IISS: A Framework to Influence Individuals through Social Signals on a Social Network

by

Tien D. Le

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

Approved April 2014 by the
Graduate Supervisory Committee:

Hari Sundaram, Co-Chair
Hasan Davulcu, Co-Chair
Huan Liu

ARIZONA STATE UNIVERSITY

May 2014

ABSTRACT

Contemporary online social platforms present individuals with social signals in the form of news feed on their peers' activities. On networks such as Facebook, Quora, network operator decides how that information is shown to an individual. Then the user, with her own interests and resource constraints selectively acts on a subset of items presented to her. The network operator again, shows that activity to a selection of peers, and thus creating a behavioral loop. That mechanism of interaction and information flow raises some very interesting questions such as: can network operator design social signals to promote a particular activity like sustainability, public health care awareness, or to promote a specific product? The focus of my thesis is to answer that question.

In this thesis, I develop a framework to personalize social signals for users to guide their activities on an online platform. As the result, we gradually nudge the activity distribution on the platform from the initial distribution p to the target distribution q. My work is particularly applicable to guiding collaborations, guiding collective actions, and online advertising.

In particular, I first propose a probabilistic model on how users behave and how information flows on the platform. The main part of this thesis after that discusses the Influence Individuals through Social Signals (IISS) framework. IISS consists of four main components: (1) Learner: it learns users' interests and characteristics from their historical activities using Bayesian model, (2) Calculator: it uses gradient descent method to compute the intermediate activity distributions, (3) Selector: it selects users who can be influenced to adopt or drop specific activities, (4) Designer: it personalizes social signals for each user.

I evaluate the performance of IISS framework by simulation on several network topologies such as preferential attachment, small world, and random. I show that the framework gradually nudges users' activities to approach the target distribution. I use both simula-

tion and mathematical method to analyse convergence properties such as how fast and how close we can approach the target distribution. When the number of activities is 3, I show that for about 45% of target distributions, we can achieve KL-divergence as low as 0.05. But for some other distributions KL-divergence can be as large as 0.5.

To my Family and Friends.

ACKNOWLEDGEMENTS

It has been an exciting and enriching journey towards my thesis. I am indebted to several people who have contributed directly and indirectly to my thesis. First and foremost, I would like to thank my advisor, Dr. Hari Sundaram for his guidance in the last two and a half years. Working with him is such a prestigious opportunity; he is a terrific scholar and always gives me insightful advice, and comments. I also would like to thank Dr. Hasan Davulcu and Dr. Huan Liu for being part of my thesis committee and reviewing my thesis.

I would like to thank my grandmother, my parents and my sister for unconditional love and support during my journey. They are always with me!

And finally, I would like to thank my professors, and friends in Reflective Living Group, and 10,000 Solutions Group for valuable discussions. Especially I would like to thank my lab mate and wonderful friend Kaushik Sarkar, discussions with him on research or any random topic are always fun and fruitful for me.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure | Page

Chapter 1

INTRODUCTION

The study of behavioral loops in social networks is of great significance in collective action challenges. On social online platforms such as Facebook, Quora, Twitter, users are being shown with activities of their connections. Sequentially, they only interacts with a subset of that activities list due to their own interests and resources constrains (e.g constraint in time). And then their activities are being shown to their connections, thus creating a stochastic behavioral loop.

Many organizations interested in collective action use social networks to engage people in public goods problems. Imagine that the City of New York has released a mobile application that will increase sustainable behavior, in the city. This can include things like carpooling for those coming in to work, doing the laundry at night, nearest composting bin, letting you know of the nearest subway or bus stop and route to use, and expected time of arrival, when hailing a cab. The application allows you to observe and comment on the social signal activities of your friends as well strangers. Since City officials know that not everyone in the city can adopt all behaviors, since people have different inclinations to participate and different constraints on resources (e.g. time and money), they will have a target distribution of behaviors that they wish to see adopted in the population.

There is noise that disrupts the reception of the social signal. Noise can arise due the individual infrequently checking the social network, or because she pays more attention to a subset of her neighbors depending on the activity. She may pay more attention to John for carpooling, for example, than Mary who lives far away from her. Since most networks organize information in a list, where the most recent message appears at the top of the list, probability of reading a piece of information falls with the rank (or more precisely the age) of the information. Network operator can alter visibility and therefore the probability of inaction of a piece of information through manipulation of information rank.

Can we create the right set of social signals overcoming the noise for each individual in the City's social network application, so that they are nudged to participate in the City's sustainability campaign? Notice that socio-information networks are feedback-control systems, with activity distribution as the key state vector. To address this question we need to formulate a fundamental problem of network control: design social signals for each individual such that we achieve the target activity distribution vector.

The following sections of this chapter are organized as follows: in section 1.1 we formally state the problem, in section 1.2 we divide the problem into sub-problems and discuss key ideas to tackle those sub-problems. Section 1.3 summarizes the main contributions of this thesis, and section 1.4 briefs on the organization of the next chapters in this thesis.

1.1 Problem Statement

Let us assume that we have a social information network with N individuals. Each individual can engage in a variety of activities—upload a post, read, and comment on posts. Each activity results in notifications as social signals to other network participants who have subscribed to her activities. On the other hand, we assume that each individual only receives notifications of activities from a specific group of other individuals whom she subscribes to, we define that group as her source of information. Without loss of generality, we assume that each article uploaded in the network belongs to one of K topics.

At the time slice t, we define $u_k(t)$ is the number of users who have activities on the topic k, and then we define activity distribution $p(t)$ as a K dimensional vector with the k^{th} element is defined as:

$$p_k(t) = \frac{u_k(t)}{\sum_{k=1}^{K} u_k(t)}; k = 1..K \tag{1.1}$$

The question that we wish address is how to re-wire the network of subscriptions (or source of information) so that we can adjust the social signals to transform an initial activity distribution p over a population to another target distribution q. Furthermore, we wish to identify the range of the population activity distribution that can be reached from any initial population activity distribution p. That is, determine Q such that we can achieve the target distribution $q, \forall q \in Q$. In this thesis, we shall use the phrase "activity distribution" to imply the population activity distribution.

1.2 Sub-problems and Discussion

In order to transform the activity distribution p to the target distribution q over a period of time we need to address three prominent sub-problems. First, we need to determine the rate at which we should attempt to change the distribution. Second, we need to select the suitable individuals who can be influenced to adopt or drop a certain activity. Finally, for each selected individual to influence, we need re-wire her source of information, that is a group of other persons in the network, who will influence the individual to act.

- How should we change the distribution?

  We use KL divergence to measure the distance between the two distributions p and q. In order to guide the distribution p toward the distribution q we should keep decreasing the KL divergence between the two distributions. Since KL divergence is a convex function, gradient descend method can be used to calculate intermediate values of q and that guarantees convergence. Details of the solution will be discussed in the section 4.3.

- How should we select individuals to influence?

  In order to change the activity distribution, we may need more users to adopt some activities while for other activities we may need some users to drop those activities. How we select potential users to influence them to adopt an activity, should we select users who are intrinsically interested in that activity or should we select users who already have many close connections adopting that activity? On the other hand, how we select potential users to drop an activity, should we select users who are less interested in that activity or should we select users who have a weak network of connections with that activity? Details of the solution will be discussed in the section 4.4

- How should we re-wire the source of information?

  After selecting potential users to influence them to either adopt or drop a particular activity, how should we re-wire their source of information so that we can achieve optimal result? People are more influenced by close friends [33, 35], and people are also more inclined to follow behavior of similar others [11, 39]. Therefore, we should consider those factors in our solution. Details of the solution will be discussed in the section 4.5.

1.3 Contributions of the Thesis

The following are the main contributions of this thesis.

- We propose a probabilistic model of how users behave and how information flows on an online platform.

- We develop a Influence Individuals through Social Signals (IISS) framework to nudge activity distribution toward the target distribution on an online platform.

1.4 Organization of the Thesis

The rest of this thesis is organized as follows:

- Chapter 2 discusses related work, we discuss five main topics which relate to our work: Social influence and Selection, Effects of Information Display, Online Advertising, Recommendation Systems, and Resource Constraint Studies.

- Chapter 3 proposes a probabilistic user model on how users behave and how information flows on an online platform.

- Chapter 4 discusses in detail the IISS framework which consists of four main parts: Learner: it learns users interests and characteristics using Bayesian model (4.1,4.1), Calculator: it calculates the intermediate activity distributions (4.3), Selector: it identifies users to influence (4.4), and Designer: it personalizes the social signals for each user(4.5)

- Chapter 5 discusses experimental results and analysis. We run the experiments for comprehensive combinations of input settings and analyse the convergent speed and property which indicate how fast and how possible that the framework can nudge the activity distribution from the initial value p to the target value q, the analysis includes both simulation results and mathematical analysis.

- Chapter 6 concludes our thesis and discusses some ideas for future extension.

Chapter 2

RELATED WORK

Our related work consists of bodies of work on social influence and selection, effects of information display, online advertising, recommendation systems, and resource constraint studies.

2.1 Social Influence and Selection

"Birds of a feather flow together"

In this section we review the study of social influence and selection from different disciplines, in both offline and online contexts. We also review the study on how people are influenced by their friends, their similar people, and strangers; and role of weak ties and strong ties.

Social influence and selection are two mechanisms under the concept of homophily, which implies that we tent to be similar to our friends. Social influence or socialization is a process in which a person's cognition, attitude, or behavior are influenced by our friends, while selection is a process in which similar people tend to make friends and bond with each other. They have been long studied in the domain of psychology, Cialdini and Trost [12] studied three core components of social influence: social norms, conformity, and compliant. Kandel [33] in a empirical, longitudinal study to find the characteristics of pairs of adolescent friends via their scholastic achievement and delinquent behavior found that both selection and socialization play roles in their friendship. Teenage friends are similar to each other, the first reason is because they make friends with people who are similar to them, and then conformity and compliant cause them to follow the group behaviors to conform with other peers in their group. Social influence not only happens within friends, Rogers [39] mentioned that social influence occurs more often between individuals who are more similar. Similarly, when talking about *social proof*, Cialdini [11] showed evidences and statistics that people are more inclined to follow the behavior of similar others. Tie strength also play an important role in social influence. Granovetter [23] found that weak ties likely form "local bridges" and for that reason weak ties play much more important role then strong ties in diffusion processes. On other hand, Krackhardt [35] found that when people are facing with uncertainties and changes, they are more dependent on strong ties than weak ties.

Study of social influence and selection is very important, Easley and Kleinberg [19] pointed out that if we understand the nature of selection and socialization of a community then we can have appropriate method to change the behavior of the community, for example, in a drug use community, if the reason of using drug is due to socialization then by changing the behavior of the influential nodes it will lead to the change of the behavior in the whole community. The idea of targeting influential nodes to optimize the effect of intervention is also used in viral marketing.

The proliferation of online websites with huge data of activities and interactions like Facebook, Twitter, Wikipedia offers unprecedented opportunities for research of social influence and selection. Wang and Chin [43] studied the effect of social influence on Flickr and showed that a user has higher number of contacts who are paid users would have higher probability be a paid user also. Similarly, Bakshy et al. [4] in a study on Facebook found that the probability of sharing a link on Facebook is proportional with number of friends who already shared the link, they also confirmed the important role of weak ties over strong ties in information diffusion process as in the study of Granovetter [23].

Crandall et al. [14]studied the role of social influence and selection in future behavior prediction. The group created two networks: interaction network and similarity network. In each network, they analyzed the probability of a user who has k friends having done some action will also do that action. Compare the probabilities from two networks will give the information about whether socialization or selection is more important in behavior of a user. Applying that method to two datasets (Wikipedia and LiveJournal) they found that in Wikipedia, social interaction is a better predictor of future behavior (future behavior is more affected by socialization) while in LiveJournal, similarity (selection) is a better predictor of future behavior (future behavior is more affected by selection).

Again, the study of social influence and selection in social network is very important. Selection and socialization also can lead to other global effects such as epidemic, and information diffusion. The formation or cessation of links basing on selection and socialization is a main cause for network dynamics and network evolution. Easley and Kleinberg [19], Holme and Newman [29]showed that from the local process of selection and socialization it can lead to global effect: the formation of communities (uniformity, segregation). Therefore, in the community discovery problem, study of social influence and homophily plays a crucial role in bottom-up approach. The study of social influence and selection is also fundamental in study of information diffusion processes.

In the next section we will review research on effects of information display which is also very relevant to the research that we have reviewed in this section about social influence and selection.

2.2 Effects of Information Display

Information display can make impactful effects on our behavior. It creates social influences, happens in offline contexts as well as online contexts, and has effects on friends, similar people, and even strangers.

In offline context, several well-known studies that indicate that revealing hidden social norms is a powerful cue for cooperation. In a study of energy use Schultz et al. [41] found that individuals were presented with affective symbols—a smiley face '☺' when an individual household's energy consumption was below the mean household consumption and a frownie face '☹' when it was above the mean—in addition to mean neighborhood consumption, there was a net decrease in energy consumption over the entire population [1] . In another study Goldstein et al. [22] demonstrated that presenting guests with the descriptive norm "most guests reuse their towels" significantly improved towel reuse in comparison to the standard environmental message.

In online world, in a study Bond et al. [7] about social influence and political mobilization, 61 million Facebook users were assigned into one of three groups: social message, information message, and control group. Users in social message group were shown with general messages as well as messages from friends' voting activities, users in information message group were only shown with general messages, and users in control group were shown with no messages. The study found that, due to social influence from friends via information display, the social message group has highest percentage participating in voting activities.

