The New England Journal of Statistics in Data Science Volume 4, 77–89 (2026)
DOI: https://doi.org/10.51387/26-NEJSDS94

Bayesian Estimation of Contagion Effect: An Application of Friendship Networks and Alcohol Behavior

Brisilda NDREKA, Dipak K. DEY, and Victor H. LACHOS*

Abstract

Social networks primarily focus on the phenomenon of the contagion effect when examining behavior patterns within specific social groups. However, the impact of peer effects is characterized by the tendency to imitate the behaviors of friends and the selection process, where individuals tend to affiliate with others sharing similar traits, significantly contributing to shaping social behaviors that are frequently interconnecting. This article presents a Bayesian approach that uses latent-space estimation methods to detect and examine contagion effects, considering the impact of social selection. The research provides a methodological explanation, followed by a sequence of simulation trials designed to explore operational functionalities and possible real-world applications. To illustrate the potential correlation between changes in alcohol use and the influence of social networks, this study concludes by presenting an example of adolescent drinking behavior.

keywords and phrases: Bayes analysis, Contagion effect, Latent space, Social selection, Friendship network.

1. INTRODUCTION

Research on social networks has primarily focused on understanding how information, trends, behaviors, and other elements are transmitted among members. This process is commonly called contagion effects, peer effects, or social influence. It has been studied across various disciplines, including economics [17], public health [21], biology [23], and physics [41]. Notable researchers like [36], [11] have conducted important studies on the structure and impact of social network analysis on social behavior. Their findings suggest that, besides the physical environment, the social environment significantly influences social behaviors. While many studies have examined relational information in network datasets, a key challenge in identifying true contagions is distinguishing social influence from homophily or social selection, which refers to the tendency of individuals with similar attributes or behaviors to form relationships with one another. [22] provides evidence suggesting that differentiating between these effects is theoretically infeasible within a static setting. Furthermore, as highlighted by [33] and [44], the failure to consider unobserved homophily leads to a significant overestimation of the influence effects.

Several research efforts have been undertaken to identify contagion effects across diverse contexts. For instance, [34] and [35] explicitly model observed homophily. [3] differentiate between influence effects and observed homophily in the adoption of mobile service applications. [10] concentrates on identifying the influence of homophily in decision-making processes. [7] utilizes indirect ties from third parties as instrumental variables. [20] scrutinize the impacts of both influence and homophily effects. Furthermore, some studies employ latent models to explore contagion effects, such as in the work conducted by [44].

Even though this field has seen much work done, more needs to be addressed. This is partly due to the limitations of different approaches. For instance, specific methods rely on strong assumptions, limiting the scope of data that can be analyzed. Other studies use homophily as an adjustable variable, which can help estimate the social influence, but may not effectively evaluate the presence of peer selection. The accurate evaluation of homophily is crucial since it directly affects the development of efficient targeting techniques. It is also important to acknowledge that while each approach can use additional data to mitigate bias, none of these methods can eliminate all sources of bias.

As an alternative to current techniques for evaluating social influence, this study proposes a hierarchical Bayesian model that simultaneously accounts for both homophily and the effect of social influence. The model effectively addresses the problem of latent homophily by using latent space positions from a latent space model analysis. Additionally, this approach can be applied to various types of network data, not limited to social network analysis. We put the model to the test using data collected at three different time points and with moderate network sizes, but it can be adapted to fit a wide range of network characteristics. Because of the

*Corresponding author: (Victor H. Lachos) 215 Glenbrook Road, Unit 4120 Storrs, Connecticut 06269-4120, U.S.A. E-mail address: hlachos@uconn.edu

nonstandard form of the posterior distribution, the computation of posterior moments estimates is a very hard task. Also, it is difficult to generate samples from these posterior data using traditional Monte Carlo methods. A reliable alternative is to develop an MCMC-type algorithm. Thus, we use `Stan` software [8], which presents a viable solution by efficiently managing the specified sampling schemes to conduct inference using the Hamiltonian Monte Carlo (HMC) algorithm. With `Stan`, we only need to specify the Bayesian model in `Stan`'s modeling language and the software returns samples from the target density. To facilitate reproducibility, the `Stan` code is available on `GitHub` at https://github.com/nbrisilda-git/Bayesian_est_peer_effect.

The remainder of the paper is organized as follows. Section 2 discusses the background and challenges associated with estimating contagion effects. The proposed estimation model is introduced, highlighting its potential to accurately identify contagion effects. Additionally, simulation studies are conducted to evaluate the performance of the proposed method. This section also demonstrates the superior performance of the proposed Bayesian approach compared to existing methods. Section 3 analyzes the longitudinal network of adolescent drinking behavior. Various aspects of model fitting are discussed and posterior predictive checks for goodness of fit are conducted. Lastly, Section 4 provides a summary and discusses future research directions.

2. A BAYESIAN LATENT TRAIT SOCIAL NETWORK MODEL

2.1 Background

Studies have shown that people on social networks tend to share similar characteristics and behaviors. This is mainly due to two factors: the contagion effect, which involves individuals influencing their friends to adopt similar characteristics and behaviors [12, 22], usually referred as influence effect; and homophily [4], which refers to the tendency for individuals with similar features or behaviors to form connections with each other, which also be called selection effect. The challenge lies in distinguishing between these two different processes due to their similar diffusion patterns. To illustrate, examine the trend of alcohol consumption among adolescents as represented in Figure 1. The diagram displays a simple network consisting of four actors, denoted by nodes i, j, k, and l. An individual attribute is assigned to each node, represented by the node's color, which can be either blue or white. The attribute, in this case, signifies the alcohol behavior of the respective individual. The ties between the actors are considered friendship. The upper network illustrates the influence process that occurs between the time points $t-1$ and t. The networks in $t-1$ consist of two dyads (i and j, l and k) connecting two nodes of different colors (blue and white). Taking this situation as a starting point, the influence process occurs if the nodes within the dyad modify their color to match that of the other one. For

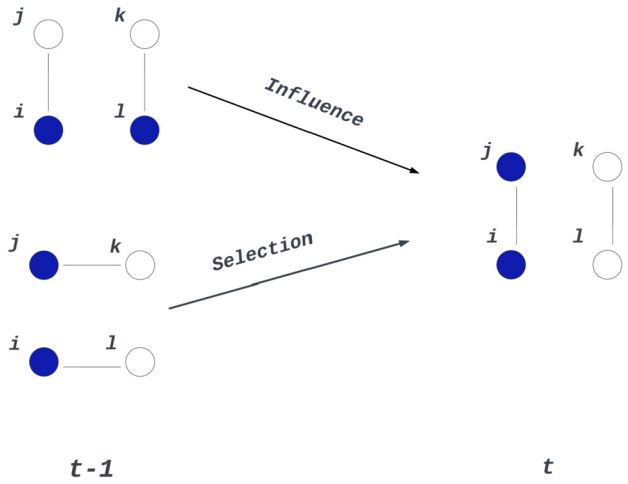


Figure 1: The entanglement between influence and selection (homophily). A network with four actors (nodes i, j, k, and l) is connected by ties, where node color (blue or white) indicates an individual attribute with two states.

example, node j transitions from white to blue, while node k might shift from blue to white. As a result, at time t, a network is formed in which two dyads show similar behaviors, indicating that the unchanged color nodes did have an influence on the other node within the dyad.

The lower network in Figure 1 represents the selection process, which involves establishing connections between nodes with identical characteristics and removing connections between nodes with dissimilar attributes. In this way, individuals who exhibited similar alcohol behaviors at $t-1$ broke their existing friendships and established new social ties with each other at time t. The same behavioral shift was observed in nodes k and l.

