Impact of Opinions in Social Networks

GARG, Priyanka

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Master of Philosophy
in
Computer Science and Engineering

The Chinese University of Hong Kong
July 2012

Thesis/Assessment Committee

Prof. LEE Pak Ching (Chair)

Prof. LYU Rung Tsong Michael (Thesis Supervisor)

Prof. KING Kuo Chin Irwin (Thesis Supervisor)

Prof. FU Wai Chee Ada (Committee Member)

Prof. WONG Raymond Hau-san (External Examiner)

Abstract of thesis entitled:
Impact of Opinions in Social Networks
Submitted by GARG, Priyanka
for the degree of Master of Philosophy
at The Chinese University of Hong Kong in July 2012

Social opinions play a crucial role in shaping both our purchase decisions and our experience. While on one hand, we are encouraged (discouraged) to adopt a product upon hearing the positive (negative) opinions; on the other hand, our opinions tend to conform to our social circle. Both of these aspects of social opinions are important in order to make precise product recommendations, to accurately predict the information flow pathways and to launch efficient viral marketing campaigns.

In this thesis, we first study the impact of polarity of opinions on our purchase decisions. For the same, we analyze the information propagation patterns of the negative and positive opinions on two real world social networks, Flixster and Epinions, and observe that the presence of negative opinions significantly reduces the number of expressed opinions. To account for the asymmetry between the two kinds of opinions, we propose extensions of the two most popular information propagation models, Independent Cascade and Linear Threshold models. The proposed extensions give a tractable influence problem and improve the prediction accuracy of future opinions, by more than 3%.

Next, we study the impact of social opinions on our expressed opinions about the products. The hypothesis is that many times our expressed opinions are not completely independent of our social circle and gets calibrated such that they are similar to the social opinions. In order to understand this phenomenon, we propose a novel formulation for the users ratings where every expressed rating is considered as a function of the social opinion along with the user preference and item characteristics. The proposed method helps in improving the prediction accuracy of users' rating by more than 2% in presence of social influence. Additionally, the learned model parameters reveal the degree of conformity of users. Detailed analysis of user social conformity show that more than 76% of users tend to conform to their friends to some extent. On an average, user ratings become more positive in presence of the social influence. We also find that the social conformers are usually not the first one to participate in an information cascade.

學位論文摘要

學位論文題目：社會網絡中的意見影響
提交人：GARG, Priyanka
學位：哲學碩士
香港中文大學，二零一二年

社會意見在塑造我們購買決策和購買經歷發揮了至關重要的作用。除了正面（或負面）的意見會鼓勵（或打消）我們購買某個產品，我們的意見更傾向於遵循我們的社交圈。社會意見的這些方面對於做出精確產品推薦、準確預測信息流向、及有效營銷活動發布極為重要。

在這篇論文中，我們首先研究極性意見對我們的購買決策的影響同時，我們分析了兩個現實世界中的社會網絡，Flixster 和 Epinions 中的消極和積極的意見的信息傳播模式。我們觀察到，否定意見的存在大大降低了表達意見的數量。考慮到這兩種意見的不對稱性，我們提出並擴展了目前最流行的兩個信息傳播模式獨立分級和線性閾值模型。我們提出的拓展模型提供了一個可處理的影響問題和並能夠提高將來意見的預測精度，超過 3％。

更進一步，我們研究了社會意見對我們表達產品意見的影響。該問題的假設是多次顯示我們表達的意見並不完全獨立於我們的社交圈，而是通過校準，使之跟社會意見相似。為了理解這一現

象，我們為用戶的評分提出了一個新型的模型。該模型中，用戶對項目的評分是由社會輿論、用戶的偏好和項目特點的一個函數該模型可以提高用戶評分的預測準確率達 2%。此外，模型中學習到的參數可展示用戶對社會意見的遵循程度。用戶對社會意見的遵循分析表明，超過 76％的用戶傾向於在一定程度上遵循他們好友的意見。平均而言，當社會影響存在的時候，用戶評分更趨於正面。我們還發現，社會的遵循者通常不是信息傳播的第一次參與者。

Acknowledgement

I would like to express my sincere thanks and gratitude to Prof. Irwin King and Prof. Michael R. Lyu for their invaluable guidance. They have been a constant source of encouragement, enthusiasm and support. I find myself highly indebted to Prof. King for providing me with freedom, critical reviews and the invaluable amount of insight and knowledge. The whole experience has been a novel one and will be cherished by me forever.

I am grateful to Haiqin for answering my endless questions and providing with critical reviews and suggestions at various stages of the work. I would also like to thank Zachary and every member of our lab for many useful discussions. I have learned a lot from each one of them, particularly from our group meetings. Many thanks to Laoshi Guo, Tom, Zachary, Haiqin, Jessie, Connie, Ning, Jingying, Siyao, Qian for introducing me to one of the world's oldest culture, the Chinese culture. My two year stay in CU would not be as rewarding and enriching without them.

I am also thankful to my colleagues in Yahoo! Labs: S. Sundararajan, Subhajit, Krishnan, Sravanthi and Sarvesh. I am especially grateful to Sundararajan for guiding me through my first research steps, for advice and for long discussions about life, universe and everything. He really set my interest in machine learning and set the course for where I am today.

I would also like to take this opportunity to thank my friends and family for their love and encouragement. I am thankful to my husband, Aakash, for his understanding and helping me settle in Hong Kong. Many thanks to my friends Amal and Rajat for lending me a patient ear and for their constant support, without which I could have never finished this journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Social networks are the graphs of individuals and relationships between them, where a relationship could be friend, colleague, follower, family, etc. The interactions in these networks, play a fundamental role in spreading information, ideas and technologies among their members. Often the decision to adapt a product is influenced by one's social connections. Such effects have been observed in many cases, when an idea or action gains sudden widespread popularity through word-of-mouth or "viral marketing" [22] effects. To name an example from recent past, Googles Gmail achieved wide usage largely through referrals, rather than direct advertising.

Study of social networks is not new and has a long history in social sciences. They have helped in identifying patterns in social networks like Milgram's famous 6-degrees of separation[68], Dunbar's threshold on the size of the well functioning community [23] etc. However, these studies were usually restricted to small sized network and the information flow was studied for one product or technology. Recently the popularity of online social networks has opened the possibility of studying the large scale social networks. Courtesy to the online networks, now we know that the social networks have small diameter, and most of the individuals have few connections but few nodes have unusually high number of connections. In fact, we know the precise shape of the degree distribution which is power-law [14]. We are also convinced that the large scale networks have a complex structure with no clear hierarchy and they are most likely structured in a core-periphery (or jelly-fish) kind of structure [61]. Researchers are also been able to propose generators which can generate networks with such properties [56].

However, peer influence in social networks is still a mystery. We know that the probability of a user to join a group or buy a product increases as a function of number of friends, and stronger ties play more important role in adoption of product or technology. But quantifying the social influence is not always an easy task. In most of the social networks, it is hard to directly observe the path of information flow; the only observable traces are the actions taken by individuals. In other words, we only know that a user has taken an action (or bought a product) but we do not know, who has influenced his/her decision. In case, multiple friends have affected the decision, then how to quantify the influence of each of his/her friend?

Recent research works have made progress in defining models to mimic the flow of information because of social influence and need for social conformity. These models have been employed for both inferring the pair-wise influence from the action logs, and for identifying the seed nodes for the viral marketing campaign. However, the efforts are required, not only to develop accurate models and efficient algorithms to solve these problems, but also to answer the following.

- How can we differentiate the social influence from homophily or external factors?
- How does the influence of a user varies with time?
- Does the influence depends on the topic?
- How does the polarity of opinion affect the product adop-

tion?

- Can social opinions also affect the user posterior evaluation of the product, i.e., the opinion the user form after experiencing the product?

These are some of the daunting questions and in this thesis, we set ourself with the goal to answer the latter two questions. We study the impact of social opinions in shaping both our purchase decisions and our experience. While on one hand, we are encouraged (discouraged) to adopt a product upon hearing the positive (negative) opinions; on the other hand, our opinions tend to conform to our social circle. Both of the studies are vital, not just from the point of curiosity, but are also vital in making precise product recommendations, accurately predicting the information flow pathways and launching efficient viral marketing campaigns.

1.1 Contributions

The main contributions of this thesis include:

- **Impact of Polarity of Opinions on Products' Adoption Probability [31]**
  - Taking examples of two real world social rating networks Flixster and Epinions, we observe that presence of negative opinions reduces the product adoption probability from 10% to 7% for Flixster and from 6% to 1% for Epinions dataset.
  - Motivated by above observation, we propose a generalize information propagation model which explicitly models the social influence as a latent state variable, while the observed state of the users are defined in terms of this latent state and the quality of product.

  - To find the social influence, we propose polarity sensitive extensions of state of art social influence functions. The usefulness of the proposed model and polarity sensitive influence functions, are demonstrated by predicting the future users' opinions using them. The prediction accuracy improves by more than 3% on Flixster and 5% on Epinions datasets.

- **Impact of Social Opinions on Posterior Evaluation of Products [30]**
  - We propose a novel formulation for user ratings that explicitly considers the users' social conformity as model parameters. The proposed formulation improves the predict accuracy of users' ratings by more than 2% in presence of social influence on Goodreads.
  - The learned social conformity parameters are also verified by qualitatively comparing the discovered most influential users with the authoritative and most socially active users.
  - Based on the learned users' degree of conformity, we find various interesting patterns on Goodreads that underline the impact of social conformity. To our surprise, the results indicate that approximately 76% socially active users tend to conform to their friends to some degree. We also find that social opinions make the user ratings more positive than negative.

1.2 Organization

The report is organized as follows:

- **Chapter 2**

This chapter provides a background on social networks and detailed survey of research progress made in developing the information flow models, in estimation of social influence and in maximizing the social influence. We also list the key observations from the empirical studies of the real social networks.

- **Chapter 3**

  This chapter studies the impact of polarity of social opinions on product's adoption probability. We study the information propagation patterns on real world dataset and based on the observations, we develop a generic framework that explicitly accounts for the asymmetry between the positive and negative opinions. The proposed framework is throughly studied on both synthetic and real world datasets.

- **Chapter 4**

  We propose a novel rating formulation that explicitly models the role of social conformity on users' product ratings. The usefulness of the proposed formulation is demonstrated by assessing its recommendation accuracy on a large scale dataset. The learned parameters are then used, to explore the nature of social conformity.

- **Chapter 5**

  The last chapter summarizes the thesis and provides potential future research directions.

☐ **End of chapter.**

Chapter 2

Background & Survey

In this chapter, we review the research progress made in understanding both the social network structure and the information flow in social networks. For both, we first present their characteristic properties and then present the models to generate them.

2.1 Network Structure

Social network is represented with a graph $G = (V, E)$ where every node $v \in V$ corresponds to an individual in the social network and an edge $(u, v) \in E$ represents the relationship between nodes u and v. The size of vertex set V is represented by n. In this section, we will first present the key structural properties of social networks and then we will briefly review the models that can generate networks with such properties.

2.1.1 Basic Definitions

Here we define the terminology related to network structures which we will be using throughout the thesis.

Directed and undirected graph: A graph is undirected if

$(u, v) \in E \implies (v, u) \in E$, i.e., edges are unordered pairs of nodes. If pairs of nodes are ordered, i.e., edges have direction, then the graph is directed.

Subgraph: A subgraph $G_s = (V_s, E_s)$ of a graph $G = (V, E)$ is a graph whose vertex set $V_s \subset V$ and edges among them E_s are subset of E.

Node degree: A node has degree d if it has d incident edges. For directed graphs we define out-degree, which is the number of edges pointing out from the node. Similarly, in-degree denotes the number of edges pointing to the node. For undirected graphs, out-degree is same as in-degree and is equal to the number of edges the node participates in.

Network Diameter: A graph has diameter o if every pair of nodes can be connected by a path of length at most o.

Triad: A triad or a triangle is a triplet of connected nodes (u, v, w) such that $(u, v), (v, w), (w, u) \in E$.

Connected component: A connected component is the maximal set of nodes such that there exists a path connecting every pair of the nodes in the set. For directed graph, one can define the weakly connected components and strongly connected components depending on the nature of connectivity between the nodes. The component is weakly connected if there exists an undirected path connecting every pair of nodes. While it is strongly connected is there exists a directed path between every pair of nodes.

2.1.2 Structural Properties of Social Networks

Research over the past few years has identified a number of properties that can be found in many real-world networks from various domains. While many patterns have been discovered, the principal ones are following.

1. **Degree distributions.** The degree distribution of many real world networks such as phone call graphs [71], citation graph [77], Epinions social graph [14], have been found to follow the power law distribution. If the number of nodes of degree d is denoted by n_{d}, then according to power law

$$n_{\mathsf{d}} \propto \mathsf{d}^{-\gamma} \text{ s.t. } \gamma > 1, \qquad (2.1)$$

where γ is called the power law degree exponent. That is, few nodes in the network have exceptionally high node degree while there are a large number of nodes with small degree. Networks with power law degree distributions, are also often refereed as scale-free networks. The name follows from the scale invariance property of power laws. The scale-free property implies that given a relation $f(x) = ax^{\gamma}$, scaling the argument x by a constant factor c causes only a proportionate scaling of the function itself. That is,

$$f(cx) = a(cx)^{\gamma} = c^{\gamma} f(x) \propto f(x). \qquad (2.2)$$

This behavior is what produces a linear relationship when logarithms are taken of both $f(x)$ and x. That is why, most commonly γ is estimated by fitting a straight line on the log-log axis and setting γ as the slope of the line. Unfortunately this method shows some bias because it violates the independence and Gaussian noise assumption of least squares linear regression [73]. One can get better estimate, by fitting a straight line on a log-log plot of the cumulative conditional distribution function (CCDF) which is more robust to the fluctuations due to the finite sample size [19]. Recently the Maximum Likelihood Estimates [19] are used for getting an unbiased estimation of γ. Based on the observed sample, γ is estimated as:

$$\hat{\gamma} = 1 + k\Big(\sum_{i=1}^{k} \log \frac{x_i}{x_{\min}}\Big)^{-1}, \tag{2.3}$$

where x_i are observed values of x such that $x_i \geq x_{\min}\ \forall i$ and k is the number of observations. For most real world datasets, the degree exponent γ typically takes values $2 < \gamma < 3$.

2. **Small diameter.** Most real-world networks have very small diameter. This observation is also known by the "small-world" or "six degrees of separation" phrases. To measure the diameter of any network, we do not use the shortest distance between every pairs of nodes because this measure is very sensitive to outliers. More robust measures such as *effective diameter* [86] are used. The integer effective diameter is the smallest number of hops at which at least 90% of all connected pairs of nodes can be reached.

   Calculating the effective diameter is very expensive for large networks at it takes $O(n^3)$ time. One possible way is sampling, i.e., sample pair of nodes in the network and calculate the length of the shortest paths between them. Another possible approach is by using an approximation algorithm ANF [75] that is based on fast approximate counting and hashing.

   The effective diameter is found to be small for many real-world social graph. The effective diameter was found to 6.6 for MSN [57] and ≈4 in Twitter [52]. Milgram [68] found that when subjects of his experiments are asked to route their letters to a lawyer in Boston only via their first-name acquaintance, then they were able to find paths with average length 6. That is, real social networks not only exhibit small diameter but they are also "navigable", where uses are able to find paths just based on their local information.

3. **Edge locality.** Most of the edges in social networks are local, i.e., it is highly likely that friend of a friend is also my friend. This transitivity between the relations (edges) is measured in terms of the *clustering coefficient*. The clustering coefficient c_v of a node v with degree d, is defined as the fraction of triads centered at node v among the $\mathsf{d}(\mathsf{d}-1)/2$ triangles that could possibly exist. The clustering coefficient of the global network c is then defined as the average c_v over all nodes v and the clustering coefficient c_d is defined as the average of c_v over all nodes v with degree d.

   The cluster coefficient c of the real social network is found to be significantly higher than for random networks [76]. Further, it has been observed that that c_d scales as power law, i.e., $\mathsf{c}_\mathsf{d} \propto \mathsf{d}^{-1}$. This observation has been used as an indication of the existence of hierarchical network structure [76]. The idea is that the low-degree nodes belong to very dense sub-graphs and those sub-graphs are connected to each other through hubs, which have high degree.

4. **Community Structure.** Social networks are naturally composed of large number of overlapping communities, where a community is a set of nodes that have more or closer connections among its members than to nodes outside the community. Mathematically, the goodness of any community $V_s \subset V$ is measured in terms of *conductance*. The conductance is defined as the fraction of number of edges between the nodes in V_s to the number of edges whose one endpoint is in V_s and other endpoint is in $V - V_s$.

Many work suggest the existence of a hierarchical structure [74] in communities, where bigger communities can be recursively split into smaller and smaller communities. Contrary to this, the existence of *core-periphery* structure has been suggested recently. Core-periphery structures, also go by the name of jellyfish or octopus structures, are composed of a large and densely linked core and a periphery. The nodes in the periphery are not connected among themselves and point towards the core. Thus, it is easier to cut periphery nodes from the rest of the network. While the core is very densely connected and is very hard to cut. Existence of such recursive core-periphery structure has been observed over many social networks like Flickr, Livejournal [61].

Further, it has been observed that the size of good communities is approximately 100. As the communities start to grow bigger, they start to gradually "blend in" with the rest of the network. This limit of community size is approximately same as Dunbar number [23] which is equal to 150 and indicates the upper limit on the size of a well-functioning human community.

Apart from above patterns, two key patterns have recently emerged as graphs evolve over time.

5. **Network densification.** Most of the networks evolve over the time as the nodes join and leave the network. It has been found that, many real networks such as paper citation and patent citation networks become *denser* as they evolve over the time [58]. That is, as a social network grows, the ratio of number of edges to number of nodes increases. Further, the densification has been found of follow the power law. At any time t, if the number of nodes are denoted by $N(t)$ and the number of edges in the network by $E(t)$, then

according to densification power law, $E(t) \propto N(t)^{\alpha}$. The densification exponent α has been observed to be greater than 1 for several real networks. This implies the real networks tend to have many more edges than nodes, and thus densify as they grow.

6. **Shrinking diameter.** It has been found that the effective diameter of many real world graphs tend to shrink (or/and stabilize) as they grow over the time [58]. That is, as the time passes the nodes become closer to each other in space. This property is not the direct result of densification but is caused by the the way the degree sequence evolves over the time.

In addition to above, several other properties have been found in social networks. For example, the real-networks have been found to be *resilient* [2] under the random node attacks. That is network's connectivity remains almost unaffected if one randomly removes nodes from the network. It has also been found that the *scree plot* obeys the power law distribution [27].

2.1.3 Network Generators

In parallel with empirical studies of large networks, there has been considerable work done to develop models for graph generation. The main motivation behind developing these graph generators is to gain better understanding of the networks, to measure the graph similarity and to be able to generate the synthetic data which can be used later for developing and testing new algorithms. Following are one of the most popular graph generators.

Random Graph Model. The earliest probabilistic generative model for graphs is a random graph model [25]. The model works by creating an edge between every pair of nodes with equal prior probability. Though this model has been used to develop a number of mathematical theories, it does not produce properties observed in the real networks. For example, the degree distribution in the random graph follows the Binomial distribution instead of the power law. Its diameter increases as the number of nodes increases in the network, which is contrary to the shrinking diameter property of social networks.

Small World Model. This family of models generate networks with small diameter and with local structures. In the small-world model [88], nodes are present in a ring structure. The local edges are created by adding edges between two hop away neighbors. To create the long range edges, some of the local edges are chosen with a fixed probability p, and their endpoints are changed such that they point to some randomly selected node in the network. Thus, when $p = 0$, this model generates very regular structures and when $p = 1$, it is equivalent to random network. That is, as p increases, both clustering coefficient and diameter of the network become smaller.

Related to the small-world models are the "navigable" network models [47]. These models strive to create networks which respect the navigable property of social networks. The navigable model assumes that nodes are present in a 2-dimensional lattice structure. Additional links are then created based on the distance between the two nodes. If the distance between the nodes u and v is $\text{dist}(u, v)$ then probability of creating an edge between them is set proportional to $\text{dist}(u, v)^{-\beta}$. For $\beta = 2$, the expected path length in the network $O(\log(n))$.

Hierarchical based network based models have also been proposed to achieve the same effect [46]. In these models, nodes are considered to reside in a hierarchy for example hierarchy based on persons' profession. Then, long range links between two nodes u and v are created based on the height of their least common ancestor. These networks also generate paths with expected length $O(\log(n))$. This hierarchical model has also been extended to case where every node belongs to multiple hierarchies e.g. profession, location [87]. In this model, the letters are routed via neighbors who are closest in any of the hierarchies.

Preferential Attachment. The discovery of degree power laws have led to the development of random graph models that exhibited such degree distributions, including the family of models based on *preferential attachment* [5]. According to these models, nodes join the network one after the other and when a node joins the network, it creates fixed number of the edges in the network. However, the edges are not created uniformly at random, but are created "preferentially". The probability that a newly arriving node creates an edge with an already existing node v is proportional to the degree of node v. This preferential attachment is also referred to as "rich getting richer" phenomenon or "cumulative advantage". This model generates the networks with power law degree distribution.

Several variants of preferential attachment have also been proposed to incorporate the node fitness [26], geography information [28] etc. The *fitness* of a node is its intrinsic ability to gather links in the network. The most fit node is able to attract more edges in comparison to the nodes with smaller fitness value. Fitness is modeled by using a fixed parameter per node and probability of new edge to any node is set proportional to its degree and its fitness parameter value. Geography information is also used to improve the models, where the intuition is that the probability of linking to node v is higher, if the node v is geographically closer. Here, a node preferentially creates edges to those nodes that belong to its local neighborhood.

