An Integrated Latent Traits Modeling Approach for Measuring Infectivity and Susceptibility of Multilaterally Interacted Individuals with Application to Fashion Contagion

Jiali Ding

North China Electric Power University, 2 Beinong Road, Beijing 102206, China
dingjiali@gsm.pku.edu.cn

Abstract

Process of contagion is driven not only by early infectious adopters but also by a critical mass of easily influenced individuals. An important research question is to statistically infer individuals' heterogeneous latent traits which can be used to rank and identify key entities and factors for precise contagion control. We propose a latent-traits modeling approach by extending IRT models to simultaneously incorporate and infer individual-level infectivity, susceptibility and baseline risk while controlling and estimating effects of environmental factors, from infection incidents and interaction data. And we offer Bayesian estimation with MCMC algorithms for the model.

The approach is illustrated with an application to fashion contagion where multilaterally connected customers' purchases across multiple products are potentially influenced by each other. The proposed model has better in-sample and out-sample fits than competing and benchmark models. The estimated results of the application show that the most infectious customers are not necessarily the most frequent buyers, nor the most connected ones; and that both the baseline adoption risks due to personal brand preferences and the sensitivities to marketing mix are highly heterogeneous among people.

Key Words: Latent traits model, Bayesian estimation, MCMC, contagion control, IRT

1. Introduction

Contagion is the spreading of a particular event, such as infectious disease, fashion, product adoption, financial shock, riot or any other contagious events, behaviors and opinions, by one susceptible entity directly or indirectly contacting another infectious entity. For example, like the spreading of infectious disease, fashion contagion happens when someone adopts a style of clothing or a way of behaving of another one. Financial contagion occurs when other institutions are affected by certain financial shocks following some initially affected institutions. It suggests that massive contagion may be driven not only by infectious individuals but also by a critical mass of susceptible individuals (Watts and Dodds 2007), while the former have more impact on the speed of the contagion process, the latter may have more impact on final size" (Goldenberg et al. 2009). The infectivity and susceptibility of humans, animals or other entities for an infectious disease can be individually heterogeneous, as different hosts might have different worm burden, genes, pathogenicity and immune levels (Keeling and Rohani

2008). For social contagion, individual entities' ability to influence or to be influenced can be variant due to certain personal traits such as being attractive, convincing, charismatic, magnetic, innovative, information-seeking, confident, and assertive, etc (Katz 1957; Weimann 1991; Chaney 2001; Thoburn, 2004; Goldenberg et al. 2009).

Measuring and ranking infectivity and susceptibility at individual level can help control the contagion process more precisely and efficiently. For example, having identified those most and least infectious and susceptible individuals who might have the potential of disproportionally affecting the spread of contagion, health professionals can optimize the use of limited medical resources by targeting them for more in-depth genetic examinations and pathological research, or more efficient vaccination and quarantine. Similarly, marketers can pick and rely on real customers, rather than expensive celebrities or models, as natural trendsetters to influence consumers' product adoption.

In epidemiology, infectivity is usually measured at the population level by incidence, which is the probability of occurrence of a given medical condition in a population within a specified period of time. At the individual level, infectivity is the ability of a specific individual to establish an infection. Intuitively, if after contacting with a particular infected individual, most people, no matter how strong they are, will get infected, we probably could infer that particular individual has high infectivity. Similarly, if many people always followed a particular individual to adopt similar clothing in various situations even when the clothing are expensive, have no brand and not in promotion, we probably could say that particular individual has high infectivity in fashion contagion. In this paper, we define an individual's infectivity as that individual's ability to increase a potentially susceptible individual's probability of being infected or behaving similarly, which means, a highly susceptible individual who contacts with a highly infectious individual will have higher probability to be infected or to behave similarly, controlling for effects of environmental factors and the baseline risk levels.

Rather than measuring infectivity without controlling for susceptibility, or measuring susceptibility without controlling for infectivity, or measuring both of them without controlling for environmental covariates, which could result in inaccurate and partial estimates, our proposed modelling approach includes all these factors into one integrative framework to capture various aspects of individual-level contagion dynamics and their synergism. Particularly, it includes (1) any two potential infectious and susceptible individuals' heterogeneous infectivity and susceptibility, (2) their infection status and contact or link strength, (3) effects from environmental or other observable factors, and (4) individuals' heterogeneous unobservable baseline infection risks. Essentially, the approach reveals individuals' persistent traits by relating them to individuals' behavioral interdependence in multilateral interactions across multiple occasions.

2. Literature Review

In social fields, questionnaires based on individual's personality are used to measure individuals' ability to affect others (Katz and Lazarsfeld 1955; King and Summers 1970; Chiders 1986; Weimann 1991; Chaney 2001; Keller and Berry 2004), but the self-report methods of questionnaires take what people claim themselves (or others) to be, and may involve high cost as the number of the surveyed people increases. Influencers cannot be identified by certain demographic or social-economic variables either. Prior research (Myers and Robertson 1972; Stern and Gould 1988; Chaney 2001) suggests that the

correlations of demographic variables with influence ability are low and insignificant in various areas.

The well-known degree centrality concepts in social network analysis (Freeman 1979) are often used to measure individuals' centrality characteristics via observed or inferred relations among people. But the "centrality" measurements based on the number of connections one have, measure "exposure" rather than "virulence", and commonly lack control for environmental factors, therefore hardly disentangle a person's intrinsic ability from confounding effects.

Most quantitative works related to contagion have primarily focused on population-scale epidemic rather than individual-level dynamics. Usually these works test the existence of certain contagion, or describe and simulate the diffusion processes in contagion using aggregate or segmental parameters of outbreak thresholds and state transferring probability, but seldom estimate heterogeneous individual-level traits. (Coleman, Katz and Menzel 1957; Case 1991; Dodds and Watts 2005; Luo et al. 2013) At individual level, some work has been done on inferring whether or not an individual entity influences his/her peers. For example, in a typical individual-level social influence framework, an individual *i*'s behavior (y_i) is modeled as dependent on the behavior of an individual *j* as his/her reference (y_j), as well as *j*'s other, usually observable, characteristics (z_j) as: $y_i = \alpha_i y_j + (\beta_i z_j) + \ldots$(Bell and Song 2007; Hartmann et al. 2008; Manchanda, Xie and Youn 2008; Nair, Manchanda and Bhatia 2010; Trusov et al. 2010) However, the focuses are to test dyad "yes-or-no" influence relationships, with few exception of measuring asymmetric two-way directional interactions (Iacobucci and Hopkins 1992).

