

Research report

Enhancing Causal Inference with Network Information

Polina Revina (8461465)

Supervisors: Erik-Jan van Kesteren, Javier Garcia-Bernardo

*Methodology and Statistics for the Behavioural, Biomedical and Social Sciences
Utrecht University*

January 16, 2025

Word count: 1863

Candidate journals: Statistical Science; Sociological Methods & Research

1 Introduction

In many experimental and observational studies, network interference is present between units, which means that the treatment of one unit affects the results of its neighbors (Cox, 1958). This happens because of transmission of exposure through social interaction or physical proximity of units. For example, students who received tutoring support can interact with other students not assigned to this program and transmit the knowledge obtained during tutoring (Forastiere, Airoldi, & Mealli, 2021). Another example is vaccine trials on infectious diseases, when treated individuals can affect the likelihood of disease transmission and influence the infection rate.

The assumption of no interference lies at the heart of causal inference, forming a fundamental component of the Stable Under Treatment Value Assumption (Rubin, 1974). As demonstrated by Sobel (2006), using methods that do not account for interference or imposing this assumption incorrectly can lead to an entirely wrong conclusion about the effect. This complicates causal estimations in settings with interdependent units, such as social networks, communities, or economic systems.

Addressing the issue of interference led to the development of two primary types of strategies: design strategies that focus on mitigating interference in experiments and inference strategies that use network data to adjust estimations. The widely used approach for estimation in these settings is to relax the no interference assumption by localizing interference within certain structures or defining its form. Under partial interference, units within each cluster can affect each other’s outcome, but between clusters, no interference occurs (Sobel, 2006). This assumption is often used with stratified interference, where the level of interference depends on the proportion of treated units within the group (Sävje, Aronow, & Hudgens, 2021). This framework provides a foundation for the development of experimental designs, such as cluster-based randomization and its extensions such as two-stage randomization (Hudgens & Halloran, 2008), graph cluster randomization (Ugander, Karrer, Backstrom, & Kleinberg, 2013).

Recent studies focused on relaxing the partial interference assumption, allowing units to interfere along more general structures, such as social networks (Manski, 2013), but suggested estimation methods require knowledge of the interference structure. To overcome this shortage, Savje et al. (2021) proposed a new estimand Expected Average Treatment Effect (EATE), which marginalizes the unit-level treatment effect over all possible reference assignments, capturing the average treatment effect while accounting for interference. The proposed estimand shows consistency under weak interference assumption without knowledge of the interference structure.

However, these assumptions about interference are rarely likely to hold, especially in high-dimensional networks, so another approach uses new estimators to adjust for network information. Forastiere et al. (2021) propose a generalized propensity score method that adjusts for network information by balancing individual and neighborhood covariates. This method aims to reduce confounding by estimating the probabilities of treatment assignment while accounting for the relationships between individuals in the network. Another significant contribution is the extension of the Targeted Maximum Likelihood Estimator, enabling it to handle situations where the treatment effect is influenced by spillovers from neighboring units (Ogburn, Sofrygin, Diaz, & Van der Laan, 2024).

Recent studies utilize modern developments of neural networks to adjust inference. Leung et al. (2022) used graph neural networks (GNN), a powerful method to capture complex dependencies in the graph to estimate the direct effect under interference. Khatami et al, (2024) combined GNN with double machine learning framework for unbiased and

efficient estimations, thus providing valid confidence intervals.

1.1 Research question

There is still a limited understanding of the degree of bias of causal inference estimators in observational studies under different exposure scenarios and network structures. Current studies are more focused on experimental design than observational studies and often rely on the partial interference assumption, which may not reflect more complex scenarios where interference can occur across groups. Existing methods for network adjustment also lack a theoretical base and their application in observational studies is limited.

This project aims to enhance causal inference methods for observational data by incorporating network information. The research question is *"To what extent does network interference produce bias in standard causal inference estimates and how can network properties be used to reduce the bias?"*

2 Data

For the analysis, I generated synthetic data to simulate how both the outcome and treatment spread through the network, reflecting the underlying structure and interactions. By controlling for relevant variables, estimation bias from not accounting for network interference can be estimated. The key parameters in the data generation process are summarized in Table 1.

Parameter	Description
Network Models	Barabási-Albert, Watts-Strogatz, Erdős-Rényi (10,000 nodes each)
Neighbor Influence	Small (0.1), Medium (0.3), Large (0.7)
Covariate	$N(5, 20)$
Treatment effect	Medium (0.3)

Table 1: Key parameters and descriptions used in the data generation process.

