

Contagion Effect Estimate Through the Lens of Different Statistical Methods

Brisilda Ndreka
Department of Statistic
University of Connecticut

December 5, 2022

- Influence model (behaviour) using dynamic linear model.
- Why is estimating contagion effect in social science considered a challenged process?
- Alternative methods which help to achieve a better estimation of the true contagion effects
- Defining the estimation method that has the best performance

Background

- Contagion effect also referred as social influence effect or peer effect, is the tendency of a person or group of individuals has to follow the behaviour of some reference group that they participate.
- The dynamic liner model used to estimate social influence effect is referred to as behaviour model(Xu, 2018; Friedkin & Johnsen, 1990)

$$Y_{it} = \beta_0 + \beta_1 Y_{it-1} + \beta_2 \frac{\sum Z_{ijt-1} Y_{it-1}}{\sum Z_{ijt-1}} + \beta_3 X_{it} + \beta_4 c_i + \epsilon_{it} .$$

Contagion effect

Y is the behavior outcome of interest, X covariates that may affect behavioral outcome Y , Z is network, c is the time-invariant unobserved trait.

Why it is Difficult to Identifying Social Influence Effect (Contagion Effect)?

- Consider the example: In daily life, friends in social networks exhibit similar personality and behaviors. Why does this phenomena happen? Answering this questions can result in two in two direction thinking:
 - Social influence effect: Individuals can influence their friends behaviour, transform their friends characteristics to be similar to theirs.
 - Homophily: In which people with similar characteristics or behaviors are more likely to be friends.
- Consequently, homophily has been identified as a major confounding factor to estimate the influence in social network effect.

This Challenge can be Treated as a Function of an Omitted Variable Bias

- Separating homophily from social influence is a challenging process, particularly when influence effect is confounded with latent homophily caused by unobserved trait.
- If an unobserved variable affects behavioral model and selection process this cause estimation problems of the social influence effect (Xu, 2018; Shalizi & Thomas 2011). Alcohol behaviour among teenager example.

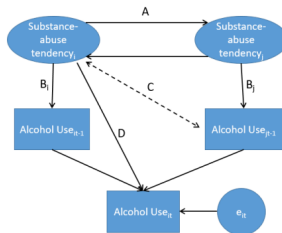


Figure: Omitted variable bias/ (Xu, 2021)

Estimation methods used to estimate latent variable :

- Considering existence of biases of the contagion effects if omitted variables in the influence/selection process exist, four estimation methods are applied:
 - ① Latent space method (Xu, 2018).
 - ② SDNE: Structural Deep Network Embedding (Wang et al, 2016).
 - ③ Node2Vec
 - ④ Latent Factor Model
- Furthermore, fixed effect estimator are used as baseline comparisons. The objective of the study is to determine the method that has the best performance and help to reduce bias of estimating peer effect.

I: Latent Space Adjusted Approach

- Assume the existence of an unobserved trait that co-determines influence and selection process.
- The information of unobserved trait from the selection process as an estimator of the latent variable in the influence model.
- Estimation of c is completed using "latent social positing" from the Latent space models (Hoff, 2002), using ergmm package in R-software.
- By including the latent positions as additional covariates in the behavioral model it will reduce the bias in the estimation of social influence effects.

II: SDNE

- SDNE is machine learning algorithm with the goal of turning a graph into a computationally absorbable format.
- Network embedding intend to capture and carry out the network structure of a network structure using low-dimensional representations of vertexes in networks.
- It is a method that is able to effectively capture the highly non-linear network structure.

Nde2vec

- Node2vec is a random walk based method, which use a walk approach to generate (sample) network neighborhoods for nodes.
- Algorithm computes a vector representation of a node based on random walks in the graph.
- Two parameters control the probability of moving in the graph.
 - return hyperparameter, p
 - inout hyperparameter, q

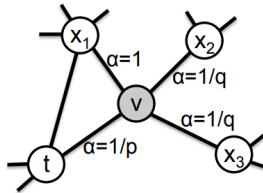


Figure: Random walk procedure in node2vec/ (Grover et al, 2016)

IV: Latent Factor Model

- Each node i has an unknown latent factor $c_i \in R^k$
- The probability of a node i and j to form a tie depends on their latent factors

$$P(Z_{ij} = 1 | c_i, c_j) = \alpha + c_i^T \Lambda c_j,$$

Λ a $K \times K$ diagonal matrix.

- To run a latent factor model, the "amen" package in R-software is used, with dimension of the multiplicative effects 1 & 3, which measure of how likely a pair of actors are to form an edge with one another.

Simulation setting

- Four estimation method considered. Each of them estimate latent variable using dimensions 1 and 3.
- Two network studied, networks with 40 nodes and a network with 80 nodes in two time point (T) each $T = 2$ and 5.
- The real value of the peer influence effect is considered taking values $\beta_2 \in \{0.1, 0.4, 0.7\}$, presenting respectively low-homofoly, mid-homofoly, high-homofoly.
- Two models involved are:
 - Influnlence model:

$$Y_{it} = \beta_0 + \beta_1 Y_{it-1} + \beta_2 \frac{\sum Z_{ijt-1} Y_{jt-1}}{\sum Z_{ijt-1}} + \beta_3 X_{it-1} + \beta_4 C_i + \epsilon_{it}.$$

- Selection model:

$$P(Z_{ij} = 1 | c_i, c_j, x_{ij}, \alpha, \beta) = \Phi(\alpha + \beta |X_i - X_j| - |C_i - C_j|)$$