Capable of inferring underlying trait and ability, item response theory based latent trait models are used for examinee-test interactions in psychometrics, but they commonly do not include environmental covariates for control (Birnbaum 1968; Hambleton and Swaminathan 1985; Kamakura, Ramaswami and Srivastava 1991; Bawa, Srinivasan and Srivastava 1997; Bradlow, Wainer and Wang 1999).

3. Model and Estimation

3.1 Model Specification

An interaction network is a set of multilaterally interacted individuals and the contact relation among them. A matrix N represents how the individual entities are linked with each other by physical contact in disease contagion or any form of communication in social contagion. If individual *i* contacts with *j*, the corresponding matrix element n_{ij} is 1, otherwise 0. The elements of the contact matrix can also be decimals or some weighted numbers to reflect the likelihood of contact between any two individuals, if the explicit and accurate contact between them is not available. For example, the spatial or temporal proximity of two individuals can be used as proxy of their contact likelihood (Robinson 2000; Yang and Allenby2003). In contagion, we also observe a sequence of infection/ adoption incidents of all the individuals, i.e., we know their infection status (infected or not infected) and the corresponding time.

We model the probability of the susceptible individual *i* infecting disease *l* (or adopting product *l*) at occasion *t* as

$$\mathrm{P}\left(\mathrm{Y}_{\mathrm{ijlt}}=1\middle|\zeta_{\mathrm{j}},\alpha_{\mathrm{i}}^{\mathrm{c}},\boldsymbol{\alpha}_{\mathrm{i}}^{-\mathrm{c}},\mathrm{x}_{\mathrm{jl(t-1)}},\mathrm{n}_{\mathrm{ijt}},\mathbf{X}_{\mathrm{ilt}}\right)=\Phi\left(\eta_{\mathrm{ijlt}}\right)$$

$$= \Phi\left(\alpha_i^c \zeta_j x_{jl(t-1)} n_{ijt} + \mathbf{X}'_{ilt}\boldsymbol{\alpha}_i^{-c}\right) = \int_{-\infty}^{\eta_{ijlt}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \qquad (1)$$

where ζ_j is the infectivity of potential infectious individual j, $x_{jl(t-1)}$ is the prior infection status or adoption behavior of j on disease or product l before occasion t, n_{ijlt} is the contact probability between i and j in term of l till occasion t, $\mathbf{X_{ilt}}=[X_{ilt1}, X_{ilt2,\ldots}, X_{iltk}]$ is i's confronted environmental factors and any relevant attributes at occasion t, α^c_i is i's susceptibility to contagion and vector α^{-c}_i is i's sensitivity to environmental factors and other attributes, and Y_{ijlt} is the incident or adoption behavior of individual i as being potentially infected by individual j, such that Y_{ijlt}=1 denotes an infected status or adoption for l on occasion t and Y_{ijlt}=0 otherwise. Different with aggregate-level models for infectious disease and product diffusion, this specification accounts for the contingent nature of contagion upon the individual-level intrinsic traits, infection/ adoption status and contact strength, as well as responses to environments and attributes. Here we allow that the infectivity and susceptibility to be continuous values, not only "yes-or-no" dichotomy or other discrete classifiers, so that they can be further used for comparison and ranking.

Sometimes an individual might get infected without contact with other individuals, and thus bears a baseline infection probability greater than zero. Similarly, in some situations, an individual may have non-negligible preference for certain products, which contributes to positive baseline adoption likelihood. We further let β_{il} denote the baseline probability of i getting infected disease (or adopting product) l due to unobservable preference and non-contagion factors. Then the probability of i getting infected or adopting product l on occasion t is

$$P\left(Y_{ijlt} = 1 \middle| \zeta_j, \alpha_i^c, \boldsymbol{\alpha}_i^{-c}, \beta_{il}, x_{jl(t-1)}, n_{ijlt}, \mathbf{X}_{ilt}\right)$$
$$= \beta_{il} + (1-\beta_{il})\Phi\left(\alpha_i^c \zeta_j x_{jl(t-1)} n_{ijlt} + \mathbf{X}'_{ilt}\boldsymbol{\alpha}_i^{-c}\right) \qquad (2)$$

This enhancement makes our model more general as we can see that if β_{il}=0, i.e., individual i has zero baseline infection probability for disease (or product) l due to unobservable and non-contagion factors, then the model reduced to Equation (1).

3.2 Bayesian Estimation for the Model

Since the model is over parameterized and the Maximum Likelihood Estimation is not identifiable. We provide a Markov Chain Monte Carlo estimation which is based on the one that proposed by Johnson and Albert (1999), but extends it to incorporate observable covariates and additional parameters in a setting of multilateral interactions among infectious and susceptible individuals.

To implement the Gibbs sampler, a latent variable W is introduced so that $W_{ijlt} \geq 0$ if Y_{ijlt} =1 and $W_{ijlt} < 0$ otherwise. Under the assumption of standard normal distributed error term, the joint posterior distribution of (W, α, ζ) conditional on the observed data is

$$p(\boldsymbol{\alpha}, \zeta, W|\text{data}) \propto p(\text{data}|W)p(W|\boldsymbol{\alpha}, \zeta)p(\boldsymbol{\alpha})p(\zeta) \qquad (3)$$

