Quantifying homophily through skewed link functions in Bayesian network models: Estimating peer influence

Dipak K. Dey
(Joint with Brisilda Ndreka & Victor H. Lachos)

Department of Statistics
University of Connecticut, Storrs, CT

Introduction

- The similarity-attraction effect is well-known in society. Social choice theory suggests that people tend to form relationships with others with similar characteristics. A phenomenon supported by numerous studies conducted in different social contexts[1][2][3][4].

- Although friends often exhibit more similarities than people without a social connection, clarifying the mechanisms underlying this phenomenon remains a central focus in the study of social relationships.

- Researchers continue to investigate whether these similarities arise from shared experiences (mutual influence), where individuals adapt their behavior to align with their friends, or from selection processes, whereby individuals with similar traits are more likely to form friendships.

---

[1] Shannon R. Bowling et al. "A Logistic Approximation to The Cumulative Normal Distribution". In: *Journal of Industrial Engineering and Management* 2 (2009), pp. 114–127.

[2] Donn Erwin Byrne. *The Attraction Paradigm*. New York: Academic Press, 1971. ISBN: 978-0-12-148650-1.

[3] R. Matthew Montoya and Robert S. Horton. "A Meta-Analytic Investigation of the Processes Underlying the Similarity-Attraction Effect". In: *Journal of Social Personal Relationships* 30.1 (2013), pp. 64–94. DOI: 10.1177/0265407512452989.

[4] Aaron D. Arndt, Kiran Karande, and Myron Glassman. "How Context Interferes with Similarity-Attraction between Customers and Service Providers". In: *Journal of Retailing and Consumer Services* 31 (2016), pp. 294–303. DOI: 10.1016/j.jretconser.2016.04.014.

...

- Many researchers consider selection and influence processes to be mutually exclusive. Therefore, it is essential to differentiate between the impacts of peer influence and homophily to advance academic research and industry practice.

- Contagion Effects (peer effects or social influence)
  - Phenomenon that people tend to mimic behavior of those with whom they have interaction in a social network.

- Homophily (social selection)
  - The tendency of individuals with similar attributes or behaviors to form relationships with one another.
  - Usually, in real-life problems, the likelihood of a connection between individuals in a social network, based on their similarities, can differ.

- In binary friendship classification, imbalanced data occur when the proportion of friendship ties equal to one (or zero) is significantly less than the proportion of corresponding real values of zero (or one).

...

- While the problem of estimating influence effects is of broad interest, the method proposed in this study is particularly motivated by the ever-increasing number of studies on social health behavior, especially concerning vulnerable populations such as teenagers. Research on the influence of peer effects on the younger generation touches on various aspects of daily life[5],[6].

- Estimating peer effects is particularly challenging due to its entanglement with the peer selection process.[7] highlighted a significant concern about the possible existence of unobserved variables that may jointly influence the probabilities of changes in network and/or behavior.

- Studies[8] have shown that when a latent trait jointly influences selection and influence in network data, the contagion effects become largely unidentifiable. This is mainly attributable to the confounding of contagion and homophily (selection) by this latent trait.

[5] Tija Ragelienė and Alice Grønhøj. "Preadolescents' healthy eating behavior: peeping through the social norms approach". In: *BMC Public Health* 20 (2020), p. 1268. DOI: 10.1186/s12889-020-09366-1.

[6] simons2010recent.

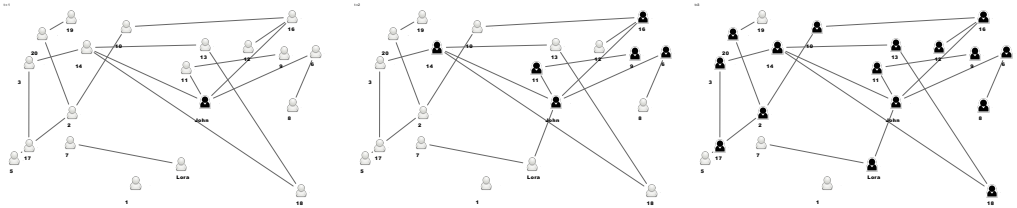
[7] Christian Steglich, Tom A. B. Snijders, and Michael Pearson. "Dynamic Networks and Behavior: Separating Selection from Influence". In: *Sociological Methodology* 40.1 (2010), pp. 329–393. DOI: 10.1111/j.1467-9531.2010.01225.x.