Effects of information display are not only limited within friends or similar people. Salganik et al. [40] conducted an online experiment in an artificial music market. Users participated in the experiment were divided into two groups: independent group, and

[1]OPower now provides this information to over a hundred power utility companies in the United States. `http://www.opower.com`.

group with social influence. In social influence group users were shown with the information about number of downloads from previous listeners. The experiment also smartly used multiple worlds design and different presentation schemes to vary the social influence strength. Result from the study shows that increase in social influence via information display leads to increase in both inequality and unpredictability of success.

In another study, Hodas and Lerman [27] analysed re-tweeting activities on Twitter and found out that, on Twitter where tweets are shown in chronological order with the most recent tweet is at the top of the screen, retweet probability is inversely proportional to the time elapsed since tweet arrival. That implies that the visibility has a significant effect on re-tweeting behavior which is a form of social contagion, social influence.

2.3 Online Advertising

In this section we focus on reviewing the study on online advertising, different types of advertising and targeting.

Online advertising is the main source of revenue for many technology companies like Google, Facebook, Twitter, Microsoft, Yahoo!. Comparing to traditional offline advertising, Goldfarb [20] argued that while they have the same purpose of presenting information and persuading customers to buy the advertised products, online advertising has a substantially lower cost of targeting, and that is the fundamental economic advantage of online advertising over offline advertising. However that argument is relatively general, Goldfarb and Tucker [21] in another study dug in more deeply and argued that online advertising has two main advantages over offline advertising: higher measurability, and higher targetability. Higher measurability because for online advertising we can easily track responses, and higher targetibility because for online advertising we can easily track at an individual level and therefore personalize ads.

There are two main types of online advertising: search advertising, and display advertising [20]. In search advertising, search engines like Google, Bing show sponsored links regarding to user's queries, search advertising is priced using auctions. In display advertising, customers are shown with text ads (for example Google's Adsense), banner ads, and video ads.

Targeting is very important for advertising, advertisers have to answer two main questions: who are target, and how to target? Because with poor performance of targeting, advertising can causes reverse effects, irrelevant advertising causes negative experience and results in avoidance of advertising [18, 30, 34]. Targeting models in search advertising and display advertising are slightly different, however they can be classified into three categories: contextual targeting, behavioral targeting, and social network based targeting.

Contextual targeting matches content of the website with customers' interest, it is commonly used in both search advertising and display advertising. In search advertising, advertisers understand customers' interest via search content and show relevant websites. In display search, advertisers display ads in relevant places: electric gadgets are advertised on cnet.com, techcrunch.com, cars are advertised on cars.com, makeup and cosmetics products are advertised on beauty care websites.

Behavioral targeting on other hand learns customers' interest from their historical activities including sites visited, interest in particular content, or purchasing activities, and then display relevant ads [5, 9]. Technically in behavioral advertising, advertisers have to collect customers' s data by storing activities of logged-in customers or by using cookies to track customers' activities.

Social network based targeting (or social advertising) uses different strategy than contextual targeting and behavioral targeting do. Social network based targeting uses information of customers' social network to allocate potential targets and then provide them with personalized social signals as the stimulus for the ads. Its principle psychologically bases on social influence and homophily that I have reviewed above. With the nature of social influence, advertisers may display a customer's network activities as social signals to lead her to a particular purchase activity. In addition to that, homophily implies that friends have tendency to do similar things, in that sense social network contains latent customer characteristics which helps advertisers to target likely adopters [42]. Hill et al. [25] in a study showed that people linked to prior customers 3-5 times more likely adopt the service than people are selected as in the baseline. Provost et al. [37] on other hand introduced concept of brand proximity which is calculated based on user-user similarity to select most relevant users with the implication that similar people tend to do similar things. The strength of social signal also play a very important role on social influence. Bakshy et al. [3] conducted two experiments on Facebook and showed that increases in number of social cues leads to increases in advertising performance, and people are more influenced by others with strong ties.

2.4 Recommendation Systems

In this section we review the study on recommendation systems, focusing on two main types of recommendation systems: collaborative filtering and content based filtering.

Recommendation Systems are used in a variety of domains: Amazon recommends an user a list of books after she bought a particular book, Netflix recommends an user a list of movies to watch basing the customer watching and rating activities, basing on what an user has viewed recently The New York Times recommends she with related articles, on a question and answer site Quora, for each new question the system recommends list of relevant users to ask for answers, on Google News an user can her customize preference to see more relevant articles. In general, there are two types of recommendation systems: collaborative filtering, and content based filtering.

For e-commerce websites, recommendation systems play a very crucial role in increasing activities and revenue. About 60 percent of Netflix subscribers select movies from the system recommendations [2]. Similarly, recommendations account for about 60 percent of video clicks from the YouTube homepage [15]. Recommendation systems are also important in the sense that they can change users' preference and therefore change the revenue of sellers significantly [3]. If a recommendation system chooses to recommend popular items, it will lead to long tail distribution, but if it recommends lesser-known items the distribution becomes more equal and it eventually increase the revenue [8]. Those important roles of recommendation systems is one of the reasons NetFlix launched a contest in 2006 with the prize of $1 million to improve its recommendation system- CineMatch- to 10 percent [4] [6].

Basically, the purpose of recommendation systems is to provide personalized information. In order to do that, many recommendation systems use customers purchase activities and explicit rates as their interests, but they also can use other customers' personal data like demographic data, personal interests [36]. In general, there are two main types of recommendation systems: collaborative filtering, content based filtering [1].

Collaborative filtering algorithm bases on user's information like behavior, activity and preferences to construct user-to-user similarity matrix. After that the algorithm predicts what users will do based on their other similar users. In more details of constructing user-to-user similarity matrix, normally the algorithm represents a user as an N-dimensional vector where N is number of features in the considering problem, e.g., for an e-commerce website recommendation system, N is number of items. Components of that vector represent the user information, for example the i^{th} component is

[2] https://netflix.hs.llnwd.net/e1/us/pdf/Consumer_Press_Kit.pdf

[3] http://www.nytimes.com/2006/06/07/technology/07leonhardt.html?ex

[4] http://www.netflixprize.com/

proportional to the number of times that user bought the i^{th} item, it is also proportional to the rating of that user about the item. Normally the vector is very sparse. After having the vector space of all users, there are a number of methods to calculate the similarity matrix; one of the popular methods is cosin similarity measurement. A key advantage of the collaborative filtering approach is that it does not require the understanding of the content of items and therefore it is very suitable in recommendation systems in which content of items are very difficult to search or retrieve such as recommendation systems of movies.

Content based filtering algorithm is based on the information about the items that are going to be recommended. In other words, the algorithm tries to recommend the items which are similar to those that a user liked in the past. The idea of the algorithm comes from information retrieval and information filtering search. The algorithm works in the case the system has rich information or "good understanding" of items and has efficient methods to search and retrieve related items for each set of items. There are recently a number of question and answer (Q&A) sites such as Yahoo! Answers, Stack Overflow, and Quora that lead to numerous studies on recommendation systems to route questions to potential answers who are interested in and capable of answering them. The main stream of those studies just study users, analyse questions and then establish user-question relationship to decide who most potential answers are for each question, and hence skipping user-user relationship Guo et al. [24], Kabutoya et al. [32], Qu et al. [38]. Other than those studies, Horowitz and Kamvar [31] present a social searsch engine called Aardvark, Aardvark takes into account both Topic expertise (user-question relationship) and Connectedness (user-user relationship) in ranking system to route questions to the most suitable answers.

2.5 Resource Constraint Studies

We all have resource constraint which results from our limited human brain's capacity, because of that it is very important to take resource constraint factor into consideration when studying social networks. As a subsequence of resource constraint, we divide our attention to different source of stimulus and interaction: we tend to pay more attention to important and more visible source of information. Backstrom et al. [2] analysed Facebook data from 2009 to 2010 and found that users paid attention unequally to their friends, for all kinds of activity: profile view, photo view, comment, message, wall post the level of attention toward the k^{th} friend are proportional to $k^{-\frac{3}{4}}$. Hodas and Lerman [28] analysed Twitter data and showed that limited attention and decaying visibility play crucial roles for retweeting behavior, the probability of retweet an URL after having a friend tweeted it at Δt ago is proportional to $\Delta t^{-1.15}$, and probability of tweeting an URL for the first time (without having any friend tweeted it) after the URL arrived at Δt ago is proportional to $\Delta t^{-0.5}$.

We have reviewed five relevant research bodies to our research: social influence and selections, effects of information display, online advertising, recommendation systems, and resource constraint studies. Our research is unique, we are not trying to improve the performance of any existing algorithm or framework, we are building a new framework to guide activities on a social network. The research on social influence and effects of information display evidence that we can personalize social signals to nudge users' activities. While research on online advertising and recommendation systems gave us initial ideas about how to select suitable users to induce, but state of the art in those researches have not explored social connectivity enough, our research explores more the social connectivity by "re-wiring" the information network to personalize social signals for each user. Because resource constraints play crucial roles on how users behave and how information flows on online platforms, our research also takes into account the importance of resource constraint studies in modelling users' behavior and in designing the framework to guide the activities.

Chapter 3

USER PROBABILISTIC MODEL

In this chapter we shall discuss our model for user characteristics, interests, and mechanism of users interactions which defines how information is exchanged within the network. The user model and the specification of how information flows in the network will be used to simulate network activity.

3.1 Basic User Characteristics

Each individual i in the network possesses a set of immutable attributes as well as set of preferences. The immutable attributes of an individual i, including gender, ethnicity, age, are represented by a normalized $D-$dimensional vector $\mathbf{a}_i; 0 \leq a_{i,j} \leq 1; j \in \{1, 2, \ldots, D\}$. Let us assume that an individual is interested in $L \subseteq K$ themes. For each theme $k \in K$ an individual i has a preference value $\beta_{i,k}; 0 \leq \beta_{i,k} \leq 1$, indicating her interest in the theme. We assume that if $k \in L$ then the preference value $\beta_{i,k}$ has a high value, and if $k \notin L$ then the preference value $\beta_{i,k}$ will be low. The preference vector $\boldsymbol{\beta}_i$ for any individual affects the her activities on the different themes.

We assume that individuals are resource constrained. This assumption is motivated by earlier work by Dunbar [16, 17]. His research indicated that maintaining social ties requires resources and that group size in humans is bounded by the number 150. More recent work by Hodas and Lerman [26] indicates that limited attention amongst participants is a compelling explanation for arresting information contagion on Twitter. To operationalize the idea of resource constraints, we assume that each individual i will browse at most z_i pieces of information from her contacts. We shall assume that z_i is drawn from an exponential distribution with parameter $\lambda_1 : \exp(-\lambda_1 X); M_1 \leq X \leq \infty$.

The lower bound M_1 is the fewest pieces of information seen by any individual. Each individual i is also constrained in terms of the number of information items c_i that she can create. Similar to the variable z_i, we shall assume that c_i is drawn from an exponential distribution $\lambda_2 \exp(-\lambda_2 X); M_2 \leq X \leq \infty$. The lower bound M_2 is the fewest pieces of information created by any individual. The items that any individual creates thus contributes to the information circulating in the network. There are two implications that arise of resource constrained individuals. First, trying to signal an individual with a large number of contacts will be hard, since the likelihood of the message being lost in the deluge will be high. Second, individuals who create lots of information (that is, very active in posting new articles or commenting on peers' articles), will likely "swamp" their less connected neighbors.

Now that we have presented the basic characteristics for each individual in the network, we turn to what individuals can do in social information networks.

3.2 The User Activity Model

A user can either create new information (link an article, or post a new idea) or act on existing information (view, comment), "like", favorite, etc.). We assume that the information about activity in a network is organized in a list, with the newer pieces of information ranked higher than older pieces of information. List based ordering is standard in many social networks, including Twitter and Facebook.

Let us first examine information access rights of an individual. When an individual i joins the network, she can access posts, that is new articles, from all of her network neighbors. When i's neighbor j comments on k's article, i can also access the original article posted by k. She can also comment on on k's article. If she does, we say that k "is aware" of i. Subsequently, if k either responds to i's comment, or comments on i's article made visible through comments by a mutual friend, we say that individual's i

and k are "mutually aware." In our network, when two individuals are mutually aware, they have access to each other's posts and comments. In this manner, the *information neighborhood* of each individual expands with time. Notice the distinction between *information neighborhood* and *network neighborhood*. The latter designates the set of friends, while the former is the the set of individuals from whom a person can access information. The proposed access model is different from networks such as Facebook, where one can only access information from friends [1] , and from Twitter, where one can access any other individual's public timeline. Accessing information based on "mutual awareness" is thus a middle ground between the two popular information access models [2] , where an individual's information neighborhood expands slowly with "mutual awareness."

An individual creates articles with the following probability:

$$P_i(create = T|k) = \gamma_c \times \beta_{i,k}. \tag{3.1}$$

Where γ_c is a constant $0 < \gamma_c < 1$; $\beta_{i,k}$ is the value of individual i's preference on topic k. The equation says that the probability of creation of a new article on topic k is proportional to her topic preference.

A user's desire to view and to act on articles in the social network will be influenced by her relationship to the individual who posts or comments as well as the topic of the article. In this work, we assume that an individual is incentivized to view the article *only* by the author of the post or the comment. To act, that is to comment or to favorite the article, she is motivated only by the content of the article. The decisions to view and subsequently act are thus assumed to be motivated by different, independent criteria.

[1] An individual on Facebook can explicitly allow any specific information to be accessible by friends of friends as well as anybody on the network.

[2] While not implemented in our model, an individual can exchange private messages only with friends.

