Network Causal Effect Estimation In Graphical Models Of Contagion And Latent Confounding

Yufeng Wu SW20@WILLIAMS.EDU
Williams College

Rohit Bhattacharya RB17@WILLIAMS.EDU
Williams College

Editors: Biwei Huang and Mathias Drton

Abstract

A key question in many network studies is whether the observed correlations between units are primarily due to contagion or latent confounding. Here, we study this question using a segregated graph (Shpitser, 2015) representation of these mechanisms, and examine how uncertainty about the true underlying mechanism impacts downstream computation of network causal effects, particularly under full interference—settings where we only have a single realization of a network and each unit may depend on any other unit in the network. Under certain assumptions about asymptotic growth of the network, we derive likelihood ratio tests that can be used to identify whether different sets of variables—confounders, treatments, and outcomes—across units exhibit dependence due to contagion or latent confounding. We then propose network causal effect estimation strategies that provide unbiased and consistent estimates if the dependence mechanisms are either known or correctly inferred using our proposed tests. Together, the proposed methods allow network effect estimation in a wider range of full interference scenarios that have not been considered in prior work. We evaluate the effectiveness of our methods with synthetic data and the validity of our assumptions using real-world networks.

Keywords: Interference, Segregated graphs, Social networks

1. Introduction

A key question that arises in many network studies is whether observed correlations between units arise primarily due to *contagion* (peer-to-peer influence) or *latent confounding* (unobserved background factors that lead to similar characteristics for connected individuals) (Jackson and López-Pintado, 2013; Rosenbusch et al., 2019). This mechanistic knowledge can inform different intervention strategies to influence network-wide changes for certain outcomes of interest. We study this question in full interference settings—settings where each unit may potentially depend on any other unit in the network, and we only have a single realization of the network—and how this difference can impact downstream computation of network causal effects.

There is a growing literature on studying interference problems through causal graphical models. Ogburn and VanderWeele (2014) proposed an approach to studying interference problems using causal directed acyclic graphs (causal DAGs). Srinivasan et al. (2023) extend this approach to also account for different types of non-ignorable missing data including missingness interference; Ogburn et al. (2020); Bhattacharya et al. (2020); Peña (2020) use various interpretations of chain graphs (CGs) to study interference. The aforementioned works focus primarily on the partial interference setting. In the full interference setting, Tchetgen Tchetgen et al. (2021) propose a chain graph representation for interference along with an estimation strategy, known as *auto-g-computation*, for

network effects. However, chain graph models do not admit representations of latent confounding, so the auto-g-computation algorithm cannot be applied in such settings. Full interference work in Ogburn et al. (2024) and Clark and Handcock (2024) have similar limitations. Bhattacharya and Sen (2024) propose a mean-field/message passing algorithms for auto-g-computation.

Here, we use causal interpretations of segregated graphical models (Shpitser, 2015) to model interference. These models allow for the representation of both contagion and latent confounding. Sherman and Shpitser (2018) provide a sound and complete identification algorithm for network effects in a certain class of segregated graphs. Their proposed estimation strategy, however, relies on directly applying auto-g-computation and so can only be applied in special cases where latent confounding can be ignored for estimation purposes. Moreover, none of the aforementioned works consider model selection procedures for separating contagion from latent confounding, a question that is of great scientific interest, but relatively understudied in the current literature (Shalizi and Thomas, 2011; Ogburn, 2018; Lee and Ogburn, 2021). We add to the existing literature as follows.

1. We propose likelihood ratio tests, based on a coding likelihood (Besag, 1974), that can be used to distinguish between dependence due to contagion and latent confounding under certain distributional assumptions and assumptions about asymptotic growth of the network.[1]
2. We also propose coding likelihood estimators for network causal effects that are consistent and asymptotically normal under different regimes of dependence due to contagion and latent confounding and operate under weaker network asymptotics than our tests.

Contribution (1) can be seen as extending the existing causal discovery literature in interference by adding new hypothesis tests for distinguishing between contagion and latent confounding in segregated graphical models under full interference. From a causal inference perspective, contribution (2) extends the auto-g-computation method of Tchetgen Tchetgen et al. (2021) to handle latent confounding. To our knowledge, neither of these directions have been pursued in prior work. We envision that (1) and (2) will often be used in conjunction with each other—the tests serve as a check to determine whether auto-g-computation is sufficient, or whether our extension of it is required.

2. Model and Problem Setup

We model interference using segregated graphs (SGs) (Shpitser, 2015), a class of loopless[2] mixed graphs where vertices can be connected by directed ($\rightarrow$), bidirected ($\leftrightarrow$), and undirected edges ($-$). A loopless mixed graph $\mathcal{G}(V)$ is an SG if, (i) Any pair of vertices in V can be connected either by a single edge of any type, or a pair of edges with one being directed and the other being bidirected; (ii) There is no vertex in V that is an endpoint of both a bidirected and undirected edge; (iii) There are no partially directed cycles—a partially directed walk starting and ending at the same vertex V_i consisting of a mix of directed and undirected edges with at least one directed edge.[3] We provide some examples of graphs that do and do not satisfy the SG property in Appendix A. For brevity, we use $V_i \circ\!\!-\!\!\circ V_j$ to denote uncertainty about whether V_i and V_j are connected via an undirected or bidirected edge and denote $\mathcal{G}(V)$ as simply $\mathcal{G}$ when the vertex set is clear from context.

1. In particular, we require that as the network grows, we also obtain a growing number of units that are at least 6 degrees of separation away from each other, a connection to the small-world principle (Watts and Strogatz, 1998). This may not hold in all network settings, and we examine its viability in Section 6 using five real-world networks.
2. Graphs that have no edges of the form $V_i *\!\!-\!\!* V_i$.
3. As an example, $A - B \rightarrow C - A$ is considered a partially directed cycle, but $A - B - C - A$ is not.

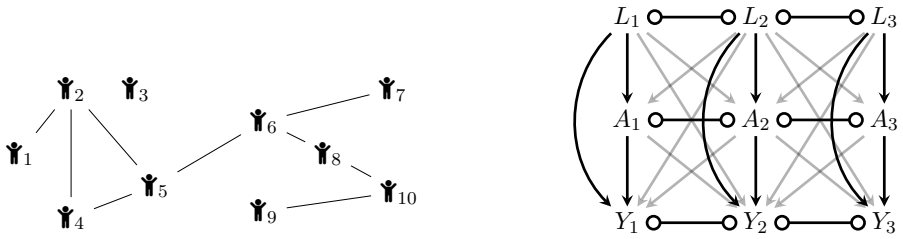

Figure 1: (a) Example of a friendship network $\mathcal{F}$ on 10 units. (b) Partially determined SG pattern $\mathcal{G}_{\circ\!-\!\circ}$ on 3 units (lighter edge colors in the figure are only used to improve readability).

2.1. Statistical models of an SG

A statistical model of an SG $\mathcal{G}(V)$ is defined as the set of distributions $p(V)$ that *segregated factorize* according to $\mathcal{G}$, or equivalently, satisfy a global Markov property defined in terms of *s-separation* (Shpitser, 2015), a generalization of the c-separation criterion (Studený, 1996) for chain graphs (segregated graphs with no bidirected edges corresponding to no latent confounding). Briefly, s-separation is similar to c-separation, but generalizes the notion of colliders and collider sections (e.g., $A \to B \leftarrow C$ and $A \to B - C \leftarrow D$) to also allow for bidirected edges (e.g., $A \leftrightarrow B \leftarrow C$). A more precise description of s-separation in terms of an augmented graph aka moralization criterion (Lauritzen, 1996) is provided in Appendix B. The global Markov property of SGs is as follows: if $X \perp\!\!\!\perp_{\text{s-sep}} Y \mid Z$ in $\mathcal{G}(V)$ then $X \perp\!\!\!\perp Y \mid Z$ in $p(V)$. In our work, to facilitate mechanism discovery, we assume that the converse is also true, which is akin to the *faithfulness* assumption made by many constraint and score-based causal discovery methods (Spirtes et al., 2000). Thus for statistical models of an SG considered here we have $X \perp\!\!\!\perp Y \mid Z$ in $p(V) \iff X \perp\!\!\!\perp_{\text{s-sep}} Y \mid Z$ in $\mathcal{G}(V)$.

2.2. Modeling interference with SGs

Lauritzen and Richardson (2002) provided a causal interpretation for segregated graphs with no bidirected edges, known as chain graphs (CGs). Under this interpretation, a directed edge $V_i \to V_j$ indicates that V_i is a potential cause of V_j while an undirected edge $V_i - V_j$ represents the possibility of a symmetric causal relationship that reaches an equilibrium state. In the interference literature, this symmetric causal relationship is often interpreted as being the result of a contagious process; see for example Ogburn et al. (2020); Bhattacharya et al. (2020); Tchetgen Tchetgen et al. (2021).

The above interpretations of directed and undirected edges naturally carry forward to causal interpretations of SGs. We now elaborate on the interpretation of bidirected edges. An SG $\mathcal{G}(V)$ can be thought of as encoding the marginal distribution of a hidden variable chain graph $\mathcal{G}(V \cup H)$ with observed variables V and hidden variables H, such that for every edge $V_i \leftrightarrow V_j$ in $\mathcal{G}(V)$, there exists an exogenous hidden variable H_{ij} connecting V_i and V_j in $\mathcal{G}(V \cup H)$ as $V_i \leftarrow H_{ij} \to V_j$ (Shpitser, 2015).[4] That is, bidirected edges in an SG represent latent confounding. Equipped with these definitions, we describe the goals and assumptions of our work.