[8] Cosma Rohilla Shalizi and Andrew C. Thomas. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies". In: *Sociological Methods Research* 40.2 (2011), pp. 211–239. DOI: 10.1177/0049124111404820.

...

- **What if your friend's spiral into overspending sparked your own battle with unhealthy spending habits?**
  - Is it possible that spending habits spread through social networks like a contagious disease?-**Contagion Effect**
  - Or is it because people who share similar interests and behaviors are more likely to become friends, thereby increasing their exposure to similar habits?-**Homophily**
  - Alternatively, might hidden traits both drive the friendship and increase the risk of excessive spending? -**Latent homophily** What if your friend's habit of overspending influenced you to start spending more than you should?
    Contagion Effect — Spending habits spread from one person to another, like catching a cold.
    Homophily — People who are alike in interests and behavior tend to become friends, so they naturally share similar spending habits.
    Latent Homophily — Hidden personal traits (like personality or values) make two people both more likely to be friends and more likely to overspend.

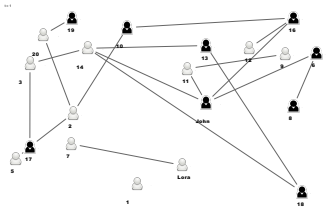
- **Contagion effect** (social influence, peer effect)— a phenomenon that people tend to mimic the behavior of those with whom they have interaction in a social network.



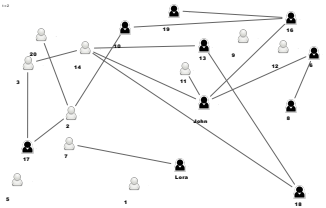
(a) t=1 (b) t=2 (c) t=3

...

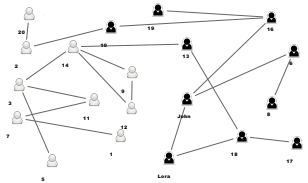
- **Homophily**(social selection) - which refers to the tendency of individuals to form friendships with others who share similar characteristics or behaviors.



(d) t=1



(e) t=2

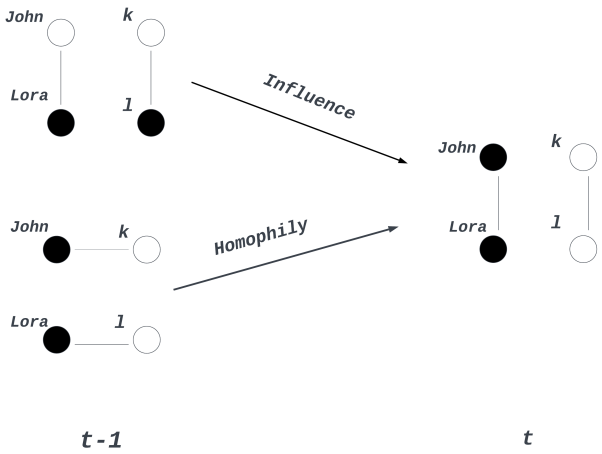


(f) t=3

Background

Why Is Identifying Contagion Effects Challenging???

- Different process, same network structure



...

- Latent homophily: More troublesome ☹
  - People become friends based on unseen traits - more work is needed to isolate peer effects from homophily
  - When a latent trait simultaneously impact both social influence and the selection process in network data, contagion effects become generaly unidentifiable.[9];
  - To get consistent estimates in model using approaches such as OLS, one key assumption is that unobserved errors have to be uncorrelated with observed variables[10]
- Latent space models[11]
  - Used to control for latent homophily in studies of[12][13][14]

[9] Cosma Rohilla Shalizi and Andrew C. Thomas. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies". In: *Sociological Methods & Research* 40.2 (2011), pp. 211–239.