It becomes evident that both selection and influence can result in similar network structures when two dyads with similar nodes are present. But the underlying processes are entirely different. While the influence process operates within the existing friendship structure and leads to adaptations in alcohol use behavior, the selection process, in contrast, alters the network structure by forming connections between individuals with similar behaviors. Differentiating between selection (homophily) and influence is crucial in social analysis. For instance, if the Substance Abuse Prevention Program aims to enhance a treatment's effectiveness by considering a patient's previous treatment, the approach varies based on the driving factor. If social influence primarily drives similarities, encouraging the patient to advocate for the treatment among friends would be advantageous. Alternatively, if homophily is the primary influencer, targeting the patient's friends with preventive measures would be beneficial. Consequently, when estimating the peer effect in a social network context, homophily is identified as a major confounding factor that needs to be considered. However,

it is challenging to distinguish between these mechanisms, particularly when peer influence is confounded with latent homophily caused by unobservable features.

Since [22], research on social influence has employed various techniques to account for homophily, exogenous variables, and factors beyond influence in general. Some of these techniques include the co-evolution of selection and influence [34, 35], or by using indirect relationships from third parties as instrumental variables [7]. Studies like [27] have investigated the impact of peer influences on prescription choices while controlling for the role of homophily by integrating individual fixed effects. [4] conducted a study exploring homophily applying propensity score matching algorithms based on observed characteristics.

In their research, [44] and [24] focused on the difficulties of estimating the contagion effect, which arises due to the entanglement of social influence effects and selection and treated the issue as an omitted variable bias, and they used latent community affiliation as a proxy for latent homophily.

Our application of latent space to control for homophily is motivated by their previous work. In this study, we propose a fully model-based approach utilizing a Bayesian framework, which allows us to infer model parameters and enhance predictive accuracy. Specifically, in the context of a social network with n nodes, the behavior of each node $i \in \{1, \ldots, n\}$ characterized by the random variable $Y_i \in \mathcal{R}$, can be effectively described using a linear structural equation model:

$$Y_i = f(Z_{ij}, Y_j, X_i, C_i), \tag{2.1}$$

where Y_i is a function of the behaviors of his/her network partners Y_j; the network relations Z_{ij} between i and j, where person i's observed characteristics is X_i, and C_i is a time-invariant unobserved trait. For example, an adolescent i's alcohol use at time $t \in \{1, \ldots, T\}$, Y_{it}, can be a function of his/her previous alcohol use $Y_{i,t-1}$, his/her close friends' previous alcohol use $Y_{j,t-1}$, his/her own cigarette use $X_{i,t}$ and a time-invariant unobserved tendency for risky behavior C_i.

Along with the influence Equation (2.1) the "true" selection model is represented as:

$$P(Z_{ij} = 1) = g(X_i, X_j, d(C_i, C_j)), \tag{2.2}$$

where the probability that person i and person j have a network relation is a function of the individual and dyadic level observed variables X_i and X_j, and a distance function of the unobserved trait C between i and j and g is an inverse link function.

Whenever latent traits co-determine influence and homophily processes, the linear model assumption may be violated due to endogeneity [44], particularly when an unobserved variable correlates with the error term. This violation will lead to biased and inconsistent estimates of contagion effects. However, if the latent variable is adequately controlled within the linear model structure, the influence can be safely evaluated. Based on this problem, the model proposed estimates the contagion effect while providing explicit control over the latent homophily. This approach enables us to quantify the impact of both social influence and homophily, simultaneously, resulting in a more comprehensive understanding of how these factors affect the network dynamics.

2.2 Model Description

This section explains the main idea behind our analysis. A fully Bayesian model for social influence assumes that there is an unobserved trait that drives both the selection and influence process. We begin by introducing the notations that will be used throughout this study. A normal distribution with mean μ and variance σ^2 is represented as $\mathrm{N}(\mu, \sigma^2)$. The probability density function (*pdf*) and cumulative distribution function (*cdf*) of this normal variate distribution are denoted as $\phi\left(\cdot \mid \mu, \sigma^2\right)$ and $\Phi(\cdot \mid \mu, \sigma^2)$, respectively. In specific instances where $\mu = 0$ and $\sigma^2 = 1$, we simplify the notation further to $\phi(\cdot)$ for the *pdf* and $\Phi(\cdot)$ for the *cdf*. Similarly, we denote the univariate Student's-t distribution with mean μ, scale σ^2 and degrees of freedom ν as $\mathrm{T}(\mu, \sigma^2, \nu)$. $\mathrm{T}_+(\mu, \sigma^2, \nu)$, denotes the half Student's-t (half-t) distribution with positive support. Finally, $Bernoulli(p)$ describes the Bernoulli distribution with success probability p.

To understand how the model estimates social influence, it is essential first to grasp the role of each mechanism and parameter involved. Similarity within the latent space model is utilized to develop a selection process that captures the evolving structure of a social network with n nodes over time, where connections are formed based on behavioral similarity. The social connections within the network are represented using an adjacency matrix (Z) with a size of $n \times n$. A value of 1 in Z_{ij} indicates a connection between individuals i and j, while a value of 0 signifies the absence of a connection. If the connections are not directed, then Z_{ij} is the same as Z_{ji}. Additionally, the model considers a d-dimensional time-invariant latent vector, C_i, which represents the position of a node i in a d-dimensional latent space. This, along with a node's observable behaviors (X_i) and network ties, determines the position of the ith node within the network. Explicitly, the latent selection model used to simulate the development of longitudinal networks with dynamic covariate effects is defined as:

$$\begin{aligned}
Z_{ij,t} &\sim Bernoulli(p_{ij,t}), \;\; i, \; j = 1, \ldots, n, \; t = 1, \ldots, T, \\
p_{ij,t} &= \Phi(\alpha_0 + \alpha_1 |C_i - C_j| + \alpha_2 |X_{i,t-1} - X_{j,t-1}|), \\
C_i &\sim N(0, \tau^2), \\
\alpha_l &\sim N(0, \sigma_l^2), \qquad l = 0, 1, 2,
\end{aligned} \tag{2.3}$$

where the probability of having a tie between actors i and j at time t is a function of the social network connection Z_{ij} at time $t-1$ to the Euclidian distance (it could also be

replaced by other distance functions [15] between i and j's time-invariant latent positions $|C_i - C_j|$ and observed covariate at time $t-1$, (*e.g.*, $|X_{i,t-1} - X_{j,t-1}|$). Essentially, i and j are more likely to have a network relation when they are close to each other in terms of latent position and observed covariates. The degree of network density is affected by the parameter α_0, whereas α_1 controls homophily based on unobserved traits.

Concerning the estimation of social influence, given that an unobserved trait co-determines influence and selection (homophily based on the unobserved trait), the behavior of interest for an individual i at time t, $Y_{i,t}$, is modeled as a liner function of a prior behavior at time $t-1$, the network behavior and to some estimated time-invariant latent variable C_i. Including latent positions as extra covariates in the behavioral model will reduce bias in estimating the effects of social influence [24]. The latent selection model's estimated positions are used as the proxies for the unobserved traits C_i that co-determine influence and selection. For a single node $i \in \{1, \ldots, n\}$, the generative model for this process becomes:

$$Y_{i,t} = \beta_0 + \beta_1 Y_{i,t-1} + \beta_2 \frac{\sum_{j=1}^{n} Z_{ij,t-1} Y_{j,t-1}}{\sum_{j=1}^{n} Z_{ij,t-1}} + \beta_3 X_{i,t} + C_i, \tag{2.4}$$

where $Y_{i,t-1}$ refers to the previous behavior of ith individual, also referred to as the lagged variable, network exposure term $\frac{\sum_{j=1}^{n} Z_{ij,t-1} Y_{j,t-1}}{\sum_{j=1}^{n} Z_{ij,t-1}} = Contag_{i,t}$ represents the weighted average behavior among the network neighbors of i, which is the contagion effect of interest, and $X_{i,t-1}$ represents the set of all other attributes for node i which affect the behavior $Y_{i,t}$.