2.1 Treatment assignment

I assume a simple contagion, meaning that one connection to the treated is enough for the effect transmission (Cencetti, Contreras, Mancastroppa, & Barrat, 2023). One of the features of this mechanism is that the rate of adoption is proportional to the number of exposures, so being in contact with two treated individuals doubles the overall rate. The assignment of treatment is random, and the split between control and treatment groups is equal. For the first simulation the medium effect (30%) is used.

2.2 Network generation

Different types of networks are used in this research to account for various phenomena. This allows us to capture a range of scenarios, such as friends groups, social media websites or purely random interaction. Moreover, as it remains unclear whether and how the type of network structure influences the estimation bias of network interference, it is essential to explore different models. For every network model, ten thousand nodes are generated. The generated network will be represented as an adjacency matrix, denoted W , which encodes the connections between nodes and serves as the foundation for further analysis.

The *Barabási-Albert model* is well suited to simulating real-world systems that exhibit scale-free properties, such as friendship groups, high school social networks, or online platforms, where a small number of nodes tend to accumulate a disproportionately large number of connections (Bertotti & Modanese, 2019).

In addition, the *Watts-Strogatz model* is generated to create small-world networks, characterized by high clustering and short average path lengths, features commonly observed in many social and biological systems. This model is useful for studying phenomena like information diffusion or community formation, as it captures the balance between local and global connectivity within a network (Song & Wang, 2014).

The *Erdős-Rényi model* generates random networks in which each possible edge between nodes has a fixed probability. This model is often used as a benchmark to understand how randomness alone can shape network structures and dynamics. While it lacks the clustering seen in the Watts-Strogatz model or the hubs in the Barabási-Albert model, it provides a baseline for comparison by modelling purely stochastic connections.

2.3 Covariate

Covariate represents individual-level characteristics, such as age or income. They are generated from a normal distribution $N(5, 20)$, its distribution does not have a theoretical foundation and is used only for data generation.

2.4 Outcome

The outcomes in the network will be simulated using a **spatial autocorrelation model** (SAR), which accounts for the dependencies incorporated in the network structure. This model captures how an individual’s outcome is influenced by their neighbours’ responses, individual-level covariates, and independent noise.

A key assumption of the SAR model is indirect interference, which implies that spillovers are proportional to the effect of direct treatment. That is, if there is no direct effect of treatment on a unit, there will be no spillovers to its neighbours. This assumption is integral to the linearity of the model and serves as a foundation for simulating outcomes in networks.

Formally, the outcome can be expressed as: $y = (I - \rho W)^{-1} X \beta + (I - \rho W)^{-1} \epsilon$, where ρ represents the influence of neighbors, W is the adjacency matrix capturing network connections, X is a matrix of covariates, and ϵ represents independent noise.

The coefficients β are sampled from a normal distribution $N(3, 10)$, and the neighbour influence parameter ρ will take on values corresponding to small (0.1), medium (0.3), and large (0.7) levels of spillover intensity. This approach allows us to model varying degrees of dependence between connected individuals and analyze the propagation of treatment effects through the network.

3 Methods

3.1 Linear regression

In this context, the estimand of interest is the Average Treatment Effect (ATE), which captures the average difference in outcomes between the treatment and control groups. The ATE provides a measure of the impact of the treatment assignment in the population when SUTVA holds.

As treatment is assigned randomly, to estimate effect I used the following linear regression: $y = \beta_0 + \beta_1 * Covariate_i + \beta_2 * Group_i + \epsilon_i$, where $Group_i$ indicates individuals' group and $Covariate_i$ individual covariates. The key expectation without interference is that the estimated treatment effect, represented by β_2 should align with the true effect used to generate the data when there is no bias.

3.2 Network adjustment

To account for the bias that arises from network interference, network information can be incorporated into estimation methods. In this way, the estimation can become more robust to network interference, providing more accurate estimates of treatment effect.

One approach is to include network features as covariates in the linear regression model. Key features that capture the structural properties of the network include:

- **Degree centrality:** the number of edges connected to a node, representing the extent of an individual's direct connections within the network.
- **Betweenness Centrality:** the frequency with which a node appears on the shortest paths between other nodes, reflecting its role as a bridge or intermediary in the network.
- **Community membership:** the group to which a node belongs, identifying subsets of individuals with greater connectivity within the group.