4. The pattern of hidden variables in the chain graph need not exactly match this description for there to be a bidirected edge. For example, a path $V_i \leftarrow H_1 \to H_2 \to V_j$ involving multiple hidden variables is also possible. However, these are all equivalent as they imply the same marginal model, so we opt for the simpler representation.

2.3. Goals and Graphical Assumptions

We observe N potentially interconnected units in a network, representing full interference. This network of connections or friendships can be represented by an undirected graph $\mathcal{F}$ with N vertices labeled $\textbf{1}, \ldots, \textbf{N}$ and edges $\textbf{i} - \textbf{j}$ for every pair of units i, j that are connected to each other. For each unit i in the network, we observe a vector of baseline covariates L_i, their treatment A_i, and their final outcome Y_i. That is, our observed data consists of (possibly) dependent realizations $(L, A, Y) = (L_1, A_1, Y_1), \ldots, (L_N, A_N, Y_N)$. In this setting, we set our target of inference to be the expected value of unit-level potential outcomes of the form $\mathbb{E}[Y_i \mid \text{do}(A = a)]$ for $i = 1, \ldots, N$. The quantity $\mathbb{E}[Y_i \mid \text{do}(a)]$ represents the expected outcome of $\textbf{i}$ when all units in the network receive the *treatment assignment vector* $A = a$. Once these are known, it is easy to compute a variety of other network effects, such as the population average overall effect and spillover effect.

To make this task feasible, we make the following simplifying assumptions: (1) We assume that the friendship network $\mathcal{F}$ is known; (2) The data $(L_1, A_1, Y_1), \ldots, (L_N, A_N, Y_N)$ are drawn from a distribution that is Markov and faithful with respect to an SG $\mathcal{G}$ that is consistent with the *partially determined* SG pattern shown in Figure 1(b) for three units, with the restriction that variables of the same type (L, A, or Y) are always connected by the same type of edge (partially determined refers to the fact that there is uncertainty about whether the ○—○ edges are bidirected or undirected). The latter restriction implies, for example, that if the outcome is contagious for two connected units, then it is also contagious for all other connected units (and similarly for latent confounding). These assumptions are similar to those adopted in prior work (Bhattacharya et al., 2020; Tchetgen Tchetgen et al., 2021; Ogburn et al., 2024), but have the benefit of allowing for certain patterns of latent confounding.

Besides implying different mechanisms, we show in Section 5 that, depending on the exact target of inference, the presence of undirected or bidirected edges in different layers[5] of the SG may also necessitate the use of different identification and estimation strategies that must be used for computing network causal effects, especially under full interference.

We end this section by noting an important limitation of our assumptions—by using an SG representation, we apriori rule out the simultaneous existence of contagion and latent confounding in the same layer of $\mathcal{G}$. While this is certainly possible in the real world, smooth globally identified parameterizations of such models have not been proposed yet, and indeed may not be possible in certain cases, as has been shown for linear Gaussian models for graphs that contain "bows," i.e., pairs of vertices such that both $V_i \to V_j$ and $V_i \leftrightarrow V_j$ exist (Drton et al., 2009; Shpitser, 2015).

3. Motivating Example

Before presenting technical details of our method, we present a hypothetical example inspired by the Networks, Norms and HIV/STI Risk Among Youth (NNAHRAY) study (Khan et al., 2009; Friedman et al., 2008), of how it might be used. Suppose we aim to examine the effect of incarceration on HIV risk. Let L_i represent an individual's past intravenous drug use (note L_i could also be a vector of baseline confounders, but we consider it to be a single variable in this example for simplicity), A_i represent past incarceration status, and Y_i denote HIV status. Connections in the network $\mathcal{F}$ on N individuals in the study population correspond to sexual or drug-use partnership, i.e. $\textbf{i} - \textbf{j}$

5. We use the term layer to refer to a collection of variables which could be L (the covariate layer), A (the treatment layer), or Y (the outcome layer).

exists in $\mathcal{F}$ if the two individuals are such partners. The underlying unknown segregated graph $\mathcal{G}$ is assumed to be one that can be represented by the pattern in Figure 1(b).

A public health policy maker might ask the following questions: (i) Does intravenous drug use in the population spread primarily due to social contagion, or is dependence due to latent factors such as household incomes? (ii) Does incarceration of an individual cause an increased likelihood of incarceration for their partners, or is this dependence also primarily explained by latent factors? Finally, (iii) Is the disease itself (HIV) contagious? The answers to each of these questions can lead to very different policy on containing future infections.

Our tests proposed in Section 4 can be used to distinguish between such mechanisms and the methods proposed in Section 5 can be used to estimate $\mathbb{E}[Y_i \mid \mathrm{do}(a)]$ and other network effects of interest, such as spillover effects of incarceration of partners. In the example above, our method would confirm background knowledge for (iii) that HIV is in fact a contagious disease that can spread via sexual contact and sharing needles. The answers to (i) and (ii) that pertain more to social science are less well established, and our test helps answer these from observational data. Moreover, even if the mechanisms can be fully determined through background knowledge, if latent confounding is present in the L or Y layer, our estimators that extend auto-g-computation are still required for unbiased estimation of the true network effects.

4. Likelihood Ratio Tests for Determining the SG

In this section we present likelihood ratio tests that can be used to distinguish between contagion and latent confounding in each layer of a partially determined SG $\mathcal{G}_{\circ\!-\!\circ}$ as shown in Figure 1(b). In order to design valid tests for this purpose, the first step is to to find an independence restriction that holds under contagion, but not under latent confounding (or vice versa). However, due to the dependent nature of our data, this alone is not sufficient to design a valid test with standard asymptotic properties. One popular approach to obtaining standard asymptotics with dependent data is to use *coding likelihood* estimators (Besag, 1974; Tchetgen Tchetgen et al., 2021)—estimators where the likelihood factorizes based on selecting conditionally independent samples from the network. We pursue this coding likelihood approach in our work, both for the design of likelihood ratio tests in this section, and for downstream estimation of network effects in the next section.

In the following, we consider contagion to be our null hypothesis and latent confounding to be the alternative. We follow this convention as contagion is often the assumed mechanism in prior work (Ogburn et al., 2020; Tchetgen Tchetgen et al., 2021; Bhattacharya et al., 2020). We design three coding likelihood ratio tests to determine the mechanism in each layer L, A, and Y of a partially determined SG $\mathcal{G}_{\circ\!-\!\circ}$. Ideally, to prevent propagation of errors, the results of each test should be independent of each other (i.e., the results of one test should not depend on the output of others). In order to design such tests, we first propose conditional independence restrictions for each layer that hold under the null but not the alternative, regardless of the mechanism in other layers. In the following, we will use $\mathcal{F}_i^{(k)}$ as short-hand for all units j that are exactly k degrees of separation away from the unit i in $\mathcal{F}$, i.e., the shortest path from i to j in $\mathcal{F}$ consists of k edges. By convention, $\mathcal{F}_i^{(0)}$ corresponds to i itself. We will also use $X_{\mathcal{F}_i^{(k,k',\ldots)}}$ as shorthand for the X variables of all units j that are $k, k', \ldots$ degrees of separation away from unit i. That is,

$$X_{\mathcal{F}_i^{(k,k',\ldots)}} = \{X_j \mid j \in \mathcal{F}_i^{(k)} \cup \mathcal{F}_i^{(k')} \cup \ldots\}.$$

Lemma 1 *Given a friendship network $\mathcal{F}$ and a corresponding partially determined SG $\mathcal{G}_{\circ\!-\!\circ}$, the following independences hold in each layer under the null hypothesis of contagion but not under the alternative hypothesis of latent confounding, regardless of the mechanism present in other layers. For any unit i in $\mathcal{F}$ for which $\mathcal{F}_i^{(1)}$ and $\mathcal{F}_i^{(2)}$ are not empty, we have,*

$$L_i \perp\!\!\!\perp L_{\mathcal{F}_i^{(2)}} \mid L_{\mathcal{F}_i^{(1)}} \tag{1}$$

$$A_i \perp\!\!\!\perp A_{\mathcal{F}_i^{(2)}} \mid A_{\mathcal{F}_i^{(1)}}, L_{\mathcal{F}_i^{(0,1,2,3)}} \tag{2}$$

$$Y_i \perp\!\!\!\perp Y_{\mathcal{F}_i^{(2)}} \mid Y_{\mathcal{F}_i^{(1)}}, A_{\mathcal{F}_i^{(0,1,2,3)}}, L_{\mathcal{F}_i^{(0,1,2,3)}}. \tag{3}$$

All proofs can be found in the Appendix. Intuitively, the distinct patterns of independence arise due to the presence/absence of colliders depending on the kind of edge that is present in each layer.

