Modeling Correlated Purchase Behavior in Large-Scale Networks – A Markov Random Field (MRF) Approach

Liye Ma

Machine Learning Data Analysis Project

May 2011

Abstract

The advent of information technology has enabled the collection of large scale network data. This presents an exciting opportunity for researchers to understand consumer behavior in a social network environment. In this study, we model the interdependence of product adoption decisions among consumers in a social network. We develop a Gaussian Markov Random Field (GMRF) model to characterize the correlation of latent product preferences among connected consumers. The GMRF approach has two desirable properties: first, it enables the modeling of arbitrary network topology; second, it encapsulates the concept of conditional independence which leads to a parsimonious specification of the correlation among consumers. Applying our model to a dataset obtained from a large Asian telecom company, we find strong and consistent evidence of positive correlation among connected consumers in their product adoption decisions. We find that the correlation is stronger between consumers who communicate more frequently. We evaluate the performance of our model on predicting consumer adoption, and find that its precision is almost twice that of a naïve model; it outperforms commonly used logistic regression-based benchmark models by 10-15%; and it outperforms support vector machine-based benchmark models as well. To our best knowledge, this study is the first to use GMRF to empirically model and estimate consumers' latent product preferences.

Keywords: Markov Random Field, Social Network, Homophily, Product Adoption, Forecast

1. Introduction

With the advent of information technology, large scale social network data is increasingly becoming available. This presents an exciting opportunity for social science researchers to leverage social network theory and machine learning techniques to investigate issues with significant practical implications. In this study, we analyze a large-scale telecom dataset, where we model the interdependence of product adoption behaviors of consumers of a major Asian telecom company, by leveraging the social network structure inferred from their communications.

Specifically, we develop a model to quantify an effect that has long been recognized in social networks, namely homophily (McPherson et al 2001). The theory of homophily indicates that people tend to connect, or form relationship, with other people with similar characteristics. Consequently, people who are connected in a social network tend to have similar traits. Consequently, their decisions, such as on whether to adopt a product, should be inter-dependent. Modeling such interdependence is not trivial, however, especially in social networks with arbitrary connection structures. In this study, we propose a model that uses Gaussian Markov Random Field, or GMRF, to quantify such interdependence.

GMRF is a specific type of Markov Random Field (MRF) model. MRFs have been used extensively in temporal and spatial analysis as well as image analysis (Rue and Held 2005). A few characteristics of GMRF make it especially attractive to study the impact of social networks on activities. These characteristics include its ability to model graphs of arbitrary topology, and the conditional independence property it elegantly encapsulates. Furthermore, model parameters of GMRF also have straightforward interpretations; this makes GMRF particularly attractive to social science researchers. Considering this, we develop a GMRF model to analyze consumers' product adoption behavior in a social network environment. Extensive research exists on the formation and implications of social networks. However, to our best knowledge, GMRF has not been used in these studies, especially on modeling consumer decisions. Our study thus contributes to the literature by introducing and evaluating this approach.

We use a unique dataset obtained from a large Asian telecom company, covering all cell phone consumers in a major city. The dataset contains the detailed phone call records of the consumers over a six-month period, and the purchase information of a product called caller-ringback tone (CRBT). In our study, we infer the social network structure using the phone call record, and model the adoption of CRBT by each individual consumer using a binary Probit model. In contrast to a standard Probit model which treats each consumer as independent, we consider the product tastes of individual consumers as being drawn from a Gaussian Markov Random Field, so that the product tastes, and in turn the adoption propensities of connected consumers, are correlated. The neighborhood structure of the GMRF is defined by social network connections inferred from the call data. Given the theory of homophily in social networks, we expect such correlations to be positive. We adopt a hierarchical Bayesian specification, and estimate the model using a Markov-Chain Monte Carlo (MCMC) method. Since the communication network is only an approximation of the true social network, we impose a threshold in identifying connections, i.e. two consumers are considered as connected only if the number of phone calls between them reaches a threshold value. We vary the threshold values to ensure robustness.

The estimation of our base model reveals significant and positive conditional correlation in product tastes among consumers who are connected. This confirms the existence of the homophily effect. Communication records are only an approximation of social connections, as they contain noise arising from non-social phone calls. With higher threshold values, more noise will be filtered out, leaving mostly true social connections. As we increase the threshold value, the correlation estimated from our model becomes stronger. This indicates that the actual correlation among consumers who are indeed connected may be even higher. In an extended model, we find the conditional correlation goes up with the communication frequency between the consumers. This further shows that not only the existence of ties, but also their strength, determine the extent to which consumers' tastes are correlated. Our estimate also shows that, somewhat unexpectedly, the precision of the estimated product taste is lower for consumers who are connected with more consumers.

To evaluate the predictive ability of the model, we estimate the model using 80% of the consumers, and apply the estimates to the remaining 20% of the consumers to predict their purchase decisions. We compare the predictive performance of our proposed models with a few benchmark models. These benchmark models use characteristics of individual consumers as well as social network measures, and make predictions using logistic regressions (LR) or support vector machine (SVM) approaches. We find that our proposed models significantly outperform all LR-based benchmark models, where the precisions of our proposed models are 10-15% better than the best benchmark model. We find that our proposed models perform better than the SVM-based benchmark models as well. Furthermore, the parameters of our models all have straightforward interpretations, while SVM-based models, especially non-linear ones, do not have such interpretability. Our modeling approach thus not only is parsimonious, but also has high potential to be applied to practical settings.

A rich literature on social and economic networks exists in fields such as sociology (Wasserman and Faust 1994), economics (Jackson 2003), marketing (Braun and Bonfrer 2010), and computer science (Kempe et al. 2003). Both the formation and the implication of networks have been the focus of existing studies. Literature on the formation of networks uses statistical and economic models. Jackson (2003) provides a comprehensive survey of economic models of network formation, where connections are in general formed according to their payoffs to related parties. Alternatively, exponential random graph models (Holland and Leinhardt 1981, Frank and Strauss 1986, Robins et al. 2007) have been developed, where the existence of connections is captured using probabilistic models. A key concept permeating network formation studies is homophily (McPherson et al. 2001), which states that people tend to associate with other people who are similar. This concept has been the focus on studies on community detections (Copic et al. 2009).

A large stream of literature also focuses on the implications of networks, since the reason to investigate networks is often to inform on specific characteristics of interest, such as opinion formation, diffusion, decision interdependence, etc. The impact of the network can be reflected either actively, where the network directly effects decision processes such as opinion formation (Golub and Jackson 2010), or passively, where the network informs decision makers, such as marketers, to predict consumer adoptions based on their associations (Hill et al 2006). Since our study focuses on understanding the interdependence of consumers' adoption decisions in a network environment, it falls into this category. There is a rich literature in this category. Choi et al. (2008) studies the imitation effect on demand propagation induced by geographic and demographic proximity across regions; Iyengar et al (2010) studies the effect of contagion on new product introduction; Hill et al. (2006) uses consumer network information to improve target marketing, with the idea that someone who is connected to an adopter is also likely to be an adopter. Our study is also related to a large stream of work on network based classifications, as predicting whether a consumer will adopt can be considered as a classification problem. Macskassy and Provost (2007) survey the research in this area, while providing a toolkit based on existing methods. These methods typically base classification on node specific attributes and attributes of related nodes, using Naïve Bayes or Logistic Regression methods or weighted averaging (e.g. Macskassy and Provost 2003). Our study is similar to these studies in that we also leverage the same type of information. We differ from these studies in that we model the interdependence as correlated latent tastes using Gaussian Markov Random Field, instead of using explicit regressions over or weighted averages of network neighbors' decisions.

GMRF is a specific type of Markov Random Field (MRF) model (Kindermann and Snell 1980). MRFs have been applied extensively in temporal and spatial analysis as well as image analysis (Rue and Held 2005). A few characteristics of GMRF make it especially attractive to study the implication of networks. These characteristics include its ability to model graphs of arbitrary topology, and the conditional independence property it elegantly encapsulates. Furthermore, model parameters of GMRF have straightforward interpretations. Given these features, GMRF has been applied to analyze genome networks lately (Wei and Li 2007), and in network mining studies (Richardson and Domingos 2002). Richardson and Domingos (2002) studies the effect of viral marketing through mining connections at knowledge-sharing sites, where adoption behavior is modeled using MRF. It is very closely related to our study. However, the study specifies the network effect as a simple average of neighbors' decisions, while we model it as correlated latent tastes, and statistically estimate the extent of correlation from data.

The rest of the paper is organized as follows: in section 2 we describe the dataset used in our study; we then discuss our model in detail in section 3; section 4 explains the estimation approach; section 5 discusses the results of model estimation, as well as the predictive performance of the proposed models and benchmark models; finally, we conclude in section 6.

