-threshold structure of the optimal control. We provide an extensive numerical study of this model, thus providing additional interesting insights.

4.1 Joint Spread of Interest and Content

In this section we describe the system that we wish to model, and motivate our assumptions and simplifications, by taking recourse to an example of a hypothetical mobile application. This application framework that we propose utilises opportunistic links between mobile personal devices and can be used for pre-release publicity of a product. The product could be an upcoming movie, a book soon to be launched, or a new gadget about to be released to the market. The actual item of content under consideration could be a digital discount coupon that could be used to pre-order the product. This may also be bundled with promotional material such as a trailer, a sample chapter, or a demo video; thus, the "coupon" carries a non-trivial quantity of data. The demand for the content could evolve

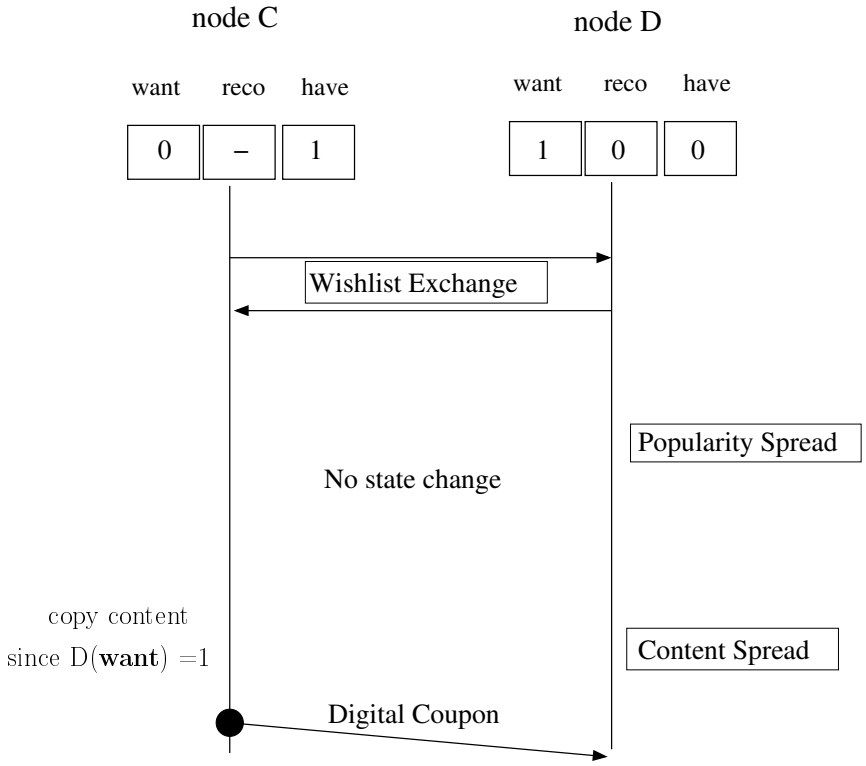


Figure 4.2: Sample exchange during a pairwise meeting of a relay with content and an inactive destination without content. There is no state change in the popularity spread phase, since neither of the nodes is an active destination.

with time and the content creator would wish to spread the digital coupons suitably to meet the demand (as it evolves).

Consider every user in the population carrying a smartphone with a device-to-device (D2D) interface turned on. Each of the phones runs the application (app), on which the users can

- Maintain a wishlist of upcoming products (movies, books, gadgets, etc.)
- Share and receive wishlists among other users in the population
- Receive promotional offers either from the central server or via pairwise contacts

By allowing users to declare the interest and enabling the spread of interest, we extend the notion of a personal wishlist (in e-commerce platforms like Amazon) to a *social wishlist*. The application also combine the services of coupon delivery (e.g., Groupon) with such wishlists, thus allowing targeted discount delivery, wherein brands can identify the users who are interested in the coupon and deliver it to them.

For simplicity we will consider a single product to explain the application framework.

Initial seeding: Let an initial subset of users add the product to their wishlist, influenced by extrinsic means such as traditional advertising, mass campaigns, etc. These will form the *initial set of active destinations*. The central app server chooses a subset of these destinations and sends them the promotional offer (digital coupon), thus yielding an *initial set of destinations who have the content*. This concludes the initial seeding phase.

Wishlist exchange: At each pairwise contact, the smartphones exchange their wishlist information. For each product (our analysis is concerned with only one such product) the wishlist information contains the following bits:

- **want** bit : **1** if the node is a destination, **0** otherwise
- **reco** bit : provided the node is a destination, **1** if it is active, **0** if it is inactive. Only active destinations aid in popularity spread.
- **have** bit : **1** if the node has the digital coupon, **0** otherwise

The **want** and **have** bits, as their names suggest, indicate whether the user wants to receive the content, and whether the user possesses the content. The **reco** bit is used to indicate whether the user is actively spreading interest for the content. As long as the reco bit of a user is 1, the app in the user's device will make this recommendation to relays that it meets. Users become inactive by switching the **reco** bit from 1 to 0 (modeled as occurring at rate β_N). Users become inactive when they lose motivation to spread the popularity after a while, due to aging of the content. These users will no longer aid the popularity spread, but still wish to receive the content. This is equivalent to users subscribing to a page on Facebook, but later revoking the rights of the page to *post on the user's behalf*.

After exchanging the wishlists, the following phases may occur depending upon the states of the two nodes in contact.

Popularity spread: If one of the nodes is an active destination (**want=1**, **reco=1**) and the other is a relay (**want=0**), then popularity spread may occur. The app of the relay user will receive a recommendation from the active destination device, and this is notified to the user by the app interface. We assume that the as yet uninterested user (relay) gets influenced with some probability, and model this by the relay switching to (**want=1**, **reco=1**) with probability Γ.

Content spread: If the two nodes that meet have different **have** states, then content (digital coupon) spread may occur. If the node without the content is a destination (**want=1**), then the content is copied with probability 1, while if it is a relay (**want=0**) then the app

decides to copy the content with probability σ.
Figures 4.1 and 4.2 show two possible pairwise interactions that can occur in this system. The process is terminated when a sufficient fraction of destinations have been given the content (digital coupon). This indicates that the content creator has successfully completed his campaign and the product is removed from the application server.
Remarks:

- Though we might be interested only in delivering the content to the destinations, there are two advantages of copying to a relay. First, a copy to a relay promotes the further spread of the content even to destinations (explored in [71]). Second, the relay we copy to now might later get influenced to become a destination.

- *Incentives and Freeriders:* There is incentive for a node in the population to declare its interest in a particular product, since it is highly likely that it will receive a promotional offer by being in the system. We also assume that the nodes forward the content without engaging in *freeriding* behavior, i.e. receiving the content but not sharing it on future pairwise meetings. The impact of freeriders is an interesting future direction of research and we suspect mechanisms suggested for existing P2P networks [100] can be adopted for this system.

- We also observe that the popularity spread is independent of the possession of content. This might be an adequate model for the spread of pre-release promotional material among an established fanbase, for example, "fans" of a particular genre of movies, or of a particular author; among fans, having the promotional material may not change the influence level. More complex coupled dynamics will arise if the (want=1, reco=1, have=1) nodes have a different rate of spreading interest in the content, as compared to the (want=1, reco=1, have=0) nodes.

4.2 CTMC Model of the Joint Evolution

We consider the dissemination of a single item of content among a population $\mathcal{N}$ of mobile nodes of fixed size N . We assume that pairs of nodes meet at the points of a Poisson process with rate $\lambda_N = \frac{\lambda}{N}$. This assumption is justified by earlier studies of several random mobility models in the literature. For instance, in the Random Waypoint Model (RWP),

the expected meeting rate between two mobile nodes [101] is given by

$$\lambda = \frac{2\overline{v}r}{A}$$

where $\overline{v}$ is the average velocity of each node, r is the radio range and A is the system area. If we assume that individual node velocities and radio ranges do not scale with N, then the assumption of meeting rate scaling inversely with N, implies the system area scales as N. This can be seen as a result of ensuring a constant node density (number of nodes per unit area).

Define $\mathcal{A}^N(t)$ to be the set of destinations at time t and let $\mathcal{S}^N(t) = \mathcal{N}\backslash\mathcal{A}^N(t)$ be the set of relays. We classify the destinations further as active ($\mathcal{D}^N(t)$) and inactive ($\mathcal{B}^N(t)$). [1] Let $A^N(t)$,$B^N(t)$ and $D^N(t)$ be the sizes of the sets $\mathcal{A}^N(t)$, $\mathcal{B}^N(t)$, and $\mathcal{D}^N(t)$ respectively. We then have $A^N(t) = B^N(t) + D^N(t)$. Only active destinations can convert other relays into destinations, i.e., *infect* them, whereas inactive destinations are still interested in the content, but no longer spread the content's popularity. The spread of popularity is governed by two parameters Γ and β_N. These are similar to the infection and recovery rates in the SIR epidemics. An active destination remains in the active state for an exponentially distributed duration with parameter β_N. Thus the rate at which an active destination gets converted to an inactive destination is β_N, while, the probability with which an active destination infects a relay it meets is given by Γ. From the Poisson process model of meetings, and the exponentially distributed active duration, it follows that the process $(B^N(t), D^N(t))$ is a continuous time Markov chain (CTMC); Figure 4.3 shows the state transitions in the content popularity process.

We aim to model the joint spread of interest in the content and of the content itself. We do this by combining the SIR model for popularity spread with a controlled-epidemic copying process for content spread (probabilistic control, similar to the SI model). We will refer to this joint model as an *SIR-SI* model. In order to model the content delivery process, we further classify the nodes depending on whether they have the content. Let $\mathcal{X}^N(t) \subseteq \mathcal{A}^N(t)$ denote the set of destinations that have the content, and $\mathcal{Y}^N(t) \subseteq \mathcal{S}^N(t)$ denote the set of relays that have the content. Let $\mathcal{X}_b^N(t)$ and $\mathcal{X}_d^N(t)$ respectively be the intersection of $\mathcal{X}^N(t)$ with the sets $\mathcal{B}^N(t)$ and $\mathcal{D}^N(t)$. For the evolution of $(\mathcal{X}^N(t), \mathcal{Y}^N(t))$, we model a content copying process based on the SI model. Whenever a node that has the content meets a node that does not, content transfer takes place based on a controlled

[1] Due to the similarity with the spread of infection, we will often use terminology from epidemic spread models, such as, susceptible, infected, recovered, etc.

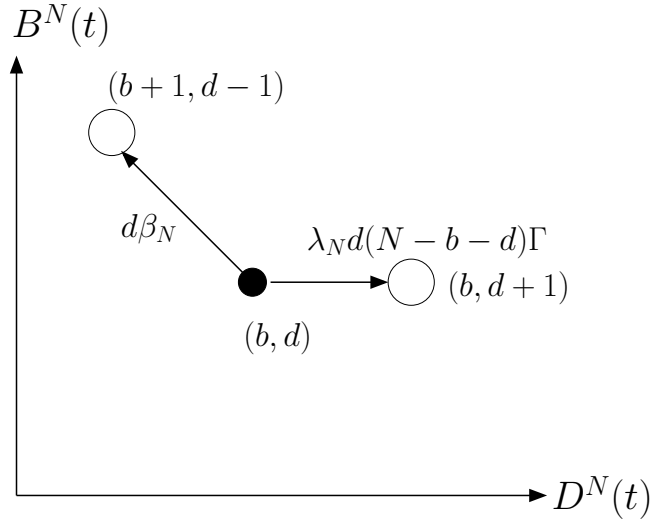


Figure 4.3: Possible state transitions for the popularity process $(B^N(t), D^N(t))$.

copying process. We always copy to a destination that does not have the content, whereas copying to a relay is controlled by a probability $\sigma \in [0, 1]$. We wish to obtain the fluid dynamics of this model. A brief summary of notation is provided in Table 4.2.1.

4.2.1 System Evolution

The set of all possible transitions among the different states of nodes is shown in Figure 4.4. Note that the process evolves at epochs t_k, $k \geq 1$ which are either pairwise meetings (occurring at rate $\lambda_N|\mathcal{N}|(|\mathcal{N}| - 1)$) or instances of recovery of an active destination (occurring at rate $\beta_N D^N(t)$). The system state is represented by the tuple $Z^N(t) = (B^N(t), D^N(t), X_b^N(t), X_d^N(t), Y^N(t))$, the sizes of the respective sets. Again, due to the Poisson process model for the meeting instants, and the exponentially distributed active durations for destination nodes, $Z^N(t)$ is a continuous time Markov chain; in Table 4.2 we show its transition structure.

4.2.2 Drift Equations

The drift rate is the expected rate of change of the process out of a given state. We can express the expected drift of the system as follows:

$$F^N(Z) = \sum_{i \in \mathcal{E}} R_i(Z)\delta_i(Z) \tag{4.1}$$

where $i \in \mathcal{E}$ indexes the epoch type, $R_i(Z)$ and $\delta_i(Z)$ are the respective transition rates and corresponding state changes given the current state of the CTMC $Z^N(t) = Z$, as given

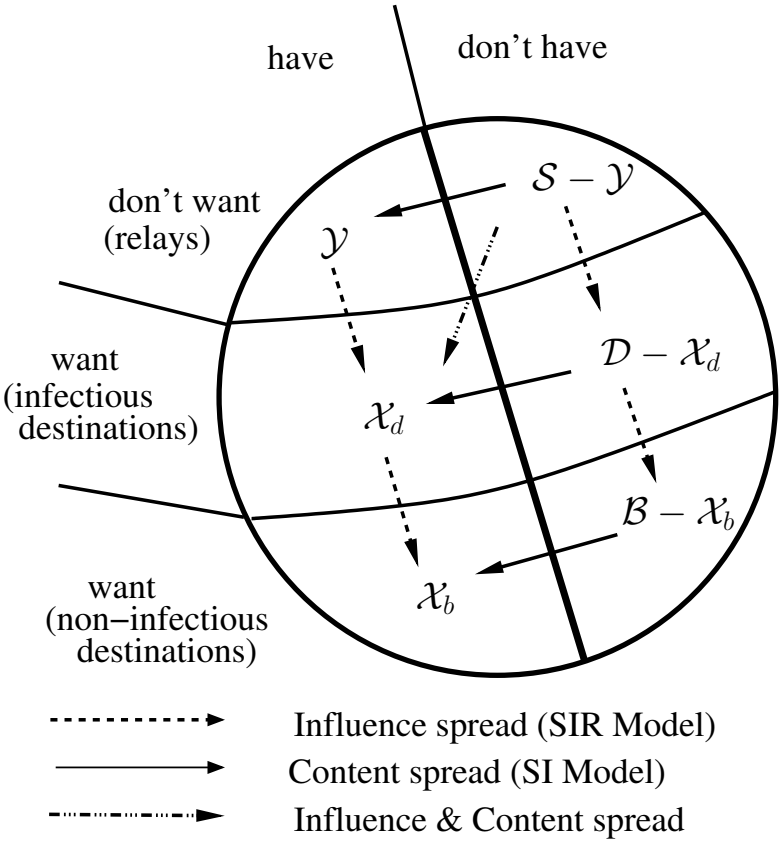


Figure 4.4: Possible transitions between the states in the SIR-SI model. Observe that, in addition to the transitions due to SIR model and SI model separately, there are instances where both can occur simultaneously. For instance, when an active destination node $i \in \mathcal{X}_d$ meets a relay node $j \in \mathcal{S}-\mathcal{Y}$, the relay might get converted to a destination and also receive the content, as indicated by the diagonal transition.

Notation	Explanation
$\mathcal{A}^N(t)$	set of destinations at time t, of size $A^N(t)$
$\mathcal{S}^N(t)$	set of relays at time t, of size $S^N(t)$
$\mathcal{B}^N(t)$	set of inactive destinations at time t, of size $B^N(t)$
$\mathcal{D}^N(t)$	set of active destinations at time t, of size $D^N(t)$
$\mathcal{X}^N(t)$	set of destinations with the content at time t, of size $X^N(t)$
$\mathcal{X}_b^N(t)$	set of inactive destinations with the content at time t, of size $X_b^N(t)$
$\mathcal{X}_d^N(t)$	set of active destinations with the content at time t, of size $X_d^N(t)$
$\mathcal{Y}^N(t)$	set of relays with the content at time t, of size $Y^N(t)$
$Z^N(t)$	$(B^N(t), D^N(t), X_b^N(t), X_d^N(t), Y^N(t))$
$\tilde{Z}^N(t)$	$\frac{Z^N(t)}{N}$

Table 4.1: A summary of the notation

Epoch type i	**Rate** $R_i(Z)$	**State update** $\delta_i(Z)$
$\mathcal{D}-\mathcal{X}_d$ recovers	$\beta_N(D-X_d)$	(1,-1,0,0,0)
$\mathcal{X}_d$ recovers	$\beta_N X_d$	(1,-1,1,-1,0)
$\mathcal{B}-\mathcal{X}_b$ meets $\mathcal{X}+\mathcal{Y}$	$\lambda_N(B-X_b)(X+Y)$	(0,0,1,0,0)
$\mathcal{D}-\mathcal{X}_d$ meets $\mathcal{Y}$	$\lambda_N(D-X_d)Y$	(0,0,0,1,0) + (0,1,0,1,-1) w.p. Γ
$\mathcal{D}-\mathcal{X}_d$ meets $\mathcal{X}$	$\lambda_N(D-X_d)X$	(0,0,0,1,0)
$\mathcal{X}_d$ meets $\mathcal{Y}$	$\lambda_N X_d Y$	(0,1,0,1,-1) w.p Γ
$\mathcal{S}-\mathcal{Y}$ meets $\mathcal{X}_b+\mathcal{Y}$	$\lambda_N(X_b+Y)(S-Y)$	(0,0,0,0,1) w.p. σ
$\mathcal{S}-\mathcal{Y}$ meets $\mathcal{D}-\mathcal{X}_d$	$\lambda_N(D-X_d)(S-Y)$	(0,1,0,0,0) w.p. Γ
$\mathcal{S}-\mathcal{Y}$ meets $\mathcal{X}_d$	$\lambda_N(X_d)(S-Y)$	(0,1,0,1,0) w.p. Γ (0,0,0,0,1) w.p$(1-\Gamma)\sigma$

Table 4.2: Transition rates and state updates for various possible epochs in the SIR-SI process, when the current state of the CTMC $Z^N(t)$ is given by $Z=(B,D,X_b,X_d,Y)$

in Table 4.2. The co-evolution process evolves at pairwise meetings or recovery instants. Given the current state of the system, the rates of each possible transition epoch type can be determined (as shown in Table 4.2), and the state update depends on the type of node(s) involved in the pairwise meeting (or node recovery). The exact expressions for the mean drift rates $F^N(Z)$ are provided in Section 4.6.1. Define $\tilde{Z}^N(t)=Z^N(t)/N$ and consider the o.d.e.s given below.

$$\dot{b} = \beta d \tag{4.2}$$

$$\dot{d} = -\beta d+\lambda\Gamma ds \tag{4.3}$$

$$\dot{x_b} = \beta x_d+\lambda(b-x_b)(x+y) \tag{4.4}$$

$$\dot{x_d} = \Gamma\lambda(d-x_d)y+\Gamma\lambda x_d s+ \tag{4.5}$$

$$\lambda(d-x_d)(x+y)-\beta x_d$$

$$\dot{y} = -\Gamma\lambda dy+\lambda\sigma(s-y)(x_b+y+(1-\Gamma)x_d) \tag{4.6}$$

where $s(t)=1-b(t)-d(t)$ and $x(t)=x_b(t)+x_d(t)$. We can then state the following result.

Theorem 4.2.1. Given the coevolution Markov process
$\tilde{Z}^N(t)=(\tilde{B}^N(t),\tilde{D}^N(t),\tilde{X}_b^N(t),\tilde{X}_d^N(t),\tilde{Y}^N(t))$ with initial conditions $\tilde{Z}^N(0)$ we have for each $T>0$ and each $\epsilon>0$,

$$\mathbb{P}\left(\sup_{0\le u\le T}\left|\left|\tilde{Z}^N(\lfloor Nu\rfloor)-z(u)\right|\right|>\epsilon\right)\overset{N\to\infty}{\to}0$$

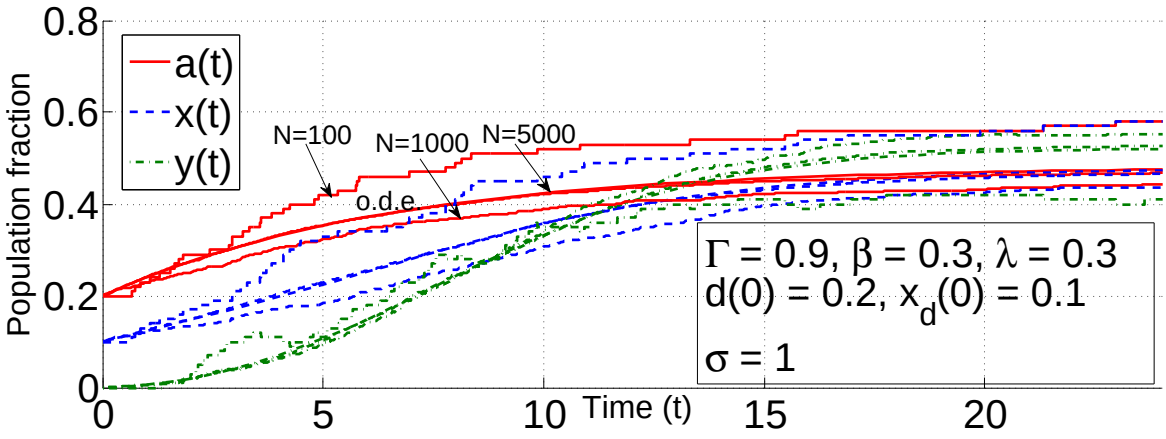


Figure 4.5: Convergence of the scaled coevolution process $\tilde{Z}^N(t)$ to the o.d.e. solutions $z(t)$ for increasing values of N with $\sigma = 1$

where $z(u) = (b(u), d(u), x_b(u), x_d(u), y(u))$ is the unique solution to the o.d.e. system given by Equations (4.2)-(4.6) with $z(0) = \tilde{Z}^N(0)$.

Proof. The proof involves verifying the conditions for applying Kurtz's theorem [88] to the SIR-SI process (see Section 4.6.1). Since the drifts are Lipschitz, the uniqueness of the solution of the o.d.e. is guaranteed once the initial condition is fixed, by the Cauchy-Lipschitz condition. □

Remark: As is evident from the o.d.e., the interest evolution and the content dissemination processes have independence in one direction, i.e., the evolution of $(b(t), d(t))$ proceeds independently, while the evolution of $(x_b(t), x_d(t), y(t))$ depends on $(b(t), d(t))$.

4.2.3 Accuracy of the O.D.E. Approximation

Figures 4.5 and 4.6 illustrate the convergence of the scaled coevolution process $\tilde{Z}^N(t)$ to the o.d.e. solutions $z(t)$ for increasing values of N for $\sigma = 1$ and $\sigma = 0.3$. We plot $a(t) = b(t) + d(t)$, $x(t) = x_b(t) + x_d(t)$ and $y(t)$ for clarity. Observe that as N increases, the approximation of the original process by the o.d.e. becomes better. Observe that when σ is increased from 0.3 to 1, there is significant increase in $y(t)$ but the contribution to $x(t)$ is not significant. Also for σ a constant, i.e., static control, $x(t)$ asymptotically reaches $a(t)$ irrespective of σ and $y(t)$ asymptotically reaches $s(t) = 1 - a(t)$ provided $\sigma > 0$. Note that from the o.d.e. equations, it is clear that, $x(t)$ is monotonically increasing, whereas in general $y(t)$ need not be monotonic.

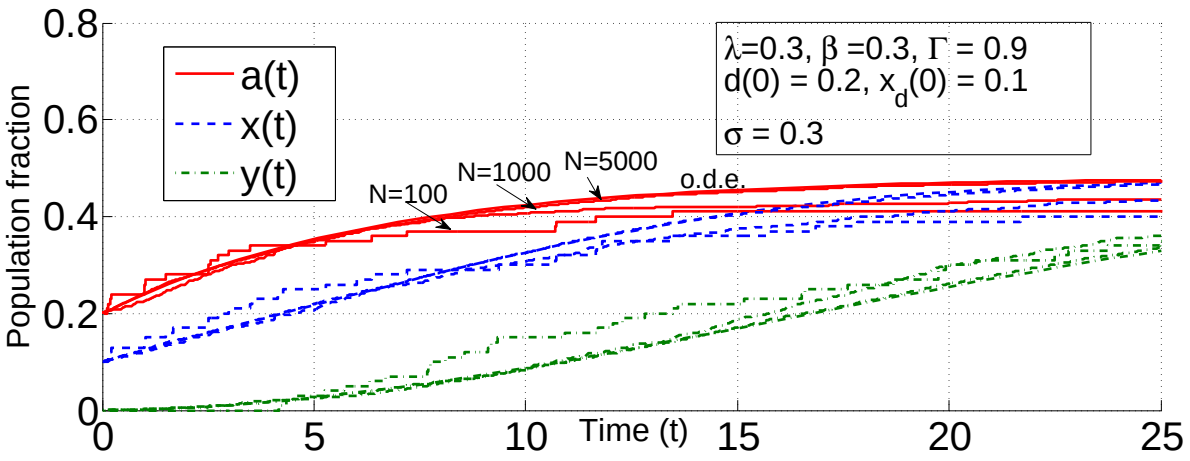


Figure 4.6: Convergence of the scaled coevolution process $\tilde{Z}^N(t)$ to the o.d.e. solutions $z(t)$ for increasing values of N with $\sigma = 0.3$

4.2.4 Copying to Relays: Static vs. Dynamic Control

In the above section we considered σ to be a static control, i.e. $\sigma(t) = \sigma$, $\forall t$. We could instead consider a dynamic copying control $\sigma(t) \in [0, 1]$. Then for each N we have a controlled continuous time Markov chain. The optimal control appears difficult to obtain in the finite size problem. We instead replace the constant copying probability in the ODE limit by $\sigma(t)$ which yields a controlled ODE. We can then obtain the optimal (deterministic) control for the controlled ODE (as in 4.3.1). In certain situations it can be shown that this control is asymptotically optimal for the finite size problem as N increases [102]. The proof given in [102], for discrete time MDPs with finite time horizon, does not directly apply here, and needs to be extended to accommodate our setting. In this work, we derive the optimal control $\sigma(t)$ for the controlled ODE, and then treat the use of this optimal control for the finite population Markov chain as a heuristic.