We assume a normal prior for the individual parameter $\zeta_j \sim N(\nu, \sigma^2)$, which means that the individuals' infectivity is normally distributed among the population. For the susceptibility and sensitivity parameter vector of the susceptible entity i, we assume a conjugate multivariable normal prior $\boldsymbol{\alpha_i} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and restrict the susceptibility parameter α_i^c to be positive which guarantees that individual i who interacts with individual j with a positive (/negative) ζ_j should have a higher (/lower) probability to

infect or adopt. With prior distributions being specified, the fully conditional distributions of *W*, $\boldsymbol{\alpha}$, and ζ for Bayesian estimation are given as below:

$$f(W_{ijlt}|*) \propto \text{TruncatedN}(\alpha_i^c \zeta_j x_{jl(t-1)} n_{ijlt} + \mathbf{X}'_{ilt}\boldsymbol{\alpha}_i^{-c}, 1) \quad (4)$$

$$f(\zeta_j|*) \propto N\left(\left(\mathbf{D}^T\mathbf{D} + \frac{1}{\sigma^2}\right)^{-1}\left(\mathbf{D}^T\mathbf{C} + \frac{v}{\sigma^2}\right), \left(\mathbf{D}^T\mathbf{D} + \frac{1}{\sigma^2}\right)^{-1}\right) \quad (5)$$

$$f(\boldsymbol{\alpha_i}|*) \propto N\left(\left(\mathbf{F}^T\mathbf{F} + \boldsymbol{\Sigma}^{-1}\right)^{-1}\left(\mathbf{F}^T\mathbf{E} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right), \left(\mathbf{F}^T\mathbf{F} + \boldsymbol{\Sigma}^{-1}\right)^{-1}\right) I(\alpha_i^c > 0) \quad (6)$$

where

$$\mathbf{C} = \left[W_{ijlt} - \mathbf{X}'_{ilt}\boldsymbol{\alpha}_i^{-c}\right] \quad (7)$$

$$\mathbf{D} = \left[\alpha_i^c x_{il(t-1)} n_{ijt}\right] \quad (8)$$

$$\mathbf{E} = \left[W_{ijlt}\right] \quad (9)$$

$$\mathbf{F} = \left[\zeta_j x_{ijl(t-1)} n_{ijt} \ \mathbf{X}'_{ilt}\right] \quad (10)$$

To estimate the enhanced model, a new latent variable Z is included such that

$$p(Z_{ijlt} = 1|Y_{ijlt} = 1) = \Phi\left(\alpha_i^c \zeta_j x_{jl(t-1)} n_{ijlt} + \mathbf{X}'_{ilt}\boldsymbol{\alpha}_i^{-c}\right); \ (Z_{ijlt} = 0|Y_{ijlt} = 1)$$
$$= \beta_{il}\left(1 - \Phi\left(\alpha_i^c \zeta_j x_{jl(t-1)} n_{ijlt} + \mathbf{X}'_{ilt}\boldsymbol{\alpha}_i^{-c}\right)\right) \quad (11)$$

$$p(Z_{ijlt} = 1|Y_{ijlt} = 0) = 0; \ p(Z_{ijlt} = 0|Y_{ijlt} = 0) = 1 \quad (12)$$

And the joint posterior distribution of (*Z*, *W*, $\boldsymbol{\alpha}$, ζ, β) conditional on the observed data is

$$p(\boldsymbol{\alpha}, \zeta, \beta, W, Z|\text{data}) \propto p(\text{data}|Z, \beta)p(Z|W)p(W|\boldsymbol{\alpha}, \zeta)p(\boldsymbol{\alpha})p(\beta)p(\zeta) \quad (13)$$

Assuming a conjugate Bata prior distribution $\beta_{il} \sim Bata(c_{il}, d_{il})$, the posterior conditional distributions of β_{il} is as below:

$$f(\beta_{il}|*) \propto \text{Bata}\left(c_{il} + \sum_{(t|Z_{ilt}=0)} Y_{ilt}, d_{il} + \sum_t I(Z_{ilt} = 0) - \sum_{(t|Z_{ilt}=0)} Y_{ilt}\right) \quad (14)$$

4. Application

We illustrate the modeling approach with an application to a context of fashion purchase contagion where the member customers who shop in a fashion mall are potentially affected by marketing mix and product attributes as well as other customers nearby. Like the spreading of infectious disease between two people, contagion occurs when a shopper purchases the similar clothing bought by others.

4.1 Data Description

From a big fashion mall we obtain the scanner data which contains information on member customers' purchases. The data is over a two year period and has no left censoring as we have product-level purchase information for individual member customers from the day of store launch. Figure 1 displays a simplified purchase sequence in which A, B, C, ..., and Z are customers and different colors represent different products.

For our analysis, we randomly pick 200 people to construct a customer network such that each of them has made a sequential same-product same-day purchase with at least one other person. This restriction is similar to that used in spatial statistics for choosing which

regions to include in the analysis. Typically, only those regions are selected that have at least one other region as a neighbor (Case 1991). For any two customers linked with such a sequential purchase, we have a focal customer (termed as customer *i*) and a potential influence source (termed as customer *j*) with the latter being the customer who purchases the same product earlier than the former in the same day. This randomly constructed customer network involves 101 unique j-type customers and 111 unique i-type customers, wherein one j-type customer may have multiple respective i-type customers and vice versa (like the one showed in Figure 2). For constructing the final data set for our model, we track all the purchase records of these customers in two years. We use the first 90% observations as the calibration data for our model building and the last 10% observations as the holdout dataset for prediction assessment.

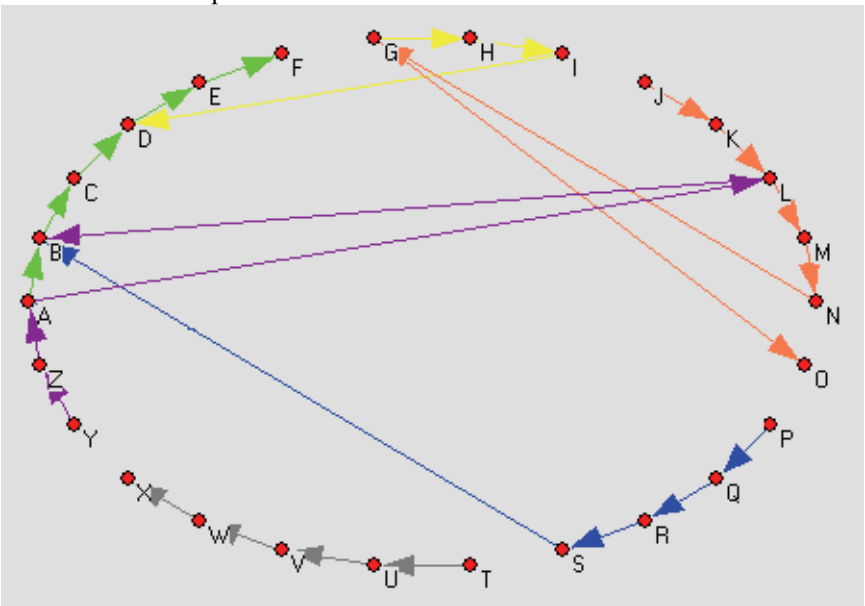


Figure 1: Purchase sequence, the directional links depict the purchase sequence among customers and different colors denote different products.