This paper proposes a model to estimate the contagion effect as a combination of a longitudinal latent selection model (which imposes a relationship between the observed network and latent space positions) and a dynamic linear model, which establishes a relationship between the outcomes at node level and the behavior of the node. The complete Bayesian formulation with proper (but diffuse) prior distributions for the influence model follows:

$$\begin{aligned}
Z_{ij,t} &\sim Bernoulli(p_{ij,t}), \quad i,j = 1,\ldots,n,\ t = 1,\ldots,T,\\
p_{ij,t} &= \Phi(\alpha_0 + \alpha_1 |C_i - C_j| + \alpha_2 |X_{i,t-1} - X_{j,t-1}|),\\
Y_{i,t}|C_i &\sim N(\beta_0 + \beta_1 Y_{i,t-1} + \beta_2 Contag_{i,t} + \beta_3 X_{it} + C_i, \sigma^2),\\
C_i &\sim N(0, \tau^2),\\
\beta_l &\sim N(0, \sigma_\beta^2), \quad l = 0,1,2,3,\\
\alpha_k &\sim N(0, \sigma_\alpha^2), \quad k = 0,1,2,\\
\tau^2 &\sim T_+(0, s_\tau^2, \nu_\tau),\\
\sigma_\beta^2 &\sim T_+(0, s_\beta^2, \nu_\beta),\\
\sigma_\alpha^2 &\sim T_+(0, s_\alpha^2, \nu_\alpha).
\end{aligned} \tag{2.5}$$

Hamiltonian Monte Carlo [28] approach was implemented where the influence parameters and homophily are computed simultaneously in order to fit this model. It is important to note that the influence model includes the estimated C_i from the selection part to control for latent homophily. The "`rstan`" package, which is an R interface to the `Stan C++` package, provides full Bayesian inference using the No-U-Turn sampler [NUTS, 8], a variant of Hamiltonian Monte Carlo (HMC) and allows users to fit `Stan` models from `R` and access the output, including posterior inferences and other quantities. Hence, we perform a Bayesian analysis using the `rstan` package.

2.3 Simulation Studies to Further Explore the Influence Model

The estimation performance is a complicated task driven by an extensive variety of factors. The focus of our work is to thoroughly examine the effectiveness of estimators in various situations that include several significant factors, such as the following.

1. The sample size denotes the total number of nodes within a network. It is also referred to as the network size.
2. The magnitude of bias in estimation can vary based on the magnitude of the network exposure.
3. The total number of time points is essential in longitudinal data, where increased time points correlate with a higher volume of information, thus enhancing the accuracy in estimating parameters.
4. The level of homophily determines the extent to which the relationship between the unobserved characteristic and network exposure is correlated. This, in turn, affects the correlation between the lag-dependent variable and network exposure. A greater degree of homophily implies a stronger correlation, and studying such correlations can help understand their impact on the magnitude of bias.

We will use 95% probability credible intervals (CI) with equal tails to measure the extent of parameter coverage to evaluate the performance of the estimators. The posterior means will then be thoroughly examined to identify potential biases, and a report will be produced summarizing the results. The principal goal of this study is to identify the specific conditions under which this model is the most advantageous, allowing for a more sophisticated and customized strategy for its subsequent uses.

The data generation process for the research methodology was established using Equations (2.3) and (2.4). It started by constructing a social network of size n, where each individual i has dynamic observed behaviors X_i, alongside time-independent unobserved characteristics $C_i \in \mathcal{R}^d$, where d was restricted to 1. It was assumed that the elements within X_i or C_i followed an independent and identically distributed (*i.i.d.*) pattern.

The formation of edges within the network followed the principle of homophily, wherein individuals sharing similar characteristics or behaviors had a higher likelihood of forging connections than those with dissimilar attributes. We first simulate latent space positions C for n nodes from a normal distribution $N(0, 2^2)$, allowing for some flexibility in the distance values. Once the latent space positions are generated, Equation (2.3) is used to quantify the probability that two nodes form an edge based on their similarity. This similarity was determined by the observed prior friendship status $Z_{ij,t-1}$ and the Euclidean differences between individual i and j in unobserved feature spaces.

Both peer influence and homophily influenced the evolution of the outcome variable. Consequently, the development of the dependent variable was approached as an iterative process. Initially, the dependent variable $Y_{i,t}$ was initialized for each node (consider the starting point $t = 1$), ensuring independent and identically distributed values. The evolution of the social network was affected by both individual influence and homophily. Given the common observation of time-dependent correlation in an individual's behavior within social networks, we conducted simulations that integrated a lagged dependent variable (prior behavior). Equation (2.4) was used to calculate the outcome variable at time t, considering the variables at time $t-1$. To maintain simplicity, we excluded additional observed variables from the model. The simulations were generally conducted with a time of $T = 3$ in most scenarios, partly excluding Subsection 2.3.4.

The selection of hyperparameters ensures the propriety of all posteriors while maintaining a relatively weakly informative nature. A vague half-t (T_+) distribution is used when determining the prior scale as recommended by [13]. The estimation model consists of:

$$\begin{aligned}
Z_{ij,t} &\sim Bernoulli(p_{ij,t}),\\
p_{ij,t} &= \Phi(\alpha_0 + \alpha_1 Z_{ijt-1} + \alpha_2 |C_i - C_j|),\\
Y_{i,t}|C_i &\sim N(\beta_1 Y_{i,t-1} + \beta_2 \frac{\sum_{j=1}^{n} Z_{ij,t-1} Y_{j,t-1}}{\sum_{j=1}^{n} Z_{ij,t-1}} + C_i, \sigma^2),\\
C_i &\sim N(0, \tau^2),\\
\beta_i &\sim N(0, 10^2),\\
\alpha_i &\sim N(0, 10^2),\\
\tau^2 &\sim T_+(0, 5, 4),\\
\sigma^2 &\sim T_+(0, 5, 4),
\end{aligned} \tag{2.6}$$

where the hyperparameters for β_l, α_k, τ^2 and σ^2 are set at $\sigma_\beta^2 = \sigma_\alpha^2 = 10^2, \forall l, k$; $s_\alpha^2 = s_\beta^2 = s_\tau^2 = 5^2$ and $\nu_\alpha = \nu_\beta = \nu_\tau = 4$. The starting density of the network is set at 0.3. Within each of the specified simulation research conditions, 100 iterations were conducted to generate replicated data sets. Subsequently, the model was applied to each data set, employing three HMC chains with 40,000 samples per chain. The initial 1,000 draws were excluded, and the subsequent draws were thin by a factor of 70.

2.3.1 Parameter Recovery and the Degree of Bias when an Unobserved Trait Fails to Be Considered

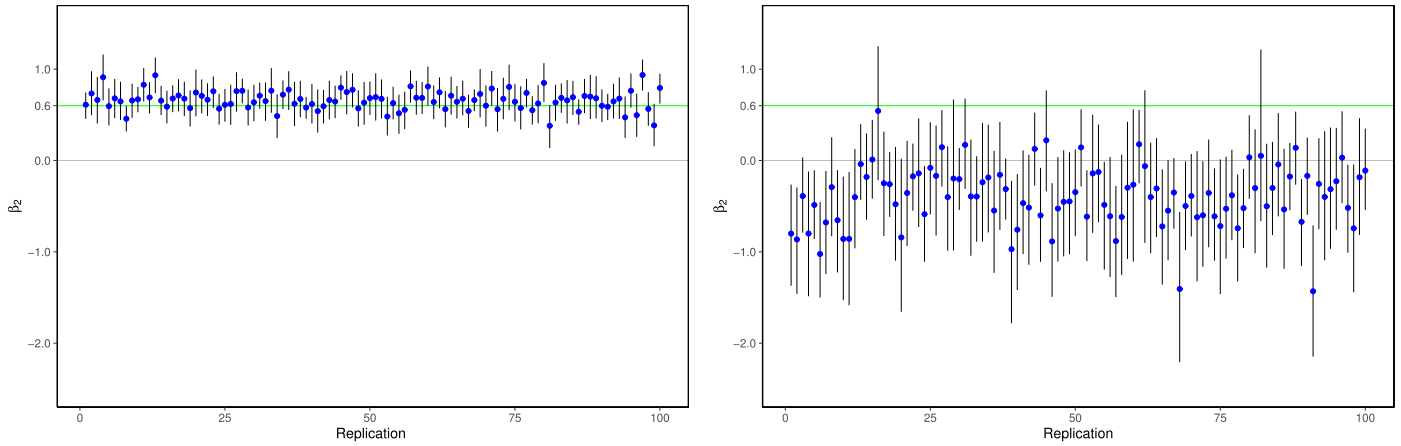
This study aims to provide intuitive insights using simulation examples to address the potential biases, particularly when a latent trait influences both selection and influence. The initial focus is to specifically examine the importance of considering latent variables in the influence model. This involves addressing potential biases in estimating contagion effects when disregarding the unobserved trait that simultaneously drives both selection and influence.