In order to justify this approach, the convergence of the co-evolution process to its o.d.e. limit is numerically demonstrated in Figure 4.7 for a time-threshold type control. In this case, $\sigma(t) = 1$, $t < 4$ and $\sigma(t) = 0$ for $t \geq 4$. Note that when $\sigma(t) = 0$, $\dot{y} \leq 0$ from Equation (4.6). Observe that the trajectory of $x(t)$ is very similar to the one obtained when $\sigma = 0.3$, but the value of $y(t)$ for large t is considerably smaller. This indicates that this threshold policy ($\tau = 4$) will perform better than the static policy $\sigma = 0.3$, for the purpose of copying to a large fraction of the destinations, while keeping in check the number of relays who get the content but do not eventually convert to destinations.

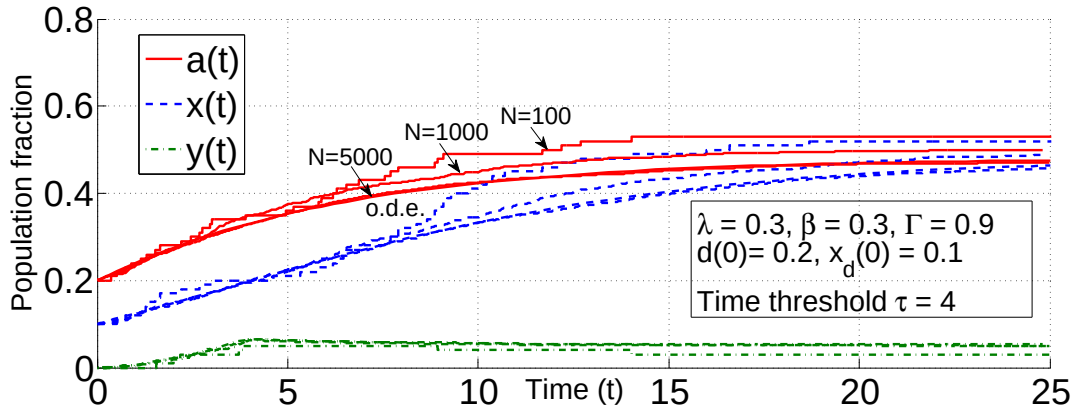


Figure 4.7: Convergence of the scaled cofor-evolution process $\tilde{Z}^N(t)$ to the o.d.e. solutions $z(t)$ for increasing values of N with dynamic $\sigma(t)$, a time-threshold function in this case, $\sigma(t) = 1$, $t < 4$ and $\sigma(t) = 0$ for $t \geq 4$.

4.3 The Optimization Problem

In [71], since the fraction of destinations was constant, it was suitable to choose the time of delivery to a fraction α of destinations as the objective to be minimized. In our setup, since the fraction of destinations is time-varying, we define the *target time*,

$$T_\sigma = \inf\{t : x_\sigma(t) \geq \alpha a(\infty)\}$$

where $a(\infty)$ is the terminal fraction of destinations as given by the SIR model for interest evolution $(b(t), d(t))$ and $x_\sigma(t)$ is the fraction of destinations that have the content at time t under the control $\sigma(t)$. Note that since $d(\infty) = 0$ and $a(t) = b(t) + d(t)$, $a(\infty) = b(\infty)$. The cost we are interested in optimizing is of the form

$$C_\sigma = \psi y_\sigma(T_\sigma) + T_\sigma = \psi y_\sigma(T_\sigma) + \int_0^{T_\sigma} 1dt \tag{4.7}$$

Here $y_\sigma(T_\sigma)$ denotes the number of relays that have the content at time T_σ and signifies the number of wasted copies, and ψ is the tradeoff parameter.

Even though the copying cost is distributed across the nodes, we treat the number of wasted copies at the target time T_σ as part of the system objective. This can be motivated by considering an incentive mechanism for the relay nodes which still hold the content but are not converted into destinations by the target time T_σ. Consider the earlier framework, where the content is a discount coupon for some service/product. The service provider, initially provides these discount coupons to a subset of its customers, who are then encouraged to spread replicas of the coupon. Nodes that are interested in the service, and are in possession of the coupon immediately claim the service. The service provider

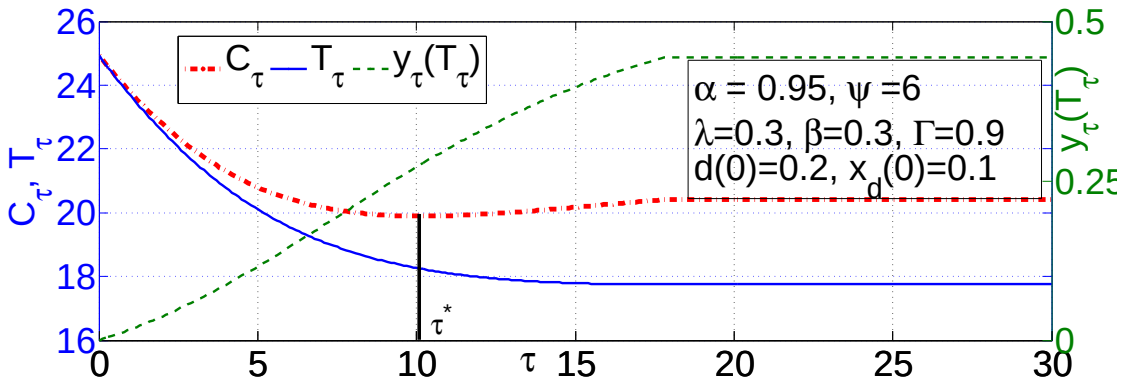


Figure 4.8: Evaluation of optimal τ for the time-threshold policy. The plot shows the behaviour of T_τ and $y_\tau(T_\tau)$ for various values of τ, for fixed system parameters.

stops accepting discount coupons once a pre-defined number (here $\alpha a(\infty)$) of coupons have been used. At this time, the relay nodes that are still in possession of the coupon will need to be provided a reimbursement for helping the spread of the coupon. In the example described in Section 4.1, this reimbursement could be in terms of credit on their cellular service, say 100 free SMSs.

4.3.1 Optimal Control

In this section we will establish the optimality of a time-threshold control for the objective given by the Equation (4.7). We will use definitions and lemmas provided in Section 4.6.2 in order to prove the following theorem.

Theorem 4.3.1. For the o.d.e system given by Equations (4.2)-(4.6) there exists an optimal control of the form,

$$\sigma_\tau(t) = \begin{cases} 1, & 0 < t < \tau \\ 0, & t \geq \tau \end{cases} \tag{4.8}$$

which optimizes the cost in Equation (4.7).

Proof. Recall $y_\sigma(T_\sigma)$ is the amount of wasted copies, at the target time T_σ. When $\sigma(t) = \sigma_\tau(t)$, a time-threshold policy (Equation (4.8)), we will denote this by $y_\tau(T_\tau)$, where τ is the time threshold associated with $\sigma_\tau(t)$. The sketch of the proof is as follows. We first establish that, the set of values taken by $y_\tau(T_\tau)$ form an interval of $[0, \rho_{\max}]$. Then, given any policy $\sigma(t)$, we show that:

- If $y_\sigma(T_\sigma) = \rho \leq \rho_{\max}$, then $\exists$ a time-threshold policy $\sigma_\tau(t)$ such that $y_\tau(T_\tau) = \rho$ and

$T_\tau \leq T_\sigma$.

- If $y_\sigma(T_\sigma) = \rho > \rho_{\max}$, we can find a time-threshold policy $\sigma_\tau(t)$ which has a smaller total cost.

Thus in either case, we have a time-threshold policy, which performs at least as good as the given policy, which proves the optimality of the time-threshold policy. Refer Section 4.6.2 for the proofs of the above claims. □

Figure 4.8 shows the variation of T_τ, $y_\tau(T_\tau)$ and C_τ as a function of τ for fixed system parameters. It can be seen that as τ is increased (as we continue to copy to relays for longer) , T_τ decreases monotonically (the target is achieved earlier), and we see an increase in the value of $y_\tau(T_\tau)$ (more wasteful copies). The optimal threshold $\tau^\star$ minimizes the total cost C_τ, by balancing the two component costs, taking the tradeoff parameter ψ into account.

4.4 Numerical results

Having observed that the optimal control is of the form 4.8, we can now numerically compute the optimal time threshold, which we shall refer to as $\tau^\star$. Recall the cost function is of the form

$$C_\sigma = T_\sigma + \psi y_\sigma(T_\sigma) = \psi y_\sigma(T_\sigma) + \int_0^{T_\sigma} 1dt$$

where T_σ is the target time to reach a fraction α of the terminal fraction of destinations under the policy $\sigma(t)$,i.e.,

$$T_\sigma = \inf\{t : x_\sigma(t) \geq \alpha a_\infty\}$$

Since in this case, $\sigma(t) = \sigma_{\tau^\star}(t)$, we shall denote the total cost by $C_{\tau^\star}$, and the two components of the cost by $T_{\tau^\star}$ and $y_{\tau^\star}(T_{\tau^\star})$. We obtain the optimal time-threshold ($\tau^\star$) by numerically sweeping over $[0, T_{\max}]$ to search for the cost minimizing value of τ (as in Figure 4.8), where $T_{\max}$ is the maximum simulation time. In the following discussion we shall numerically study the effect of various cost parameters, system parameters and initial conditions on $\tau^\star$,$C_{\tau^\star}$, $T_{\tau^\star}$ and $y_{\tau^\star}(T_{\tau^\star})$.

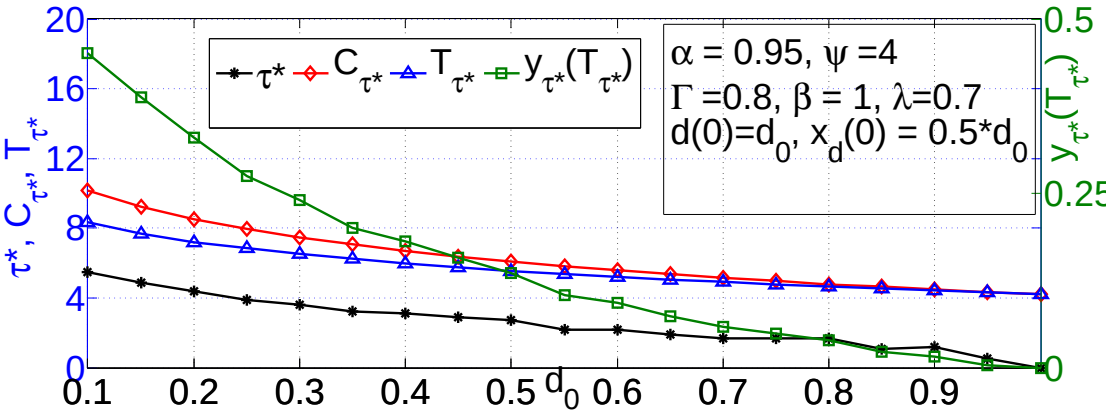


Figure 4.9: Effect of d_0 on the optimal cost and optimal τ

4.4.1 Effect of initial conditions

Figure (4.9) shows the effect of d_0 on the optimal $\tau^\star$. We see that as we increase d_0 keeping the fraction of $\frac{x_{d_0}}{d_0}$ constant, there are more destinations that have the content. Further, there are fewer relays in the population, and with fixed Γ and β, there is less chance for them to get converted into destinations. This prevents us copying to more relays and hence there is a decrease in the value of $\tau^\star$. Note that when $d_0 = 1$, the entire population consists of only destinations, and hence the optimal $\tau = 0$. Figure (4.10) shows the effect of x_{d_0} on the optimal $\tau^\star$. As the fraction $\frac{x_{d_0}}{d_0}$ is increased, since a larger number of destinations have the content, the newly infected relays also obtain the content. Observe that $\frac{x_{d_0}}{d_0} = 1$ implies that all the initial destinations are given the content. This implies that any destination converted in the future, will automatically receive the content. This alleviates the need to copy to any of the relays, since the main purpose of copying to relays is to serve the destinations without the content (either because they were destinations not given the content initially or were converted by other destinations without the content at a later time).

4.4.2 Effect of system parameters

Figure (4.11) shows the effect of Γ on $\tau^\star$. We see that as Γ increases, it increases the rate at which relays get converted into destinations, without affecting the rate at which content copying occurs (dependent on the meeting rate λ and copying probability $\sigma(t)$). We do not see much effect on $\tau^\star$, and $T_{\tau^\star}$. This is because, though increasing Γ increases the terminal set of destinations (and thus the target $\alpha a(\infty)$), it also helps in achieving the target by the conversion of relays with the contents into destinations with the content.

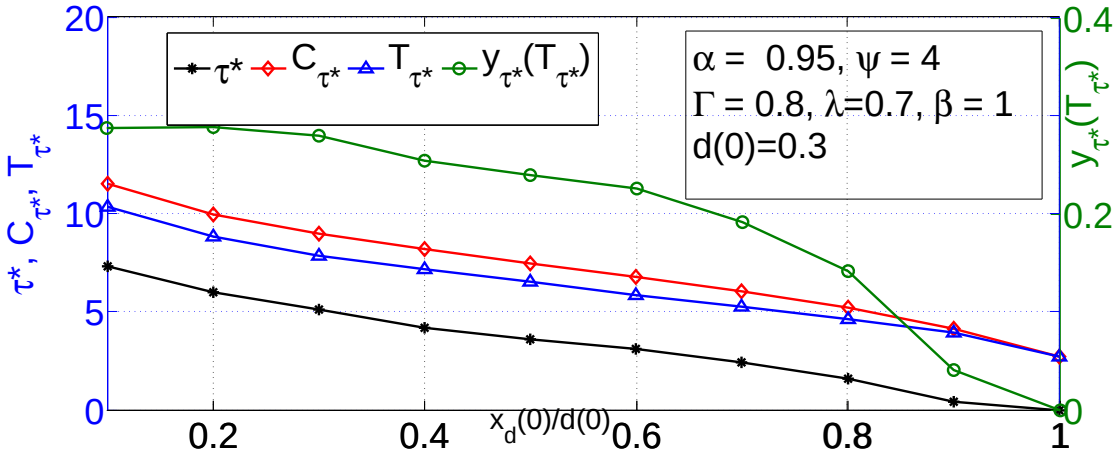


Figure 4.10: Effect of $\frac{x_{d_0}}{d_0}$ on the optimal cost and optimal τ

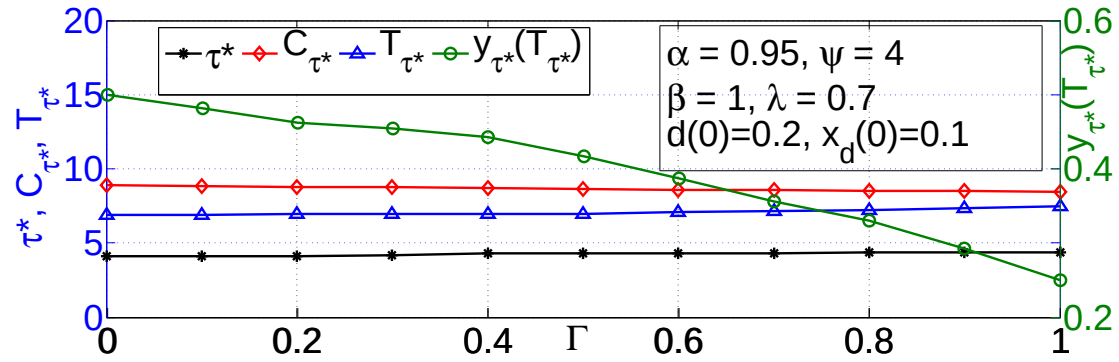


Figure 4.11: Effect of Γ on the optimal cost and optimal τ

Also due to this conversion, we observe a decrease in $y_\tau(T_\tau)$, further indicated by the fact that $y(t) \leq s(t)$ and the fraction of relays $s(t)$ is decreasing rapidly.

Figure (4.12) shows the effect of β on $\tau^\star$. We observe that when β is very low, destinations remain active for a longer duration before recovery. This leads to a faster decay in $s(t)$, the fraction of relays, and also an increase in the target $\alpha a(\infty)$. The larger target requires that copying to relays must continue for longer, and the faster decay in the fraction of relays prevents an excessive number of copies being eventually wasted in spite of copying for longer. Thus, a small β results in a large value of τ^*. For large β, since active destinations stay active for a shorter duration, the eventual population of destinations is small. Hence, the population of destinations carrying the content is small, thus requiring the relays to play a greater role in the spread of the content. However, again, since β is large these relays are less likely to get converted into destinations, thus explaining the increase in $y_\tau(T_\tau)$ for large values of β. Figure (4.13) shows the effect of λ on the optimal

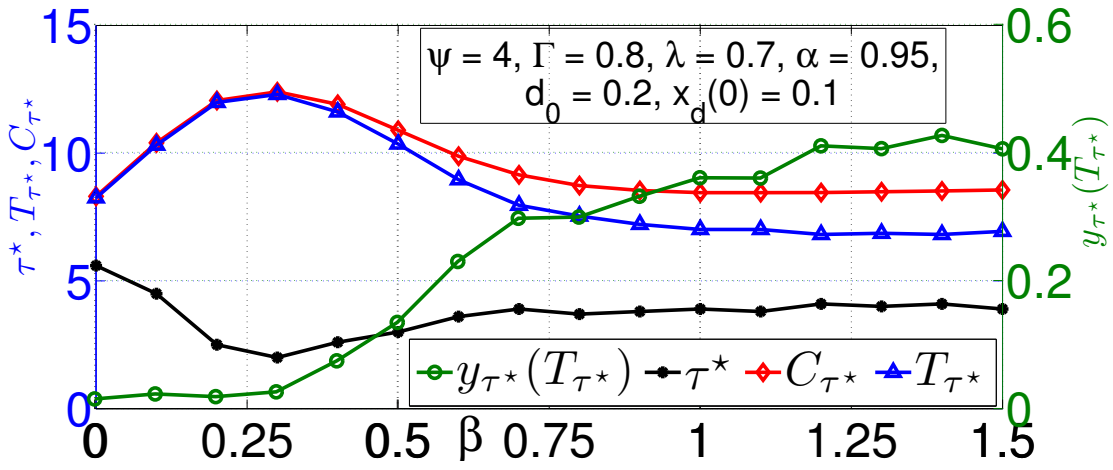


Figure 4.12: Effect of β on the optimal cost and optimal τ

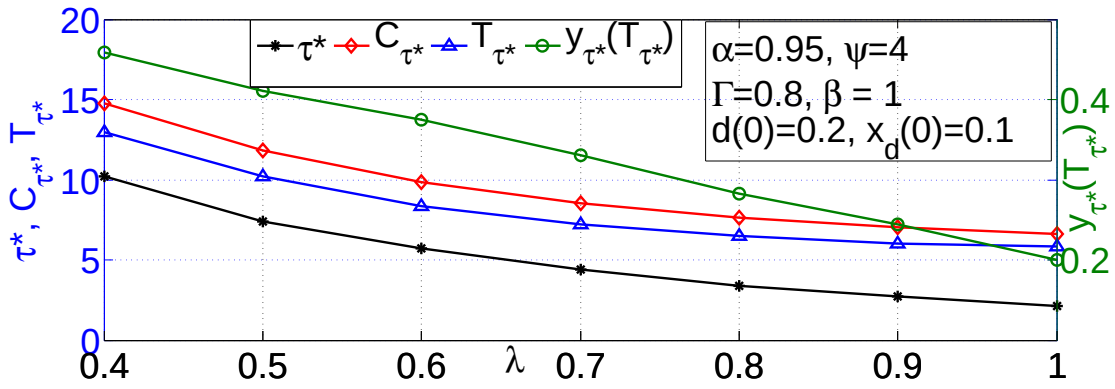


Figure 4.13: Effect of λ on the optimal cost and optimal τ

$\tau^\star$. As λ increases, the rate at which content spread increases, and thus a destination in need of the content is highly likely to obtain it from another destination. This results in a more passive policy of copying to relays, leading to a decrease in the optimal $\tau^\star$ and the corresponding costs.

4.4.3 Effect of cost parameters

Figure (4.14) shows the effect of α on $\tau^\star$. For the given system parameters, $a_\infty = 0.4352$ and hence for $\alpha < 0.23$, we see that the target is already achieved at time zero (since the initial seeding $x_d(0) = 0.1$). Thus for $\alpha < 0.23$, $\tau^\star = 0$,$T_{\tau^\star} = 0$ and $y_{\tau^\star}(T_{\tau^\star}) = 0$. As α increases, we see that $\tau^\star$ increases and in turn increases the costs, since we need to continue copying to relays for a longer duration to achieve the target. It is also to be noted that as $\alpha \to 1$, T_τ approaches ∞, since by definition $x(t)$ reaches $a(\infty)$ only asymptotically. Figure (4.15) shows the effect of ψ on $\tau^\star$. As ψ increases, we see that the emphasis on the

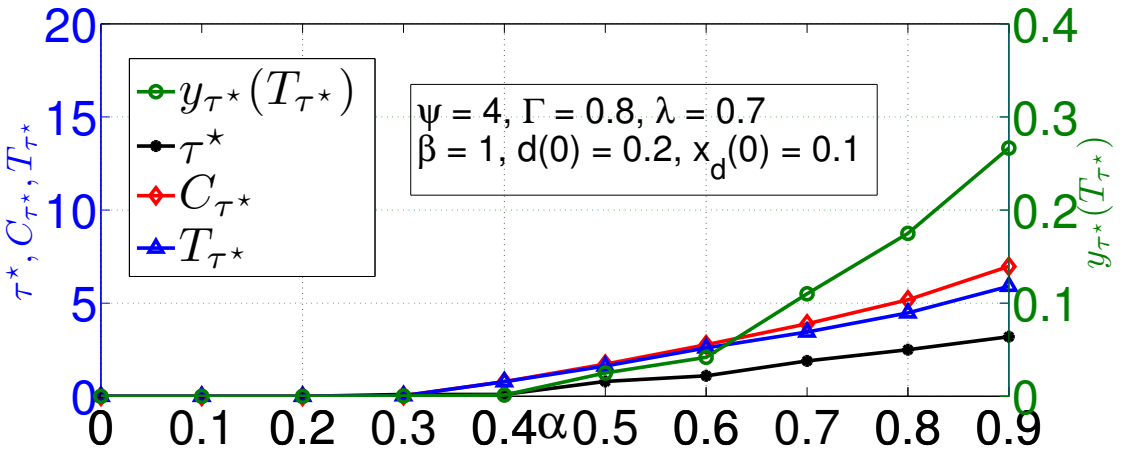


Figure 4.14: Effect of α on the optimal cost and optimal τ

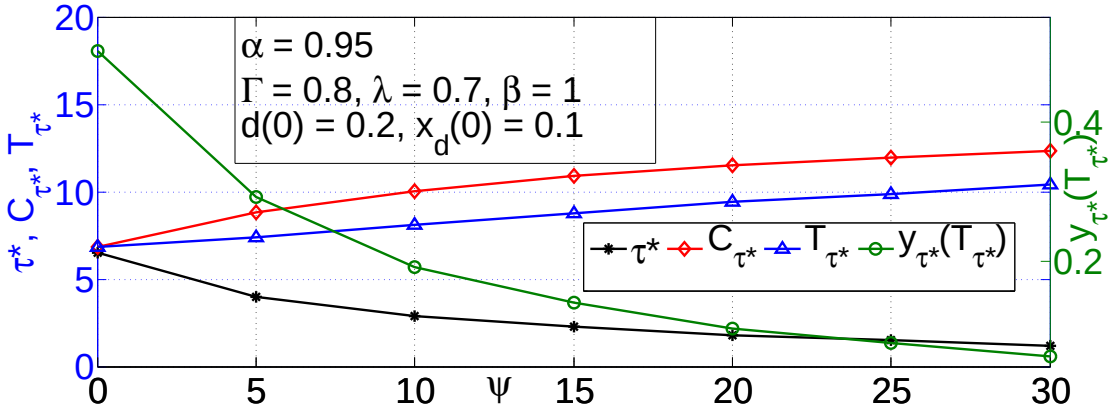


Figure 4.15: Effect of ψ on the optimal cost and optimal τ

wasted copies, $y(T)$, increases, and hence the control needs to be *less aggressive*. Thus we see a decrease in $\tau^\star$ as ψ increases, and this leads to a decrease in the wasted copies $(y_{\tau^\star}(T_{\tau^\star}))$ and an increase in the delay of reaching the target $(T_{\tau^\star})$. We see an increase in the total cost $C_{\tau^\star}$ since both the terms in the cost function ($T_{\tau^\star}$ and $\psi y_{\tau^\star}(T_{\tau^\star})$) are increasing.

4.5 Summary

In this chapter we studied the joint evolution of popularity and spread of content in a mobile opportunistic setting. We proposed a possible application framework for such a system. We developed a continuous time Markov model for the joint evolution and derived its fluid limit. Finally we showed that a time-threshold policy is an optimal copying policy for the joint evolution model, to optimize the combined cost of target time and

the amount of wasted copies. We also have reported several interesting insights into the evolution of popularity and the co-evolution of popularity and content delivery, which will help the content producers and distributors understand the interplay of various system parameters. By allowing users to declare the interest and enabling the spread of interest, we extend the notion of a personal wishlist (in e-commerce platforms like Amazon) to a social wishlist. The application also combine the services of coupon delivery (Groupon) with such wishlists, thus allowing targeted discount delivery, wherein brands can identify the users who are interested in the coupon and deliver it to them.