Individuals are persuaded to adopt the behaviors of friends or of strangers with whom they can identify, a phenomena known as " the social proof" [Chapter 4, 10]. With this in mind, in our simulation, individuals are persuaded to view articles when they see strangers with whom they can identify, or with friends. The similarity score between two individuals i and j, including strangers is computed as follows:

$$s_{i,j} = 1 - \frac{||\boldsymbol{a}_i - \boldsymbol{a}_j||}{\sqrt{D}}. \tag{3.2}$$

Where $\mathbf{a}_i$ is the immutable attribute vector for individual i and $||\cdot||$ is the familiar euclidean norm, and D is the number of dimensions in the attribute vector. The strength of the tie between two friends j and k is computed as follows:

$$f_{j,k} = \frac{|\mathcal{I}_j \cap \mathcal{I}_k|}{|\mathcal{I}_j \cup \mathcal{I}_k|}. \tag{3.3}$$

Where, $\mathcal{I}_k$ is the information neighborhood of individual i. The equation says that the tie strength between two individuals is highest, when their information neighborhoods completely overlap. Now, the probability that individual i views an article is determined as follows:

$$P_i(view = T|origin = j) = \begin{cases} \gamma_f \times (1 - \alpha^{1+f_{i,j}+s_{i,j}}) & \text{friends,} \\ \gamma_s \times (1 - \alpha^{1+f_{i,j}+s_{i,j}}) & \text{strangers.} \end{cases} \tag{3.4}$$

Where, γ_f and γ_s are two constants with $0 \leq \gamma_s \ll \gamma_f \leq 1$. The constant α, where $0 \leq \alpha < 1$ ensures that the probability of viewing an arbitrary article lies above a threshold; that is $P_i \geq \gamma_s \times (1 - \alpha)$. The equation says that probability of viewing a piece of information for an individual i *only* depends on the origin or the article, and increases with increasing tie strength $f_{i,j}$ and increase in attribute similarity $s_{i,j}$. Since $\gamma_f \gg \gamma_s$ the probability for viewing an article from a stranger is guaranteed to be lower than for an information neighbor, for the same tie strength and same degree of similarity.

The probability of acting on an article after viewing it, *only* depends on the content of the article. In this paper, we assume that the topic of the article is a good proxy for the content. Thus,

$$P_i(action = T | view = T, topic = k) = \gamma_a \times \beta_{i,k}, \tag{3.5}$$

where, γ_a is a constant, and $\beta_{i,k}$ is the preference value for individual i on topic k. The equation says that the probability of acting on an article after viewing it is proportional to the individuals interest in the topic.

Notice that the user model has several "universal" constants: γ_f, γ_s, γ_a and α. These constants are universal in the sense that each agent's model in the simulation will have the same constants. One can relax this assumption with the consequence that the number of activity distributions over the population will increase.

In this chapter we presented the user model which simulates how information is exchange in a social network. In the next chapter we will present a framework which learns latent users' preferences, characteristics from the system perspective and then ultimately personalizes social signals for each user to nudge the activity distribution from the initial value $p(0)$ toward the target distribution q.

Chapter 4

IISS: INFLUENCE INDIVIDUALS THROUGH SOCIAL SIGNALS FRAMEWORK

In this chapter we discuss the IISS framework which is used to influence users via social signals to achieve a target activity distribution. In particular, we shall discuss how to alter the activity distribution within a network from a starting distribution p to a target distribution q. First, we shall estimate users preferences, identify active individuals. Then, we shall determine the extent to which we shall change the distribution of individuals active on a particular topic. Then, we shall identify the size of the subset of individuals to influence. Finally, we will show how to 're-wire' the information network, so that each individual receives an optimal social signal.

4.1 Estimating User Interests

Let us assume that we can keep track of individual actions on each topic k, including the number of views $v_{i,k}$ and the number of comments $c_{i,k}$. Then, the probability that an individual i comments on the article is

$$\widehat{P}_i(action = T | view = T, topic = k) = \frac{\widehat{P}_i(action = T, view = T, topic = k)}{\widehat{P}_i(view = T, topic = k)} \tag{4.1}$$

$$= \frac{c_{i,k}}{v_{i,k}}. \tag{4.2}$$

Where, $\widehat{P}_i$ represents the system estimate of the probability. The equations say that the probability of acting on an article for an individual is simply the ratio of the number of her comments $c_{i,k}$ on topic k to the number of her views $v_{i,k}$ on the same topic.

It is possible to develop a algorithm for estimating the viewing probability and the action probability for each individual, through data likelihood maximization techniques for estimating model parameters including γ_f, γ_s and γ_a. In this paper, we have chosen not do so, since in real-world scenarios, these constants will vary with individual. Estimating these parameters for each individual will require significant number of observations per individual for stable statistical estimates.

Having discussed the basic user characteristics, the model for each agent, and how the system estimates probabilities of viewing and of actions on each topic, we now discuss how to identify active users.

4.2 Identifying Active Individuals

The first step towards changing the distribution of activities in a network is to identify individuals $u_k(t)$ active on a topic k. As before, we define $u_k(t)$ as the number of individuals active on topic k at time t. That is, these individuals comment on, or create, posts related to topic k. Measuring the number of active individuals on a topic at any time slice t is inefficient, since the number of individuals active on a topic is a random variable. Instead of a single snapshot, at time t, we measure the number of active users over a time period $(t, t-\Delta)$, to have a better estimate of the average number of individuals $\hat{u}_k(t)$ active on topic k. That is,

$$\hat{u}_k(t) = \frac{1}{\Delta} \sum_{t'=t-\Delta+1}^{t'=t} u_k(t'). \tag{4.3}$$

The equation says that the value of the number of active individuals at any time t is averaged value over the prior Δ observations. In the remainder of the paper, we shall use the variable $u_k(t)$ to imply the averaged value $\hat{u}_k(t)$. Let us assume that for any individual i, $n_k(i, t; \Delta)$ represents the total number of creations and comments on topic k over a time period Δ. To identify "active" individuals, we sort individuals in descending order of $n_k(i,t;\Delta)$. The top $\widehat{u}_k(t)$ users are labeled active on topic k at time t; all others are labeled "inactive" on topic k.

Now that we have identified sets of individuals active on topic k at each time t, we discuss how to optimally change the activity distribution $p(t)$.

4.3 Changing the Activity Distribution

The activity distribution $p(t)$ is the distribution of *activities*—different from the distribution of active individuals—over each topic k. We propose a gradient descent algorithm to determine incremental changes to the activity distribution $p(t)$. Hence at any time t, we have:

$$\hat{p}(t+1) = p(t) - \gamma \nabla D(p(t)||q), \tag{4.4}$$

where, $\hat{p}(t+1)$ is the desired activity distribution, q is the target activity distribution, $D(p(t)||q)$ is the familiar Kullback-Liebler divergence between two distributions, and where ∇ is the gradient operator. At each time t, we ensure two properties for $\hat{p}_k(t)$, the proportion of activity devoted to topic k. First, $0 \leq \hat{p}_k(t) \leq 1$. Second, we normalize the target activity distribution over the set of topics k: $\sum_k \hat{p}_k(t) = 1$.

At the time slice t, total number of active users $u(t)$ (over all K topics) is given as

$$u(t) = \sum_{k=1}^{K} u_k(t), \tag{4.5}$$

where, $u_k(t)$ is the number of individuals active on topic k. Assuming that the *total* number of active individuals is stable over time, the desired number of individuals active on topic k is computed as follows:

$$\hat{u}_k(t+1) = u(t) \times \hat{p}_k(t+1). \tag{4.6}$$

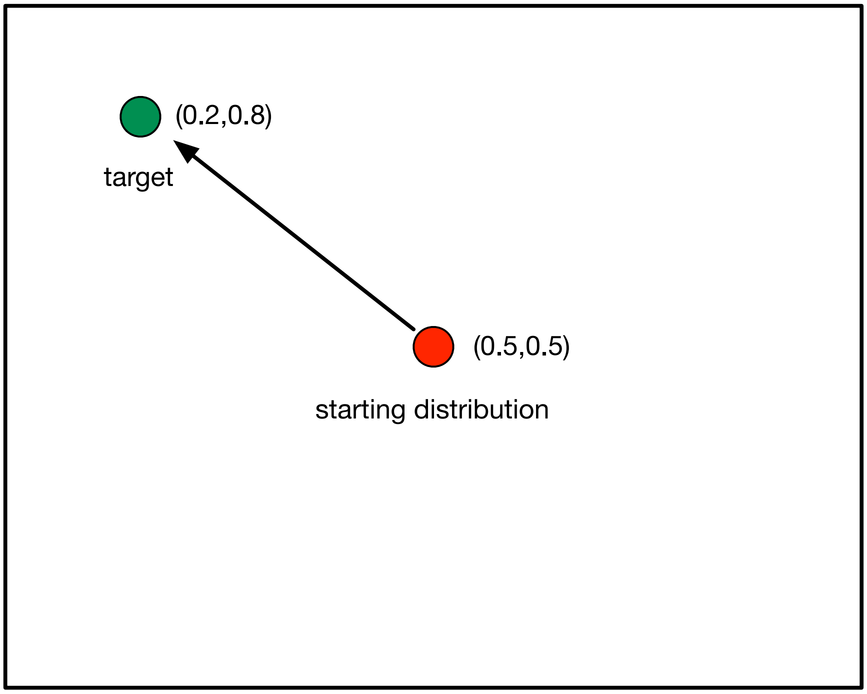

Figure 4.1: Assume that that the goal is to alter an initial distribution $p = (0.5, 0.5)$ to a target distribution $q = (0.2, 0.8)$. While there are several methods to incrementally change the current distribution $p(t)$, we adopt a gradient descent approach, where we take the gradient of the KL divergence of the current activity distribution $p(t)$ with the target distribution q. That is, the change to the activity distribution is proportional to $\nabla D(p(t)||q)$.

So for topic k, the desired change in number of active users on topic k is simply proportional to the changes in the desired activity distribution. That is,

$$\Delta \hat{u}_k(t) = \hat{u}_k(t+1) - u_k(t), \tag{4.7}$$

$$\propto \Delta p_k(t). \tag{4.8}$$

We would like to remind the reader of the difference between the two variables $p(t)$ and $u(t)$. The former refers to the *distribution* of activities—comments, views, creations—across topics. The latter refers to the *number* of active individuals across topics. Notice that at any time, the number of activities on a topic will be greater than or equal to the number of individuals active on the same topic.

Since the KL divergence $D(p||q)$ is convex in the pair (p,q) [Theorem 2.7.2, 13], in principle, using gradient descent should guarantee convergence to target distribution q. However, there are a few things of note here. We are proposing an incremental change to the activity distribution. To achieve the increment, we need to provide the appropriate social signals to each individual in the network. Therefore, there is no guarantee that we can achieve this incremental change, since individuals have to view the signals *and* subsequently act.

Having identified the change in the number of individuals active on each topic, we now address the problem of identifying individuals to influence for each topic k.

4.4 Identifying Individuals to Influence

The identification of suitable individuals to influence rests on two variables: the required change in the number of individuals for each topic $\Delta\hat{u}_k(t)$ (ref. Equation 4.7) and the estimate of an individual i's interest in a particular topic $\hat{\beta}(i,k)$. To estimate the latter, we make use of our estimate of $\widehat{P}_i(action = T|view = T, topic = k)$, Equation 4.1. We shall set $\hat{\beta}(i,k) = \widehat{P}_i(action = T|view = T, topic = k)$.

The second variable $\Delta u_k(t)$, requires us to consider two cases. In the *first* case $\Delta u_k(t) > 0$. This implies that we need to increase the number of individuals performing activity k. We sort individuals inactive on topic k (see Section 4.2), in *descending* order of $\hat{\beta}(i,k)$. Then, we select the top $\Delta u_k(t)$ individuals from this list. These individuals are most likely to act on information on topic k. In the *second* case $\Delta u_k(t) < 0$. In this case, we need to decrease activity on topic k. In this case, we sort all users active on topic k in *ascending* order of $\hat{\beta}(i,k)$. Then, we select the top $\Delta u_k(t)$ individuals from this list. These individuals are least likely to act on information on topic k. Figure 4.2 illustrates the process of adding to, and deleting from, sets of individuals active on a topic.

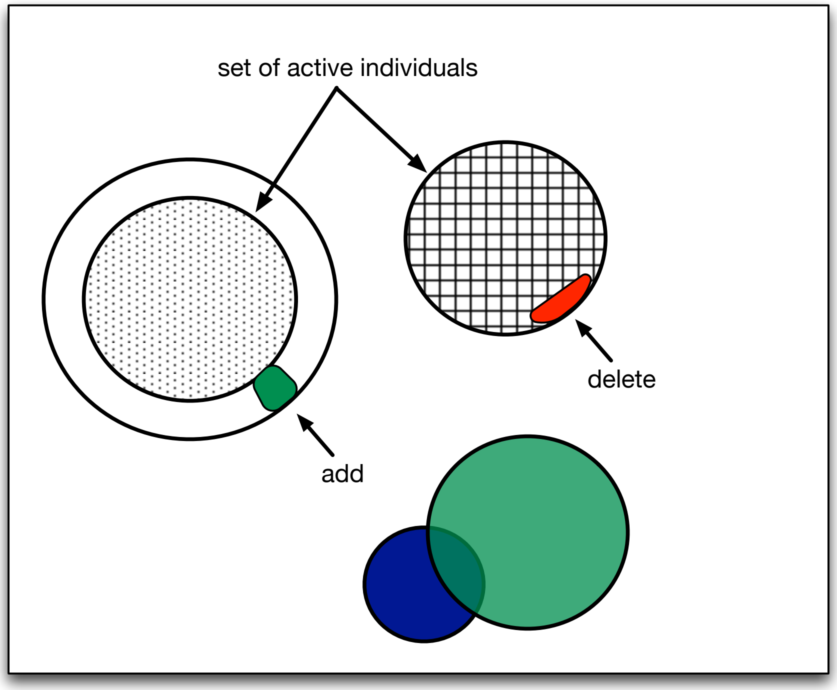

Figure 4.2: The figure shows two cases. In the first example, we wish to increase the set of individuals active on a topic. We do so by identifying those individuals not currently active, but who have the next highest activity on that topic. In the second example, we want to decrease the number of individuals active on a topic. In this case, we identify those individuals who are least active on that topic.