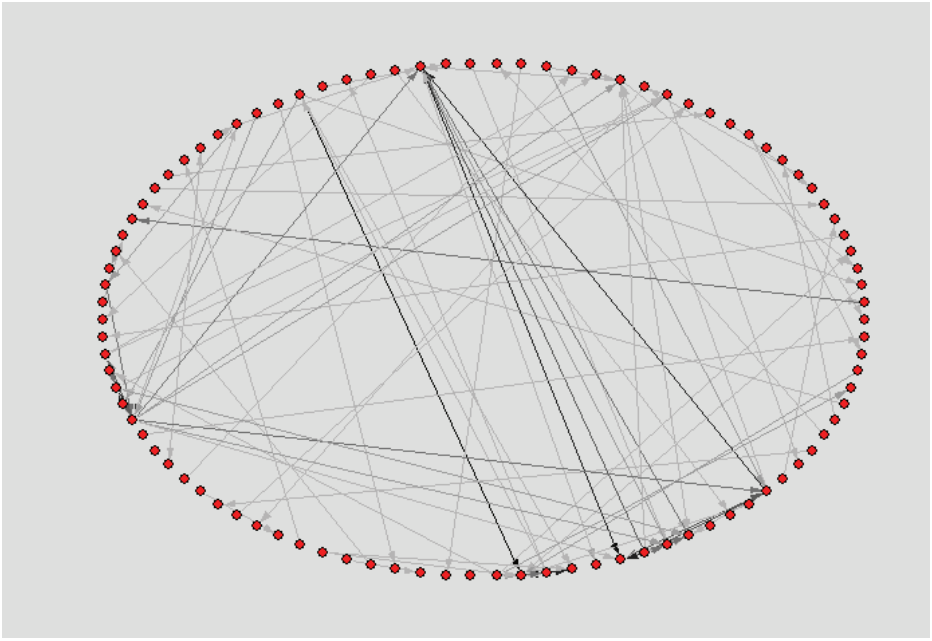
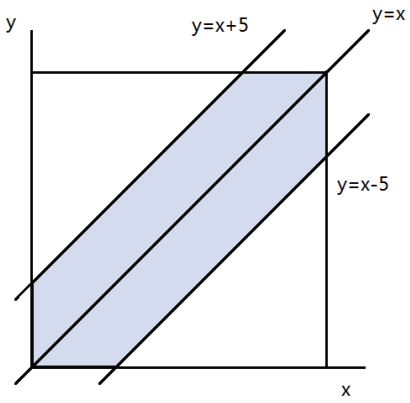


Figure 2: Multilaterally connectted customers

For a customer *i*, we calculate the exposure to his or her respective customer *j* using the construct of meeting probability, i.e., the probability that they might see each other in the store. We use their checkout times from the retail store to ascertain this probability. The higher is the meeting probability, the greater is the chance that customer *i* observes customer *j*'s behavior in the store. Figure 3 illustrates how to calculate the meeting probability via the use of geometry. In our context, we know the checkout times of each customer and also how long they spend between intermediate checkouts.[1] Therefore, we can estimate the total time each customer spends for a shopping trip. In addition, we know the number of items a customer purchases in the trip. This gives the average time a customer spends on purchasing one unit of product. We use this information from each customer to find the meeting probability between any two customers. Figure 4 shows three examples of meeting probability calculated from the data.



[1] The fashion mall has a store-within-a-store format and customers typically checkout multiple times within a single visit.

Figure 3: Geometric representation for calculating the meeting probability between two customers, the x-axis denotes the arrival times for customer j and the y-axis denotes the arrival times for customer i. The shaded area represents the intersection of the two customers within a 5 minute interval.

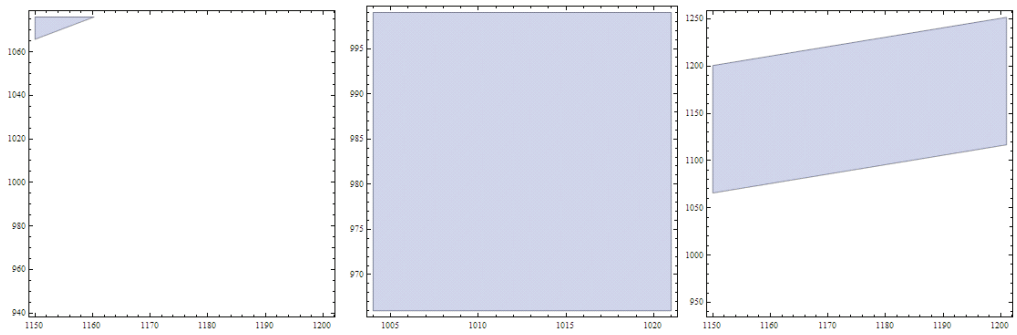


Figure 4: Three examples of meeting probability calculated from the data

4.2 Model Comparison

We ran the two proposed models specified as Equation (1) and Equation (2), six alternative models with slight variations (with or without the infectivity trait parameter ζ_j and the contact strength measurement MP_{ijlt}) to these two models, and six benchmark models such as fixed effects model, random intercept model and standard Probit model with their slight variations. For example, one alternative model has individual random parameters $\boldsymbol{\alpha_i}$ and has neither ζ_j nor MP_{ijlt}.