4.6 Appendix

4.6.1 Verification of Kurtz's theorem conditions for SIR-SI model

Kurtz's theorem [88] is used to approximate pure jump Markov processes by ordinary differential equations in the limit (usually as system size $N \to \infty$). To do so, we must derive the mean drift rates of an appropriately scaled version of the system, and study the limit of these mean drift rates as the scaling goes to ∞. In our case, for the joint evolution model described in Section 4.2, we can use the Equation 4.1 and Table 4.2 to write down the mean drift rates. Let $Z = (B, D, X_b, X_d, Y)$ denote the current state of the CTMC $Z^N(t)$. Recall the equation,

$$F^N(Z) = \sum_{i \in \mathcal{E}} R_i(Z)\delta_i(Z)$$

where $i \in \mathcal{E}$ indexes the epoch type, $R_i(Z)$ and $\delta_i(Z)$ are the respective transition rates and corresponding state changes given the current state of the CTMC $Z^N(t) = Z$, as given in Table 4.3. Let $\tilde{Z} = \frac{Z}{N}$. The expected drift rate of the CTMC $\tilde{Z}^N(t)$, can then be written as:

$$\begin{aligned}
F_b^N(\tilde{Z}) &= \beta_N \tilde{D} \\
F_d^N(\tilde{Z}) &= -\beta_N \tilde{D} + N\lambda_N \Gamma \tilde{D}\tilde{S} \\
F_{x_b}^N(\tilde{Z}) &= \beta_N \tilde{X}_d + N\lambda_N(\tilde{B} - \tilde{X}_b)(\tilde{X} + \tilde{Y}) \\
F_{x_d}^N(\tilde{Z}) &= \Gamma N\lambda_N(\tilde{D} - \tilde{X}_d)\tilde{Y} - \beta_N \tilde{X}_d + \Gamma N\lambda_N \tilde{X}_d\tilde{S} + N\lambda_N(\tilde{D} - \tilde{X}_d)(\tilde{X} + \tilde{Y}) \\
F_y^N(\tilde{Z}) &= -\Gamma N\lambda_N \tilde{D}\tilde{Y} + N\lambda_N \sigma(\tilde{S} - \tilde{Y})(\tilde{X}_b + \tilde{Y} + (1 - \Gamma)\tilde{X}_d)
\end{aligned}$$

Epoch type i	**Rate** $R_i(Z)$	**State update** $\delta_i(Z)$
$\mathcal{D}-\mathcal{X}_d$ recovers	$\beta_N(D-X_d)$	(1,-1,0,0,0)
$\mathcal{X}_d$ recovers	$\beta_N X_d$	(1,-1,1,-1,0)
$\mathcal{B}-\mathcal{X}_b$ meets $\mathcal{X}+\mathcal{Y}$	$\lambda_N(B-X_b)(X+Y)$	(0,0,1,0,0)
$\mathcal{D}-\mathcal{X}_d$ meets $\mathcal{Y}$	$\lambda_N(D-X_d)Y$	(0,0,0,1,0) + (0,1,0,1,-1) w.p. Γ
$\mathcal{D}-\mathcal{X}_d$ meets $\mathcal{X}$	$\lambda_N(D-X_d)X$	(0,0,0,1,0)
$\mathcal{X}_d$ meets $\mathcal{Y}$	$\lambda_N X_d Y$	(0,1,0,1,-1) w.p Γ
$\mathcal{S}-\mathcal{Y}$ meets $\mathcal{X}_b+\mathcal{Y}$	$\lambda_N(X_b+Y)(S-Y)$	(0,0,0,0,1) w.p. σ
$\mathcal{S}-\mathcal{Y}$ meets $\mathcal{D}-\mathcal{X}_d$	$\lambda_N(D-X_d)(S-Y)$	(0,1,0,0,0) w.p. Γ
$\mathcal{S}-\mathcal{Y}$ meets $\mathcal{X}_d$	$\lambda_N(X_d)(S-Y)$	(0,1,0,1,0) w.p. Γ (0,0,0,0,1) w.p$(1-\Gamma)\sigma$

Table 4.3: Transition rates and state updates for various possible epochs in the SIR-SI process, when the current state of the CTMC $Z^N(t)$ is given by $Z=(B,D,X_b,X_d,Y)$

and the corresponding o.d.e. equations,

$$\dot{b} = \beta d \tag{4.9}$$

$$\dot{d} = -\beta d + \lambda\Gamma d s \tag{4.10}$$

$$\dot{x_b} = \beta x_d + \lambda(b-x_b)(x+y) \tag{4.11}$$

$$\dot{x_d} = \Gamma\lambda(d-x_d)y + \Gamma\lambda x_d s + \lambda(d-x_d)(x+y) - \beta x_d \tag{4.12}$$

$$\dot{y} = -\Gamma\lambda d y + \lambda\sigma(s-y)(x_b+y+(1-\Gamma)x_d) \tag{4.13}$$

where $s(t)=1-b(t)-d(t)$ and $x(t)=x_b(t)+x_d(t)$. Let $f_b, f_d, f_{x_b}, f_{x_d}, f_y$ denote the right hand sides of the above fluid limit equations and let $f:=(f_b,f_d,f_{x_b},f_{x_d},f_y)$. In [98], the following four conditions are shown to be equivalent to the necessary conditions for Kurtz's theorem, and thus we verify them in order to prove Theorem 4.2.1.

- $f(z)$ is Lipschitz continuous.
- $F^N(z)$ uniformly converges to $f(z)$ in z
- In the scaled process $\tilde{Z}^N(t)$, the jump rates are $O(N)$ and the drifts are $O(K^{-1})$
- $\lim_{N\to\infty}\mathbb{P}\big(||(\tilde{Z}^N(0)-z(0))||>\epsilon\big)=0$

(i) **Lipschitz property** It is straightforward to compute the entries of the Jacobian matrix $Df(\cdot)$ i.e., the partial derivatives $\frac{\partial f_u}{\partial v}$ where $u,v\in(b,d,x_b,x_d,y)$, and observe

that they are bounded when $(b, d, x_b, x_d, y) \in [0,1] \times [0, 1-b] \times [0, b] \times [0, d] \times [0, 1-b-d]$. Thus the norm of Jacobian $||D\ f(\cdot)||$ is uniformly bounded, and it follows that $f(z)$ is Lipschitz.

(ii) **Uniform Convergence** Taking $N\lambda_N \to \lambda$ and $\beta_N \to \beta$ we see that the uniform convergence of $F^N(z)$ to $f(z)$ is straightforward.

(iii) **Bounded Noise variance** Since in the co-evolution system the maximum jump rate out of a state is bounded $\lambda_N N(N-1) + \beta_N N$, and since $N\lambda_N \to \lambda$ and $\beta_N \to \beta$, the jump rate is $O(N)$; Table 4.3 lists the drifts for the $Z^N(t)$ process, and for the $\tilde{Z}^N(t)$ they need to be scaled down by N. Thus, it can be seen that the increments for the scaled process $\tilde{Z}^N(t)$ are $O(N^{-1})$. This is referred to as "hydrodynamic scaling" [98]. This ensures that the necessary noise convergence conditions (as described in [98]) are met.

(iv) **Convergence of initial conditions** The initial conditions are chosen such that $(\tilde{B}^N(0), \tilde{D}^N(0), \tilde{X}_b^N(0), \tilde{X}_d^N(0), \tilde{Y}^N(0)) = (b(0), d(0), x_b(0), x_d(0), y(0))$

Since $F^N(.)$ and $f(.)$ satisfy the above four conditions, by [98], Theorem 4.2.1 follows. □

4.6.2 Kamke-dominance

Let $\mathbf{w}(t)$ be the solution of the o.d.e $\dot{\mathbf{w}} = f(\mathbf{w}; z)$ with piecewise Lipschitz continuous control $z(t)$, where f is continuously differentiable and Lipschitz in $\mathbf{w}$ and z. Let $\mathbf{w}^{(1)}(t)$ and $\mathbf{w}^{(2)}(t)$ be the trajectories corresponding to two controls $z^{(1)}(t)$ and $z^{(2)}(t)$ respectively, i.e.,

$$\dot{\mathbf{w}}^{(1)} = f(\mathbf{w}^{(1)}; z^{(1)}) \tag{4.14}$$

$$\dot{\mathbf{w}}^{(2)} = f(\mathbf{w}^{(2)}; z^{(2)}) \tag{4.15}$$

Motivated by the Kamke condition (see [103]) we define Kamke-dominance between two controls $z^{(1)}(t)$ and $z^{(2)}(t)$ as follows.

Definition 1. We say $z^{(1)}(t)$ *Kamke-dominates* $z^{(2)}(t)$ *in the system* $f(\mathbf{w}; z)$ if for each t_0 where $\mathbf{w}^{(1)}(t_0) \geq \mathbf{w}^{(2)}(t_0)$ and $\exists i$ with $w_i^{(1)}(t_0) = w_i^{(2)}(t_0)$ then

$$f_i(\mathbf{w}^{(1)}(t_0); z^{(1)}(t_0)) \geq f_i(\mathbf{w}^{(2)}(t_0); z^{(2)}(t_0))$$

Lemma 4.6.1. Let $\mathbf{w}^{(1)}(t)$ and $\mathbf{w}^{(2)}(t)$ be as in Equations (4.14) and (4.15), and $z^{(1)}(t)$ Kamke-dominate $z^{(2)}(t)$ in the system $f(\mathbf{w}; z)$. If $\mathbf{w}^{(1)}(0) \geq \mathbf{w}^{(2)}(0)$ then $\forall t$, $\mathbf{w}^{(1)}(t) \geq \mathbf{w}^{(2)}(t)$.

Proof. The proof of Proposition 1.1 in [103] can be adopted directly in writing this proof. Since f is continuously differentiable and Lipschitz in $\mathbf{w}$ and z, by the Cauchy-Lipschitz condition, we have a unique solution given the initial conditions. Let $\phi_t^{(1)}(\mathbf{w}^{(1)}(0))$ denote the solution of (4.14) with initial condition $\mathbf{w}^{(1)}(0)$, and $\phi_t^{(2)}(\mathbf{w}^{(2)}(0))$ denote the solution of (4.15) with initial condition $\mathbf{w}^{(2)}(0)$. We wish to show that $\phi_t^{(1)}(\mathbf{w}^{(1)}(0)) \geq \phi_t^{(2)}(\mathbf{w}^{(2)}(0))$. Consider m an integer, and let $\phi_t^{(1),m}(\mathbf{w}^{(1)}(0))$ be the solution corresponding to

$$\dot{\mathbf{w}}^{(1)} = f(\mathbf{w}^{(1)}; z^{(1)}) + \big(\frac{1}{m}\big)\mathbf{e}$$

Then, by continuity of o.d.e. solutions with respect to drift and initial conditions [104, Chap.1, Lemma 3.1], $\phi_s^{(1),m}(\mathbf{w}^{(1)}(0) + \frac{\mathbf{e}}{m})$ defined on $0 \leq s \leq t$ for all large m, say $m > M$ and

$$\phi_s^{(1),m}(\mathbf{w}^{(1)}(0) + \frac{\mathbf{e}}{m}) \to \phi_s^{(1)}(\mathbf{w}^{(1)}(0)) \tag{4.16}$$

as $m \to \infty$, uniformly in $s \in [0, t]$, where $\mathbf{e} = (1, 1, \cdots, 1)$. We claim that for $0 \leq s \leq t$, for all $m > M$,

$$\phi_s^{(1)m}(\mathbf{w}^{(1)}(0) + \frac{\mathbf{e}}{m}) >> \phi_s^{(2)}(\mathbf{w}^{(2)}(0)) \tag{4.17}$$

where $x >> y$ implies $x_i > y_i$, $\forall i$.

Proof of claim Fix $m > M$. We know that $\mathbf{w}^{(1)}(0) + \frac{\mathbf{e}}{m} = \phi_0^{(1),m}(\mathbf{w}^{(1)}(0) + \frac{\mathbf{e}}{m}) >> \mathbf{w}^{(2)}(0) = \phi_0^{(2)}(\mathbf{w}^{(2)}(0))$. By continuity of $\phi^{(1)}$ and $\phi^{(2)}$, the claim (4.17) holds for small s. If the claim (4.17) were false, $\exists\ 0 < t_0 \leq t$ such that (4.17) holds for $0 \leq s < t_0$, and an index i such that,

$$\big(\phi_{t_0}^{(1),m}(\mathbf{w}^{(1)}(0) + \frac{\mathbf{e}}{m})\big)_i = \big(\phi_{t_0}^{(2)}(\mathbf{w}^{(2)}(0))\big)_i$$

and

$$\frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=t_0} \big(\phi_s^{(1),m}(\mathbf{w}^{(1)}(0) + \frac{\mathbf{e}}{m})\big)_i \leq \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=t_0} \big(\phi_s^{(2)}(\mathbf{w}^{(2)}(0))\big)_i \tag{4.18}$$

But since $\forall j \neq i$

$$\big(\phi_{t_0}^{(1)m}(\mathbf{w}^{(1)}(0) + \frac{\mathbf{e}}{m})\big)_j \geq \big(\phi_{t_0}^{(2)}(\mathbf{w}^{(2)}(0))\big)_j$$

and $z^{(1)}(t)$ Kamke-dominates $z^{(2)}(t)$ in the system $f(\mathbf{w}, z)$, we have

$$\begin{aligned} & f_i(\phi_{t_0}^{(1),m}(\mathbf{w}^{(1)}(0) + \frac{\mathbf{e}}{m}); z^{(1)}(t_0)) + \frac{1}{m} \\ & > f_i(\phi_{t_0}^{(1),m}(\mathbf{w}^{(1)}(0) + \frac{\mathbf{e}}{m}); z^{(1)}(t_0)) \\ & \geq f_i(\phi_{t_0}^{(2)}(\mathbf{w}^{(2)}(0)); z^{(2)}(t_0)) \end{aligned}$$

which implies

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=t_0} \left(\phi_s^{(1),m}(\mathbf{w}^{(1)}(0) + \frac{\mathbf{e}}{m})\right)_i > \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=t_0} \left(\phi_s^{(2)}(\mathbf{w}^{(2)}(0))\right)_i$$

This contradicts (4.18) and hence proves the claim (4.17). Applying (4.16) to this claim completes the proof of the lemma. □

Consider the system of equations representing the evolution of $(x_b(t), x_d(t), y(t))$ in the SIR-SI process.

$$\dot{x_b} = \beta x_d + \lambda(b - x_b)(x + y) \tag{4.19}$$

$$\dot{x_d} = \Gamma\lambda(d - x_d)y + \Gamma\lambda x_d s + \lambda(d - x_d)(x + y) - \beta x_d \tag{4.20}$$

$$\dot{y} = -\Gamma\lambda dy + \lambda\sigma(s - y)(x_b + y + (1 - \Gamma)x_d) \tag{4.21}$$

For identifying the optimal control $\sigma(t)$ it is sufficient to restrict our attention to Equations (4.19)-(4.21), since by the definition of the SIR-SI process, $(b(t), d(t))$ are independent of the control, i.e., the popularity evolution is fixed and not influenced by the content dissemination. Let $w^{(1)} = (x_b^{(1)}, x_d^{(1)}, y^{(1)})$ and $w^{(2)} = (x_b^{(2)}, x_d^{(2)}, y^{(2)})$ be the solutions respectively for the controls $\sigma^{(1)}(t)$ and $\sigma^{(2)}(t)$ for the above system with $\sigma^{(1)}(t) \leq \sigma^{(2)}(t)$. Define

$$(u_b, u_d, v) := (x_b^{(1)}, x_d^{(1)}, y^{(1)}) - (x_b^{(2)}, x_d^{(2)}, y^{(2)})$$

and $\Delta\sigma(t) = \sigma^{(1)}(t) - \sigma^{(2)}(t)$. Then we can write,

$$\dot{u_b}|_{u_b=0} = \beta u_d + \lambda(b - x_b^{(1)})(u_d + v) \tag{4.22}$$

$$\dot{u_d}|_{u_d=0} = \Gamma\lambda(d - x_d^{(1)})v + \lambda(d - x_d^{(1)})(u_b + v) \tag{4.23}$$

$$\dot{v}|_{v=0} = \lambda(s - y^{(1)})[\Delta\sigma y^{(1)} + \sigma^{(1)}u_b + \Delta\sigma x_b^{(2)} + (1 - \Gamma)(\sigma^{(1)}u_d + \Delta\sigma x_d^{(2)})] \tag{4.24}$$

Lemma 4.6.2. Control Domination

1. If $\forall t$, $\sigma^{(1)}(t) \geq \sigma^{(2)}(t)$, then $\sigma^{(1)}(t)$ Kamke-dominates $\sigma^{(2)}(t)$ in the system given by equations (4.19)-(4.21).
2. If $\forall t$, $y^{(1)}(t) \geq y^{(2)}(t)$, then $y^{(1)}(t)$ Kamke-dominates $y^{(2)}(t)$ in the system given by equations (4.19)-(4.20).

Proof. 1. From equations (4.22)-(4.24). Since $\Delta\sigma \geq 0$, we find that,

- if $u_d, v \geq 0$ and $u_b = 0$, then $\dot{u}_b \geq 0$
- if $u_b, v \geq 0$ and $u_d = 0$, then $\dot{u}_d \geq 0$
- if $u_b, u_d \geq 0$ and $v = 0$ $\dot{v} \geq 0$

This verifies the conditions of Definition 1, and thus proves (a).

2. From equations (4.22)-(4.23). Since $v \geq 0$, we find that,

- if $u_d \geq 0$ and $u_b = 0$, then $\dot{u}_b \geq 0$
- if $u_b \geq 0$ and $u_d = 0$, then $\dot{u}_d \geq 0$

This verifies the conditions of Definition 1, and thus proves (b).

□

Lemma 4.6.3. [(a)]

1. If $\forall t$, $\sigma^{(1)}(t) \geq \sigma^{(2)}(t)$, and $w^{(1)}(0) \geq w^{(2)}(0)$, then $\forall t$, $w^{(1)}(t) \geq w^{(2)}(t)$
2. If $\forall t$, $y^{(1)}(t) \geq y^{(2)}(t)$, and $(x_b^{(1)}(0), x_d^{(1)}(0)) \geq (x_b^{(2)}(0), x_d^{(2)}(0))$, then $\forall t$,

$$(x_b^{(1)}(t), x_d^{(1)}(t)) \geq (x_b^{(2)}(t), x_d^{(2)}(t))$$

Proof. The proof follows by applying Lemma 4.6.1 to the systems in Lemma 4.6.2 (a) and (b). □

Define $\mathbf{w}(t) = (x_b(t), x_d(t), y(t))$ We will now use the above lemmas to prove the claims in Section 4.3.1, in order to establish the optimality of a time-threshold policy. Recall that α has been fixed earlier. Recall the definitions of T_σ and $y_\sigma(T_\sigma)$. We shall replace them with T_τ and $y_\tau(T_\tau)$ whenever $\sigma(t) = \sigma_\tau(t)$, a time-threshold policy.

Consider τ such that $\tau \geq T_\tau =: \tau'$. Evidently, $T_{\tau'} = \tau'$, and $y_\tau(T_\tau) = y_{\tau'}(T_{\tau'})$, $\forall \tau \geq \tau'$.

Define $\mathcal{T} = \{\tau : \tau \geq 0\}$ and $\underline{\mathcal{T}} = \{\tau : \tau \leq T_\tau\} \subset \mathcal{T}$. Observe that $\{\rho : \exists\ \tau \in \mathcal{T}, y_\tau(T_\tau) = \rho\} = \{\rho : \exists\ \tau \in \underline{\mathcal{T}}, y_\tau(T_\tau) = \rho\}$. Hence we limit our discussion to $\tau \in \underline{\mathcal{T}}$.

Lemma 4.6.4. The set $\{\rho : \exists\ \tau \in \underline{\mathcal{T}}, y_\tau(T_\tau) = \rho\}$ forms an interval of the form $[0, \rho_{\max}]$.

Proof. Setting $\tau = 0$, i.e., $\sigma(t) = 0$, $\forall t$, we see from Equation (4.21), since $y(0) = 0$, we have $y(t) = 0$, $\forall t$. This establishes that $\rho = 0$ belongs to the set. It is also easy to observe that $\rho_{\max}$ is achieved by τ such that $\tau = T_\tau$. By the intermediate value theorem, it now suffices to show that $y_\tau(T_\tau)$ is a continuous function of τ.

Consider $\tau, \tau + \delta \in \underline{\mathcal{T}}$, i.e., $\tau \leq T_\tau$, $\tau + \delta \leq T_{\tau+\delta}$. Observe from Lemma 4.6.3 , for $\delta > 0$, $T_{\tau+\delta} \leq T_\tau$. With the above constraints, evidently the only case to be considered is

$$\tau < \tau + \delta \leq T_{\tau+\delta} \leq T_\tau$$

Let $\mathbf{w}_\tau(.)$ and $\mathbf{w}_{\tau+\delta}(.)$ be the trajectories corresponding to $\sigma_\tau(t)$ and $\sigma_{\tau+\delta}(t)$, with $\mathbf{w}_\tau(0) = \mathbf{w}_{\tau+\delta}(0)$. Since $\sigma_\tau(t) = \sigma_{\tau+\delta}(t)$, $0 \leq t \leq \tau$, it follows that $\mathbf{w}_\tau(\tau) = \mathbf{w}_{\tau+\delta}(\tau)$. Further, from Equations (4.19)-(4.21) we observe that,

$||\mathbf{w}_\tau(\tau + \delta) - \mathbf{w}_{\tau+\delta}(\tau + \delta)||_2 \leq C_1\delta$, where

$$C_1 = \sqrt{(\beta + \lambda)^2 + (2\Gamma\lambda + \lambda + \beta)^2 + (\lambda + \Gamma\lambda)^2}$$

$\forall t > \tau + \delta$, $\sigma_\tau(t) = \sigma_{\tau+\delta}(t)$ and hence by continuity of o.d.e. system w.r.t initial conditions at $\tau + \delta$ implied by the Lipschitz nature of f w.r.t $\mathbf{w}$ (see [105]), $\mathbf{w}_\tau(.)$ is continuous w.r.t τ.

Recall that $\tau < \tau + \delta \leq T_{\tau+\delta} \leq T_\tau$. By the just observed continuity, for $\epsilon > 0$, we can obtain $\delta > 0$ such that, $|x_{\tau+\delta}(T_{\tau+\delta}) - x_\tau(T_{\tau+\delta})| \leq \epsilon$, i.e.,

$$|\alpha a(\infty) - x_\tau(T_{\tau+\delta})| \leq \epsilon$$

However over the interval $(T_{\tau+\delta}, T_\tau]$, the rate of increase of $x_\tau(.)$ is bounded below by $\lambda a(t)^2(1 - \alpha)\alpha$ (from Equations (4.19)-(4.21)). Hence,

$$|T_{\tau+\delta} - T_\tau| \leq \frac{\epsilon}{\lambda a(t)^2(1 - \alpha)\alpha}$$

which can be made small by choosing an appropriate $\delta > 0$. We have thus shown that T_τ continuous w.r.t τ. $\tau \in \underline{\mathcal{T}}$.

To show the continuity of $y_\tau(T_\tau)$, consider,

$$|y_\tau(T_\tau) - y_{\tau+\delta}(T_{\tau+\delta})| \leq$$
$$|y_\tau(T_\tau) - y_\tau(T_{\tau+\delta})| + |y_\tau(T_{\tau+\delta}) - y_{\tau+\delta}(T_{\tau+\delta})|$$

In the above equation, the first term on the right hand side can be made arbitrarily small by the continuity of T_τ w.r.t τ and the fact that $y_\tau(.)$ is a continuous trajectory, and the second term can be made arbitrarily small by the continuity of $\mathbf{w}_\tau(.)$ w.r.t τ. Thus $y_\tau(T_\tau)$ is a continuous function of τ, in the space of threshold policies. And since $y_\tau(T_\tau) : \tau \to \rho$ and $\tau \in [0, \infty)$, we see that $\rho \in [0, \rho_{max}]$. □

Lemma 4.6.5. Let $\sigma(t)$ be a policy such that $y_\sigma(T_\sigma) < \rho_{\max}$. Then there exists a threshold policy $\sigma_\tau(t)$ whose cost is no worse than that of $\sigma(t)$.

Proof. Since $\rho := y_\sigma(T_\sigma) < \rho_{\max}$, there exists $\tau \geq 0$ such that $y_\tau(T_\tau) = \rho$. We will argue that $\sigma_\tau(t)$ is such that $T_\tau \leq T_\sigma$.

Let $\mathbf{w}_\tau(t)$ and $\mathbf{w}_\sigma(t)$ be the system trajectories for the controls $\sigma_\tau(t)$ and $\sigma(t)$ respectively with $\mathbf{w}_\tau(0) = \mathbf{w}_\sigma(0)$. By definition, for $t \leq \tau$, $\sigma_\tau(t) \geq \sigma(t)$ and hence, by Lemma 4.6.3, $\mathbf{w}_\tau(\tau) \geq \mathbf{w}_\sigma(\tau)$.

Suppose, contrary to the claim, $T_\tau > T_\sigma$. Evidently, this cannot happen if for all t, $0 \leq t \leq T_\sigma$, $y_\tau(t) \geq y_\sigma(t)$. This is because, by Lemma 4.6.3, we will have $x_\tau(t) \geq x_\sigma(t)$ and hence $T_\tau \leq T_\sigma$.