Copying Model. Another set of models which generate the power law degree distribution are the copying model [48] and forest fire model [59]. According to the copying model, nodes arrive one at a time. When a new node joins the network, it randomly selects a node v and then either copies all its edges or creates random edges. The edges of v are copied by creating edges to all the neighbors of v. If β represents the probability of copying the edges then this model generates power law degree distribution with exponent $\gamma = 1/(1-\beta)$.

However, the diameter of the networks generated by the copying model increases as the graph evolve. This observation has lead to the development of *forest fire* model [59]. According to forest fire model, when a node joins the network, it first randomly selects a node v. Then, it randomly selects edges of v with a fixed probability and copy or *burn* them. The edges going out of the end points of the burnt edges are burned in the next step. The process continues until no more edges are selected for burning. Thus, burning of edges starts at node v and spreads like a fire in the forest.

Affiliation Networks. The underlying idea behind the *affiliation network* model [53] is that in social networks, there are two types of entities - individuals and societies. Individuals belongs to multiple societies by affiliation such as a particular football club, same location. This relation between the individuals and societies can be structurally viewed as a bipartite graph. The social network can then be generated by creating relations between the nodes with common affiliations. If the bipartite graph is generated as the scale-free bipartite graph then the generated social network has power law degree distribution. Further, it preserves local structure and respects the evolving graph properties such as shrinking diameter and densification.

Self-Replicating Networks. The key idea behind the self-replication models is to recursively produce networks with self-similarity. One such type of the model is the *Kronecker network* [56] which uses the Kronecker product recursively to replicate the self-similar structure. The basic structure is represented by a matrix $K1$. A 2-dimensional matrix $K1 = [\mathfrak{a}\ \mathfrak{b}; \mathfrak{c}\ \mathfrak{d}]$ represents the edge density within and in-between the two components. $\mathfrak{a}$ and $\mathfrak{d}$ represent density of edges within the components 1 and 2 respectively. While $\mathfrak{b}$ and $\mathfrak{c}$ represent the density of edges in between the two components. If $\mathfrak{a}$ is high and $\mathfrak{d}$ is very small, then the recursive Kronecker product of $K1$ produces networks with core-periphery structure. Further, it displays most of the structural properties of the real social networks such as power law degree distribution, small diameter, densification and shrinking diameter. Since this model is a parametric model with $K1$ as its parameter, it can be fit over any real network data to generate the network signature.

The properties of above generators are summarized in Table 2.1. For further reading on the graph laws and generators, we refer readers to a detailed survey in [13].

2.2 Information Diffusion in Social Networks

There are many situations when people's decision making process is influenced by other's behavior and decisions, for example, their product purchase decisions, their political opinions, their interests, their usage of a particular technology. This social influence sometimes gives rise to a network wide diffusion of information. Thus, the study of social influence is important from two point of views. One view is to understand the guiding principles behind individual's decision making process. Another view is the micro level view and look at the population-wide

Models	**Degree Distribution**	**Small Diameter**	**Edge Locality**	**Densification**	**Shrinking Diameter**	**Work as a process**
Random Graph	Binomial	No	No	No	No	Yes
Small World Model	Dirac delta function	Yes	Yes	No	No	Yes
Preferential Attachment	Power Law	Yes	Yes	No	No	Yes
Copying Model	Power Law	Yes	Yes	No	No	Yes
Forest Fire	Power Law	Yes	Yes	Yes	Yes	Yes
Affiliation Network	Power Law	Yes	Yes	Yes	Yes	Yes
Kronecker Network	Power Law	Yes	Yes	Yes	Yes	No

Table 2.1: Properties of social network generators

effect produced by every individual actions.

Even though, the information diffusion has been studied for many years by sociologists, most of the empirical studies are carried out on small sized datasets and usually in restrictive environments. However, with the advent of internet age, a number of online social networks have become popular recently and thus a large volume of data related to information cascades have been made available. Most of the online social networking sites work by letting their users to perform some actions or express their opinions/ideas. These actions or opinions or ideas are then made visible to users' friends/peers in the network. Twitter, Flixster, Goodreads, Facebook are few examples of popular online social networks. Twitter is a micro-blogging site where users can express ideas/information in 140 text characters. On Flixster and Goodreads users can rate movies and books online and share their reviews with their friends. While Facebook is a general social networking site where users can share pictures, videos, message, blogs with their friends.

The main interest here is to under some of basic principles which govern individual's decision making process and then leverage them to achieve or observe some of the aggregate effect in the network, for example, to launch effective viral marketing campaign and to accurately detect a new event. Next, we introduce basic terminologies related to information diffusion in social networks and then list the key understanding of social influence among the researchers. The models developed for information diffusion based on these patterns, are discussed next in Section 2.2.3. The related topic of estimation of social influence in the network is then taken in Section 2.2.4. One of the main application of social influence analysis, viral marketing is then discussed in depth in Section 2.2.5. It is important that we do not confuse the social influence from homophily and this is a topic of discussion in Section 2.2.6.

2.2.1 Basic Terminologies

Here we define some of the basic terminologies that will be used throughout this thesis.

Social Influence. The term social influence refers to the phenomenon where individual adopts a product because of the actions taken by others. Social influence is known by different terms like *conformity*, *peer pressure* etc. In general individuals' decisions can be classified as *rational* or *irrational* decisions. However, many times it is not easy to tell them apart.

Information Cascade. Information cascade of a product, is the process where the product is adopted by the individuals in the network because of social influence.

Early adopter. An early adopter is an user who adopts the product in early stages of its information cascade. This person is also referred to as a *trend-setter*.

Homophily. Homophily is the tendency of individuals to make friends or associate with people who are similar to them. This principle explains why our friends do not look like a random sample of population, but are similar to us in terms of age, race, geographic locations, interests. This behavior is often captured by "birds of a feather flock together" saying.

2.2.2 Principles governing the Decision-Making process

A large number of social studies have established the basic principles that have been used to model the decision making process of individuals in the social networks. Following are the key principles found by the *diffusion of innovation* work.

- **Informational Effect.** Many times individuals derive indirect information by observing the actions taken by others. This derived information then lead them to take the similar actions regardless of their prior opinion. This is known as informational effect. For example, consider a scenario where you have to chose to dine in a new town and based on your research, you find restaurant $\mathfrak{A}$ is better than $\mathfrak{B}$. However, when you reach there, you find that no one is eating in the restaurant $\mathfrak{A}$ and there is a long queue outside the restaurant $\mathfrak{B}$. Then, if you believe that other people also have similar taste as yours then you might decide to join the crowd and go to restaurant $\mathfrak{B}$. Under the informational effect principle, your decision is seen as a rational decision if everyone is assumed to have independent but imperfect information about the two restaurants [24].

- **Direct Benefit Effect.** Contrary to information effect is the direct benefit effect, where people get a direct payoff by copying others' actions. For example, consider a choice to join a social network. Your utility is directly depended on the number of friends who have already joined a particular network. Further, every time another friend joins the network, you get a direct payoff.

- **Bandwagon Effect.** According to the bandwagon effect, as the number of individuals who believe in something increases, others tend to disregards their own opinions and also "hop on the bandwagon" [32]. That is, tendency to follow others opinions is directly proportional to the number of individuals holding the same opinion. Both the direct-benefit and informational effect produce the bandwagon effect. In case of direct-benefit effect, as the number of friends adopting a product increases, you get a direct benefit from that product. While in case of information effect, your believe in the goodness of a product will increase if lots of people are adopting.

- **Strength of Ties.** The interpersonal ties between the individuals are generally classified as: strong ties and weak ties. The weak ties act as bridges between the tightly knitted communities and are responsible for spreading the information about a product across the network. On the other hands, the actions taken by strong ties (family and close friends) build more trust in a product and thus reduce the resistance or threshold of the user to take an action. Thus, weak ties help in spreading a word about the product but the probability of buying the product increases if your strong ties also buy the product.

  Brown et al. [10] interviewed the families of students being instructed by three piano teachers, in order to find out the network of referrals. They found that strong ties, those between family or friends, were more likely to be activated for information flow and were also more influential than weak ties between acquaintances.

2.2.3 Information Cascade Models

Based on the above principles, several information cascade models have been proposed to mimic the way information flows in the social networks. The different models with their key characteristics are show in Figure 2.1. Next we will describe each one of them in details.

Direct-benefit Effect Models. Networks based on the direct-benefit effect are usually modeled using the coordinated games

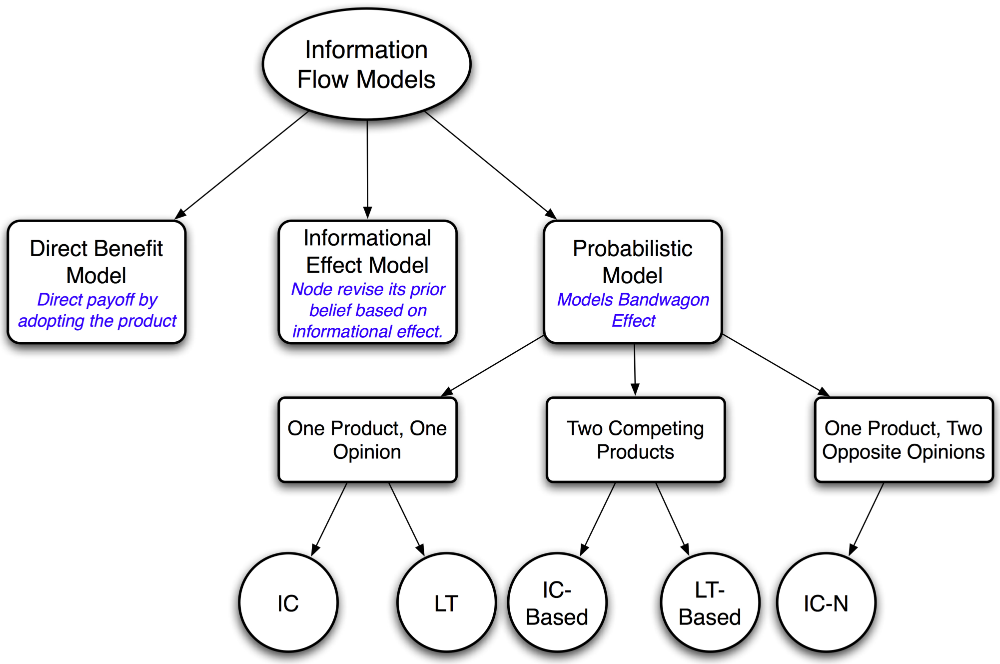


Figure 2.1: Information flow models

[69]. Each node in the network has the choice between two possible behaviors, $\mathfrak{A}$ and $\mathfrak{B}$. If two nodes u and v are linked in the network then they gain direct incentive by adopting the similar behavior. The payoffs are then defined as follows

		v	
		$\mathfrak{A}$	$\mathfrak{B}$
u	$\mathfrak{A}$	a,a	0,0
	$\mathfrak{B}$	0,0	b,b

where a and b are positive payoffs. Thus if p fraction of u's neighbors adopt $\mathfrak{A}$ and rest of them adopt $\mathfrak{B}$, then u will get a payoff of $p \cdot a$ by adopting behavior $\mathfrak{A}$ while a payoff of $(1-p) \cdot b$ by adopting $\mathfrak{B}$. Thus, $\mathfrak{A}$ will be a better choice if $p \geq a/(a+b)$. This basic model has been extended to account for the different payoffs for each users and to consider the user resistance to adopt any behavior. This resistance is modeled by setting a threshold parameter for each node. In such setting, a behavior is adopted only when p is greater than the node's threshold. The node threshold can be imagined to be small among the early adopters.

Informational Effect Models. These models consider a situation where nodes make decisions sequentially one after the other and these decisions can be seen by other nodes in the network [24]. Every node u has a prior belief about a behavior and as it observes the actions of others, it derives the information from it and revise its belief. It is assumed that the node u does not know prior belief of other nodes. In the example of selection between the two restaurants $\mathfrak{A}$ and $\mathfrak{B}$, the two choices can be seen as two opposing behaviors, where selection of one implies the rejection of the other. If the belief of the node u in the goodness of restaurant $\mathfrak{A}$ is denoted by probability p_u then the goodness of restaurant $\mathfrak{B}$ can be written as $(1-p_u)$. Let us assume that the probability of a node to visit a restaurant given it is good, is q s.t. $q > 1/2$. According to the information effect, when the node u observes that other node v has visited the restaurant $\mathfrak{B}$, then it revises its estimation based on its prior belief (recall that u does not know the prior belief of node v). According to the Bayes theorem,

$$\begin{aligned} p(\mathfrak{A} \text{ is good} | v \text{ choses } \mathfrak{B}) &= \frac{P(\mathfrak{A} \text{ is good}) p(v \text{ choses } \mathfrak{B} | \mathfrak{A} \text{ is good})}{p(v \text{ choses } \mathfrak{B})} \\ &= \frac{p_u(1-q)}{p_u(1-q) + (1-p_u)q} \\ &\leq \frac{p_u(1-q)}{p_u(1-q) + (1-p_u)(1-q)} \because \; q > 1/2 \\ &= p_u. \end{aligned} \tag{2.4}$$

Thus, the initial belief of the node u in $\mathfrak{A}$ reduces. Similarly, we can see that the user's belief in goodness of $\mathfrak{B}$ increases when it observes the other node v has visited the restaurant $\mathfrak{B}$.

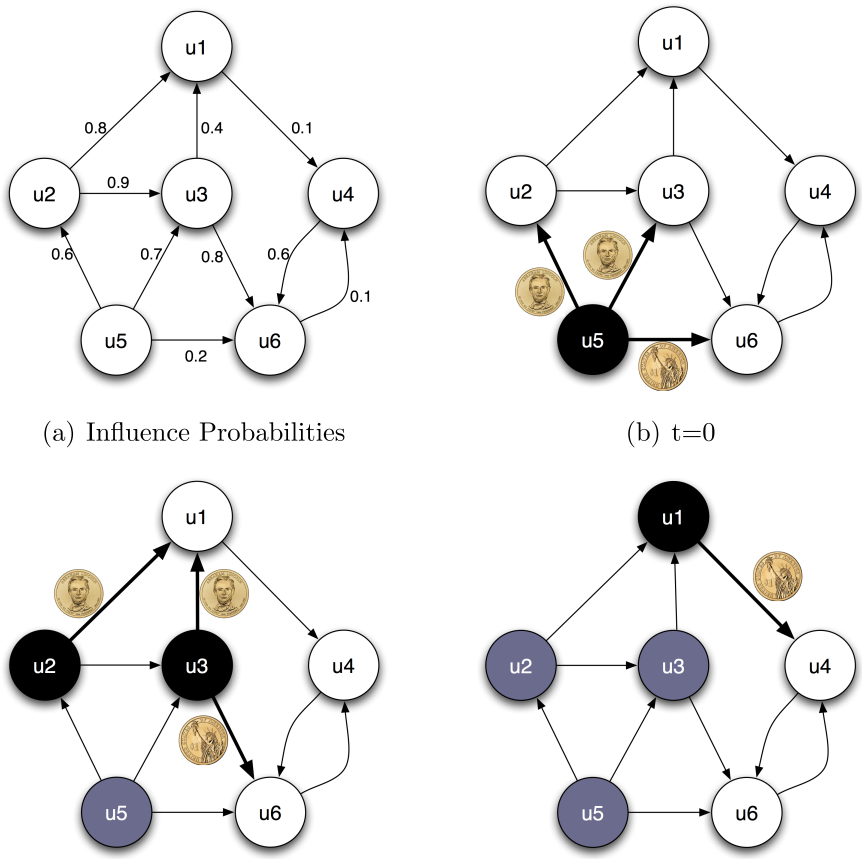
$$\begin{aligned}
p(\mathfrak{B} \text{ is good}|v \text{ choses } \mathfrak{B}) &= \frac{P(\mathfrak{B} \text{ is good})p(v \text{ choses } \mathfrak{B}|\mathfrak{B} \text{ is good})}{p(v \text{ choses } \mathfrak{B})} \\
&= \frac{(1-p_u)q}{p_u(1-q)+(1-p_u)q} \\
&\geq \frac{(1-p_u)q}{p_uq+(1-p_u)q} \because q > 1/2 \\
&= (1-p_u).
\end{aligned} \tag{2.5}$$

It is important to note that cascades generated under this model can lead to wrong cascades, where the wrong selection of behavior by initial nodes can lead to future wrong choices. This can happen when every node has a different prior information. If every node in the network makes decision independently based only on its prior belief then on an average they will make the correct decision. However, the wrong decision of initial nodes and informational effect can lead to lots of wrong choices. This effect is many times exploited by marketers where they distribute free samples of a new product to few people for adoption. By seeing them, others in the network can also adopt the product, even though the new product is not better than the existing ones. This kind of strategy works best when people can see the adoption behavior but not how satisfied they are. If the reviews from previous customers are known then such kind of wrong cascades can be prevented.

Probabilistic Social Influence Models. The models in this category do not explicitly model the reason behind a persons' decision, but consider all different types of social influence by directly modeling the bandwagon effect. These models consider the adoption of a product as a step-wise process. The social influence of nodes who have already adopted the product, influence the new nodes to adopt the product in the next time step. Thus, the dynamics of every information cascade in considered to unfold in discrete steps. The nodes who adopt the product are called the *active nodes*. Initially, at time $t = 0$, few seed nodes are active. Influence of these seed nodes can activate their neighbors at $t = 1$ and the influence of newly activated nodes can activate their neighbors in next time step. Thus, the process of activation continues until there are no new activations. These kind of models can be broadly divided in two categories - the Independent Cascade and the Linear Threshold models.

- **Independent Cascade model (IC).** According to this model, as soon as a node v becomes active, it takes a single chance to activate its inactive neighbor u. If it succeeds then the node u becomes active else it remains inactive. Thus every time a neighbor of u purchases a product, there is a chance that u will also decide to purchase the product. Specifically, every directed edge (u, v) in the network is associated with a fixed probability $p_{v,u}$ which indicates the influence of node v on the node u. Larger the value of $p_{v,u}$, stronger is the strength of influence of v on u. When the node u becomes active at time t then it flips a biased coin with *head* probability $p_{v,u}$, if the outcome is head then the edge (v, u) is considered an *active edge*. Note that the direction of active edges or influence is opposite of the original edges.

  Thus, the node u becomes active in the next time step $(t+1)$ if it has at least one active edge incident on it. Formally, given the set of neighbors $\mathcal{A}(u)$ of the node u that got activated at time t, the node u's probability of activation can be written as



(a) Influence Probabilities (b) t=0

(c) t=1 (d) t=2

Figure 2.2: An example of IC cascades.

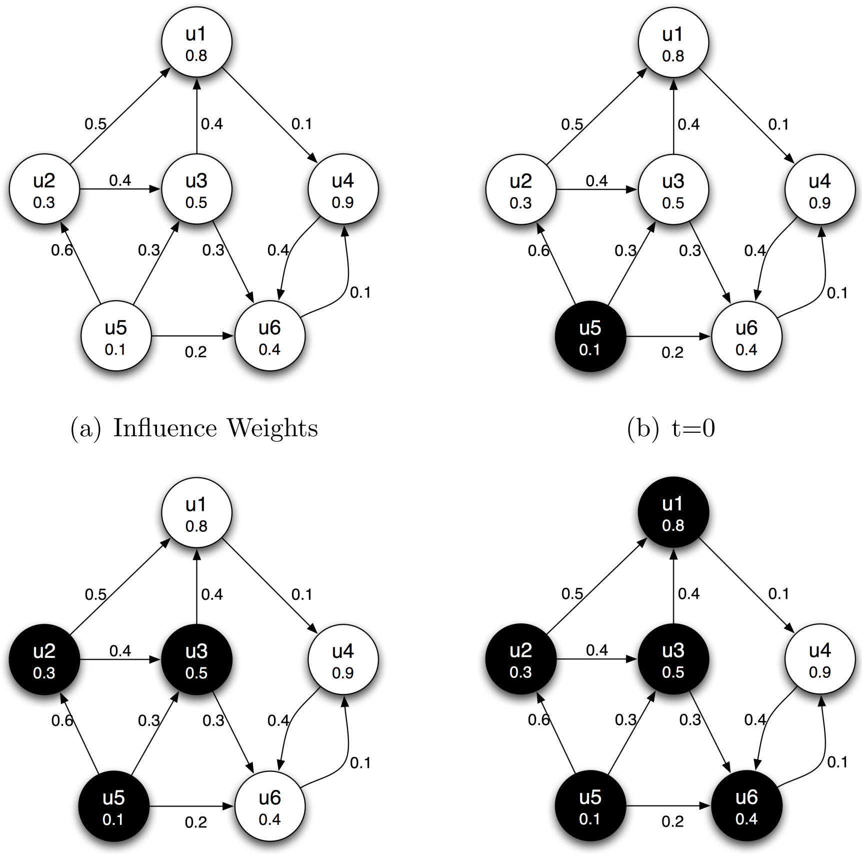
$$1 - \prod_{v \in \mathcal{A}(u)} (1 - p_{v,u}). \tag{2.6}$$

Figure 2.2 shows an example of a network with six nodes and how information flows in it under the IC model. The influence probabilities are indicated next to the corresponding edges. At any time, the nodes activated in that time step are indicated by the black circles and the nodes activated before that are shown by gray circles. Initially, at time $t = 0$ only node u_5 is active. It flips the coins for its all outgoing edges according to the edge influence probabilities and since the outcome for edges (u_5, u_2) and (u_5, u_3) is head, the nodes u_2 and u_3 become active at $t = 1$. While the node u_6 remains inactive. After $t = 0$ the node u_5 is not given any chance to active any other nodes. At time $t = 1$, both the newly active nodes u_2 and u_3 are successfully able to activate the node u_1. Node u_3 tries to activate the node u_6 but fails. Thus, u_6 continues to remain inactive. Information propagation stops at time $t = 2$, when only active node u_1 fails to activate u_4.