2. Data

The dataset used in this study is obtained from a large Asian telecom company. The data covers all mobile phone customers in a major city for six months. The dataset contains detailed phone call data. The information contained in each communication record includes the customer who placed the call, identified by the phone number, the customer who was called, the timestamp of the phone call, and the duration of the call. Certain demographic information, including gender and age, is available for each customer. Also contained in the dataset are the purchase records of a product called caller-ringback-tone (CRBT). This product is popular in many Asian countries. The product works in a way similar to but different from that of a ringtone: if a consumer A has purchased a CRBT, and consumer B calls A, then *consumer B* will hear the tone played over the phone before A picks up. In our study, we use the phone call records to identify connections among consumers, and we model the adoption decisions of CRBT by leveraging the social network structure constructed this way.

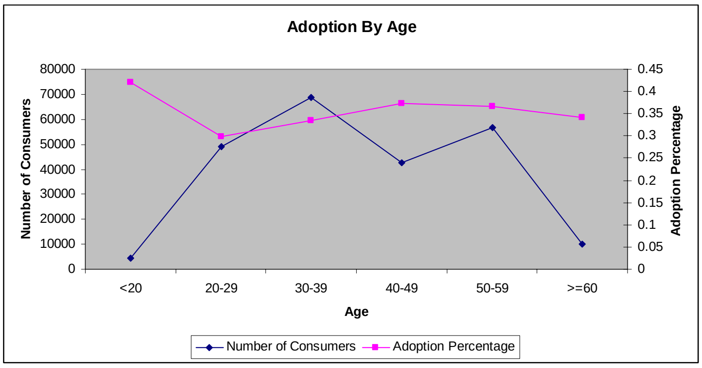
Table 2-1: Descriptive Statistics

	Mean	SD	Min	Max
Gender	Male	218017	Female	13399
Age	40.56	13.67		
Number of Consumers Called by Each Consumer	13.73	22.9	1	2858
Number of Phone Calls Per Consumer	410.4	942.7	1	59016
	Number	Adoption Percentage		
Number of Consumers	231416			
Number of Consumers Who Adopted CRBT	79505	34.36%		
Number of Consumers with Neighbors Who Adopted CRBT	200970			
Number of Such Consumers Who Adopted CRBT	73623	36.63%		
Number of Consumers with 3+ Neighbors Who Adopted CRBT	126294			
Number of Such Consumers Who Adopted CRBT	53904	42.68%		
Adoption Percentage by Gender	Male	34.50%	Female	31.89%

We first look at the descriptive statistics, which are reported in Table 2-1. There are a total of 231,416 consumers in the dataset. The average age of the consumers is 40.56. Each consumer on average called 13.73 other consumers over the six month period, placing an average of 410.4 phone calls. There is a wide dispersion on number of phone calls and the number of other consumers called, where the maximum are 59,016 and 2,858, respectively. The consumers are predominantly recorded as male, although this may not be an accurate reflection of reality, since those who purchase the phone may be different from those who use it (e.g. husband opens an account for wife to use).

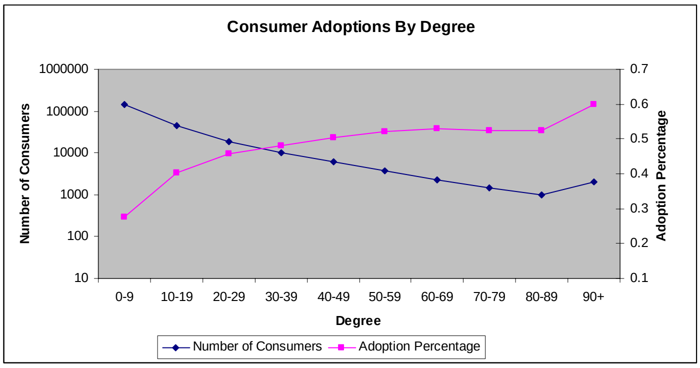
Among all the consumers, there are 79,505 who have adopted CRBT, leading to an adoption percentage of 34.36%. To further look into the adoption decisions, we calculate the adoption percentage by gender. As shown in the table, the adoption percentage of male consumers is 34.50%, slightly higher than that of female consumers, 31.89%, though the difference is minor. We then look at the adoption decisions of the age groups, shown in Figure 2-1. Most of the consumers are between ages of 20 and 60. The adoption percentage, however, does not exhibit a clear pattern based on age. The age group 20-29 has the lowest adoption percentage, 29.76%, while the age group younger than 20 has the highest adoption percentage, 42.14%.

Figure 2-1: CRBT Adoption by Age



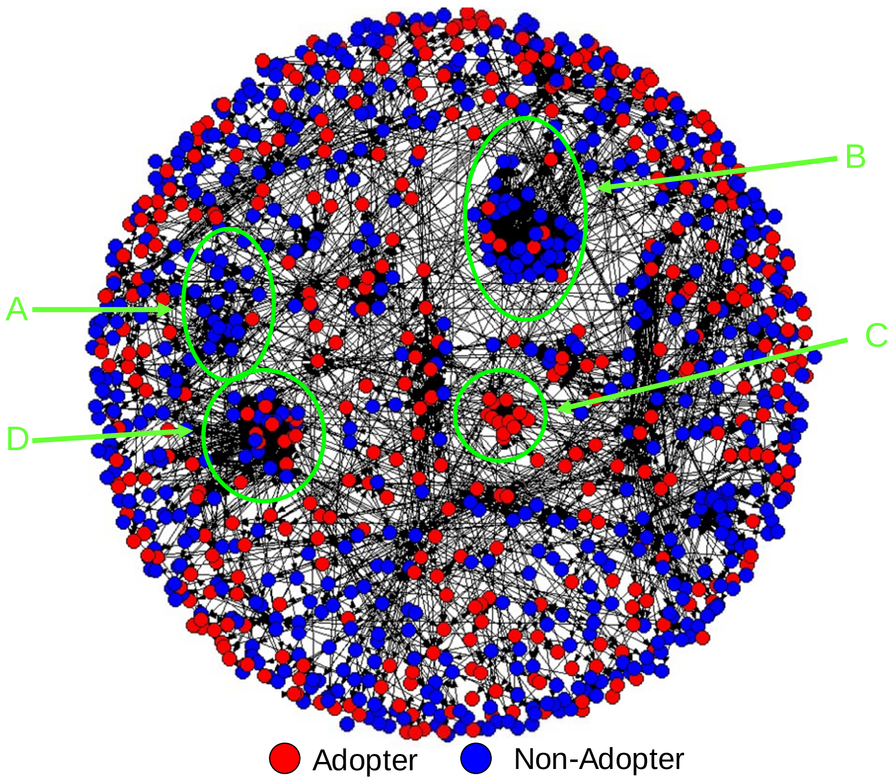
Neither gender nor age appears to affect CRBT adoption decisions greatly. To better understand adoption, we turn to characteristics related to the consumer network. Figure 2-2 plots the adoption decision by consumers' "degrees". The degree of a consumer is simply the number of other consumers who communicated with the consumer over the phone over the six month period. Two features are evident from the chart. First, most consumers communicate with only a very small number of other consumers – more than 80% of consumers communicated with less than 20 other consumers over the six month period (the left Y-axis is in logarithm scale). Second and more importantly, the adoption percentage goes up with degree: for consumers with degrees less than 10, the adoption percentage is 27.65%; whereas for the 2087 consumers who have degrees of 90 or higher, 59.94% adopt the product. It is thus quite promising to use network characteristics to inform on adoption decisions.

Figure 2-2: CRBT Adoption by Degree



Incorporating the degree of a consumer, however, is only the first step in this direction. To better leverage information of the network, we should also look at the decisions among connected consumers. Visualization is an effective way to discern patterns, especially for studies related to networks. With hundreds of thousands of consumers, however, the size of the dataset poses a significant challenge. Figure 2-3 is the graph constructed for 1000 consumers selected from the dataset through snowball sampling. Each node in the graph represents a consumer, and is color-coded based on his adoption decision. The connections between nodes represent the existence of phone calls between the corresponding consumers. Even for this small sample, with less than 1% consumers of the entire dataset, the connection structure is already very complex (visualizing the entire dataset would require specialized tools). Nonetheless, some observations can be made. As shown in the graph, some consumers seem to form tightly connected subgroups, and similar adoption decisions are apparent in some of these groups. For example, both group A and group B in the graph consist mostly of non-adopters. In contrast, group C consists primarily of adopters. This suggests that adoption decisions of connected consumers are likely correlated. To obtain further evidence, we calculated the number of consumers who have communicated with at least one other consumer who has adopted CRBT, the result of which is presented in Table 2-1. There are 200,970 such consumers with at least one neighbor who adopted. Among them, 73,623 consumers also adopted CRBT themselves. This implies a conditional adoption percentage of 36.63%, slightly higher than the population level average. Furthermore, there are 126,294 consumers who each have at least three neighbors who adopted CRBT. Among them, 53,904 consumers also adopted themselves. This indicates a conditional adoption percentage of 42.68%, significantly higher than that of the population. These descriptive statistics are further evidence that consumers' decisions to adopt CRBT are indeed correlated, and provide empirical motivation to our study. Such patterns, it should be pointed out, are not universal. As shown in Figure 2-3, group D is evenly divided on the adoption decisions, even though they are fairly closely connected. In-depth modeling is thus needed to further our understanding.