Hence, there exists t_0, $\tau \leq t_0 < T_\sigma$, such that $y_\tau(t_0) = y_\sigma(t_0)$. Then we have $y_\tau(t_0) = y_\sigma(t_0) > \rho$ (since $y_\tau(T_\tau) = \rho$ and $\tau \leq t_0 < T_\tau$, and for $t \geq \tau$, $\dot{y}_\tau(t) < 0$). Also, for $t \geq t_0$, $\sigma_\tau(t) = 0 \leq \sigma(t)$. Thus from (4.21), $\forall s \in \{s : s > t_0, y_\tau(s) = y_\sigma(s)\}$, $\dot{y}_\tau(s) \leq \dot{y}_\sigma(s)$. Thus for $s \geq t_0$, $y_\tau(s) \leq y_\sigma(s)$. Since $y_\tau(T_\tau) = y_\sigma(T_\sigma) = \rho$, this implies $T_\tau \leq T_\sigma$.

□

Lemma 4.6.6. Consider a non-threshold policy $\sigma(t)$ which achieves $\rho > \rho_{max}$. Consider the time-threshold policy $\hat{\sigma}(t)$ of the form,

$$\hat{\sigma}(t) = \begin{cases} 1, & 0 \leq t < \sup\{t : \sigma(t) > 0\} \\ 0, & \text{otherwise} \end{cases}$$

Then $\hat{\sigma}(t)$ policy achieves a smaller total cost (given by Equation 4.7) than $\sigma(t)$.

Proof. Let $\sigma(t)$ and $\sigma_{\hat{\tau}}(t)$ be as above. Let $\mathbf{w}_{\hat{\tau}}(t)$ and $T_{\hat{\tau}}$ be as defined earlier for time-threshold policies, corresponding to $\sigma_{\hat{\tau}}(t)$. Then by Lemma 4.6.4, $y_{\hat{\tau}}(T_{\hat{\tau}}) \leq \rho_{max} < \rho =$

$y_\sigma(T_\sigma)$, and by Lemma 4.6.3, $T_{\hat{\tau}} \leq T_\sigma$, and hence the $\sigma_{\hat{\tau}}(t)$ policy achieves a smaller total cost than $\sigma(t)$. □

CHAPTER 5

Content Competition on Social Networks

Recall that Chapters 2 and 3 were focused on studying the spread of a single item of content or influence. In Chapter 4, we studied a co-evolution model, where we considered multilayered diffusion, i.e., interest evolution at the lower layer, and the content spread (dependent on interest evolution) at the higher layer. However, in each layer, the diffusion process spread in isolation, i.e., we did not model competing diffusions. In the first part of this chapter (Sections 5.1), we study competing epidemics, spreading on a common population. Though such models have been studied in the past (see Section 1.1.4 for a brief literature survey), all of them assume that the users are part of a single social network. However in the real world, users might spend time on several social networks, and in turn the content creators need to optimally allocate their resources across multiple social networks. In Sections 5.1, we look at one such model for competition between content creators over multiple social networks.

Another aspect of content competition that is largely unexplored in the context of online social networks, is how an end user receives the content, and thereby gets influenced by it. Most online social networks display the contents reaching an end user as a reverse-chronological list of items, which we shall refer to as the *timeline*. Though such a structure is present on most online social networks, not much attention has been given in prior literature to modeling the user behavior in processing the timeline. In the latter part of this chapter (Sections 5.2), we propose two models (for timelines of finite size and infinite size), to study such a competition over visibility in a user's timeline.

5.1 Competition over Multiple Social Networks

In this work, we characterize the competition between two content creators for a common user base, via two social networks. Each social network is characterized by its level of activity and its popularity among the consumers. There is a maximum number of resources that each content creator can use in each of the networks, and the aim is to optimize the net budget for competition. We study the non-cooperative game between the two content creators and characterize the best response, which exhibits a threshold behavior. We obtain the Nash equilibria from the intersections of the best response functions and study the impact of cost parameters on the equilibrium budget allocations of each content creator.

5.1.1 System Model

In this section, we will introduce the system model and justify the associated assumptions by considering examples from real world networks. We begin with the description of the competition framework and then describe the social network.

Competition Framework: We restrict our attention to publish-subscribe based networks, i.e., networks in which there are sources which generate content and users follow the sources. Consider M content sources trying to spread their content to a consumer population through K different such networks. We primarily deal with digital content, say news articles, popular song videos or podcasts. We also assume that the content sources specialize in the same genre of content, and hence there is a competition among them to attract a larger fraction of the common user base. We also assume that the contents are exclusive in nature, i.e., a particular consumer would adopt the first content he comes across, and is no longer interested in others. This is true in cases such as news podcasts (CNN, BBC,etc.) covering the same sports event, media streaming sites (Youtube, Vimeo, etc.) providing the same music video, etc. Figure 5.1 shows the various components (content sources, social network and the user population) in the system. We look at the scenario where the users could be present in one of several such networks at any given point of time. These networks vary in their popularity in the user population, and also in the level of interaction within the network. Thus each network j is characterized by a visit probability per user, γ_j, and a meeting rate within the network, λ_j. Note that $\sum_{j \in K} \gamma_j \leq 1$. The system parameters γ_j and λ_j can be interpreted as follows: in each

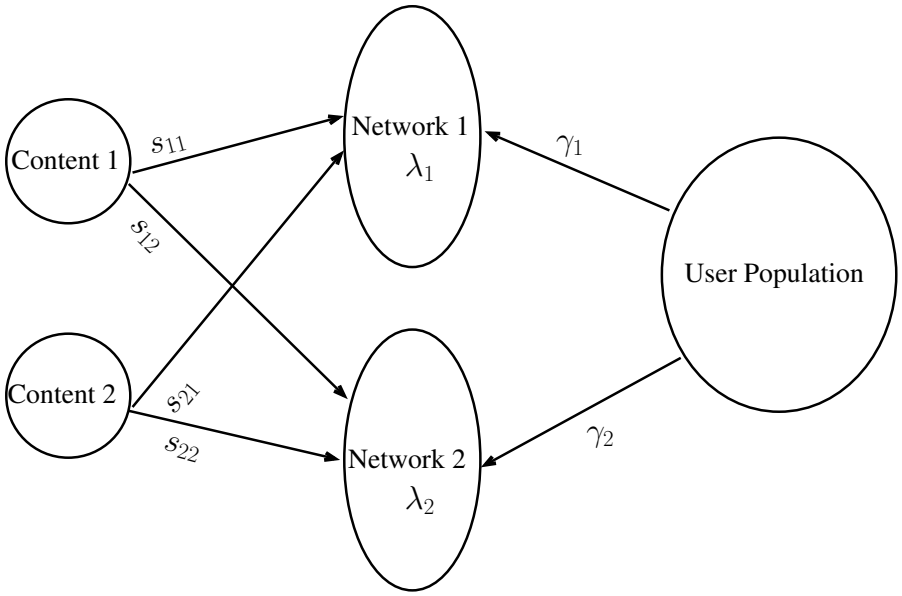


Figure 5.1: Interaction between various components of the system. Sources recruit promotional accounts (s_{ij}) in each of the social networks, and users visit each network with a probability γ_j. Once inside the network, they receive updates from the promotional accounts at the rate λ_j

time slot, every user chooses which network to be present in, according to the probability γ_j, and once in Network j, he/she visits the sources' pages at the rate λ_j.

Though there are several online social networks, each of them have evolved to provide a unique service, and hence it is common for users to maintain accounts simultaneously on multiple social networks. For instance, Gmail is primarily intended for personal and professional communications among a small group of individuals. Facebook is used to stay in touch with our social circle (acquaintances and friends) and broadcast contents to all (or a majority) of friends. Twitter is primarily used as a content-sharing network, where users follow (unidirectional) content sources for latest updates. γ_j represents how users divide their attention across these various social networks. This will highly depend on individual preferences, but for simplicity we have assumed this to be constant for the user population under consideration.

λ_j is dependent on the network, since different networks may operate on different time scales and have varying mechanisms by which users access the content. For instance, it has been empirically observed [24], that the dynamics of news cycles on Facebook and Twitter occur at different rates, with Twitter being more up to date. This may be due to the different reasons for which the networks have evolved, as hinted earlier. Consider a single user who has subscribed to a mailing list on Gmail, has *liked* a page on Facebook

(thus receives updates) and also follows a particular account on Twitter. Receiving half a dozen updates about a content from a single account over the period of a day may be perceived as normal on Twitter, while on Facebook it might be annoying and on Gmail subscriptions it might border on spamming. Thus content sources are restricted to using a λ_j as dictated by the time-scale at which the network operates.

Each content source creates aliases in each of the networks, to spread his particular content. This can be done by creating accounts/pages on these networks to promote their content, and recruit personnel to manage and update these accounts/pages. We denote by s_{ij}, the fraction of promoters for content i in network j. The net budget spent by a given content source i is then $\sum_j s_{ij}$. We will also assume that there is a cost per unit budget for each source and is given by ϕ_i. We also assume there is a maximum limit to the number of promotion accounts that can be created on each network. s_{ij}'s are hence normalized with respect to the user population, i.e., $s_{ij} \leq 1$. The aim of the content provider would then be to optimize the net budget to maximize his revenue. The exact utility to be optimized is provided in Section 5.1.1.

For this work, we will restrict our attention to two content sources and social networks (M=2, K=2), though the methodology and the results in this work can be generalized for the (M, K) case.

Social Network: Consider the social network as shown in Figure 5.2. We have considered Network 1 for example. s_{i1} denotes the promotion accounts promoting content from source i. The total set of users in the population is normalized to 1 and the fraction of users currently present in Network 1 is given by γ_1. Among these, $\gamma_1 x_1$ fraction of users have already consumed content of source 1 and $\gamma_1 x_2$ fraction of users have consumed content of source 2. Hence the competition between the sources is for the fraction of users who haven't consumed either of the content and are currently present in Network 1, i.e., $\gamma_1(1 - x_1 - x_2)$. These users receive content updates from each of the sources (in s_{11} or s_{21}) at rate λ_j and depending on which source reaches them first, they get converted to x_1 or x_2. We can use this to write down the system dynamics.

We will now write down the system of o.d.e.'s for the system evolution under fluid limit assumptions [88] with the total user population normalized to 1. Let x_1, x_2 denote the fraction of users who have seen content 1 and 2 respectively ($x_1 + x_2 \leq 1$). The system

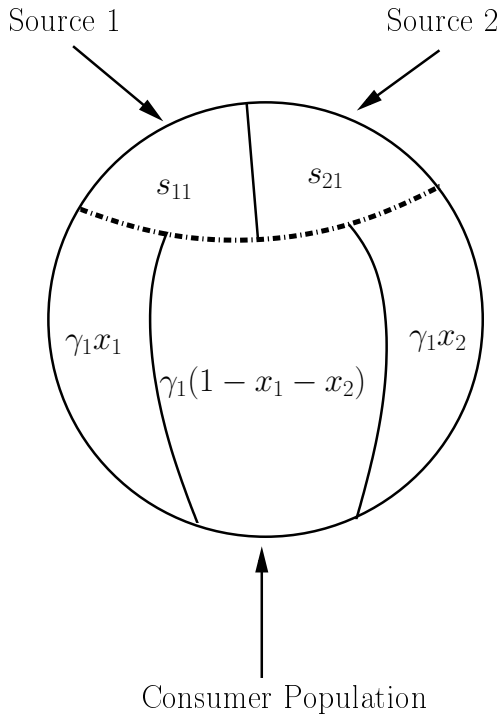


Figure 5.2: The different subsets of users within a given social network (Social Network 1 is considered as an example). s_{i1} represents the promotion accounts related to source i and x_i denotes the fraction of total user population that has consumed content i. γ_1 is the probability with which a user is present in Network 1.

evolution can then be given by,

$$\dot{x}_1 = (s_{11}\lambda_1\gamma_1 + s_{12}\lambda_2\gamma_2)(1 - x_1 - x_2) \tag{5.1}$$

$$\dot{x}_2 = (s_{21}\lambda_1\gamma_1 + s_{22}\lambda_2\gamma_2)(1 - x_1 - x_2) \tag{5.2}$$

In the above equations, the first term represents the fraction of nodes delivered the content in Network 1 (by s_{i1}), and the second term represents the fraction of nodes delivered the content in Network 2 (by s_{i2}). Let $a = \lambda_1\gamma_1$ and $b = \lambda_2\gamma_2$. The system evolution o.d.e.'s then become:

$$\dot{x}_1 = (as_{11} + bs_{12})(1 - x_1 - x_2) \tag{5.3}$$

$$\dot{x}_2 = (as_{21} + bs_{22})(1 - x_1 - x_2) \tag{5.4}$$

Let X_i represent the final fraction of users who have seen content i, i.e., $X_i := x_i(\infty)$ from

equations (5.3) and (5.4). By solving the above system we get,

$$X_i = \frac{as_{i1} + bs_{i2}}{as_{i1} + bs_{i2} + as_{j1} + bs_{j2}} \tag{5.5}$$

Note that since we are restricting ourselves to publish-subscribe based networks, we do not consider content sharing among users in this work. Modeling information spread in such system taking into account content sharing is a promising future direction. This will require a realistic modeling of the graph structure within each of these networks. This would also involve addressing issues such as, users sharing contents across the social networks (sharing an article read on Facebook to his followers on Twitter), and the distinction between user sharing and promotion from sources (users generally tend to share only once, since they are not interested in maximizing the visibility of any particular content).

Game Formulation: We model the competition between the two content creators (players) as a non-cooperative game, with strategies (s_{11}, s_{12}) for player 1 and (s_{21}, s_{22}) for player 2. s_{ij}'s are normalized with the user population, hence $s_{ij} \leq 1$. The general utility function for player i is of the form,

$$U_i = X_i - \phi_i(s_{i1} + s_{i2})$$

where X_i represent the final fraction of users who have seen content i, as given in Equation 5.5 and ϕ_i is the cost per unit budget for content source i. We have made the cost depend only on the source, assuming that he does not pay differently, the personnel involved in spreading his content on different networks. Section 5.1.3 has discussions on a variant of the cost structure, which depends both on the source and the corresponding social network.

We see that $\lambda_j\gamma_j$ serves as a measure of how efficient a network is for spreading information quickly to the common pool of consumers. This is intuitive, since λ_j indicates the rate at which you can generate content in Network j and γ_j indicates how often users are present in Network j. Without loss of generality, assume $a, b > 0$, with $a > b$ i.e. $\lambda_1\gamma_1 > \lambda_2\gamma_2$. Then if we fix the total budget, say $B_i = s_{i1} + s_{i2}$, then we can show that in the optimal allocation of the net budget (assuming a cost structure independent of the network) is, $s_{i1} = \min(B_1, 1)$ and $s_{i2} = \min((B_1 - 1)^+, 1)$. This implies that since Network 1 is *better* than Network 2 ($a > b$), given a total budget, it is optimal to invest in Network 1 until the maximum limit is reached, and then invest in Network 2. It is evident that,

even in the case of a general K network system, the networks can be ordered in decreasing order of $\lambda_j \gamma_j$ and the budget can be allocated beginning with the *best* network and proceeding to the next network once the maximum limit is reached.

Define,

$$z_1 = a \min(B_1, 1) + b \min((B_1 - 1)^+, 1)$$

$$z_2 = a \min(B_2, 1) + b \min((B_2 - 1)^+, 1)$$

We can then rewrite the utilities as function of the budgets (B_1, B_2).

$$U_1 = \frac{z_1}{z_1 + z_2} - \phi_1 B_1$$

$$U_2 = \frac{z_2}{z_1 + z_2} - \phi_2 B_2$$

Hence we have,

$$U_i = \begin{cases} \frac{aB_i}{aB_i + z_j} - \phi_i B_i, & B_i \leq 1 \\ \frac{a - b + bB_i}{a - b + bB_i + z_j} - \phi_i B_i, & 1 \leq B_1 \leq 2 \end{cases} \tag{5.6}$$

Thus, by observing the optimal allocation across the social networks, we have reformulated the utility in terms of a single strategy for each user, i.e., the net budget B_i. We can now solve for the Nash equilibrium of the above non-cooperative game by studying the best response dynamics of (B_1, B_2).

Observe that when $B_2 = 0$, the strategy $B_1 = \epsilon > 0$ yields a utility of $U_1 = 1 - \phi_1 \epsilon$ for player 1 and $\forall \epsilon > 0$, $B_1 = \frac{\epsilon}{2}$ yields a higher utility. But, by the system definition, $B_1 = 0$ yields zero utility. Thus there is no best response for $B_2 = 0$, and the utility function U_1 is discontinuous at $B_1 = 0$. Hence we require $B_i \geq \delta > 0$, the minimum budget for each player. Also, since $s_{ij} \leq 1$, it is sufficient to restrict the strategy set of each user to $B_i \in [\delta, 2]$.

5.1.2 Best Response Dynamics

In this section, we first establish the existence of a Nash equilibrium and then proceed to characterize the best response function for each player. We can then obtain the Nash equilibrium as the intersection of the best response functions.

5.1.3 Existence of PSNE

Theorem 5.1.1 (Debreu, Glicksberg, Fan [89])**.** Consider a strategic form game $< \mathcal{I}; (S_i); (u_i) >$, where $\mathcal{I}$ is a finite set. Assume that the following conditions hold for each $i \in \mathcal{I}$.

- S_i is a non-empty, convex, and compact subset of a finite-dimensional Euclidean space.
- $u_i(s)$ is continuous in s.
- $u_i(s_i, s_{-i})$ is quasi-concave in s_i

Then the game $< \mathcal{I}; (S_i); (u_i) >$ has a pure strategy Nash equilibrium.

Since the set of strategies (the net budget), $B_i \in [\delta, 2]$ the first condition is satisfied. From the utility functions (5.6), the remaining two conditions can be verified, and hence the game has a pure strategy Nash equilibrium.

Characterizing the Best Response Function: Let $BR_i(B_j)$ be the best response of player i corresponding to the strategy B_j by player j. Let $U_i^*(B_j)$ be the corresponding maximum utility. We then have the following cases:

Case (i) $BR_i(B_j) = \delta$, $U_i^*(B_j) = \frac{a\delta}{a\delta + z_2} - \phi_1\delta$

Case (ii): $0 < BR_i(B_j) < 1$

Setting $\frac{dU_i}{dB_i} = 0$ in (5.6) for $B_i \leq 1$, we get

$$BR_i(B_j) = \frac{1}{a}\left(\sqrt{\frac{z_2 a}{\phi_1}} - z_2\right)$$

$$U_i^*(B_j) = \left(1 - \sqrt{\frac{z_2\phi_1}{a}}\right)^2$$

Case (iii): $BR_i(B_j) = 1$, $U_i^*(B_j) = \frac{a}{a+z_2} - \phi_1$

Case (iv): $1 < BR_i(B_j) < 2$

Setting $\frac{dU_i}{dB_i} = 0$ in (5.6) for $1 \leq B_i \leq 2$, we get

$$BR_i(B_j) = \frac{1}{b}\left(\sqrt{\frac{z_2 b}{\phi_1}} - a + b - z_2\right)$$

$$U_i^*(B_j) = \left(1 - \sqrt{\frac{z_2\phi_1}{b}}\right)^2 + \phi_1\frac{a-b}{b}$$

Case (v): $BR_i(B_j) = 2$, $U_i^*(B_j) = \frac{a+b}{a+b+z_2} - 2\phi_1$

We note that the best response can belong to any of the above cases, and hence by comparing the maximum utility $U_i^*(B_j)$ obtained in each of the cases, we can construct the best response function.

Define,

$$f_1 = \frac{1}{a}\left(\sqrt{\frac{z_2 a}{\phi_1}} - z_2\right)$$

$$g_1 = \frac{1}{b}\left(\sqrt{\frac{z_2 b}{\phi_1}} - a + b - z_2\right)$$

If $f_1 \geq g_1$ (which is true for sufficiently large ϕ_1), then we have the following structure for the best response function.

Let Θ_{f_1}, Θ'_{f_1} be the values of B_2 for which $f_1 = 1$ and similarly Θ_{g_1}, Θ'_{g_1} be the values of B_2 for which $g_1 = 1$. The best response function $BR_1(B_2)$ is of the following form

$$BR_1(B_2) = \begin{cases} \max(f_1, \delta), & 0 < B_2 < \Theta_{f_1} \\ 1, & \Theta_{f_1} < B_2 < \Theta_{g_1} \\ \min(g_1, 2), & \Theta_{g_1} < B_2 < \Theta'_{f_1} \\ 1, & \Theta'_{g_1} < B_2 < \Theta'_{f_1} \\ \max(f_1, \delta), & B_2 > \Theta'_{f_1} \end{cases}$$

The max and min terms feature, in order to ensure that the budget remains in the feasible set $[\delta, 2]$. We can similarly characterize $BR_2(B_1)$. The PSNE of the game can then be obtained as the intersection of the best response curves.

Threshold Behavior of Resource Allocation: Figure 5.3 shows the structure of the best response function. An interesting feature that arises out of this structure is that, in the intervals $\Theta_{f_1} < B_2 < \Theta_{g_1}$ and $\Theta'_{g_1} < B_2 < \Theta'_{f_1}$, it is optimal to use the budget $B_1 = 1$, i.e., exploit the resources in Network 1 to the maximum, but refrain from investing in Network 2. Especially as the opponent (player 2) increases his budget from δ to 2, player 1 increases his investments in Network 1. When Network 1 is saturated for Player 1, i.e.,

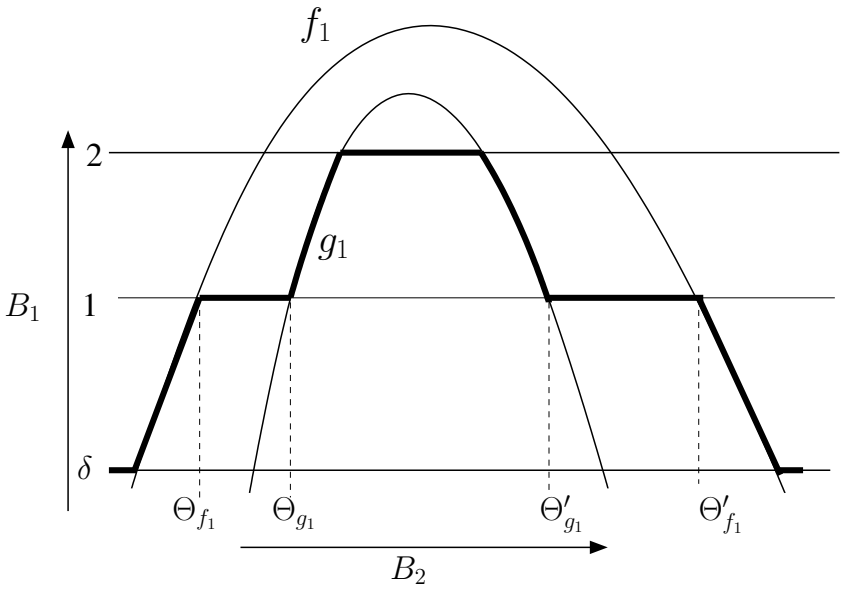


Figure 5.3: Structure of the Best Response Function

$s_{11} = 1$, he does not immediately begin investing in Network 2. Thus we observe there is an interval $\Theta_{f_1} < B_2 < \Theta_{g_1}$ when the optimal response remains at $B_1 = 1$. A similar behavior is observed as Player 2 begins approaching his maximum budget $B_2 = 2$. In such a scenario, it is no longer optimal to maintain $B_1 = 2$ and hence Player 1 starts lowering his budget (thus reducing his investment in Network 2). Once he relieves all his resources in Network 2, it is observed that Player 1 does not immediately begin reducing resource usage in Network 1. Instead there is an interval $\Theta'_{g_1} < B_2 < \Theta'_{f_1}$, where it is still optimal to operate his resources in Network 1 at full capacity. Such a behavior is similar to hysteresis curves generally observed in magnetism and thermodynamics. There are also several observed instances of hysteresis in economics [106].

We can also infer that the intervals become larger, as the difference in the networks' efficiency ($\lambda_1\gamma_1 - \lambda_2\gamma_2$) increases. For instance, when the networks are equally efficient ($\lambda_1\gamma_1 = \lambda_2\gamma_2$), then the problem is similar to a Cournot duopoly model [89] with the inverse demand function $P(B_1, B_2) = \frac{1}{B_1+B_2}$ and a budget constraint. Such a demand function is termed as *iso-elastic*, since the quantity demanded is reciprocal to price, and thus reflects a case where the consumers spend a constant amount on the commodity, irrespective of the price [107]. Also, for instance, when $\lambda_2\gamma_2 = 0$, and $\lambda_1\gamma_1 > 0$, we can see that the players will never invest in Network 2.