As an example, a conditional independence statement that holds under the null in the A layer for 🧍_4 in the friendship network $\mathcal{F}$ in Figure 1(a) and its corresponding SG is,[6]

$$A_4 \perp\!\!\!\perp A_{\mathcal{F}_4^{(2)}} \mid A_{\mathcal{F}_4^{(1)}}, L_{\mathcal{F}_4^{(0,1,2,3)}} \implies A_4 \perp\!\!\!\perp \{A_1, A_6\} \mid \{A_2, A_5\}, \{L_4, L_1, L_2, L_5, L_6, L_7, L_8\}.$$

Lemma 1 suggests we can design a likelihood ratio test for each layer by proposing nested models parameterized by a set of parameters β and γ respectively, where the model parameterized by β encodes a set of restrictions in Lemma 1 (depending on the layer being tested) and the model parameterized by γ is a strict supermodel that does not. Under full interference, however, proposing such models for the conditional densities of each layer—$p(L), p(A \mid L)$, and $p(Y \mid A, L)$—is infeasible without some dimension reduction assumptions. Here, we will assume parameters for the null and alternative models for each layer do not grow as a function of the size of the network and are shared across all units in the network. For example, in the case of binary treatments in the A layer, we may propose the following nested parametric models,

$$p(A_i = 1 \mid A_{\mathcal{F}_i^{(1)}}, L_{\mathcal{F}_i^{(0,1,2,3)}}; \beta) = \text{expit}(\beta_0 + \beta_1 \sum_{k \in \mathcal{F}_i^{(1)}} A_k + \beta_1 L_i + \beta_2 \sum_{k \in \mathcal{F}_i^{(1,2,3)}} L_k), \tag{4}$$

$$p(A_i = 1 \mid A_{\mathcal{F}_i^{(1,2)}}, L_{\mathcal{F}_i^{(0,1,2,3)}}; \gamma) = \text{expit}(\gamma_0 + \gamma_1 \sum_{k \in \mathcal{F}_i^{(1)}} A_k + \gamma_1 L_i + \gamma_2 \sum_{k \in \mathcal{F}_i^{(1,2,3)}} L_k + \gamma_3 \sum_{k \in \mathcal{F}_i^{(2)}} A_k), \tag{5}$$

where the parameters β and γ are shared across all units i in the network. We note that other parametric forms for these models are possible, e.g., having separate parameters for modeling dependence of A_i on L_k for $k \in \mathcal{F}_i^{(1)}, \mathcal{F}_i^{(2)}$, and $\mathcal{F}_i^{(3)}$ (we use this more flexible model in our experiments), as long as the number of parameters are not a function of the size of the network N and the null and alternatives are nested hypotheses as in (4) and (5).

We now show that the parameters can be consistently estimated using a coding likelihood that factorizes based on selecting units that are at least six degrees of separation away from each other in $\mathcal{F}$. We call such a set a 6-degree separated set and denote it as $\mathcal{S}^6(\mathcal{F})$, i.e., every distinct $\text{🧍}_i, \text{🧍}_j \in \mathcal{S}^6(\mathcal{F})$ are at least six degrees of separation apart. Later, we will also use $\mathcal{S}^k(\mathcal{F})$ to denote k-degree separated sets.[7] Below we present factorizations of coding likelihoods for the null and alternative models and show they hold regardless of the true underlying mechanism—this is important as we want standard asymptotics for estimation of the null even if the alternative is true and vice versa.

6. A simpler conditional independence $A_i \perp\!\!\!\perp A_{\mathcal{F}_i^{(2)}} \mid A_{\mathcal{F}_i^{(1)}}, L_{\mathcal{F}_i^{(0,1)}}$ also holds under the null but not the alternative. However, such independences cannot be used to factorize a coding likelihood, as seen in the proof for Lemma 2.

7. As concrete examples, $\{\text{🧍}_1, \text{🧍}_9\}$ is an $\mathcal{S}^6(\mathcal{F})$ set for Figure 1(a), while $\{\text{🧍}_1, \text{🧍}_3, \text{🧍}_4, \text{🧍}_6, \text{🧍}_{10}\}$ is an $\mathcal{S}^2(\mathcal{F})$ set.

Lemma 2 *Given data (L, A, Y) drawn from a distribution that is Markov and faithful wrt to a partially determined SG $\mathcal{G}_{\circ\!-\!\circ}$, the coding likelihood functions for the null and alternative models factorize as follows regardless of the true underlying mechanisms in each layer of $\mathcal{G}_{\circ\!-\!\circ}$,*

$$\mathcal{CL}(\beta_L) = \prod_{i \in \mathcal{S}^6(\mathcal{F})} p(L_i \mid L_{\mathcal{F}_i^{(1)}}; \beta_L) \qquad \mathcal{CL}(\gamma_L) = \prod_{i \in \mathcal{S}^6(\mathcal{F})} p(L_i \mid \cdot, L_{\mathcal{F}_i^{(2)}}; \gamma_L) \tag{6}$$

$$\mathcal{CL}(\beta_A) = \prod_{i \in \mathcal{S}^6(\mathcal{F})} p(A_i \mid A_{\mathcal{F}_i^{(1)}}, L_{\mathcal{F}_i^{(0,1,2,3)}}; \beta_A) \qquad \mathcal{CL}(\gamma_A) = \prod_{i \in \mathcal{S}^6(\mathcal{F})} p(A_i \mid \cdot, A_{\mathcal{F}_i^{(2)}}; \gamma_A) \tag{7}$$

$$\mathcal{CL}(\beta_Y) = \prod_{i \in \mathcal{S}^6(\mathcal{F})} p(Y_i \mid Y_{\mathcal{F}_i^{(1)}}, \{A \cup L\}_{\mathcal{F}_i^{(0,1,2,3)}}; \beta_Y) \qquad \mathcal{CL}(\gamma_Y) = \prod_{i \in \mathcal{S}^6(\mathcal{F})} p(Y_i \mid \cdot, Y_{\mathcal{F}_i^{(2)}}; \gamma_Y) \tag{8}$$

where $(\cdot,)$ in each case above denotes conditioning on the same set in the null.

Intuition for the factorizations is as follows: for any i, we condition on a set of variables in $\mathcal{F}_i^{(0,1,2,3)}$. Thus, the rest of the variables appearing in the coding likelihood functions belong to units that are at least 4 degrees of separation away from i (by the antecedent that all units in $\mathcal{S}^6(\mathcal{F})$ are at least 6 degrees separated from each other). Thus, factorization of, for e.g., $\mathcal{CL}(\beta_Y)$ is established by showing that $Y_i \perp\!\!\!\perp L_j, A_j, Y_j \mid \cdot$ for all j that are at least 4 degrees of separation from i. Having established factorization of the coding likelihood functions, we can now rely on standard maximum likelihood asymptotic theory (similar to Besag (1974)) for constructing coding likelihood ratio tests for each layer, as long as the number of units in $\mathcal{S}^6(\mathcal{F})$ grows with the size of the network. This leads us to a statement about our assumptions on network asymptotics for designing our tests.

Network asymptotics for testing Let $\mathcal{G}_{\circ\!-\!\circ\, N}$ be a sequence of partially determined SGs with associated friendship networks $\mathcal{F}_N$ for $N = 1, 2, \ldots$ such that as $N \to \infty$, the size of at least one 6-degree separated set obtained from $\mathcal{F}_N$ also tends to infinity, i.e. $|\mathcal{S}(\mathcal{F}_N)| \to \infty$.

Under the above network asymptotics and mild regularity conditions stated in the Appendices of Tchetgen Tchetgen et al. (2021), the estimates for the unknown parameters of correctly specified null and alternative models obtained by maximizing the coding likelihood functions in Lemma 2 are consistent and asymptotically normal. In Algorithm 1 we then propose our final coding likelihood ratio tests for determining the mechanism in each layer of a partially determined SG $\mathcal{G}_{\circ\!-\!\circ}$.

Based on Lemma 1 and Lemma 2 and the consistency of the coding likelihood ratio estimates for the parameters of correctly specified null and alternative models, the test for each layer possess the desired asymptotic properties: type-I error control at a specified significance level α and power tending to 1 as $N \to \infty$. This is also confirmed empirically with our experiments in Section 6.

Note that many possible $\mathcal{S}^6(\mathcal{F})$ sets of potentially different sizes can be obtained from a single network $\mathcal{F}$. Ideally, we would want to find the largest such set for estimating our parameters. However, finding the maximum sized k-degree separated set is well-known to be NP-hard (Robson, 1986; Myrvold and Fowler, 2013; Eto et al., 2014). Here, we opt for a greedy approach to finding maximal k-degree separated sets and find that this works well in practice. We also note a well-known tradeoff of coding likelihood estimators is that asymptotic guarantees come at the cost of sample efficiency (Besag, 1974). The network asymptotics for our tests are particularly interesting in this regard—the small-world principle is a popular hypothesis that in real-world networks every pair of units is no more than 6 degrees of separation apart (Watts and Strogatz, 1998). Fortunately, our tests make use of units in $\mathcal{S}^6(\mathcal{F})$ that exactly meet this upper bound, presenting an interesting connection between the connectedness of networks and the testability of mechanisms.