Figure 2-3: Call Graph and CRBT Adoption – 1000 Consumers



3. Model

We now discuss the model used in our study. There are I consumers, each indexed by $i, i=1..I$. Each consumer has a vector of observed characteristics, denoted as X_i. This vector includes gender, age, and degree.

Each consumer in the model is connected with a subset of other consumers. A connection is a two-way relationship, i.e. if A is connected to B, then B is connected to A. Therefore, the connections can be collectively represented using an undirected graph, the adjacency matrix of which is denoted as:

(1) $$C=[c_{ij}]$$

In equation (1), C is an $I \times I$ matrix, where:

(2) $$c_{ij}=\begin{cases}1 & \text{if consumers } i \text{ and } j \text{ are connected}\\ 0 & \text{otherwise}\end{cases}$$

Each consumer makes a decision on whether to adopt a product. In our dataset, the product is the CRBT. Denote the decision made by consumer i as D_i:

(3) $$D_i = \begin{cases} 1 & \text{if consumers } i \text{ adopts the product} \\ 0 & \text{otherwise} \end{cases}$$

We model the adoption decision using a binary Probit formulation:

(4) $\Pr(D_i = 1) = \Pr(U_i >= 0)$, where

(5) $$U_i = \alpha_i + \beta X_i + \varepsilon_i$$

In equation (5), β is the parameter that captures the importance of each observed characteristics in the adoption decision. $\varepsilon_i \sim N(0,1)$ is assumed to be an i.i.d. random disturbance, the variance of which is normalized to 1 for identification of the Probit model.

Of most interest in the model is the parameter α_i. This is the individual-specific, unobserved product taste parameter. The focus of this study is to leverage the network structure in understanding and predicting people's behavior. We do so by modeling α_i using a Gaussian Markov Random Field (GMRF).

3.1 Gaussian Markov Random Fields (GMRF)

The origin of Markov Random Fields, or MRF, (Kindermann and Snell 1980) can be traced back to the Ising model. Over the years, it has received widespread use in fields such as physics, economics, and sociology. A specific case of MRF, Gaussian Markov Random Fields (GMRF), is defined as follows (Rue 2008):

Definition (GMRF): A random vector $\vec{x} = (x_1,...x_n)^T$ is called GMRF w.r.t. the undirected graph $G = (V = \{1..n\}, E)$ with mean $\vec{\mu}$ and precision matrix $Q > 0$ if and only if its density has the form:

$$\pi(\vec{x}) = (2\pi)^{-n/2} |Q|^{1/2} \exp(-\frac{1}{2}(\vec{x} - \vec{\mu})^T Q(\vec{x} - \vec{\mu}))$$

And

$$Q_{ij} \neq 0 \Leftrightarrow \{i, j\} \in E, \forall i, j$$

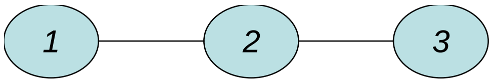
In another word, a GMRF is a multivariate-normal vector with the connection structure encoded by its precision matrix, so that non-zero off-diagonal elements in the matrix correspond to the existence of a connection between the two corresponding variables.

A seemingly simple formulation, GMRF has two desirable properties for modeling interdependence. The first is its ability to model arbitrary connection topology, as there is a simple correspondence between connections encoded in an adjacency matrix and non-zero off-diagonal element in the precision matrix of a GMRF.

The second desirable property of GMRF is the concept of conditional independence (CI) it encapsulates: if i and j are not directly connected, then conditional on every other variables, x_i and x_j are independent:

$$x_i \perp x_j \mid x_{-ij} \Leftrightarrow Q_{ij} = 0, \forall i, j$$

The property of CI can be easily seen from the density of the multivariate normal, where x_i and x_j can be factorized conditional on other elements if $Q_{ij} = 0$. The above two features make GMRF especially attractive for modeling the interdependence in consumer decisions. With this setup, we achieve the conditional independence property: if i and j are not directly connected, then their characteristics, conditional on the characteristics of all other consumers in the network, are independent.



To further illustrate conditional independence, consider the graph above. There are three consumers, with consumer 2 connected with both 1 and 3, while consumers 1 and 3 are not directly connected. In this group, we can expect the characteristics of 2 to be correlated with both that of 1 and that of 3 due to homophily. Because of this, we can also expect the characteristics of consumers 1 and 3 to be correlated. Since the correlation comes from consumer 2, however, we can expect that once the characteristic of 2 is known, those of 1 and of 3 should be independent. This property is elegantly encapsulated in the GMRF formulation.

Specific to our study, we assume that the intrinsic product taste parameters of all consumers are collectively randomly drawn from a GMRF:

$$\text{(6)} \qquad \begin{pmatrix} \alpha_1 \\ \ldots \\ \alpha_I \end{pmatrix} \sim N(\begin{pmatrix} \bar{\alpha} \\ \ldots \\ \bar{\alpha} \end{pmatrix}, Q^{-1})$$

In addition to the standard multivariate normal setup, the special property of GMRF is encoded in the precision matrix $Q = [q_{ij}]$, where $q_{ij} = 0$ if $c_{ij} = 0$. That is, the off-diagonal element of the precision matrix is zero for any pair of consumers who are not directly connected.

3.2 Parameterization and Interpretation of the Precision Matrix

In addition to enabling the modeling of arbitrary topology and encoding conditional independence, GMRF also has desirable interpretation properties. We now further detail the parameterization of our model and explain the interpretation.

We parameterize Q as follows:

$$(7) \qquad Q = \begin{pmatrix} \kappa & -r\kappa & 0 & \ldots & -r\kappa \\ -r\kappa & \kappa & 0 & \ldots & 0 \\ 0 & 0 & \kappa & \ldots & -r\kappa \\ \ldots & \ldots & \ldots & \ddots & \ldots \\ -r\kappa & 0 & -r\kappa & \ldots & \kappa \end{pmatrix}$$

Note: the off-diagonal element is 0 if the two nodes are not directly connected

This formulation is driven by the interpretation of GMRF. The precision matrix $Q = [q_{ij}]_{i,j=1..I}$ of a GMRF has straightforward interpretations. Each diagonal element represents the precision (inverse of variance) of the corresponding variable conditional on all other variables:

$$(8) \qquad \text{Precision}(\alpha_i \mid \alpha_{-i}) = q_{ii}$$

Meanwhile, the conditional correlation between each pair of variables is the negative of the corresponding off-diagonal element divided by the square root of the two corresponding diagonal elements:

$$(9) \qquad \text{Cor}(\alpha_i, \alpha_j \mid \alpha_{-ij}) = -q_{ij} / \sqrt{q_{ii} q_{jj}}$$

Given this, the parameter r in the precision matrix in our model is the conditional correlation of α_i and α_j, between two consumers i and j who are directly connected, while κ is the precision of α_i conditional on the intrinsic product taste of all others.

In addition to the direct interpretation of the precision matrix, the covariance matrix

$$(10) \qquad \Sigma = Q^{-1}$$

can also be interpreted normally, with the off-diagonal element of the matrix representing the (unconditional) covariance between the pair of consumers.

3.3 Alternative Specifications

The base model discussed above is a parsimonious formulation that incorporates the key features of GMRF. When applied to actual datasets, it could be extended to include additional properties. We now discuss two extensions of the base model. We will refer to the base model as Model B, while these two alternative models as Model AI and Model AII, respectively.