Remarks on Cost Structure: We have assumed that a source i pays ϕ_i per unit budget. This assumption is valid, when the payment/cost incurred per promotional account

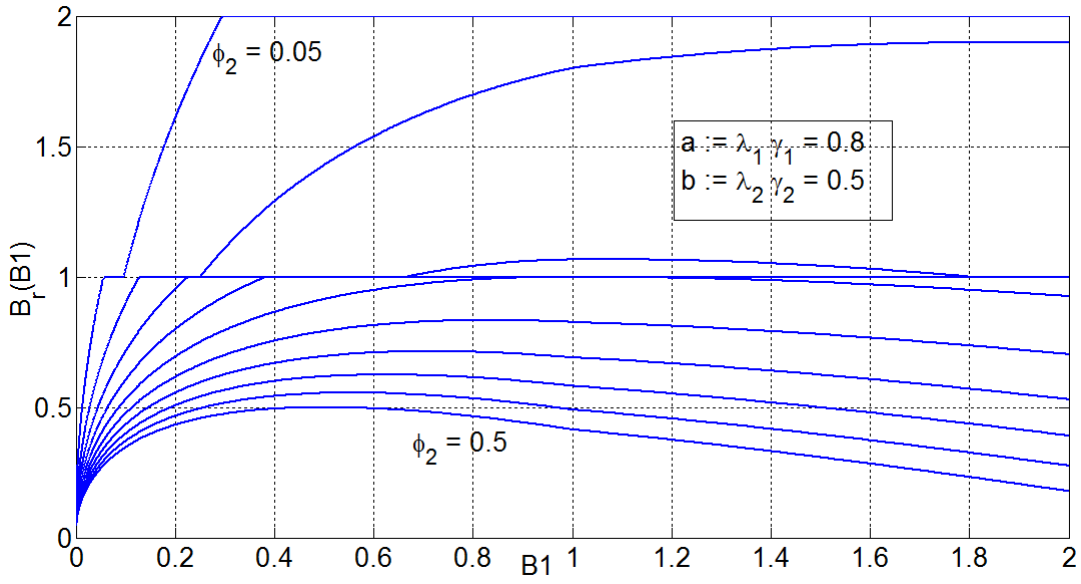


Figure 5.4: Best response function for player 2 $BR_2(B_1)$ for various values of ϕ_2

is independent of the social network. This may not be the case, and in general, various social networks could charge differently for promotional accounts on their site. Twitter and Facebook have their own mechanisms for promoting content spread, and charge the users per account. With slight abuse of notation, let ϕ_j denote the cost per unit budget, in the social network j. In such a case, the total cost incurred by source i would be $\phi_1 s_{i1} + \phi_2 s_{i2}$. As a special case, when $\phi_j = C\lambda_j\gamma_j$, i.e. proportional to the network efficiency, then the utility functions will be of the form,

$$U_i = \frac{Q_i}{Q_i + Q_{i'}} - CQ_i \tag{5.7}$$

where $Q_i = \lambda_1\gamma_1 s_{i1} + \lambda_2\gamma_2 s_{i2}$. Thus we again get a variation on the Cournot duopoly game [89] with inverse demand function $P(Q_1, Q_2) = \frac{1}{Q_1+Q_2}$ and a budget constraint.

5.1.4 Numerical Results

In this section, we numerically study the best response function for various values of cost per unit budget ϕ_i and also compute the Nash equilibrium numerically.

Best response function: Figure 5.4 shows the best response function for player 2 $BR_2(B_1)$ for various values of player 2's cost per unit budget ϕ_2. We observe that for increasing values of ϕ_2, player 2 becomes more conservative, i.e., if $\phi_2' < \phi_2^o$, $BR_2'(B_1) \geq BR_2^o(B_1)$ for all values of B_1. We can also see that the simulations verify the threshold

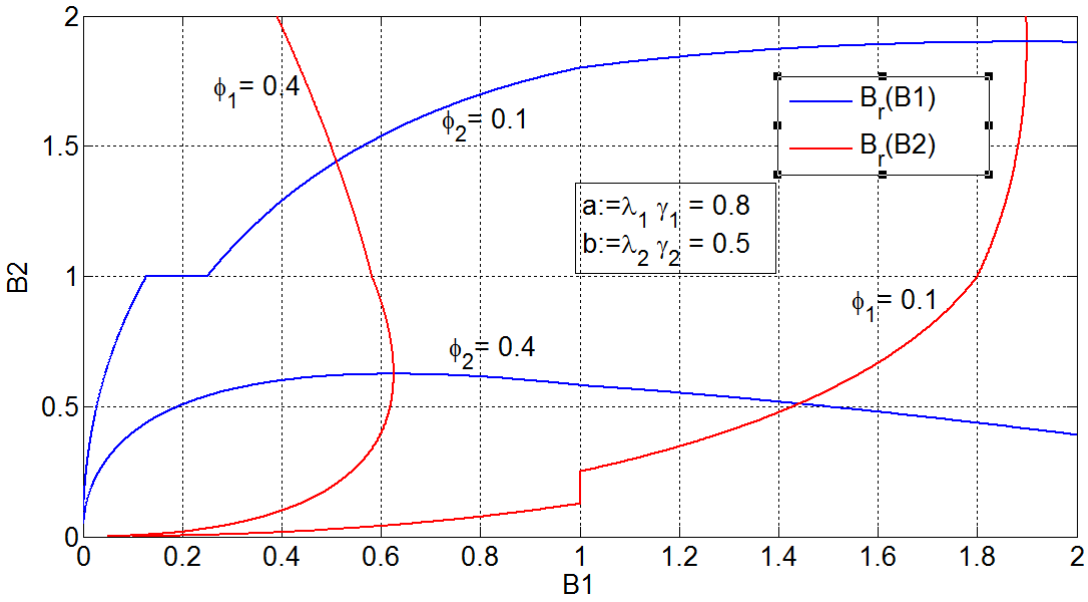


Figure 5.5: Pure strategy Nash equilibrium obtained as the intersection of the best response curves. The blue curves represent the best responses of player 2 $BR_2(B_1)$ and the red curves represent the best responses of player 1 $BR_1(B_2)$.

behavior observed in Section 5.1.2.

Nash equilibrium: Figure 5.5 shows the computation of Nash equilibrium obtained by intersection of the best response functions. We can see that, for fixed cost per unit budget of player 2 ($\phi_2 = 0.4$) the equilibrium budget of player 1 is greater for $\phi_1 = 0.1$ than when $\phi_1 = 0.4$. This implies that the equilibrium budget for competition increases with decrease in the player's own cost per unit budget (as observed in Section 5.1.4). Also, we see that for fixed cost per unit budget of player 1 ($\phi_1 = 0.4$), the equilibrium budget of player 1 is greater for $\phi_2 = 0.4$ than when $\phi_2 = 0.1$. This implies that the equilibrium budget for competition increases with increase in the opponent's cost per unit budget.

5.2 Competition over a Timeline

In this section, we study competition for visibility and influence between several contents on a single user's timeline. We begin with a timeline of fixed size in Section 5.2.1, and derive the occupancy distribution of the timeline in Section 5.2.2. We then formulate a non-cooperative game among N content creators, whose utility is the probability of finding their content on the timeline in Section 5.2.3. We derive explicit equilibrium rate expressions under a symmetric cost scenario, and show that it exhibits non-monotonic behavior, with increase in number of players N.

We then proceed to model influence evolution for a timeline of infinite size (Section 5.2.4), with a behavioral model for user scanning the timeline. We obtain integral equations, capturing the evolution of influence for a single content creator in isolation (Section 5.2.5), and in the presence of competing contents (Section 5.2.6). We show that the system behaves like an $M/G/1$ queue *with immediate discount*. We then finally study non-cooperative games among the content creators, and provide some interesting insights.

5.2.1 Finite Timeline model

Though practically every user in the social network can create and share content, we will restrict our attention in this work to the setting where we have clear delineation between consumers and content creators (brands and organizations). Thus, we characterise the content creator-consumer interaction as a bipartite graph, with $\mathcal{I}$ the set of consumers and $\mathcal{C}$ the set of creators. The set of followers (or subscribers) of creator c is denoted by $M_c \in \mathcal{I}$. Content from source c is created according to a Poisson process with rate/intensity parameter ν_c and sent to the timelines of all members in M_c. Each consumer in the social network, say i, has a set $N_i \in \mathcal{C}$ of sources that it follows (see Figure (5.6)). In this work, we will be considering only a single user's timeline, and hence in this work, $N_i = \mathcal{C}$, and let $N = |\mathcal{C}|$. Define the total rate of content creation in the social network which will appear on the timeline of user i as

$$\nu = \sum_{c \in \mathcal{C}} \nu_c$$

Define further

$$\nu_{[-c]} = \sum_{c' \in \mathcal{C}, c' \neq c} \nu_{c'}$$

We are interested in the stationary probability $\pi(i, c)$ of finding a content of source c on the timeline of a user i. Assume that at the beginning, i.e., at time 0, there is an arrival of some content from c to i. Assume the timeline of user i has a constant size of K. Each future arrival to i from $\mathcal{C}$, will push the c's content further down, and after K such shifts the content will no longer be visible on the timeline of user i. In the meantime, if another post arrives from c to i, there is still a presence of content from c on the timeline of user i. We call an ic-busy period the time period that starts when a content from c first arrives to the timeline of user i until the instant at which there is no more content from c on the timeline of user i. We denote this by $\hat{T}_{ic}$. Before characterising the ic-busy period, we introduce the following notation. When content from c arrives, it is said to be

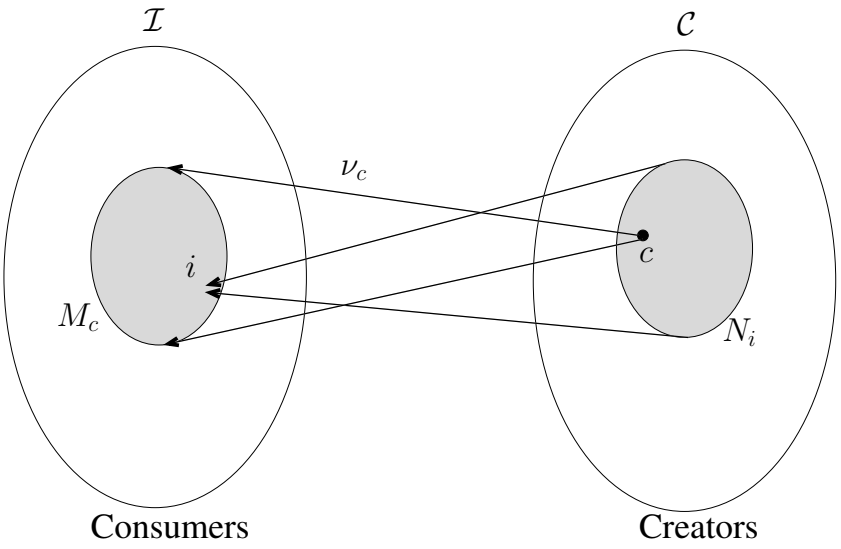


Figure 5.6: System model.

on the top of the timeline, i.e., at position K, and at each new arrival, its position gets decremented/shifted by 1. Let $T_{ic}(k)$ be the time from the instant when the content from c on the timeline of i is at position k on the timeline, until the instant when the timeline of user i does not have any content from c. Then $\hat{T}_{ic} = T_{ic}(K)$.

5.2.2 Occupancy of the finite timeline

In this section we obtain the occupancy distribution of the timeline positions. In order to do that, we represent the timeline as a continuous time Markov chain (CTMC) $\mathbf{C}(t)$ taking values in $\mathcal{C}^K$. $\mathbf{C}(t)$ denotes the vector of contents that occupy the K locations in the timeline of i at time t. Then we can state the following theorem.

Theorem 5.2.1. The stationary probability distribution for the CTMC $\mathbf{C}(t)$ is given by,

$$\pi_{\mathbf{c}} = \Pi_{k=1}^{K(i)} \frac{\nu_{c_k}}{\nu} \tag{5.8}$$

where c_k is the content at the kth position on the timeline.

Proof. Let Q denote the transition rate matrix for the CTMC $\mathbf{C}(t)$. We wish to show that the proposed $\pi_{\mathbf{c}}$ satisfies $\forall \mathbf{c}' \in \mathcal{C}^K$

$$\sum_{\mathbf{c} \in \mathcal{C}^K} \pi_{\mathbf{c}} q_{\mathbf{c},\mathbf{c}'} = 0$$

or equivalently,

$$\pi_{\mathbf{c}'} a_{\mathbf{c}'} = \sum_{\substack{\mathbf{c} \in \mathcal{C}^K \\ \mathbf{c} \neq \mathbf{c}'}} \pi_{\mathbf{c}} q_{\mathbf{c},\mathbf{c}'} \tag{5.9}$$

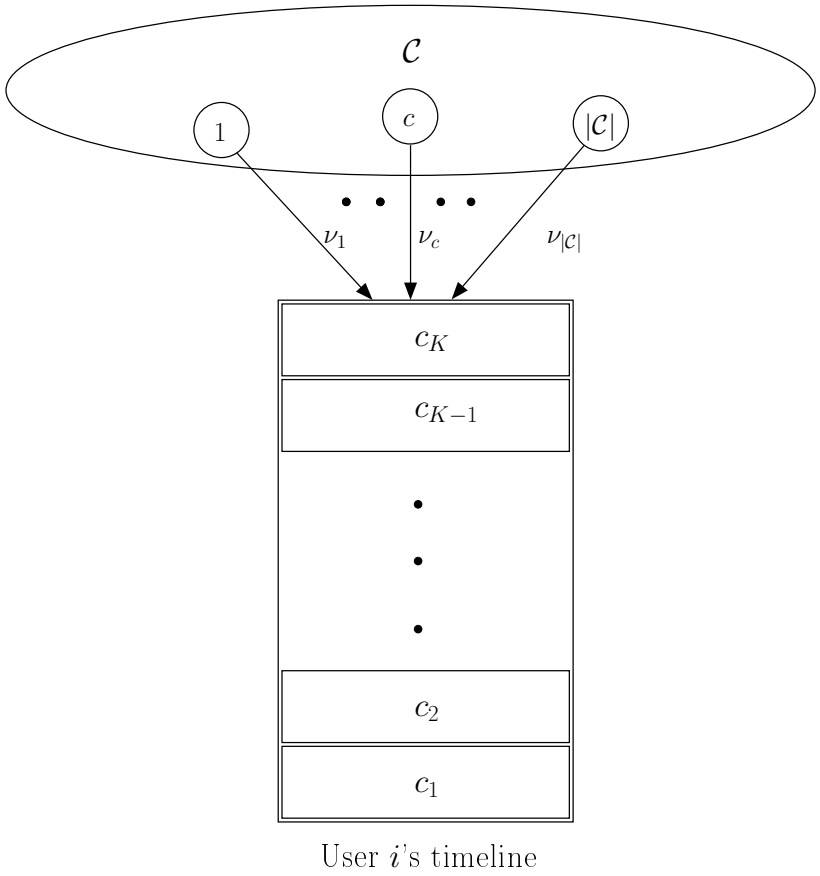


Figure 5.7: Schematic of the model for the timeline of user i.

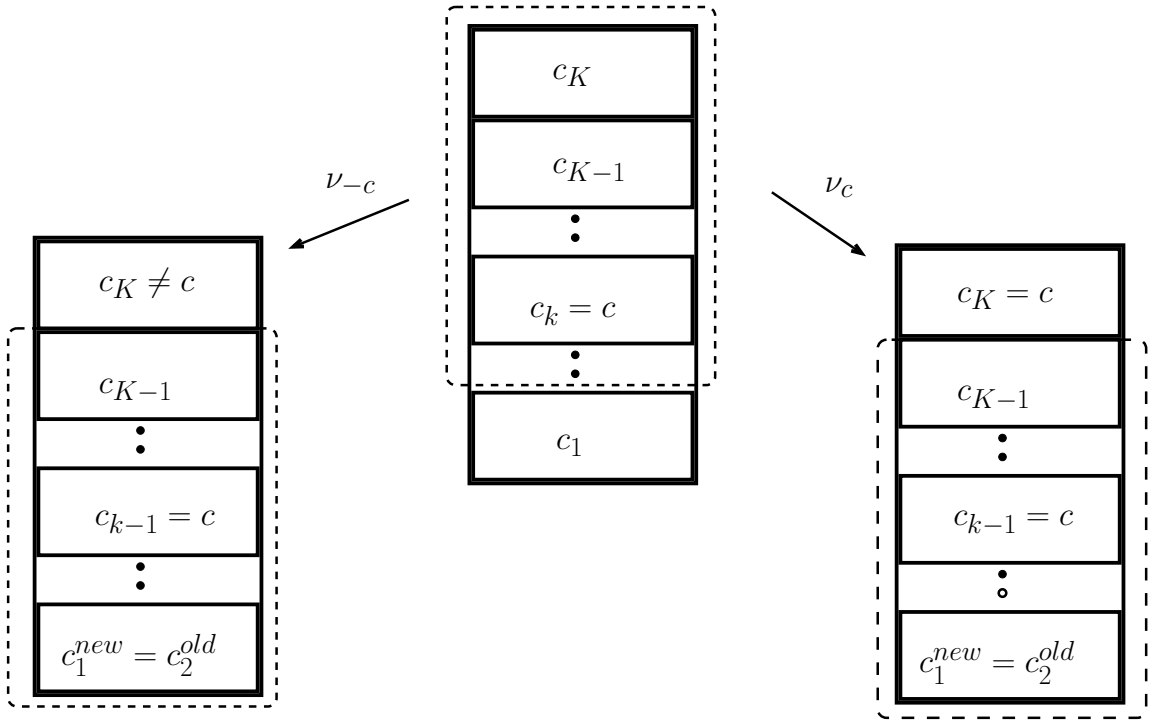


Figure 5.8: Evolution of user i's timeline.

where $a_{\mathbf{c}}$ is the rate of exiting state $\mathbf{c}$. The argument that follows holds for the case where not all the elements of $\mathbf{c}'$ are the same. Observe from Figure (5.8) that there is a possible transition from state $\mathbf{c}$ to $\mathbf{c}'$ if and only if the first $(K-1)$ entries of $\mathbf{c}$ coincide with the last $(K-1)$ entries of $\mathbf{c}'$. If a given $\mathbf{c}'$ satifies this condition then we say $\mathbf{c}' \in s(\mathbf{c})$. Let $l := c'_K$, then the rate $q_{\mathbf{c},\mathbf{c}'} = \nu_l$ if $\mathbf{c}' \in s(\mathbf{c})$ and 0 otherwise. The right hand side of Equation (5.9) then becomes

$$\begin{aligned}
\sum_{\mathbf{c}\in\mathbf{C}^K,\mathbf{c}\neq\mathbf{c}'} \pi_{\mathbf{c}} q_{\mathbf{c},\mathbf{c}'} &= \sum_{\substack{\mathbf{c}\in\mathcal{C}^K,\mathbf{c}\neq\mathbf{c}' \\ \mathbf{c}'\in s(\mathbf{c})}} \left(\Pi_{k=1}^{K(i)} \frac{\nu_{c_k}}{\nu}\right) \nu_l \\
&= \nu_l \left[\frac{\nu_1}{\nu} + \cdots + \frac{\nu_{|\mathcal{J}|}}{\nu}\right] \left(\Pi_{k=1}^{K(i)-1} \frac{\nu_{c_k}}{\nu}\right) \\
&= \nu_l \Pi_{k=1}^{K(i)-1} \frac{\nu_{\mathbf{c}_k}}{\nu} \\
&= \pi_{\mathbf{c}'} a_{\mathbf{c}'}
\end{aligned}$$

The first equality arises since there are exactly $|\mathcal{C}|$ possible terms in the summation, and $\sum_{l=1}^{|\mathcal{C}|} \nu_l = \nu$. If all the elements of $\mathbf{c}'$ are the same, say ℓ, then the following happens:

$$\begin{aligned}
\pi_{\mathbf{c}'} a_{\mathbf{c}'} &\stackrel{?}{=} \sum_{\substack{\mathbf{c}\in\mathcal{C} \\ \mathbf{c}\neq\mathbf{c}'}} \pi_{\mathbf{c}} q_{\mathbf{c},\mathbf{c}'} \\
\nu_l^K \sum_{j\neq l} \nu_j &\stackrel{?}{=} \left(\sum_{\substack{j=1 \\ j\neq l}}^{|\mathcal{C}|} \nu_l^{(K-1)} \nu_j\right) \nu_l
\end{aligned}$$

which is true. In the above equation, there are only $(|\mathcal{C}|-1)$ entries on the right hand side, since $\mathbf{c} \neq \mathbf{c}'$, i.e., the last position in $\mathbf{c}$ cannot be l . □

Expected *ic*-busy period: In this section, we characterise the expected *ic*-busy period by using a recursion argument. It is also to be noted that, the occupancy distribution derived in the preceding section can also be used to arrive at the same result.

Theorem 5.2.2. The expected *ic*-busy period, the duration for which content from creator c stays on user i's timeline of size K, is given by

$$E[T_{ic}(K)] = \frac{1}{\nu_{-c}} \left(\frac{1-\alpha^{-K}}{1-\alpha^{-1}}\right) \tag{5.10}$$

where

$$\alpha = \frac{\nu_{-c]}}{\nu}.$$

Proof. We have the following recursion for $E[T_{ic}(k)]$:

$$E[T_{ic}(0)] = 0,$$

$$E[T_{ic}(k+1)] = \frac{1}{\nu} + \frac{\nu_j}{\nu} E[T_{ic}(K)] + \frac{\nu_{-c}}{\nu} E[T_{ic}(k)]$$

By taking $k = (K-1)$ we obtain after some algebra,

$$E[T_{ic}(K)] - E[T_{ic}(K-1)] = \frac{1}{\nu_{-c}} \tag{5.11}$$

To solve the recursion, let us introduce for $k = 1, ..., K$

$$S_{ic}(k) := E[T_{ic}(k)] - E[T_{ic}(k-1)]$$

This satisfies the simple recursion $S_{ic}(k+1) = \alpha S_{ic}(k)$, hence, $S_{ic}(k+1) = \alpha^k S_{ic}(1)$. From Equation (5.11) we have that $S_{ic}(K) = \frac{1}{\nu_{-c}}$ and thus

$$S_{ic}(1) = \frac{1}{\nu_{-c}} \alpha^{-K+1}.$$

We combine this with the relation,

$$E[T_{ic}(k+1)] = S_{ic}(1) \sum_{n=0}^{k} \alpha^n,$$

and thus we get,

$$\begin{aligned} E[T_{ic}(K)] &= \frac{1}{\nu_{-c}} \sum_{n=0}^{K-1} \alpha^{-n} \\ &= \frac{1}{\nu_{-c}} \left(\frac{1 - \alpha^{-K}}{1 - \alpha^{-1}} \right). \end{aligned} \tag{5.12}$$

□

Note that the expected time during which there is no content from c on the timeline of user i is simply $1/\nu_c$. Hence the fraction of time that there is content from c on the timeline of user i is given as

$$p_{ic} = \frac{E[T_{ic}(K)]}{E[T_{ic}(K)] + 1/\nu_c}$$

$$
\begin{aligned}
&= 1 - \frac{1}{\nu_c E[T_{ic}(K)] + 1} \\
&= 1 - \frac{1}{q_c \sum_{n=0}^{K-1} (1+q_c)^n + 1} \\
&= 1 - \frac{1}{(1+q_c)^K}
\end{aligned} \tag{5.13}
$$

where $q_c := \frac{\nu_c}{\nu_{-c}}$. Note that by definition, $1 + q_c = 1/\alpha$, and thus we have

$$p_{ic} = 1 - \alpha^K \tag{5.14}$$

Note that for $K = 1$, $p_{ic} = \nu_c/\nu$.

5.2.3 Timeline game

The fraction of time during which content from creator c is present on the timeline of user i, is shown in Equation (5.14), which depends on the rate/intensity parameters $\nu_c, c \in \mathcal{C}$, with which the content is created. Therefore, we assume that for each content creator c, the ν_c are decision variables, satisfying $\nu_c \geq \phi_c$, where ϕ_c is the minimum value of the rate of the content created by source c. Player c can put some effort in order to increase the basic arrival rate ϕ_c of its content to ν_c at some cost. This together with the rates of the competitors of c will determine the duration of presence of content from c on i's timeline. We shall seek for the equilibrium values of the variables ν_c, $c \in \mathcal{C}$. ν_c at equilibrium is characterised by the fact that it maximises W_c^i over $\nu_c \geq \phi_c$ where

$$W_c^i := p_{ic} - \gamma_c (\nu_c - \phi_c) \tag{5.15}$$

The Lagrange relaxation gives the minimization of the Lagrangian, substituting for p_{ic} from Equation (5.14) gives us

$$L_c^i = 1 - \alpha^K - \gamma_c (\nu_c - \phi_c) + \beta_c (\nu_c - \phi_c), \tag{5.16}$$

where the Lagrange multipliers β_c are non-negative. For the best response, we solve

$$\frac{\partial L_c^i}{\partial \nu_c} = -K\alpha^{(K-1)} \left(-\frac{\nu_{-c}}{(\nu)^2} \right) - (\gamma_c - \beta_c) = 0 \tag{5.17}$$

This yields:

$$\frac{K}{\nu} \left(1 - \frac{\nu_c}{\nu}\right)^K - (\gamma_c - \beta_c) = 0$$

Assume that the minimum is not attained at $\nu_c = \phi_c$, then by the complementary slackness conditions, $\beta_c = 0$.Thus, the solution of the best response problem is given as

$$\nu_c = \nu \left[1 - \left(\frac{\gamma_c \nu}{K}\right)^{\frac{1}{K}}\right] \tag{5.18}$$

if it is larger than or equal to ϕ_c, otherwise $\nu_c = \phi_c$.

Characterising the equilibrium: Consider the game in which the set of sources $\mathcal{C}$ compete over the timeline of user i. For simplicity of notation, let $N = |\mathcal{C}|$. Each source is a player and we aim to obtain the Nash equilibrium. Consider the symmetric game where the cost parameters are the same for all players, i.e., $\gamma_c = \gamma$. Then we have the following theorem.