- **Linear Threshold model (LT).** In LT model, a node becomes active when sufficient number of its friends have become active. Specifically, every node u in the network is associated with a threshold $\theta_u \in [0, 1]$. This threshold indicates the resistance of the user u to adopt a new product. The larger the value of θ_u, larger amount of social influence is required to convince the user about the product. The social influence of neighbor v on the node u is denoted by weight $w_{v,u}$, such that $\sum_{(u,v)\in E} w_{v,u} \leq 1$. When the sum of social influence weights of the active neighbors of the node u exceeds its threshold, the node becomes active. Formally, given a set of all active neighbors $\mathcal{A}(u)$ of the node u, the node becomes active when

$$\theta_u \leq \sum_{v\in\mathcal{A}(u)} w_{v,u}. \tag{2.7}$$

An example of how information cascade unfolds under the LT model is shown in Figure 2.3. In this example, there are six nodes which are connected to each other via directed edges. The threshold of each node is indicated along with its label and the influence weights are indicated next to the corresponding edges. All the active nodes are shown



(a) Influence Weights (b) t=0

(c) t=1 (d) t=2

Figure 2.3: An example of LT cascades.

by the black circle. The information cascade starts from the node u_5, which is able to influence the node u_2 and u_3 as their internal thresholds are less than the influence weights. These nodes also become active in time step $t = 1$. The combined influence of the node u_5 and u_3 are able to activate the node u_6 in step $t = 2$. It should be noted that the node u_5 failed to activate the node u_6 in step $t = 1$ because the influence weight was less than 0.4.

The two models have been generalized to model the complex decision functions [45].

Information Flow in presence of Competition. Both the IC and LT models have been extended for the situations where instead of one product, multiple products compete in the network for adoption. Like in case of the single product, the cascade unfolds in discrete steps and every node is either in active or inactive state. Additionally, every active node also has a color associated with it which denotes the product the node has adopted. Thus, initially only few seed nodes are active with specific color. The activation processes of new nodes is then modeled by following models.

- **IC-Based Model.** Like the IC based model, as soon as a node v becomes active in the network, it flips coins with head probability $p_{v,u}$ to activate its neighbors u. It the outcome is head then the node u becomes active in the next time step. Further, the color of v is also copied to the node u. However, the problem arises when more than one neighbor with different colors, try to activate their common neighbor u at the same time t. If they all get the coin flip outcome as head, then it is not clear which color should be copied to node u. To resolve this situation, one of the active edge is considered (at random) as the influencing edge and the color of its node is copied to the node u [7]. This model has been used to study the second-mover strategy where given the competitor's choice of early adopters for product $\mathfrak{B}$, the task is to select a set of early adopters for the product $\mathfrak{A}$, such that the expected spread of the product $\mathfrak{A}$ can be maximized.

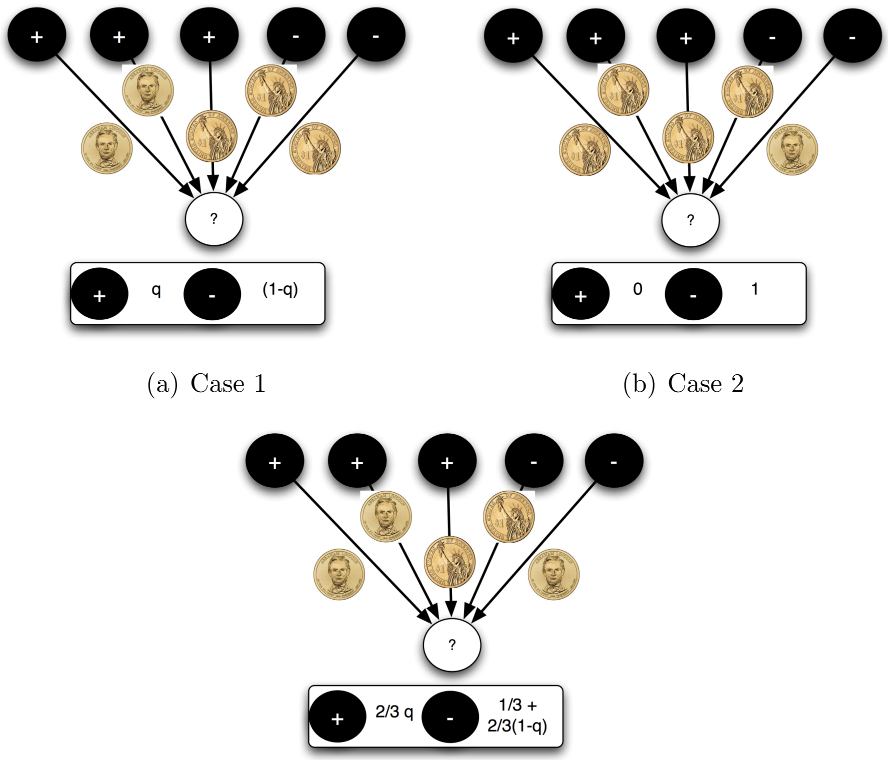
- **LT-Based Model.** LT model has been naturally extended to handle the competitive product adoption scenario [8]. The *weight-proportional* model considers that a node u becomes active if the sum of weights from neighbors exceeds its threshold. However, the color of the node u is decided based on the relative weight of the social influence of each color. Thus, the probability of node u selecting the color $\mathfrak{A}$ is proportional to the sum of the weights of neighbors with color $\mathfrak{A}$. In the *separated-threshold* model, every node has separate internal threshold for different products. If the sum of weights from neighbors with the color $\mathfrak{A}$, exceeds the corresponding threshold of u, then u becomes active with color $\mathfrak{A}$. In case, more than one color is able to activate the node, then the ties are broken at broken. The *or* model is a theoretical model which first assumes that each color diffuses in the network unhindered from the competing color. Then, once the information cascades for all the colors stop, a tie-breaking stage takes place where the colors of the active nodes are decided. If a node is active with more than one color, then its final color is selected by randomly picking from the active colors set.

Information Flow in presence of Negative Opinions. All of the above models assume that people can observe just the adoption decision of others. However, in many real world situations, we not only know the adoption decision of others but also their experience with the product. Further, in most of the cases, the negative opinions do not exist in the beginning but they arise as the information cascade unfolds and people try the product. In such situations, it is important to consider the emergence and propagation of polarity of opinions. Among the few works done in this direction is by Ma et al. [65] and Chen et al. [15]. However, the former work assumes that both kinds of opinions exist from the beginning, thus is closer to the competitive scenario where two opinions compete with each other.

Recently, the IC model has been extended to model the emergence and propagation of negative opinions. This model is been referred as IC-N (IC model when Negative opinions may emerge)

[15]. The motivating scenario for IC-N is the restaurant's review, where people decide to dine in a new restaurant based on their friends' reviews. If the reviews are positive then they decide to try that restaurant and based on their experience, they will either hold a positive opinion or negative opinion. However, if the friends' reviews have extremely negative and surprising voice to it, then they may decide not to go to the restaurant. For example, if a review says "I found cockroach in my meal.", then you surely will not go to that restaurant and in turn tell your friends, not to try it. Thus, the negative opinions will emerge from the product faults and will spread virally in the network. Specifically, the IC-N model associates a quality factor q with every product, where a high value of indicates that the product has a very good quality. Every node v in the network influence its neighbor u with probability $p_{v,u,+}$ when it is positively opinionated. The influence probability is $p_{v,u,-}$ when the node v is negatively opinionated. The information cascade starts with few seed nodes as active nodes, who choses their opinion as positive with probability q or become negative with probability $(1 - q)$. Then like the IC-model, they try to activate their neighbors u according to the influence probabilities. If there are more than one active edge on the node u, one of them is chosen as influencing node and its opinion is considered as the influencing opinion. If the influencing opinion is negative, then it is directly copied to the node u. However, if it is positive, then the node u goes and tries the product (restaurant in our example) and thus, becomes positive with probability q and negative with probability $1 - q$. The newly active node influence their neighbors in the next time step. Thus, the process of activation continues until there are no new activations.

The decision making process of node u for different cases is shown in Figure 2.4. There are five active neighbors, out of them three are positively active and two are negatively active.



(a) Case 1 (b) Case 2

(c) Case 3

Figure 2.4: Example of decision making process under the IC-N model.

In each case, every active node flips a coin according to edge influence probabilities. The figure shows the different outcome for different outcomes of coin flip. In case 1, when the attempts of only positively active nodes are successful, the node becomes positively active with probability q and negative with probability $(1-q)$. While in case 2, when only successful activation is by the negatively active node, the influencing node becomes negative with probability 1. The case 3 is tricky where two out of three edges are positively active and one is negatively active. Since the influencing opinion is found by breaking ties at random, there is 1/3 probability that negative edge is chosen, in which case the opinion is replicated by the influencing node. Thus, there is $1/3$ probability that the node becomes negatively active. However the influencing opinion is positive with $2/3$ probability. In this case, the node will be positive with probability q, thus the overall probability of being positive is $2q/3$ and probability of being negative is $1/3 + 2/3(1-q)$.

There are a number of problems with this model. First it assumes that the negative opinions flow virally in the network, which could be true in case of extremely negative and shocking news; but in most of cases negative opinions do not get echoed by the others and stops the information diffusion. For example, comments like "food was cold ", are not likely to spread virally in the network. Secondly, the propose model gives rise to a complex inference problem. That is, if we try to estimate the influence probabilities using this model, the resultant form is very complex to solve. This problem arises even if we assume that every active node v influences its neighbor with p_+ when v is positively opinionated and influence probability is p_- when v is negatively opinionated. In such case, the total probability of a node u getting activated with positive opinion is

$$q \sum_{s'_+=1}^{s_+} \sum_{s'_-=0}^{s_-} \frac{s'_+}{s'_+ + s'_-} \binom{s_+}{s'_+} \binom{s_-}{s'_-} p_+^{s'_+} (1-p_+)^{(s_+-s'_+)} p_-^{s'_-} (1-p_-)^{(s_--s'_-)}, \tag{2.8}$$

where s_+ is total number of positive neighbors and s_- is total number of negative neighbors of the node u. The fraction term which is the first component inside the sum, is the probability of selection of positively active edge given s'_+ edges are positively active and s'_- edges are negatively active. The rest of the term corresponds to the probability of activation of s'_+ out of s_+ edges from positive neighbors and the probability of activation of s'_- out of s_- edges from negatively opinionated neighbors. The complexity of this expression shows the difficulty involved in estimating the model parameters p_+ and p_-. We address both of the above problems in Chapter 3.

2.2.4 Influence Estimation

All the information models proposed in the previous section, have a set of parameters. Specifically the probabilistic models have the probability/weight of social influence parameters for every edge in the network. Usually, these parameters are chosen to be some function of frequency and duration of interactions and number of common attributes between two persons [90]. The following is the rational behind it: when two individuals interact with each other more frequently and their interaction duration is high, then it is highly likely that the two are close to each other and hence the tie between them is strong. Another, factor which is considered is number of interests and friends the two persons share, because similar interests implies stronger friendship. However, these heuristics do not have a principled approach behind them and they often confuse the social influence with homophily (discussed in Section 2.2.6).

Another idea is to directly observe who influence whom in the given social networks. However, such signals can never be observed directly. Usually, the activities in any social network, can be recorded at the node level, not at the edge level. For example, it can be recorded when a user buy a product but not whose recommendation has motivated the user. Further, many times it is not just one friend who influence the user's decisions. In such cases, we need to find out the relative weight of influence of each friends.

Recently few approaches have been proposed to estimate these influence parameters using just the action log data. The action log data contains the details of the time an action is taken (in our case adopts a product) by a user. These logs do not have the information of who influenced the user to take the action. Formally, if the set of information cascades is represented by $\mathcal{C}$, where every cascade $c \in \mathcal{C}$ represents the information cascade for a particular product and $\mathcal{A}^c$ represents the the set of active nodes in the cascade c. Then, the action log records the time t_u^c when every user $u \in \mathcal{A}^c$ takes the action in the cascade c.

The key idea behind these influence estimation approaches is following: If a user's actions are frequently followed by the actions' of her friend, then it implies that the user has high influence on her friend. Most of the proposed approaches learn the influence probabilities by maximizing the likelihood of observing the action log. One notable exception is [29] where authors maximize the f-measure instead of the likelihood. The approaches differ from each other in two ways: 1) The formulation of likelihood function and 2) the assumed information cascade model. Following are the key approaches.

- **IC-Based Approach.** This is the first approach proposed to learn the influence parameters from the action log. The approach [80] assumes that the information propagates strictly according to IC model and divides the likelihood of observing an action log in three components, which are as follows.

    1. The conditional probability that the nodes $u \in \mathcal{A}^c$ get activated at time t_u^c in a cascade c, given the set of neighbors v activated at time $t_v^c = t_u^c - 1$.
    2. The conditional probability that these active nodes $u \in \mathcal{A}^c$ do not get activated before time $t < t_u^c$, given the set of neighbors v activated at time $t_v^c = t - 1$.
    3. The conditional probability that the inactive nodes $u \notin \mathcal{A}^c$ do not get activated at any time step t, given the set of neighbors v activated at time $t_v^c = t - 1$.

Thus, the likelihood of an action log is written as,

$$L = \prod_{c \in \mathcal{C}} \prod_{u \in \mathcal{A}^c} \Big[\Big(1 - \prod_{(u,v) \in E; t_v^c = t_u^c - 1} (1 - p_{v,u})\Big) \cdot \prod_{t < t_u^c} \prod_{(u,v) \in E; t_v^c = t-1} (1 - p_{v,u})\Big] \cdot \prod_{u \notin \mathcal{A}^c} \prod_{t} \prod_{(u,v) \in E; t_v^c = t-1} (1 - p_{v,u}), \tag{2.9}$$

The expectation maximization (EM) algorithm is then used for maximizing the likelihood function.

- **IC with transmission delay.** Unlike the above approach, which strictly follows the IC model, approaches in this category consider a variation of IC model which also accounts for the *transmission* delays or *incubation* delays. The transmission delay specifies the time between the time of influence and time of an action of a node. This delay occurs because taking actions take time, for example to buy a product or to write a review. Thus, for most real world settings transmission delays can not be assumed to be zero.

  The modified IC model works as following. When a node v gets activated at time t_u and it attempts to activate its neighbors u. If the attempt is successful, then the node u will become active at time $t_u + \tau_{u,v}$, where $\tau_{u,v}$ is the transmission delay and is sampled from a transmission delay model. Rest of the process of the information flow is same as the IC model. The likelihood of observing the entire action log (over various products) is then divided in two components, which are as follows.

  1. The conditional probability that the nodes $u \in \mathcal{A}^c$ get activated at time t_u^c by its active neighbor v and selects $t_u^c - t_v^c$ as the transmission delay.

2. The conditional probability that the inactive nodes $u \notin \mathcal{A}^c$ do not get activated at any time step t, given the set of all active neighbors v.

ConNIe [70] assumes that all transmission delays are sampled from a fixed transmission model and writes the likelihood of the action log as,

$$L = \prod_{c \in \mathcal{C}} \prod_{u \in \mathcal{A}^c} \Big(1 - \prod_{(u,v) \in E; t_v^c < t_u^c} (1 - \dot{w}(t_u^c - t_v^c) p_{v,u})\Big) \cdot \prod_{u \notin \mathcal{A}^c} \prod_{(u,v) \in E; v \in \mathcal{A}^c} \Big(1 - p_{v,u}\Big), \tag{2.10}$$

where $\dot{w}(t)$ is the probability of selecting t as the transmission delay. Note that above expression for likelihood ignores the probability of the nodes not getting activated before their recorded time of activation.

Since maximizing log of a function also maximizes the original function, we maximize the log likelihood function. Thus, we have the following optimization problem

$$\begin{aligned} \text{maximize} & \sum_{c \in \mathcal{C}} \sum_{u \in \mathcal{A}^c} \hat{\gamma}_u^c + \sum_{u \notin \mathcal{A}^c} \sum_{(u,v) \in E; v \in \mathcal{A}^c;} \hat{\hat{p}}_{v,u} \\ \text{subject to} & \\ & \hat{\hat{p}}_{v,u} \leq 0, \\ & \hat{\gamma}_u^c \leq 0, \\ \sum_{(v,u) \in E; t_v^c < t_u^c} & \log(1 - \dot{w}(t_u^c - t_v^c) + \dot{w}(t_u^c - t_v^c) \exp \hat{\hat{p}}_{v,u}) \\ & \leq \log(1 - \exp\ \hat{\gamma}_u^c), \end{aligned} \tag{2.11}$$

where $\hat{\gamma}_u^c = \log(1 - \prod_{(v,u) \in E; t_v^c < t_u^c} (1 - \dot{w}(t_u^c - t_v^c) p_{v,u})$ and $\hat{\hat{p}}_{v,u} = \log(1 - p_{v,u})$. Above optimization problem is a separable convex optimization problem because the objective function is linear and the last constraint is the sum of non-linear functions who depend on only one variable (recall $\dot{w}(t_u^c - t_v^c)$ is a constant). Thus, it can be solved efficiently using Mosek [1] and gives solution with global maximum.

ConNIe has been extended by considering different transmission delays for every edge. NetRate [33] considers same functional form for every edge but with different transmission rate parameter. These parameters are then learned along with influence probabilities by maximizing the log likelihood function. It has been shown that the optimization problem is convex for the exponential, power-law and Rayleigh transmission models.

Apart from above, many variations of the problem of influence probabilities, have also been explored. Gomez-Rodrigue et al. [34] assume the same influence probability for every edge and focus on learning the existence of influential edges in the network. They exploit the sub-modularity property to efficiently find the most likely spanning tree for each cascade. In [91], authors consider the problem of finding the over all influence of the nodes, instead of finding the edge-wise influence. They propose an efficient method to estimate the influence of the nodes as the function of time. Specifically, they assume that influence of any node v as a long vector along the time axis, where every k^{th} entry indicates the expected number of nodes directly activated by the node v at k time units passed the time of activation of the node t_v. Thus, this vector represents the influence of the node v as the function of time. Then, the idea is to consider the number of new activations in the network at any time t, to be the sum of $(t - t_v)^{th}$ entries of influence vectors of all the active nodes v. Thus, the nodes influence vectors are learned by minimizing the square error between the actual number of new activations and the proposed functional form. The problem of learning the topic-specific influence weights has also been considered [85].

Recently researchers have started questioning the above approaches because they attribute every activation to social influence. Many times correlation in time of actions can also result from the homophily (similarity between ties) or from the external factors [3]. Aral et al. [4] observe that ignoring these factors results in 300-700% overestimation of social influence in Yahoo! Go data. To add to complexity, Crandall et al. [20] observe a feedback effect between the homphily and the social influence, where social influence makes current friends more similar while homophily results in creation of new ties.

2.2.5 Viral Marketing

Viral marketing is unanimously agreed key application of information flow in social networks. It refers to a marketing technique that uses the social network connections to produce increase in the brand awareness or to increase product sales. It uses the self-replicating word-of-mouth publicity, which is analogous to the spread of viruses or computer viruses. The underlying goal of marketers is to identify individuals with high social influence that can spread a word about the product such that it reaches to many people. However, these individuals do not always produce non-overlapping set of influenced nodes, thus it is required to select these individuals such that the overlapping can be minimized. Formally, this is formulated as the following influence maximization problem [22].

Influence Maximization Problem. Given a directed and edge-weighted social graph G, a propagation model m, and a number $k \leq n$, find a set $S \subseteq V, |S| = k$, such that the expected number of active nodes at the end of the information diffusion

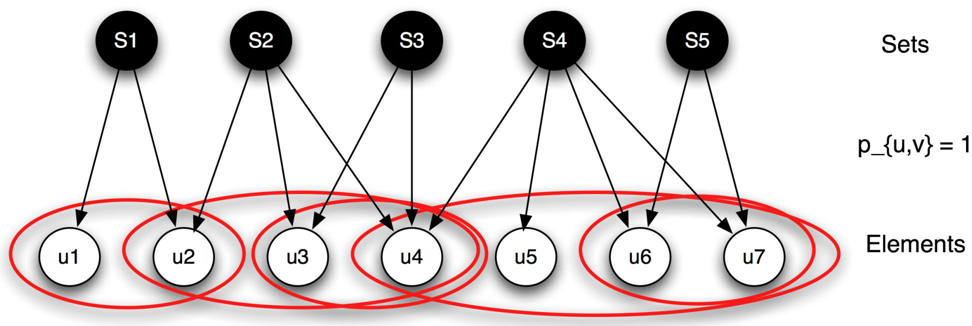


Figure 2.5: Reduction of NP-complete Set Cover problem to Influence Maximization problem

process, denoted by $\sigma_m(S)$, is maximum.