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

[11] Peter D Hoff, Adrian E Raftery, and Mark S Handcock. "Latent space approaches to social network analysis". In: *Journal of the american Statistical association* 97.460 (2002), pp. 1090–1098.

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

[13] Cosma Rohilla Shalizi and Andrew C. Thomas. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies". In: *Sociological Methods & Research* 40.2 (2011), pp. 211–239.

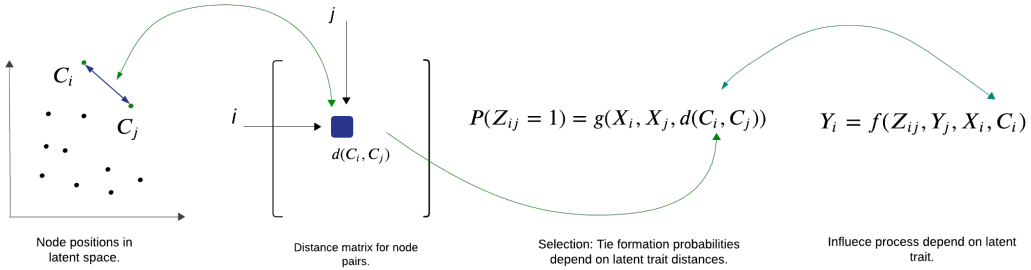
[14] Joseph P Davin, Sunil Gupta, and Mikolaj Jan Piskorski. *Separating homophily and peer influence with latent space*. Harvard Business School, 2014.

Research Problem

- Computationally efficient Bayesian hierarchical modeling
  - high dimensional social network data
  - incorporation of prior information
- Extend current methods by introducing latent space which can control for observed as well as latent homophily
- Robust and interpretable estimation and feature extraction
  - achieve highly accurate estimates and predictions
- Flexible with respect to imbalances in binary-level data
- Apply methods to social network measurements and beyond
  - predict influence predisposition to social health problems

Aims and Goals

- Address complexities of social network data via Bayesian hierarchical modeling approach
- Investigate peer effect while controlling for latent homophily



Method

1. Positions of the sample nodes in the latent space

$$C_i \sim N(0, \sigma^2 I)$$

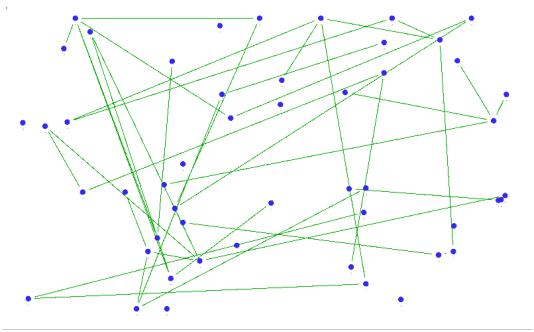
2. Fit selection process:Tie formation probabilities depend on latent trait distances.

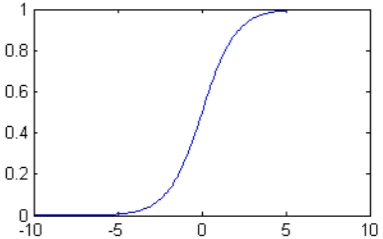
$$P(Z_{ij} = 1) = g(X_i, X_j, d(C_i, C_j))$$

3. Involve estimated latent traits to the influence model.

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

4. Simultaneous estimation of selection and influence dynamics in one Bayes set up





Previour work

Influence model

Selection model: Probit link

"Friendship" network

$$Z_{ijt} \sim Bernoulli(p_{ijt}), \qquad i,j = 1,\ldots,N;\ t = 1,\ldots,T,$$

$$p_{ijt} = \Phi(\alpha_0 + \alpha_1|C_i - C_j| + \alpha_2|X_{it} - X_{jt}|),$$

$$Y_{it}|C_i \sim N(\beta_0 + \beta_1 Y_{it-1} + \beta_2 \frac{\sum_{j=1}^{n} Z_{ijt-1} Y_{jt-1}}{\sum_{j=1}^{n} Z_{ijt-1}} + \beta_3 X_{it} + C_i, \sigma^2),$$