Having presented a framework to add and subtract from a set of individuals active on a topic, we now discuss how design each individual's social signal so that they are are persuaded to act.

4.5 Designing the Social Signals

Personalized social signals that reveal the social norm [22, 41] are key to persuade individuals to adopt. We have developed a framework based on the idea of the "social proof" [Chapter 4 10], for generating personalized social signals for each resource constrained individual so that they are incentivized to adopt a set of behaviors. More formally, since the system controls the information that each individual in the network views, the system can change the information neighborhood for each individual. Let

$S_i(t)$ refer to the information neighborhood of individual i. In other words, i is notified of activities by all individuals in set $S_i(t)$. The information generated by the individuals in the set $S_i(t)$ constitutes the "social signal" of i. Figure 4.3 shows an example of an information neighborhood.

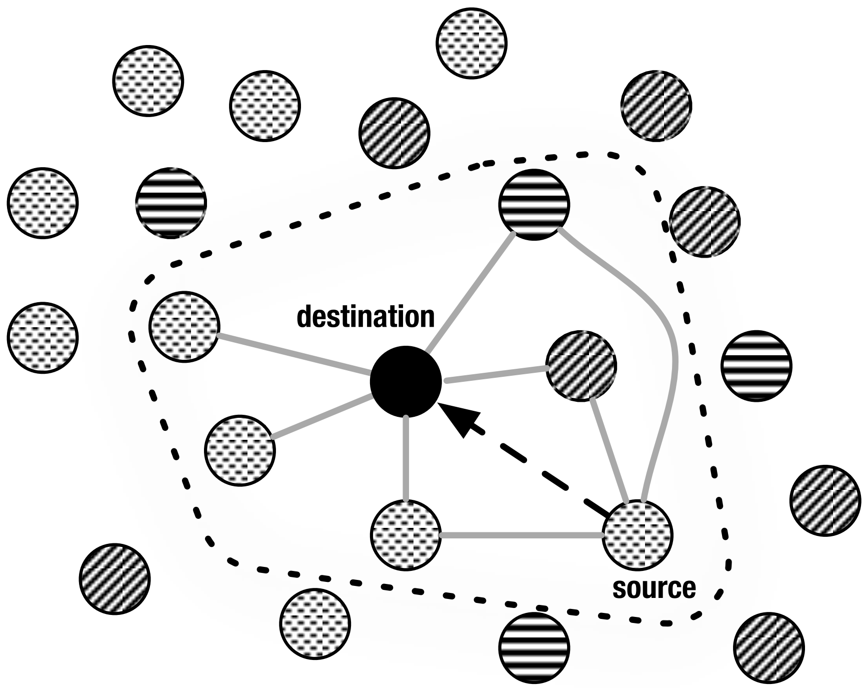

Figure 4.3: The figure shows the information neighborhood for an individual (black circle), indicated by the dotted region, in the the network. The information neighborhood of an individual is the set of people from whom she receives information; this set can include individuals with whom she is friends, or with whom she is similar along her immutable attributes. The figure shows her receiving information from another person; they share two individuals in their information neighborhood.

In Section 4.4, for each topic k, we showed how to identify the individuals to be influenced. In the case $\Delta u_k(t) > 0$, the individuals identified to be influenced for topic k, need to be persuaded to become active in that topic. In the second case, when $\Delta u_k(t) < 0$, the individuals identified to be influenced for topic k, need to be persuaded to become inactive in that topic.

To persuade an individual i to become active on topic k, we first identify a set of relevant friends and other individuals, highly active on topic k, with whom they share significant personal attributes. We shall term this group $L^p_{i,k}$ as "positive influentials" for individual i. In this paper, we assume that the size of the set of positive influentials $|L^p_{i,k}|$ is a constant across all individuals i. To identify individuals belonging to the set $L^p_{i,k}$, we proceed as follows.

1. Identify all *active* users $J_{k,i}$ on topic k, excluding those in $S_i(t)$. That is, $J_{k,i} = \{j | j \in u_k(t) \wedge j \notin S_i(t)\}$.

2. For each $j \in J_{k,i}$, rank in *descending* order of the value of their appropriateness $r(j)$ to be part of $L^p_{i,k}$. Where, $r(j)$ is defined as follows:

$$r(j) = \begin{cases} \hat{\beta}_{j,k} + \delta_f \times f_{i,j} + s_{i,j} & \text{if } i, j \text{ are friends,} \\ \hat{\beta}_{j,k} + \delta_s \times f_{i,j} + s_{i,j} & \text{if } i, j \text{ are strangers.} \end{cases} \tag{4.9}$$

   Where, $0 \leq \delta_s < \delta_f \leq 1$ and represent the system biases towards strangers (δ_s) and friends (δ_f). The parameter $f_{i,j}$ represents the overlap in the information neighborhoods between individuals i and j (ref. Equation 3.3), and where $s_{i,j}$ represents the similarity value between the two individuals (ref. Equation 3.2. The equation says that j is an appropriate positive influence for individual i, if j is either highly active (high value of $\hat{\beta}_{j,k}$) on topic k, has a high overlap in their information neighborhood $f_{i,j}$ or is highly similar in terms of their immutable attributes $s_{i,j}$.

3. Select the top $|L^p_{i,k}|$ individuals to be part of the set of positive influentials for individual i.

To persuade an individual to become less active on a topic, we shall introduce "negative influentials" in her social signal. Intuitively, the set of negative influentials $L^n_{i,k}$ are less active on topic k. We proceed as follows:

1. Identify all *inactive* users $\bar{J}_{k,i}$ on topic k, excluding those in $S_i(t)$. That is, $\bar{J}_{k,i} = \{j | j \notin u_k(t) \wedge j \notin S_i(t)\}$.
2. For each $j \in \bar{J}_{k,i}$, rank in *ascending* order of their appropriateness $r(j)$ to be part of $L^n_{i,k}$. In this case, the appropriateness $r(j)$ is defined as follows: $r(j) = \hat{\beta}(j,k)$.
3. Select the top $|L^n_{i,k}|$ individuals to be part of the set of positive influentials for individual i.

Notice that for any individual i, we may need to strengthen the social signal along topic k but weaken the signal along another topic l. Thus the updated set of sources of social signals for any individual i is determined as follows:

$$\hat{S}_i(t+1) = S_i(t) \bigcup_{k=1}^{K} \left\{ L^p_{i,k} \cup L^n_{i,k} \right\}. \tag{4.10}$$

Where, $\hat{S}_i(t+1)$ is the updated set comprising the set of sources for social signals. $L^p_{i,k} = \emptyset$ when we are unconcerned with positively influencing individual i to act on topic k. Similarly, $L^n_{i,k} = \emptyset$ when we are unconcerned with negatively influencing individual i to act on topic k. Notice that it is possible that

$$L^p_{i,k} = L^n_{i,k} = \emptyset,$$

when we are uninterested in influencing individual i on topic k. Since we fix the size λ of the size of set of sources of social signals to be the same across individuals, we randomly sample λ times from the set $\hat{S}_i(t+1)$ without replacement to generate the social signal

$S_i(t+1)$. Thus the updated social signal is equally distributed across signals from the previous social signal as well as positive and negative influentials.

In this section we discussed several key ideas. First, we introduced the idea of the set of individuals active on a topic. Then, proposed the use of gradient descent to incrementally change the activity distribution and determine the size of the change. Then, we showed how to identify individuals to influence for each topic k. Finally, for each such individual, we design a personalized social signal comprising of positive and negative influentials as well as prior social signals, that would persuade them to act. Having discussed the construction of the social signals, in the next chapter we will present our experimental results and analysis.

Chapter 5

EXPERIMENTAL RESULTS AND ANALYSIS

In this chapter we will present experimental results and analysis. Our framework was implemented in Repast Simphony 2.0 [1], a popular agent based modeling software. In the next section, we shall discuss the values set for the parameters used in our framework. Then, we shall discuss the initial phase of the experiments, without social signals to set up interaction within the network. We shall introduce the social signals after 100 iterations.

5.1 Experimental Parameters

The total number of individuals N is fixed to 400. At the beginning of the simulation, individuals are connected with an initial network topology (e.g. small world). The total number of topics in the network K is fixed.

We now discuss the parameters of each individual's activity model. Each individual is resource constrained affecting the number of article she browses ($z_i; M_1 = 4; \lambda_1 = 0.5$) and the number of articles that she creates ($c_i; M_2 = 4; \lambda_2 = 0.5$). Each individual is interested in $L \leq K$ topics. We determine L using a normal distribution $N(\mu, \sigma)$, where $\mu = \frac{K}{2}$ and where $\sigma = \frac{\mu-1}{2}$. The preference vector $\boldsymbol{\beta}_i$ for each individual i over the K topics is defined as $\beta_{i,k}$ follows the following bimodal distribution:

$$\beta_{i,k} \sim \frac{L}{K}\mathcal{N}(\mu_1, \sigma_1^2) + \frac{K-L}{K}\mathcal{N}(\mu_2, \sigma_2^2) \tag{5.1}$$

Note that the RHS is independent from k, that means for all topic k, the interest $\beta_{i,k}$ is drawn from the same distribution. We choose large μ_1 and small μ_2 to implement that

[1] `http://repast.sourceforge.net`

$\beta_{i,k}$ is high in case i is interested in topic k and $\beta_{i,k}$ is low otherwise. RHS distribution with large μ_1 and small μ_2 also convey that overall i is interested in L out of K topics. In the simulation we choose $\mu_1 = 0.8, \mu_2 = 0.2, \sigma_1 = \sigma_2 = 0.4$

The constant γ_c that modulates the article creation probability is set to 0.2. The number of dimensions D of the immutable attributes $\boldsymbol{a}_i$ is set to 5.We set the universal constants $\gamma_f, \gamma_s, \alpha$ as follows. We set $\gamma_f = 1.0$, $\gamma_s = 0.5$, $\alpha = 0.5$. These three constants control the probability that an individual i reads an article posted by a person in her network neighborhood. We set $\gamma_a = 1.0$; the parameter controls the probability than an individual will act on an article after reading it.

Let us now discuss the parameters relevant to the system signaling model. We set the update parameter $\Delta = 20$; that is, the system 're-wires' or updates each individual's information neighborhood every 20 time steps. The system biases towards friends $\delta_f = 1.0$ and towards strangers $\delta_s = 0.5$ (ref. Equation 4.9) are set so as to bias the appropriateness values towards friends. The information neighborhood for each individual is fixed to be 10. The number of negative influencers is the same as the size of the positive influencers and fixed to be 5 each. Table5.1 summarizes list of parameters and their values used in our experiments .

5.2 Simulating the initial phase without social signals

We run the experiments for three types of initial network topologies (Preferential Attachment Network, Small World Network, and Random Network). The network topologies were generated using Jung library [2] as follows, using `BarabasiAlbertGenerator` to generate Preferential Attachment Network, using `KleinbergSmallWorldGenerator` ($p = 0.2$) to generate Small World Network, and using `ErdosRenyiGenerator` ($p = 0.03$) to generate a Random Network.

[2] `http://jung.sourceforge.net`

Parameter	Value
λ_1	0.5
λ_2	0.5
M_1	4
M_2	1, 4
a_i^k	U(0,1)
γ_f	1
γ_s	0.5
δ_f	1
δ_s	0.5
α	0.5
γ_a	1
γ_c	0.2

Table 5.1: Parameters and their values in the experiments

In the initial phase, we construct a network of information neighborhoods, without any social signals with the purpose of creating networks with individual activities prior to introduction of social signals. In this phase, each user only receives activity notifications from her closest neighbors who are identified based on mutual awareness. The mutual awareness between two individuals $a_{i,j}$ is initialized to 1 if they are network neighbors, and to 0 otherwise. Afterward, if i comments on j's article, then $a_{i,j}$ is incremented by 1 and $a_{j,i}$ is incremented by 0.5. For any individual i, we pick the top 5 neighbors with the highest mutual awareness. S_n is updated each iteration.

5.3 Experimental Results

We now present our experimental results. We conducted experiments for three types of networks (Preferential Attachment, Small World and Random) and comprehensive settings of other parameters.

To obtain each graph, we evaluated the network activity for 600 time steps, and we report the average of 500 simulations. In each case the initial set-up phase lasted for 100 time steps, and the social signals begin at the the 101^{st} time step. We measure the performance of the social signaling scheme through the KL divergence between the target distribution q and the current activity distribution $p(t)$, we also measure the change in source of information.

We varied the number of topics K to take on a small value ($K = 3$) and a large one ($K = 25$). In both cases, we assumed that the number of topics L in which an individual was interested varied as a normal distribution $N(\mu, \sigma)$, where $\mu = \frac{k}{2}$ and where $\sigma = \frac{\mu - 1}{2}$. We constrain L to lie between $0 \leq L \leq K$. Each individual is thus assigned L topics by uniformly sampling the set of K topics L times without replacement.