In this scenario, the contagion effect was estimated using two different approaches with the proposed model. First, was considered the latent trait C in the influence model, as shown in Equation (2.6). Secondly, the unobserved trait was disregarded. The data generation process remains consistent with the explanation provided in Section 2.3. Specifically, the simulation configuration involves a study using $n = 40$ nodes and a sequence of $T = 3$ time points. The parameters for the selection model have been set as follows: $\alpha_0 = -0.1$, $\alpha_1 = -0.45$, and $\alpha_2 = -0.5$. Additionally, in the influence model, the values for β_1 and β_2 are maintained at 0.3 and 0.6, respectively.The findings were obtained by comparing the contagion effect parameter estimates between a correctly specified model and a misspecified model. Results are presented in Figure 2, which provides a graphical representation of posterior means and 95% confidence intervals for the β_2 parameter. As anticipated, the recovery of parameters diminishes when the latent feature is not adequately incorporated into the model. The preceding results indicate that an unaccounted, unobserved feature, which either influences selection and influence or solely affects influence, leads to biased estimations in the influence model for the contagion effects. Consequently, this could potentially yield incorrect inferences. Although there exist other elements that may impact estimator performance, our attention is directed toward the accuracy of the social influence model in the subsequent scenarios.

2.3.2 Varying Network Size

The initial aspect under evaluation is the dimension of the network. To explore this was performed a simulation incorporating network and influence data sourced from Equation (2.6), following the same data generation process as previously described. The network's dimensions were systematically altered, spanning from 40 to 80 and up to 160. All other parameters were maintained at their initial values for a true value of peer effect 0.6. Figure 3 illustrates the posterior mean and the 95% equal-tailed credible interval for β_2 across 100 replications.

This graphical representation illustrates the reciprocal correlation between network size and posterior variance,



(a) Controlling the unobserved trait

(b) Ignoring the unobserved trait

Figure 2: Posterior means and 95% credible intervals for β_2 in two scenarios: considering the latent trait in the influence model (left panel) and excluding it (right panel).

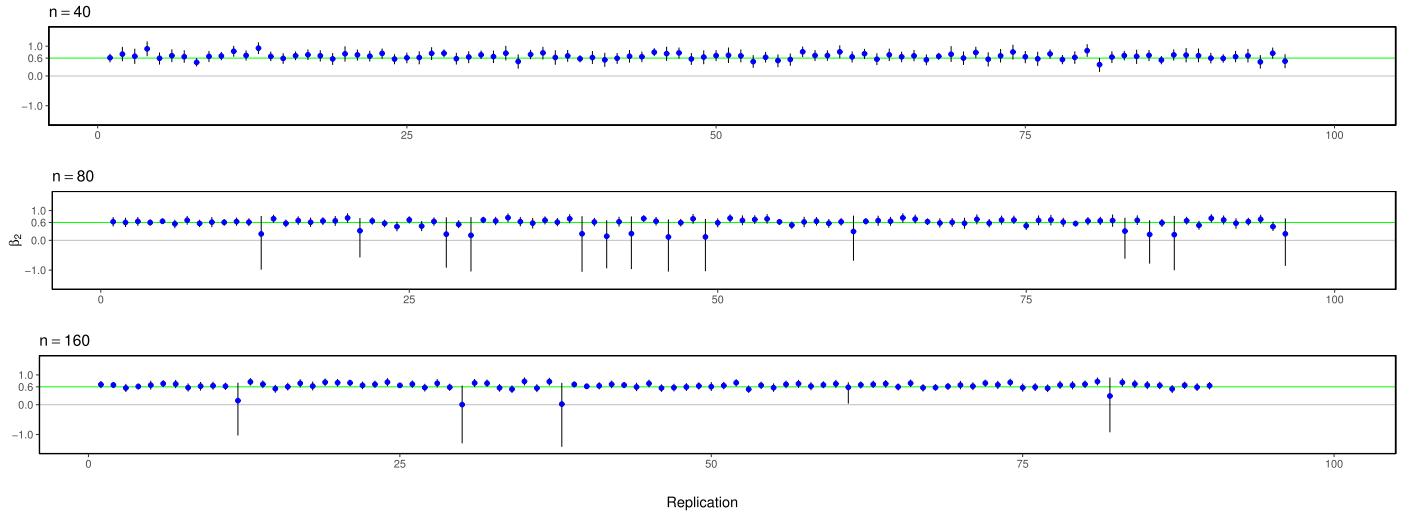


Figure 3: The impact of network size on the estimation of contagion effect was examined. Data from multiple networks of different sizes were generated and used to calculate posterior means and 95% credible intervals.

where an expansion in network size decreases posterior variance. As the network size expands, there is minimal variation in the capacity to estimate the posterior mean accurately. The bias of the posterior mean is approximately zero across all network sizes, indicating minimal bias and low impact of network size on accurately estimating the true amount of contagion effect. Specifically, the average bias for the β_2 was found to be $0.06, -0.03$, and 0.02 for networks consisting of 40, 80, and 160 units, respectively The accuracy of parameter recovery, signifying the proportion of replications encompassing the true value of β_2 within the 95% equal-tailed credible interval, diminished as the network size enlarged, dropping to 90% for $n = 160$ from over 95% for $n = 40$ and 80. This decline was anticipated, as the increasing network size makes precise estimation of the posterior mean more challenging due to the growing number of parameters to be estimated. Furthermore, the observations revealed that an increased network size diminishes the bias in estimating the β_1; specifically, the mean bias for the lag effect was determined as $-0.02, 0.04$, and 0.002 for network sizes of 40, 80, and 160, respectively.

2.3.3 Varying Number of Time Points

This simulation section explores the impact of time points within longitudinal data analysis. Data were generated at two distinct values of T, namely $T = 3$ and $T = 6$, specifically intended to facilitate the identification of variations across different conditions. This disparity was ob-

Table 1. Contagion Effect Simulation Results. CP indicates the probability of coverage.

		β_2 - Contagion Effect	
		Bias	*CP*
$n = 40$, $T = 3$	$\beta_2 = 0.2$	0.039	0.90
	$\beta_2 = 0.6$	0.06	0.90
	$\beta_2 = 0.9$	0.03	0.90
$n = 40$, $T = 6$	$\beta_2 = 0.2$	0.006	0.90
	$\beta_2 = 0.6$	0.005	0.97
	$\beta_2 = 0.9$	0.004	0.94
$n = 160$, $T = 3$	$\beta_2 = 0.2$	-0.012	0.92
	$\beta_2 = 0.6$	0.027	0.87
	$\beta_2 = 0.9$	0.025	0.86
$n = 160$, $T = 6$	$\beta_2 = 0.2$	0.003	0.94
	$\beta_2 = 0.6$	0.018	0.92
	$\beta_2 = 0.9$	0.001	0.95

served across multiple scenarios encompassing diverse network sample sizes (n) and various settings of the contagion effect parameter β_2, as illustrated in Table 1.

The relationship between the capacity to estimate the influence effect and the time points reveals an anticipated positive association, where an increase in the time points consistently correlates with an enhanced ability to forecast the influence impact. This correlation persists consistently despite fluctuations in parameter magnitudes and network sizes. A higher number of time points in a longitudinal analysis provides more information for prediction, resulting in reduced bias and an elevated parameter recovery rate.