Contagion effect

$i-th$ latent trait

$$C_i \sim N(0,\tau^2),$$
$$\beta_l \sim N(0,\sigma_l^2), \qquad l = 0,\ldots,3,$$
$$\alpha_k \sim N(0,\sigma_k^2), \qquad k = 0,\ldots,2,$$
$$\tau^2 \sim T_+(0, s_\tau^2, \nu_\tau),$$
$$\sigma^2 \sim T_+(0, s_\sigma^2, \nu_\sigma).$$

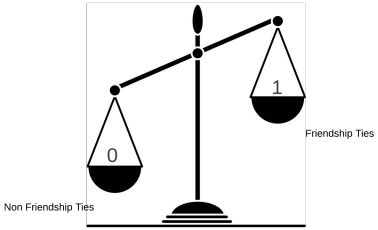
Imbalance Data: Selection Through Skew-Link

- In binary regression contexts, symmetrical links become inadequate when the probability of the response variable approaches 0 at a different rate than it approaches 1. This imbalance can introduce significant bias in the mean response estimate[15].
- This study employs a latent space model to address confounding traits while simultaneously modeling peer effects. The method offers a statistical framework for analyzing peer influence in imbalanced network data—a common issue in real-world networks where, for instance, students form friendships with only 10–20% of peers. The use of skewed link functions provides a novel and effective approach to handle such network imbalances.

[15] Alex de la Cruz Huayanay et al. "Performance of asymmetric links and correction methods for imbalanced data in binary regression". In: *Journal of Statistical Computation and Simulation* 89.13 (2019), pp. 2549–2569. DOI: 10.1080/00949655.2019.1593984. URL: https://doi.org/10.1080/00949655.2019.1593984.

Link Functions Role in Estimation and Prediction

- In social networks, individuals typically have an imbalance between friends and non-friends.



- When the response variable's probability approaches 0 at a different rate than it does 1 (imbalanced data), symmetrical links are especially inappropriate.
  - This can result in significant bias in the estimated mean response[16],[17],[18]
- An appropriate selection model enhances the accuracy of estimating C-latent traits, subsequently improving the influence model

[16] Ming-Hui Chen, Dipak K Dey, and Qi-Man Shao. "A new skewed link model for dichotomous quantal response data". In: *Journal of the American Statistical Association* 94.448 (1999), pp. 1172–1186.

[17] Amalia Luque et al. "The impact of class imbalance in classification performance metrics based on the binary confusion matrix". In: *Pattern Recognition* 91 (2019), pp. 216–231. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2019.02.023. URL: https://www.sciencedirect.com/science/article/pii/S0031320319300950.

[18] Mehrdad Fatourechi et al. "Comparison of Evaluation Metrics in Classification Applications with Imbalanced Datasets". In: *2008 Seventh International Conference on Machine Learning and Applications*. 2008, pp. 777–782.

- Homoplily via Power Cauchy (PC) link[19]

Selection model: Power Cauchy

$$Z_{ijt} \sim Bernoulli(p_{ijt}), \qquad i,j = 1,\ldots,N;\ t = 1,\ldots,T,$$

$$p_{ijt} = (0.5 + \frac{arctan(\alpha_0+\alpha_1|C_i - C_j|+\alpha_2|X_{it}-X_{jt}|)}{\pi})^{\lambda},$$

$$Y_{it}|C_i \sim N(\beta_0 + \beta_1 Y_{it-1} + \beta_2 Contagion_{it} + \beta_3 X_{it} + C_i, \sigma^2),$$