Theorem 5.2.3. At equilibrium each player sends either

$$\lambda_c = \frac{K}{\gamma} \frac{(N-1)^K}{N^{(K+1)}} \tag{5.19}$$

or ϕ, whichever is larger.

refsec:cost-structure

Proof. The condition of symmetry gives us $\nu = N\nu$. Thus

$$\nu_c = N\nu_c \left[1 - \left(\frac{\gamma N \nu_c}{K}\right)^{\frac{1}{K}}\right]$$

which leads to Equation (5.19). □

Also observe that at equilibrium by symmetry, we have $\alpha = \frac{N-1}{N}$, thus $\forall\ c$,

$$p^*_{ic} = 1 - \left(1 - \frac{1}{N}\right)^K \tag{5.20}$$

Numerical results In this section we study the effect of various system parameters on the equilibrium rate ν. We will consider the timeline size K, number of players N and the cost parameter γ. Note that ν is a measure of how aggressive the players are, and in real world, the rate of sharing content on the social networks can be measured in units like posts/hour, posts/day, etc. which will depend on the duration for which the user is active on the social network. The tradeoff parameter γ is indicative of the cost paid by the competitors to operate at higher rates than the system minimum ϕ (we set $\phi = 0$ for

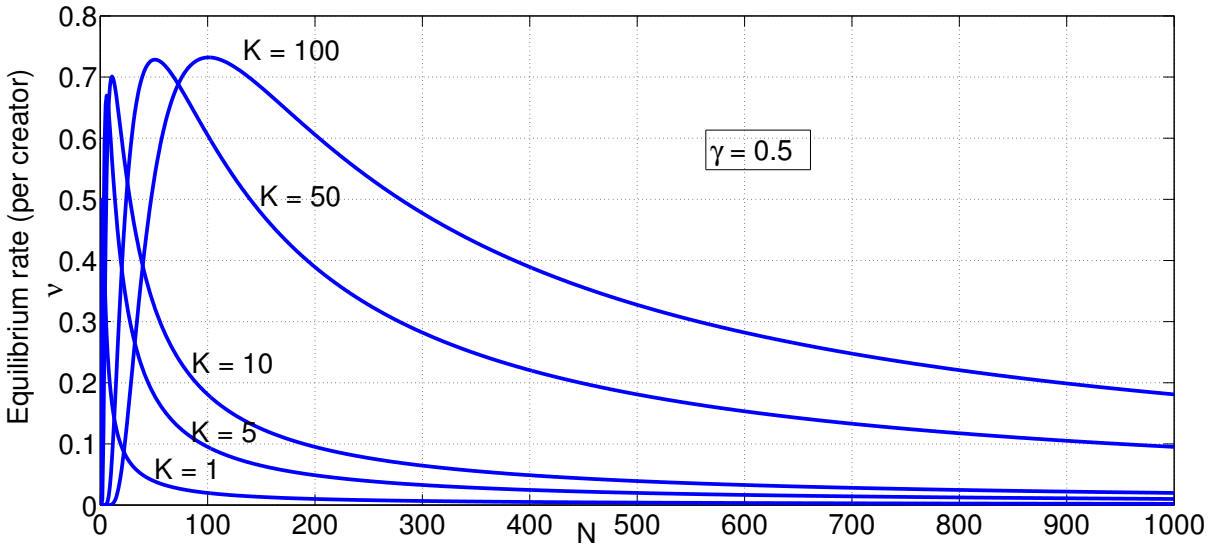


Figure 5.9: Variation of equilibrium rate ν against N for $K = 1, 5, 10, 50, 100$.

the simulations). With respect to the tradeoff parameter, it is straightforward from the expression that $\nu \propto \frac{1}{\gamma}$, which implies as the cost per unit rate increases, the nodes become less aggressive.

The variations with respect to K and N show interesting behavior. Figure (5.9) shows the effect of number of players N at the equilibrium rate ν. We observe that for fixed K, the players are initially less aggressive for small N (less competition) and accelerate their aggression as N increases to a critical value. Beyond the critical N, players become less aggressive since at large N, there are too many competitors for the timeline, that the payoffs (p_{ic}) become considerably smaller. The acceleration and deceleration of aggressive behavior on either side of the critical N are more pronounced for smaller K.

Figure (5.10) shows the effect of timeline size K at the equilibrium rate ν. For fixed N, we see that the players are less aggressive for smaller K since there are fewer slots on the timeline to compete for and the payoffs are lesser compared to the cost. As the number of slots on the timeline increase, the players become more aggressive until a critical K. Beyond the critical K, since there are enough slots on the timeline (for fixed number of players N), players become less aggressive. As earlier, the acceleration and deceleration of aggressive behavior on either side of the critical K are more pronounced for smaller N.

Observe from the plots that there are critical values of N and K for which the players' aggression is maximum. Hence, if a social planner has control over N (by dictating the number of players in the system) and K (by varying the number of posts visible to a consumer in a session), then the social planner can also aim to maximise the total revenue

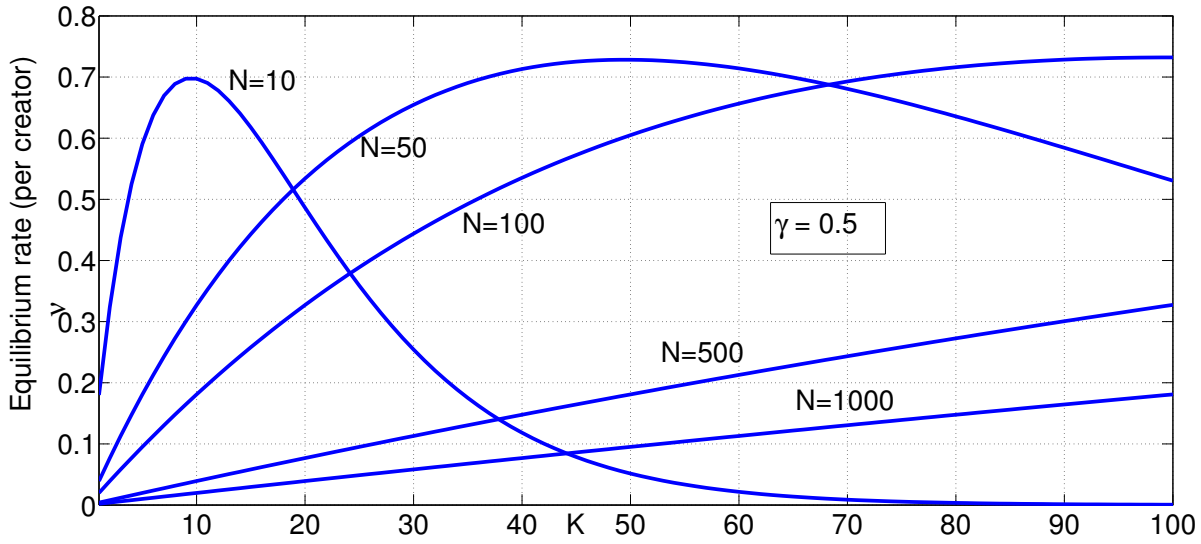


Figure 5.10: Variation of equilibrium rate ν against K for $N = 10, 50, 100, 500, 1000$.

(which is $N\gamma(\lambda - \phi)$).

5.2.4 Infinite Timeline Model

Though the above analysis gave some insights into the competition for visibility between contents, on a single user's timeline, the assumption of a finite timeline is quite restrictive. The timeline could be potentially infinite, but one should somehow also account for the diminishing chances of a user viewing a content further down the timeline. Also, different content creators might generate contents which have different influence weight on the user, which may be interpreted as perceived quality. The model to be discussed addresses the content creator's ability to vary both the quantity (rate of content generation) and the quality (influence weight) of the content, in an infinite timeline setting.

Consider the timeline as shown in Figure 5.11. As earlier, let $\mathcal{C}$ represent the set of content creators with $C = |\mathcal{C}|$. The timeline is reverse-chronologically ordered, and the item of content at position k is identified by (c_k, b_k), where c_k is the source of the content, i.e., $c_k \in \mathcal{C}$, and b_k is the *influence weight* of the item of content. Since we are considering an infinite timeline, the indexing begins at the top (as against the bottom-up indexing in the finite timeline case in Section 5.2.1). In this section, we will restrict ourselves to the case of a single content creator, $(C = 1)$, to aid better understanding of the model. We can easily extend the analysis to a setup with multiple content creators, as shown in Section 5.2.6.

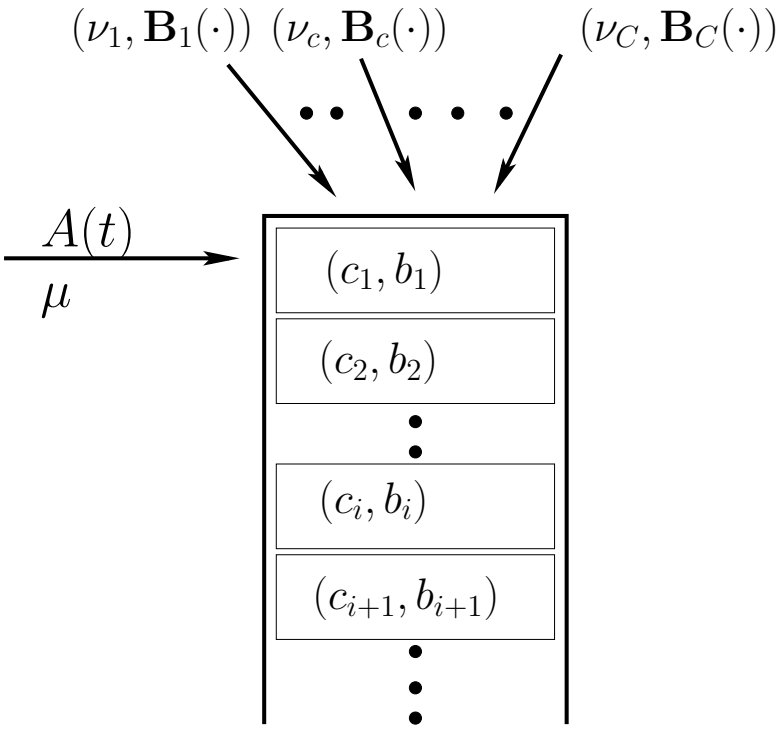


Figure 5.11: Model of an online social network user's timeline

Let new items of content arrive into the timeline in a Poisson process of rate ν. $T_k, k \geq 1$, denote the points of this Poisson process of rate ν. Define $T_0 = 0$, and, for $k \geq 1$, let $Z_k = T_k - T_{k-1}$ denote the inter-arrival times of the new items of content. Let $B_k, k \geq 1$, be the influence brought in by item k which are assumed to be i.i.d. nonnegative random variables. By an abuse of notation, we will denote the common c.d.f. of the random variables $B_k, k \geq 1$ by $B(\cdot)$, keeping in mind the fact that in this paper we will never need to refer to the random variable as a function over the sample space. We assume that $B(\cdot)$ has finite expectation.

The user visits the timeline at instants that constitute the points of a homogeneous Poisson process of rate μ; the counting process for this Poisson process will be denoted by $A(t), t \geq 0$. When the user visits the timeline, we assume that he scans the timeline beginning at the top and terminating after a random number of posts, which is geometrically distributed. In other words, he stops at position j , $j \geq 1$ with probability $\alpha^{j-1}(1-\alpha)$, where $0 < \alpha < 1$ is a given parameter. We refer to α as the *discount factor*, for reasons which will be shortly evident.

It follows that if the influence of the item at position $j, j \geq 1$, (with $j = 1$ denoting the top of the timeline), is b_j, then the expected influence when the user visits the timeline is $v = \sum_{j=1}^{\infty} b_j \alpha^{j-1}$ where, the expectation is over the number of timeline entries seen by the user. Thus, α discounts the influence of the content in position j by α^{j-1}. *We will use this measure to quantify the influence of the timeline at any given time* t. In particular,

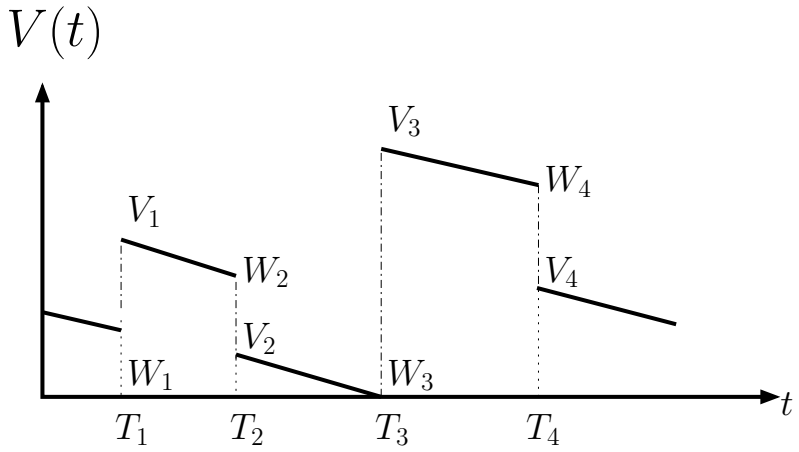


Figure 5.12: Sample trajectory of the $V(t)$ process

if the current influence of the timeline is v, we note that if a new item with influence b is added to the head of this timeline, the new influence of the timeline is easily computed as $b + \alpha v$.

$$\lim_{t\to\infty} \frac{1}{A(t)} \int_0^t V(u)\mathrm{d}A(u)$$

At time t, let $V(t)$ denote the influence of the timeline. This random process accounts for all the arrivals of new items in $[0, t]$, and the timeline scanning model of the user. We observe that random changes in $V(t)$ take place only at the content arrival instants $T_k, k \geq 1$, and in between these instants we assume that the value of $V(t)$ *decreases at a constant rate* (1 after appropriate scaling of time). This decrease in $V(t)$ is used to model the decay of influence of a particular item of content with time and thus the time critical nature of the item of content, i.e., older content contributes lower influence. This also ensures that, if the timeline has not changed between two visits of a user, the latter visit produces less influence than the former. We assume in this work that this decay is linear with time, but the analysis can be extended to other forms of decay, for instance, exponential decay with time.

Thus we see that a content's influence can decrease due to (a) arrival of new content, or (b) due to passage of time. Figure 5.12 shows a sample trajectory of the influence process for a given content. The user's visits do not affect the process $V(t)$, but these visits result in the user being influenced by the contents of the timeline. For example, one measure of influence could be the average influence on the user over the visits to the timeline, i.e.,

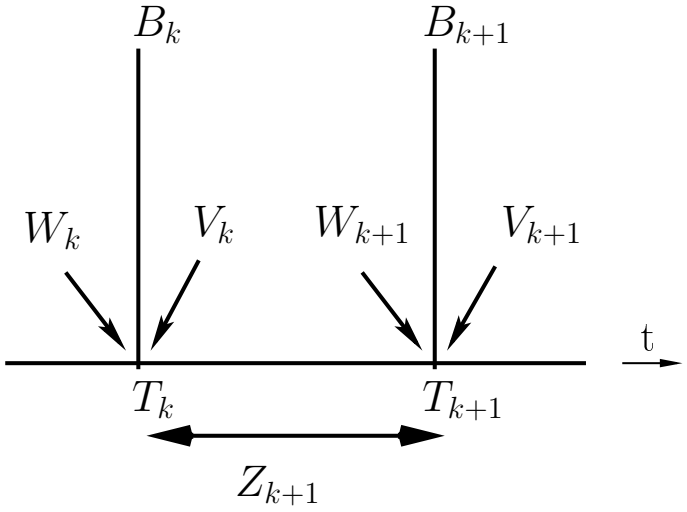


Figure 5.13: Timing of the embedded Markov chains V_k and W_k

Note that $V(t)$ is a piecewise deterministic Markov process, i.e., Markov processes with deterministic trajectories between random jumps. Such processes have been studied in [108], [109]. If $V(t)$ can be shown to be an asymptotically stationary and ergodic process with stationary probability distribution, then, using PASTA (Poisson arrivals see time averages), the above limit can be obtained almost surely as the expectation with respect to the stationary distribution.

5.2.5 Analysis of the $V(t)$ Process

Due to the assumption that new items arrive in a Poisson process and bring i.i.d. influence, it is clear that $V(t)$ is a continuous time Markov process, taking nonnegative values, and with right continuous sample paths with left hand limits. Define discrete parameter embedded processes as follows, for $k \geq 1$,

$$\begin{aligned} V_k &= V(T_k) \\ W_k &= V(T_k-) \end{aligned}$$

i.e., V_k is the value of the influence *after* the influence of the arrival at T_k is taken into account, whereas W_k is the value of the influence *found* by the arrival at T_k (see Figure 5.13). The following dynamics are evident:

$$V_{k+1} = \alpha(V_k - Z_{k+1})^+ + B_{k+1} \tag{5.21}$$

$$W_{k+1} = (\alpha W_k + B_k - Z_{k+1})^+ \tag{5.22}$$

Theorem 5.2.4. For $0 \leq \alpha < 1$, the following hold for the process $V(t)$:

1. There exists a probability distribution $F(\cdot)$ on $[0, \frac{b_{\max}}{1-\alpha})$, such that, $\forall v \in [0, \frac{b_{\max}}{1-\alpha})$, almost surely,

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t I_{\{V(u)\leq v\}} \mathrm{d}u = F(v)$$

2. Almost surely,

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t V(u) \mathrm{d}u = \int_0^\infty (1 - F(u)) \mathrm{d}u$$

Proof: The processes V_k and W_k are embedded at the epochs of the Poisson arrival process of rate ν. The process $V(t)$ is obtained from these embedded Markov processes by appropriately "filling" in the time between the Poisson epochs by straight line segments of slope -1. Note that this *temporal decay of influence* is used to model the time-critical nature of the content. We use regenerative analysis to carry out the proof. Let $T_{k_j}, j \geq 1$, denote the epochs $\{T_k\}$ such that $W_{T_{k_j}} = 0$. Due to the Poisson arrival model for the arrival of new items and the i.i.d. model for the influence they carry, it is clear that $T_{k_j}, j \geq 1$, are regeneration epochs for the random process $V(t), t \geq 0$. By Lemma 1 (Appendix) the embedded Markov chain is ϕ-irreducible and positive Harris recurrent. Since $\phi(\{0\}) > 0$, positive Harris recurrence implies that the mean time of return to $\{0\}$ in the Markov chain $\{W_k\}$ is finite. Thus, $V(t)$ is a nonnegative regenerative process with finite mean regeneration time. Both conclusions of the theorem follow from this observation (see [110]). □*Remark:* It can be shown in a similar manner as in Lemma 5.4.1 (Appendix 5.4.1) that the embedded Markov chain $V_k, k \geq 1$, is also positive Harris recurrent. Let $\Pi(\cdot)$ denote the invariant probability distribution of $\{V_k\}$. Since this embedded Markov chain is embedded at the epochs of a Poisson process of rate ν, and between the Poisson epochs $V(t)$ decreases linearly at rate 1, from Markov regenerative theory (see [111]) it can be seen that

$$F(v) = \frac{\int_v^\infty \int_{(x-v)}^\infty (u - (x - v)) \nu e^{-\nu u} \mathrm{d}u \; \mathrm{d}\Pi(x)}{\frac{1}{\nu}}$$

This can also be viewed as an alternative proof for the first part of Theorem 5.2.4. However, due to the difficulty in obtaining an expression for $\Pi(\cdot)$, this expression is not directly useful. Instead, we will resort to a level crossing argument [112] to characterise $F(\cdot)$.

Having established the existence of the steady-state probability distribution $F(\cdot)$, let p_0 be the point mass at 0 for the average influence process. Let $f(x)$ be the density of the distribution $F(\cdot)$, with $\int_0^\infty f(u)\mathrm{d}u = 1 - p_0$. Now we will use the level crossing approach [112], equating the total long-run average rates of downcrossings and upcrossings of a level x, which yields a renewal-type integral equation as follows:

$$f(x) + \nu \int_x^{\frac{x}{\alpha}} B(x - \alpha y) f(y)\mathrm{d}y = \nu p_0 B^c(x) + \nu \int_0^x F_B^c(x - \alpha y) f(y)\mathrm{d}y \tag{5.23}$$

In the above equation, the downcrossing and upcrossing rates of level x are listed on the LHS and RHS respectively. The terms on the LHS, in order, represent the downcrossings of level x due to:

1. unit rate of decay with time
2. decrease caused by arrival of new content (this occurs in case the incoming influence does not compensate enough for the discount due to α)

The terms on the RHS represent upcrossings of level x due to arrival of new content which see the influence process at level 0 (with probability p_0) or at level y (and the incoming weight is sufficiently large enough to cause an upcrossing of x). Define $\tilde{f}(s) = \int_0^\infty e^{-sx} f(x)\mathrm{d}x$. Similarly, $\tilde{b}(s) = \int_0^\infty e^{-sx} b(x)\mathrm{d}x$, where $b(x)$ is the probability density function corresponding to $B(\cdot)$. We assume, for now, that b does not have a point mass at 0 (Later we will extend to multiple content case, by relaxing this assumption). On rearranging terms in Equation (5.23), and taking the Laplace transform, we get

$$\tilde{f}(s)(s - \nu) + \nu \tilde{b}(s)\tilde{f}(\alpha s) = \nu p_0 (1 - \tilde{b}(s)) \tag{5.24}$$

Equation (5.24) can be used to get the moments of the $V(t)$ process. By differentiating Equation (5.24), writing out the Taylor expansion at $s = 0$ and setting $s = 0$, we get

$$p_0 = 1 - \nu EB + \nu(1 - \alpha)EV \tag{5.25}$$

where EV is the expected value of the $V(t)$ process. Similarly, on differentiating Equation (5.24) twice, writing out the Taylor expansion at $s = 0$ and setting $s = 0$ and using p_0 from Equation (5.25), we get

$$EV = \frac{\nu EB^2 - \nu(1 - \alpha^2)EV^2}{2(1 - \alpha\nu EB)} \tag{5.26}$$

Observe that, when $\alpha = 1$ in the expressions for p_0 and EV, they become equal to the probability of finding the queue empty and the expected work-in-system of an $M/G/1$ queue. This equivalence provides an interesting interpretation of the average influence process, which can be thought of as the work-in-system of an $M/G/1$ queue with *immediate discount at an arrival*, i.e., the work in system is discounted by a factor α on every new arrival. It is interesting to note, however, that whereas an M/G/1 queue with arrival rate ν and service requirement distribution $B(\cdot)$ is stable only if $\nu EB < 1$, the influence process that we have developed has a steady state for all ν if $\alpha < 1$.

From Equation (5.25), it can also be interpreted that each arrival (occuring at rate ν) brings in an average influence of $EB - (1-\alpha)EV$, i.e., $EV \leq \frac{EB}{1-\alpha}$. According to this interpretation, the average arriving influence depends on the existing influence, and thus it is difficult to obtain closed form expressions for EV as in the case of $M/G/1$ queues. It can be seen from Equation (5.26) that computation of EV requires EV^2. Thus, we will resort to other means (successive approximation) to numerically obtain EV (see Section 5.2.7).

5.2.6 Analysis of $V(t)$ with Multiple Content Creators

In this section, we will extend the level crossing analysis for the multiple content case. Consider $\mathcal{C}$, the set of content creators each with rates ν_c. Define $\nu := \sum_{c\in\mathcal{C}} \nu_c$ and $\nu_{-c} = \nu - \nu_c$. Let $V_c(t)$ process denote the average influence process, for content c. Note that whenever a content $c' \neq c$ arrives, the net influence of content c is scaled down by α. From the perspective of content c, this is equivalent to an arrival with 0 influence weight. Thus, for each content, we can reduce the analysis to the single content case, by introducing point mass in the arrival influence distribution, and normalizing accordingly. If $b_c(x)$ is the probability density function of arrival influence of content c, then the modified distribution would be $\frac{\nu_{-c}}{\nu}\delta(x) + \frac{\nu_c}{\nu} b_c(x)$, where $\delta(x)$ indicates point mass at 0. We can then write the level crossing expressions as follows:

$$\begin{aligned} & f_c(x) + \nu_{-c} \int_x^{\frac{x}{\alpha}} f_c(y)\mathrm{d}y + \nu_c \int_x^{\frac{x}{\alpha}} B_c(x-\alpha y) f(y)\mathrm{d}y \\ = \;& \nu_c f_{c,0} B_c^c(x) + \nu_c \int_0^x B_c^c(x-\alpha y) f_c(y)\mathrm{d}y \end{aligned} \tag{5.27}$$

As earlier, in the above equation, the downcrossing and upcrossing rates of level x are listed on the LHS and RHS respectively. The terms on the LHS, in order, represent the downcrossings of level x by the $V_c(t)$ process due to:

1. unit rate of decay with time
2. decrease caused by arrival of content $\hat{c} \in \mathcal{C}$, $\hat{c} \neq c$
3. decrease caused by arrival of content c (this occurs in case the incoming influence does not compensate enough for the discount due to α)

The terms on the RHS represent upcrossings of level x due to arrival of new content which see the influence process at level 0 (with probability p_0) or at level y (and the incoming weight is sufficiently large enough to cause an upcrossing of x). Proceeding with an analysis similar to Section 5.2.5, we get

$$\tilde{f}_c(s)(s-\nu) + (\nu_c \tilde{b}_c(s) + \nu_{-c})\tilde{f}(\alpha s) = \nu_c f_{c,0}(1 - \tilde{b}_c(s)) \tag{5.28}$$

$$f_{c,0} = 1 - \nu_c EB_c + \nu(1-\alpha)EV_c \tag{5.29}$$

$$EV_c = \frac{\nu_c EB_c^2 - \nu(1-\alpha^2)EV_c^2}{2(1-\alpha\nu_c EB_c)} \tag{5.30}$$

Note that Equation (5.28) can be obtained from Equation (5.24) by replacing $\tilde{b}(s)$ with $\frac{\nu_{-c}}{\nu} + \frac{\nu_c}{\nu}\tilde{b}_c(s)$ (which is the Laplace transform of $\frac{\nu_{-c}}{\nu}\delta(x) + \frac{\nu_c}{\nu}b_c(x)$). Also similar to p_0, $f_{c,0}$ can be interpreted in terms of the average load brought in by an arrival. Rearranging terms we can write,

$$f_{c,0} = 1 - \nu\Big(\frac{\nu_c}{\nu}EB_c - (1-\alpha)EV_c\Big)$$

We see that content arrives at rate ν (including competing content), and on an average bring in an influence of $\frac{\nu_c}{\nu}EB_c - (1-\alpha)EV_c$. Similar to the single content case, this expression does not yield an explicit formula for EV_c and we have to resort to successive approximation based techniques to numerically solve the integral equation.

5.2.7 Numerical Results

In this section, we will use successive approximation to compute the expected influence for a single content case. We will then proceed to study various game formulations for N-player and 2-player settings. For simplicity, we carry out the numerical analysis assuming the deterministic influence distribution, i.e.,

$$\begin{aligned} B_c(x) &= 0, \text{if } x < b_c \\ &\quad 1, \text{if } x \geq b_c \end{aligned} \tag{5.31}$$

It is then straightforward to see that F is supported in the interval $[0, \frac{b_c}{1-\alpha}]$. We also have $EB = b_c$ and $EB^2 = b_c^2$. Note also that in this case, since the all arriving contents carry the same influence, *we will never have downward jumps in the $V(t)$ process for a single content case* (see Figure 5.12), i.e., the second term on the LHS of Equation (5.23) will be absent.