5.3.1 Small K, K=3

In the results that follow, we vary the target distribution $\boldsymbol{q}$. In particular, we vary the value of $\boldsymbol{q}_1$ to assume one of values in the set $\{0.1, 0.2, 0.3\}$; the value for $\boldsymbol{q}_2$ is fixed to be 0.3 and thus $\boldsymbol{q}_3$ is easily determined since $\boldsymbol{q}_3 = 1 - \boldsymbol{q}_1 - \boldsymbol{q}_2$. The starting distribution was established as $p(0) = (1/3, 1/3, 1/3)$. Notice that since each individual was assigned L topics by uniformly sampling the space of topics it is straightforward to achieve a uniform initial activity distribution

5.3.1.1 Different Initial Network Topologies

We run the experiments with different initial network topologies: Preferential Attachment, Small World, and Random.

Figure 5.1 shows the variation in the divergence between the activity distribution $\boldsymbol{p}(t)$ and the target distribution $\boldsymbol{q}$ as a function of time for three different network topologies: small world (solid line), power-law (dotted line, generated through preferential attachment) and random (dashed line). In all three cases, the initial information neighborhood is constructed by picking five network neighbors with highest mutual awareness. We introduce social signals at $t = 100$. The starting activity distribution is set $\boldsymbol{p}(0) = (1/3, 1/3, 1/3)$; the target activity distribution is $\boldsymbol{q} = (0.1, 0.3, 0.6)$. The value of the M_2 parameter, which controls the minimum rate of creation for each individual, is set to 1 and the step size γ is set to 0.1. The figure shows that there is slightly higher rate of convergence for power law networks in comparison to small world and random graph networks. The differences amongst the curves are minor, and we can show that all three will tend towards the same asymptote, since final distribution is affected only by the innate characteristics of article creation.

5.3.1.2 Different Values of Target Distribution

We run the experiments with different values of the target distribution q.

Figure 5.2 shows the results of the simulation with a small world network. It shows the variation of KL divergence between the activity distribution $\boldsymbol{p}(t)$ and the target distribution $\boldsymbol{q}$ as a function of time starting from an initial activity distribution of $\boldsymbol{p}(0) = (1/3, 1/3, 1/3)$. We construct the initial information neighborhood by picking five network neighbors with highest mutual awareness. We introduce social signals at $t = 100$. The initial distribution $\boldsymbol{p}(0) = (1/3, 1/3, 1/3)$, and there are three curves corresponding to three different target distributions q. The value of the M_2 parameter, which controls the minimum rate of creation for each individual, is set to 1 and the step size γ is set to 0.1

Notice that amongst the curves, the initial KL divergence is highest when the target distribution $\boldsymbol{q} = (0.1, 0.3, 0.6)$ is most skewed and all three curve tends to asymptotes. The asymptotes are in general different from $D = 0$ since not all target distributions can be reached. The reason is straightforward—the signaling schemes alter the responses to information posted on the network, but we conservatively assume that the social signals

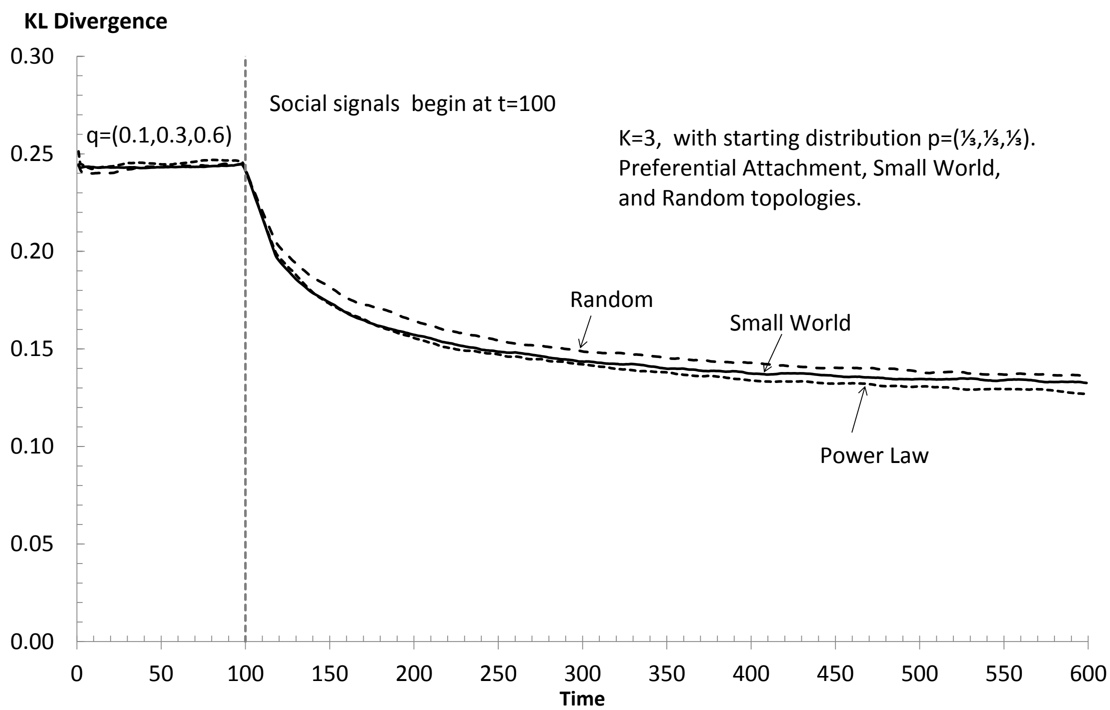

Figure 5.1: The figure shows the variation of KL divergence between the activity distribution $\boldsymbol{p}(t)$ and the target distribution $\boldsymbol{q}$ as a function of time. The results show the variation in the divergence for three different network topologies: small world (solid line), power-law (dotted line, generated through preferential attachment) and random (dashed line). In all three cases, the initial information neighborhood is constructed by picking five network neighbors with highest mutual awareness. We introduce social signals at $t = 100$. The starting activity distribution is set $\boldsymbol{p}(0) = (1/3, 1/3, 1/3)$; the target activity distribution is $\boldsymbol{q} = (0.1, 0.3, 0.6)$. The value of the M_2 parameter, which controls the minimum rate of creation for each individual, is set to 1 and the step size γ is set to 0.1. The figure shows that there is slightly higher rate of convergence for power law networks in comparison to small world and random graph networks.

do not alter the innate interests held by each individual. Thus, if the network is very active in topics different from what we target, it may become impossible to achieve those distributions. We shall examine reachability in more detail in Section 5.4.3.

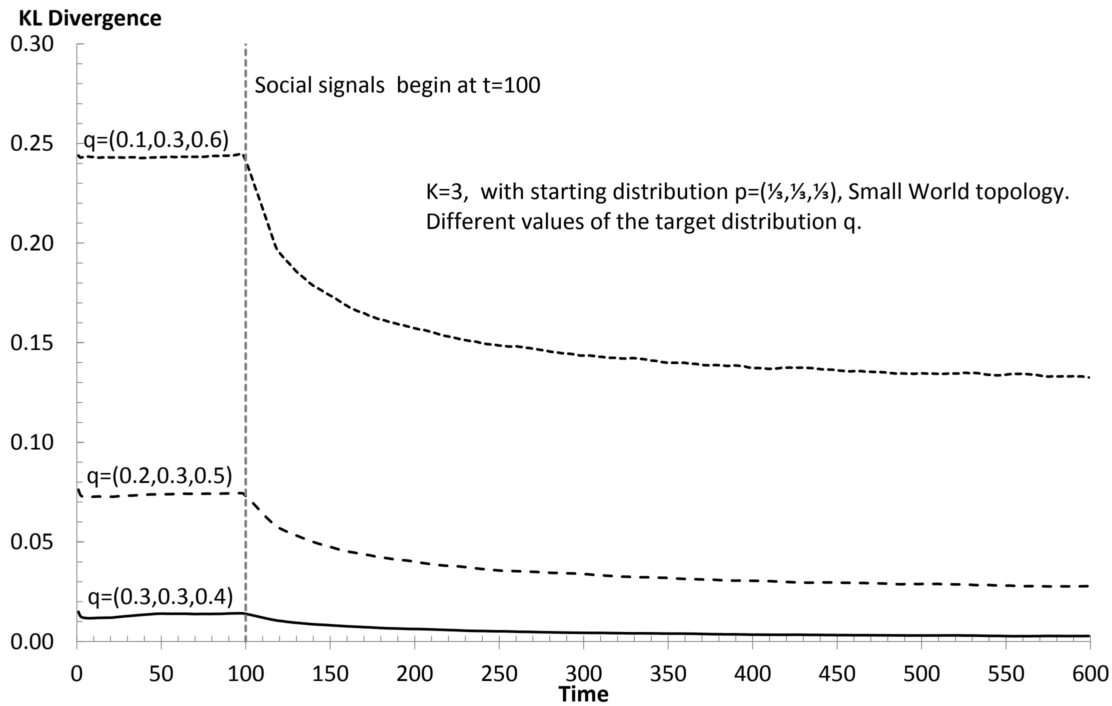

Figure 5.2: The figure shows the variation of KL divergence between the activity distribution $\boldsymbol{p}(t)$ and the target distribution $\boldsymbol{q}$ as a function of time. The results are for a Small World initial topology and the initial information neighborhood is constructed by picking five network neighbors with highest mutual awareness. We introduce social signals at $t = 100$. The initial distribution $\boldsymbol{p}(0) = (1/3, 1/3, 1/3)$, and there are three curves corresponding to three different target distributions q. The value of the M_2 parameter, which controls the minimum rate of creation for each individual, is set to 1 and the step size γ is set to 0.1. Three curves correspond to three values of the target distribution q, the initial KL divergence is highest when the target distribution $\boldsymbol{q} = (0.1, 0.3, 0.6)$.

5.3.1.3 Convergence Speed

In the previous sections (5.3.1.1,5.3.1.2), we showed KL divergence over the time for different network topologies, different target distributions, and we fix the user creation constraint parameter M_2 to 1, and step size γ in the gradient descent method to 0.01. In this section we show results for different values of M_2 and γ and show how those parameters affect convergence speed. In addition to measuring KL divergence, we also measure another output called $pChange$ which indicates how much change in the source of information in the network, more details about $pChange$ will be discussed in the section 3

1. Different Values of M_2

   Figure 5.3 shows the results of the simulation with a small world network. It shows the variation of KL divergence between the activity distribution $\boldsymbol{p}(t)$ and the target distribution $\boldsymbol{q}$ as a function of time starting from an initial activity distribution of $\boldsymbol{p}(0) = (1/3, 1/3, 1/3)$. We construct the initial information neighborhood by picking five network neighbors with highest mutual awareness. We introduce social signals at $t = 100$. We run the experiments with two target distributions $q = (0.1, 0.3, 0.6), q = (0.2, 0.3, 0.5)$ and two different settings for M_2, $M_2 = 1, M_2 = 4$, the step size γ is fixed to 0.1. We can see that in each group of the target distribution q, when parameter M_2 increases, the convergence to the target distribution is more difficult.

2. Different Values of γ

   Figure 5.4 shows the results of the simulation with a small world network. It shows the variation of KL divergence between the activity distribution $\boldsymbol{p}(t)$ and the target distribution $\boldsymbol{q}$ as a function of time starting from an initial activity distribution of

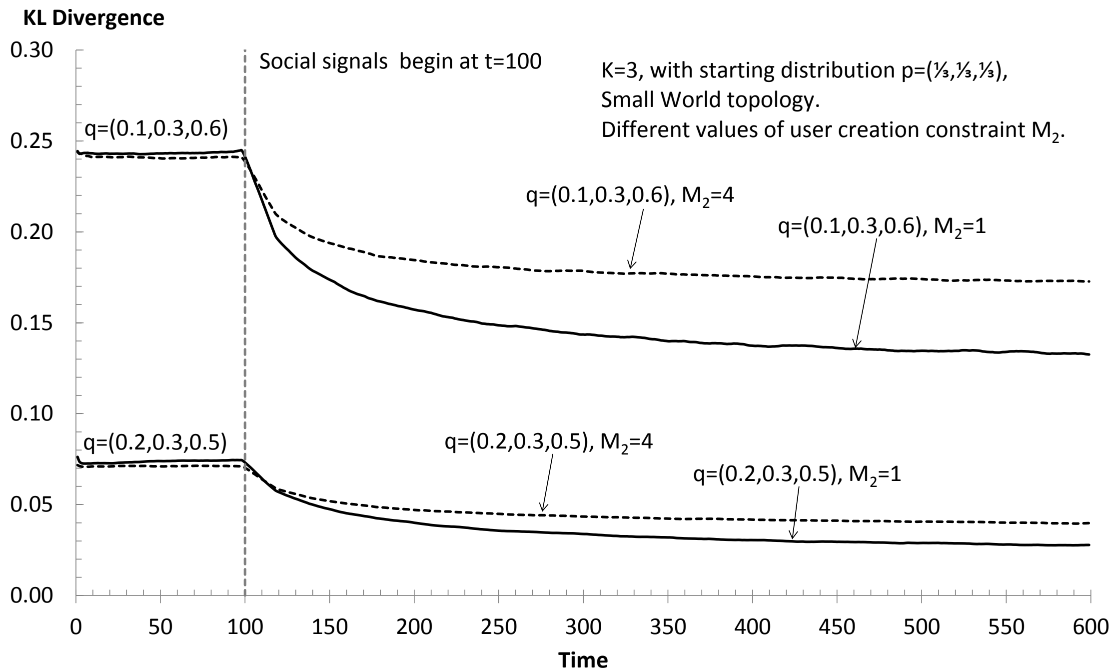

Figure 5.3: The figure shows the variation of KL divergence between the activity distribution $\boldsymbol{p}(t)$ and the target distribution $\boldsymbol{q}$ as a function of time. The results are for a Small World initial topology and the initial information neighborhood is constructed by picking five network neighbors with highest mutual awareness. We introduce social signals at $t = 100$. There are four curves organized into two groups. Each group refers to a different target distribution; in each case we use the same initial activity distribution of $\boldsymbol{p}(0) = (1/3, 1/3, 1/3)$ and the step size is fixed to 0.1. Within each group, the solid curve is for the case $M_2 = 1$ and the dashed curve is for the case $M_2 = 4$. A higher value of M_2 increases the activity of each individual. The two cases refer to two different parameters that affect the number of articles created by each individual. See Section 3.1 for more details. Within each group, we see that when parameter M_2 increases, thereby increasing the activity of each individual, the convergence to the target distribution is more difficult.