$$C_i \sim N(0,\tau^2),$$

$$\beta_l \sim N(0,\sigma_l^2), \qquad l = 0,\ldots,3,$$

$$\alpha_k \sim N(0,\sigma_k^2), \qquad k = 0,\ldots,2,$$

$$\tau^2 \sim T_+(0, s_\tau^2, \nu_\tau),$$

$$\sigma^2 \sim T_+(0, s_\sigma^2, \nu_\sigma),$$

$$\lambda \sim Gamma(1, s_\lambda^2).$$

[19] Jorge Luis Bazán Guzmán et al. "Power and reversal power links for binary regressions: an application for motor insurance policyholders". In: *Applied Stochastic Models in Business and Industry* (2017). DOI: 10.1002/asmb.2215.

Simulation Study

- A simulation study was conducted to evaluate the performance of asymmetric links for unbalanced data in comparison with symmetric links.
  - The skewness parameter $\gamma = f(\lambda, \alpha)$ represents the proportion of friendship ties in the sociomatrix.

- The unbalanced data are generated using a Probit power link function

$$Z_{ijt} \sim Bernoulli(p_{ijt})$$

where

$$p_{ijt} = \Phi(\alpha_0 + \alpha_1|C_i - C_j| + \alpha_2 Z_{ijt-1})^{\lambda}$$

- Other parameters where set as:
  - $\beta = (\beta_0, \beta_1, \beta_2)^T = (0, 0.5, 0.8)$;
  - $\alpha = (\alpha_0, \alpha_1, \alpha_2)^T = (0, -0.25, -0.5)$;
  - $T = 3$;
  - $N = 20, 50, 100$;
  - $\lambda = 3, 2, 1, 1/5$
  - totally 7 scenarios, each one with 100 replicas

Results/ Varying Skew Level

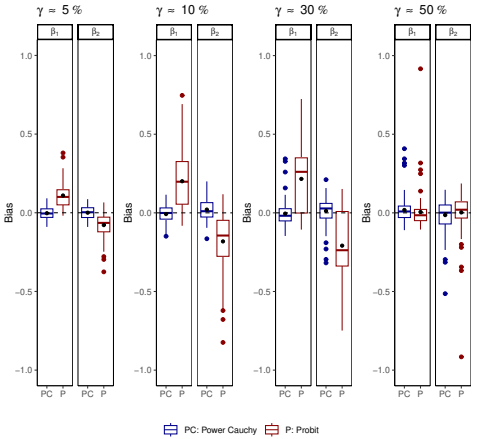


Figure 1: Simulation study—bias of posterior mean of influence parameters considering different skewness parameter values for network sizes $N = 70$.

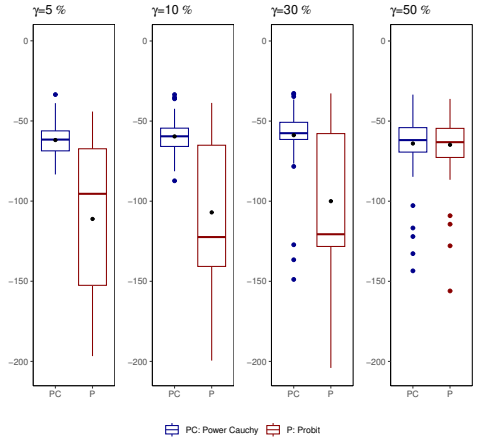


Figure 2: Simulation study—the Bayesian LOO (leave-one-out cross-validation) estimate of the expected log point-wise predictive density.

Results/Varying Network Size

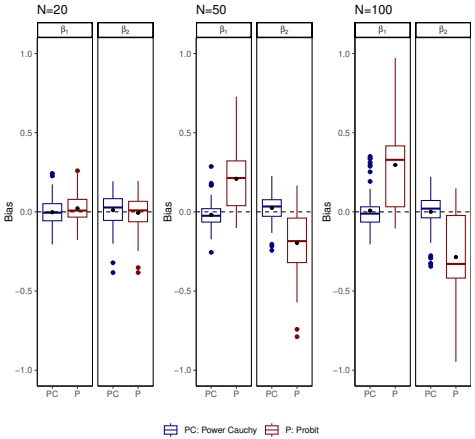


Figure 3: Simulation study—bias of posterior mean of influence parameters considering different network size.