We compare the model goodness of fit of our proposed models with alternative and benchmark models using calibration data and hold-out data. For comparing the models, the log-marginal likelihoods (Newton and Raftery 1994) which measures in-sample fit, the mean absolute deviation of estimated choice probability and actual choice (Yang and Allenby 2003) which measures the out-sample fit, and the Bayesian deviance information criterion (DIC) which measures Bayesian model's fit and penalizes for model complexity (Spiegelhalter et al. 2002) of each model are calculated. The results for model comparison based on these three criteria are reported in Table 1.

Table 1: Model Goodness of Fit Comparison

Model Specification	*LML*	*Out-sample Fit*	*DIC*
$\eta_{ilt}=\alpha_i^0+\alpha_i^1\zeta_j Y_{jl(t-1)}MP_{ijlt}+...;\quad P(Y_{ilt}=1)=\Phi(\eta_{ilt})+\beta_{ib(l)}[1-\Phi(\eta_{ilt})]$	-11609	**0.0373**	**23497**
$\eta_{ilt}=\alpha_i^0+\alpha_i^1 Y_{jl(t-1)}MP_{ijlt}+...;\quad P(Y_{ilt}=1)=\Phi(\eta_{ilt})+\beta_{ib(l)}[1-\Phi(\eta_{ilt})]$	-11652	0.0374	23592
$\eta_{ilt}=\alpha_i^0+\alpha_i^1\zeta_j Y_{jl(t-1)}+...;\quad P(Y_{ilt}=1)=\Phi(\eta_{ilt})+\beta_{ib(l)}[1-\Phi(\eta_{ilt})]$	-11691	0.0378	23607
$\eta_{ilt}=\alpha_i^0+\alpha_i^1 Y_{jl(t-1)}+...;\quad P(Y_{ilt}=1)=\Phi(\eta_{ilt})+\beta_{ib(l)}[1-\Phi(\eta_{ilt})]$	-11702	0.0375	23659
$\eta_{ilt}=\alpha_i^0+\alpha_i^1\zeta_j Y_{jl(t-1)}MP_{ijlt}+...;\quad P(Y_{ilt}=1)=\Phi(\eta_{ilt})$	**-11578**	0.0394	23627
$\eta_{ilt}=\alpha_i^0+\alpha_i^1 Y_{jl(t-1)}MP_{ijlt}+...;\quad P(Y_{ilt}=1)=\Phi(\eta_{ilt})$	-11605	0.0396	23650

$\eta_{ilt} = \alpha_i^0 + \alpha_i^1 \zeta_j Y_{jl(t-1)} + \ldots; \quad P(Y_{ilt} = 1) = \Phi(\eta_{ilt})$	-11631	0.0408	23729
$\eta_{ilt} = \alpha_i^0 + \alpha_i^1 Y_{jl(t-1)} + \ldots; \quad P(Y_{ilt} = 1) = \Phi(\eta_{ilt})$	-11653	0.0402	23814
$\eta_{ilt} = \alpha_i + \alpha^0 + \alpha^1 Y_{jl(t-1)} MP_{ilt} + \ldots; \quad P(Y_{ilt} = 1) = \Phi(\eta_{ilt})$	-11890	0.0406	N.A.
$\eta_{ilt} = \alpha_i + \alpha^0 + \alpha^1 Y_{jl(t-1)} + \ldots; \quad P(Y_{ilt} = 1) = \Phi(\eta_{ilt})$	-11915	0.0407	N.A.
$\eta_{ilt} = \alpha^0 + u_i + \alpha^1 Y_{jl(t-1)} MP_{ilt} + \ldots; \quad P(Y_{ilt} = 1) = \Phi(\eta_{ilt}); \quad u_i \sim N(0,1)$	-12020	0.0407	N.A.
$\eta_{ilt} = \alpha^0 + u_i + \alpha^1 Y_{jl(t-1)} + \ldots; \quad P(Y_{ilt} = 1) = \Phi(\eta_{ilt}); \quad u_i \sim N(0,1)$	-12047	0.0408	N.A.
$\eta_{ilt} = \alpha^0 + \alpha^1 Y_{jl(t-1)} MP_{ijlt} + \ldots; \quad P(Y_{ilt} = 1) = \Phi(\eta_{ilt})$	-12299	0.0413	N.A.
$\eta_{ilt} = \alpha^0 + \alpha^1 Y_{jl(t-1)} + \ldots; \quad P(Y_{ilt} = 1) = \Phi(\eta_{ilt})$	-12330	0.0415	N.A.

The comparison shows that the proposed models Equation (1) and (2) perform better than alternative and benchmark models for both the calibration and the hold-out datasets. The model specified as Equation (1) has the best in-sample fit, while the general model specified as Equation (2) has the second best in-sample fit but the best out-sample fit as well as best DIC. This suggests that incorporating the personal brand preference (βib(l)) into the model improves out-sample prediction. The comparison also reveals that including infectivity (ζ_j) and the contact strength (MP_{ijlt}) into the model improves both in-sample and out-sample fit, and that the contact strength measurement helps to improve fit performance for the benchmark models too.

4.3 Results and Analysis

The estimated results for the proposed models show that, among our randomly constructed sample customer network with 111 *i*-type customers and 101 *j*-type customers, one has statistically significant and positive estimate of infectivity, and five customers have statistically significant estimate of susceptibility toward other customers' infection.

Table 2 reports the posterior estimate of infectivity of the 20 *j*-type customers with the smallest and largest ζ_j estimates, as well as the total number of purchases and connectors of each of them. It's worth noting that the customers who are statistically infectious are not necessarily the most frequent buyers nor the most connected customers.