In general, the influence maximization problem is NP-hard [45] and the reduction follows from the *Set-cover* problem. To understand it, consider an instance of NP-complete set-cover problem where $U = \{u_1, u_2, \ldots, u_n\}$ is the set of nodes and $S_1, S_2, \ldots, S_m$ are the subsets of U and we want to find k subsets such that they cover all nodes in U. This problem can be viewed as a special case of influence maximization by constructing a social graph as following. Add every node in U and every subset S_i, as a node in the graph. Thus there are total $n+m$ nodes. Then, for every node corresponding to set S_i, add directed edges to the nodes u_j such that $u_j \in S_i$. Set the activation probabilities of these edges to 1. Thus, finding the set cover of size k is equivalent to finding the set of nodes in this graph with $\sigma_m \geq (n+k)$ under IC model. The graph construction is shown in Figure 2.5.

However, it has been shown that, under the IC and LT model, the influence maximization problem can be solved using the greedy algorithm, to achieve $(1 - 1/e)$ approximate solution [45]. The result follows from the following well known result:

if f is a monotonic increasing and submodular function, the problem of finding a set S of size k such that $f(S)$ is maximum, can be approximated to within a factor of $(1 - 1/e)$ by a greedy algorithm [72]. A function f is submodular if $f(S \cup v) - f(S) \geq f(T \cup v) - f(T)$ whenever $S \subseteq T$. To utilize this result, authors have proved that, for both IC and LT models the function $\sigma_m(S)$ is monotone and *submodular*. Intuitively, the submodularity of influence function σ_m implies that the probability of a node v to influence u, does not increase if more number of nodes are active.

The submodularity of influence function under IC model, can be proved as follows. Let S be the set of seed nodes which are active at time $t = 0$. Recall that these nodes v activate their neighbors u by flipping coins with probability $p_{v,u}$. If outcome is head then edge (v, u) is considered active and u is marked active. The newly activated nodes do the same thing and the process continues. Thus, at the end of cascade, every active node can be reachable from the seed set via active edges. However, instead of unfolding the cascade in steps, we can find the active edges by pre-fliping all the edge coins, thereby revealing the results immediately. In this case too, the set of reachable nodes from seed set S will be same as the step wise information propogation. Thus, the influence of a set of seed nodes can be found by pre-fliping the coins for every edge and finding the number of reachable nodes. Let us denote the non-probabilistic graph generated by one possible pre-flipping the coins, by g. Then the expected influence of the set S can be written as

$$\sigma(S) = \sum_{g} P(g).\sigma_g(S) \tag{2.12}$$

where $P(g)$ is the probability of generating the non-probabilistic graph g and $\sigma_g(S)$ is the number of reachable nodes from S in the graph g. Since non-negative linear sum of submodular is also submodular, it is sufficient to prove that $\sigma_g(S)$ is submodular.

Algorithm 1 Greedy Algorithm for Influence Maximization

Require: G, k, σ_m

1: $S \leftarrow \phi$
2: **while** $|S| < k$ **do**
3: $\quad u \leftarrow \arg\ \max_{v \in (V-S)} \sigma_m(S+v) - \sigma_m(S)$
4: $\quad S \leftarrow S \cup u$
5: **end while**
6: **return** seed set S

To prove the latter, consider two seed sets T and S such that $S \subseteq T$. Now consider the set of nodes which are reachable from $(S+v)$ but not from S in g. This set is always going to be superset of the set of nodes which are reachable from $(T+v)$ but not from T in g. That is, $\sigma_g(S \cup v) - \sigma_g(S) \geq \sigma_g(T \cup v) - \sigma_g(T)$. Thus, the influence function $\sigma(S)$ is submodular under the IC model.

The submodularity of influence function under LT model, can be proved by observing that, corresponding to every LT model, an IC model can be constructed such that the distribution of reachable nodes from the seed set S is same in both models [45]. The greedy algorithm for the influence maximization for monotonic increasing and submodular influence function σ_m is presented in Algorithm 1.

However, even if we use the greedy algorithm, the complexity of evaluating line 3 (finding the node with the largest marginal gain $\sigma_m(S+v) - \sigma_m(S)$) is very high. In fact, computing the expected spread given a seed set $\sigma_m(S)$ is #P-hard under both the IC model [16, 17] and the LT model [18]. One possible way to estimate $\sigma_m(S)$ by running Monte Carlo (MC) simulations to generate different worlds g and then, calculate the number of reachable nodes from the seed set S in each of the world [45]. Then the average of number of reachable nodes can be used as an estimate of the expected spread. However, to get an reliable estimate, a large number of such worlds needs to be generated (the authors report 10,000 trials). This makes the step 3 computationally very expensive.

Significant amount of work has been done to find efficient solutions to influence maximization. The submodularity of influence function has been further exploited [35, 60] to reduce the number of calls to the marginal gain module. The key observation is that, the marginal gain of a node r in the current iteration cannot be more than its marginal gain in previous iterations because of the submodularity property. Thus if there is a node w whose marginal gain in the current iteration is greater than that of marginal gain of r in the previous iteration, then there is no need to calculate the marginal gain of r in the current iteration. This significantly reduces the number of calls to the spread estimation module. Leskovec et al. [60] have reported an improvement of over 700 times in the running time.

Recently many works have focused on replacing the MC simulations used to calculate $\sigma(S)$ with efficient heuristics [16, 17, 18]. For IC models, one of the heuristic uses Maximum Influence Paths (MIP) on the original graph G instead of finding shortest path in every sampled graph g [17]. An MIP between a pair of nodes (v, u) is the path with the maximum propagation probability from v to u. For LT model, the expected influence of the seed set S can be calculated by enumerating all *simple* paths (with no cycles) starting from S and summing their weight. However, this problem is #-P hard. One of the heuristic to get tractable solution is to ignore the paths with small probability and thus, terminate the path traversals if probability of the path is very small [36]. Another proposed heuristic is to construct local DAGs for each node and consider influence only within it [18] because computing $\sigma(S)$ over DAGs can be done in linear time while otherwise is #P-hard for general graphs. While none of heuristics offer theoretical guarantees, they are empirically found to obtain solutions which are very similar to the greedy algorithm output.

The influence maximization problem has also been analyzed theoretically for specific types of graph, especially for the scale free graphs [93].

2.2.6 Influence vs. Homophily

It has been well established fact that human behaviors tend to cluster in both network space and time. That is, the behavior of nodes who are close to each other behave similarly. Additionally their time of actions is also close to each other. For example, obesity in humans. However, this pattern can be explained by a number of factors apart from social influence, such as homophily or other confounding factors. Recall that homophily is the tendency of people to chose friends who are similar to them. Thus, it can be expected that behavior of friends is correlated in space. In fact, this observation has lead to development to various social recommender systems [43, 66]. The homophily can many times also result in friends to take their actions almost at similar time, thereby giving high correlation in time space too. Thus, the correlation in space and time can be attributed to both influence and homophily. Differentiating between the two is thus equivalent to differentiate between the causation and correlation, which is known to be notoriously difficult problem.

It is important that we should differentiate between the social influence and homophily, because different marketing strategies are effective for each one of them. When social influence is prominent, then viral marketing is more effective. While traditional market segmentation strategy based on observable characteristics of consumers, is best when homophily is prominent.

To differentiate between the two, a simple randomized time shuffle test has been designed [3]. The key idea is that, if social influence is abesent then the activation time of a user should be independent of the activation time of his friends. The test works by assuming a simple influence model, where the activation probability of the nodes increases as number of friends who have already active increases. Formally, the probability of activation of a node who has $\mathcal{A}$ active neighbors is

$$\frac{e^{a\log(|\mathcal{A}|+1)+b}}{1+e^{a\log(|\mathcal{A}|-1)+b}}, \tag{2.13}$$

where a is the social influence and b is a constant. Then the shuffle test estimates the values of a on two datasets, the original data and on the shuffled time data,. If the values of a are same both the datasets, then it implies that the time of action of nodes are independent of their friends' time of action. Thus, there is no social influence present in the network. This test has been used to exclude the presence of social influence in tagging behavior on Flickr [3].

In general, it has been noticed that, it is very difficult to differentiate between the latent homophily and social influence without making strong parametric assumptions [83].

2.2.7 Results from Large Scale Empirical Studies

Several empirical studies have also been carried out to understand the flow of information in large scale social networks like Flickr [12], Digg [84], Blogosphere [37], Twitter [54] etc. It has been observed that most of the cascades are short and "stars" to be most common shape in blogoshpere [62]. Studies on real information cascades like Digg [84], recommendation referral program [55] show that the probability of a node getting activated from repeated exposure, quickly saturates. This is in odd with the IC and LT model, where probability of activation increases as a function of number of active neighbors. The effect of repeated exposures for different topics has been studied on Twitter [78]. It has been observed that the effect varies with topics. The repeated exposure results in more number of adoptions for hashtags related to politically controversial topic. While for hashtags related to idioms and neologisms, the effect of multiple exposure decays rapidly. The transmission delay between the nodes has been analyzed on a large corporate email network [51], where it has been found that the transmission delay is not same for every edge in the network.

□ **End of chapter.**

Chapter 3

Impact on Product Purchase Decision

3.1 Introduction

Several probabilistic information flow models [45] have been developed to mimic the way information spreads in a social network. They attempt to predict the probability of a user to adopt a product given its friends' recommendations. The underlying belief is that social recommendations increases the user's trust on the product and thereby increases the probability of the user to adopt it. For example, positive friends reviews about a movie encourages us to watch it.

However, most of the existing models ignore the polarity of opinions, which is one of the important aspect of the information. In real world, both positive and negative opinions critically affect one's decision. While on one hand, positive opinions promote a product, on the other hand, negative opinions discourage its adoption. Further, negative opinions usually dominate the positive opinions in shaping one's decision [6]. Even slight hint of product faults, are sometimes sufficient to change our purchase decisions.

The two kind of opinions differ from each other, not just in terms of affecting the purchase decision, but also in their propagation patterns. Many work in psychology, hypothesize the negative opinions propagate contagiously in the network [79]. Such kind of pattern can be expected in case of shocking news, for example, comments such as “food poisoning from a restaurant food” are likely to be echoed in the network even though a user has not dined there. However, on social rating networks, the negative opinions do not get spread at all. For example, bad reviews about a movie discourage us from watching it buy it is less likely that we pass the negative comments to other friends without watching the movie by ourselves. In fact, we do observe that the presence of negative opinions reduces the number of expressed opinions (either positive or negative) on two real world datasets Flixster[1] and Epinions[2].

Motivated by above observations, we propose **polarity sensitive extensions** of both IC and LT to model the flow of information in social rating networks. We explicitly consider every opinion expressed by users to be a two step process. First step is the social influence which drives users to consider or not to consider the product. While the second step, describes the expressed opinion given the outcome of first step. The latter step considers various scenarios where negative opinions may emerge from the product faults or may not get published by the node. The usefulness of the proposed models are demonstrated by predicting the future users’ opinions using them. On both Flixster and Epinions datasets, polarity sensitive functions are able to predict the future opinions more accurately. Further, the best accuracy is achieved using the LT based extension.

The rest of this chapter is organized as follows. First we briefly present the closely related works in Section 3.2 and then present the precise problem formulation in Section 3.3, follow that with the description of the datasets used in this paper and

[1]http://www.flixster.com/
[2]http://www.epinions.com/

present our key observations in Section 3.4. Next, we describe our proposed model and the method to infer the social influence in Section 3.5 and Section 3.6. This is followed by a detailed empirical study on both synthetic and real datasets in Section 3.7 and Section 3.8 respectively. Finally, Section 3.9 summarizes our main results.

3.2 Related Work

Heat diffusion [65] and IC-N [15] are among the few works, which consider the polarity of opinions in information flow. However the former assumes that both kinds of opinions exist from the beginning of the information cascade and uses the same propagation behavior for both kind of opinions. Therefore, it is closer to the competitive model rather than the polarity sensitive information flow models. The IC-N model does consider the possiblity of emergence of the negative opinions from the product faults but it gives rise to a complex inference problem (the details are available in Section 2.2.3).

The polarity sensitive information diffusion shares some similarities with competitive information flow where more than one products compete within the social network for adoption [7, 8]. However, the two kind of opinions can not be treated as two competing products, because the impact of two kind of opinions is not symmetric. Further, when two polarities are present at the same time, they can cancel out the impact of each other, and thereby reduce the overall probability of getting influenced with either of opinions.

Recently, few studies have been carried out to study the propagation patterns of negative sentiments [39, 89]. However, negative sentiment is different from negative opinions, because negative opinions emerge from product qualities and work against the positive opinion for adoption of the underlying product.

3.3 Problem Definition

In this section, we first give necessary definitions and present the problem formulation.

Definition 1 *Let the* **activation state** s_u^c *be the state of a node* u *in cascade* c*. If the polarity of its published opinion is positive, then the node is called the* **positively active** *and its state value* $s_u^c = +$*. While if the polarity of its opinion is negative, then it is called* **negatively active** *and its state value* $s_u^c = -$*. The nodes who do not publish any opinion are referred to as* **inactive** *nodes and their state value* $s_u^c = 0$*.*

Definition 2 *Let the* **activation time** t_u^c *be the time when the node* u *publishes its opinion in cascade* c*.*

Definition 3 *Let the set of* **active neighbors** $\mathcal{A}^c(u,t)$ *be the set of already active neighbors of the node* u *in cascade* c *at time* t*, i.e.,* $\mathcal{A}^c(u,t) = \{v | v \in V \ \& \ (u,v) \in E \ \& \ s_v^c \in \{+,-\} \ \& \ t_v^c < t\}$*.*

Learning Task. Given the set of already active neighbors $\mathcal{A}^c(u,t)$ of user u at time t, the task is to learn the probability distribution over the activation state $p(s_u | \mathcal{A}^c(u,t))$ of the node u.

Next, we observe patterns that arise because of the presence of negative opinions in two real world rating networks, Flixster and Epinions. Based on the observed patterns, we propose various functional forms for the probability distribution over the activation state. The method to learn the parameters of the proposed functions is described next.

3.4 Data and Observations

3.4.1 Data Collection

Flixster. Flixster[3] is a popular social movie website which allows users to rate movies and share them with their friends. Users can rate movies by giving them score between 0 to 50, with 50 being the best. To construct the dataset, we have collected user ratings for all the movies released from Jan, 2005 to Dec, 2010. Only users who have rated at least 50 movies from this set, are kept and their friendship network is crawled. The raw data has 16,041 movies, 85,209 users and 5,71,505 edges. All the ratings of a movie are considered as one information cascade and ratings from 00 to 25 are assumed to be negative while 25 to 50 are assumed to be positive ratings.

Epinions. Epinions[4] allows users to post review articles about

Data set	Users	Edges	Products	Ratings
Flixster	85,209	5,71,505	16,049	10,086,362
Epinions	1,32,000	71,76,671	560,144	13,668,319

Table 3.1: Data statistics

products items from different categories (software, music, etc). While other users are allowed to rate their articles on the scale of 1 to 5, with 5 being most helpful. Recently Epinion has published an extended dataset [67] which also contains the date when people have rated the review articles. The dataset contains 1,32,000 users, 71,76,671 trust edges, 560,144 articles and 13,668,319 article ratings across 29 product categories. We consider all the ratings of a review article as one information cascade. Further, we assume rating 1, 2 as negative and 3-5 as

[3]http://www.flixster.com/
[4]http://www.epinions.com/

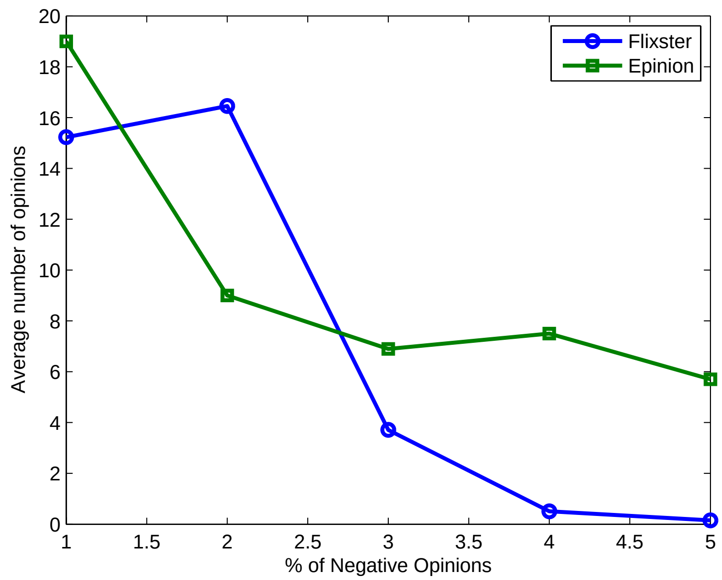


Figure 3.1: Average number of expressed opinions as percentage of negative opinions increases. For Flixster dataset, cascade length is scaled (divided by 100) to fit the data in plot.

positive rating.

3.4.2 Observations

As a first step, we qualitatively study the impact of polarity of opinions/ratings on the information flow. To get the first qualitative measure, we study the variation of cascade length with the percentage of negative opinions, where cascade length is defined as the number of active nodes in that cascade. For the same, we categorize cascades based on the ratio of number of negative opinions and total number of opinions expressed in that cascade. Then, we plot the average cascade length in each category against the percentage of negative opinions expressed. The plot is shown in Figure 3.1. We can observe that cascades with higher percentage of negative opinions, are usually shorter than the one with lower percentage of negative opinions.

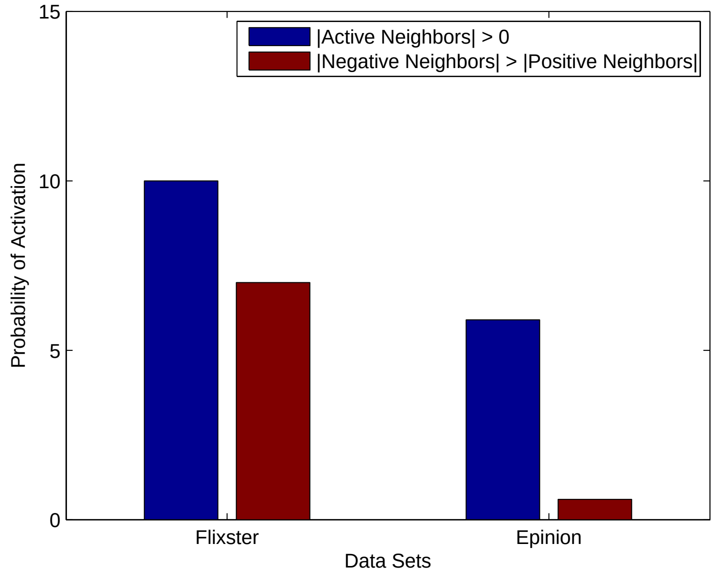


Figure 3.2: Average probability of activation, when at least one the neighbor is active vs when more than 50% of neighbors are negatively active.

Next, we look at the probability of activation of users (either with positive or negative opinion). We compare the probability of activation of users who have at least one of their neighbor active, with the probability of activation of users whose more than 50% of neighbors are negatively active. The plot is shown in Figure 3.2. It can be observed that the probability of activation decreases by a large amount when more than 50% of neighbors are negatively opinionated.

Both the observations clearly demonstrate the asymmetry between the positive and negative opinions. They suggest that,

- The presence of negatively opinionated neighbors significantly **reduce the activation probability** of a node.

- Cascades with negative opinions have relatively less number of participants. That is, negative opinions **prevent information spread** in the network.

Both the observations underline that the presence of negative opinions discourage users to look at the product (movie in case of Flixster and article in case of Epinion) and thereby rate them.

3.5 Polarity-Sensitive Information Flow Model

Here, we define $p(s_u^c|\mathcal{A}^c(u,t))$ such that it accounts for the difference in the propagation patterns and impact of two kinds of polarities. Since we will be considering only one cascade in this section, we will write s_u^c as s_u and $\mathcal{A}^c = \mathcal{A}$ for reading convince.

We consider the activation of any node u at time t as a two step process. Step 1 models neighbors' influence on node u, while in step 2, node u decides to publish its own opinion based on its neighbors' recommendations and its own experience. For the same, we introduce a **hidden** state variable $\tilde{s}_u$ to represent the social influence on node u, where $\tilde{s}_u \in \{+, -, 0\}$ with + indicating positive influence, - indicating negative influence and 0 indicating absence of any influencing opinion. Given the set of already active neighbors $\mathcal{A}(u,t)$ of node u,

$$p(s_u|A(u,t)) = \sum_{\tilde{s}_u} p(\tilde{s}_u|A(u,t)).p(s_u|\tilde{s}_u). \qquad (3.1)$$

Thus, we need efforts in two directions:

1. Define the functional form of $p(\tilde{s}_u|A(u,t))$, i.e., the probability of the node u to get influenced with a particular polarity of the opinion (positive or negative), given the neighbors' opinions and their influence probabilities/weights.

2. Define $p(s_u|\tilde{s}_u)$, in other words, decide the activation state of the node given the influencing opinion. In a simplistic scenario, the node can simply mirror its neighbors' opinion. Although other scenarios are also possible. A node can become negatively active even if it is influenced with

positive opinion by its neighbors. For example, consider a user who buys a product after reading positive opinions from its neighbors, but finds it to be a disappointment. In such cases negative opinions may emerge. Thus, this step is needed to decide if a node publishes any opinion and its polarity.

Next, we discuss each of the above steps in detail.

3.5.1 Social Influence Function

In this subsection, we define various functions over the neighbors' influence to decide whether a node is influenced by its neighbor and the polarity of opinion, it is influenced with. For the same we modify, the two most popular information propagation models, IC and LT models.