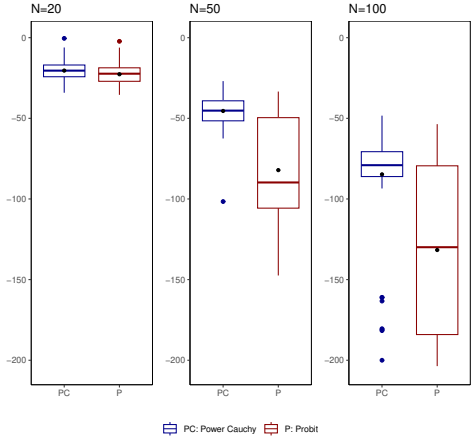


Figure 4: Simulation study—the Bayesian LOO (leave-one-out cross-validation) estimate of the expected log pointwise predictive density.

Results/Including Covariates in a Model

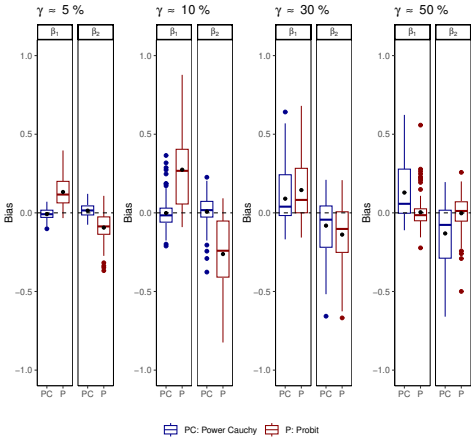


Figure 5: Simulation study—bias of posterior mean of influence parameters considering different skewness parameter values for network sizes $N = 70$.

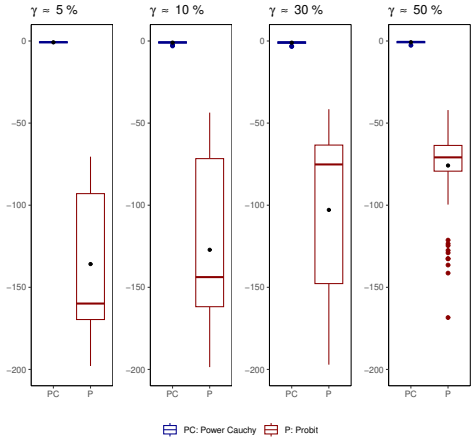


Figure 6: Simulation study—the Bayesian LOO (leave-one-out cross-validation) estimate of the expected log point wise predictive density.