We compute the expected influence numerically by using successive approximation on the integral equation obtained from level crossing analysis, (Equation (5.23) or (5.27)). We will demonstrate it for the single content case. We begin with $p_0^{(0)} = \max(0, 1-\nu EB)$, $f^{(0)} = (1 - p_0^{(0)})\mathcal{U}[0, \frac{b}{1-\alpha}]$ where $\mathcal{U}[x_1, x_2]$ is the uniform distribution over the interval $[x_1, x_2]$. We then use Equation (5.26) to obtain $EV^{(1)}$ and subsequently use Equation (5.25) to obtain $p_0^{(1)}$. We can then obtain $f^{(1)}$ by using $(p_0^{(1)}, f^{(0)})$ in Equation (5.23). The process is then repeated, obtaining $(p_0^{(k+1)}, f^{(k+1)})$ from $(p_0^{(k)}, f^{(k)})$ via $EV^{(k+1)}$. We iterate until the total variation distance between successive iterations of (p_0, f) is sufficiently small. Figures 5.14 – 5.16 demonstrate that the EV values obtained by numerically solving the integral equation match well with the results of $V(t)$ simulation.

Computation of Expected Influence Figure 5.14 shows the variation of EV with the discount factor α. We recall that α^k is the probability that the user scans at least up to the kth item in the timeline. Observe from Figur 5.14 that for small values of α the expected influence is below $b_c = 0.8$. This is because, for small α, when the user samples his timeline, with a high probability, he just looks at the first item; also, if the content arrival rate is small, then the item at the top of the timeline is "dated" content, and, hence, contributes low influence. We further observe that as α increases, the number of entries seen by the user increases, thus increasing the expected influence of the content. Figures 5.15 and 5.16 show the variation of EV with the influence weight b and the rate of content generation ν. It is interesting to observe numerically that, while EV exhibits diminishing returns for increasing ν, it exhibits increasing returns for small ranges of b and for large ranges of b it is linear. For large ranges of b and fixed ν, the $V(t)$ process never hits 0 (i.e., $p_0 = 0$), and hence any investment into b reflects linearly in EV. These insights will be useful while analyzing the game formulations to obtain the optimal rate and optimal influence weight. Also, for high ν, EV approaches $\frac{EB}{1-\alpha}$, with p_0 approaching 0 (Equation 5.25).

Competing Contents In this section, we will study various competition scenarios that may arise in a system with multiple content creators. The utility derived by each user

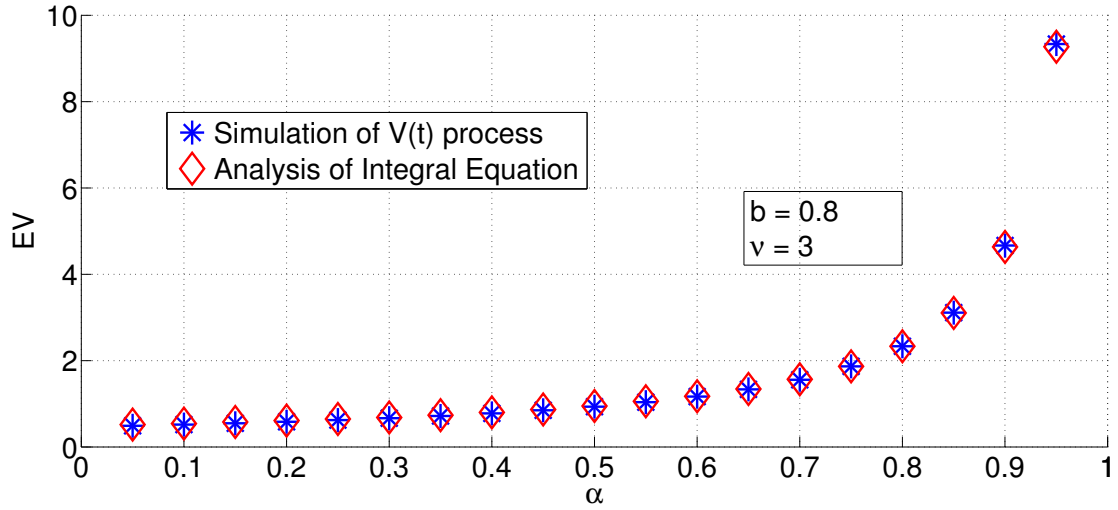


Figure 5.14: Variation of the expected influence EV with the discount factor, α.

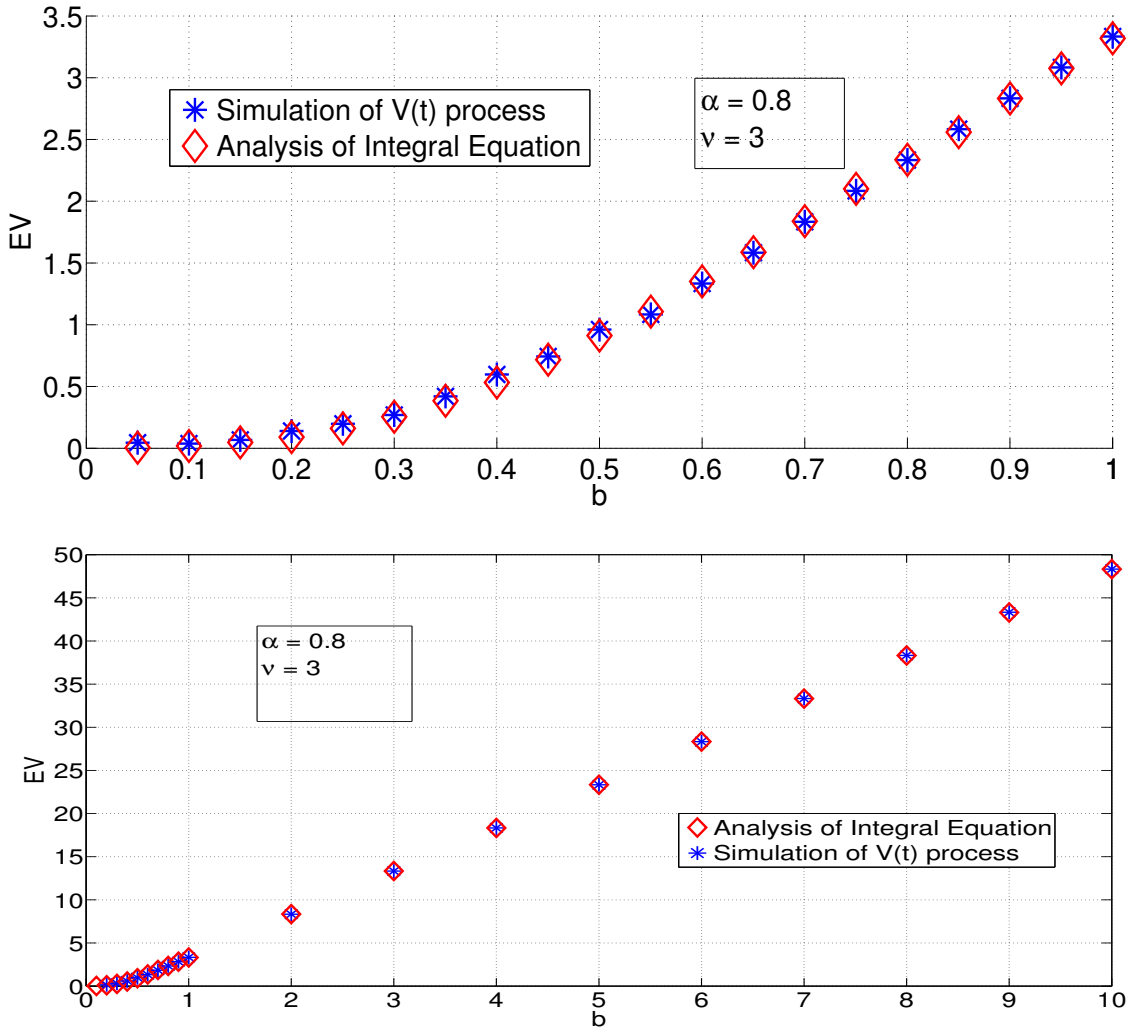


Figure 5.15: Variation of the expected influence EV with the influence weight for small and large ranges of b. One can see that For small values of b, EV yields increasing returns (convex), and for larger values, EV is linear with b

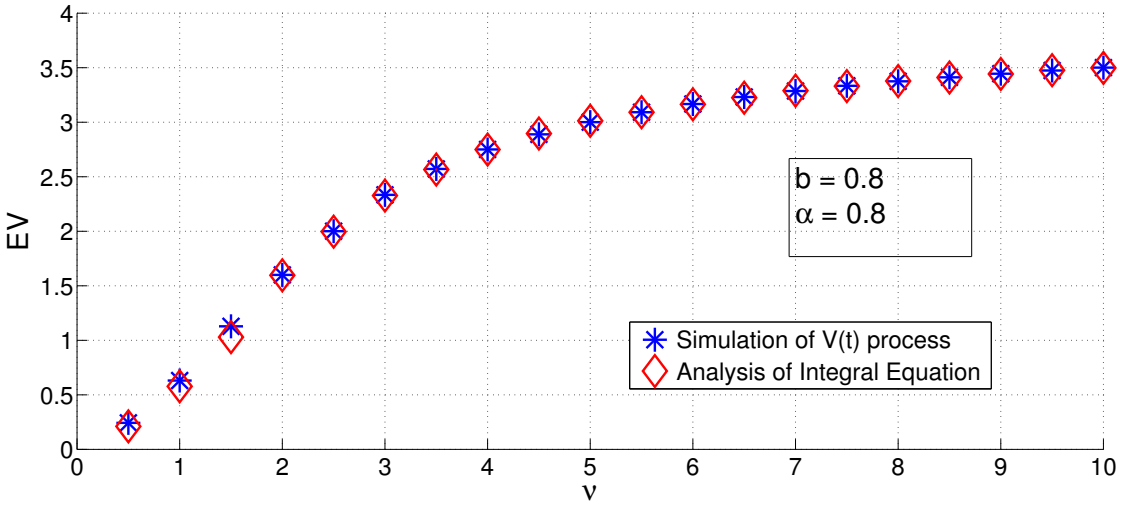


Figure 5.16: Variation of the expected influence EV with the content generation rate, ν.

is the expected influence for his content on the timeline minus the cost incurred due to operating at (ν, b). We first study an N-player game with symmetric costs and numerically obtain the optimal rate (with influence weight fixed), and optimal influence weight (with rate fixed). We then proceed to study a two-player formulation, and study the effect of cost parameter η and maximum influence weight b.

Symmetric N-player game Consider the system with N players (content creators) with $(\nu_i, b_i), i \in \mathcal{C}$ as the rates and influence weights. Let EV_i be the expected influence derived by player i, and let the utility of player i be given as:

$$U_i = EV_i - \eta_i \nu_i b_i \tag{5.32}$$

where η_i is the cost parameter. We also note that the cost is dependent on ν_i and b_i via $\nu_i b_i$. This is motivated by the assumption that, given a fixed budget of operation, the rate of generation of content (quantity) is inversely proportional to the influence generated by the content (quality).

Consider an N-player game with symmetric costs, i.e., $\eta_i = \eta$, $\forall i \in \mathcal{C}$. We can then look at the optimal rate of operation, given all players generate the same amount of influence. Figure 5.17 shows the variation of ν^{opt} the optimal rate of content generation with increasing number of competitors. This is to be contrasted with the earlier result with fixed timeline size, and the probability of finding the content on the timeline as the objective, the optimal rate exhibited a non-monotonic behavior with increasing number of players N.

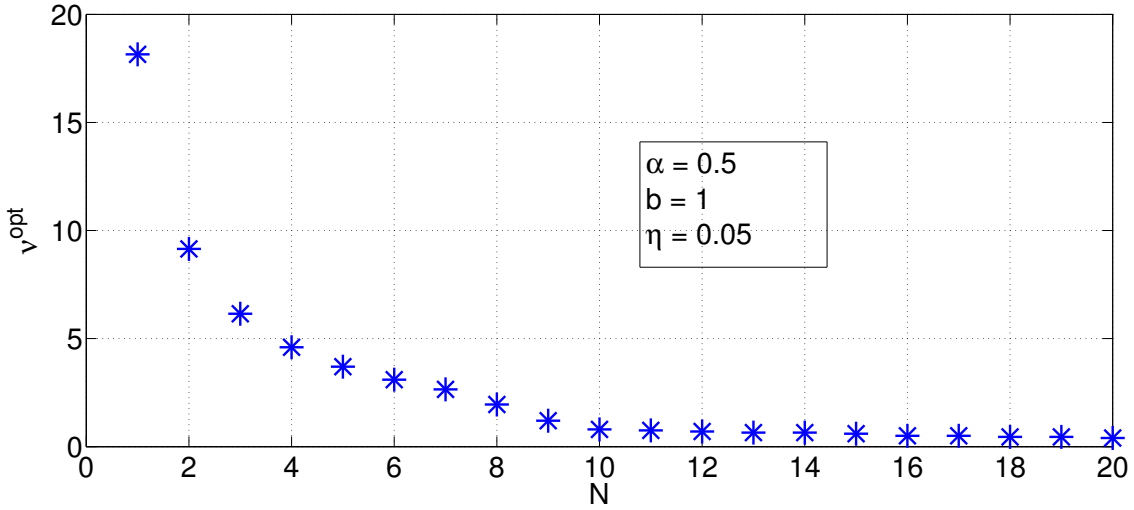


Figure 5.17: Variation of optimal rate of content generation in a symmetric game among N content providers (with symmetric influence b and costs η), with increasing N.

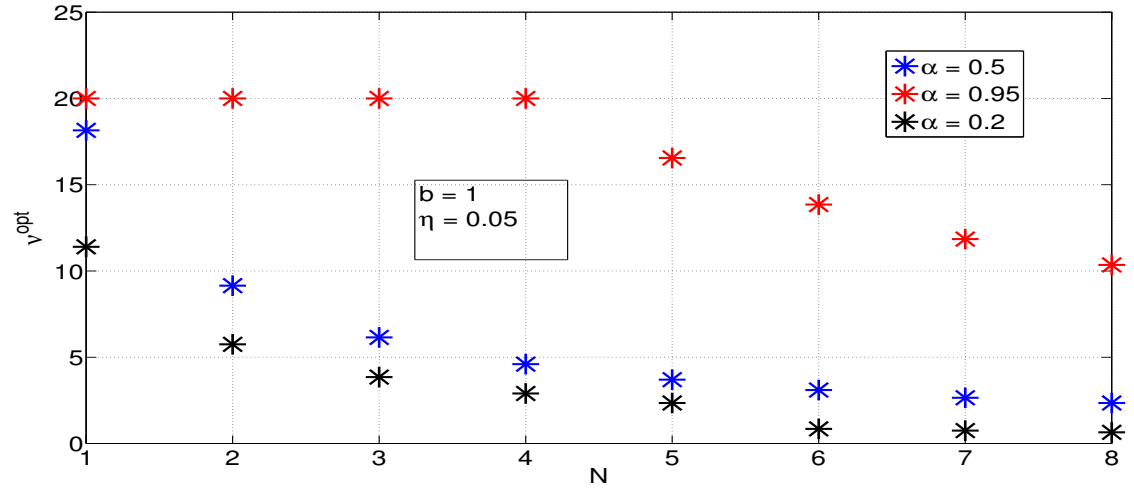


Figure 5.18: Variation of optimal rate with number of players for different values of α. In this example, $\nu \in [0.1, 20]$

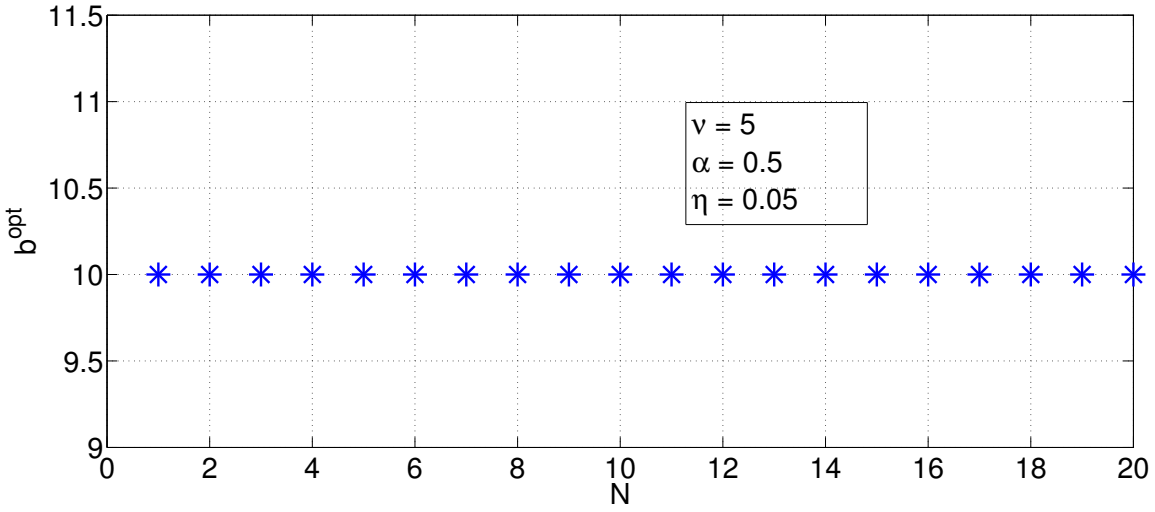


Figure 5.19: Variation of optimal influence weight in a symmetric game among N content providers (with symmetric content generation rate ν and symmetric costs η), with increasing N. The maximum influence weight in this example is, $b_{\max} = 10$.

Figure 5.18 shows the variation of optimal rate with number of players for different values of α. For large α, the user sees more entries in the timeline, hence increasing ν ensures that the content occupies more entries in the timeline (leading to higher expected influence).

We then consider a N-player symmetric game of optimal influence weight, under the symmetric cost setting, given all players generate content at the same rate $\nu_i = \nu$, $\forall i \in \mathcal{C}$. Recall that we are working with the deterministic influence setup. For small values of b_c, EV_c yields increasing returns, and larger values EV_c is linear with b_c (see Figure 5.15). Thus the slope $\frac{\partial EV_c}{\partial b}$ starts at 0 and increases to the value at the linear portion of the curve. Let the maximum influence weight that each player's content can generate be $b_{\max}$. In the N-player game, we could have one of the following cases:

- $\frac{\partial EV_c}{\partial b} < \eta\nu$ for all $b \in [b_{\min}, b_{\max}]$. This implies U_c will always be less than zero. Thus optimal point of operation is $b = b_{\min}$ (still yielding a negative utility). Thus the case is not individually rational, i.e., the content creator is better off not participating in the competition.

- $\frac{\partial EV_c}{\partial b} > \eta\nu$ for some $b \in [b_{\min}, b_{\max}]$. In this case, it is clear that operating at $b_{\max}$ is optimal, provided it results in positive utility.

Thus, assuming $b_{\max}$ provides positive utility, observe from Figure 5.19, the increase in the number of players has no effect on b. Thus, a linear cost to quality (influence weight), causes the content creators to operate at the maximum possible influence weight, $b_{\max}$.

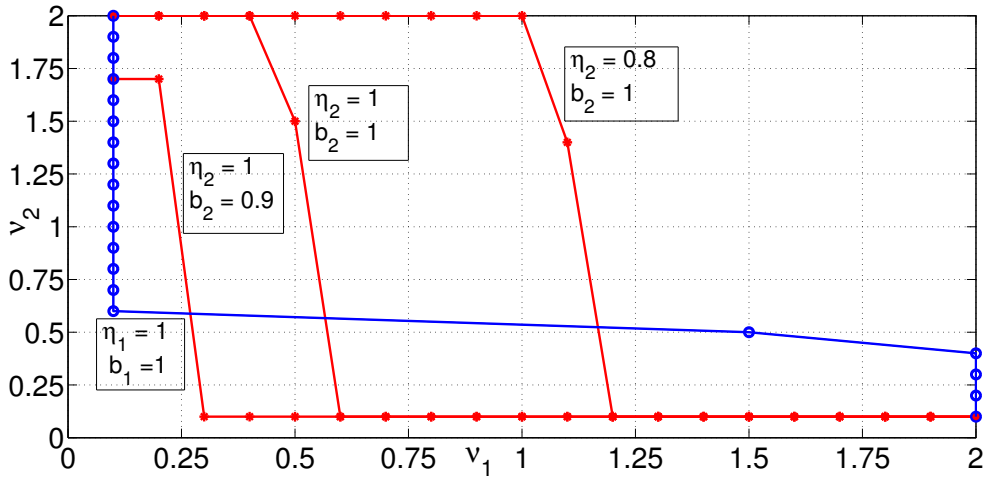


Figure 5.20: This shows the best response curves of player 1 (blue) and player 2 (red) for various values of (η_2, b_2) and $(\eta_1 = 1, b = 1)$. In this simulation, $\nu_{\min} = 0.1$ and $\nu_{\max} = 2$.

Two-player rate game In this section we will study a two-player game formulation with (ν_1, b_1) and (ν_2, b_2) as the strategies. From the previous section, it can be seen that given a fixed budget, it is always optimal to operate at the maximum possible influence weight $b_{\max}$. Hence in the two-player formulation, we will assume that the players fix their b_i and optimize only over ν_i. This can also be motivated by the fact that, in real world systems, the influence generated b is usually a function of established reputation, which cannot be changed as easily as changing the content generation rate. Given the utility functions (Equation (5.32)), we can plot the best response curves for both players 1 and 2, and the Nash equilibrium can be obtained as the intersection of the best response curves. From Figure 5.20 it can be seen that we can get three different equilibria, i.e.,

1. Player 1 operates close to ν_{max} and Player 2 close to ν_{min}
2. Player 2 operates close to ν_{max} and Player 2 close to ν_{min}
3. Non-trivial equilibrium with $\nu_1^* \in (\nu_{min}, \nu_{max})$ and $\nu_2^* \in (\nu_{min}, \nu_{max})$

Firstly it is obvious that, when $\eta_1 = \eta_2$ and $b_1 = b_2$, the non-trivial equilibrium is symmetric. Note that a decrease in η_2 allows player 2 to use a higher rate ν_2 as best response to a given ν_1. This is because, for lower η_2, player 2 faces less cost, and hence can be more aggressive. This shifts the player 2's best response curve to the right, thus in the non-trivial equilibrium, causing player 1 to operate at a higher equilibrium rate. This can be understood as follows: *the presence of an opponent with low cost causes the player to increase his equilibrium rate.*

Another implication from Figure 5.20 is that, a decrease in b_2 causes player 2's best response curve to shift to the left (restricting him to using lower rates as best response

to a given ν_1). This, then results in a non-trivial equilibrium where player 1 can operate at a lower equilibrium rate. This can be summarized as follows: *the presence of an opponent with lower influence weight (quality) allows the player to reduce his equilibrium rate (quantity).*

5.3 Discussion

In this chapter, we studied competition between contents in two different contexts: (a) when users spend time across multiple social networks, (b) competition for visibility in a single user's timeline. While, the first problem involved spread of influence, the latter model focused entirely on how a single user processes the contents reaching him. In the competition across multiple social networks, we studied the equilibrium budget allocations of competing players in each of the social networks, and observed that the best response functions exhibited a threshold behavior. We also obtained the Nash equilibria as intersections of the best response function and observed the effect of cost parameters on the equilibrium budget allocation.

We then considered models of a single social network user's timeline, of both the finite and infinite size. In the finite timeline case, we derived explicit equilibrium rates for a symmetric game among content creators, and numerically observed the non-monotonic behavior with respect to N, the number of content creators. For the infinite timeline case, we obtained integral equations capturing the evolution of average influence of multiple contents on the timeline, and showed that it behaves similar to an $M/G/1$ queue with *immediate discount*. We used it to study competition among content creators, and observed the impact of competing content's quality (influence weight) and quantity (rate of content generation) on the equilibrium strategies of a given content creator.

5.4 Appendix

5.4.1 Analysis of $V(t)$

Lemma 5.4.1. With $0 \leq \alpha < 1$, and if the support of $B(\cdot)$ is $[0, b_{\max}]$, then

1. $W_k, k \geq 1$, is a ψ-irreducible Markov process on $[0, \frac{b_{\max}}{1-\alpha})$.

2. $W_k, k \geq 1$, is positive Harris recurrent.

Proof: *Outline:* We first observe that if the influence brought in by a content is $\leq b_{\max}$ then $V(t) < \frac{b_{\max}}{1-\alpha}$, $\forall t$; hence, $W_k \in [0, \frac{b_{\max}}{1-\alpha})$. The proof of Part 1 utilizes the fact that the set $\{0\}$ is reachable from any $w \in [0, \frac{b_{\max}}{1-\alpha})$, and any Borel set in $[0, \frac{b_{\max}}{1-\alpha})$ is similarly reachable from the state 0. The proof of Part 2 follows by using a drift argument to show that the process $\{W_k\}$ returns to a petite set $[0, c]$ infinitely often with probability 1.

The dynamics in Equation 5.22 indicates that the irreducibility measure will give positive mass to the atom at $\{0\}$, and to every Borel set in $[0, \frac{b_{\max}}{1-\alpha})$ with positive Lebesgue measure.