Note that in the base model, we specify the same conditional precision κ and correlation r parameters for all nodes and pairs. In the first alternative model, model AI, we relax the assumption on the precision parameter by making it depend on the node degree:

(11) $$Q^I = \begin{pmatrix} \kappa_{d_1} & -r\sqrt{\kappa_{d_1}\kappa_{d_2}} & 0 & \dots & -r\sqrt{\kappa_{d_1}\kappa_{d_I}} \\ -r\sqrt{\kappa_{d_1}\kappa_{d_2}} & \kappa_{d_2} & 0 & \dots & 0 \\ 0 & 0 & \kappa_{d_3} & \dots & -r\sqrt{\kappa_{d_3}\kappa_{d_I}} \\ \dots & \dots & \dots & \ddots & \dots \\ -r\sqrt{\kappa_{d_1}\kappa_{d_I}} & 0 & -r\sqrt{\kappa_{d_3}\kappa_{d_I}} & \dots & \kappa_{d_I} \end{pmatrix}$$

In this formulation, d_i is the degree of node i, i.e. the number of consumers to whom consumer i is connected. We make this extension because intuitively, the more connections we know about a consumer, the more we should know about her, hence higher precision. That is, we expect $\kappa_m > \kappa_n$ if $m > n$. To preserve the parsimony of the model, we specify the following functional relationship between the conditional precision and the degree:

(12) $$\kappa_d = \kappa_0 + \kappa_1 \cdot \log(d+1)$$

That is, the conditional precision is a linear function of the log of the node degree.

All relations are not the same, with some stronger or closer than others. Considering this, in our second model extension, Model AII, we make the conditional correlation parameter depend on the observed closeness between the two consumers.

(13) $$Q^{II} = \begin{pmatrix} \kappa_{d_1} & -r_{21}\sqrt{\kappa_{d_1}\kappa_{d_2}} & 0 & \dots & -r_{I1}\sqrt{\kappa_{d_1}\kappa_{d_I}} \\ -r_{21}\sqrt{\kappa_{d_1}\kappa_{d_2}} & \kappa_{d_2} & 0 & \dots & 0 \\ 0 & 0 & \kappa_{d_3} & \dots & -r_{I3}\sqrt{\kappa_{d_3}\kappa_{d_I}} \\ \dots & \dots & \dots & \ddots & \dots \\ -r_{I1}\sqrt{\kappa_{d_1}\kappa_{d_I}} & 0 & -r_{I3}\sqrt{\kappa_{d_3}\kappa_{d_I}} & \dots & \kappa_{d_I} \end{pmatrix}$$

To account for the closeness, we specify the correlation as a function of the number of communications between the two consumers:

(14) $$r_{ij} = r_0 + r_1 \cdot \log(Call_{ij})$$

That is, the conditional correlation is a linear function of the log of the number of phone calls between the two consumers.

4. Estimation

The base model is parameterized at the population level with $\Theta = \{\overline{\alpha}, \beta, r, \kappa\}$. Given the observed connection structure C, and individual characteristics $X = \{X_i\}_{i=1..I}$, the likelihood of observing the adoption decision $D = \{D_i\}_{i=1..I}$ is:

$$(15) \qquad L(D \mid \Theta, C, X) = \iiint\limits_{\alpha_1 \ldots \alpha_I} \prod_{i=1}^{I} L(D_i \mid X_i, \alpha_i, \beta) L(\alpha_1..\alpha_I \mid C, \overline{\alpha}, r, \kappa) d\alpha_1..d\alpha_I$$

Unfortunately, this likelihood function is computationally infeasible to compute, as it involves high-dimensional integration and calculating the density of a high-dimensional multivariate distribution. To address these two problems, we resort to a hierarchical Bayesian setup, where we estimate the model parameters together with the individual level propensity parameters, i.e. $\alpha_1..\alpha_I$, using Markov-Chain Monte Carlo (MCMC) method. Furthermore, to address the issue of computing the high-dimensional normal density, we follow a pseudo-likelihood approach (Besag 1986) for computing the model parameters.

4.1 MCMC Draws

Specifically, we take MCMC draws in a hybrid Metropolis-Gibbs fashion as follows:

4.1.1 Draw α_i:

$$(16) \qquad f(\alpha_i \mid \alpha_{-i}, \beta, \overline{\alpha}, r, \kappa, X_i, D_i, C) \propto \varphi(\alpha_i \mid \alpha_{N(i)}, \overline{\alpha}, r, \kappa) L(D_i \mid \alpha_i, \beta, X_i, D_i)$$

In equation (16), $\varphi(.)$ is the conditional density of a normal random variable given the other random variables; $N(i)$ is the neighbors of i as encoded in C – the other consumers are conditionally independent from i given $N(i)$; $L(.)$ is the Probit-likelihood of the adoption decision given the parameters and data. We sample α_i through Metropolis with random walk, where the random walk step is taken from normal distribution $N(0, 0.05)$.

4.1.2 Draw $\overline{\alpha}$:

$$(17) \qquad f(\overline{\alpha} \mid \alpha_i : i = 1..I) \propto \phi((I + V_\alpha)^{-1}(\sum_{i=1}^{I} \alpha_i + V_\alpha \overline{\overline{\alpha}}), (I + V_\alpha)^{-1})$$

In equation (17), $\varphi(.)$ is the density of normal distribution. We choose diffuse conjugate hyper-priors $\overline{\overline{\alpha}} = 0$ and $V_\alpha = 10000$

4.1.3 Draw β:

(18) $$f(\beta \mid \alpha_i : i = 1..I, X_i, D_i) \propto \pi(\beta)\prod_{i=1}^{I} L(D_i \mid \alpha_i, \beta, X_i, D_i)$$

We choose a diffuse improper prior of uniform distribution for each element in β. We sample each element in β through Metropolis with random walk, where the random walk step is taken from normal distribution $N(0,0.05)$.

4.1.4 Draw r:

(19) $$f(r \mid \alpha_i : i = 1..I, \beta, \overline{\alpha}, \kappa, C) \propto \pi(r)\prod_{i=1}^{I} \varphi(\alpha_i \mid \alpha_{N(i)}, \overline{\alpha}, r, \kappa)$$

We choose a flat prior for r. We sample r through Metropolis with random walk, where the random walk step is taken from normal distribution $N(0,0.05)$.

4.1.6 Draw κ:

(20) $$f(\kappa \mid \alpha_i : i = 1..I, \beta, \overline{\alpha}, r, C) \propto \pi(\kappa)\prod_{i=1}^{I} \varphi(\alpha_i \mid \alpha_{N(i)}, \overline{\alpha}, r, \kappa)$$

We choose a flat prior for r. We sample r through Metropolis with random walk, where the random walk step is taken from normal distribution $N(0,0.05)$.

The estimation of the two extended models is very similar to that of the base model, with slight modifications of the MCMC draws to incorporate the new functional form. Flat priors are used for parameters κ_0, κ_1, r_0 and r_1.

5. Empirical Results

In this section, we discuss the empirical analysis of applying the model and estimation methods to the mobile phone dataset. We first discuss an important step to operationalize the model, namely identifying the connection structure. We then present the estimation results and examine predictive performance of the proposed models.

5.1 Identifying Connections

Our model assumes the knowledge of social connections among consumers, yet what is actually contained in the dataset is the communication network. Ideally, a generative model of the communication network should be used. That is, it is best to have a statistical model which specifies the distribution of a certain true underlying characteristics of people, and the likelihood they communicate with one another given their respective characteristics. An example is the latent space formulation (Hoff et al. 2002). Such a generative model, however, has two important difficulties in our setting. First, existing generative models, such as latent space approach, cannot be easily scaled to networks of large size, as estimating such a generative model minimally calls for repeatedly computing the likelihood of each edge, which is of the order $O(n^2)$, where n is the number of nodes in the networks. Second and more importantly, existing generative models mainly focus on explaining the communication network alone, instead of drawing implications from the network structure, e.g. understanding individual purchase behavior as in our study. It is unclear how a latent characteristics approach, which is of dyadic nature, can be extended to modeling individual behaviors. For example, when using the latent space approach, the actual coordinate of a point in the latent space is of no meaning; only the distance between two points matters. However, to explain purchase behavior we need a generative model where the coordinates of points are interpretable.

Considering this, we use the communication networks to approximate the underlying social network. Since it is possible that the data contains random phone calls among people who are not closely related, we use a threshold approach, by considering two people connected only if the number of times they communicated with each other reaches a threshold. The setting of the threshold value is a balance between precision and recall: setting a low threshold would admit connections that are actually non-social phone calls, while setting a high threshold would drop true social connections between two people who do not communicate frequently. To ensure robustness, we vary the threshold values and examine the sensitivity of results to such thresholds.

5.2 Estimation Result

We now discuss the result of the model estimation. We randomly choose 80% of the consumers for estimation, leaving the remaining 20% to evaluate predictive performance. Since our model is a static one, this random partition is used to approximate the situation where a company first learns from existing customers, then applies the model to make prediction for new customers. The result is generated from taking 10,000 MCMC draws, with the first 5,000 discarded as the burn-in draws.