2.3.4 Varying the Social Influence's Effect

Finally, our exploration focuses on examining the impact of the range or margin of values associated with β_2 on the parameter recovery process and the characteristics exhibited within the posterior distribution of β_2. Furthermore, the aim is to extend this analysis to understand better the nuanced implications and dependencies arising from the varying margins or ranges of β_2 values on both the parameter recovery process and the resulting posterior distribution of β_2.

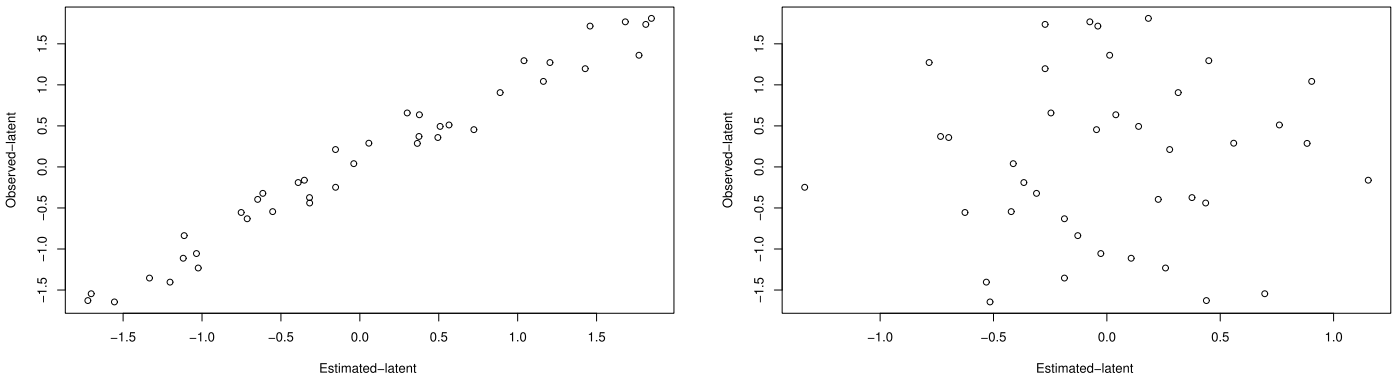
To conduct this study, 100 iterations were executed, each generating data for varying β_2 values (specifically, $\beta_2 = (0.2, 0.6, 0.9)$) in conjunction with different β_1 values (specifically, $\beta_1 = (0.7, 0.3, 0.1)$) The remaining parameters were held constant, following the specifications outlined in Section 2.3. This process was replicated in four unique scenarios, each characterized by differing network sizes, specifically a relatively small network of $n = 40$ and a relatively larger network of $n = 160$. Furthermore, the analysis considered two distinct time-point configurations denoted as $T = 3$ and $T = 6$. Table 1 presents the recovered parameter values, the proportion of occurrences indicating a positive influential effect, and the average bias of the posterior mean across replications. The bias of the posterior mean tends to zero across all conditions, this indicates minimal bias and a limited impact on the value of β_2 on the bias. Notably, the parameter recovery rates exceed 90%. However, as the size of the network expands to 160, this rate experiences a reduction. In contrast, this rate increases for all conditions with an increase in the number of time points (T). As anticipated, there is a parallel increase in the proportion of instances inferring a positive network influence with the increase of the β_2 value. Table 1 shows the probability of coverage (CP), the mean bias (between replications) of the posterior mean, and the percentage of scenarios in which it would be concluded that there was a positive social impact effect.

2.3.5 Varying Homophily Level

In cases where homophily functions through the unobserved trait, increased homophily levels signify a stronger correlation between network exposure and the unobserved trait and an elevated correlation between network exposure and the lagged dependent variable. In this section, our focus is on examining the impact of homophily on bias magnitude and beyond. Specifically, the following simulation configurations were considered: (1) The number of nodes is fixed at 80, and the number of time points at 3. (2) Homophily levels are varied under three conditions: (i) low homophily ($\alpha_0 = -0.3, \alpha_1 = -0.65, \alpha_2 = -0.1$), ($ii$) moderate homophily ($\alpha_0 = -0.1, \alpha_1 = -0.45, \alpha_2 = -0.5$), ($iii$) high homophily ($\alpha_0 = -0.1, \alpha_1 = -0.55, \alpha_2 = -0.8$). Note that α_2 controls the level of homophily based on the unobserved attribute, while α_1 reflects homophily based on previous friendships observed and α_0 affects the density of the overall network. The summation of the α's remains constant. (3) Fixed values for $\beta_1 = 0.3$ and $\beta_2 = 0.6$. Further model configurations were maintained as detailed in Section 2.3. The Mean biases for the network exposure term are shown in Table 2. Various considerations arise regarding the influence effect variable term. First, the study shows that an increase in the homophily controlled by latent traits leads to a higher magnitude of bias. This is not surprising, as latent homophily affects the selection process and influences the determination of the contagion effect. As a result, more information is missing, which requires estimation and leads to more significant errors in estimating the contagion effect. Secondly, an increase in homophily leads to a decrease

Table 2. Homophily level. Higher homophily in simulated data indicates a stronger correlation between network exposure and the unobserved trait and between network exposure and the lagged dependent variable.

$n = 80$, $T = 3$			
	Low	Mid	High
Mean bias (β_2)	0.026	-0.030	0.036
CP (β_2)	0.99	0.95	0.93



(a) Bayesian Estimation C_i

(b) Two-Step Estimation C_i

Figure 4: Simulation study. (a) Latent variable recovered from the Bayesian model, and (b) two-step frequentist estimation of latent trait.

in coverage probability. In other words, the model's ability to recover network influence effects decreases when an unspecified parameter governs the process. However, even in situations with high levels of unobserved factors, the model performs well, with a bias close to zero. Finally, we found that the model accurately estimates the lagged term, represented by β_1, with mean biases of -0.003, 0.04, and -0.03 for scenarios of low, moderate, and high levels of homophily, respectively.

2.3.6 A Comparison of Two-Step Frequentist vs. Bayesian of Contagion Estimation

In a final simulation study, the proposed Bayesian model was compared with a frequentist framework in the context of estimating social effects. As previously mentioned, earlier research utilizing latent space models to control for homophily employed a frequentist two-step estimation approach [44], which may have implications for the accuracy of standard errors. This investigation encompasses two main aspects. First, the data were generated using the influence model, as described in Section 2.2, taking into account the influence parameters $\beta_1 = 0.5$ and $\beta_2 = 0.8$, as well as the selection model, which is only considering the intercept $\alpha_0 = -0.2$ and the latent trait $\alpha_1 = -0.25$. Subsequently, the data is fitted using the model proposed model 2.6 and the two-step approach using the frequentist version of the influence model. Specifically, start using the "`ergmm`" function [8, 19] to estimate the latent space models. Once the latent space models have been estimated, the latent positions will be extracted and used as additional variables to estimate the behavioral/influence model. To align with Bayesian models, time-invariant latent positions are estimated at a single time point. The latent variable recovery from both models is seen in Figure 4. The linear relationship between estimated C_i and true C_i values from the Bayesian model suggests a successful recovery of the latent trait. This relationship was not observed when using a two-step frequentist approach, indicating that the Bayesian influence model proposed may be more reliable for estimating latent variables. Moreover, an analysis was performed to assess the bias in the calculation of the contagion effect and the coverage probability. The results of our study indicate that the average bias and CP of the peer estimate were determined to be -0.007 and 94% for the Bayes approach and -0.05 and 85% for the frequentist approach, respectively. The disparity noticed pertains to the larger size of the mean bias when the parameter was derived using a two-step frequentist model.

Table 3. Bayesian vs two-step Frequentist estimation in a simulation study. CP is the Coverage Probability.

$n = 80, T = 3$		
	Bayes setup	Frequentist setup
	β_2	β_2
Mean bias	-0.007	-0.05
CP	0.94	0.85

3. AN APPLICATION TO ALCOHOL-FRIENDSHIP NETWORK

Having established an approach that reduces bias in social influence studies, the next important question to explore is whether peers influence alcohol use within a social network.