4 Preliminary results

The results of the analysis of network structures with varying levels of neighbor influence show that, in general, interference bias is on average 32.4%. As can be seen from Table 2 there is also considerable variation in the effect differences between different types of networks. Overall, the interference effect affects the estimation more with increasing neighbor influence. In the Erdos-Renyi network this dependency is less clear likely due to the more dispersed nature of connection between units. Figure 1 shows that the range of estimated and true effect also differs across simulations.

With network adjustment, the bias is reduced to 20.5%, which is true for all network types and neighbour influence. This reduction highlights the effectiveness of adjustment techniques in controlling interference bias.

Network	Neighbour influence	Effect difference	Effect difference (with adjustment)
Barabasi-Albert (n_edges=3)	0.3	6.2%	4%
Barabasi-Albert (n_edges=3)	0.6	8.4%	6%
Barabasi-Albert (n_edges=3)	0.9	20%	15%
Watts Srogatz (n_edges=3, p=0.7)	0.3	8%	6%
Watts Srogatz (n_edges=3, p=0.7)	0.6	15%	12%
Watts Srogatz (n_edges=3, p=0.7)	0.9	148%	72%
Erdos Renyi (n_edges=3, p=0.1)	0.3	40%	32%
Erdos Renyi (n_edges=3, p=0.1)	0.6	8%	8%
Erdos Renyi (n_edges=3, p=0.1)	0.9	38%	30%

Table 2: Comparison of effect differences with and without adjustments across different networks.

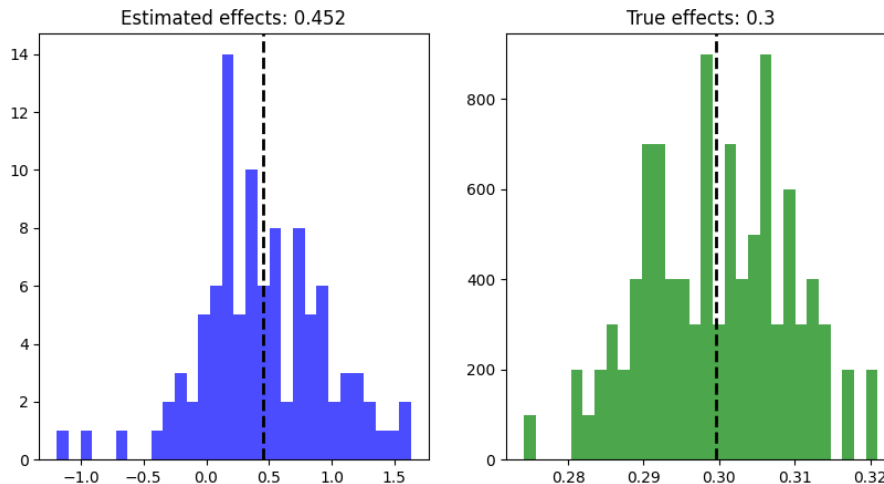


Figure 1: Distribution of average estimated and true effects.

5 Discussion

One limitation of this study is that it relies on synthetic data and idealized network structures, such as the adjacency matrix, which may not accurately reflect the complexity of real-world networks. In real-world scenarios, network structures are often dynamic, and detailed information about the connections between units may not be available. The study also focuses on the effects of network interference within static networks and does not account for evolving interactions or complex contagion. This shows the limits of applications of the provided adjustment method.

To address these limitations and enhance the applicability of the findings, future research will focus on several key directions. First, we will explore nonrandom assignment mechanisms to better reflect real-world settings where treatment assignment is often influenced by underlying network characteristics or external factors. Second, advanced representation learning techniques, such as Node2Vec, will be integrated to better capture node representations. These extensions will enable a deeper understanding of the estimations under network interference in complex networks.

References

- Agarwal, C., Lakkaraju, H., & Zitnik, M. (2021). Towards a unified framework for fair and stable graph representation learning. In *Uncertainty in artificial intelligence* (pp. 2114–2124).
- Aronow, P. M., & Samii, C. (2017). Estimating average causal effects under general interference, with application to a social network experiment.
- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, 46(3), 399–424.
- Bertotti, M. L., & Modanese, G. (2019). The configuration model for barabasi-albert networks. *Applied Network Science*, 4, 1–13.
- Bhattacharya, R., Malinsky, D., & Shpitser, I. (2020). Causal inference under interference and network uncertainty. In *Uncertainty in artificial intelligence* (pp. 1028–1038).