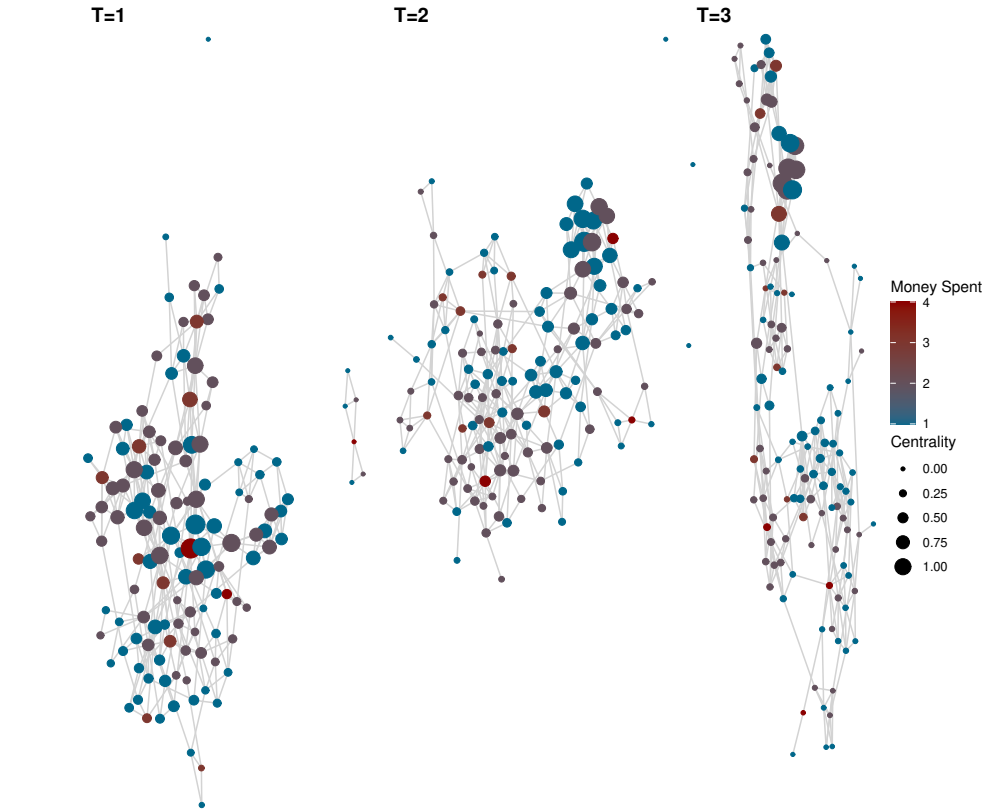
Comments

- Including a skew link in the selection model when modeling imbalanced data significantly reduces bias in estimating social influence effects.
- It is necessary to control for latent homophily when estimating social effects and consider the data's nondiscriminatory imbalance nature.
- Using the standard symmetric link to model selection effects can result in significant inaccuracies when estimating social influence effects, thereby decreasing the predictive accuracy of the influence model

Application/Financial Behaviours

- Imbalance data
  - **Friendship ties** 7% – 12%
- Other data information
  - N = 123, T=3
  - Node color - Money spent in a week (0 £- 100£)
  - Variable of Interest: Weekly Financial Spending
  - Observed Covariates: cannabis consumption, sports behavior, leisure time
- Selection process:
  - via Power Cauchy link
  - covariates involve: Z_{ijt-1}, $d(sport_{it}, sport_{jt})$, $d(drug_{it}, drug_{jt})$, $d(C_i, C_j)$
- Influence process:
  - covariates involve: $money_{it-1}$, $cotagion_{it}$ $canabis_{it}$, $sport_{it}$, $inactive_{it}$.

Friendship Networks



Parameter	Post.mean	Post. SD	2.5%	97.5%	$\hat{R}$	Neff
β_0 – *intercept*	9.3	0.87	7.24	10.69	1	661
β_1 – *prior behaviour*	-0.12	0.01	-0.13	-0.1	1	1209
β_2 – *contagion effect*	-0.07	0.01	-0.09	-0.05	1	1208
β_3 – *sport use*	-2.82	0.06	-2.94	-2.69	1	1276
β_4 – *canabis*	3.62	0.07	3.58	3.75	1	1150
β_5 – *nonactive*	1.08	0.06	0.97	1.19	1	1312

Table 1: Money-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.

- Hamiltonian Monte Carlo (HMC) method has successfully achieved convergence
  - potential scale reduction factor $\hat{R}$ of 1
  - effective number of samples (Neff) is over 100
- Empirical evidence suggesting the impact of the network effect on money-spending behavior

Skew vs. Symmetric Link

- Fit the data using a symmetric Probit link for the selection process
- A superior model with the skew Power Cauchy link is shown by a lower WAIC and higher ELPD

	Probit Model	Power Cauchy Model
ELPD	-1017.0	-903.2
WAIC	2035.5	1806.4

Future Potential Application

- Contagion effects can be used in social network studies to understand and analyze various behaviors, such as:
  - Health Behaviors
  - Emotional Contagion
  - Technology Adoption
  - Consumer Behavior
  - Educational Performance
  - Public Health
  - Workplace Dynamics
- Other domains with comparable features and spatio-temporal structure as:
  - Epidemiology
  - Finance
  - Urban Studies
  - Political Science

THANK YOU!