Algorithm 1 for determining edge types in $\mathcal{G}_{\circ\!\!-\!\!\circ}$

Inputs: Partially determined SG $\mathcal{G}_{\circ\!\!-\!\!\circ}$; Network $\mathcal{F}$; Data L, A, Y; Significance level α.
Obtain a maximal $\mathcal{S}^6(\mathcal{F})$ such that for each i in the set, $\mathcal{F}_i^{(1)}$ and $\mathcal{F}_i^{(2)}$ are not empty
for each type of variable X in "L", "A", "Y" **do**
    $\widehat{\beta}_X \leftarrow \text{argmax}_{\beta_X} \mathcal{CL}(\beta_X)$ and $\widehat{\gamma}_X \leftarrow \text{argmax}_{\gamma_X} \mathcal{CL}(\gamma_X)$
    $\Lambda \leftarrow -2(\log \mathcal{CL}(\widehat{\beta}_X) - \log \mathcal{CL}(\widehat{\gamma}_X))$ and $k \leftarrow |\beta_X| - |\gamma_X|$
    if $P\left(\chi_k^2 \geq \Lambda\right) < \alpha)$ **then** replace each $X_i \circ\!\!-\!\!\circ X_j$ in $\mathcal{G}_{\circ\!\!-\!\!\circ}$ with $X_i \leftrightarrow X_j$
    else replace each $X_i \circ\!\!-\!\!\circ X_j$ in $\mathcal{G}_{\circ\!\!-\!\!\circ}$ with $X_i - X_j$
return the now fully determined $\mathcal{G}_{\circ\!\!-\!\!\circ}$

5. Identification and Estimation of Network Effects

The identification and estimation methods presented in this section assume a known SG $\mathcal{G}$, or one that is correctly inferred using Algorithm 1. We start with an identification result for the target.

Theorem 3 *Given a hidden variable CG $\mathcal{G}(V \cup H)$ whose observed margin can be represented by any SG $\mathcal{G}(V)$ that is compatible with the assumptions of Figure 1(b), the target parameter is identified as,* $\mathbb{E}[Y_i \mid \text{do}(A = a)] = \sum_{L,Y} p(L) \times p(Y \mid A = a, L) \times Y_i$.

That is, network ignorability (Tchetgen Tchetgen et al., 2021) is satisfied regardless of mechanisms present at each layer. While the identifying functional for our target parameter does not depend on mechanisms, we see in Section 5.1 that mechanistic knowledge, particularly in the L and Y layer, is still essential for estimation purposes. For completeness, we also provide an example of a different target parameter where the identifying functional can change depending on mechanisms.

Consider the unit level potential outcome where we intervene on the outcomes of all other units except i, i.e., $\mathbb{E}[Y_i \mid \text{do}(Y_{-i} = y_{-i})]$. When contagion is present in the Y layer, the identifying functional for this query is, $\sum_{L,A,Y_i} p(L, A) \times p(Y_i \mid Y_{-i} = y_{-i}, A, L) \times Y_i$; when latent confounding is present in the Y layer, the identifiying functional is simply $\sum_{Y_i} p(Y_i) \times Y_i$. These functionals align with the interpretation of undirected and bidirected edges—interventions on Y_{-i} may have an effect on Y_i under contagion but not latent confounding. The proofs are in Appendix F.

5.1. Extending auto-g-computation to account for latent confounding

We now present a new method for network effect estimation that extends the auto-g-computation method of Tchetgen Tchetgen et al. (2021) to account for latent confounding. We first provide a brief description of auto-g-computation before describing our modifications. We return to our target of inference, identified in Theorem 3, which depends on densities $p(L)$ and $p(Y \mid A, L)$.

Auto-g-computation overcomes the challenge of having access to only a single realization of the network under full interference by drawing a number of independent samples of L and Y from the densities $p(L)$ and $p(Y \mid A, L)$. The method proposes a Gibbs sampler that draws independent realizations of L and Y from these densities given access to just univariate conditionals $p(L_i \mid L_{-i})$ and $p(Y_i \mid Y_{-i}, A, L)$ for each i in the network. To make estimation of these conditionals feasible from a single realization, Tchetgen Tchetgen et al. (2021) assumes that the L and Y layer are connected by undirected edges, which simplifies these conditionals to depend only on adjacent units

Algorithm 2 for estimating $\mathbb{E}[Y_i \mid \text{do}(A = a)]$ for every i in the network

1: **Inputs:** SG $\mathcal{G}$; Friendship network $\mathcal{F}$; Data L, A, Y; Treatment assignment vector a.
2: Obtain a maximal $\mathcal{S}^2(\mathcal{F})$ set
3: **if** each $L_i \circ\!\!-\!\!\circ L_j$ in $\mathcal{G}$ is undirected **then**
4: $\quad \widehat{\theta}_{L_i - L_j} \leftarrow \text{argmax}_{\theta_{L_i - L_j}} \mathcal{CL}(\theta_{L_i - L_j})$ and $L_{\text{draws}} \leftarrow$ Gibbs sampler $L\big(\mathcal{G}; \widehat{\theta}_{L_i - L_j}\big)$
5: **else**
6: $\quad$ Obtain a maximal $\widetilde{\mathcal{S}}^2(\mathcal{F})$ set of isomorphic local structures
7: $\quad \widehat{\theta}_{L_i \leftrightarrow L_j} \leftarrow \text{argmax}_{\theta_{L_i \leftrightarrow L_j}} \mathcal{CL}(\theta_{L_i \leftrightarrow L_j})$ and $L_{\text{draws}} \leftarrow$ draws from $p(L; \widehat{\theta}_{L_i \leftrightarrow L_j})$
8: **if** each $Y_i \circ\!\!-\!\!\circ Y_j$ in $\mathcal{G}$ is undirected **then**
9: $\quad \widehat{\theta}_{Y_i - Y_j} \leftarrow \text{argmax}_{\theta_{Y_i - Y_j}} \mathcal{CL}(\theta_{Y_i - Y_j})$ and $Y_{\text{draws}} \leftarrow$ Gibbs sampler $Y\big(\mathcal{G}; \widehat{\theta}_{Y_i - Y_j}; L_{\text{draws}}, a\big)$
10: **else**
11: $\quad$ Fit an outcome regression model $\mathbb{E}[Y_i \mid A_{\mathcal{F}_i^{(0,1)}}, L_{\mathcal{F}_i^{(0,1)}}; \theta_{Y_i \leftrightarrow Y_j}]$ using $\mathcal{CL}(\theta_{Y_i \leftrightarrow Y_j})$
12: $\quad$ // Below $L^{(m)}$ denotes the m^{th} row of L_{draws}
13: $\quad Y_{\text{draws}} \leftarrow M \times N$ matrix where entry $m, i = \mathbb{E}[Y_i \mid a, L^{(m)}; \widehat{\theta}_{Y_i \leftrightarrow Y_j}]$
14: **return** $\widehat{\mathbb{E}}[Y_i|\text{do}(a)], \ldots, \widehat{\mathbb{E}}[Y_N|\text{do}(a)] \leftarrow$ empirical averages of outcomes for each i in Y_{draws}

in the network, i.e. $p(L_i \mid L_{\mathcal{F}_i^{(1)}})$ and $p(Y_i \mid Y_{\mathcal{F}_i^{(1)}}, A_{\mathcal{F}_i^{(0,1)}}, L_{\mathcal{F}_i^{(0,1)}})$ respectively. Finally, similar to our assumptions in Section 4, auto-g-computation assumes that the parameters $\theta_{L_i - L_j}$ and $\theta_{Y_i - Y_j}$ used to parameterize these univariate conditional distributions are shared across the network.

Thus, to estimate $\mathbb{E}[Y_i \mid \text{do}(a)]$ for each i, auto-g-computation proceeds by (i) Estimating parameters of the univariate conditionals via maximizing coding likelihoods; (ii) Using Gibbs sampling to draw several, say M, independent realizations of L and Y from $p(L)$ and $p(Y \mid A = a, L)$, where each realization from the sampler is an $N \times 1$ vector corresponding to values of L and Y for every unit in the network under a treatment assignment $A = a$; and (iii) Taking the empirical average of each Y_i from these M draws as the final estimate for each $\mathbb{E}[Y_i \mid \text{do}(a)]$.

While auto-g works well when edges among L and Y are undirected, it leads to biased estimates if either layer contains bidirected edges. This is because the simplification of the univariate conditionals used for Gibbs sampling no longer holds due to conditioning on colliders. By a simple s-separation argument, $p(L_i \mid L_{-i}) \neq p(L_i \mid L_{\mathcal{F}^{(1)}})$ when the L layer contains bidirected edges and $p(Y_i \mid Y_{-i}, A, L) \neq p(Y_i \mid Y_{\mathcal{F}_i^{(1)}}, A_{\mathcal{F}_i^{(0,1)}}, L_{\mathcal{F}_i^{(0,1)}})$ when the the Y layer contains bidirected edges. We propose modifications to auto-g-computation to account for these cases. Our full procedure is described in Algorithm 2; the Gibbs samplers are described in Appendix H. If the else statements in our algorithm do not execute, then there are undirected edges in both the L and Y layer, and our procedure reduces to the original auto-g-computation method ($\mathcal{CL}(\theta_{L_i - L_j})$ and $\mathcal{CL}(\theta_{Y_i - Y_j})$ in lines 4 and 9 are coding likelihood functions used to compute parameter estimates when the edges in both layers are undirected). Thus, we focus on explaining the else statements in Algorithm 2.