The result for the base model, Model B, is presented in Table 5.2-1. We tested six different threshold values for identifying connections. In the case of the lowest threshold, we consider two consumers as connected as long as there has been at least one phone call between them. In the case of the highest threshold, we consider two as connected only when they have called each other at least 20 times. As the table shows, both κ, the conditional precision parameter, and r, the conditional correlation parameter, are highly statistically significant (the standard deviations of the draws are quite low for both parameters, as there are hundreds of thousands of observations). The conditional precision parameter κ is estimated to be quite similar across all threshold values.

The conditional correlation parameter r is positive for all threshold values. This clearly shows that the product tastes of connected consumers are positively correlated. This confirms the motivation of the model, and shows there is potential to identify likely adoption targets based on social network information. Furthermore, the result in table 5.2-1 shows that as threshold value increase, this conditional correlation increases significantly: from 0.0225 for threshold value 1 to 0.0595 for threshold value 20, a 164% increase. This shows that identifying connection is important for social network based inferences. When threshold value is low, many random phone calls will be picked up as representing connections, and such noise dilutes the correlation between consumers who are truly connected. As threshold value goes up, more noise will be filtered out, leaving mostly true connections, and the estimated correlation becomes stronger.

Table 5.2-1: Parameter Estimation – Model B

Threshold	κ		r	
	Mean	SD	Mean	SD
1	0.0991	0.00036	0.0225	0.00012
3	0.0978	0.00064	0.0303	0.0004
5	0.0964	0.00044	0.0385	0.00072
8	0.0951	0.00059	0.0464	0.00075
10	0.0952	0.00074	0.0471	0.00088
20	0.0934	0.00051	0.0595	0.00104

The result for the first alternative model, Model AI, is presented in Table 5.2-2. This model extends the base model by making the conditional precision depend on node degree. The motivation for this extension is that as more neighbors of a consumer are known, we should know more about the consumer, i.e., the condition precision should be higher. Interestingly, however, we see that the coefficient for the linear parameter, κ_1, has a negative sign for all threshold values. This means the conditional precision is lower for consumers who have more connections. An explanation is that those consumers who are more connected, and more active, have higher variation in their tastes for the product. Thus although more of their neighbors are known, it is not enough to pinpoint their taste with higher precision. An in-depth understanding of this is in interesting topic for further research. In this model specification, we again see positive conditional correlation for all threshold values, and that the correlation goes up with threshold value.

Table 5.2-2: Parameter Estimation – Model AI

Threshold	κ_0		κ_1		r	
	Mean	SD	Mean	SD	Mean	SD
1	0.129	0.0011	-0.013	0.00031	0.0227	0.00038
3	0.115	0.00093	-0.0097	0.00037	0.03487	0.0006
5	0.113	0.00153	-0.0094	0.00061	0.03912	0.00079
8	0.108	0.0011	-0.008	0.00075	0.0469	0.00088
10	0.1043	0.0015	-0.0063	0.00084	0.0536	0.00094
20	0.101	0.0016	-0.0054	0.00091	0.0607	0.0012

The result for the second alternative model, Model AII, is reported in Table 5.2-3. In this extension, the conditional correlation parameter is defined to be a function of the strength of the ties, measured using the number of calls between the two consumers. As the result shows, the linear parameter, r_1, is positive and statistically significant for all threshold values. This shows that in additional to the existence of a connection, the strength of the tie also indicates the extent to which the two consumers are correlated in their tastes. Two consumers who are more strongly connected, indicated by their more frequent communications, have higher correlation in their product taste. In this table, we again see that the conditional correlation go up as the threshold value goes up, indicating it is important to filter out noise in the data.

Table 5.2-3: Parameter Estimation – Model AII

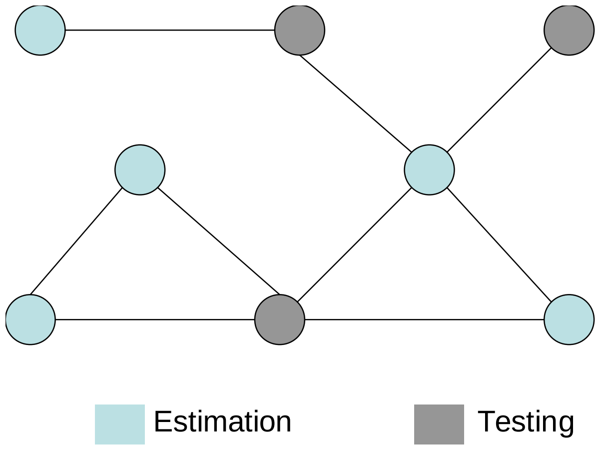
Threshold	κ_0		κ_1		r_0		r_1	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	0.129	0.0011	-0.0127	0.0004	-0.0013	0.000832	0.0128	0.0004
3	0.117	0.0008	-0.0099	0.0004	-0.021	0.0022	0.0183	0.0007
5	0.11	0.0012	-0.0078	0.0006	-0.025	0.0034	0.0199	0.001
8	0.1077	0.0016	-0.0074	0.0008	-0.0476	0.0036	0.0253	0.0009
10	0.1051	0.0011	-0.0063	0.0006	-0.0444	0.0047	0.0242	0.0012
20	0.0994	0.0014	-0.004	0.00087	-0.056	0.0061	0.0283	0.0014

In summary, the estimated results of the three models for various threshold values confirm the correlation in product taste between the consumers who are connected. Such correlation increases as we increase the threshold value for identifying connections, as more noise is filtered out with higher threshold values. Furthermore, the correlation is stronger for consumers who communicated with each other more frequently, so the strength of the tie also matters in addition to its existence.

5.3 Predictive Performance

We now evaluate the predictive performance of the model. From marketer's perspective, it is important to accurately evaluate a consumer's likelihood to adopt, to improve the effectiveness of target marketing efforts. We perform this evaluation by randomly choosing 80% of the consumers as training sample for estimation. We then apply the estimated model on the remaining 20% of the consumers to predict their adoption. The individualized prediction comes from the knowledge of their connections with the customers in the training dataset. This is illustrated in Figure 5.3-1. We exclude consumers who are isolated, i.e. those who are not connected to any other consumers. Since our objective is to leverage social network knowledge, isolated consumers are irrelevant to our study. The number of remaining test cases is reported in Table 5.3-2.

Figure 5.3-1: Dividing Training and Testing Data



We perform two types of predictions. In the first, we predict for each consumer in the testing dataset based on its estimate alone. That is, we predict that the consumer will adopt if the model calculates an adoption probability higher than 0.5. In the second, we follow a “top-k” approach, by predicting that a fixed number of consumers, N_a will adopt. To do this, we rank consumers based on their predicted adoption probability, and choose the first N_a consumers based on this ranking. The first prediction approach is fairly standard, while the second approach is done with a focus on effectiveness. For example, if the company wants to target 100 consumers for marketing, it would look at the 100 consumers who rank the highest in adoption probability.

For the first prediction approach, we look at two metrics for evaluation. The first is the percentage of correct predictions, while the second is the percentage of correct predictions where the prediction is to adopt (that is, the *precision* of adoption prediction). The first is a standard evaluation metric, while the second is in a sense more managerially relevant. Obviously for the second prediction approach (the top-k approach), only the precision metric is applicable.

We compare the performance of our three proposed models against five benchmark models. The first benchmark model, denoted as BM1, uses individual demographics characteristics observed, namely gender and age, to predict adoption using logistic regression. The second benchmark model, denoted as BM2, uses the demographics characteristics plus the degree of consumers for the logistic regression. Adding degree as an explanatory variable can potentially improve performance, as the data patterns show higher adoption percentage of consumers of higher degrees. The degree of a consumer is a feature in the social network context. Thus the second benchmark model already leverages social network information. However, this model does not leverage the adoption decisions of other consumers. The third benchmark model, denoted as BM3, leverages the decisions of other consumers, by including in the covariates not only the degree of the consumer, but also the percentage of neighbors who adopted. This is similar to the weighted-vote relational neighbor classifier in Macskassy and Provost (2003).

Existing network classification models mostly use Logistic Regression methods (or Naïve Bayes which is equivalent). The logistic regression approach is somewhat similar to our proposed models, as they both can be grounded to a utility framework. To also compare with models that are "non-nested", we introduce two more benchmark models, denoted as BM4 and BM5, that use the support vector machine (SVM) method. These models treat the adoption prediction as a classification task. Both models use gender, age, degree, and percentage of neighbors who adopted as feature variables. Model BM4 uses a linear kernel, while model BM5 uses a polynomial kernel. These benchmark models are summarized in Table 5.3-1.