$\boldsymbol{p}(0) = (1/3, 1/3, 1/3)$. We construct the initial information neighborhood by picking five network neighbors with highest mutual awareness. We introduce social signals at $t = 100$. We run the experiments with two target distributions $q = (0.1, 0.3, 0.6)$, $q = (0.2, 0.3, 0.5)$ and two different settings for γ, $\gamma = 0.05, \gamma = 0.1$, the user creation constraint M_2 is fixed to 1. We see can see that in each group of the target distribution q, when γ increases, the convergence to target distribution is faster. However, to ensure the convergence the step size γ must satisfy the Wolfe conditions, that bounds the condition on how large γ can be.

3. Changing in the Source of Information

   In the section 1, 2, we showed that different values of user creation constraint M_2 and step size γ can affect the convergence speed. In this section we will show the how much change in the source of information for each user which leads to the change in KL divergence. To qualify that, we measure $pChange$ as follow.

   Remember that at each time stamp t, user i only receive social signals from her source of information $S_i(t)$, we measure $p_i(t)$ as:

$$p_i(t) = \begin{cases} 1 & \text{if } S_i(t) \neq S_i(t-1) \\ 0 & \text{if } S_i(t) = S_i(t-1) \end{cases} \tag{5.2}$$

   Basically $p(i)_t$ indicates whether source of information $S(i)$ has changed from time stamp $t-1$ to t. Overall, we average that change over all users.

$$pChange(t) = \frac{\sum_{i=1}^{N} p_i(t)}{N} \tag{5.3}$$

   Figure 5.5 shows the variation of $pChange$ over time. We construct the initial information neighborhood by picking five network neighbors with highest mutual

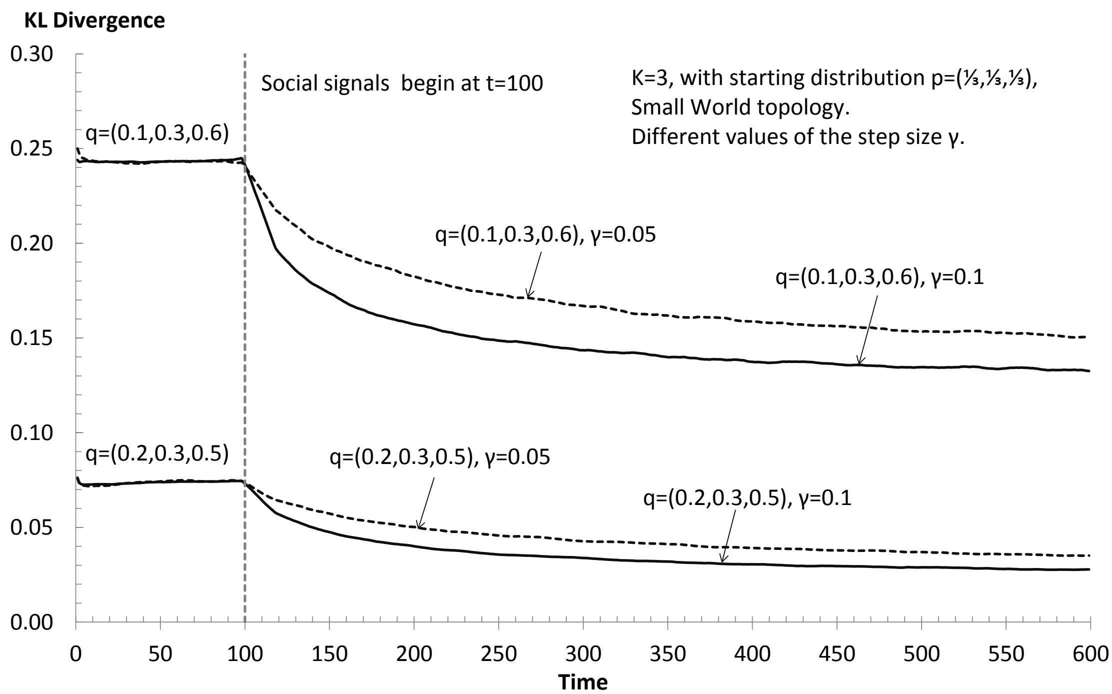

Figure 5.4: The figure shows the variation of KL divergence between the activity distribution $\boldsymbol{p}(t)$ and the target distribution $\boldsymbol{q}$ as a function of time. The results are for a Small World initial topology and the initial information neighborhood is constructed by picking five network neighbors with highest mutual awareness. We introduce social signals at $t = 100$. There are four curves organized into two groups. Each group refers to a different target distribution; in each case we use the same initial activity distribution of $\boldsymbol{p}(0) = (1/3, 1/3, 1/3)$ and the user creation constraint M_2 is fixed to 1. Within each group, the solid curve is for the case $\gamma = 0.1$ and the dashed curve is for the case $\gamma = 0.05$. We see can see that in each group of the target distribution q, when γ increases, the convergence to target distribution is faster.

awareness. We introduce social signals at $t = 100$. We run the simulations with Small World initial network topology, initial distribution $\boldsymbol{p}(0) = (1/3, 1/3, 1/3)$, and target distribution $q = (0.1, 0.3, 0.6)$. The default setting is $M_2 = 1, \gamma = 0.1$ and then we vary $M_2 = 4$ and $\gamma = 0.05$. Notice that before the social signals begin, in all cases $pChange$ keeps decreasing because users start interactions and forming groups, the connections within groups become stronger over time and users have tendency to interact within their group. Therefore, their source of information (social circle) becomes more stable over time and as the result $pChange$ keeps decreasing. Notice that we introduce social signals at $t = 100$ and the system re-wires each individual's source of information every $\Delta = 20$ time steps, therefore $pChange > 0$ in every step that the system re-wires the source of information for each user, other than that there is no change in the network and $pChange = 0$.

Compare two cases, $M_1 = 1, \gamma = 0.1$(square line) and $M_1 = 4, \gamma = 0.1$ (triangle line) we can see that in case users have higher creation rate ($M_2 = 4$) the system has to do more work in re-wiring the source of information for users (higher $pChange$), however the convergence speed still lower than in the case M_1, refer to the figure 5.3.

Compare two cases, $M_1 = 1, \gamma = 0.1$(square line) and $M_1 = 1, \gamma = 0.05$ (diamond line) we can see that if user creation constraint M_2 is fixed then if system does more work in re-wiring the source of information for users ($\gamma = 0.1$) the convergence will happen faster, refer to the figure 5.4.

5.3.2 Large K, K=25

We also run experiments in case number of topics K is large, $K = 25$. In this case, we start with the equal initial distribution $p(0) = (1/25, ..., 1/25)$ and the target distribution $q = (q_0, q_1, .., q_{24})$ where $q_i = 0.02$ with $i = 0, .., 9$, $q_i = 0.04$ with $i = 10, .., 14$, and

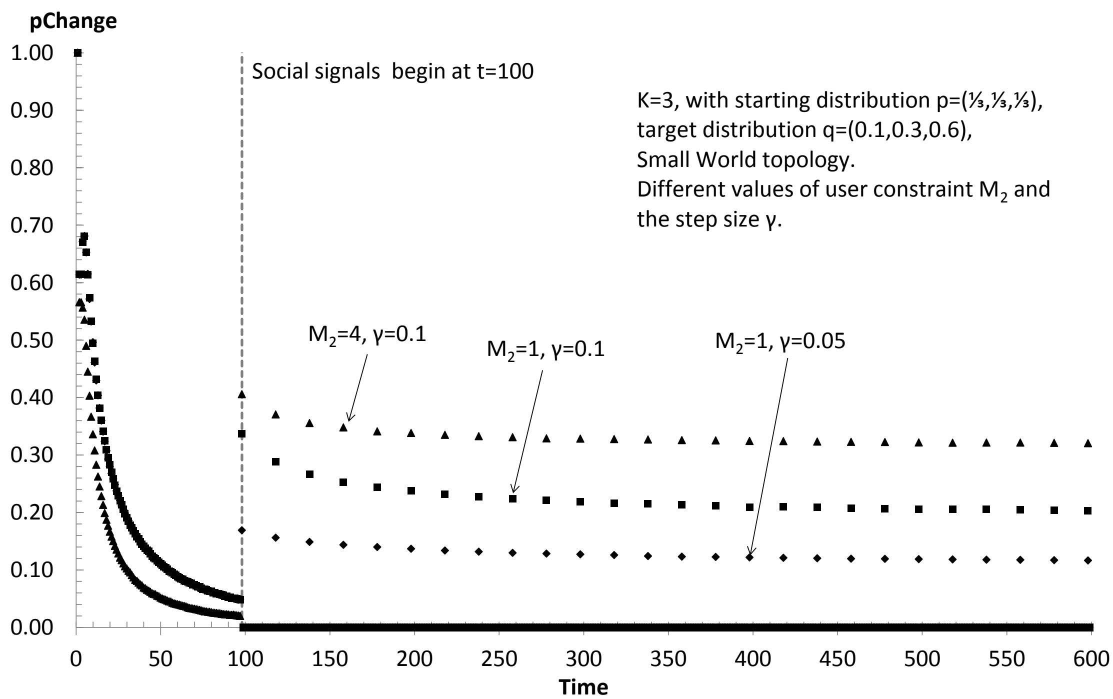

Figure 5.5: The figure shows the variation of $pChange$ over time. We construct the initial information neighborhood by picking five network neighbors with highest mutual awareness. We introduce social signals at $t = 100$. We run the simulations with Small World initial network topology, initial distribution $\boldsymbol{p}(0) = (1/3, 1/3, 1/3)$, and target distribution $q = (0.1, 0.3, 0.6)$. Notice that before the social signals start $pChange$ keeps decreasing in all cases, after social signals start $pChange > 0$ once the system re-wires social signals, and $pChange = 0$ otherwise. Compare three cases we can see that the system has to do more work in re-wiring social signals in cases of higher M_2 and higher γ

$q_i = 0.06$ with $i = 15,..,24$. As in the case of $K = 3$ we also run the simulation with three different network topologies Preferential Attachment (PA), Small World (SW), and Random (RD).

Figure 5.6 shows the variation in the divergence between the activity distribution $\boldsymbol{p}(t)$ and the target distribution $\boldsymbol{q}$ as a function of time for three different network topologies: small world (solid line), power-law (dotted line, generated through preferential attachment) and random (dashed line). In all three cases, the initial information neighborhood is constructed by picking five network neighbors with highest mutual awareness. We introduce social signals at $t = 100$. We can see that after the social signals start, KL divergence keeps decreasing, that means the distribution $p(t)$ goes toward the target distribution q.

5.4 Analysis

In this section we will analyse the convergent property of KL divergence between the two distributions $p(t)$ and q. We discuss both experimental results and mathematical analysis.

5.4.1 KL Divergence and Reachable Area

Figures 5.1,5.2,5.6 show that $D(p(t)||q)$ decreases over time, that is as expected as in gradient descent method. But the question is if there is a lower bound for $D(p(t)||q)$? Ideally gradient descent method grants a convergence, that is.

$$\lim_{t \to +\infty} p(t) = q$$

or

$$\lim_{t \to +\infty} D(p(t)||q) = 0$$

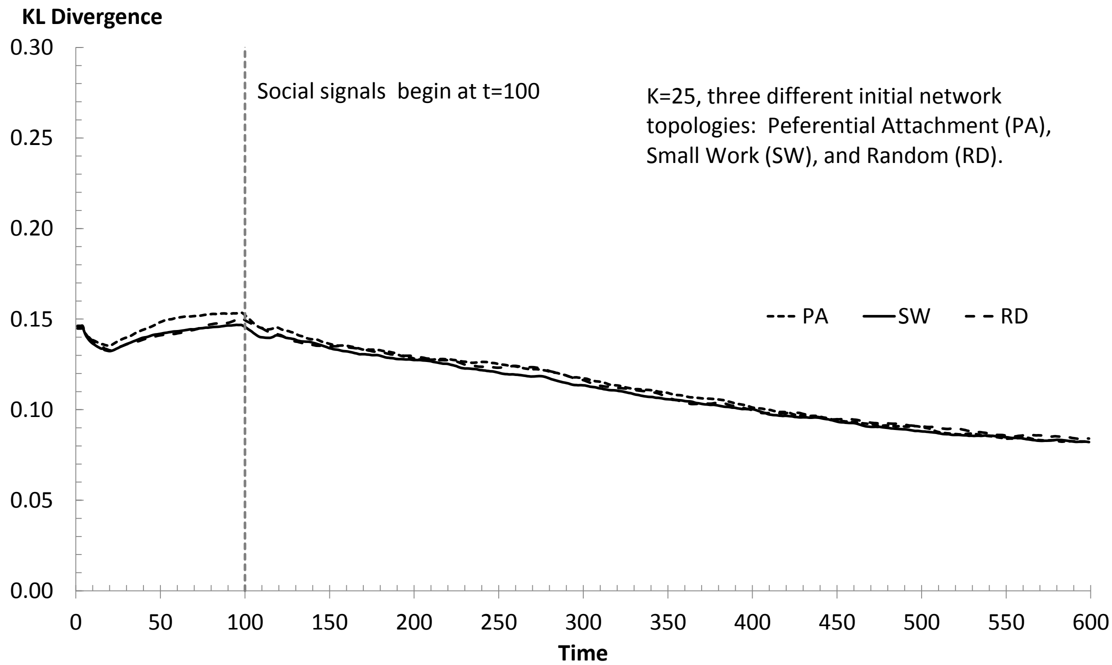

Figure 5.6: The figure shows the variation of KL divergence between the activity distribution $\boldsymbol{p}(t)$ and the target distribution $\boldsymbol{q}$ as a function of time in case $K = 25$. The results show the variation in the divergence for three different network topologies: small world (solid line), power-law (dotted line, generated through preferential attachment) and random (dashed line). In all three cases, the initial information neighborhood is constructed by picking five network neighbors with highest mutual awareness. We introduce social signals at $t = 100$. After social signals starts,KL divergence keeps decreasing, that means the distribution $p(t)$ goes toward the target distribution q.