3.5.1.1 Polarity-Sensitive IC Model

Like IC model, we assume every node $u_i \in \mathcal{A}(u,t)$ influences node u independently with probability p_{v_i,u,o_i} when v_i is opinionated with opinion o_i. However, considering the completely independent model (considered by IC-N [15]) gives a very complex form for $p(\tilde{s}_v|\mathcal{A}(u,t))$; even when we assume that every v_i has same $p_{v_i,u,+}$ and $p_{v_i,u,-}$. Hence, next we propose two simple functions which can be seen as approximation of IC-N.

Independent Activation (IA). Here we first consider influence form all positively opinionated neighbors and all negatively opinionated neighbors, separately by following the IC model. For the same, we combine all the active neighbors with positive opinions and represent them by a super positive node *sp*. While all negatively opinionated active neighbors are combined to represent a super negative node *sn*. The probability of *sp*

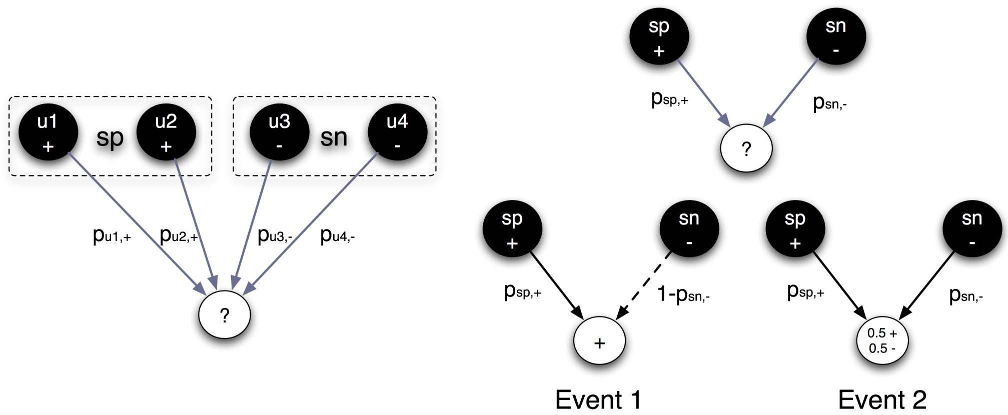


Figure 3.3: IA influence function

to influence the node$u$ with positive opinion, is defined as the probability that at least one of the positive active node is able to influence u and is equal to

$$p_{sp,u,+} = 1 - \prod_{v_i \in \mathcal{A}(u,t), o_i = +} (1 - p_{v_i,u,o_i}). \tag{3.2}$$

Similarly the probability of sn to influence u with negative opinion is:

$$p_{sn,u,-} = 1 - \prod_{v_i \in \mathcal{A}(u,t), o_i = -} (1 - p_{v_i,u,o_i}). \tag{3.3}$$

Then like IC-N model, both sp and sn independently flip coins with probability $p_{sp,u,+}$ and $p_{sn,u,-}$ respectively. If outcome is head then the corresponding edge to node v is considered to be an *active edge*. If both (sp, u) and (sn, u) are active edges then tie is broken by choosing one of the edges uniformly at random as the influencing node. Then, the probability of node u influenced by node sp is equal to the probability of following two events: E_1: (sp, u) is active and (sn, u) is inactive, E_2: both (sp, u) and (sn, u) are active and (sp, u) is chosen as the influencing opinion.

Thus,

$$\begin{aligned} p(\tilde{s}_u = +|\mathcal{A}(u,t)) &= p(E_1) + p(E_2) \\ &= p_{sp,u,+}(1 - p_{sn,u,-}) + \frac{1}{2}\, p_{sp,u,+}\, p_{sn,u,-} \\ &= \frac{1}{2}\, p_{sp,u,+}(1 + p_{sn,u,-}). \end{aligned} \tag{3.4}$$

Similarly probability of getting influenced with negative opinion $p(\tilde{s}_u = -|\mathcal{A}(u,t))$ can be written as $\frac{1}{2}\, p_{sn,u-}(1 + p_{sp,u,+})$. An example of decision making process under IA is shown in Figure 3.3.

Weight Proportional (WP). Here we define the probability of node u getting influenced with positive opinion as the ratio of expected number of active edges with positive opinions and expected number of total active edges. The expected number of active edges is simply the sum of their influence probabilities. Thus,

$$p(\tilde{s}_u = +|\mathcal{A}(u,t)) \propto \frac{\sum_{v_i \in \mathcal{A}(u,t), o_i = +} p_{v_i,u,+}}{\sum_{v_i \in \mathcal{A}(v,t)} p_{v_i,u,o_i}}. \tag{3.5}$$

However, this excludes the possibility that none of the neighbors are able to influence the node u. Hence we multiply the above quantity by the probability of having atleast one active edge. Thus,

$$p(\tilde{s}_u = +|\mathcal{A}(u,t)) = \frac{\sum_{v_i \in \mathcal{A}(u,t), o_i = +} p_{v_i,u,+}}{\sum_{v_i \in \mathcal{A}(u,t)} p_{v_i,u,o_i}} \left(1 - \prod_{v_i \in \mathcal{A}(u,t)} (1 - p_{v_i,u,o_i})\right). \tag{3.6}$$

Similarly one can write $p(\tilde{s}_u = -|\mathcal{A}(u,t))$.

3.5.1.2 Polarity-Sensitive LT Model (LT-PS)

Next, we extend the influence function used in LT model to account for the polarity of opinions. To the best of our knowledge, this is the **first** extension of the LT model which considers the polarity of opinions. Like in LT model, each node $u \in V$ is associated with an internal threshold $\theta_u \in [0, 1]$ which represents the minimum amount of social influence required for node u to get influenced. Lets assume that $w_{v_i,u,o_i} \in [0, 1]$ is weight of influence of node v_i on u when v_i has o_i opinion. Then, we define the social influence on node u as the difference between the sum of influence from positive opinionated nodes and the sum of influence from the negatively opinionated nodes. If the difference is greater than zero then the probability of u getting influenced with positive opinion is defined as

$$p(\tilde{s}_u = +|\mathcal{A}(u, t)) = g(b.(\theta_u - f^+(u))), \tag{3.7}$$

where g is a sigmoid function and is used to keep the probabilities between 0 and 1. The constant b is a hyper-parameter and controls the slope of sigmoid function. The function $f^+(u)$ is defined as

$$f^+(u) = \sum_{v_i \in \mathcal{A}(u,t), o_i = +} p_{v_i,u,+} - \sum_{v_i \in \mathcal{A}(u,t), o_i = -} p_{v_i,u,-}. \tag{3.8}$$

Similarly one can defined $p(\tilde{s}_u = 0|\mathcal{A}(u, t))$ if the difference between influence from positive opinionated nodes and the negatively opinionated node is less than zero. The probability of not getting with any influence is simply $p(\tilde{s}_u = 0|\mathcal{A}(u, t)) = 1 - p(\tilde{s}_u = +|\mathcal{A}(u, t)) - p(\tilde{s}_u = -|\mathcal{A}(u, t))$.

It is important to note that unlike IC based influence functions (IC-N, IA, WP), in LT-PS influence function, probability of a node u getting influenced (either with positive or negative opinion) does not always increase as the number of active neighbors increases. For example, lets assume that w_{v_i,u,o_i} (p_{v_i,u,o_i} for IC-N) is same for all the nodes and let it be p. Consider the two scenarios: S_1 when there is only one active neighbor and has positive opinion and S_2: when there are three active neighbors and two out of them are positive and one is negative. According to LT-PS, $p(\tilde{s}_u|\mathcal{A}(u,t))$ remain same in both scenarios. However for IC-N, the probability of influencing node u will increase from p to $(1-(1-p)^3)$ with $p(\tilde{s}_u = +|\mathcal{A}(u,t)) = \frac{2}{3}(1-(1-p)^3)$ and $p(\tilde{s}_u = -|\mathcal{A}(u,t)) = \frac{1}{3}(1-(1-p)^3)$. This property of LT-PS function makes it more suitable for modeling social influence for real world data, because in real world the presence of both positive and negative opinions cancel each others influence and there by reduces the overall probability of getting influenced.

3.5.2 Activation State of Influenced Node

Having decided the polarity of opinion which has influenced the node u, the next step is to decide state of the node u. Depending upon the nature of information, there are several possibilities:

Scenario 1 (Echo). The influenced node mirrors its neighbors' opinions as they are. Opinions related to ethical or political campaign, government policies are examples where neighbors opinions are echoed by their followers. This is similar to the competitive information propagation where two kinds of opinions exist at the start of the information cascade and compete with each other [7].

Scenario 2 (Emerge). In this case, opinions are expressed strictly based on one's experience with a product, while the positive (negative) social opinions play role in encouraging (discouraging) the user to buy the product. Thus, negative opinions

	Negative opinions exist from the beginning	**Negative opinions emerge**
Negative Opinion echoed	Scenario 1 (Echo)	Scenario 3 (Emerge-Echo)
Negative Opinion not echoed	-	Scenario 2 (Emerge)

Table 3.2: Various possibilities for propagation and emergence of negative opinions.

can emerge but they are not mirrored. The negative opinions create negative impression about the product and thus stopping users from buying it and thereby, no opinion is expressed by the user. For example, in case of movie reviews, when a user reads negative opinion, then it is unlikely that the user will watch that movie, but at the same time, it is unlikely that user publish negative opinions about the movie without watching it.

Scenario 3 (Emerge-Echo). Unlike Scenario 2, in some cases negative opinions propagate even without one's own experience with the product. For example, comments such as "food poisoning from a restaurant food." are likely to be echoed in the network even though a user has not dined in that restaurant. This scenario is studied in detail in [15].

The above scenarios are summarized in Table 3.2. They can be generalized by using two variables product quality factor, $q \in \{0, 1\}$ and virality of negative opinion constant $\rho \in [0, 1]$, where q quantifies his experience with a product and ρ represents the probability by which negative opinions from neighbors are echoed by a user. Thus, after getting positively influenced, a user becomes positively active with probability q and with $(1 - q)$, he becomes negatively active. Thus, the probability of positive activation is

$$p(s_u = +|\mathcal{A}(u,t)) = p(\tilde{s}_u = +|\mathcal{A}(u,t)) * q, \tag{3.9}$$

while probability of negative activation is

$$\begin{aligned} p(s_u = -|\mathcal{A}(u,t)) =& p(\tilde{s}_u = -|\mathcal{A}(u,t)) * \rho + \\ & p(\tilde{s}_u = +|\mathcal{A}(u,t)) * (1-q). \end{aligned} \tag{3.10}$$

For Scenario 2, ρ is equal to 0 while for Scenario 3, ρ is equal to 1. Further if we set $q = 1$ and $\rho = 1$ it will be equivalent to Scenario 1. The Scenario 2 best describes the information propagation is social rating networks. Since we work with datasets from Epinions and Flixster (which are social rating networks), we will use Scenario 2 to demonstrate the goodness of the proposed model.

3.6 Influence Estimation

In order to make predictions using $p(s_u^c|A^c(u,t))$, we need to estimate the values of the pair-wise influence parameters $p_{v,u,+/-}$ ($w_{v,u,+/-}$, θ_u for LT-PS). To learn these values, we use the historical information cascades and maximize the likelihood of observing them. Lets assume that C represents the set of historical information cascade. Then, the log likelihood LL of observing the set of cascades C can be written as the sum of log likelihood of each cascade $c \in C$.

$$\begin{aligned} LL(C) = \sum_{c \in C} \Big(& \sum_{o_u^c \in \{+,-\}} \log\ p(s_u^c = o_u^c|\mathcal{A}^c(u,t_u^c)) \\ & + \sum_{o_u^c = 0} \log\ p(s_u^c = 0|\mathcal{A}^c(u,t)) \Big) \end{aligned} \tag{3.11}$$

Here t is the end of the observation time window of cascades. Our objective is to chose model parameters which maximize the

$LL(C)$ and generalize well on the unseen data. Thus, for IC based model, we write our objective function as

$$\arg\ \max\ LL(C) - \lambda \sum_{u,v} (p^2_{v,u,+} + p^2_{v,u,-}), \tag{3.12}$$

and for LT-PS model

$$\arg\ \max\ LL(C) - \lambda \sum_{u} \theta^2_u - \lambda \sum_{u,v} (w^2_{v,u,+} + w^2_{v,u,-}), \tag{3.13}$$

where λ is a hyper-parameter that controls the amount of regularization. It can be noted that for Scenario 2, the quality factor q of products gets observed as part of a constant because it is assumed to be constant for each product (cascade). Thus, our objective function is independent of q.

This big objective function can be minimized, by independently minimizing the objective function for every node $u \in V$; because the parameter set $p_{v,u,+/-}$ (($w_{v,u,+/-}$, θ_u for LT-PS) for every node u, are different from other nodes' parameter set. This makes the inference problem scalable. We minimize each of the sub-problems using the steepest gradient descent method.

3.7 Experiments on Synthetic Data

The aim of this set of experiments is three folds.

1. Measure the effectiveness of IA and WP influence functions as approximation of IC-N model.

2. Evaluate the quality of model parameters estimated using the inference method proposed in Section 3.6.

3. Assess the ability of different influence functions to predict the next hop neighbor's activation status.

The synthetic data is generated by first synthetically generating the social network and then generating the synthetic cascades. We use the two state-of-art generators Scale Free Network [5] and Kronecker Graph [56], to generate the social graph. For the Kronecker graph, the core-periphery network is generated because real word networks are believed to have core-periphery. For the same, SNAP[5] is used as the tool for generation with kronecker parameter matrix [0.962 0.535;0.535 0.107] . To generate the scale-free network, we used NetworkX [38] with 512 nodes and 1024 edges.

The synthetic cascades for a given network are generated using the IC-N and LT-N as influence function. In both the sets, activation status of any influenced node is decided based on the Scenario 2. Further, we assume there is no incubation delay. The numbers representing the probabilities to influence others when the node is positively opinionated and negatively opinionated respectively are generated by sampling uniformly from 0.05 to 0.99 intervals.

3.7.1 IA and WP as Approximation of IC-N

Here we measure, how well IA and WP approximate the IC-N influence functions? For the same, we generate 1000 cascades using the IC-N model for different values of product quality q and network type (scale-free and kronecker graph). Then we measure the performance of the three influence functions assuming they know the actual model parameters $p_{v,u,+}, p_{v,u,-} \forall (u, v)$.

Performance Measure. To measure the performance of each influence function, we calculate the likelihood of observing the activation state of every influenced node with respect to the three influence functions IA, WP and IC-N and actual model

[5] http://snap.stanford.edu

parameters.

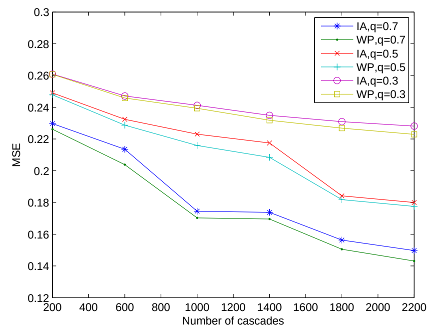
Network	q	IA	WP	IC-N
Scale Free Network	0.9	0.9604	0.9599	0.9391
	0.7	0.7865	0.7856	0.7667
	0.5	0.6396	0.6388	0.6214
	0.3	0.5536	0.5534	0.5422
Kronecker Network	0.9	0.9098	0.9073	0.8229
	0.7	0.7435	0.7412	0.6763
	0.5	0.6305	0.6292	0.5720
	0.3	0.5272	0.5264	0.4824

Table 3.3: Negative log likelihood averaged over all the influenced nodes when the model parameters are known.

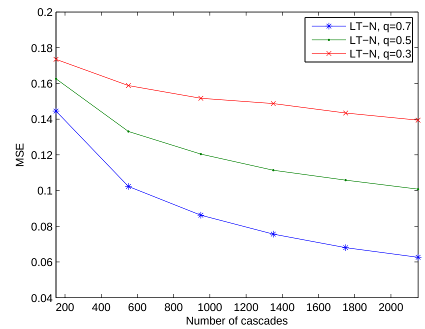
Observations. The negative log likelihood averaged over the influenced nodes for two types of network is presented in Table 3.3. It can be seen that the likelihood obtained from both IA and WP influence functions is very close to IC-N function with WP is a slightly better approximation. Further, following two trends can be noticed. As the product quality q decreases, the difference between the proposed approximations and IC-N becomes narrower. Additionally for kronecker network, the difference between the two approximations and IC-N is larger than that of the scale-free network. The probable reason is that number of influencing neighbors with different opinions are more in kronecker network and when q is large.

3.7.2 Quality of the Estimated Parameters

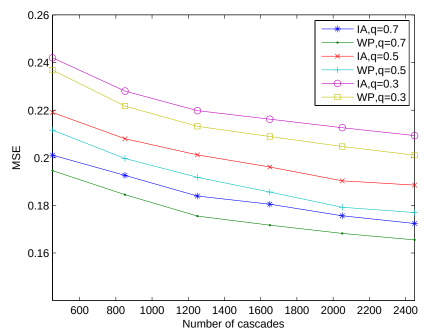
Next we evaluate the goodness of model parameters obtained using the inference method proposed in Section 3.6. For the same, the synthetically generated IC-N cascades are used as training set and IA and WP are used as the influence functions to estimate the latent model parameters of IC-N model. While LT-N



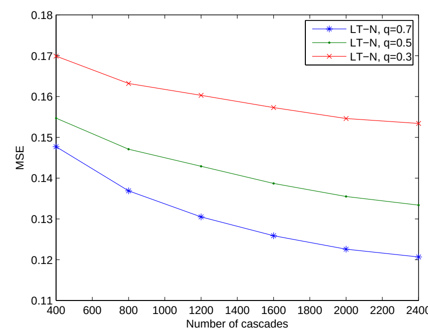
(a) IC-N Cascades, Scale Free Network



(b) LT-N Cascades, Scale Free Network



(c) IC-N Cascades, Kronecker Network



(d) LT-N Cascades, Kronecker Network

Figure 3.4: Mean Square Error between the estimated and actual model parameter, as the number of cascades increases.

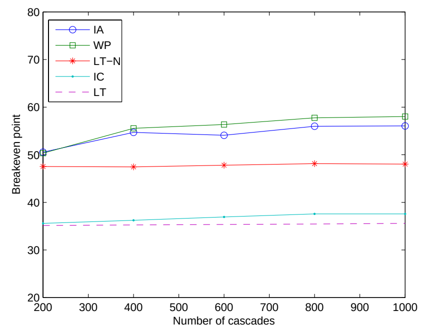
model parameters are estimated by using the LT-N cascades as training set and LT-N as influence function.

Performance Measure. The quality of estimated parameters is measured in terms of mean square error (MSE); where MSE is defined as the mean of square difference between the estimated and actual model parameters. Recall that, the IC-N model has two model parameters are $p_{v,u,+}$, $p_{v,u,-}$ for every $(u, v) \in E$, while the LT-N model has one more parameter θ_u for every node in addition to $w_{v,u,+}$, $w_{v,u,-}$ for every $(u, v) \in E$.

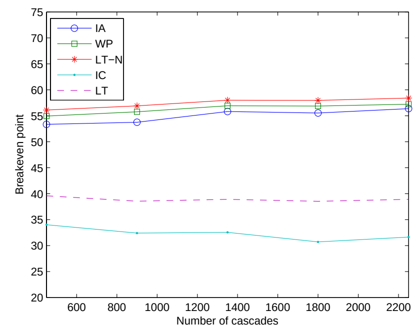
Observations. Figure 3.4(a) and Figure 3.4(c) present MSE in estimating the IC-N model parameters. The variation in quality of estimated parameters is plotted against the number of cascades increases for the scale free and kroneker graph. It can be observed that as the number of cascades increases MSE decreases. Further MSE obtained using the WP is slightly lower than that of IA. Additionally, it can be noted that, lower product quality q requires more number of cascades to achieve a given level of MSE. The probable reason is that there are relatively lesser number of activation per cascade when q is smaller, thus it requires more number of cascades to learn the positive activation probabilities. Similar results can be observed for LT-N model, when parameters are estimated using the LT-N influence function. Results for the two types of network are shown in Figure 3.4(b) and Figure 3.4(d).