Define a *local structure* S as a frequently occurring isomorphic subgraph of $\mathcal{F}$ (i.e., a subgraph with the same structure if the labels of the nodes are ignored). Examples of local structures include dyads and triangles (cliques consisting of 3 units, e.g., $2, $4, $5 in Figure 1(a)). For a chosen granularity of local structures, we then propose a parameterization of local structures $p(L_S; \theta_{L_i \leftrightarrow L_j})$ that is shared across the network. For example, under a multivariate normal assumption on $p(L)$ where we parameterize dyad local structures, we would propose $\theta_{L_i \leftrightarrow L_j}$ to be the parameter vector

$(\mu_i, \sigma_{ii}, \sigma_{ij})$ corresponding to the mean, variance, and covariance shared across all units (this encodes marginal independences between non-adjacent units, as the covariance between non-adjacent units is 0 in the model definition). In the case of discrete data, one could also use log-linear parameterizations described in Rudas et al. (2010); Forcina et al. (2010); Evans and Richardson (2013).[8]

Consistently estimating the parameters $p(L_S; \theta_{L_i \leftrightarrow L_j})$ allows us to draw independent realizations of L from a distribution $p(L)$ that is Markov wrt $\mathcal{G}_L(V)$ when the L layer has bidirected edges. This is done in line 7 of Algorithm 2 after obtaining maximum coding likelihood estimates of the parameters. The coding likelihood $\mathcal{CL}(\theta_{L_i \leftrightarrow L_j})$ in line 7 is defined over a collection of local structures $\widetilde{\mathcal{S}}^2(\mathcal{F})$ such that for any $S, S' \in \widetilde{\mathcal{S}}^2(\mathcal{F})$ and for every pair of units, $i \in S$ and $j \in S'$, i and j are at least 2 degrees of separation away from each other.[9]

After line 7 of Algorithm 2, we have M independent draws of L from either an undirected or bidirected model of $p(L)$. The remaining steps do not depend on which one we drew from. We now focus on how the algorithm finishes effect estimation when the Y layer has bidirected edges. Notice that the functional in Theorem 3 can be further simplified to $\sum_L p(L) \times \mathbb{E}[Y_i \mid A = a, L]$. Given M independent draws of L, the only task that remains then is to estimate the outcome regression $\mathbb{E}[Y_i \mid A, L]$ from which we can get M predictions Y_i given each draw of L and $A = a$ and average these predictions to obtain an estimate of $\mathbb{E}[Y_i \mid \text{do}(a)]$. When the Y layer contains bidirected edges, we can easily obtain maximum coding likelihood estimates for this outcome regression model using a function $\mathcal{CL}(\theta_{Y_i \leftrightarrow Y_j})$ defined on units in $\mathcal{S}^2(\mathcal{F})$. The estimation and prediction steps are executed in lines 11 and 13. For completeness, we present the factorizations of all coding likelihood functions in Algorithm 2 in the following Lemma. The difference in factorizations below emphasize the need to know the underlying mechanisms in the L and Y layer, but not necessarily the A layer.

Lemma 4 *Given data (L, A, Y) drawn from a distribution that is Markov wrt to an SG $\mathcal{G}$ that is compatible with the assumptions in Figure 1(b), the coding likelihood function for modeling $p(L)$ and $p(Y \mid A, L)$ factorize depending on the edges present in the L and Y layers as,*

$$\mathcal{CL}(\theta_{L_i - L_j}) = \prod_{i \in \mathcal{S}^2(\mathcal{F})} p(L_i \mid L_{\mathcal{F}_i^{(1)}}; \theta_{L_i - L_j}) \qquad \mathcal{CL}(\theta_{Y_i - Y_j}) = \prod_{i \in \mathcal{S}^2(\mathcal{F})} p(Y_i \mid Y_{\mathcal{F}_i^{(1)}}, \{A \cup L\}_{\mathcal{F}_i(0,1)}; \theta_{Y_i - Y_j})$$

$$\mathcal{CL}(\theta_{L_i \leftrightarrow L_j}) = \prod_{S \in \widetilde{\mathcal{S}}^2(\mathcal{F})} p(L_S; \theta_{L_i \leftrightarrow L_j}) \qquad \mathcal{CL}(\theta_{Y_i \leftrightarrow Y_j}) = \prod_{i \in \mathcal{S}^2(\mathcal{F})} p(Y_i \mid \{A \cup L\}_{\mathcal{F}_i(0,1)}; \theta_{Y_i \leftrightarrow Y_j})$$

Maximizing the above functions gives us consistent and asymptotically normal estimates of the corresponding parameters, and thus the target parameters through Algorithm 2, as long as we have the following asymptotic growth of the network (which could be considered weaker than the one proposed for our tests, since it only relies on 2-degree separated sets and local structures).

Network asymptotics for estimation Let $\mathcal{G}_N$ be a sequence of SGs compatible with Figure 1(b) and with associated friendship networks $\mathcal{F}_N$ for $N = 1, 2, \ldots \infty$ such that as $N \to \infty$, the sizes of at least one 2-degree separated set of individuals and one 2-degree separated set of isomorphic local structures obtained from $\mathcal{F}_N$ also tend to infinity, i.e. $|\mathcal{S}^2(\mathcal{F}_N)| \to \infty$ and $|\widetilde{\mathcal{S}}^2(\mathcal{F}_N)| \to \infty$.

8. As the size of the local structure increases, so does the size of the parameter vector, corresponding to more flexible modeling assumptions. In our experiments, we use the simplest possible local structure of dyads, however, we only require that the size of the parameter vector does not grow as a function of the size of the network.
9. An example of $\widetilde{\mathcal{S}}^2(\mathcal{F})$ in Figure 1(a) when considering dyads as local structures is $\{1 - 2, 6 - 7, 9 - 10\}$.

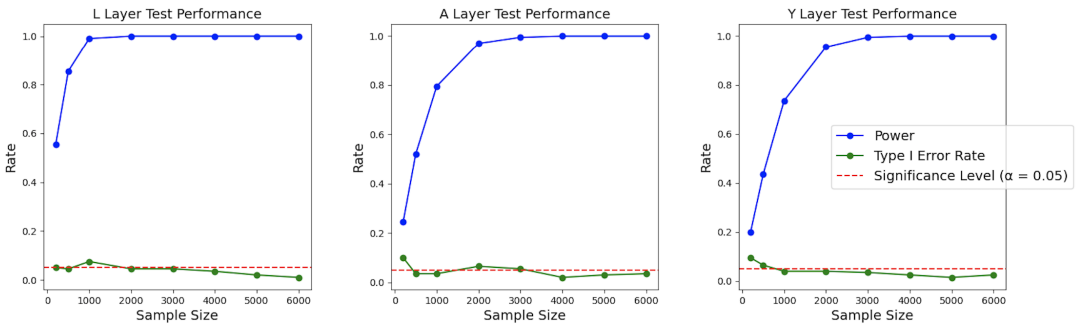

Figure 2: Performance of coding likelihood ratio tests with varying sample sizes.

6. Experiments, Discussion, and Conclusions

We performed experiments to evaluate the correctness of the proposed methods using both synthetic data and semi-synthetic data with known ground truths[10]. Additionally, we tested the validity of our key network assumption—that sufficiently large sets $\mathcal{S}^6(\mathcal{F})$ can be found—using various real-world social networks. Code for our experiments is available in this GitHub repository.

6.1. Evaluation of Likelihood Ratio Tests

We first evaluate our likelihood ratio tests in Algorithm 1 using effective sample sizes of $200, 500,$ $1000, 2000, 3000,$ $4000,$ and 5000. These numbers represent sizes of a maximal $\mathcal{S}^6(\mathcal{F})$ obtained from a synthetic network of $200,000$ units, where each unit has 1 to 6 randomly assigned neighbors. We create two configurations of SGs compatible with the partially determined SG in Figure 1(b), one with undirected edges in all layers and another with bidirected edges in all layers; details of the underlying data generating processes are in Appendix I.

We run 200 trials at each effective sample size to calculate power and type-I error rate. Figure 2 shows the results. The results show that our tests are well calibrated, with power approaching 1 quite quickly as the size of $\mathcal{S}^6(\mathcal{F})$ increases and type-I error controlled at the desired level of $\alpha = 0.05$.

Since real-world networks can be smaller and denser, we also evaluate whether sufficiently large $\mathcal{S}^6(\mathcal{F})$ can be retrieved in practice by analyzing 10 real networks from the Stanford Large Network Dataset Collection (SNAP) (Leskovec and Krevl, 2014). In each network, we identify 100 maximal 6-degree separated sets, discard units for whom $\mathcal{F}_i{}^{(1)}$ and $\mathcal{F}_i{}^{(2)}$ are empty, as our tests rely on data from these neighborhoods, and retain the largest maximal set, denoted as $|\mathcal{S}^6(\mathcal{F})|^*$. Table 1 shows that node usage varies significantly with network density. In denser networks like Twitch Gamers, we could use only 0.02% of units as effective samples for our tests, but in sparser networks such as Deezer RO, we could use 2.85% of the units. These findings indicate that our tests perform best in networks that are both large and relatively sparse. There is also some evidence that virtual social networks are more dense than non-virtual ones (Bakhshandeh et al., 2011; Cheng, 2010; Edunov et al., 2016) due to the ease of forming connections on the internet, so our method may have more samples in non-virtual settings (such as our motivating example).