Table 2: Posterior estimates for infectivity of the 20 *j*-type customers with the smallest and largest ζ_j, as well as their total numbers of purchases and connectors. The numbers in bold indicate that the 0 lies outside of the 95% interval of the estimates

Customer ID	*Model (2)*	*Model (1)*	*Num of Purchases*	*Num of Connectors*
1	0.599	0.624	16	1
2	-0.245	-0.440	39	1
3	0.959	0.989	6	1
4	1.013	0.964	25	1
5	0.566	0.567	24	1
6	-0.457	-0.566	34	1

7	-0.211	-0.230	6	1
8	-0.368	-0.424	17	1
9	1.409	0.833	147	21
10	1.171	**1.230**	24	4
11	-0.285	-0.309	23	1
12	-0.774	-0.561	39	1
13	1.013	0.797	20	4
14	0.990	1.078	52	1
15	-0.428	-0.334	65	1
16	-0.576	-0.622	44	4
17	0.896	0.931	11	1
18	-0.362	-0.317	27	1
19	-0.268	-0.302	62	1
20	**1.867**	**1.210**	91	13

Table 3 provides the posterior estimates of susceptibility and sensitivity parameters of the 18 *i*-type customers with smallest and largest estimated α^1_i for the proposed model Equation (1) and (2)..

Table 3: Posterior estimates of susceptibility and sensitivity parameters for Equation 2 (upper) and Equation 1 (lower). The numbers in bold indicate that the 0 lies outside of the 95% interval of the estimates.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
α^1_i	**35.02**	6.67	32.85	34.35	10.85	15.01	1.07	5.04	**41.16**	**40.77**	35.38	17.12	10.66	15.67	**37.57**	16.93	**37.85**	32.84
α^2_i	0.16	**1.60**	**3.40**	**3.80**	0.27	**3.58**	**1.33**	**0.73**	1.06	-0.13	3.74	**5.83**	0.19	2.56	1.73	**7.09**	0.24	**8.23**
α^3_i	-0.29	**-0.49**	-0.19	-0.10	**2.09**	-0.25	-0.03	**-0.18**	-0.01	-0.61	1.07	-0.27	**-0.71**	0.37	-1.12	-1.26	-0.90	-0.44
α^4_i	-0.73	**0.61**	0.64	0.58	7.91	0.54	**4.76**	**1.15**	**3.15**	5.62	4.52	**11.38**	**0.82**	-0.24	0.44	**6.60**	0.46	**1.42**
α^5_i	-0.42	**0.41**	0.07	0.27	**-2.19**	0.54	-0.20	**0.17**	0.12	-0.35	-0.62	0.73	**0.54**	-0.32	1.30	**1.53**	-0.75	0.45
α^6_i	0.09	0.11	0.06	-0.05	0.18	0.51	-2.74	**-5.46**	0.29	-0.72	-0.21	0.35	**0.10**	1.83	-1.29	1.45	-0.07	-0.31
α^1_i	**25.05**	7.88	35.92	32.05	8.07	17.15	1.02	4.72	**41.35**	28.04	16.17	16.95	19.58	13.43	**39.73**	37.13	26.45	30.82
α^2_i	0.39	**1.60**	**3.37**	**3.80**	0.38	**3.56**	**1.25**	**0.74**	1.07	-0.41	4.32	**5.55**	0.18	3.32	**1.79**	**6.72**	0.95	**8.30**
α^3_i	0.47	**-0.49**	-0.16	-0.09	**1.57**	-0.24	-0.04	**-0.18**	0.02	-0.36	0.24	-0.50	**-0.71**	0.75	-0.79	**-1.34**	0.85	-0.45
α^4_i	-0.16	**0.62**	0.69	0.57	**3.23**	0.38	**6.39**	**1.15**	**2.57**	2.97	**4.24**	**2.88**	**0.81**	-1.30	0.28	**4.16**	-4.75	**1.33**
α^5_i	-0.52	0.40	0.03	0.27	-1.72	0.54	-0.21	0.16	-0.02	-0.43	0.15	0.73	0.54	-0.89	0.95	1.46	-1.36	0.44
α^6_i	0.44	0.12	0.03	0.13	0.29	0.48	**-1.55**	**-3.89**	0.13	-1.84	0.23	**0.63**	0.11	0.91	-1.50	0.62	0.14	-0.24

The plot of these individual-level estimates of susceptibility and sensitivities to marketing mix (α^1_i, α^3_i, α^4_i) for all the 111 *i*-type customers are presented in Figure 5. We notice that some people with high price and promotion sensitivities are very susceptible to others' influence. We also notice that there are people who are extremely insensitive to promotion while very susceptible to influence. This suggests that marketers should offer differentiated incentives according to individual customer's unique traits profiles.

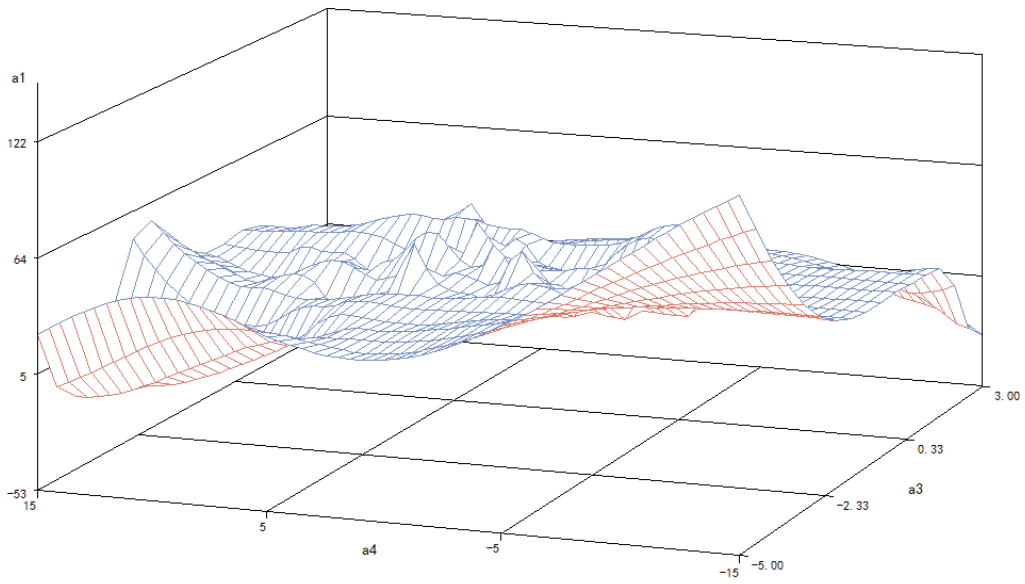


Figure 5: Plot of posterior estimates of susceptibility to influence (α^1_i) and sensitivities to price and promotion (α^3_i, α^4_i)

Figure 6 shows the plots of posterior estimate of personal brands preferences of the 111 *i*-type customers, which suggests that personal preferences toward brands are highly heterogeneous among different individuals and across different brands even for the same individual.