3.7.3 Prediction Accuracy

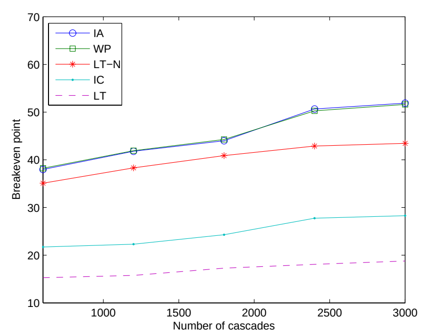
We compare the polarity sensitive influence functions with the polarity-insensitive influence functions IC and LT. For the same, we randomly divide the cascades in five parts and use every four parts as the training set and remaining part as the test set. The parameters of each of the influence functions are then learnt on the training set. The learnt parameters and activation state of one-hop-neighbor nodes are taken as input to predict the activation (either as positive or negative) of a node in test set cascades. Thus for both polarity-sensitive and polarity-insensitive influence functions, prediction target is same (node's activation). The prediction quality is then averaged over all the five test parts. The parameters for IA, WP, LT-N and LT are learnt using the method proposed in Section 3.6 while for IC we use Mosek [1] implementation of the state-of-art method Connie [70]. Connie provides a convex objective function to estimate the IC model parameters and thereby is guaranteed to give global



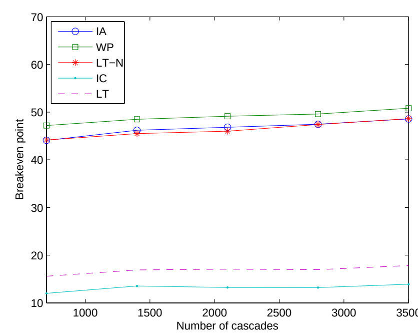
(a) IC-N Cascades, Scale Free Network, $q = 0.7$



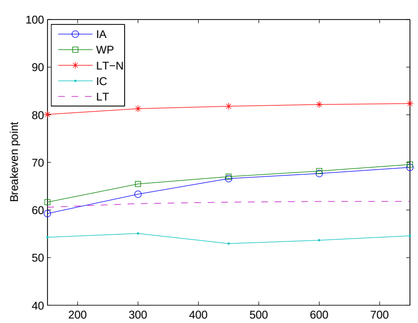
(b) IC-N Cascades, Kronecker Network, $q = 0.7$



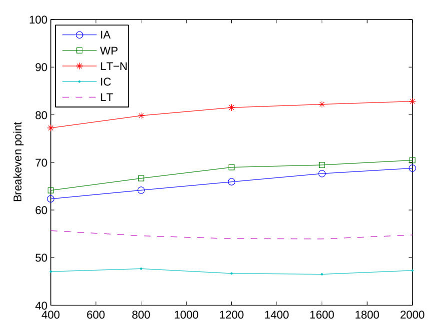
(c) IC-N Cascades, Scale Free Network, $q = 0.3$



(d) IC-N Cascades, Kronecker Network, $q = 0.3$



(e) LT-N Cascades, Scale Free Network, $q = 0.7$



(f) LT-N Cascades, Kronecker Network, $q = 0.7$

Figure 3.5: The variation of break even point (to predict the activation state given the neighbors activation status) as the number of cascades.

optimal solution.

Performance Measure. We assess the quality of prediction accuracy in terms of the break even point on the cross-validation set. The break even point is the point at which both precision and recall are equal. If tp is the number of true positives, fp is number of the false positives and tn is the number of true negatives, then precision is defined as $= \frac{tp}{tp+fp}$ and recall as $\frac{tp}{tp+fn}$.

Observations. The variation of break even point as the number of cascades for different settings of networks, cascades and q is shown in Figure 3.5. It can be observed that in all cases, ignoring the polarity of opinions significantly reduce the performance. Further, the performance of IC-N approximations WP and IA perform better than LT-N model when cascades are generated using the IC-N model. While the opposite is true when cascades are generated using the LT-N model. However, the performance difference between the two is much larger when cascades are generated using the LT-N model. This shows that IC-N based influence functions are not suitable when cascades generation follow the LT-N model. Additionally, it can be noticed that the performance of LT-N models does not increase by a large amount as the number of cascades increases. This shows that LT-N can learn the underlying the model parameters with fewer number of cascades.

3.8 Experiments on Real Data

In this section, we evaluate the proposed polarity sensitive influence functions in terms of their ability to predict the future activations on real datasets, Flixster and Epinions. In addition to comparing IA, WP and LT-PS with polarity insensitive IC and LT model, we also compare them with other 2 base line methods. Base line 1 and 2 are the global influence functions, where activation of nodes depend on the number of active users (not necessarily neighbors) and their ratings.

1. **BL1.** The base line 1 (BL1) sets the activation probability proportional to total number of active users. If n^+ is the total number of positively active nodes and n^- is the number of negatively active nodes at time $(t_u^c - 1)$ then, BL1 sets $p(s_u^c \in \{+, -\}) \propto (n^+ + n^-)/n$.

2. **BL2.** The base line 2 (BL2) respects the polarity of opinions. It sets $p(s_u^c = +) \propto (n^+ - n^-) \cdot (n^+ + n^-)$ if $n^+ > n^-$ and $p(s_u^c = -) \propto (n^- - n^+) \cdot (n^+ + n^-)$ if $n^+ < n^-$. Thus, any improvement over BL1 and BL2, can be attributed to the influence from friends.

3.8.1 Experimental Setup

For all IC and LT based methods, we modify the users' network $G(V, E)$ by adding external node for every node in the graph. This is required to account for the tendency of users to become active irrespective of the social influence. Each external node has positive polarity and has only one connection and that is to the node it corresponds to. Thus the modified network $G'(V', E')$ has twice the number of nodes of in original network $G(V, E)$ and $|V|$ additional edges. Further, to consider the global influence on nodes, for every cascade, we add a node which influences other nodes according to a fixed probability. For polarity insensitive functions, this probability is defined as in BL1. While for polarity sensitive functions, this probability is same as in BL2.

In Epinions dataset, we select 500 most active users for each of the top 10 product categories. Here most active users are the set of users who have rated maximum number of articles while

Dataset	BL1	BL2	IC	LT	IA	WP	LT-PS
Flixster	28.68	32.08	41.03	41.62	42.59	42.44	44.63
Epinions	17.5	21.16	22.55	22.33	24.19	24.44	27.57

Table 3.4: Breakeven point for Flixster and Epinions

the top product categories are the categories with maximum number of review article ratings. For each product category, the set of information cascades is selected by stratified sampling on the cascade length and number of negative ratings. On an average 3,000 cascades were selected per product category.

Performance Measure. All the different influence functions are evaluated in terms of break even point on the test data. For each dataset 20% of the cascades are randomly selected as the test set and rest of the 80% cascades are used as the train set.

3.8.2 Observations

The results on Flixster and Epinions dataset are presented in Table 3.4. For Epinions data, every entry shows the average breakeven point over top 10 its most popular product categories. The breakeven point for every category is shown in Table 3.5. It can be observe that the break even point is lowest for BL1. By incorporating the polarity of opinions, BL2 improves over BL1 on every dataset. This shows that considering polarity of opinions is very important, even if we just consider global influence. Next we can notice that, both IC and LT outperform BL1 by incorporating the friends' influence. Though, in most cases, IC and LT improves over BL2, in some cases the improvement is not statistically significant; for example in case of Epionion's category 1, 2 and 3. Further, there is not much difference in the performance of IC and LT.

Next, we can note that IA and WP models improve the per-

Category	BL1	BL2	IC	LT	IA	WP	LT-PS
1	26.07	31.02	29.55	29.65	32.08	32.81	36.62
2	16.07	19.91	20.85	20.54	21.86	22.11	24.96
3	18.48	24.55	25.44	25.29	27.60	27.68	30.64
4	17.00	21.91	23.30	23.31	24.86	25.12	28.45
5	13.57	17.47	19.88	19.78	21.01	21.05	24.55
6	21.16	21.14	21.96	21.59	23.25	23.59	27.17
7	21.17	23.72	24.59	24.29	26.04	26.38	29.50
8	15.10	18.37	18.70	18.62	19.99	20.08	23.47
9	13.25	18.27	21.00	20.03	23.28	23.30	24.69
10	13.44	15.28	20.25	20.25	22.00	22.27	25.70

Table 3.5: Breakeven point on Epinions Categories

formance by 1.5% on Flixster and 2% on Epinions as compared to the IC model. It highlights the fact that negative opinions do not spread contiguously in the social rating networks such as Flixster and Epinions. Recall that, in the IA and WP model, a node gets activated only when it is influenced by the positively active neighbor. However, in the IC model, a node can get activated by any (positively or negatively) active neighbor.

Among all the models, LT-PS achieves the best prediction accuracy by accurately modeling the behavior of polarity of opinions. In LT-PS model, when some neighbors of a user are positively active and some negatively active, then the influence of two kind of polarities cancel out each other, and thereby reduces the overall probability of getting activated.

In summary, we observe that accounting for the asymmetry in propagation patterns of two kind of opinions improve the prediction accuracy. Further, the presence of two kind of opinions work against each other and thereby reduces the total probability of activation. This is unlike the competitive information models where the two kind of products compete in the network for adoption [7, 8].

3.9 Summary

In this work, we have extended both LT and IC information flow models to incorporate the polarity of opinions and have studied the impact of negative opinions with respect to them. Experimental results show that the proposed polarity-sensitive influence functions are able to predict the activation state of nodes more accurately. Moreover, LT-N outperforms all, owning to its unique property to reduce the overall influence on a node when both kinds of opinions are expressed by its neighbors. Additionally, our analysis on the Epinions data set clearly shows that the influence of negative opinions is much stronger than the positive opinions, however they remain local near the point of origination and are not echoed by the influenced nodes.

□ **End of chapter.**

Chapter 4

Impact on Posterior Evaluation

4.1 Introduction

Several probabilistic information flow models [45] have also been developed to mimic the way information spreads in a social network. They attempt to predict the probability of a user to adopt a product given its friends' recommendations. The underlying belief is that social recommendations increases the user's trust on the product and thereby increases the probability of the user to adopt it. For example, positive friends reviews about a book encourages us to read it.

However most of the existing information flow models have largely ignored the effect of social opinions on the **posterior** users evaluation of products, i.e., the opinion the user form after experiencing the product. They either assume that the expressed opinion is same as the influencing opinion [45] or they are assumed to depend strictly on the product quality [15]. However many times, user's evaluation of the product, is not completely independent of her social circle and she tends to conform with social opinions. For example, a user reads a book and does not like it much. However lots of friends praise it and call it a really insightful book or "5/5", then this might change the user's opinion slightly and user might rate the book as "3". Had she not interacted with her friend, she might have given a rating of "2". This behavior usually arise because of the presence of **social pressure** and the **innate difficulty in providing an absolute numerical rating** to a product [42]. In such cases, social opinions can act as a reference rating and calibrate the user ratings such that they are not very different from the prevalent social opinion. We call this behavior as **social conformity** and the users who changes their rating as *social conformers*. Recently, this effect has been shown to exist on Goodreads and Douban [41].

Understanding this behavior is important not just from the point of curiosity, but it is also crucial in improving the accuracy of personalized recommender systems and in developing better information flow models. The recommendation systems can boost the quality of recommendation by removing the social conformity bias, thus making the recommendation better tailored to users' preference. While the information flow models can more accurately predict the further information cascade by accurately predicting the users' opinions.

However, it is a very difficult task to quantify the social conformity as for a given product we *never get to know the two ratings, one under the social influence and one without it*. All that is known is a single opinion expressed by the user. Thus, the key challenge is to identify what component of any rating corresponds to the user's preference and what component corresponds to the social conformity.

In this chapter, we propose a novel formulation for the users' ratings which explicitly considers the social conformity. The proposed formulation represents every user rating as a function of social conformity and social opinion along with user's preference and item's characteristics. The social conformity **down-weighs the user's preference** such that as the number of influential friends increases, the user's rating become more **similar to the social opinion**. The model parameters provide an intuitive interpretation of the social conformity behavior which reflect the degree a user conforms to her friend. Using this model, we explore the presence of social conformity on a real large scale dataset, Goodreads[1]. To our surprise, the results indicate that approximately **76% socially active users tend to conform to their friends** to some degree. We also find that social opinions make the user ratings more positive than negative.

The key contributions of this work are following.

1. We propose a novel formulation for user ratings that explicitly considers the social conformity. The proposed model improves the predict accuracy of users' ratings by more than 2% in presence of social influence. The learned social conformity parameters are also verified by qualitatively comparing the discovered most influential users with the authoritative and most socially active users.

2. Based on the learned users' degree of conformity, we find various interesting patterns on Goodreads that underline the impact of social conformity.

The rest of the chapter is organized as follows. Section 4.3 presents the proposed approach to quantify the social conformity. The proposed approach is verified in Section 4.4. The patterns of conformity are then analyzed in depth in Section 4.5. Finally we conclude the chapter with potential future work directions in Section 4.6.

4.2 Related Work

Several models for information diffusion have been proposed to mimic the way information or technological innovations propagate via word-of-mouth publicity. Among them, LT and IC

[1] http://www.goodreads.com/

models are the most popular ones [45]. Both LT and IC models have been extended for the competitive information diffusion, where more than one products compete within the social network for adoption [7, 8]. However like their predecessors, they also assume that the influenced user's opinion is same as the influencing opinions. Chen et al. [15] first time modeled the negative opinions as being emerged from the product faults. They extended the IC model to explicitly incorporate the emergence of negative opinions. However they assume that if influencing opinion is positive, then the influenced user's opinion is strictly governed by the product quality. While if influencing opinion is negative, then influenced user also holds negative opinion against the product.

Huang et al. [41] are one of few who have studied the effect of social influence on the user's product evaluation on large scale data. They design statistical tests and show that social recommendations tend to make improve our posterior evaluation of the product. However, they do not study the impact of the social opinions on predicting the user ratings.

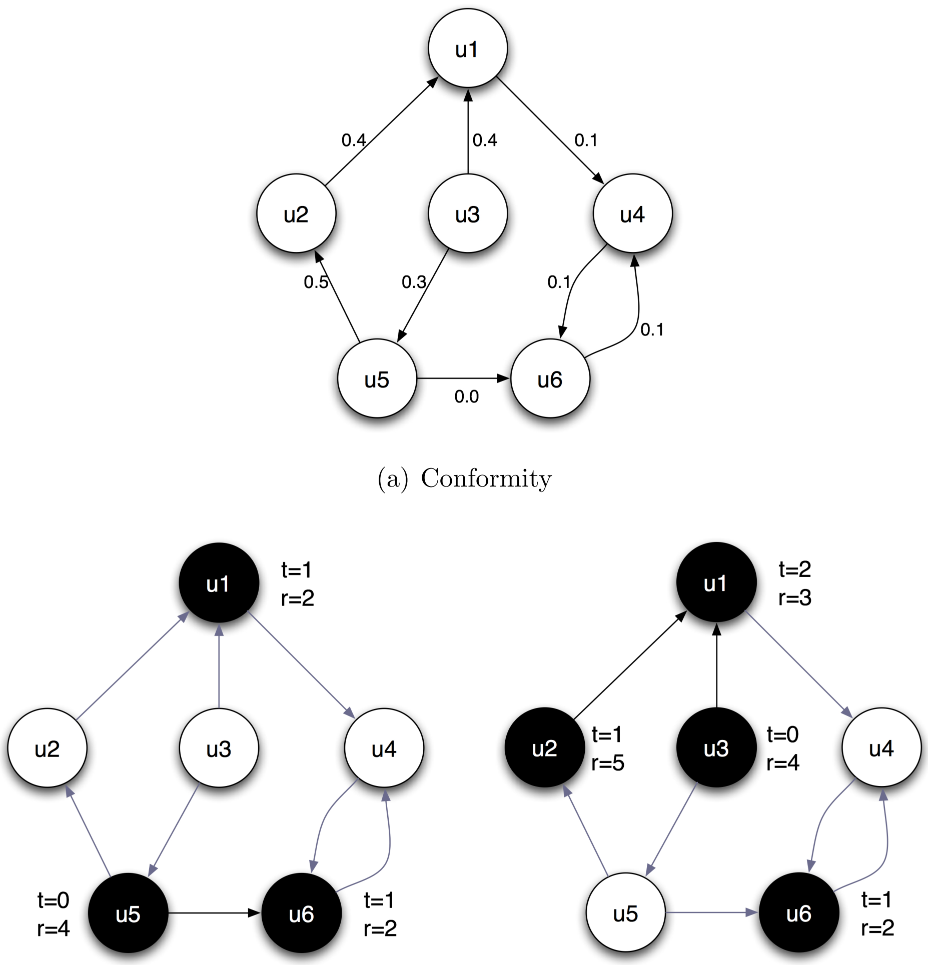
A different line of work which is also very closely related to this work is the collaborative filtering based recommendation systems. The task of these systems is to predict the rating of a user for an item, given the user-item rating matrix. The user-item rating matrix contains product ratings given by different users. These recommendation systems can broadly be classified in two categories: neighbor-based approaches and model-based approaches. Neighbor-based approaches predict a user's rating either based on the past user's ratings for similar products [21, 63, 82] or based on the ratings given to the item by similar users [9, 44]. The model based approaches work by learning a compact model from the user-item rating matrix [64]. These approaches offer large scale efficiency as they do not need to keep the entire rating matrix in memory, once the model is trained. This category includes the clustering model [49], the aspect model [40], the latent factor model [11], the Bayesian hierarchical model [92], the ranking model [64] and low-rank matrix factorization methods [50, 81]. The low-rank matrix factorization methods are the most popular once among them, because of their ability to deal with large scale data efficiently. Probability Matrix Factorization model is one of the low-rank matrix factorization and is regarded as the state of art model [81]. The underlying idea is that only a small number of factors govern the users' preferences and items characteristics. To learn the model, the method approximates the user-item matrix with its low rank matrix factorization.

The recommendation systems have been extended to account for the social *homophily*. Most of the social recommendation approaches leverage the social homophily to widen/regularize the users preference based on their friends [43, 66]. Contrary to these homophily based approaches, our proposed extension of the PMF model considers change in user ratings of a particular item due to the social pressure.

Many times social influence can be confused with homophily or other confounding factors which can also result in correlation of actions in space and time. In this paper, we use PMF model to best characterize the latent user preference and then look for the presence of the social conformity.

4.3 Ratings under social conformity

Here, we study the change in user ratings caused by her social circle. It is different from the existing information flow models, which study the effect of friends on the product adoption probability. Also different from the homophily based recommendation systems, we focus on the change of ratings at item level. Contrary to our approach, homophily based recommendation methods, try to learn user preference based on their friends.



(a) Conformity

(b) item 1

(c) item 2

Figure 4.1: An example of influence of social opinions on users' ratings on two very similar items. User u_1 behaves like a social conformer. While the ratings of user u_6 are independent of its friends u_5.

Many times users tend to give different ratings to products in social settings. Sometimes this user behavior is a result of social pressure and sometimes this is caused by the innate difficulty in providing an absolute rating to the product [42]. In both cases, the social opinions act as a reference rating for the product and change the user rating such that it is not very different from them. Figure 4.1 shows an example scenario. There are 6 users with their social connections shown by edges between them. User ratings and the time of ratings are shown for two very similar items/products in Figure 4.1(b) and Figure 4.1(c). Dark nodes represent the users who have rated the item. In this example, u_1 behaves like a social conformer while the behavior of u_6 is independent from its social circle. In case of item 1, u_1 gives a rating of 2, when none of its friends have rated the item before it. However, it gives a rating of 3 to item 2 when two of its friends give high ratings to the item. Here we can see that u_1 changes its rating in presence of social opinions. On the other hand, user u_6 gives same rating to both items whether its friend u_5 have rated the item (in case of item 2) or not (in case of item 1). Thus, for social conformer like user u_1, we find differences in its ratings of items even when the two items are very similar. Figure 4.1(a) shows the actual degree of conformity of each user with her friends. These values are unknown and we aim to uncover them.

4.3.1 Problem Definition & Notations

Notations. Let $G = (V, E)$ be a directed graph where every node $u \in V$ corresponds to an individual in the social network and edge $(u, f) \in E$ exists if node f is a friend of u. Nodes in the network, post their opinions related to several items $\mathcal{I}$, where every opinion takes a real number and is called rating. The user's u rating for item i is represented by $r_{u,i}$ and time of the rating is denoted by $t_{u,i}$.

Definition 4 *Let the set of* **active neighbors** *who have posted their rating for item i before user u be* $\mathcal{A}(u, i)$. *Formally* $\mathcal{A}(u, i) = \{f \in V | (u, f) \in E, t_{f,i} < t_{u,i}\}$.

Problem Definition. Predict the rating $r_{u,i}$ for item i given by user u, given the set of active neighbors $\mathcal{A}(u, i)$, the preferences of the user u and the characteristics of the item i.

4.3.2 Conformer's Ratings

In this section, we propose a new rating function which also accounts for the social conformity of users along with their preference and characteristics of items. Lets represent $r^0_{u,i}$ be the **inner rating** given by user u to item i, had there been no social influence. The $r^0_{u,i}$ strictly a function of user's preference and item's characteristics. The social opinions calibrate this inner ratings of users such that they are not very different from the prevailing social opinions. To account for such social behavior, we propose the following social conformity based rating model.

$$\begin{aligned}\hat{r}_{u,i} &= r^0_{u,i} + conf \cdot (social\_opinion - r^0_{u,i}) \\ &= \Big(1 - conf\Big) r^0_{u,i} + conf \cdot social\_opinion, \qquad (4.1)\end{aligned}$$

where both $conf$ and $social\_opinion$ are the social functions and depend on the set of active friends $A_{u,i}$. $conf$ defines the degree by which a user conforms to the social opinion. $social\_opinion$ is the social opinion about the item i before the user u rates it, i.e., at $(t_{u,i} - 1)$ time. The rewritten form in Eq. (4.1) can also be seen as down-weighing the user's personal preference and giving higher weight to the friends' opinions. That is, if the user u has extremely high degree of social conformity then user u will change her rating such that it becomes same as the social opinion.