10. While it would be ideal to evaluate our methods with real-world data, accompanying ground-truths are typically not available in the absence of a gold standard randomized trial (see Keith et al. (2023) for discussion). Finding publicly available network trial data, however, is challenging, and we were unable to find a suitable one for our experiments.

Dataset	# of Nodes	# of Edges	$\lvert\mathcal{S}^6(\mathcal{F})\rvert^*$	Node Usage
Social Circles: FB (Leskovec and Mcauley, 2012)	4,039	88,234	3	0.07%
GitHub Social Network (Rozemberczki et al., 2019a)	37,700	289,003	211	0.56%
Deezer Europe (Rozemberczki and Sarkar, 2020)	28,281	92,752	476	1.68%
Deezer HR (Rozemberczki et al., 2019b)	54,573	498,202	327	0.60%
Deezer HU (Rozemberczki et al., 2019b)	47,538	222,887	488	1.03%
Deezer RO (Rozemberczki et al., 2019b)	41,773	125,826	1,192	2.85%
GEMSEC FB Artists (Rozemberczki et al., 2019b)	50,515	819,306	206	0.41%
GEMSEC FB Atheletes (Rozemberczki et al., 2019b)	13,866	86,858	96	0.69%
LastFM Asia (Rozemberczki and Sarkar, 2020)	7,624	27,806	159	2.09%
Twitch Gamers (Rozemberczki and Sarkar, 2021)	168,114	6,797,557	29	0.02%

Table 1: Maximal $\mathcal{S}^6(\mathcal{F})$ sets in SNAP networks. Node Usage $= \lvert\mathcal{S}^6(\mathcal{F})\rvert^*/$# of Nodes.

		L Layer		A Layer		Y Layer	
Network	Average $\lvert\mathcal{S}^6(\mathcal{F})\rvert^*$	Type I	Power	Type I	Power	Type I	Power
LastFM Asia (Rozemberczki and Sarkar, 2020)	139.4	0.056	0.416	0.058	0.198	0.038	0.124
Deezer HR (Rozemberczki et al., 2019b)	297.7	0.046	0.490	0.028	0.358	0.052	0.044
Deezer Europe (Rozemberczki and Sarkar, 2020)	451.1	0.044	0.392	0.046	0.324	0.056	0.416
Deezer HU (Rozemberczki et al., 2019b)	464.6	0.058	0.878	0.058	0.398	0.058	0.652
Deezer RO (Rozemberczki et al., 2019b)	1147.0	0.050	0.998	0.030	0.916	0.038	0.872

Table 2: Performance of Likelihood Ratio Tests on semi-synthetic data. Average $\lvert\mathcal{S}^6(\mathcal{F})\rvert^*$ is simply the mean of effective sample sizes across. Type I represents type I error rate.

To evaluate the effectiveness of our likelihood ratio tests on a variety of network topologies, we also conducted semi-synthetic experiments using five real-world networks from Table 1. For each network, we generate synthetic data on L, A, and Y using the same rules as in the synthetic version, and we run Algorithm 1 500 times. The results are summarized in Table 2.

We make the following observations: 1) Type I error rate is well controlled at the desired significance level ($\alpha = 0.05$) across all networks; 2) For a given layer, the power of our test is generally higher for networks where we are able to obtain larger effective samples; and 3) For a given network, the power of our test is generally the highest at the L layer. Distinguishing latent confounding from contagion is slightly more challenging for the A layer and the most challenging for the Y layer.

These results are consistent with the synthetic experiment results in Figure 2, showing that our likelihood ratio tests work well on a variety of network topologies. Our synthetic experiments generate networks where each unit has 1 to 6 randomly assigned neighbors. In real-world networks, however, the maximal degree and average degree are much higher. For example, in Deezer HR, the maximum and average degrees are 420 and 18.26; in LastFM Asia, these numbers are 216 and 7.29.

6.2. Evaluation of Causal Effect Estimation Method

There are a total of eight possible SGs compatible with the partially determined SG $\mathcal{G}_{\circ\!-\!\circ}$ shown in Figure 1(b). For brevity, we denote these SGs using a three-letter code, where each letter corresponds to a layer and is either 'U' (for undirected) or 'B' (for bidirected). For example, BUB represents the case where the L layer is connected by $\leftrightarrow$, the A layer by $-$, and the Y layer by $\leftrightarrow$.

We do not evaluate the UUU case because this is already studied in prior work. For each of the other cases, we repeat the following 100 times: 1) Randomly generate friendship networks of different sizes ($500, 1000, 2000, 3000, 4000$, and 5000) where units have an average of 5 and at most

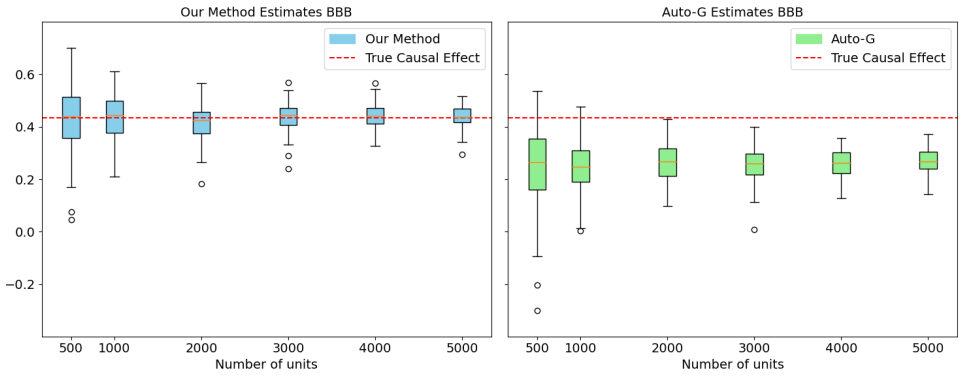

Figure 3: Network effect estimation results for our method and auto-g-computation.

10 neighbors, and 2) Using data generated from these networks, estimate the population average overall effect as the contrast $\mathbb{E}[Y \mid \text{do}(\mathbf{1})] - \mathbb{E}[Y \mid \text{do}(\mathbf{0})] = \frac{1}{N}\sum_{i=1}^{N}\big(\mathbb{E}[Y_i \mid \text{do}(\mathbf{1})] - \mathbb{E}[Y_i \mid \text{do}(\mathbf{0})]\big)$, where bold-faced $\mathbf{1}$ indicates a vector of treatment assignments where everyone in the network receives $A_i = 1$, and similarly for $\mathbf{0}$, using both Algorithm 2 and standard auto-g-computation.

Figure 3 illustrates the consistency of our proposed method in the BBB case. In contrast, the auto-g-computation yields biased results due to violations of its assumptions. Experimental results for other cases are deferred to the Appendix, as they are quite similar.

6.3. Discussion and Conclusion

In this work we developed hypothesis tests to differentiate between associations due to contagion and latent confounding in networks with full interference. We also extend auto-g-computation to settings where sets of variables may be dependent due to either contagion or latent confounding. We demonstrated our method's effectiveness using a mixture of synthetic and real network data.

Future work on this topic include exploring more sample efficient estimators (for example, those designed using pseudolikelihood functions Besag (1974)) and methods that are robust to uncertainty in the friendship network $\mathcal{F}$. An interesting future direction would also be to explore the feasibility of mechanism testing and network effect identification and estimation in models where contagion and latent confounding co-occur in the same layer. Sadeghi (2016) studies separation criteria for graphs depicting such co-occurrences, however, to our knowledge, it is currently unknown whether smooth globally identifiable models for this class of graphs is possible in general. If global identification is not possible, developing sensitivity analysis methods for the assumption that contagion and latent confounding cannot co-occur would be an interesting future direction.

Acknowledgments

RB acknowledges support from the NSF CRII grant 2348287. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

Reza Bakhshandeh, Mehdi Samadi, Zohreh Azimifar, and Jonathan Schaeffer. Degrees of separation in social networks. In *Proceedings of the International Symposium on Combinatorial Search*, volume 2, pages 18–23, 2011.

Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225, 1974.

Rohit Bhattacharya, Daniel Malinsky, and Ilya Shpitser. Causal inference under interference and network uncertainty. In *Uncertainty in Artificial Intelligence*, pages 1028–1038. PMLR, 2020.

Sohom Bhattacharya and Subhabrata Sen. Causal effect estimation under network interference with mean-field methods. *arXiv preprint arXiv:2407.19613*, 2024.

Alex Cheng. Sysomos — six degrees of separation, Twitter style. https://web.archive.org/web/20100502005723/http://sysomos.com/insidetwitter/sixdegrees/, 2010. [Accessed 13-02-2025].

Duncan A Clark and Mark S Handcock. Causal inference over stochastic networks. *Journal of the Royal Statistical Society Series A: Statistics in Society*, page qnae001, 2024.

Mathias Drton, Michael Eichler, and Thomas S Richardson. Computing maximum likelihood estimates in recursive linear models with correlated errors. *Journal of Machine Learning Research*, 10(10), 2009.