- Bhattacharya, S., & Sen, S. (2024). Causal effect estimation under network interference with mean-field methods. *arXiv preprint arXiv:2407.19613*.
- Cencetti, G., Contreras, D. A., Mancastropa, M., & Barrat, A. (2023). Distinguishing simple and complex contagion processes on networks. *Physical Review Letters*, 130(24), 247401.
- Cox, D. R. (1958). Planning of experiments.
- Forastiere, L., Airoidi, E. M., & Mealli, F. (2021). Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association*, 116(534), 901–918.
- Hagberg, A., Swart, P. J., & Schult, D. A. (2008). *Exploring network structure, dynamics, and function using networkx* (Tech. Rep.). Los Alamos National Laboratory (LANL), Los Alamos, NM (United States).
- Higgott, O. (2022). Pymatching: A python package for decoding quantum codes with minimum-weight perfect matching. *ACM Transactions on Quantum Computing*, 3(3), 1–16.
- Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*, 101(475), 901–910.
- Hudgens, M. G., & Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482), 832–842.
- Khatami, S. B., Parikh, H., Chen, H., Roy, S., & Salimi, B. (2024). Graph neural network based double machine learning estimator of network causal effects. *arXiv preprint arXiv:2403.11332*.
- Leung, M. P., & Loupos, P. (2022). Graph neural networks for causal inference under network confounding. *arXiv preprint arXiv:2211.07823*.
- Li, W., Sussman, D. L., & Kolaczyk, E. D. (2021). Causal inference under network interference with noise. *arXiv preprint arXiv:2105.04518*.
- Ma, J., Wan, M., Yang, L., Li, J., Hecht, B., & Teevan, J. (2022). Learning causal effects on hypergraphs. In *Proceedings of the 28th acm sigkdd conference on knowledge discovery and data mining* (pp. 1202–1212).
- Ma, Y., & Tresp, V. (2021). Causal inference under networked interference and intervention policy enhancement. In *International conference on artificial intelligence and statistics* (pp. 3700–3708).
- Manski, C. F. (2013). Identification of treatment response with social interactions. *The Econometrics Journal*, 16(1), S1–S23.
- Ogburn, E. L., Sofrygin, O., Diaz, I., & Van der Laan, M. J. (2024). Causal inference for social network data. *Journal of the American Statistical Association*, 119(545), 597–611.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5), 688.
- Sävje, F., Aronow, P., & Hudgens, M. (2021). Average treatment effects in the presence of unknown interference. *Annals of statistics*, 49(2), 673.
- Sobel, M. E. (2006). What do randomized studies of housing mobility demonstrate? causal inference in the face of interference. *Journal of the American Statistical Association*, 101(476), 1398–1407.
- Sofrygin, O., & van der Laan, M. J. (2017). Semi-parametric estimation and inference for the mean outcome of the single time-point intervention in a causally connected population. *Journal of causal inference*, 5(1), 20160003.
- Song, H. F., & Wang, X.-J. (2014). Simple, distance-dependent formulation of the watts-

- strogatz model for directed and undirected small-world networks. *Physical Review E*, 90(6), 062801.
- Ugander, J., Karrer, B., Backstrom, L., & Kleinberg, J. (2013). Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the 19th acm sigkdd international conference on knowledge discovery and data mining* (pp. 329–337).
- Van der Laan, M. J. (2014). Causal inference for a population of causally connected units. *Journal of Causal Inference*, 2(1), 13–74.
- Wang, M., Zheng, D., Ye, Z., Gan, Q., Li, M., Song, X., ... others (2019). Deep graph library: A graph-centric, highly-performant package for graph neural networks. *arXiv preprint arXiv:1909.01315*.
- Zhao, Z., Bai, Y., Xiong, R., Cao, Q., Ma, C., Jiang, N., ... Kuang, K. (2024). Learning individual treatment effects under heterogeneous interference in networks. *ACM Transactions on Knowledge Discovery from Data*, 18(8), 1–21.
- Zhu, H., Zhang, S., Su, Y., Zhao, Z., & Chen, N. (2024). Integrating active learning in causal inference with interference: A novel approach in online experiments. *arXiv preprint arXiv:2402.12710*.