We expect the degree of conformity $conf$ to take large values as the number of friends who have already rated the item increases. This phenomenon is known as the **bandwagon effect** in social sciences [32]. According to the bandwagon effect, as the number of individuals who believe in something increases, others tend to disregard their own opinions and also “hop on the bandwagon”. That is, the social conformity is directly proportional to the number of friends with similar opinions. Thus, $conf$ can be written as constant times the number of active friends $A_{u,i}$, because only active friends (friends who have posted their rating of item i before user u) can affect the user's rating for the item. However, one can expect that users do not conform to all their friends equally. The friends who are regarded highly in the user's eyes, tend to affect their rating more. Hence, we introduce a parameter $\kappa_{f,u}$ corresponding to every user and her friend pair. This parameter defines the degree by which user u conforms to the rating of its friend f. As the number of friends with high $\kappa_{f,u}$ increases, the $conf$ can be expected to increase. Thus, we write

$$conf = \sum_{f \in \mathcal{A}(u,i)} \kappa_{f,u}. \tag{4.2}$$

Since $conf$ can take maximum value of 1, we constraint $\kappa_{f,u}$ such that $\sum_f \kappa_{f,u} \leq 1$. Such linear forms of social influence have also been used in Linear threshold model [45] where the adoption probability of a product depends linearly on the active friends' influence.

We write the $social\_opinion$ as the sum of friends' opinions weighted according to $\kappa_{f,u}$. This is because the opinions of friends with high $\kappa_{f,u}$ affect the user's rating by the most amount. Thus, we have

$$social\_opinion = \frac{\sum_{f \in \mathcal{A}(u,i)} \kappa_{f,u} \cdot r_{f,i}}{\sum_{f \in \mathcal{A}(u,i)} \kappa_{f,u}}. \tag{4.3}$$

Substituting values of $conf$ and $social\_opinion$ from Eq. (4.2) and Eq. (4.3) in Eq. (4.1), we get

$$\hat{r}_{u,i} = \Big(1 - \sum_{f\in\mathcal{A}(u,i)} \kappa_{f,u}\Big) r^0_{u,i} + \sum_{f\in\mathcal{A}(u,i)} \kappa_{f,u} * r_{f,i} \tag{4.4}$$

To represent the user's inner rating $r^0_{u,i}$, we user one of the state of art recommendation models, Probability Matrix Factorization (PMF) method [81]. PMF model uses a small number of factors to represent the preference of users and item characteristics. The preference of users $q_u \in R^K$ and item characteristics $p_i \in R^K$ are represented by low dimensional vectors in latent space of dimensionality K. Then every rating is written as

$$\hat{r}^0_{u,i} = \mu + b_i + b_u + q_u^T \cdot p_i, \tag{4.5}$$

where μ is average user-item rating, b_i is item bias and b_u is user bias.

By substituting $r^0_{u,i}$ in Eq. (4.4) by $\hat{r}^0_{u,i}$ from Eq. (4.5), we finally have

$$\begin{aligned}\hat{r}_{u,i} =& \Big(1 - \sum_{f\in\mathcal{A}(u,i)} \kappa_{f,u}\Big)(\mu + b_i + b_u + q_u^T \cdot p_i) \\ &+ \sum_{f\in\mathcal{A}(u,i)} \kappa_{f,u} * r_{f,i}.\end{aligned} \tag{4.6}$$

4.3.3 Parameter Estimation

To learn the model proposed in previous section, we need to estimate the model parameters b_i, b_u, q_u, p_i, $\kappa_{f,u}$. Like the PMF model, we construct the objective function such that it minimizes the square of difference between observed user rating $r_{u,i}$ and estimated rating $\hat{r}_{u,i}$. Additionally, all parameters are regularized to avoid over fitting on the train dataset. Thus, our objective function is

$$\begin{aligned}
\min \sum_{u,i} & (r_{u,i} - \hat{r}_{u,i})^2 \\
& + \lambda_1 \Big(\sum_u ||b_u||^2 + \sum_i ||b_i||^2 + \sum_u ||q_u||^2 + \sum_i ||p_i||^2 \Big) \\
& + \lambda_2 \sum_u \sum_f \kappa_{f,u}^2 \\
\text{subject to } & \kappa_{f,u} \geq 0 \ \forall u, f; \ \sum_f \kappa_{f,u} \leq 1 \ \forall u,
\end{aligned} \tag{4.7}$$

where λ_1 and λ_2 are the hyper-parameters which control the amount of regularization. The objective function is minimized by using the alternating minimization. In every first alternating step, we minimize the function with respect to the PMF model parameters b_i, b_u, q_u and p_i, using the steepest gradient decent method. Then in the second alternating step, we minimize the function with respect to $\kappa_{f,u}$. Given the estimate $\hat{r}^0_{u,i}$ from first step, the objective function in the latter step can be written as the sum of small subproblem, each corresponding to one user. That is

$$\begin{aligned}
\min \sum_u & \Bigg(\sum_i \Big(r_{u,i} - \Big(1 - \sum_{f \in \mathcal{A}(u,i)} \kappa_{f,u}\Big) \hat{r}^0_{u,i} - \sum_{f \in \mathcal{A}(u,i)} \kappa_{f,u} r_{f,i} \Big)^2 \\
& \quad + \lambda_2 \sum_f \kappa_{f,u}^2 \Bigg) \\
\text{subject to } & \kappa_{f,u} \geq 0 \ \forall u, f; \ \sum_f \kappa_{f,u} \leq 1 \ \forall u.
\end{aligned} \tag{4.8}$$

Since the set of parameters $\kappa_{f,u}$ of every subproblem are different from the others, the objective function can be minimized by minimizing each of the sub problems separately. Thus, each of the sub problem can be minimized in parallel, using the gradient descent method.

Time Complexity. The computation time required to minimize the objective function is equal to the number of descent steps times the time required to calculate the gradient in every step. The time required to calculate the gradient in first alternating step is same as that of PMF and is $O(\rho_R)$, where ρ_R is the number of nonzero entries in the training set. In the second alternating step, gradient calculation for a subproblem corresponding to a user u, takes $O(d_u m_u)$, where d_u is the degree of the user and m_u is number of items rated by the user. Thus, the total complexity of gradient calculation in the second alternating step is $O(ndm_{max})$, where n is number of users, d is the maximum user degree and m_{max} is the maximum number of ratings by any user. This complexity analysis shows that our proposed approach is very efficient and can be used with very large datasets.

The learned $\kappa_{f,u}$ parameters give an estimate on the social conformity. The higher value of $\kappa_{f,u}$ indicates that user u tends to conform friend f while zero value indicate that the rating of user u are independent of its friends f.

4.4 Evaluation

Before exploring the characteristics of the social conformers, we evaluate the goodness of the proposed approach. Since there is no ground truth available about the social conformity of users, it is not possible to directly test if the users with nonzero $\kappa_{f,u}$ are truly conformer. Hence, we indirectly evaluate the approach by using two criterions. First criterion is to measure the improvement in the rating prediction accuracy attained by the proposed rating model. Second criterion is to qualitatively analyze the list of users who affect their friends rating by most amounts. These users can be considered as the social influencer and are expected to have high authority. For example, book authors, most popular users on Goodreads. For brevity, we refer to our proposed Conformity Rating Model as **CRM**.

In this section, we seek to answer the following questions.

- Does incorporating the social conformity, help in improving the prediction accuracy of users' ratings?
- What is the effect of hyper-parameters λ_2 and K on the accuracy?
- How well does the list of social influencer, match with the users with high authority on Goodreads?

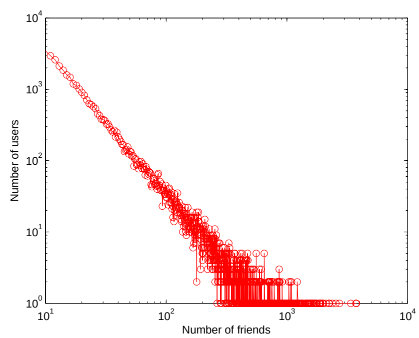
4.4.1 Goodreads Dataset

We use the Goodreads dataset to explore the impact of social opinions on users' ratings. Goodreads is an online social books cataloging website. It permits users to register books and provide their ratings. Also, it allows users to add friends and to view their reading list. When a user rates a book, all her friends get notification. Books' ratings take values from 0 to 5 stars, with 5 stars being the best. We use the dataset crawled by authors in [41]. The dataset contains the friendship graph of users and user item ratings. Every rating is also associated with its time of rating (available at the date granularity).

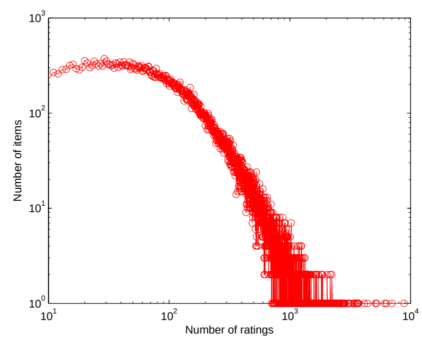
Users	Items	Edges	Ratings
55,654	120,703	1,757,568	9,462,016

Table 4.1: Data Statistics of Goodreads data

We filter the items and users such that each item has at least 10 ratings and every user has rated at least 5 books and have at least 10 friends. Further, users who have rated books on less than 5 different dates are also removed from the dataset. This is to make sure the selected users are active users of Goodreads. The statistics of the two datasets is presented in Table 4.1. The user-item rating matrix is very sparse and has density 0.0014. The degree distribution (number of friends) and the distribution of popularity of items (number of times an item is rated) are shown in Figure 4.2(a) and Figure 4.2(b). The degree distribution follows the power law. The average degree of users is 31.5 and the average popularity of items is 170.



(a) Degree Distribution



(b) Popularity Distribution

Figure 4.2: Goodreads data

In addition to this data, we also crawl the profile pages of all the selected 55,654 users. Users who have also authored books are marked as the authors. In our dataset, we have 5,078 authors which is approximately 9% of all the users. These users are typically seen as the users with high authority among their friends.

		Test ratings	PMF	CRM
$K = 10$	All ratings	1,789,663	0.8556	**0.8520**
	Ratings with $conf > 0.1$	208,852	0.8476	**0.8254**
$K = 5$	All ratings	1,789,663	0.8472	**0.8441**
	Ratings with $conf > 0.1$	196,855	0.8471	**0.8280**

Table 4.2: RMSE when $\lambda_1 = 1, \lambda_2 = 0.1$

4.4.2 Prediction Accuracy

Here we use the CRM to predict the users' ratings and compare its accuracy with the PMF method. The idea is to check if considering social conformity can help in improving the prediction accuracy of user ratings. For the same, we train the model parameters of both the PMF and the CRM on a train set and calculate their performance on a test set. The train and test sets are constructed by splitting the user-item ratings in 4:1 ratio. Specifically, for every item, we select a user rating with probability 0.8 and put it in the train set. Rest of the ratings are used to construct the test set.

Performance Measure. We use the Root Mean Square Error (RMSE) metric to measure the prediction accuracy. RMSE is defined as

$$RMSE = \sqrt{\frac{\sum_{u,i}(r_{u,i} - \hat{r}_{u,i})^2}{m}}, \tag{4.9}$$

where m is the number of ratings in the test set.

Observations. The results are presented in Table 4.2. The first row presents RMSE value over all test ratings when $\lambda_2 = 0.01, K = 10$. It can be seen that RMSE improves by more than 0.4% when social conformity is taken into consideration. The second row shows RMSE results when $conf > 0.1$. The reason is to narrow the focus to only those ratings who are potentially affected by the social influence. There are more than 11% such

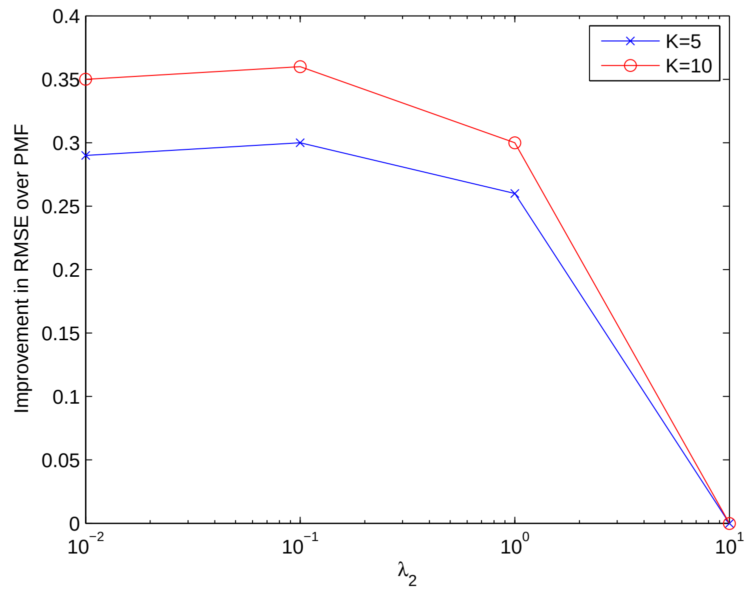


Figure 4.3: Improvement in RMSE over PMF, as λ_2 changes, $\lambda_1 = 1$

ratings in the test set for $K = 10$. It can be noticed that the RMSE improves by more than 2.6% in such cases. This demonstrates the importance of considering the social conformity and effectiveness of our proposed formulation. Similar results can be observed when $K = 5$ though improvement is slightly less as compared to $K = 10$.

Sensitivity to λ_2 and K. Next we study the impact of λ_2 (hyper-parameter to control the regularization) and K (dimensionality of the latent space) on prediction accuracy. It can be noticed from Figure 4.3 that CRM consistently outperforms over the PMF model. Further, the performance improvement is larger when $K = 10$. Note that, for the PMF model, the RMSE value at $K = 10$ is smaller than the RMSE value at $K = 5$. (as can be seen in Table 4.2). This might be because a large value of latent space dimensionality K, can lead to over fitting on the training set. However, larger improvement in CRM performance when $K = 10$, underlines the robust performance of CRM.

Effect of λ_2 on the RMSE values is as per the expectations. The performance gets hurt if λ_2 is too large (≥ 1) or when it is too small (≤ 0.01). Higher value of λ_2 forces the selection of small social conformity factors $\kappa_{f,u}$ and thereby under fits the model. While very small value leads to over-fitting on the training set. The best RMSE value is achieved when $\lambda_2 = 0.1$, though performance is reasonably robust around this value.

We also compare the proposed approach with one of the state of art homophily based recommendation system proposed in [43]. However we did not observe any improvement in RMSE over the PMF model. We suspect that the strength of this method lies in solving the cold start problem. However, in our dataset, we do not have any user with less than 10 ratings.

4.4.3 Influencers Quality

Next we analyze the quality of the learned $\kappa_{f,u}$ parameters which denote the conformity of user u to its friend f. Since, there is no direct ground truth present to evaluate them; we instead qualitatively analyze the quality of most influential users. In our setting, the users who have maximum effect on their friends' ratings can be considered as the social influencers. We define social influence of such users as $\sum_f \kappa_{u,f}$. Notice that it is defined by reversing the conformity $\kappa_{f,u}$ direction.

To evaluate the goodness of the quality of influencers found by CRM, we rank the users based on their social influence and study the authority and the degree of top most influential users. We expect that the top influencers should have very high authority and very higher number of social connections. The average degree (number of friends) of top x most influential users as function of x is plotted in Figure 4.4. It can be noticed that top 200 influential users have the highest average degree and as we consider more and more top users as the influential ones, their

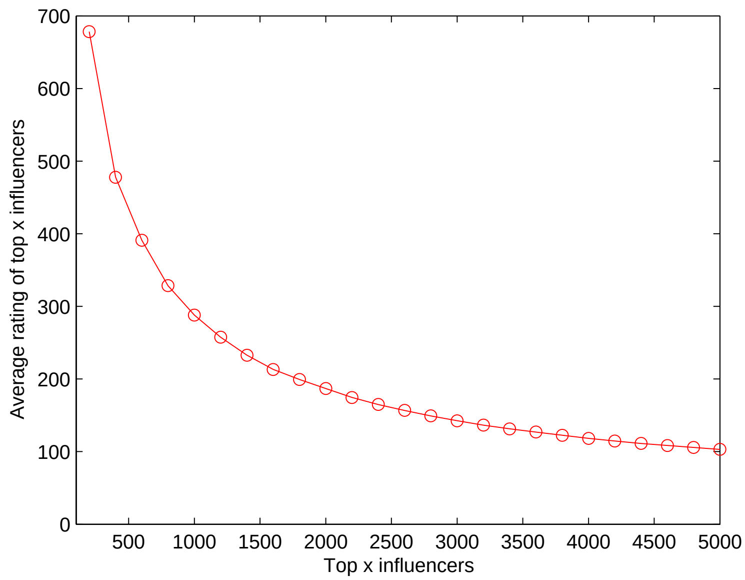


Figure 4.4: Average degree of top influencer

average degree starts to fall.

We also validate the authority of the influential users by checking if they have also authored the books. This is a reasonable criterion as the book authors have higher perceived authority among their friends. The plot of percentage of authors among the top x most influential users is shown in Figure 4.5. More than 45% authors appear in the top 200 influencers. It can be noted that the percentage of authors falls sharply in the beginning and there are only 12% authors among the top 5000 users. The entire dataset has approximately 9% authors. Thus, we can see that as we keep widening the value of x, the ratio of authors to non-authors approaches to the ratio of entire dataset.

The above two results clearly show that the proposed method is effectively able to give high rank to the influencers with high authority and high degree. Few of the most influential users' profiles are shown in Figure 4.4.3 along with their rank. It can be seen that all these authors are very socially active and have

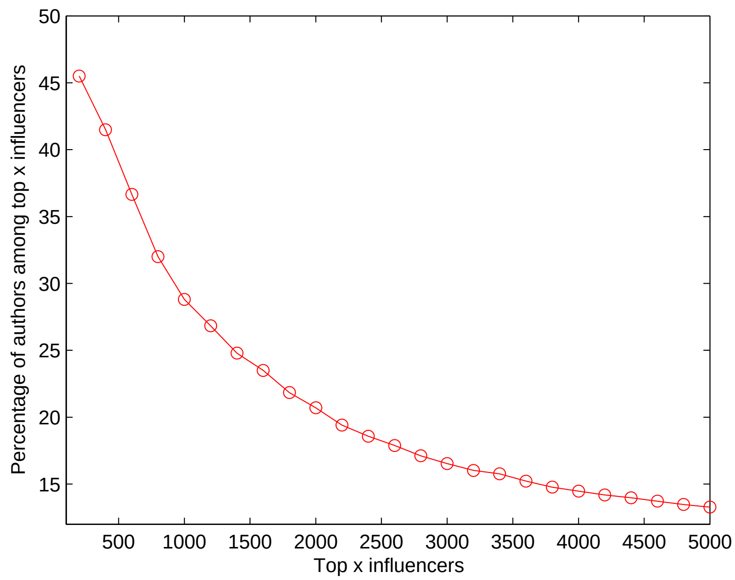


Figure 4.5: Percentages of authors among top influencer

very high authority.

4.5 Social Conformity Analysis

Having verified CRM model, in this section we explore the nature of social conformity. We seek to answer following the questions.

1. How many users conform to their friends?
2. How many friends to a user conform to?
3. By how much amount the user's ratings get calibrated because of the social opinions?
4. When does the conformity prevail the most?

We define the degree of conformity of a user u to be $\kappa_u = \sum_f \kappa_{f,u}$. The distribution of user conformity κ_u is presented in

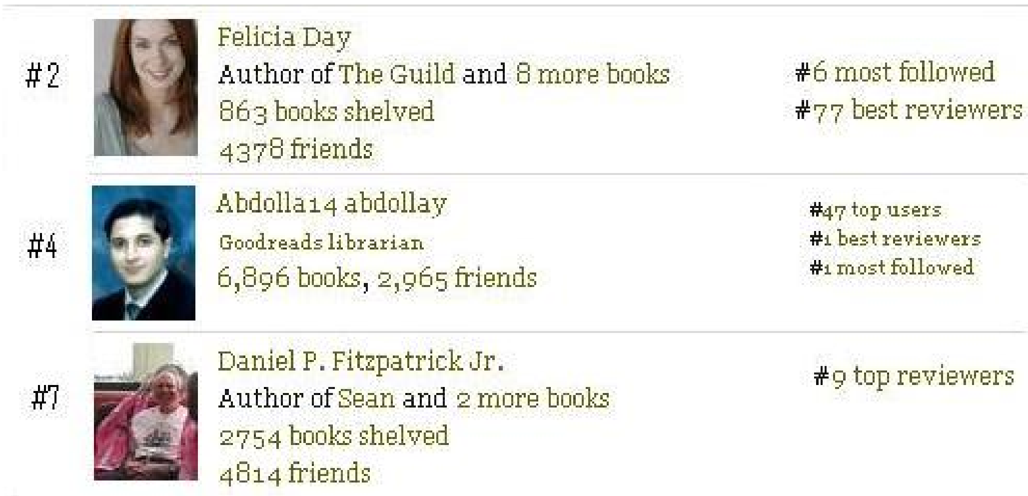
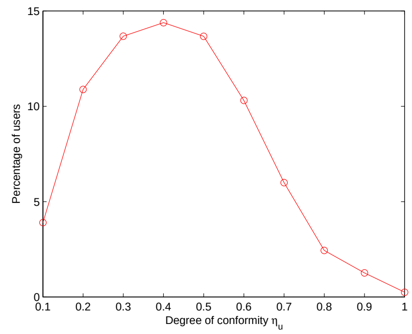


Figure 4.6: Users appeared in the top influencers list

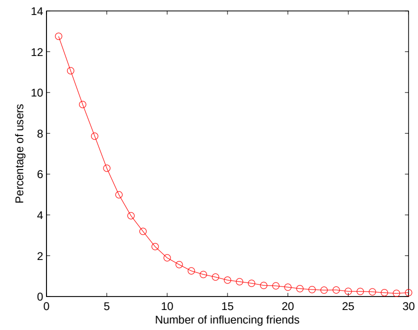
Figure 4.7(a). In this plot, we have considered κ_u intervals of size 0.1 and labeled each of the interval with its highest value. Then for every interval, we plot the percentage of users whose κ_u lie in that interval. It can be noticed that more than 76% users have $\kappa_u > 0$. Among these users, most of them belong to the 0.2 to 0.6 interval. That is, most of the users in the network display some sort of conformity to their friends.