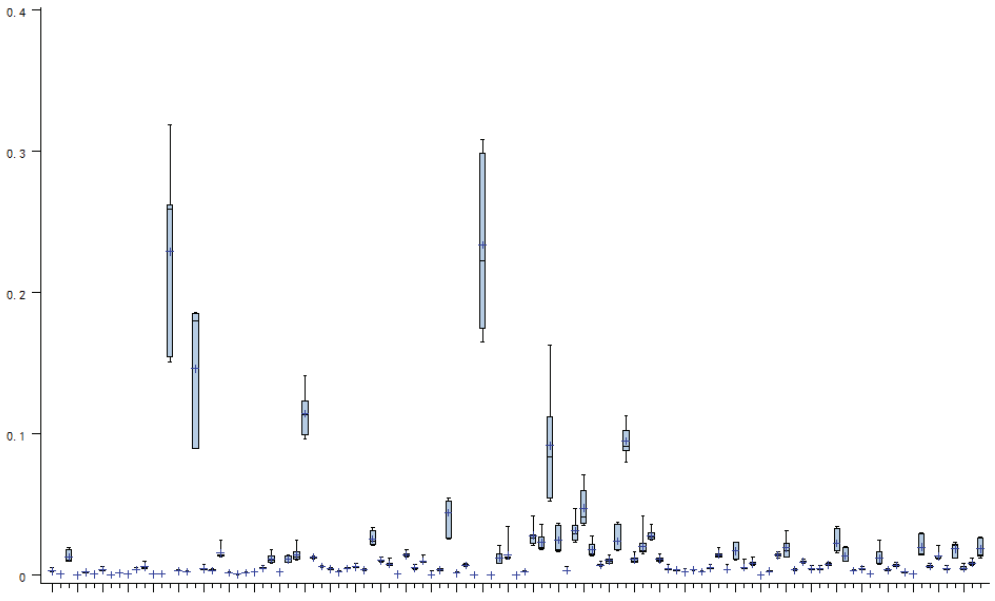


Figure 6: Plots of posterior estimates of individuals' brand preferences of 111 *i*-type customers

Using the estimated parameters for each individual, we further calculate the difference in the potential susecptible customer's purchase probability when the respective potential

infectious member buys or does not buy. We find that, across all purchase occasions involving the one most infectious *j*-type customer (with a statistically significant positive estimate of ζ_j, called J) in the dataset, there is an average increase of 0.0371 in the probability of purchase from her respective *i*-type customers if J buys, as compared to an average purchase probability of 0.0359 when J does not buy. The increase in individual purchase probability, combined with product prices, results in an increase of about $5,977 in the overall expenditure from the *i*-type customers. This increase also indicates that, the most infectious *j*-type customer, J, by exerting influence on respective *i*-type customers in the sample, is worth $5,977 more than what she herself has purchased ($3,227) in two years, which translates to a network value of around 185% of her intrinsic value. When scaled up roughly based on the number of customers in our sample (about 200 persons) and the member customer pool (the fashion mall has about 20000 members), the total economic impact on the fashion mall due to contagion for the two year period is as high as $597,700.

5. Conclusion and Discussion

We propose an integrated modeling framework for simultaneously estimating individually heterogeneous infectivity, susceptibility and sensitivities as well as baseline risks or preferences for interacted individuals based on their infection/ adoption incidents and interaction data. The approach extends IRT model with covariates, infection status and contact matrix, and captures the contingent and complex nature of dynamics in contagion. And we offer Bayesian estimation for the proposed model.

We illustrate this approach by applying it to fashion adoption where customers are potentially affected by other consumers in physical and temporal proximity. Without direct data for customer interactions, we introduce a measure to infer the likelihood of interaction among individual customers. Our proposed model performs better for both the in-sample and the holdout data predictions than alternative and benchmark models. We discover that the individuals who are statistically infectious are not necessarily the most frequent buyers nor the most connected customers in the network, and customers are highly heterogeneous in terms of their baseline risks or preferences as well sensitivities to environmental factors. Marketers can optimize offerings to individual customer based on their unique latent traits profile.

In this illustration application the interaction (i.e., meeting probability) matrix is inferred due to data unavailability, and the inclusion of it has increased model goodness of fit. However, the inference and construct about social network structure is not this paper's research object. The proposed modeling framework can be applied to contexts in which the interaction links among individuals are known.

We have restricted the susceptibility parameter to be positive and allow the infectivity parameter to be positive or negative. As prior research has found that the desirability of an influence source (i.e., whether being an attractive person) will moderates a consumer's evaluations toward the focal product. For example, a customer will have a higher (/lower) evaluation toward a product if he/she observed the product's being touched by another attractive (/undesirable) customer (Morales and Fitzsimons 2007; Argo, Dahl and Morales 2008). There are other possible restriction choices according to specific application contexts. And we notice that, based on our application context, the model

estimation finds no *j*-type customer who has a statistically significant negative estimate of latent infectivity in our sample data, we also calculate and find that across all purchase occasions involving the 47 unattractive j-type customers (with an insignificant negative estimate of ζj) in the dataset, there is an average decrease of 0.0003 in the probability of purchase from the i-type customers if the unattractive *j*-type customers buy, as compared to an average base probability of 0.0279 when they do not buy. The individual decrease in purchase probability, combined with product prices, results in a decrease of about $340 in the overall expenditure from the i-type customers or about $34,000 for the fashion mall when scaled up. Clearly, the positive contagion effect is much greater than the negative effect for this application.

References

Argo, Jennifer J., Darren W. Dahl and Andrea C. Morales (2008), "Positive Consumer Contagion: Responses to Attractive Others in a Retail Context," *Journal of Marketing Research*, Vol. 45, No. 6, 690-701.