Sergey Edunov, Smriti Bhagat, Moira Burke, Carlos Diuk, and Ismail Onur Filiz. Three and a half degrees of separation - Meta Research — Meta Research — research.facebook.com. https://research.facebook.com/blog/2016/2/three-and-a-half-degrees-of-separation/, 2016. [Accessed 13-02-2025].

Hiroshi Eto, Fengrui Guo, and Eiji Miyano. Distance-d independent set problems for bipartite and chordal graphs. *Journal of Combinatorial Optimization*, 27(1):88–99, 2014.

Robin J Evans and Thomas S Richardson. Marginal log-linear parameters for graphical markov models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75(4):743–768, 2013.

Antonio Forcina, Monia Lupparelli, and Giovanni M Marchetti. Marginal parameterizations of discrete models defined by a set of conditional independencies. *Journal of Multivariate Analysis*, 101(10):2519–2527, 2010.

SR Friedman, M Bolyard, M Sandoval, P Mateu-Gelabert, C Maslow, and J Zenilman. Relative prevalence of different sexually transmitted infections in HIV-discordant sexual partnerships: Data from a risk network study in a high-risk New York neighbourhood. *Sexually Transmitted Infections*, 84(1):17–18, 2008.

Matthew O Jackson and Dunia López-Pintado. Diffusion and contagion in networks with heterogeneous agents and homophily. *Network Science*, 1(1):49–67, 2013.

Katherine A Keith, Sergey Feldman, David Jurgens, Jonathan Bragg, and Rohit Bhattacharya. RCT rejection sampling for causal estimation evaluation. *Transactions of Machine Learning Research*, 2023.

Maria R Khan, Irene A Doherty, Victor J Schoenbach, Eboni M Taylor, Matthew W Epperson, and Adaora A Adimora. Incarceration and high-risk sex partnerships among men in the united states. *Journal of Urban Health*, 86:584–601, 2009.

Steffen L Lauritzen. *Graphical Models*, volume 17. Clarendon Press, 1996.

Steffen L Lauritzen and Thomas S Richardson. Chain graph models and their causal interpretations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 64(3):321–348, 2002.

Youjin Lee and Elizabeth L Ogburn. Network dependence can lead to spurious associations and invalid inference. *Journal of the American Statistical Association*, 116(535):1060–1074, 2021.

Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data, June 2014.

Jure Leskovec and Julian Mcauley. Learning to discover social circles in ego networks. *Advances in Neural Information Processing Systems*, 25, 2012.

W Myrvold and PW Fowler. Fast enumeration of all independent sets up to isomorphism. *Preprint*, 2013.

Elizabeth L Ogburn. Challenges to estimating contagion effects from observational data. *Complex Spreading Phenomena in Social Systems*, page 47, 2018.

Elizabeth L Ogburn and Tyler J VanderWeele. Causal diagrams for interference. 2014.

Elizabeth L Ogburn, Ilya Shpitser, and Youjin Lee. Causal inference, social networks and chain graphs. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 183(4):1659–1676, 2020.

Elizabeth L Ogburn, Oleg Sofrygin, Ivan Diaz, and Mark J van der Laan. Causal inference for social network data. *Journal of the American Statistical Association*, 119(545):597–611, 2024.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Jose M Peña. Unifying Gaussian LWF and AMP chain graphs to model interference. *Journal of Causal Inference*, 8(1):1–21, 2020.

James M Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7 (9-12):1393–1512, 1986.

John Michael Robson. Algorithms for maximum independent sets. *Journal of Algorithms*, 7(3): 425–440, 1986.

Hannes Rosenbusch, Anthony M Evans, and Marcel Zeelenberg. Multilevel emotion transfer on YouTube: Disentangling the effects of emotional contagion and homophily on video audiences. *Social Psychological and Personality Science*, 10(8):1028–1035, 2019.

Benedek Rozemberczki and Rik Sarkar. Characteristic Functions on Graphs: Birds of a Feather, from Statistical Descriptors to Parametric Models. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20)*, page 1325–1334. ACM, 2020.

Benedek Rozemberczki and Rik Sarkar. Twitch gamers: a dataset for evaluating proximity preserving and structural role-based node embeddings, 2021.

Benedek Rozemberczki, Carl Allen, and Rik Sarkar. Multi-scale attributed node embedding, 2019a.

Benedek Rozemberczki, Ryan Davies, Rik Sarkar, and Charles Sutton. Gemsec: Graph embedding with self clustering. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2019*, pages 65–72. ACM, 2019b.

Tamás Rudas, Wicher P Bergsma, and Renáta Németh. Marginal log-linear parameterization of conditional independence models. *Biometrika*, 97(4):1006–1012, 2010.

Kayvan Sadeghi. Marginalization and conditioning for LWF chain graphs. *The Annals of Statistics*, pages 1792–1816, 2016.

Cosma R Shalizi and Andrew C Thomas. Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, 40(2):211–239, 2011.

Eli Sherman and Ilya Shpitser. Identification and estimation of causal effects from dependent data. *Advances in Neural Information Processing Systems*, 31, 2018.

Ilya Shpitser. Segregated graphs and marginals of chain graph models. *Advances in Neural Information Processing Systems*, 28, 2015.

Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT press, 2000.

Ranjani Srinivasan, Rohit Bhattacharya, Razieh Nabi, Elizabeth L Ogburn, and Ilya Shpitser. Graphical models of entangled missingness. *arXiv preprint arXiv:2304.01953*, 2023.

Milan Studený. On separation criterion and recovery algorithm for chain graphs. In *Proceedings of the Twelfth international conference on Uncertainty in Artificial Intelligence*, pages 509–516, 1996.

Eric J Tchetgen Tchetgen, Isabel R Fulcher, and Ilya Shpitser. Auto-g-computation of causal effects on a network. *Journal of the American Statistical Association*, 116(534):833–844, 2021.

Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *Nature*, 393 (6684):440–442, 1998.

Appendix A. Examples and Non-Examples of Segregated Graphs

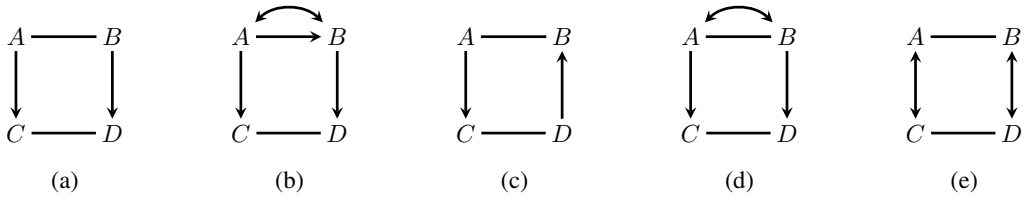

Figure 4: The graphs shown in (a) and (b) are examples of segregated graphs. The graphs shown in (c), (d), and (e) are not segregated graphs.

The two graphs in Figures 4(a) and (b) are indeed SGs as they satisfy properties (i-iii) listed in the definition of SGs in Section 2. The graph in Figure 4(c) is not an SG due to the existence of a partially directed cycle, violating property (iii). The graph in Figure 4(d) is not an SG due to A and B being connected by both a bidirected and undirected edge, violating property (i). Finally, Figure 4(e) is not an SG as there exist vertices that serve as the endpoints of both a bidirected and undirected edge, violating property (ii).

Appendix B. Detailed Description of s-separation

To describe s-separation in an SG $\mathcal{G}(V)$ with vertices V, we require the following graphical definitions. For brevity, we sometimes use $*\!\!-\!\!*$ to denote situations where any of the three types of edges is valid and $*\!\!\rightarrow$ to indicate that only a directed or bidirected edge is valid.

B.1. Graph terminology

A *walk* in $\mathcal{G}(V)$ is an alternating sequence of vertices and edges $V_1 *\!\!-\!\!* V_2 \cdots *\!\!-\!\!* V_K$, where every $V_k *\!\!-\!\!* V_{k+1}$ is an edge present in $\mathcal{G}$. A *path* is a walk consisting of unique vertices and edges. A *section* of a walk is defined as a maximal subwalk that only consists of undirected edges (by definition a section could consist of just a single vertex). Two vertices V_i and V_j are said to be connected by a *collider section* if there exists a walk between V_i and V_j of the form $V_i *\!\!\rightarrow \sigma \leftarrow\!\!* V_j$, where σ is a section. A *partially directed walk* between vertices V_i and V_j is a walk starting at V_i and ending at V_j where every edge is either an undirected edge or a directed edge pointing from V_k to V_{k+1}. The *anterior* of a set of variables S denoted as $\mathrm{ant}_{\mathcal{G}}(S)$, is the set S as well as any vertices in $\mathcal{G}$ that have a partially directed walk to any vertex in S. We use $\mathcal{G}_S$ to denote a *subgraph* of $\mathcal{G}(V)$ consisting of only vertices in S and the edges between them. Finally, given a subgraph $\mathcal{G}_S$ we use $\mathcal{G}_S^a$ to denote an undirected graph—sometimes called an *augmented graph*—containing vertices in S and edges $S_i - S_j$ if any edge $S_i *\!\!-\!\!* S_j$ exists in $\mathcal{G}_S$ or S_i and S_j are connected by a walk in $\mathcal{G}_S$ consisting of just collider sections.[11]

Finally, given disjoint sets of vertices X, Y, Z such that $(X \cup Y \cup Z) \subseteq V$, the sets X and Y are said to be *s-separated* given Z in an SG $\mathcal{G}(V)$ if at least one vertex in Z is present in each path between X and Y in $\mathcal{G}^a_{\mathrm{ant}_{\mathcal{G}}(X \cup Y \cup Z)}$. In the main text, we described the relation between s-separation

11. Two examples of such walks are $S_i \leftrightarrow A \leftrightarrow B \leftrightarrow S_j$ and $S_i \rightarrow A - B - C \leftarrow S_j$.

and the global Markov property of SGs. In our proofs, we also occasionally invoke the following standard result in undirected graphs (these undirected graphs arise as a consequence of augmentation when checking s-separation queries). The Markov blanket of a vertex V_i in the undirected graph $\mathcal{G}(V)$, denoted $\mathrm{mb}_{\mathcal{G}}(V_i)$, is defined as the set of all vertices V_j that share an undirected edge with V_i. In an undirected graph $\mathcal{G}(V)$, we have $V_i \perp\!\!\!\perp V \setminus \mathrm{mb}_{\mathcal{G}}(V_i) \cup \{V_i\} \mid \mathrm{mb}_{\mathcal{G}}(V_i)$ in $p(V)$.