The influence model integrates latent space to determine the impact of peer behavior on alcohol consumption patterns among teenagers. One significant finding is the prevalence of alcohol-based selection among adolescents who consume alcohol. This phenomenon suggests that they tend to form friendships with peers who exhibit similar alcohol-related behaviors.

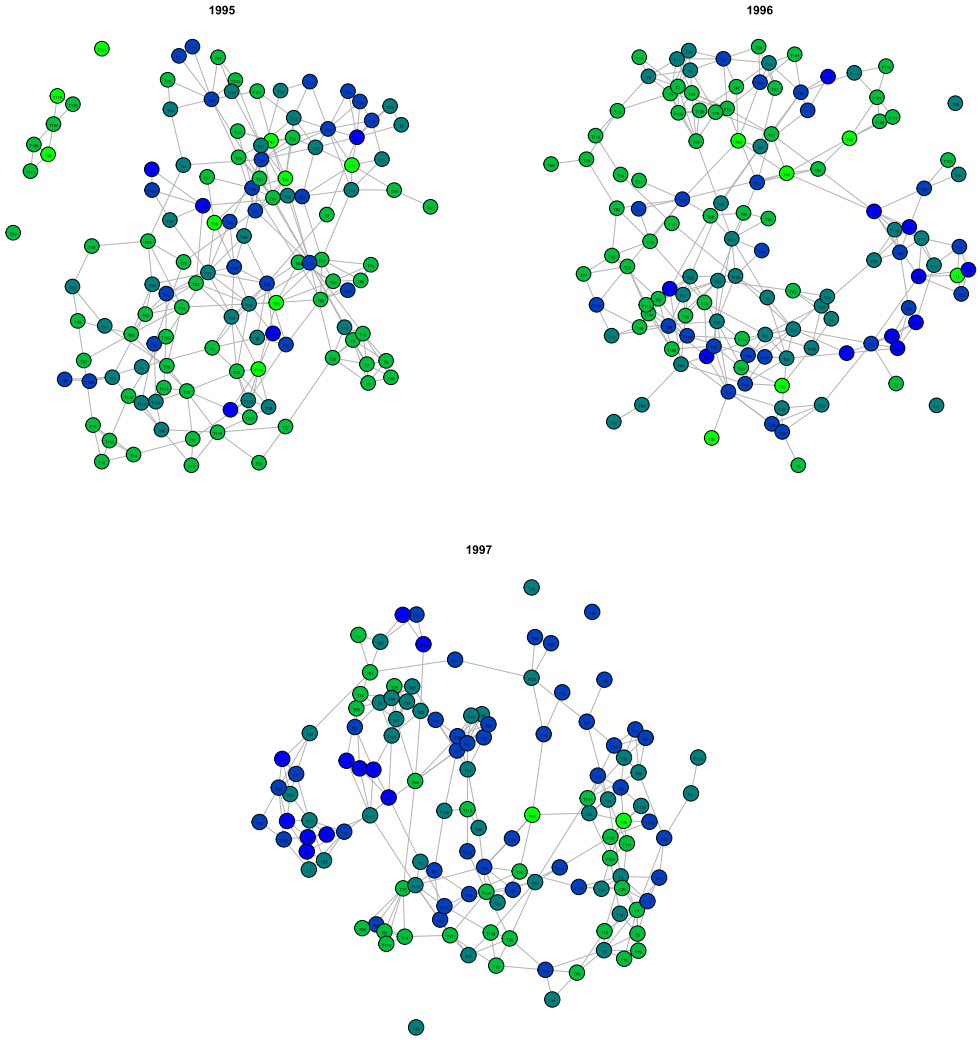


Figure 5: Alcohol-friendship network data. Temporal depiction of alcohol-related clusters within the adolescent social network from 1995 to 1997. Nodes signify individuals, while the connections represent relationships. Node color reflects alcohol use, with darker blue shades signifying higher alcohol consumption and lighter green shades indicating lower consumption (various shades of green represent intermediate levels).

3.1 Dataset

The analysis utilizes longitudinal data gathered from friendship groups within the 'Teenage Friends and Lifestyle Study'. The study involved a cohort of 160 female adolescents, averaging 13 years of age, investigating several dimensions of their lifestyle, family dynamics, smoking habits, participation in sports, alcohol consumption, and engagement in drug-related behaviors. Out of the initial 160 participants, we excluded individuals who did not complete the questionnaire. Consequently, the final sample size for the study was reduced to 123 participants. The data collection spanned a three-year period, specifically from 1995 to 1997. Questionnaires were distributed within the educational environment to all individuals who were involved in the study and fell within the age range of 12 to 13 years. Subsequent assessments were carried out at yearly intervals during the periods when the students were aged 13-14 and 14-15 years.

The measurement involved consists of *Friendship ties,* to construct friendship-seeking, participants were asked to identify up to twelve individuals they considered their best friends. They were then instructed to indicate whether a friendship existed between the ith and jth teenager by assigning a value of 1 or if no friendship existed by assigning a value of 0. Figure 5 illustrates the alterations in the friendship network and alcohol behavior of the girls over three years. A node represents each individual. The presence of lines connecting nodes serves to denote a relationship between them. The color of nodes represents the level of alcohol consumption, where darker shades of blue indicate a higher frequency of alcohol use and lighter shades of green indicate lower consumption. Intermediate shades of green

Table 4. Alcohol-friendship network data. Posterior means(Post.means) of the coefficients within the influence, along with the corresponding posterior standard deviation(Post. SD), 95% credible intervals, $\hat{R}$ and Neff for various influence covariates.

Parameter	Post.mean	Post. SD	2.5%	97.5%	$\hat{R}$	Neff
$\beta_0 - intercept$	0.87	0.21	0.47	1.28	1	1214
$\beta_1 - prior\ behaviour$	0.52	0.06	0.41	0.63	1	2068
$\beta_2 - contagion\ effect$	0.13	0.06	0.01	0.25	1	1607
$\beta_3 - canabis\ use$	0.29	0.07	0.16	0.42	1	1913
$\beta_4 - smoking$	0.13	0.08	-0.03	0.28	1	1851
$\beta_5 - sport$	-0.16	0.05	-0.26	-0.07	1	1571

are used to represent levels of consumption between these extremes. The observation can be made that there is a considerable proportion of adolescents engaging in alcohol consumption on a weekly basis, while all students partake in alcohol consumption at least once annually.

Adolescent alcohol behavior was evaluated through a singular inquiry that requested adolescents to self-report their alcohol consumption(1 - non-alcohol user, 2- drink alcohol twice a year, 3- once a month, 4- once a week, 5- drink alcohol more than once a week).

Several variables recognized to have associations with adolescent alcohol consumption were included as covariates:

sporting behavior (1- nonregular sports activity, 2 regular sport activity), *tobacco use* (1 -no smoker, 2- occasional, 3- regular), *cannabis consumption* (1 - nonuser, 2 -tried once, 3 -occasional, 4 -regular).

The covariate that reflects the average behaviors among the neighbors within the network was established following experts.