Bawa, Kapil, Srini S. Srinivasan and Rajendra K. Srivastava (1997), "Coupon Attractiveness and Coupon Proneness: A Framework for Modeling Coupon Redemption," *Journal of Marketing Research*, 34 (4), 517-525.

Bell, David R. and Sangyoung Song (2007), "Neighborhood effects and trial on the Internet: Evidence from Online Grocery Retailing," *Quantitative Marketing and Economics*, 5(4), 361-400.

Birnbaum, A., (1968), "Some latent trait models and their use in inferring an examinee's ability," In Lord, F.M. & Novick, M.R. (Eds.), Statistical Theories of Mental Test Scores. Reading, MA: Addison–Wesley,

Bradlow, E. T., Wainer, H., and Wang, X (1999), "A Bayesian Random Effects Model for Testlets, " *Psychometrika*, 64, 153-168.

Chaney, Isabella (2001), "Opinion Leaders as a Segment for Marketing Communications," *Marketing Intelligence and Planning*, 19(5), 302-308

Coleman, James, Elihu Katz and Herbert Menzel (1957), "the Diffusion of an Innovation among Physicians," *Sociometry*, 20 (December), 253-270.

Dodds and Watts (2005), "A Generalized Model of Social and Biological Contagion," *Journal of Theoretical Biology*, 232 (4), 587–604.

Case, Anne C. (1991), "Spatial Patterns in Household Demand," *Econometrica*, 59 (4), 953–965.

Chiders, Terry L. (1986), "Assessment of the Psychometric Properties of an Opinion Leadership Scale," *Journal of Marketing Research*, 23, 184-188.

Coleman, James, Elihu Katz and Herbert Menzel (1957), "The Diffusion of an Innovation among Physicians," *Sociometry*, 20 (December), 253-270.

Freeman, L., (1979), "Centrality in Social Networks: Conceptual Clarification," *Social Networks*, 1, 215-239.

Goldenberg, Jacob, Sangman Han, Donald R. Lehmann and Jae Weon Hong (2009), "The Role of Hubs in the Adoption Processes," *Journal of Marketing*, 73, 1-13.

Hambleton and Swaminathan, Item Response Theory: Principles and Applications. Boston: Kluwer Academic Publishers, 1985.

Hartmann, Wesley R., Puneet Manchanda, Harikesh Nair, Matthew Bothner, Peter Dodds, David Godes, Kartik Hosanagar and Catherine Tucker (2008), "Modeling Social Interactions: Identification, Empirical Methods and Policy Implications," *Marketing Letters,* 19(3), 287-304.

Iacobucci and Hopkins 1992 "Modeling Dyadic Interactions and Networks in Marketing," *Journal of Marketing Research*; Feb 1992; 29, 1,5-17

Johnson, Valen E. and James H. Albert (1999), Ordinal Data Modeling, New York: Springer.

Kamakura, Wagner A., Sridhar N. Ramaswami and Rajendra K. Srivastava (1991), "Applying Latent Trait Analysis in the Evaluation of Prospects for Cross-Selling of Financial Services," *International Journal of Research in Marketing*, 8(4), 329-349.

Katz, Elihu (1957), "The two-step flow of communication: An Up-to-date Report on an Hypothesis," *Public Opinion Quart*, 21, 61–78

Katz, Elihu and Paul F. Lazarsfeld (1955), Personal Influence: the Part Played by People in the Plow of Mass Communications, New York: Free Press.

Keeling and Rohani (2008), Modeling Infectious Diseases in Humans and Animals. NJ, USA.

Keller, Edward and Jonathan Berry (2003), The Influentials, New York: Free Press.

King, Charles W. and John O. Summers (1970), "Overlap of Opinion Leadership Across Consumer Product Categories," *Journal of Marketing Research*, 7, 43-50.

Luo et al, "Identifying Infection Sources and Regions in Large Networks", *Signal Processing, IEEE Trans*, 61 (11), 2013.

Manchanda, Puneet, Ying Xie and Nara Youn (2008), "The Role of Targeted Communication and Contagion in New Product Adoption," *Marketing Science*, 27(6), 950-961.

Myers, James H. and Thomas S. Robertson (1972), "Dimensions of Opinion Leadership," *Journal of Marketing Research*, 41-46.

Morales, Andrea C. and Gavan J. Fitzsimons (2007), "Product Contagion: Changing Consumer Evaluations through Physical Contact with 'Disgusting' Products," Journal of Marketing Research, 44 (2), 272-283.

Nair H, Manchanda P and Bhatia T. (2010). "Asymmetric Social Interactions in Prescription Behavior: The Role of Opinion Leaders," *Journal of Marketing Research*, Vol.47 (5), 883-895..

Newton, Michael A. and Adrian E. Raftery (1994), "Approximating Bayesian Inference with the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society (B)*, 56 (1), 3–48.

Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin and Angelika van der Linde (2002), "Bayesian Measures of Model Complexity and Fit (with Discussion)," *Journal of the Royal Statistical Society, Series B (Statistical Methodology), 64 (4), 583–639.*

Stern, B. and S.J. Gould (1988), "The Consumer as Financial Opinion Leader," *Journal of Retail Banking*, 10 (summer), 47-49.

Thoburn Patrick (2004), "Who are the Influencers," *Marketing*, 109 (29), 21

Trusov, Michael, Anand Bodapati and Randolph E. Bucklin (2010) , "Determining Influential Users in Internet Social Networks," *Journal of Marketing Research*, Vol. 47 (4), 643-658.

Watts, Duncan J. and Peter Sheridan Dodds (2007), "Influentials, Networks, and Public Opinion Formation" *Journal of Consumer Research*, 34(4), 441-458

Weimann, Gabriel (1991), "The Influentials: Back to the Concept of Opinion Leaders?" *Public Opinion Quarterly*, 55(2), 267-279.

Yang, Sha and Greg M. Allenby (2003), "Modeling Interdependent Consumer Preferences," *Journal of Marketing Research*, 40(3), 282–294.