Next, we plot the distribution of number of friends a user conforms to. Results are presented in Figure 4.7(b). On x-axis, we plot the number of influencing friends x and on y-axis the percentage of users with x number of influencing friends. It can be seen that, when x is small, the number of users with x influencing friends falls sharply as x increases and becomes less than 2% when $x > 9$. Then it starts to decay at much slower rate. Further, most of the users conform to only one friend and less than 9% users conform to more than 14 friends.

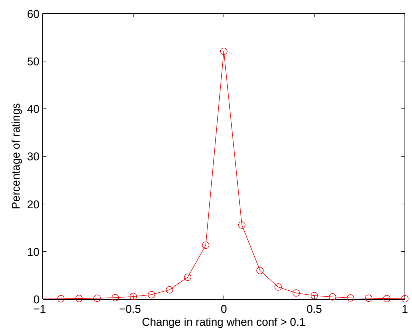
To understand by how much amount the conformity can change a user rating, we select those ratings for whom $conf > 0.1$ at the time of rating. Recall that $conf$ is the sum of the



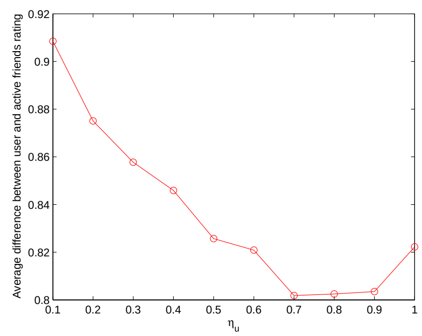
(a) Distribution of conformity



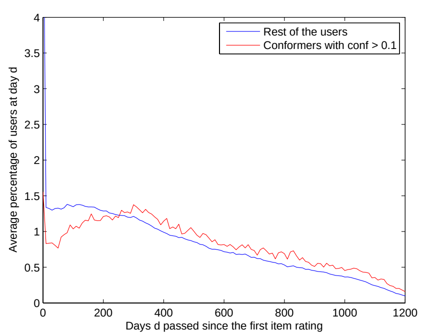
(b) Distribution of number of influencing friends



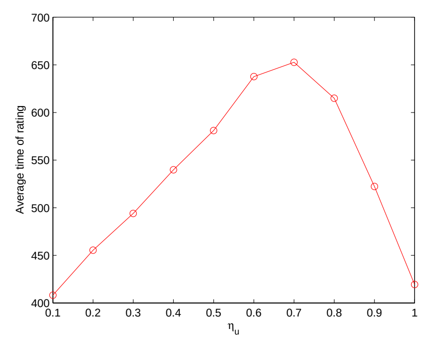
(c) Change in ratings caused by social opinions



(d) Average difference between user-item rating and average active friends rating



(e) Distribution of conformers as the item cascade unfolds



(f) Average time of users' ratings as κ_u varies

Figure 4.7: Patterns of social conformity

social conformity of a user to her active friend. There are more than 11% such ratings in the test set for $K = 10, \lambda_2 = 0.1$. For these ratings, we plot the distribution of change in ratings caused by the social opinions. For the same, we calculate the difference between the ratings predicted by the CRM $\hat{r}_{u,i}$ and $\hat{r}^0_{u,i}$. Results are presented in Figure 4.7(c). We can note that, most of the ratings change is between -0.5 and 0.5. Additionally, it is interesting that more ratings change by positive factor than by the negative factor. 15% rating changes by +0.1 amount while only 12% ratings changes to -0.1 amounts. This is similar to the results published in [41], where authors find that for recommended items, the distribution of user ratings shifts in positive direction.

We also check if there is any relation between conformers' ratings and their friends' rating. For the same, for every κ_u interval, we plot the average difference between user's rating and user's active friends' rating. Note that, the average is not weighted by $\kappa_{f,u}$ and also includes the opinions of active friends with $\kappa_{f,u} = 0$. Results are presented in Figure 4.7(d). It can be seen that the difference decreases linearly as κ_u increases with exception when $\kappa_u > 0.7$. In other words, users with high κ_u values behave more similar to their friends.

Next we explore, when do the ratings with social conformity appears during an item information cascade? An item information cascade is the life cycle of the item since the first rating is given to it. We want to understand if most of the ratings with social conformity appear towards the end of the information cascade? And how is their behavior different from other ratings? For the same, for each item, we find the percentage of social ratings with $conf > 0.1$ that appeared in between day d and day $d + 10$ since first rating is posted. Then, we calculate their average over all the items and plot them against day d where d is incremented in steps of 10. Similarly we plot the ratings with $conf \leq 0.1$. Results are presented in Figure 4.7(e). The ratings with $conf > 0.1$ are shown by red colored line while other ratings are shown by the blue line. It can be seen that two kinds of ratings follow very different patterns. The ratings with $conf \leq 0.1$ have maximum presence during the start of the information cascade and their percentage decays slowly as the time passes by. While the ratings with $conf > 0.1$ have relatively smaller presence at the start. As the time passes by their percentage increases and peaks at around 300 days passed. After that, their percentage falls with time and follows similar pattern as the other ratings. This is similar to the pattern observed by authors in [4], where authors discover on Yahoo! Go dataset that early adopters are typically trend setters and other users start adopting the product as social influence increases.

In general, we find that users with higher value of κ_u tend to post their ratings during the later part of the item life-cycle. The results are presented in Figure 4.7(f), where we plot the average time of rating users for each of the κ_u interval. It can be noted that as the κ_u increases, the average time of users in that interval increases linearly, with exception when $\kappa_u > 0.7$. We suspect such high values of κ_u are a result of over-fitting over the train data.

To summarize, we find the following patterns.

- More than 76% users display some sort of conformity with their friends. Additionally most of them do not conform to more than 9 of their friends.
- Owning to social influence, most of the ratings change by -0.5 to +0.5 amounts. However the effect is asymmetric. More ratings change by positive factor than the negative factor.
- The social conformity prevails the most approximately 300 days passed since the first item ratings. Then it starts to

decay slowly.

- Users with high degree of conformity κ_u post their rating relative late as compared to the users with smaller κ_u.

4.6 Summary

In this work, we propose a novel formulation for the user ratings CRM that explicitly considers the social opinions. Specifically, the CRM introduces a set of parameters to denote the social conformity of each user with her friends. The CRM is shown to be effective in both improving the prediction accuracy of user's rating and in accurately identifying the social influencers. The prediction accuracy is found to improve by more than 2% in presence of social influence. Based on the detailed analysis of learned social conformity, several interesting patterns have emerged. We find that most of the users show some degree of conformity with their friends; however they do not tend to conform to lots of their friends. The analysis also sheds light on the change in ratings caused by the social opinions. To our surprise, more number of user ratings become positive than negative in presence of social opinions. That is, our friends opinion makes our posterior evaluation of the product more positive, which is certainly a good news for the viral marketing strategy. Further, similar to the item product adopters, the users with high conformity tend to post their rating during the later part of the information cascade.

□ **End of chapter.**

Chapter 5

Summary & Future Work

5.1 Summary

In this work, we have reviewed the current state of research on the information flow in social networks. We study the impact of social opinions on both users' product adoption behavior and the posterior evaluation of the product. In particular, we observe that the negative and positive opinions have asymmetric impact on the product adoption behavior, while the positive opinions encourage users to purchase a product, the negative opinions discourage the people. The two kind of opinions also show different patterns of propagation, where negative opinions tend to localize near the point of originations on social rating networks. To capture this asymmetry, we propose models that explicitly accounts for this asymmetry and thereby, improves the accuracy of prediction of future opinions.

We also study the impact of social opinions on posterior product evaluation of the product. In order to quantify this impact, we extend the rating prediction models to explicitly account for social conformity along with the users' and products' characteristics. Using the proposed model, we discover various interesting patterns of social conformity on a large real dataset.

5.2 Future Work

We consider this work as a step towards understanding the social behavior of users and hope that it would help in developing better recommendation systems and information propagation models. We think following will be worthwhile directions for future work.

- Our work studies one of the aspect of the opinion. In real world, many times opinions are expressed at attribute level of products (e.g. great story but bad direction) or opinions are not binary (positive or negative) but a degree of liking/disliking is associated with the polarity (e.g. good movie, awesome movie). In such cases, considering only polarity may not be sufficient and we would like to extend our information flow model, to incorporate these aspects.

- Viral marketing is one of the main application of the information flow in social networks. We would like to explore algorithms where we can maximize the social influence in the presence of polarity of opinions.

- Currently, we have assumed that a person has same social influence over every topic. However, in general the person's perceived expertise over different topics, affect his social influence. In future, it will be worth incorporating the topical expertise in the information flow models.

- It will be interesting to study if there is any difference between the characteristics of the users, with most influence on product adoption and that of the users with most influence on product posterior evaluation.

- It has been observed that, as time passes by, friends become more similar to each other. Thus, given users current preference and similarity, one can build models to predict the user similarity in future. These predictions can be given as an input signal to the recommender systems.

☐ **End of chapter.**

Bibliography

[1] The MOSEK optimization software. http://www.mosek.com/.

[2] Reka Albert, Hawoong Jeong, and Albert-Laszlo Barabasi. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

[3] Aris Anagnostopoulos, Ravi Kumar, and Mohammad Mahdian. Influence and correlation in social networks. In *KDD*, 2008.

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

[5] Albert-László Barabási and Réka Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, 1999.

[6] Roy F. Baumeister, Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D. Vohs. Bad is stronger than good. *Review of General Psychology*, 5:323–370, 2001.

[7] Shishir Bharathi, David Kempe, and Mahyar Salek. Competitive influence maximization in social networks. 4858:306–311, 2007.

[8] Allan Borodin, Yuval Filmus, and Joel Oren. Threshold models for competitive influence in social networks. In *WINE*, 2010.

[9] John S. Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *UAI*, 1998.

[10] Jacqueline J. Brown and Peter H. Reingen. Social Ties and Word-of-Mouth Referral Behavior. *The Journal of Consumer Research*, 14(3):350–362, 1987.

[11] John Canny. Collaborative filtering with privacy via factor analysis. In *SIGIR*, 2002.

[12] Meeyoung Cha, Alan Mislove, and Krishna P. Gummadi. A measurement-driven analysis of information propagation in the flickr social network. In *WWW*, 2009.

[13] Deepayan Chakrabarti and Christos Faloutsos. Graph mining: Laws, generators, and algorithms. *ACM Computer Survey*, 38, 2006.

[14] Deepayan Chakrabarti, Yiping Zhan, and Christos Faloutsos. R-MAT: A Recursive Model for Graph Mining. In *SDM*, 2004.

[15] Wei Chen, Allex Collins, Rachel Cummings, Te Ke, Zhenming Liu, David Rincon, Xiaorui Sun, Yajun Wang, Wei Wei, and Yifei Yuan. Influence maximization in social networks when negative opinion may emerge and propagate. In *SDM*, 2011.

[16] Wei Chen, Chi Wang, and Yajun Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD*, 2010.

[17] Wei Chen, Yajun Wang, and Siyu Yang. Efficient influence maximization in social networks. In *KDD*, 2009.

[18] Wei Chen, Yifei Yuan, and Li Zhang. Scalable influence maximization in social networks under the linear threshold model. In *ICDM*, 2010.

[19] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[20] David Crandall, Jon Kleinberg, Dan Cosley, Siddharth Suri, and Daniel Huttenlocher. Feedback effects between similarity and social influence in online communities. In *KDD*, 2008.

[21] Mukund Deshpande and George Karypis. Item-based top-n recommendation algorithms. *ACM Transaction on Information Systems*, 22(1):143–177, 2004.

[22] Pedro Domingos and Matt Richardson. Mining the network value of customers. In *KDD*, 2001.

[23] Robin Dunbar. Grooming, gossip, and the evolution of language. 1998.

[24] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2010.

[25] Paul Erdös and Alfréd Rényi. On the evolution of random graphs. In *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, pages 17–61, 1960.

[26] Guler Ergün and Geoff Rodgers. Growing random networks with fitness. *Physica A Statistical Mechanics and its Applications*, 303:261–272, 2002.

[27] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. *SIGCOMM Computer Communication Review*, 29(4):251–262, 1999.

[28] Abraham D. Flaxman, Alan M. Frieze, and Juan Vera. A geometric preferential attachment model of networks. In *Algorithms and Models for the Web-Graph: 3rd international Workshop*, 2004.

[29] Wojciech Galuba, Karl Aberer, Dipanjan Chakraborty, Zoran Despotovic, and Wolfgang Kellerer. Outtweeting the twitterers - predicting information cascades in microblogs. In *Proceedings of the 3rd International Conference on Online Social Networks*, 2010.

[30] Priyanka Garg, Irwin King, and Michael R. Lyu. Are you a social conformer? In *CIKM*, 2012. Under review.

[31] Priyanka Garg, Irwin King, and Michael R. Lyu. Cikm. In *CIKM*, 2012. Under review.

[32] Micha Gisser, James E. McClure, Giray Okten, and Gary Santoni. Some anomalies arising from bandwagons that impart upward sloping segments to market demand. *Economic Journal Watch*, 6(1):21–34, 2009.

[33] Manuel Gomez-Rodriguez, David Balduzzi, and Bernhard Schölkopf. Uncovering the temporal dynamics of diffusion networks. In *ICML*, 2011.

[34] Manuel Gomez-Rodriguez, Jure Leskovec, and Andreas Krause. Inferring networks of diffusion and influence. In *KDD*, 2010.

[35] Amit Goyal, Wei Lu, and Laks V.S. Lakshmanan. Celf++: optimizing the greedy algorithm for influence maximization in social networks. In *WWW*, 2011.

[36] Amit Goyal, Wei Lu, and Laks V.S. Lakshmanan. Simpath: An efficient algorithm for influence maximization under the linear threshold model. In *ICDM*, 2011.

[37] Daniel Gruhl, Ramanathan V. Guha, David Liben-Nowell, and Andrew Tomkins. Information diffusion through blogspace. In *WWW*, 2004.

[38] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference*, 2008.

[39] Lars Kai Hansen, Adam Arvidsson, Finn Årup Nielsen, Elanor Colleoni, and Michael Etter. Good friends, bad news - affect and virality in twitter. *CoRR*, abs/1101.0510, 2011.

[40] Thomas Hofmann. Latent semantic models for collaborative filtering. *ACM Transaction on Information Systems*, 22(1):89–115, 2004.

[41] Junming Huang, Xue-Qi Cheng, Hua-Wei Shen, Tao Zhou, and Xiaolong Jin. Exploring social influence via posterior effect of word-of-mouth recommendations. In *WSDM*, 2012.

[42] Seung-Won Hwang and Mu-Woong Lee. A uncertainty perspective on qualitative preference. In *UAI*, 2009.

[43] Mohsen Jamali and Martin Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *RecSys*, 2010.

[44] Rong Jin, Joyce Y. Chai, and Luo Si. An automatic weighting scheme for collaborative filtering. In *SIGIR*, 2004.

[45] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *KDD*, 2003.

[46] Jon Kleinberg. Small-world phenomena and the dynamics of information. In *NIPS*, 2001.

[47] Jon M. Kleinberg. Navigation in a small world. *Nature*, 406(6798):845, 2000.

[48] Jon M. Kleinberg, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew S. Tomkins. The web as a graph: measurements, models, and methods. In *Proceedings of the 5th Annual International Conference on Computing and Combinatorics*, pages 1–17, 1999.

[49] Arnd Kohrs and Bernard Merialdo. Clustering for collaborative filtering applications. In *Proceedings of Computational Intelligence for Modeling, Control and Automation*, 1999.

[50] Yehuda Koren. Collaborative filtering with temporal dynamics. In *KDD*, 2009.

[51] Gueorgi Kossinets, Jon Kleinberg, and Duncan Watts. The Structure of Information Pathways in a Social Communication Network. In *KDD*, 2008.

[52] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *WWW*, 2010.

[53] Silvio Lattanzi and D. Sivakumar. Affiliation networks. In *Proceedings of the 41st annual ACM symposium on Theory of Computing*, 2009.

[54] Kristina Lerman and Rumi Ghosh. Information Contagion: an Empirical Study of the Spread of News on Digg and Twitter Social Networks. In *ICWSM*, 2010.

[55] Jure Leskovec, Lada A. Adamic, and Bernardo A. Huberman. The dynamics of viral marketing. In *Proceedings of the 7th ACM Conference on Electronic Commerce*, 2006.

[56] Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg, Christos Faloutsos, and Zoubin Ghahramani. Kronecker Graphs: An Approach to Modeling Networks. *Journal of Machine learning Research*, 11:985–1042, 2010.

[57] Jure Leskovec and Eric Horvitz. Planetary-scale views on a large instant-messaging network. In *WWW*, 2008.

[58] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *KDD*, 2005.

[59] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data*, 1(1), 2007.

[60] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-effective outbreak detection in networks. In *KDD*, 2007.

[61] Jure Leskovec, Kevin J. Lang, Anirban Dasgupta, and Michael W. Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1):29–123, 2009.

[62] Jure Leskovec, Mary Mcglohon, Christos Faloutsos, Natalie Glance, and Matthew Hurst. Cascading behavior in large blog graphs. In *SDM*, 2007.

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

[64] Nathan N. Liu and Qiang Yang. Eigenrank: a ranking-oriented approach to collaborative filtering. In *SIGIR*, 2008.

[65] Hao Ma, Haixuan Yang, Michael R. Lyu, and Irwin King. Mining social networks using heat diffusion processes for marketing candidates selection. In *CIKM*, 2008.

[66] Hao Ma, Dengyong Zhou, Chao Liu, Michael R. Lyu, and Irwin King. Recommender systems with social regularization. In *WSDM*, 2011.

[67] Paolo Massa and Paolo Avesani. Trust metrics in recommender systems. In *Computing with Social Trust*, Human-Computer Interaction Series, pages 259–285. 2009.

[68] Stanley Milgram. The Small World Problem. *Psychology Today*, 2:60–67, 1967.

[69] Stephen Morris. Contagion. *Review of Economic Studies*, 67:57–78, 2000.

[70] Seth A. Myers and Jure Leskovec. On the convexity of latent social network inference. In *NIPS*, 2010.

[71] Amit A. Nanavati, Siva Gurumurthy, Gautam Das, Dipanjan Chakraborty, Koustuv Dasgupta, Sougata Mukherjea, and Anupam Joshi. On the structural properties of massive telecom call graphs: findings and implications. In *CIKM*, 2006.

[72] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming*, 14(1):265–294, 1978.

[73] M. E. J. Newman. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5):323–351, 2005.

[74] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National academy of sciences of USA*, 103(23):8577–8582, 2006.

[75] Christopher R. Palmer, Phillip B. Gibbons, and Christos Faloutsos. Anf: a fast and scalable tool for data mining in massive graphs. In *KDD*, 2002.

[76] Erzsébet Ravasz and Albert-László Barabási. Hierarchical organization in complex networks. *Physical Review E*, 67:026112, 2003.

[77] S. Redner. How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B - Condensed Matter and Complex Systems*, 1998.

[78] Daniel M. Romero, Brendan Meeder, and Jon Kleinberg. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *WWW*, 2011.

[79] Paul Rozin and Edward B. Royzman. Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 4:296–320, 2001.

[80] Kazumi Saito, Ryohei Nakano, and Masahiro Kimura. Prediction of information diffusion probabilities for independent cascade model. In *Proceedings of the 12th International Conference on Knowledge-Based Intelligent Information and Engineering Systems, Part III*, 2008.

[81] Ruslan Salakhutdinov and Andriy Mnih. Probabilistic matrix factorization. *Advances in Neural Information Processing Systems*, 20, 2008.

[82] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In *WWW*, 2001.

[83] Cosma R. Shalizi and Andrew C. Thomas. Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods and Research*, 40:211–239, 2011.

[84] Greg Ver Steeg, Rumi Ghosh, and Kristina Lerman. What stops social epidemics? In *ICWSM*, 2011.

[85] Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social influence analysis in large-scale networks. In *KDD*, 2009.

[86] Sudhir L. Tauro, Christopher Palmer, Georgos Siganos, and Michalis Faloutsos. A simple conceptual model for the internet topology. In *Global Internet*, pages 1667–1671, 2001.

[87] Duncan J. Watts, Peter Sheridan Dodds, and M. E. J. Newman. Identity and search in social networks. *Science*, 296:1302–1305, 2002.

[88] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[89] Shaomei Wu, Chenhao Tan, Jon Kleinberg, and Michael Macy. Does bad news go away faster. In *ICWSM*, 2011.

[90] Rongjing Xiang, Jennifer Neville, and Monica Rogati. Modeling relationship strength in online social networks. In *WWW*, 2010.

[91] Jaewon Yang and Jure Leskovec. Modeling Information Diffusion in Implicit Networks. In *ICDM*, 2010.

[92] Yi Zhang and Jonathan Koren. Efficient bayesian hierarchical user modeling for recommendation system. In *SIGIR*, 2007.

[93] Bridge Q. Zhao, Y.K. Li, John C.S. Lui, and Dah-Ming Chiu. Mathematical modeling of advertisement and influence spread in social networks'. In *Proceedings of ACM Workshop on the Economics of Networked Systems*, 2009.