3.2 Model Specification

We utilized a one-dimensional latent space to fit the Bayes model on longitudinal data concerning friendship-seeking networks. To address the directed nature of relationships within these networks, we integrated time-invariant latent traits that jointly determine friendship selection and the variable of interest, namely alcohol behavior. Additionally, a covariate influencing these dynamics was incorporated. The resulting model can be expressed as follows:

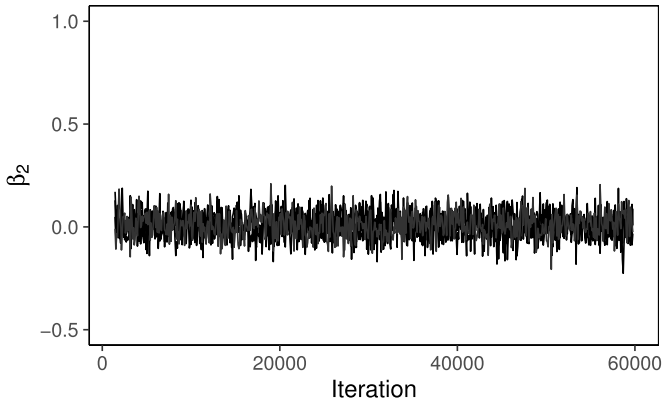
$$\begin{aligned} Z_{ij,t} &\sim Bernoulli(p_{ij,t}) \quad i,j = 1,\dots,123,\ t = 1,2,3, \\ p_{ij,t} &= \Phi(\alpha_0 + \alpha_1|C_i - C_j| + \alpha_2|X_{i,t-1} - X_{j,t-1}|), \\ Y_{i,t}|C_i &\sim N(\beta_0 + \beta_1 Y_{i,t-1} + \beta_2 Contag_{i,t} + \beta_3 X_{i,t} + C_i, \sigma^2), \\ C_i &\sim N(0,\tau^2) \end{aligned} \tag{3.1}$$

and the priors for β_k, $k = 0,1,2,3$, α_l, $l = 0,1,2,3$, as well as for τ^2 and σ^2, are specified as in Equation (2.6), where $Y_{i,t}$ signifies the alcohol behavior of the ith teenager at time t, where i ranges from 1 to 123 and t spans from 1995 to 1997. The parameter $Contag_{i,t} = \frac{\sum_{j=1}^{n} Z_{ij,t-1} Y_{j,t-1}}{\sum_{j=1}^{n} Z_{ij,t-1}}$ denotes the average lagged behavior in alcohol among ith adolescents. The vector $X_{i,t}$ represents the observed characteristics of adolescents, which might exert an influence on behavioral outcomes, particularly pertaining to smoking, drug use, and participation in sporting activities. Additionally, C_i represents the latent variable utilized to manage latent homophily effects. The observed ties proportion ranges from 0.07 to 0.15, suggesting relatively sparse networks, warranting consideration of the approximate probit model as a viable option for selection. While relatively flat priors are set for the coefficients (β_l and α_k), more constrained priors are applied for the intercept of the influence β_0, T. The dynamic influence model is presented accordingly. The models were fitted using three Hamilton Monte Carlo (HMC) chains, each comprising 60,000 iterations. A burn-in period of 1000 iterations was implemented, and the samples were thinned by a factor of 90, resulting in a total of 594 posterior samples. Provided below are summaries of the posterior distribution. It's crucial to note that the introduction of covariance introduces noise, whereas the color remains constant over time.

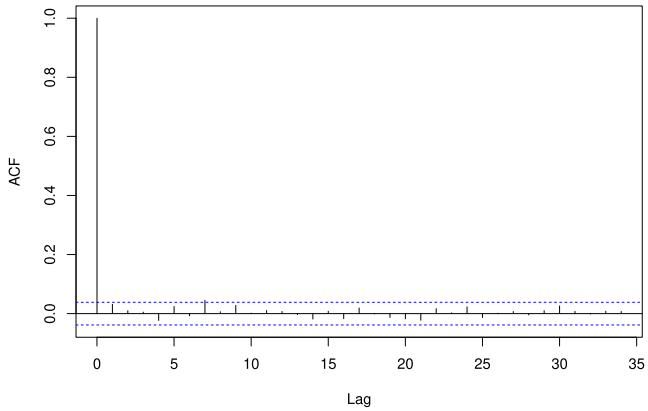
3.3 Results

The Hamiltonian Monte Carlo (HMC) method has successfully achieved convergence, as shown by a potential scale reduction factor ($\hat{R}$) of 1 in Table 4. Additionally, the effective number of samples (Neff) is over 100, further indicating a well-converged HMC process. The autocorrelation analysis in the left panel of Figure 6 shows that the autocorrelation values approach zero after a lag of around four for the samples obtained from `Stan`. This observation suggests that the sampler effectively explored the posterior distribution. The posterior samples obtained by the HMC at each iteration show a clear central tendency, accompanied by evenly dispersed deviations. This pattern indicates minimum inter-sample correlation and an effective exploration of the posterior distribution by the sampling algorithm. Additionally, it is important that the chains do not exhibit any vertical displacement, but rather maintain a consistent value.

After confirming the convergence of the sampling process towards the posterior, our focus shifts to analyzing the parameter estimates to ascertain the model's fit with the observed data. In Table 4, the posterior means of the coefficients are displayed, accompanied by their respective 95%



(a) MCMC chain of β_2



(b) Autocorrelation function of the β_2

Figure 6: Alcohol-friendship network data. Posterior samples of the contagion parameter within the HMC chain and its corresponding ACF plot.

equal-tailed credible intervals. These coefficients represent the change in behavior observed for a one-unit change in the covariate value at that specific time (t) within the model. In the realm of adolescent friendships, there is empirical evidence suggesting the impact of the effect of the network on alcohol behavior during this phase. The statistical analysis reveals a significant contagion effect, which is also evident from Table 4. In addition, a positive association is observed between current alcohol behavior and previous alcohol consumption. Past experiences with alcohol might influence an individual's future behaviors, extending to their choices regarding various aspects. Moreover, existing models suggest a significant correlation between cannabis consumption and alcohol use. Additionally, healthy habits such as participating in sports can prevent teenagers from using alcohol.

The selection model posits that the establishment of social connections is significantly influenced by prior friendships. It follows that individuals are more likely to form connections with people they have previously interacted with, as there is an established level of familiarity and trust. However, the selection model also highlights the importance of unobserved variables in driving friendship selection. These variables may include factors such as shared interests, values, and personality traits, economic status as well as other factors that are not considered in the study.

4. CONCLUSIONS

The presented model provides a framework for the quantitative analysis and prediction of how socially contagious phenomena influence social behaviors. By employing longitudinal measurements of behavior outcomes and social network data, the proposed model enables the examination of the dynamics of a social behavior trend, specifically in relation to the behavior of friends as well as other relevant characteristics. Relationships and interactions between a group of people are represented by social network data. Including network data in models of individual outcomes enables researchers to consider the impact of social influence, as changes in behavior are seldom independent events. By modeling social influence, researchers can determine the degree to which social networks influence particular outcomes, in addition to understanding which outcomes are potentially influenced. The current approach aims to effectively integrate selection processes with impact processes by utilizing a latent variable notion to assess the degree of influence between nodes. This approach is based on the assumption that both the selection of peers and the influence of those social relationships are equally significant factors in shaping particular behaviors.

In order to get a deeper comprehension of the operational attributes of our proposed model, a number of simulation experiments were undertaken to investigate the viability and effectiveness of our model. The parameter coverage was seen to be at or close to 100% in models that were appropriately stated. The capacity to identify network influence was shown to be increased when there was an increase in either the size of the network or the number of time points, as well as when the genuine effect exhibited an increase. Our analysis revealed that the failure to account for latent variables in our model resulted in model misspecification. Furthermore, our models that account for the simultaneous estimation of latent traits and peer effects appear to be more accurate in reducing the implications for standard errors that may arise when using two-step frequentist estimation.

In summary, we conducted an empirical investigation wherein we employed the aforementioned methodology to assess the influence of social factors on the formation of alcohol-related behaviors in adolescents. There is evidence that the 13-year-old teenager was significantly affected by the use of alcohol in friendship drinking behavior. As pointed

out by an anonymous referee, the proposed model bears resemblance to the well-known linear-in-means framework commonly used in causal inference. While there is indeed a structural similarity in terms of peer influence, our focus differs substantially. In contrast to works such as [22], [24] and [40], which emphasize the identification of causal effects, our approach is primarily concerned with estimating contagion dynamics without aiming to establish causal claims.

Future research objectives are oriented towards a more in-depth exploration of the mechanisms governing influence processes and the investigation of behaviors within social selection. In the context of the selection model, symmetric link functions exhibit notable inadequacy, particularly in the presence of imbalanced data. This asymmetry has the potential to introduce substantial bias into the estimation of the mean response. We contend that the integration of skewed link functions, in conjunction with an examination of network density and latent space dimensions, will exert a significant impact on the estimation of social influence effects. Finally, we acknowledge that further theoretical investigation of the proposed approach is important and constitutes a promising direction for future research, although it lies beyond the scope of the present study.