However in the model that is not always achievable. For a better intuition we consider a simpler case when $K = 2$, each user is interested in equally random L out of K topics, ($L = 0$, 1, or 2), therefore $p(0) = (0.5, 0.5)$. In this case we definitely can go to $q = (0.49, 0.51)$ but because there are always users who are interested in topic 1, they will always have article creations on topic 1 therefore we are definitely unable to go to the extreme distribution $q' = (0, 1)$. That intuition gives us an idea that there is a reachable area Q of q such that:

$$\lim_{t \to +\infty} D(p(t) \| q) = 0 \text{ if } q \in Q$$

and

$$\lim_{t \to +\infty} D(p(t) \| q) = \epsilon_q \text{ if } q \notin Q$$

In the next two sections we will estimate activity bound using experimental method and mathematical method in case $K = 3$.

5.4.2 Experimental Estimation

In this section we try to estimate Q and estimate ϵ_q in case $q \notin Q$.

- Estimation of Q

  Assume that $q = (q_1, q_2, q_3) \in Q$. First we estimate the range of q_1, $Min \leq q_1 \leq Max$. To estimate Min we run the simulation in the extreme case $q = (0, 0.5, 0.5)$ and to estimate Max we run the simulation in the extreme case $q = (1, 0, 0)$. For those two extreme distributions, $D(p(t) \| q)$ gets stable at some value >0. Basing on those system stable states, we obtain an approximation $Min = 0.1$ and $Max = 0.7$. That means, the necessary condition for $q \in Q$ is $0.1 < q_1, q_2, q_3 < 0.7$. We

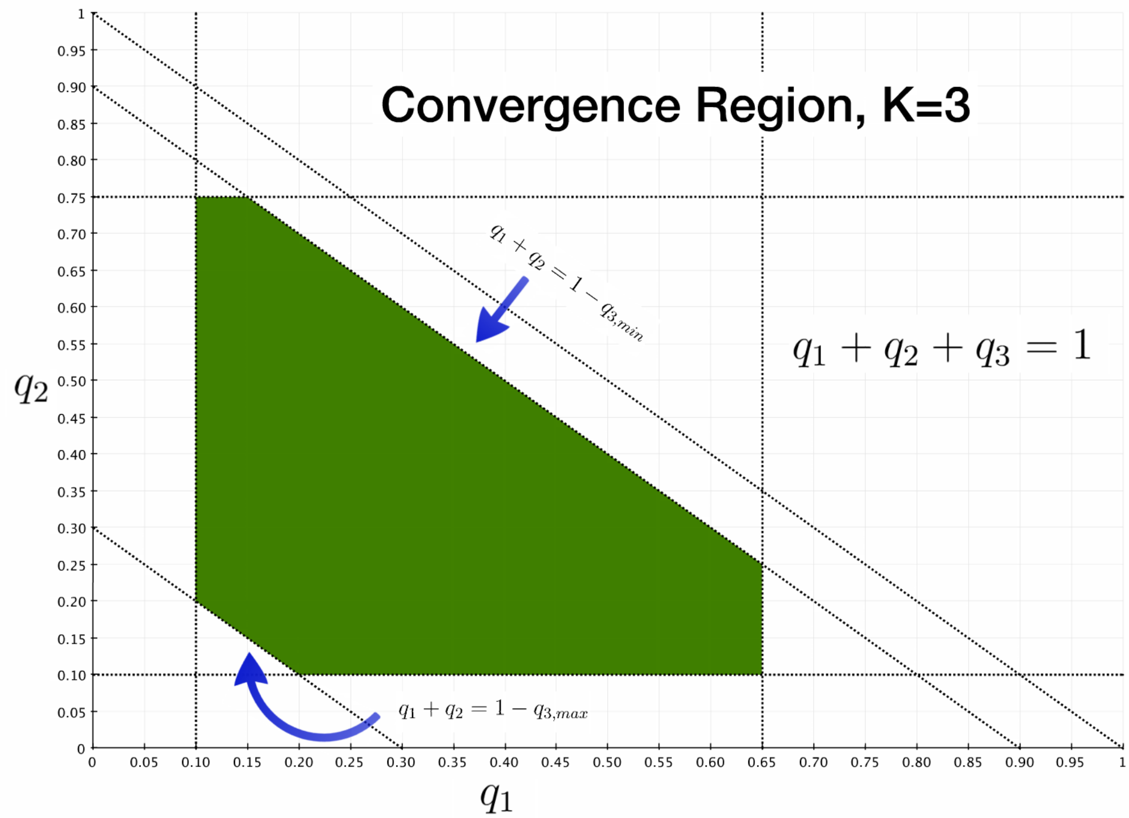

Figure 5.7: Reachable Area Q (Convergence Region) which is about 45% of the entire domain of q. IISS framework can guide the distribution $p(t)$ as close as possible toward q in case $q \in Q$. However if $q \notin Q$, even though IISS framework is able to nudge $p(t)$ toward q, there is always a lower bound ϵ_q that limits how close that $p(t)$ can approach q

conjecture that is also the sufficient condition for $q \in Q$. Since $q_1 + q_2 + q_3 = 1$, figure 5.7 visualizes the conjectured Q when being projected in (q_1, q_2) plane.

- Estimation of ϵ_q in case $q \notin Q$

  To do so we run the simulation with extensive rages of q and run longer time. In previous section we only run the simulation for 600 iterations, in this section we run the simulation for 2000 iterations. Figure 5.8 visualizes the divergence $D(p(t)||q)$ for different values of q after 2000 iterations. If $q \in Q$ the corresponding divergence is much smaller in compared with the divergence in case $q \notin Q$.

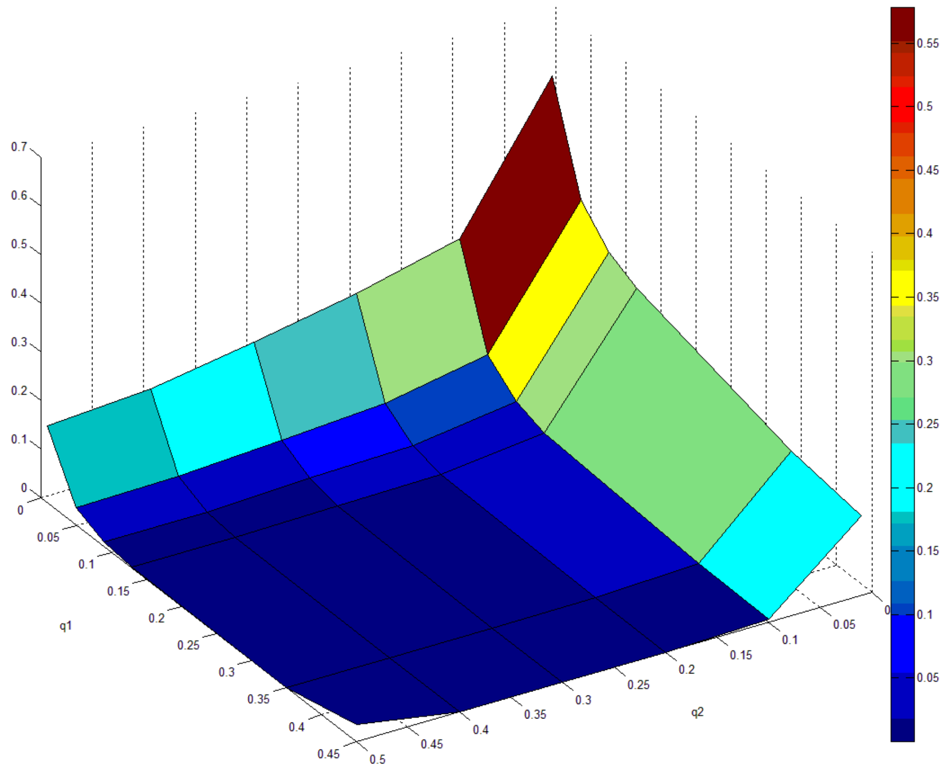

Figure 5.8: Divergences after 2000 iterations for different values of q. In case $q \in Q$, the divergence can be small as 0.05 while in case $q \notin Q$, the divergence can be as large as 0.5.

5.4.3 Mathematical Estimation

We now discuss the issue of target activity distributions in a little more detail. In particular, we discuss the upper and lower bounds on an activity and show that not all target activity distributions $\boldsymbol{q}$ are achievable by altering the social signals.

Let us examine the case when there are exactly three topics (i.e. $k = 3$) in the network. Let us assume that there are $N = 100$ people in the network, and where the parameters for all the model constants are the same as those described in Section 5.1. We set the parameter that controls the lower bound of the creation activity as follows: $M_2 = 4$. Let $x^c_{i,k}$ be the average number of articles created on topic k, by individual i. And let

$x^a_{i,k}$ be the average number of *notifications* created on topic k, by individual i by though comments. One way to minimize activity on topic 1, that is, minimize the value of $\boldsymbol{q}_1$, is to re-wire the social signals to show notifications *only* from topics 2 and 3. Thus, everyone on the network would see lots of activity on topics 2 and 3 and will comment on them if they find the article of interest, thus further increasing the activity on these two topics. The total number of notifications on topic $\boldsymbol{q}_1$, denoted as x_1, including creations on the activity and the notifications due to comments on an article is computed as follows:

$$x_1 = \sum_{i=1}^{N} \left(x^c_{i,1} + x^a_{i,1} \right), \tag{5.4}$$

where, $x^c_{i,1}$ refers to the number of articles created on topic 1 by individual i, and where $x^a_{i,1}$ refers to the total number of comments on topic 1. Thus the value of $\boldsymbol{q}_1$ can be computed as follows:

$$\boldsymbol{q}_1 = \frac{\overbrace{x_1}^{minimize}}{x_1 + \underbrace{x_2 + x_3}_{maximize}}, \tag{5.5}$$

where, to to estimate the lower bound for $\boldsymbol{q}_1$ we need to minimize x_1 and maximize the sum of $x_2 + x_3$. Now expanding $\boldsymbol{q}_1$ using Equation 5.4:

$$\boldsymbol{q}_1 = \frac{\sum_{i=1}^{N} x^c_{i,1} + x^a_{i,1}}{\sum_{i=1}^{N} x^c_{i,1} + x^a_{i,1} + \sum_{i=1}^{N} \sum_{k=2}^{3} x^c_{i,k} + x^a_{i,k}}. \tag{5.6}$$

Now to minimize x_1, we need to ensure that activity on topic 1 is limited to creation. That is, $x^a_{i,1} = 0, \forall i$. We can achieve this state by ensuring that there are no social signals related to activity on topic 1. That is, $x^a_{i,1} = 0, \forall i$. Rewriting Equation 5.6, we have:

$$\boldsymbol{q}_{1,\min} = \min_{x^a_{i,k}} \frac{\sum_{i=1}^{N} x^c_{i,1}}{\sum_{i=1}^{N} x^c_{i,1} + \sum_{i=1}^{N} \sum_{k=2}^{3} x^c_{i,k} + x^a_{i,k}}. \tag{5.7}$$

Now, replacing summations over the population with expectations over all the individuals:

$$\boldsymbol{q}_{1,\min} = \min_{x^a_{i,k}} \frac{E[x^c_{i,1}]}{E[x^c_{i,1}] + \sum_{k=2}^{3} E[x^c_{i,k} + x^a_{i,k}]} \tag{5.8}$$

where, E is the expectation operator. While $E[x^c_{i,k}]$, the average number of creations on topic k, is straightforward to calculate since the topics are equally distribution over the population, computing the expectation of $E[x^a_{i,k}]$, the expected number of comments on topic k requires more work.

$$E_i[x^c_{i,k}] = E[c_i] \cdot \gamma_c \cdot E[\beta_{i,k}], \tag{5.9}$$

$$= (M_2 + \frac{1}{\lambda_2}) \cdot \gamma_c \cdot \frac{1}{K}. \tag{5.10}$$

since c_i, the creative capacity, is exponentially distributed with parameter λ_2 and has a lower bound of M_2, and $\beta_{i,k}$ is identically distributed with a uniform prior over all k. The equation says that on average, all topics are equally likely to be created, with the number of articles created proportional to the creative capacity.