Table 5.3-1: Benchmark Models

Model	Explanatory Variables	Mechanism
BM1	Gender, Age	Logistic Regression
BM2	Gender, Age, Degree	Logistic Regression
BM3	Gender, Age, Degree, Percentage of Neighbors who Adopt	Logistic Regression
BM4	Gender, Age, Degree, Percentage of Neighbors who Adopt	Suppor Vector Machine, Linear Kernel
BM5	Gender, Age, Degree, Percentage of Neighbors who Adopt	Suppor Vector Machine, Polynomial Kernel

Table 5.3-2 reports the statistics of the test cases, and the percentage of correct predictions of the three proposed models, corresponding to the different threshold values imposed for identifying connections. As the table shows, the number of test cases reduces as the threshold goes up. This is because the higher the threshold value we impose, the more consumers become isolated and are thus eliminated from the test dataset. About 35% of consumers are eliminated as we increase the threshold from 1 to 20. The table also shows that the higher the threshold, the higher the percentage of consumers who adopt CRBT – the adoption percent increases from 34.18% to 38.6% as we increase the threshold from 1 to 20. This is as expected, as the data patterns show that consumers who are more connected are more likely to adopt CRBT. The intuition is that those who communicate more frequently are also more active consumers who may be more prone to purchase related products.

Interestingly, the table shows that the percentage of correct predictions goes down as the threshold value increases. This is surprising at first glance, since as threshold value increases the "noise" in the data goes down, so the performance should improve. To understand this pattern, then, recall that the adoption percentage of the test dataset goes up with threshold value. Since the adoption is below 50%, this means that the prediction task gets harder as the threshold increases. Consider the column for the "Naive" model in the table. This is simply a predictor that always predicts no adoption. Since fewer than half of consumers adopt, such a dummy predictor will already be correct in more than 50% of the cases. As the threshold value increases, we can see the accuracy of this dummy model goes down, showing that the task gets harder. Note that the three proposed models all have higher predictive accuracy than this naïve model, which serves as a first validation of the performance of the proposed models.

Table 5.3-2: Predictive Accuracy – Individual Predictions

				Percent of Correct Prediction			
Threshold	Total Test Cases	Total Adoption	Adoption Percent	Mode B	Model AI	Model AII	"Naive" Model
1	46092	15752	34.18%	66.82%	66.71%	67.14%	65.82%
3	42675	15205	35.63%	65.93%	66.10%	66.52%	64.37%
5	39575	14234	35.97%	65.35%	65.24%	66.06%	64.03%
8	36715	13674	37.24%	64.52%	64.97%	65.49%	62.76%
10	35290	13103	37.13%	64.38%	63.84%	64.79%	62.87%
20	29846	11520	38.60%	63.11%	63.20%	63.74%	61.40%

The percentage of correct prediction is the first metric to look at, but it is not the most interesting one. From a marketer's perspective, it is more important to know how well the model performs when it predicts adoption. That is, how often is the model correct when it predicts that a consumer will adopt. Table 5.3-3 reports this result, when the model predicts adoption on an individual basis, i.e. predicting adoption when the calculated probability is greater than 0.5.

We first note that all three models have higher than 50% precision for all threshold values, ranging from 52.76% to 57.48%. Comparing this with a naïve predictor which always predicts adoption, the precision of which ranges from 34.18% to 38.6% depending on the threshold value, we can see that the three proposed models have strong predictive performance, with about 50% improvement in precision from a naïve, no information, model. Comparing the three proposed models, we can see that Model AI in general is slightly better than model B, while Model AII is better than both Model B and Model AI. This is as expected, as the models are successively nested.

Table 5.3-3: Precision – Individual Predictions

	Model B		Model AI		Model AII	
Threshold	Predicted Adoption	Correct Percentage	Predicted Adoption	Correct Percentage	Predicted Adoption	Correct Percentage
1	8385	52.88%	7671	52.76%	8129	53.72%
3	5658	55.07%	6439	55.71%	6752	56.80%
5	6609	54.18%	6359	55.56%	6672	56.01%
8	6707	54.96%	6333	55.35%	6700	57.48%
10	6182	55.26%	7344	54.10%	6242	55.43%
20	6213	54.45%	5977	55.19%	6693	55.22%

Table 5.3-4 reports the precisions of the logistic regression-based benchmark models when prediction is made on an individual basis, i.e. predict adoption when the calculated probability is above 0.5. When predicting using the benchmark model BM1, which uses only gender and age for prediction, the predicted adoption probability is less than 0.5 for all consumers in the test dataset. Thus the model does not predict adoption for any consumer, and is excluded from this evaluation. The precision of BM2 and BM3 are in general comparable with the three proposed models, and in several cases the precision of the benchmark models is better. However, we note a crucial distinction here: the

benchmark models in general make many fewer predictions of adoption than our proposed models – the three proposed models mostly make more than 6000 predictions of adoptions for all threshold values, whereas the two benchmark models, especially BM3, make only a little over 2000 such predictions. Precision is expected to be "diluted" when more predictions are made, so this must be taken into account when comparing the models. The three proposed models can retrieve more "leads" than the two benchmark models. Only BM2 makes more than 6000 predictions when threshold value is 10 or 20. And in both these cases, its precision is lower than those of the three proposed models. Figure 5.3-2 further highlights this point by comparing model AII with BM3. As the chart shows, although both models have similar overall precision, model AII has significantly higher recall.

Figure 5.3-2: Precision/Recall – Model AII vs. Model BM3

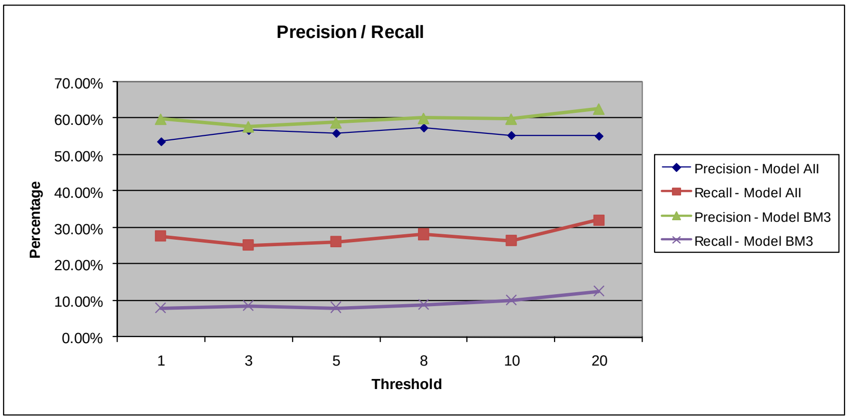


Table 5.3-4: Benchmark Model (LR) Precision – Individual Predictions

	Model BM2		Model BM3	
Threshold	Predicted Adoption	Correct Percentage	Predicted Adoption	Correct Percentage
1	2006	56.23%	2089	59.89%
3	2060	54.13%	2226	57.77%
5	4142	56.78%	1951	58.89%
8	5475	55.87%	2015	60.10%
10	7124	52.91%	2176	59.93%
20	10939	48.43%	2289	62.69%

Table 5.3-5 reports the precisions of the support vector machine-based benchmark models, when predictions are made individually. The result shows that the SVM-based models have higher precisions than the logistic regression-based models. These precisions are also about 10% higher than the three proposed models. Similar to the logistic regression-based models, however, the SVM-based models also make many fewer predictions of adoptions than the three proposed models do. Thus again the direct comparison of precision may be misleading.

Table 5.3-5: Benchmark Model (SVM) Precision – Individual Predictions

	Model BM4		Model BM5	
Threshold	Predicted Adoption	Correct Percentage	Predicted Adoption	Correct Percentage
1	3470	62.07%	1654	68.50%
3	3718	61.97%	1946	65.83%
5	3371	62.06%	2529	64.41%
8	4383	62.03%	2977	65.10%
10	4712	60.36%	3474	63.27%
20	4688	60.30%	3403	62.83%

Since it is not easy to compare performance when different models make different numbers of predictions, we turn to making only a fixed number of adoption predictions for each model, i.e. following a "top-k" approach. For each model, we order all consumers based on the calculated probability of the consumer adopting (for SVM-based models, the distance to the classification boundary), and predict adoption only for the top N_a consumers. This helps us make a fairer comparison across models. This approach also has more direct practical applications, as marketing managers often work under capacity constraint, e.g. a manager may need to distribute a fixed number of coupons to consumers.