Simulation results: Contagion effect

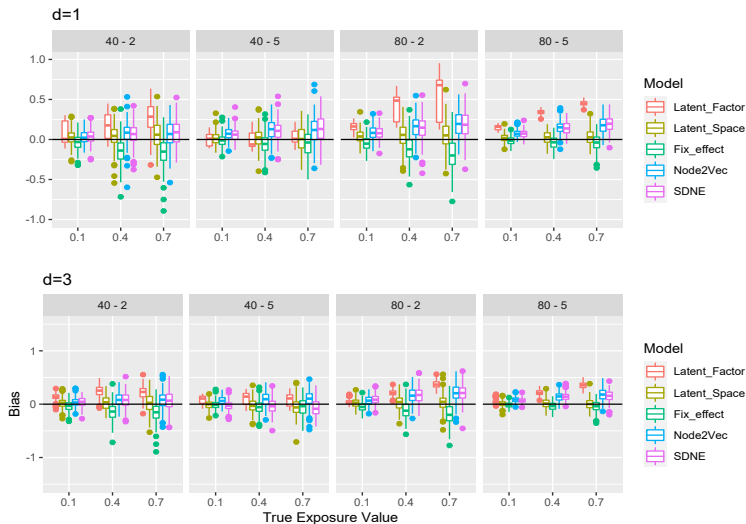


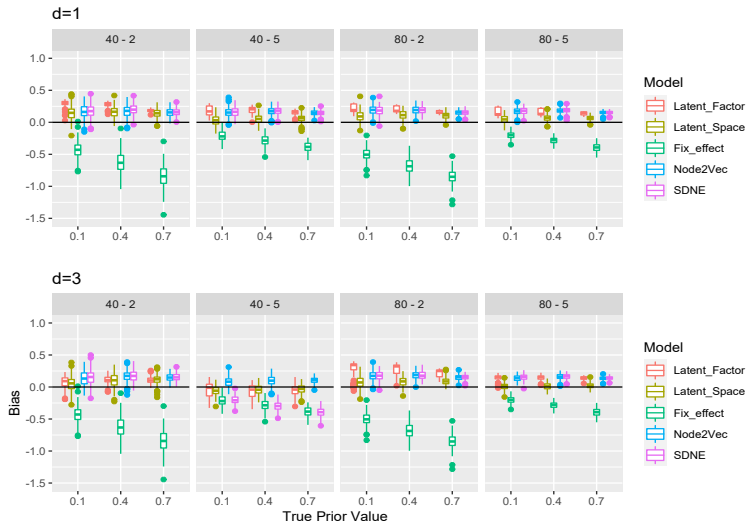
Figure: Bias Distribution for Contagion Effect

Results interpretation

- The magnitude of bias is smaller for all estimation methods when we include more time points ($T=5$ bias estimate smaller than $T=2$).
- Latent space estimates perform compare to other methods in terms of estimating contagion effects, leading in smallest bias.
- Overall, Node2Vec estimates have small bias, but performance is better when contagion effect is small, (presence of low level of homophily) and network size. Moving in larger network method decreases the performance.
- Latent factor estimates are more unstable, with higher bias mostly for small dimension and small time points. Performance improved when this two factors increased.

- SDNE archived good results when network size was relatively small (almost the same bias estimation as latent space), but lost the efficiency in large networks.
- Fixed effects estimates produce the biggest bias among all estimators.
- We should point out that the latent space and latent factor model require a large computing time, since both methods use MCMC (Markov Chain Monte Carlo Algorithm) estimation process.

Simulation Results: Latent Trait Estimation Also Help With Reducing Bias of Y_{t-1} also.



Simulation interpretation:

- Latent space adjusted approach produces much smaller bias in estimating the independent variable Y_{n-1} (previous behaviour) compare to other methods.
- As in the contagion effect the bias value is smaller for all estimation methods when we include more time points (compare for $T=2$ & $T=5$).
- Latent factor estimates show small bias when the true coefficient large $\beta_1 = 0.7$, but is unstable for other scenarios.
- Fixed effects estimates have the larger bias and are always negatively biased.
- In general Node2vec estimates produce a small bias in contradiction to SDNE that was more efficient with contagion estimate.

Reference

- [1] **Sh. Minhas and P. D. Hoff and M. D. Ward**
Inferential Approaches for Network Analysis: AMEN for Latent Factor Models
Political Analysis; 208–222, 2019.
- [2] **R. Xu**
Alternative estimation methods for identifying contagion effects in dynamic social networks: A latent-space adjusted approach
Social Networks; 101–117, 2018.
- [3] **N. E. Friedkin and E. Johnsen**
Social influence and opinions
Journal of Mathematical Sociology; 193–206, 1990.
- [4] **P. D. Hoff and A. E. Raftery and M. S. Handcock**
Latent Space Approaches to Social Network Analysis
Journal of the American Statistical Association; 1090–1098, 2002.
- [5] **R. Xu**
Estimating Social Influence Effects in Networks Using a Latent Space Adjusted Approach in R
The R Journal; 57–69, 2021.
- [6] **C. R. Shalizi and A. C. Thomas**
Homophily and Contagion Are Generically Confounded in Observational Social Network Studies
Sociological Methods & Research; 211–239, 2011.
- [7] **A. Grover and J. Leskovec**
node2vec: Scalable Feature Learning for Networks
CoRR; abs/1607.00653, 2016.
- [8] **D. Wang and P. Cui and W. Zhu**
Structural Deep Network Embedding
Association for Computing Machinery; 1225–1234, 2016.
- [9]