ACKNOWLEDGEMENTS

The authors sincerely thank the guest editors and anonymous reviewer for their valuable feedback and constructive criticism, which have significantly enhanced the presentation and quality of this paper.

FUNDING

Victor Lachos acknowledges partial financial support from the Research Excellence Program (REP) at UConn-OVPR.

Accepted 26 September 2025

REFERENCES

[1] An, W. (2015). Instrumental variables estimates of peer effects in social networks. *Social Science Research* **50** 382–394.
[2] Anagnostopoulos, A., Kumar, R. and Mahdian, M. (2008). *Influence and Correlation in Social Networks*. Association for Computing Machinery, New York, NY, USA.
[3] Aral, S. and Walker, D. (2011). Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks. *Management Science* **57**(9) 1623–1639.
[4] Aral, S., Muchnik, L. and Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences* **106**(51) 21544–21549.
[5] Asch, S. E. (1955). Opinions and Social Pressure. *Scientific American* **193**(5) 31–35.
[6] Bell, D. and Song, S. (2007). Neighborhood effects and trial on the internet: Evidence from online grocery retailing. *Quantitative Marketing and Economics (QME)* **5**(4) 361–400.
[7] Bramoullé, Y., Djebbari, H. and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics* **150**(1) 41–55. https://doi.org/10.1016/j.jeconom.2008.12.021. MR2525993
[8] Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. and Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software* **76**(1) 1–32.
[9] Christakis, N. A. and Fowler, J. H. (2007). The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine* **357**(4) 370–379.
[10] Christakis, N. A. and Fowler, J. H. (2013). Social contagion theory: examining dynamic social networks and human behavior. *Statistics in Medicine* **32**(4) 556–577. https://doi.org/10.1002/sim.5408. MR3042499
[11] Cialdini, R. B. and Goldstein, N. J. (2004). Social influence: compliance and conformity. *Annual Review of Psychology* **55**(1) 591–621.
[12] Friedkin, N. E. and Johnsen, E. (1990). Social influence and opinions. *Journal of Mathematical Sociology* 193–206.
[13] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis* **1**(3) 515–534. https://doi.org/10.1214/06-BA117A. MR2221284
[14] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013). *Bayesian Data Analysis. Bayesian Data Analysis* 675. MR3235677
[15] Hoff, P. D. (2009). Multiplicative Latent Factor Models for Description and Prediction of Social Networks. *Comput. Math. Organ. Theory* **15**(4) 261–272.
[16] Hoff, P. D., Raftery, A. E. and Handcock, M. S. (2002). Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association* **97**(460) 1090–1098. https://doi.org/10.1198/016214502388618906. MR1951262
[17] Jackson, M. O. (2008) *Social and Economic Networks*. Princeton University Press. MR2435744
[18] Krivitsky, P. N. and Handcock, M. S. (2008). Fitting Latent Cluster Models for Networks with latentnet. *Journal of Statistical Software* **24**(5) 1–23.
[19] Krivitsky, P. N., Handcock, M. S., Raftery, A. E. and Hoff, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks* **31**(3) 204–213.
[20] La Fond, T. and Neville, J. (2010). *Randomization Tests for Distinguishing Social Influence and Homophily Effects* 601–610.
[21] Luke, D. A. and Harris, J. K. (2007). Network analysis in public health: history, methods, and applications. *Annual Review of Public Health* **28** 69–93.
[22] Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies* **60**(3) 531–542. https://doi.org/10.2307/2298123. MR1236836
[23] May, R. M. (2006). Network structure and the biology of populations. *Trends in Ecology & Evolution* **21**(7) 394–399.
[24] McFowland, E. and Shalizi, C. R. (2021). Estimating Causal Peer Influence in Homophilous Social Networks by Inferring Latent Locations. *Journal of the American Statistical Association* **118**(541) 707–718. https://doi.org/10.1080/01621459.2021.1953506. MR4571152
[25] McPherson, M., Smith-Lovin, L. and Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* **27**(1) 415–444.
[26] Michael Pearson, L. M. (2000). Smoke Rings: social network analysis of friendship groups, smoking and drug-taking. *Drugs: Education, Prevention and Policy* **7**(1) 21–37.
[27] Nair, H. S., Manchanda, P. and Bhatia, T. (2010). Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders. *Journal of Marketing Research* **47**(5) 883–895.
[28] Neal, R. M. et al. (2011). MCMC using Hamiltonian dynamics.

Handbook of Markov Chain Monte Carlo **2**(11) 2. MR2858447

[29] Pearson, M. and Michell, L. (2000). Smoke Rings: Social network analysis of friendship groups, smoking and drug-taking. *Drugs: Education Prevention and Policy* **7**.

[30] Pearson, M. and West, P. (2003). Drifting smoke rings: Social network analysis and Markov processes in a longitudinal study of friendship groups and risk-taking. *Connections* **25**.

[31] Roosmalen, E. and Mcdaniel, S. (1989). Peer group influence as a factor in smoking behavior of adolescents. *Adolescence* **24** 801–816.

[32] Serge Moscovici, E. v. A. Gabriel Mugny (1985) *Frontmatter. European Studies in Social Psychology*. Cambridge University Press.

[33] Shalizi, C. R. and Thomas, A. C. (2011). Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *Sociological Methods & Research* **40**(2) 211–239. https://doi.org/10.1177/0049124111404820. MR2767833

[34] Snijders, T., Steglich, C. and Schweinberger, M. (2017). Modeling the coevolution of networks and behavior. In *Longitudinal Models in the Behavioral and Related Sciences* 41–71 Routledge.

[35] Steglich, C., Snijders, T. A. B. and Pearson, M. (2010). Dynamic Networks and Behavior: Separating Selection from Influence. *Sociological Methodology* **40**(1) 329–393.

[36] Steglich, C., Sinclair, P., Holliday, J. and Moore, L. (2012). Actor-based analysis of peer influence in A Stop Smoking In Schools Trial (ASSIST). *Social Networks* **34**(3) 359–369. Dynamics of Social Networks (2).

[37] Sweet, T. and Adhikari, S. (2020). A latent space network model for social influence. *Psychometrika* **85**(2) 251–274. https://doi.org/10.1007/s11336-020-09700-x. MR4128061

[38] Team, S. D. (2017). *shinystan: Interactive Visual and Numerical Diagnostics and Posterior Analysis for Bayesian Models*. R package version 2.4.0. https://mc-stan.org.

[39] Thorlindsson, T. and Vilhjalmsson, R. (1991). Factors related to cigarette smoking and alcohol use among adolescents. *Adolescence* **26**(102) 399–418.

[40] Toulis, P. and Kao, E. (2013). Estimation of causal peer influence effects. In *International conference on machine learning* 1489–1497. PMLR.

[41] Watts, D. J. (2004). The "New" Science of Networks. *Annual Review of Sociology* **30**(1) 243–270.

[42] West, P. and Sweeting, H. (2003). Fifteen, Female and Stressed: Changing Patterns of Psychological Distress Over Time. *Journal of Child Psychology and Psychiatry, and Allied Disciplines* **44** 399–411.

[43] Wooldridge, J. M. (2010) *Econometric analysis of cross section and panel data*. MIT press. MR2768559

[44] Xu, R. (2018). Alternative estimation methods for identifying contagion effects in dynamic social networks: A latent-space adjusted approach. *Social Networks* **54** 101–117.

[45] Xu, R. (2021). Estimating social influence effects in networks using a latent space adjusted approach in R. *The R Journal* **13**(2) 57–69.

[46] Zeger, S. L., Liang, K. -Y. and Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44** 1049–1060. https://doi.org/10.2307/2531734. MR0980999

Brisilda Ndreka. University of Connecticut, USA.
E-mail address: brisilda.ndreka@uconn.edu

Dipak K. Dey. University of Connecticut, USA.
E-mail address: dipak.dey@uconn.edu

Victor H. Lachos. University of Connecticut, USA.
E-mail address: hlachos@uconn.edu