Table 5.3-6: Precision – Top 1000/2000 Consumers

	Model B		Model AI		Model AII	
Threshold	Top 1000	Top 2000	Top 1000	Top 2000	Top 1000	Top 2000
1	66.00%	65.80%	65.90%	62.25%	66.30%	65.35%
3	69.80%	64.60%	68.60%	64.90%	72.00%	68.00%
5	69.80%	67.00%	69.60%	65.10%	73.10%	68.75%
8	71.10%	67.05%	67.50%	64.65%	73.80%	68.55%
10	71.40%	65.55%	68.70%	65.25%	71.70%	67.40%
20	70.50%	66.40%	73.50%	66.90%	72.40%	67.10%

The precision for the three models is reported in table 5.3-6. We tested two values, $N_a = 1000$ and $N_a = 2000$. As the table shows, the precision is above 60% for all three models and all threshold values, and reaches as high as 73.8%. Recall that only about 35% of the consumers in the dataset actually adopt, this represents almost double the precision of a naïve predictor, a very strong performance. The precision for the top 1000 consumers is higher than that for the top 2000 consumers for each model and each threshold value. This is as expected, as we have higher confidence the higher the predicted probability is. This further shows that the models are stable and well behaved.

The precision for Model AII is higher than that for both AI and B, for each threshold value and for both top 1000 and 2000 consumers, with the only exception of top 1000 consumers with threshold 20 for AI. This suggests that accounting for varying strengths in connections indeed can improve model fit and predictive performance. This improvement in performance from Model B and AI to AII, however, becomes smaller as threshold increases. This is as expected – as threshold values go up the connections become more consistently strong, so further differentiating connection strengths do not add much additional value. The precision for model AI is not higher than that for model

B. This is not surprising, either, as the model estimate does not show significant relationship between the conditional precision and the degree of a node. It is unclear, however, why in several cases the precision of model B is higher than that of model AI, as the former is nested in the latter. One explanation is this may be due to data noise when we look at a relatively small sample. When we compare model AI with model B with respect to individual predictions, where the sample size is larger, the performance are almost the same.

Looking along the threshold value dimension, we can see that the performance first goes up when the threshold value increases from 1 to 3, and from 3 to 5. After that, however, the performance does not further increase. In fact it decreases for Model AII. Increasing threshold value has both positive and negative implications. On one hand, it filters out noise, so that what is left is more likely true connections. On the other hand, it also filters out true connections when the communication is of low frequency, leaving fewer neighbors for each consumer to base inference on. The result suggests that a threshold value of 5 is reasonable compromise between these two factors.

We now compare the proposed models with the benchmark models. The results the three logistic regression-based models are reported in table 5.3-7. Among the three benchmark models, BM2 has significantly higher precision than BM1, while the performance of BM1 is not much better than that of a naïve predictor. This suggests that individual demographics information is not valuable in predicting the adoption decisions, while incorporating social network related metrics, such as number of connected consumers, can dramatically increase predictive performance. Model BM3 has roughly a further 10% improvement over BM2. This suggests that the adoption decisions of neighbors can directly improve predictive performance, in addition to other social network related measures. Comparing table 5.3-7 with 5.3-6, we can see that the three proposed models have clearly superior performance over the three logistic regression-based benchmark models – average precisions are 67.92%, 66.90%, and 69.54% for the three proposed models, versus 36.25%, 54.68%, and 60.76% for the three benchmark models. Even the proposed base model, Model B, has better performance than the best benchmark model, Model BM3.

Table 5.3-7: Benchmark Model (LR) Precision – Top 1000/2000 Consumers

	Model BM1		Model BM2		Model BM3	
Threshold	Top 1000	Top 2000	Top 1000	Top 2000	Top 1000	Top 2000
1	34.20%	34.05%	59.60%	56.25%	62.20%	60.25%
3	36.10%	35.90%	55.70%	53.90%	60.50%	57.90%
5	35.80%	35.80%	54.50%	52.45%	61.50%	59.00%
8	35.70%	37.75%	55.50%	53.90%	61.40%	60.00%
10	36.00%	38.70%	54.10%	53.25%	60.50%	59.45%
20	36.80%	38.15%	54.90%	52.15%	63.60%	62.85%

The results for the two SVM-based benchmark models are reported in Table 5.3-8. The two SVM-based models have much better performance than the three logistic regression-based models. The model using polynomial kernel, BM5, has slightly better performance than the one using linear kernel, BM4. The performances of these two benchmark models are similar to two of the proposed models, model B and AI, but worse than that of model AII. Among all the models, the proposed model AII has the best predictive performance when evaluated using this top-k method, which can also be seen from Figures 5.3-3 and 5.3-4, where the performances of all models are plotted. This clearly shows the value of our modeling approach.

Table 5.3-8: Benchmark Model (SVM) Precision – Top 1000/2000 Consumers

	Model BM4		Model BM5	
Threshold	Top 1000	Top 2000	Top 1000	Top 2000
1	68.10%	66.25%	71.10%	67.05%
3	69.30%	65.25%	70.10%	65.90%
5	70.50%	65.70%	71.80%	66.70%
8	67.10%	66.80%	69.70%	67.50%
10	68.80%	65.60%	70.40%	66.80%
20	70.30%	68.25%	74.60%	67.40%

Figure 5.3-3: Predictive Precision – Top 1000 Consumers

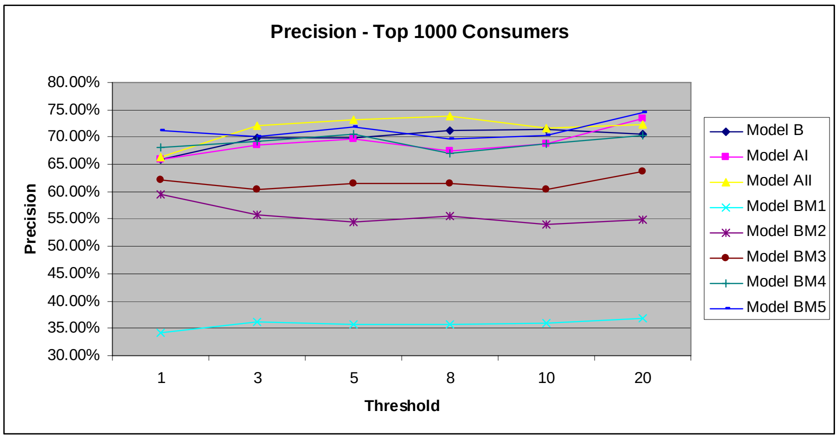
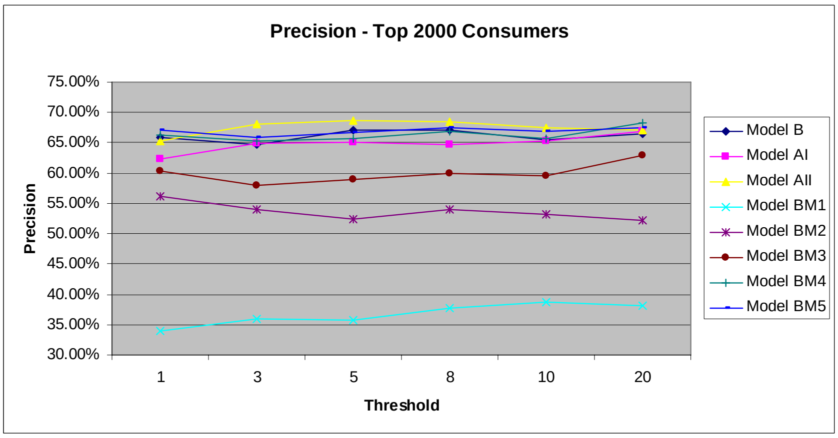


Figure 5.3-4: Predictive Precision – Top 1000 Consumers



In summary, although our proposed models and the benchmark model use the same set of information, the proposed models, under the GMRF framework, place higher emphasis on “local” characteristics, by incorporating a correlation structure directly between neighbors. The result shows that such local emphasis lead to both higher number of predicted adoptions (Table 5.3-3) and higher precision (Table 5.3-6), making the proposed models superior than the traditional logistic regression or SVM-based approaches. Although the performances of the SVM-based models are close to those of the proposed model, the proposed models are superior in terms of interpretability as well as extensibility: compared with the SVM-based models, the proposed models have parsimonious specifications, and the coefficients can be explained intuitively. Furthermore, the proposed models can be extended to handle more complex decisions, such as repeat purchase over time, or product selection, while extending the SVM-based models for those purposes would be challenging.

5.3.1 The Role of Training Data Size

To further evaluate the predictive performance of our proposed models, we check its robustness by varying the amount of training data used for estimating the model. The prediction task becomes more difficult as the amount of training data reduces, so it is important to investigate the sensitivity of the methods with respect to the data availability. As discussed earlier, Model AII has the best predictive performance among the three proposed models, while Model BM5 is the best among the benchmark models. We thus further compare these two models by varying the portion of the data used for training, from 10% to 90%. We use threshold value 5 for these tests. The result is presented in Table 5.3.1-1.