To compute $E[x^a_{i,k}]$, we need to estimate the number of notifications that any individual sees and the likelihood of commenting on them. Only the act of commenting produces measurable activity.

$$E_i[x^a_{i,k}] = E[z_i] \cdot E[P(topic = k)] \cdot E[P_i(action = T|topic = k)] \tag{5.11}$$

$$= (M_1 + \frac{1}{\lambda_1}) \cdot \gamma_a \cdot E[\beta_{i,k}] \cdot E[P(topic = k)] \times$$

$$\times E\left[\sum_{j=1}^{N} P_i(view = T|origin = j) \cdot P(origin = j)\right] \tag{5.12}$$

$$= (M_1 + \frac{1}{\lambda_1}) \cdot \gamma_a \cdot \frac{1}{K} \cdot E[P(topic = k)] \times$$

$$\times \sum_{j=1}^{N} E\left[P_i(view = T|origin = j) \cdot P(origin = j)\right] \tag{5.13}$$

If we assume that the social signal can come from anywhere in the networks, then $P(origin = j) = 1/N$. Thus, equation 5.13 simplifies to:

$$E_i(x^a_{i,k}) = (M_1 + \frac{1}{\lambda_1}) \cdot \gamma_a \cdot \frac{1}{K} \cdot E[P(topic = k)] \cdot E[P_i(view = T|origin = j)] \tag{5.14}$$

In case we minimize value of q_1, we re-wire the social signals to only show notifications from topic 2 and 3. Therefore, $P(topic = k) = 0$ in case $k = 1$ and $P(topic = k) = 0.5$ in case $k = 2, 3$. Note that $P_i(view = T|origin = j)$ is defined in the equation 3.4. From equations 5.10, and 5.14, we can simplify equation 5.8 as:

$$\boldsymbol{q}_{1,\min} = \frac{(M_2 + \frac{1}{\lambda_2}) \cdot \gamma_c \cdot \frac{1}{K}}{(M_2 + \frac{1}{\lambda_2}) \cdot \gamma_c + (M_1 + \frac{1}{\lambda_1}) \cdot \gamma_a \cdot \frac{1}{K} \cdot E[P_i(view = T|origin = j)]} \tag{5.15}$$

In order to maximize q_1, we re-wire the social signals to only show notifications from topic 1. With the similar logic as above discussed, we can estimate $q_{1,\max}$ in an equation which is similar to the equation 5.7

$$\boldsymbol{q}_{1,\max} = \max x^a_{i,k} \frac{\sum_{i=1}^{N} x^c_{i,1} + x^a_{i,1}}{\sum_{i=1}^{N} x^c_{i,1} + x^a_{i,1} + \sum_{i=1}^{N} \sum_{k=2}^{3} x^c_{i,k}}. \tag{5.16}$$

Finally we have:

$$\boldsymbol{q}_{1,\max} = \frac{(M_2 + \frac{1}{\lambda_2}) \cdot \gamma_c \cdot \frac{1}{K} + (M_1 + \frac{1}{\lambda_1}) \cdot \gamma_a \cdot \frac{1}{K} \cdot E[P_i(view = T|origin = j)]}{(M_2 + \frac{1}{\lambda_2}) \cdot \gamma_c + (M_1 + \frac{1}{\lambda_1}) \cdot \gamma_a \cdot \frac{1}{K} \cdot E[P_i(view = T|origin = j)]} \tag{5.17}$$

Equations 5.15, 5.17 show mathematical analysis for the range of q_1, since users' interests are equally distributed over K topics, therefore those equations are also applicable to the cases q_2 and q_3.

Chapter 6

CONCLUSION AND FUTURE WORK

In this chapter we first summarize the work has been done and presented in this thesis, and then we discuss some ideas for future extension to our work.

In this thesis we solve a very interesting problem of guiding the activities on a social network platform so that the activity distribution can gradually go from the initial value $p(0)$ toward the target value q. Our research deeply roots from previous research on social influence and selection which evidence that humanity is influenced by social signals. In order to guide users's activities, using similar logic with existing online advertising and recommendation systems, first and foremost our framework has to understand them by learning their interests, characteristics and connections (4.1,4.2). After that we use gradient descent method to update the distribution (4.3), in this case since KL divergence is a convex function so gradient descent guarantees a convergence. With understanding of users, we then select suitable users to influence (4.4) and personalize social signals for each user (4.5). Both selecting users and designing social signals are very challenging because we are guiding a system in which users act probabilistically based on their interests and interconnections that results in a complex stochastic process. Experimental results in chapter 5 show that our framework is able to nudge the distribution toward the target distribution q in all cases of initial network topologies (Preferential Attachment, Small World, Random) and for all values of target q. The convergence speed and property (of KL divergence between the distribution p and distribution q) depend on lots of factors. If users have more tendency to post articles independently with social signals presented to them (larger value of M_2), then the system has to do more work in re-wiring the network and the convergence speed is still lower. If system uses larger step size γ in the gradient descent method, it will do more work in re-wiring the network and is able to achieve faster speed of convergence, however γ cannot be arbitrarily large since it is bounded by Wolfe conditions to ensure convergence. There is a reachable range Q in which for all values of q the framework is able to guide the activity distribution $p(t)$ toward q as close as we want, however if $q \notin Q$, then even though the framework is able to guide $p(t)$ toward q, there is always positive lower abound on how close that $p(t)$ can go to q.

In this thesis we made two assumptions, the first assumption is the interest $\beta_{i,k}$ of user i on topic k is fixed during the experimental process, and the second assumption is all users participate on the platform at all time stamps. If we consider a short period of some special campaign like an online advertising campaign, those assumptions are reasonable, however during a longer span of time, those assumptions become very rigid. That suggests some ideas for future extensions. First, we can consider that user interests are changeable over time, that will make the problem more interesting and more challenging. Second, we can consider that users has different participation patterns, user i may participate on the platform every Δ_i iterations instead of participating on the platform every iteration, again this change will make the problem more challenging.

REFERENCES

[1] Asim Ansari, Skander Essegaier, and Rajeev Kohli. Internet recommendation systems. *Journal of Marketing research*, 37(3):363–375, 2000.

[2] Lars Backstrom, Eytan Bakshy, Jon M Kleinberg, Thomas M Lento, and Itamar Rosenn. Center of attention: How facebook users allocate attention across friends. *ICWSM*, 11:23, 2011.

[3] Eytan Bakshy, Dean Eckles, Rong Yan, and Itamar Rosenn. Social influence in social advertising: evidence from field experiments. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, pages 146–161, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1415-2. doi: 10.1145/2229012.2229027. URL `http://doi.acm.org/10.1145/2229012.2229027`.

[4] Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*, pages 519–528. ACM, 2012.

[5] Howard Beales. The value of behavioral targeting. *Network Advertising Initiative*, 2010.

[6] James Bennett and Stan Lanning. The netflix prize. In *Proceedings of KDD cup and workshop*, volume 2007, page 35, 2007.

[7] Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298, September 2012. ISSN 0028-0836. doi: 10.1038/nature11421. URL `http://dx.doi.org/10.1038/nature11421`.

[8] Òscar Celma and Pedro Cano. From hits to niches?: or how popular artists can bias music recommendation and discovery. In *Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition*, page 5. ACM, 2008.

[9] Ye Chen, Dmitry Pavlov, and John F Canny. Large-scale behavioral targeting. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 209–218. ACM, 2009.

[10] Robert B Cialdini. Influence: The psychology of persuasion. *New York, Morrow*, 1993.

[11] Robert B. Cialdini. *Influence: science and practice*. Allyn & Bacon, 2000.

[12] Robert B Cialdini and Melanie R Trost. Social influence: Social norms, conformity and compliance. 1998.

[13] Thomas M Cover and Joy A Thomas. *Elements of information theory*. Wiley-interscience, 2012.

[14] David Crandall, Dan Cosley, Daniel Huttenlocher, Jon Kleinberg, and Siddharth Suri. Feedback effects between similarity and social influence in online communities. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 160–168. ACM, 2008.

[15] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, et al. The youtube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 293–296. ACM, 2010.

[16] R. I. M. Dunbar. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, 16:681–694, 11 1993. ISSN 1469-1825. doi: 10.1017/S0140525X00032325.

[17] R. I. M. Dunbar. Cognitive constraints on the structure and dynamics of social networks. *Group Dynamics: Theory, Research, and Practice*, 12(1):7 – 16, 2008. ISSN 1089-2699.

[18] Mohan J Dutta-Bergman. The demographic and psychographic antecedents of attitude toward advertising. *Journal of Advertising Research*, 46(1):102–112, 2006.

[19] David Easley and Jon Kleinberg. Networks, crowds, and markets. *Cambridge Univ Press*, 6(1):6–1, 2010.

[20] Avi Goldfarb. What is different about online advertising? *Review of Industrial Organization*, pages 1–15, 2013.

[21] Avi Goldfarb and Catherine Tucker. Online advertising. *Advances in Computers*, 81:289–315, 2011.

[22] Noah J. Goldstein, Robert B. Cialdini, and Vladas Griskevicius. A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35(3):472 – 482, 2008. ISSN 00935301. URL `http://offcampus.lib.washington.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=ufh&AN=34412781&site=ehost-live`.

[23] Mark Granovetter. The strength of weak ties. *American journal of sociology*, 78(6): l, 1973.

[24] Jinwen Guo, Shengliang Xu, Shenghua Bao, and Yong Yu. Tapping on the potential of q&a community by recommending answer providers. In *Proceedings of the 17th ACM conference on Information and knowledge management*, CIKM '08, pages 921–930, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-991-3. doi: 10.1145/1458082.1458204. URL `http://doi.acm.org/10.1145/1458082.1458204`.

[25] Shawndra Hill, Foster Provost, and Chris Volinsky. Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science*, pages 256–276, 2006.

[26] Nathan Oken Hodas and Kristina Lerman. How visibility and divided attention constrain social contagion. 05 2012. URL `http://arxiv.org/abs/1205.2736`.

[27] Nathan Oken Hodas and Kristina Lerman. How visibility and divided attention constrain social contagion. In *Proceedings of the 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust*, SOCIALCOM-PASSAT '12, pages 249–257, Washington, DC, USA, 2012. IEEE Computer Society. ISBN 978-0-7695-4848-7. doi: 10.1109/SocialCom-PASSAT.2012.129. URL `http://dx.doi.org/10.1109/SocialCom-PASSAT.2012.129`.

[28] Nathan Oken Hodas and Kristina Lerman. How visibility and divided attention constrain social contagion. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 249–257. IEEE, 2012.

[29] Petter Holme and Mark EJ Newman. Nonequilibrium phase transition in the coevolution of networks and opinions. *Physical Review E*, 74(5):056108, 2006.

[30] Pamela Miles Homer. Relationships among ad-induced affect, beliefs, and attitudes: Another look. *Journal of Advertising*, 35(1):35–51, 2006.

[31] Damon Horowitz and Sepandar D. Kamvar. The anatomy of a large-scale social search engine. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 431–440, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-799-8. doi: 10.1145/1772690.1772735. URL `http://doi.acm.org/10.1145/1772690.1772735`.

[32] Yutaka Kabutoya, Tomoharu Iwata, Hisako Shiohara, and Ko Fujimura. Effective question recommendation based on multiple features for question answering communities, 2010. URL `http://aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1471`.

[33] Denise B Kandel. Homophily, selection, and socialization in adolescent friendships. *American journal of Sociology*, pages 427–436, 1978.

[34] Louise Kelly, Gayle Kerr, and Judy Drennan. Avoidance of advertising in social networking sites: The teenage perspective. *Journal of Interactive Advertising*, 10(2): 16–27, 2010.

[35] David Krackhardt. The strength of strong ties: The importance of philos in organizations. *Networks and organizations: Structure, form, and action*, 216:239, 1992.

[36] Greg Linden, Brent Smith, and Jeremy York. Amazon. com recommendations: Item-to-item collaborative filtering. *Internet Computing, IEEE*, 7(1):76–80, 2003.

[37] Foster Provost, Brian Dalessandro, Rod Hook, Xiaohan Zhang, and Alan Murray. Audience selection for on-line brand advertising: privacy-friendly social network targeting. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '09, pages 707–716, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-495-9. doi: 10.1145/1557019.1557098. URL `http://doi.acm.org/10.1145/1557019.1557098`.

[38] Mingcheng Qu, Guang Qiu, Xiaofei He, Cheng Zhang, Hao Wu, Jiajun Bu, and Chun Chen. Probabilistic question recommendation for question answering communities. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 1229–1230, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-487-4. doi: 10.1145/1526709.1526942. URL `http://doi.acm.org/10.1145/1526709.1526942`.

[39] Everett M Rogers. *Diffusion of Innovations, 5th Edition*. Free Press, 5th edition, August 2003. ISBN 0743222091. URL `http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0743222091`.

[40] M. J. Salganik, P. S. Dodds, and D. J. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854, 2006.

[41] P Wesley Schultz, Jessica M Nolan, Robert B Cialdini, Noah J Goldstein, and Vladas Griskevicius. The constructive, destructive, and reconstructive power of social norms. *Psychological science*, 18(5):429–434, 2007.

[42] Catherine Tucker. Social advertising. *SSRN eLibrary*, 2012.

[43] Hao Wang and Alvin Chin. Social influence on being a pay user in freemium-based social networks. In *Advanced Information Networking and Applications (AINA), 2011 IEEE International Conference on*, pages 526–533. IEEE, 2011.