Let $Q^{(n)}(w, A)$ denote the n-step transition probability measure for $\{W_k\}$; when $n = 1$, the superscript is suppressed. We follow the discussion in [113, Section 4.3.1]. From the dynamics shown in Equation 5.22, it follows that for $x > 0$, $Q(x, \{0\}) > 0$. Indeed,

$$Q(x, \{0\}) > \int_0^\infty e^{-\nu(\alpha x + y)} \mathrm{d}B(y)$$

We now show that $Q(0, A) > 0$ for all Borel sets A in $[0, \frac{b_{\max}}{1-\alpha})$ which are of positive Lebesgue measure. Such a set must have an intersection of positive Lebesgue measure with an interval $[0, (\frac{b_{\max}}{1-\alpha}) - \epsilon)$ for some $\epsilon > 0$. We can obtain an n_ϵ such that

$$b_{\max} \frac{1 - \alpha^{n_\epsilon}}{1 - \alpha} > \frac{b_{\max}}{1 - \alpha} - \frac{\epsilon}{2}$$

Thus, if n_ϵ contents arrive over an interval of length $\frac{\epsilon}{2}$ after the $\{W_k\}$ process enters $\{0\}$, then the resulting value of the process will exceed $\frac{b_{\max}}{1-\alpha} - \epsilon$. It follows that

$$Q^{(n_\epsilon)}\left(0, \left(\frac{b_{\max}}{1-\alpha} - \frac{\epsilon}{2}, \frac{b_{\max}}{1-\alpha}\right)\right) > \frac{(\nu\epsilon)^{n_\epsilon}}{n_\epsilon!} e^{-\nu\epsilon}$$

Now having crossed above $\frac{b_{\max}}{1-\alpha} - \epsilon$ (in n_ϵ steps), the reduction in value due to the next interarrival time can be utilized to complete the argument (see Equation 5.22). Since the exponential interarrival time distribution gives positive support to all Borel sets on $[0, \frac{b_{\max}}{1-\alpha})$, the original Borel set A (which has an intersection of positive Lebesgue measure with $[0, (\frac{b_{\max}}{1-\alpha}) - \epsilon)$) will be entered with positive probability.

We will employ Theorem 11.3.4 in [113]. To this end, we see that

$$\begin{aligned} & E\left((W_{k+1} - W_k) | W_k = w\right) \\ = \;& E\left(((\alpha W_k + B_k - Z_{k+1})^+ - W_k) | W_k = w\right) \end{aligned}$$

$$\begin{aligned} &\leq E((\alpha\frac{b_{\max}}{1-\alpha}+b_{\max}-Z_{k+1})^+-w) \\ &= E((\frac{b_{\max}}{1-\alpha}-Z_{k+1})^+-w) \\ &< -\delta \quad \text{for} \quad w>c \end{aligned}$$

for a suitably chosen c. Further,

$$E\left((W_{k+1}-W_k)|W_k=w\right) \quad \leq \quad EB_1 \quad \text{for all} \quad w\geq 0$$

It follows that

$$E\left((W_{k+1}-W_k)|W_k=w\right) \leq -\delta+(EB_1+\delta)I_{\{[0,c]\}}(w)$$

We further show that the set $[0,c]$ is petite. It suffices to show that there exists a positive integer m such that for all Borel sets $A\in[0,\frac{b_{\max}}{1-\alpha})$ it holds that $Q^{(m)}(x,A)\geq\beta(A)$ for all $x\in[0,c]$, where β is a non-trivial measure. We notice that

$$\begin{aligned} Q^{(2)}(x,A) &\geq Q(x,\{0\})Q(0,A) \\ &= \left(\int_0^\infty B(z-\alpha x)\nu e^{-\nu z}\mathrm{d}z\right)Q(0,A) \\ &\geq \left(\int_0^\infty B(z-\alpha c)\nu e^{-\nu z}\mathrm{d}z\right)Q(0,A) \end{aligned}$$

Writing $\delta_c:=\int_0^\infty B(z-\alpha c)\nu e^{-\nu z}\mathrm{d}z$, we conclude that

$$Q^{(2)}(x,A) \quad \geq \quad \delta_c Q(0,A)$$

Since we already know from the first part of the proof that $Q(0,A)$ is positive for all Borel sets in $[0,\frac{b_{\max}}{1-\alpha})$, we take the right hand side of the previous equality as $\beta(A)$. It then follows from Theorem 11.3.4 in [113] that $\{W_k\}$ is positive Harris recurrent. □

CHAPTER 6

Conclusion

In this thesis, we have provided an analytical study of various models of influence dynamics on social networks. We have addressed problems ranging from influence maximization, mobile content delivery, content competition on social networks, etc. We briefly summarize the key results.

Influence maximization: We analytically characterized the influence spread under the Linear Threshold (LT) model [1], and provided an equivalent interpretation in terms of acyclic path probabilities in a Markov chain. We provided explicit characterization of optimal seed set for influence maximization for some toy networks, such as ring, star and node-degree based models. We also studied the performance of Pagerank [2] under various scenarios, and provided insights into its performance comparison with the greedy algorithm. Finally, using the insights from the analytical characterization, we propose the computationally efficient *G1-sieving* algorithm for the influence maximization problem, and show it performs on par with the greedy algorithm.

Fluid limit of the LT model: We derived fluid limits of a homogeneous variant of the LT model, under a general threshold distribution. We showed that the threshold distribution features in the fluid limit via its hazard function. We also showed that, under the exponential threshold distribution, the LT model becomes equivalent to the well known SIR model from epidemiology. We then considered the LT model under specific distributions, such as Weibull and loglogistic, and showed that qualitatively different influence dynamics can arise due to different threshold distributions, even under a homogeneous

setting. We also showed that our approach is amenable to a heterogeneous LT model with communities.

Coevolution of Influence and Content in MONETs: We modeled the evolution of popularity and the spread of content by models inspired from epidemiology. While the former is similar to an uncontrolled SIR (Susceptible-Infected-Recovered) process, the latter is modeled as a controlled SI (Susceptible-Infected) process, thus yielding what we call the *SIR-SI* model. The evolutions are modeled as coupled continuous time Markov chains, and using [88] we derived the fluid limit of the joint evolution process. We then obtained a time-threshold policy while copying to relays, that is delay-cost optimal for the problem of mobile content delivery.

Coevolution over Multiple Social Networks: We modeled the competition between two content providers across multiple social networks as a non-cooperative game. We modeled the influence dynamics using epidemic models, and formulated the non-cooperative game in terms of the content creators' budgets. We obtained the best response functions for each of the users, thus characterizing the Nash equilibria, as the intersection of the best response functions. We also provided interesting remarks on the threshold structure of the best response function. We finally discussed some variants of the cost structure and discussed its similarities to competition models such as the Cournot duopoly model [89].

Competition over Timelines in Social Networks: We studied competition between content creators for visibility in a single social network user's timeline. We began with a timeline of fixed size, and formulated a non-cooperative game among N content creators, whose utility is the probability of finding their content on the timeline. We derived explicit equilibrium rate expressions under a symmetric cost scenario, and showed that it exhibits non-monotonic behavior, with increase in number of players N. We then developed a model for influence evolution on a timeline of infinite size, with a behavioral model for user scanning the timeline. We also modeled the quality variability among content creators via the influence weight of a content. We obtained integral equations, capturing the evolution of influence for a single content creator in isolation, and in the presence of competing contents. We showed that the system behaves like an $M/G/1$ queue *with immediate discount*. We then finally studied non-cooperative games among the content creators, and provided some interesting insights into competitions involving quality and quantity.

List of Publications

Journals

- Srinivasan Venkatramanan and Anurag Kumar, *"Spread of Influence and Content in Mobile Opportunistic Networks,"* to appear in IEEE Transactions on Mobile Computing.

- Srinivasan Venkatramanan and Anurag Kumar, *"Fluid Limit Characterization of the Linear Threshold Model for Influence Spread on Social Networks,"* in preparation, 2014

Conference/Workshop Publications

- Srinivasan Venkatramanan and Anurag Kumar, *"Competition for Content Spread over Multiple Social Networks,"* Workshop on Social Networks in Science and Engineering (SCINSE'14), co-held with 6th Intl. Conference on Communication Systems and Networks (COMSNETS), Bangalore, India, 2014. (Recipient of the Best presentation award, sponsored by Adobe Research)

- Eitan Altman, Parmod Kumar, Srinivasan Venkatramanan and Anurag Kumar, *"Competition over Timeline in Social Networks,"* Workshop on Social Network Analysis and Algorithms (SNAA), co-held with IEEE/ACM ASONAM, Niagara Falls, Canada, 2013.

- Srinivasan Venkatramanan and Anurag Kumar, *"Co-evolution of Content Popularity and Delivery in Mobile P2P Networks,"* IEEE Infocom 2012 mini-Conference, Orlando, FL, 2012.

- Srinivasan Venkatramanan and Anurag Kumar, *"Information Dissemination in Socially Aware Networks under the Linear Threshold model,"* National Conference on Communication(NCC), IISc. Bangalore, 2011.

Technical Reports

- Srinivasan Venkatramanan and Anurag Kumar, *"Co-evolution of Content Popularity and Delivery in Mobile P2P Networks,"* arXiv:1107.5851

- Srinivasan Venkatramanan and Anurag Kumar, *"New Insights from an Analysis of Social Influence Networks under the Linear Threshold model,"* arXiv:1002.1335

Bibliography

[1] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *ACM SIGKDD*, 2003.

[2] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." 1999.

[3] N. T. Bailey, *The mathematical theory of infectious diseases and its applications*. Charles Griffin, 1975.

[4] M. Granovetter, "The strength of weak ties," *American journal of sociology*, vol. 78, no. 6, p. l, 1973.

[5] S. Eubank, H. Guclu, V. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, "Modelling disease outbreaks in realistic urban social networks," *Nature*, vol. 429, no. 6988, pp. 180–184, 2004.

[6] E. M. Rogers, *Diffusion of innovations*. New York: Free Press, 1962.

[7] P. Domingos and M. Richardson, "Mining the network value of customers," in *ACM SIGKDD*, 2001.

[8] C. A. Sheingold, "Social networks and voting: the resurrection of a research agenda," *American Sociological Review*, pp. 712–720, 1973.

[9] L. Coviello, Y. Sohn, A. D. Kramer, C. Marlow, M. Franceschetti, N. A. Christakis, and J. H. Fowler, "Detecting emotional contagion in massive social networks," *PloS one*, vol. 9, no. 3, p. e90315, 2014.

[10] J. Travers and S. Milgram, "An experimental study of the small world problem," *Sociometry*, vol. 32, no. 4, pp. 425–443, 1969.

[11] P. Erdős and A. Rényi, "On the evolution of random graphs," *Magyar Tud. Akad. Mat. Kutató Int. Közl*, vol. 5, pp. 17–61, 1960.

[12] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-worldnetworks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[13] A.-L. Barabási, R. Albert, and H. Jeong, "Mean-field theory for scale-free random networks," *Physica A: Statistical Mechanics and its Applications*, vol. 272, no. 1, pp. 173–187, 1999.

[14] T. G. Lewis, *Network science: Theory and applications*. John Wiley & Sons, 2011.

[15] N. Ganguly, A. Deutsch, and A. Mukherjee, *Dynamics on and of complex networks: applications to biology, computer science, and the social sciences*. Springer, 2009.

[16] M. Granovetter, "Threshold models of collective behaviour," *American Journal of Sociology*, 1978.

[17] T. W. Valente, "Social network thresholds in the diffusion of innovations," *Social networks*, vol. 18, no. 1, pp. 69–89, 1996.

[18] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *ACM SIGKDD*, 2002.

[19] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, "Everyone's an influencer: quantifying influence on twitter," in *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 2011, pp. 65–74.

[20] L. Hong, O. Dan, and B. D. Davison, "Predicting popular messages in twitter," in *Proceedings of the 20th international conference companion on World wide web*. ACM, 2011, pp. 57–58.

[21] G. Szabo and B. A. Huberman, "Predicting the popularity of online content," *Communications of the ACM*, vol. 53, no. 8, pp. 80–88, 2010.

[22] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999.

[23] R. Crane, D. Sornette *et al.*, "Viral, quality, and junk videos on youtube: Separating content from noise in an information-rich environment." in *AAAI Spring Symposium: Social Information Processing*, 2008, pp. 18–20.

[24] J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-tracking and the dynamics of the news cycle," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 497–506.

[25] S. Muralidharan, L. Rasmussen, D. Patterson, and J.-H. Shin, "Hope for haiti: An analysis of facebook and twitter usage during the earthquake relief efforts," *Public Relations Review*, vol. 37, no. 2, pp. 175–177, 2011.

[26] R. Thomson, N. Ito, H. Suda, F. Lin, Y. Liu, R. Hayasaka, R. Isochi, and Z. Wang, "Trusting tweets: The fukushima disaster and information source credibility on twitter," in *Proceedings of the 9th International ISCRAM Conference*, 2012.

[27] G. Lotan, E. Graeff, M. Ananny, D. Gaffney, I. Pearce *et al.*, "The arab spring— the revolutions were tweeted: Information flows during the 2011 tunisian and egyptian revolutions," *International Journal of Communication*, vol. 5, p. 31, 2011.

[28] M. D. Conover, E. Ferrara, F. Menczer, and A. Flammini, "The digital evolution of occupy wall street," *PloS one*, vol. 8, no. 5, p. e64679, 2013.

[29] S. Paul-Choudhury, "Digital legacy: the fate of your online soul," *New Scientist*, vol. 210, no. 2809, pp. 41–43, 2011.

[30] K. Lerman and R. Ghosh, "Information contagion: An empirical study of the spread of news on digg and twitter social networks." *ICWSM*, vol. 10, pp. 90–97, 2010.

[31] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: A survey," *ACM SIGMOD Record*, vol. 42, no. 1, pp. 17–28, 2013.

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

[33] S. Bharathi, D. Kempe, and M. Salek, "Competitive influence maximization in social networks," in *Internet and Network Economics*. Springer, 2007, pp. 306–311.

[34] S. Kumar, R. Zafarani, and H. Liu, "Understanding user migration patterns in social media." 2011.

[35] L. Weng, A. Flammini, A. Vespignani, and F. Menczer, "Competition among memes in a world with limited attention," *Scientific Reports*, vol. 2, 2012.

[36] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," in *Proceedings of the Royal Society of London*, ser. A, Containing Papers of a Mathematical and Physical Character, vol. 115, no. 772. Royal Society, 1927, pp. 700–721.

[37] J. Satsuma, R. Willox, A. Ramani, B. Grammaticos, and A. Carstea, "Extending the sir epidemic model," *Physica A: Statistical Mechanics and its Applications*, vol. 336, no. 3, pp. 369–375, 2004.

[38] A. S. Klovdahl, "Social networks and the spread of infectious diseases: the aids example," *Social science & medicine*, vol. 21, no. 11, pp. 1203–1216, 1985.

[39] R. M. Anderson, R. M. May, and B. Anderson, *Infectious diseases of humans: dynamics and control*. Wiley Online Library, 1992, vol. 28.

[40] M. E. J. Newman, "Spread of epidemic disease on networks," *Physical Review*, 2002.

[41] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical review letters*, vol. 86, no. 14, p. 3200, 2001.

[42] L. Danon, A. P. Ford, T. House, C. P. Jewell, M. J. Keeling, G. O. Roberts, J. V. Ross, and M. C. Vernon, "Networks and the epidemiology of infectious disease," *Interdisciplinary perspectives on infectious diseases*, vol. 2011, 2011.

[43] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry, "Epidemic algorithms for replicated database maintenance," in *Proceedings of the sixth annual ACM Symposium on Principles of distributed computing*. ACM, 1987, pp. 1–12.

[44] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on*. IEEE, 2003, pp. 482–491.

[45] F. Chierichetti, S. Lattanzi, and A. Panconesi, "Rumor spreading in social networks," in *Automata, Languages and Programming*. Springer, 2009, pp. 375–386.

[46] ——, "Rumour spreading and graph conductance," in *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, 2010, pp. 1657–1663.

[47] D. J. Aldous, "Some inequalities for reversible markov chains," *Journal of the London Mathematical Society*, vol. 2, no. 3, pp. 564–576, 1982.

[48] M. Draief, A. Ganesh, and L. Massoulié, "Thresholds for virus spread on networks," in *Proceedings of the 1st international conference on Performance evaluation methodolgies and tools*. ACM, 2006, p. 51.

[49] C. Asavathiratham, "The influence model : A tractable representation of the dynamics of networked markov chains," Ph.D. dissertation, MIT, 2000.

[50] M. E. Newman, S. Forrest, and J. Balthrop, "Email networks and the spread of computer viruses," *Physical Review E*, vol. 66, no. 3, p. 035101, 2002.

[51] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," *ACM Transactions on the Web (TWEB)*, vol. 1, no. 1, p. 5, 2007.

[52] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. Vanbriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *ACM SIGKDD*, 2007.

[53] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 199–208.

[54] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. IEEE, 2010, pp. 88–97.

[55] R. Narayanam and Y. Narahari, "A shapley value-based approach to discover influential nodes in social networks," *IEEE Transactions on Automation Science and Engineering*, no. 99, pp. 1–18, 2010.

[56] E. Altman, F. D. Pellegrini, R. El-Azouzi, D. Miorandi, and T. Jimenez, "Emergence of equilibria from individual strategies in online content diffusion," in *Fifth International Workshop on Network Science for Communication Networks (NetSciCom)*, 2013.

[57] P. S. Dodds and D. J. Watts, "Universal behavior in a generalized model of contagion," *Physical review letters*, vol. 92, no. 21, p. 218701, 2004.

[58] D. J. Watts, "A simple model of global cascades on random networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 9, pp. 5766–5771, 2002.

[59] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless video content delivery through distributed caching helpers," in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 1107–1115.

[60] T. Repantis and V. Kalogeraki, "Data dissemination in mobile peer-to-peer networks," in *Proceedings of the 6th international conference on Mobile data management*. ACM, 2005, pp. 211–219.

[61] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and human mobility in conference environments," in *SIGCOMM'05 Workshop*.

[62] D. Kotz and T. Henderson, "A community resource for archiving wireless data at dartmouth," in *IEEE Pervasive Computing*, 2005.

[63] M. Musolesi and C. Mascolo, "A community mobility model for ad-hoc network research," in *ACM SIGMOBILE*, 2006.

[64] C. Boldrini, M. Conti, A. Passarella, and V. Moruzzi, "Impact of social mobility on routing protocols for opportunistic networks," 2008.

[65] N. Sastry, K. Sollins, and J. Crowcroft, "Delivery properties of human social networks," 2009.

[66] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: Social-based forwarding in delay tolerant networks," in *MobiHoc*, 2008.

[67] Q. Li, S. Zhu, and G. Cao, "Routing in socially selfish delay tolerant networks," in *InfoComm*, 2010.

[68] Ó. R. Helgason, E. A. Yavuz, S. T. Kouyoumdjieva, L. Pajevic, and G. Karlsson, "A mobile peer-to-peer system for opportunistic content-centric networking," in *Proceedings of the second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds*. ACM, 2010, pp. 21–26.

[69] R. Ramanathan, R. Hansen, P. Basu, R. Rosales-Hain, and R. Krishnan, "Prioritized epidemic routing for opportunistic networks," in *Proceedings of the 1st international MobiSys workshop on Mobile opportunistic networking*. ACM, 2007, pp. 62–66.

[70] S. Shakkottai and R. Johari, "Demand-aware content distribution on the internet," in *IEEE/ACM Transactions on Networking (TON)*, vol. 18, no. 2. IEEE Press, 2010, pp. 476–489.

[71] C. Singh, E. Altman, A. Kumar, and R. Sundaresan, "Optimal forwarding in delay-tolerant networks with multiple destinations," *Networking, IEEE/ACM Transactions on*, vol. PP, no. 99, pp. 1–1, 2013.

[72] M. Khouzani, S. Sarkar, and E. Altman, "Dispatch then stop: Optimal dissemination of security patches in mobile wireless networks," in *49th IEEE Conference on Decision and Control (CDC), 2010*. IEEE, 2010, pp. 2354–2359.

[73] S. A. Myers and J. Leskovec, "Clash of the contagions: Cooperation and competition in information diffusion," in *Data Mining (ICDM), 2012 IEEE 12th International Conference on*. IEEE, 2012, pp. 539–548.

[74] B. Karrer and M. Newman, "Competing epidemics on complex networks," *Physical Review E*, vol. 84, no. 3, p. 036106, 2011.

[75] A. Beutel, B. A. Prakash, R. Rosenfeld, and C. Faloutsos, "Interacting viruses in networks: can both survive?" in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 426–434.

[76] H. Hotelling, *Stability in competition*. Springer, 1990.

[77] H.-K. Ahn, S.-W. Cheng, O. Cheong, M. Golin, and R. Van Oostrum, "Competitive facility location: the voronoi game," *Theoretical Computer Science*, vol. 310, no. 1, pp. 457–467, 2004.

[78] S. Goyal and M. Kearns, "Competitive contagion in networks," in *Proceedings of the 44th symposium on Theory of Computing*. ACM, 2012, pp. 759–774.

[79] E. Altman, "A semi-dynamic model for competition over popularity and over advertisement space in social networks," in *VALUETOOLS*, 2012, pp. 273–279.

[80] ——, "A stochastic game approach for competition over popularity in social networks," Sep. 2012, available at http://hal.inria.fr/hal-00681959.

[81] O. Baron, O. Berman, and D. Perry, "Shelf space management when demand depends on the inventory level," vol. 20, no. 5. Wiley Online Library, 2011, pp. 714–726.

[82] F. Wu and B. A. Huberman, "Novelty and collective attention," *Proceedings of the National Academy of Sciences*, vol. 104, no. 45, pp. 17 599–17 601, 2007.

[83] L. Backstrom, E. Bakshy, J. M. Kleinberg, T. M. Lento, and I. Rosenn, "Center of attention: How facebook users allocate attention across friends." *ICWSM*, vol. 11, p. 23, 2011.

[84] B. Jiang, N. Hegde, L. Massoulié, and D. Towsley, "How to optimally allocate your budget of attention in social networks," in *INFOCOM, 2013 Proceedings IEEE*. IEEE, 2013, pp. 2373–2381.

[85] J. Nielsen, "Scrolling and attention," 2010, available at http://www.nngroup.com/articles/scrolling-and-attention/.

[86] R. Hassin and M. Haviv, *To queue or not to queue: Equilibrium behavior in queueing systems*. Springer, 2003, vol. 59.

[87] D. R. Cox, *Renewal Theory*. Metheun & Co. Ltd. Science Paperbacks, 1961.

[88] T. G. Kurtz, "Solutions of ordinary differential equations as limits of pure jump markov processes," 1970.

[89] M. J. Osborne and A. Rubinstein, "A course in game theory," *Cambridge/MA*, 1994.

[90] A. Goyal, F. Bonchi, and L. V. Lakshmanan, "Learning influence probabilities in social networks," in *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 2010, pp. 241–250.

[91] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web," Stanford University Computer Science Department, Tech. Rep., 1998.

[92] M. E. J. Newman, "The structure of scientific collaboration networks," *PNAS*, vol. 98, no. 2, pp. 404–409, 2001.

[93] ——, "Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality," *Physical review*, vol. 64, 2001.

[94] J. Morton, J. Dups, N. Anthony, and J. Dwyer, "Epidemic curve and hazard function for occurrence of clinical equine influenza in a closed population of horses at a 3-day event in southern queensland, australia, 2007," *Australian veterinary Journal*, vol. 89, no. s1, pp. 86–88, 2011.

[95] D. J. Daley and J. Gani, *Epidemic Modelling: An Introduction.* Cambridge University Press, 2001.

[96] R. V. Gould, "Collective action and network structure," *American Sociological Review*, pp. 182–196, 1993.

[97] A. P. C. Bordenave, D. McDonald, "Random multi-access algorithms, a mean field analysis," in *43rd Allerton Conference*, 2005.

[98] R. Darling, "Fluid limits of pure jump markov processes: A practical guide," 2002, available at arxiv.org/pdf/math/0210109.

[99] R. Watson, "On an epidemic in a stratified population," *Journal of Applied Probability*, pp. 659–666, 1972.

[100] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, "Incentives for combatting freeriding on p2p networks," in *Euro-Par 2003 Parallel Processing.* Springer, 2003, pp. 1273–1279.

[101] R. Groenevelt, P. Nain, and G. Koole, "The message delay in mobile ad hoc networks," *Performance Evaluation*, vol. 62, no. 1, pp. 210–228, 2005.

[102] N. Gast, B. Gaujal, and J. Boudec, "Mean field for markov decision processes: from discrete to continuous optimization," *Arxiv preprint arXiv:1004.2342*, 2010.

[103] H. Smith, *Monotone dynamical systems: An introduction to the theory of competitive and cooperative systems.* American Mathematical Soc., 2008.

[104] J. Hale, *Ordinary Differential Equations.* RE Krieger, Malabar, FL, 1980.

[105] V. Borkar, *Stochastic approximation: a dynamical systems viewpoint.* Cambridge Univ Pr, 2008.

[106] M. Gocke, "Various concepts of hysteresis applied in economics," *Journal of Economic Surveys*, vol. 16, no. 2, pp. 167–188, 2002.

[107] T. Puu, "Chaos in duopoly pricing," *Chaos, Solitons & Fractals*, vol. 1, no. 6, pp. 573–581, 1991.

[108] O. L. V. Costa and F. Dufour, "Stability and ergodicity of piecewise deterministic markov processes," *SIAM Journal on Control and Optimization*, vol. 47, no. 2, pp. 1053–1077, 2008.

[109] M. H. Davis, "Piecewise-deterministic markov processes: A general class of non-diffusion stochastic models," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 353–388, 1984.

[110] R. W. Wolff, *Stochastic Modelling and the Theory of Queues*. Englewood Cliffs, New Jersey: Prentice Hall, 1989.

[111] V. G. Kulkarni, *Modeling and Analysis of Stochastic Systems*. London, UK: Chapman and Hall, 1995.

[112] P. Brill and M. Posner, "Level crossings in point processes applied to queues: single-server case," *Operations Research*, vol. 25, no. 4, pp. 662–674, 1977.

[113] S. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability*. Springer-Verlag, 1993.