Table 5.3.1-1: Vary Training Data Size

	Model AII			Model BM5		
TrainingPortion	Individual	Top 1000	Top 2000	Individual	Top 1000	Top 2000
90%	56.85%	69.40%	62.20%	64.55%	66.10%	61.55%
80%	56.17%	71.60%	68.05%	66.11%	73.70%	67.55%
70%	55.30%	73.10%	69.25%	65.03%	72.10%	68.60%
60%	54.83%	74.90%	70.30%	63.46%	71.80%	68.55%
50%	53.86%	74.60%	71.85%	63.14%	73.90%	69.55%
40%	54.32%	76.50%	73.80%	61.31%	74.20%	70.90%
30%	53.64%	73.60%	69.75%	61.74%	74.40%	70.35%
20%	52.86%	72.30%	69.70%	61.92%	72.80%	69.25%
10%	52.74%	69.70%	68.40%	56.17%	69.30%	64.80%

As expected, as the training data portion decreases, the prediction precision where predictions are made individually also decreases. As shown in the table, when the portion of training data is reduced from 90% to 10%, the precision of model AII decreases from 56.85% to 52.74%, while that of BM5 decreases from 64.55% to 56.17%. The performance decline is only moderate, however, suggesting that the dataset is large enough to base inference on, even when only a small portion is used for training the model. Interestingly, when the predictions are made only for the top 1000 or 2000 consumers, the precisions for both Model AII and BM5 have an “Inverted-U” shape, with the precision highest when 40% of the data are used for training, while the precision lower when either more or less data are used for training. It may at first look surprising that precision is lower when more data are used for training. The answer lies with the test dataset: not all consumers are “good” candidates for predicting adoption (“good” here means they have high probability to adopt). As a larger portion of the data is used for training, the testing dataset becomes smaller, so there may be fewer such good candidates, meaning the top-1000 or top-2000 is a more difficult task, hence the lower precision. It is interesting to see the interactions of these two opposing factors: more training data lead to better estimate, but leave fewer candidates. Certainly in a practical situation the marketer may not be able to dictate the training and testing data size. But if the marketer can, such as through randomly promoting a product first to a subset of consumers and then learning the adoption propensity, he should be cognizant of these two forces and achieve a balance in between. Finally, we note that the relative performance between Model AII and BM5 is similar to what we discussed previously, that AII has better performance for the top-1000 and top-2000 consumers, whereas BM5 has higher precision for individual-based predictions, although it makes much fewer predictions of adoption. This validates the robustness of the previous result.

6. Discussions and Conclusion

Social network has been a focus of study by researchers in social science, economics, marketing, and computer science. The advent of information technology enables the collection of large-scale network data, which provide an exciting research opportunity. It has long been recognized that people tend to associate with other people who are similar, a phenomenon known as homophily. Consequently, people who are close to one another often have similar traits and make similar decisions.

The correlation among connected people means marketers can potentially leverage the knowledge of the social network among consumers to better identify prospects for target marketing. Our study contributes to the literature on this topic, by introducing a model based on Gaussian Markov Random Field (GMRF), which has received widespread applications in image processing, physics, and biology. We model the product tastes of consumers in a social network as a GMRF parameterized by the conditional precision of each node and the conditional correlation between each connected pair. This approach allows us to leverage two desirable properties of GMRF, its ability to model networks of arbitrary topology and its parsimonious specification with the property of conditional independence.

Applying the model to a large mobile phone dataset, where we study the adoption decisions of caller ring-back tones, we find strong and consistent evidence of correlation in product tastes among connected consumers. We find that such correlation becomes stronger as we increase threshold to filter out more noise in the dataset. Furthermore, we find that the correlation is also higher between consumers who communicate more frequently, suggesting the variation of tie strengthen also leads to different degrees of correlation among connected consumers. Our GMRF-based models have consistently superior predictive performance over the benchmark models, while working with the same set of information.

This study is a first step towards an in-depth modeling of correlated consumer behaviors, and is limited in several ways. First, the model is static, and treats the purchase of ring-back tones as a one-shot adoption decision. Many real world settings are dynamic, as consumers form and sever ties over time, and purchase products repeatedly over time as well. Two extensions are possible here. One is to keep the network structure as static, while modeling the repeated purchase decisions over time for consumers. Temporal dependence can be introduced in such a framework. The other is to treat both network and purchase as dynamically evolving, and use the up-to-date network information for inference.

Secondly, the model accounts for the correlation of adoption decisions among consumers who are connected, but more work can be done to uncover the source of such correlation. Although we motivated the correlation using homophily, the correlation in adoption decisions may also arise from social influence, e.g., a person calls another person, finds out the callee has a certain ring-back tone, and decides to purchase one himself. While it is hard to separate homophily from influence with a static model, it is possible with a dynamic one, especially with the detailed communication data in the dataset. Our modeling framework also lends itself to extensions which include social influence.

Finally, we use a threshold approach to identify social connections from the communication data. While intuitively appealing, it is subjective on the researcher’s side. It is more desirable to have an integrated data generation process, where one statistical model is used to explain both the calling behavior and the purchase decisions. This is a challenging task, as calling behavior is dyadic in nature, i.e. involves two people, while purchase decisions is at individual consumer level. Existing literature has either modeled the calling behaviors, e.g. using latent space approach, or modeled purchase decisions, e.g. using spatial regression, but has not done both together. This presents a fascinating question for future study.

Reference

Besag, J., (1986), "On the Statistical Analysis of Dirty Pictures", Journal of the Royal Statistical Society, 48(3) 259–302

Braun, M. and A. Bonfrer, (2010), "Scalable Inference of Customer Similarities from Interactions Data Using Dirichlet Processes", Marketing Science, forthcoming

Choi, J., S.K. Hui and D.R. Bell, (2008), "Bayesian Spatio-Temporal Analysis of Imitation Behavior Across New Buyers at an Online Grocery Retailer", Journal of Marketing Research, Volume XLV

Copic, J., M.O. Jackson and A. Kirman, (2009), "Identifying Community Structures from Network Data via Maximum Likelihood Methods", The B.E. Journal of Theoretical Economics, 9(1)

Frank, O. and D. Strauss, (1986), "Markov graphs", Journal of the American Statistical Association 81, 832–842

Golub, B. and M.O. Jackson, (2010), "How Homophily Affects the Speed of Contagion, Best Response and Learning Dynamics", Working Paper

Hill, S., F. Provost and C. Volinsky, (2006), "Network-Based Marketing: Identifying Likely Adopters via Consumer Networks", Statistical Science, 21(2), 256-276

Hoff, P.D., A.E. Raftery and M.S. Handcock, (2002), "Latent Space Approaches to Social Network Analysis", Journal of the American Statistical Association, 97(460), 1090–1098

Holland, P.W. and S. Leinhardt, (1981), "An Exponential Family of Probability Distributions for Directed Graphs", Journal of the American Statistical Association 76, 33–65

Iyengar, R., C. Van den Bulte and T.W. Valente, (2010), "Opinion Leadership and Social Contagion in New Product Diffusion", Marketing Science, forthcoming

Jackson, M.O, (2003), "A Survey of Models of Network Formation: Stability and Efficiency", Group Formation in Economics: Networks, Clubs, and Coalitions, Cambridge University Press

Kempe D., J. Kleinberg and E. Tardos, (2005), "Influential Nodes in a Diffusion Model for Social Networks", Proceedings of the 32nd International Colloquium on Automata, Languages and Programming (ICALP), 1127-1138

Kindermann R. and L. Snell (1980). "Markov Random Fields and their Applications", Amer. Math. Soc., Providence, R.I.

Macskassky, S.A. and F. Provost, (2007), "Classification in Networked Data: A Toolkit and a Univariate Case Study", Journal of Machine Learning Research, 8, 935-983

Macskassy, S.A. and F. Provost. (2003), "A Simple Relational Classifier", In Proceedings of the Multi-Relational Data Mining Workshop (MRDM), Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

McPherson, M., L. Smith-Lovin and J.M. Cook, (2001), "Birds of a Feather: Homophily in Social Networks", Annual Review of Sociology, Vol. 27, 415-444

Richardson, M. and P. Domingos, (2002), "Mining Knowledge-Sharing Sites for Viral Marketing", SIGKDD 02, 61-70

Robins, G., P. Pattison, Y. Kalish and D. Lusher, (2007), "An Introduction to Exponential Random Graph (p*) Models for Social Networks", Social Networks, 29(2), 173-191

Rue H. and L. Held, “Gaussian Markov Random Fields – Theory and Applications”, 2005, Chapman & Hall/CRC

Rue, H., (2008) “Gaussian Markov Random Fields – Theory and Applications”, Lecture slides available online.

Wasserman, S. and K. Faust, (1994), “Social Network Analysis: Methods and Applications”, Cambridge University Press

Wei Z. and H. Li, (2007), “A Markov Random Field Model for Network-based Analysis of Genomic Data”, Bioinformatics