Reference I

[1] Aaron D. Arndt, Kiran Karande, and Myron Glassman. "How Context Interferes with Similarity-Attraction between Customers and Service Providers". In: *Journal of Retailing and Consumer Services* 31 (2016), pp. 294–303. DOI: 10.1016/j.jretconser.2016.04.014.

[2] Jorge Luis Bazán Guzmán et al. "Power and reversal power links for binary regressions: an application for motor insurance policyholders". In: *Applied Stochastic Models in Business and Industry* (2017). DOI: 10.1002/asmb.2215.

[3] Shannon R. Bowling et al. "A Logistic Approximation to The Cumulative Normal Distribution". In: *Journal of Industrial Engineering and Management* 2 (2009), pp. 114–127.

[4] Donn Erwin Byrne. *The Attraction Paradigm*. New York: Academic Press, 1971. ISBN: 978-0-12-148650-1.

[5] Ming-Hui Chen, Dipak K Dey, and Qi-Man Shao. "A new skewed link model for dichotomous quantal response data". In: *Journal of the American Statistical Association* 94.448 (1999), pp. 1172–1186.

Reference II

[6] Alex de la Cruz Huayanay et al. "Performance of asymmetric links and correction methods for imbalanced data in binary regression". In: *Journal of Statistical Computation and Simulation* 89.13 (2019), pp. 2549–2569. DOI: 10.1080/00949655.2019.1593984. URL: https://doi.org/10.1080/00949655.2019.1593984.

[7] Joseph P Davin, Sunil Gupta, and Mikolaj Jan Piskorski. *Separating homophily and peer influence with latent space*. Harvard Business School, 2014.

[8] Mehrdad Fatourechi et al. "Comparison of Evaluation Metrics in Classification Applications with Imbalanced Datasets". In: *2008 Seventh International Conference on Machine Learning and Applications*. 2008, pp. 777–782.

[9] Peter D Hoff, Adrian E Raftery, and Mark S Handcock. "Latent space approaches to social network analysis". In: *Journal of the american Statistical association* 97.460 (2002), pp. 1090–1098.

Reference III

[10] Amalia Luque et al. "The impact of class imbalance in classification performance metrics based on the binary confusion matrix". In: *Pattern Recognition* 91 (2019), pp. 216–231. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2019.02.023. URL: https://www.sciencedirect.com/science/article/pii/S0031320319300950.

[11] R. Matthew Montoya and Robert S. Horton. "A Meta-Analytic Investigation of the Processes Underlying the Similarity-Attraction Effect". In: *Journal of Social Personal Relationships* 30.1 (2013), pp. 64–94. DOI: 10.1177/0265407512452989.

[12] Tija Ragelienė and Alice Grønhøj. "Preadolescents' healthy eating behavior: peeping through the social norms approach". In: *BMC Public Health* 20 (2020), p. 1268. DOI: 10.1186/s12889-020-09366-1.

[13] Cosma Rohilla Shalizi and Andrew C. Thomas. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies". In: *Sociological Methods Research* 40.2 (2011), pp. 211–239. DOI: 10.1177/0049124111404820.

Reference IV

[14] Cosma Rohilla Shalizi and Andrew C. Thomas. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies". In: *Sociological Methods & Research* 40.2 (2011), pp. 211–239.

[15] Christian Steglich, Tom A. B. Snijders, and Michael Pearson. "Dynamic Networks and Behavior: Separating Selection from Influence". In: *Sociological Methodology* 40.1 (2010), pp. 329–393. DOI: 10.1111/j.1467-9531.2010.01225.x.

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