Appendix C. Proof of Lemma 1

Proof We first prove (3) holds under the null hypothesis, i.e., edges in the Y layer are all undirected. Let $\mathrm{ant}_{\mathcal{G}}(\cdot)$ denote the anterior of the vertices in the independence query (3). Under the null, (regardless of connections in other layers) there are no walks from Y_i to any other vertex in $\mathrm{ant}_{\mathcal{G}}(\cdot)$ that consist of just collider sections, since all walks starting at Y_i must start with either $Y_i \leftarrow \cdots$ or $Y_i - \cdots$ which leads to at least one non-collider section. Based on this observation, the only edges incident to Y_i in the augmented graph $\mathcal{G}^a_{\mathrm{ant}_{\mathcal{G}}(\cdot)}$ are undirected versions of those that already exist in $\mathcal{G}_{\mathrm{ant}_{\mathcal{G}}(\cdot)}$. This leads to the following edges in the augmented graph: edges $Y_i - L_k \; \forall \; L_k \in L_{\mathcal{F}_i^{(0,1)}}$, edges $Y_i - A_k \; \forall \; A_k \in A_{\mathcal{F}_i^{(0,1)}}$, and edges $Y_i - Y_k \; \forall \; Y_k \in Y_{\mathcal{F}_i^{(1)}}$. The conditioning set in (3) is then a superset of vertices that have edges incident to Y_i. Further, no $Y_j \in Y_{\mathcal{F}_i^{(2)}}$ is contained in this set. The result then follows, since Y_i and in $Y_{\mathcal{F}_i^{(2)}}$ are separated in $\mathcal{G}^a_{\mathrm{ant}_{\mathcal{G}}(\cdot)}$ given the conditioning set.

We now show that (3) does not hold under the alternative hypothesis, i.e., edges in the Y layer are all bidirected. In this case, there exist walks in $\mathcal{G}_{\mathrm{ant}_{\mathcal{G}}(\cdot)}$ from Y_i to $Y_j \in Y_{\mathcal{F}_i^{(2)}}$ consisting of just collider sections—for example, ones of the form $Y_i \leftrightarrow Y_k \leftrightarrow Y_j$, where $Y_k \in Y_{\mathcal{F}^{(1)}}$. Under the assumption of faithfulness, the result follows immediately as there now exists $Y_i - Y_j$ in $\mathcal{G}^a_{\mathrm{ant}_{\mathcal{G}}(\cdot)}$ and the two vertices are not s-separated based on the proposed conditioning set.

The proofs for (1) and (2) are nearly identical using the appropriate $\mathcal{G}_{\mathrm{ant}_{\mathcal{G}}(\cdot)}$ and $\mathcal{G}^a_{\mathrm{ant}_{\mathcal{G}}(\cdot)}$ based on vertices provided in the respective conditional independence queries. ■

Appendix D. Proof of Lemma 2

Proof We first show that the factorizations in (8) hold under both the null and alternative hypotheses. Let $L_{\mathcal{CL}}, A_{\mathcal{CL}}, Y_{\mathcal{CL}}$ be variables that appear in the coding likelihoods in (8), and for each $🧍_i \in \mathcal{S}^6(\mathcal{F})$ let $\mathcal{O}_i = L_{\mathcal{F}_i^{(0,1,2,3)}} \cup A_{\mathcal{F}_i^{(0,1,2,3)}} \cup Y_{\mathcal{F}_i^{(0,1,2)}}$ and $\mathcal{O}_{-i} = (L_{\mathcal{CL}} \cup A_{\mathcal{CL}} \cup Y_{\mathcal{CL}}) \setminus \mathcal{O}_i$. Notice for any $🧍_i \in \mathcal{S}^6(\mathcal{F})$, the variables $\mathcal{O}_{-i}$ belong to units at least 4 degrees of separation away from $🧍_i$. If this were not the case, there would be a path between $🧍_i$ and $🧍_j$ in $\mathcal{F}$ that is shorter than 6 edges, contradicting the antecedent that they are 6 degrees apart. Thus, to show factorizations of the coding likelihood function, it is sufficient to show that for each $🧍_i \in \mathcal{S}^6(\mathcal{F})$ and any $🧍_j$ more than 3 degrees of separation away from $🧍_i$, we have that $Y_i \perp\!\!\!\perp L_j, A_j, Y_j \mid \cdot$ and $Y_i \perp\!\!\!\perp L_j, A_j, Y_j \mid \cdot, Y_{\mathcal{F}_i^{(2)}}$ under both the null and alternative hypotheses.

Under the null, we have only undirected edges in the Y layer so the anterior for Y_i could potentially be the entire graph. However, similar to the proof for Lemma 1, the vertices in $Y_{\mathcal{F}_i^{(1)}} \cup A_{\mathcal{F}_i^{(0,1)}} \cup L_{\mathcal{F}_i^{(0,1)}}$ form the Markov blanket of Y_i in $\mathcal{G}^a_{(L \cup A \cup Y)}$. As the conditioning sets in (8) are supersets of the Markov blankets of each Y_i and variables of $🧍_j$ are not included in them, the aforementioned independences for factorization hold.

Under the alternative, we have only bidirected edges in the Y layer. Thus, for some $i \in \mathcal{S}^6(\mathcal{F})$ and any j more than 3 degrees of separation away from i in $\mathcal{F}$, $Y_{\mathcal{F}_i^{(3)}}$ is not included in the following two augmented graphs: $\mathcal{G}_1^a \equiv \mathcal{G}^a_{\text{ant}_\mathcal{G}(Y_i, L_j, A_j, Y_j, \cdot)}$ and $\mathcal{G}_2^a \equiv \mathcal{G}^a_{\text{ant}_\mathcal{G}(Y_i, L_j, A_j, Y_j, \cdot, Y_{\mathcal{F}_i^{(2)}})}$. In $\mathcal{G}_1$, the variable Y_i has a walk consisting of just collider sections to $L_k \in L_{\mathcal{F}_i^{(2)}}$ and $A_k \in A_{\mathcal{F}_i^{(2)}}$, which implies the existence of edges $Y_i - L_k$ and $Y_i - A_k$ in the augmented graph $\mathcal{G}_1^a$. Together with variables $\{V_k \mid Y_i \leftarrow\!* V_k \in \mathcal{G}_1\}$, the neighbors of Y_i in $\mathcal{G}_1^a$ are: $L_k \in L_{\mathcal{F}_i^{(0,1,2)}}$, $A_k \in A_{\mathcal{F}_i^{(0,1,2)}}$, and $Y_k \in Y_{\mathcal{F}_i^{(1)}}$. Since they are a subset of '$\cdot$' in the query $Y_i \perp\!\!\!\perp L_j, A_j, Y_j \mid \cdot$, the coding likelihood $\mathcal{CL}(\beta_Y)$ factorizes under the alternative. Similarly in $\mathcal{G}_2$, Y_i has a walk consisting of just collider sections to $L_k \in L_{\mathcal{F}_i^{(2,3)}}$, $A_k \in A_{\mathcal{F}_i^{(2,3)}}$, and $Y_k \in Y_{\mathcal{F}_i^{(2)}}$ (but not $Y_k \in Y_{\mathcal{F}_i^{(3)}}$ as mentioned earlier). Thus, $\text{mb}_{\mathcal{G}_2^a}(Y_i) = L_{\mathcal{F}_i^{(0,1,2,3)}} \cup A_{\mathcal{F}_i^{(0,1,2,3)}} \cup Y_{\mathcal{F}_i^{(1,2)}}$. These neighbors is exactly the conditioning set in $Y_i \perp\!\!\!\perp L_j, A_j, Y_j \mid \cdot, Y_{\mathcal{F}_i^{(2)}}$, which proves the independence property needed to factorize $\mathcal{CL}(\gamma_Y)$ under the alternative.

The proofs for (6) and (7) are nearly identical using the appropriate independence queries based on the variables that appear in the (6) and (7). ■

Appendix E. Proof of Theorem 3

Proof Let $\mathcal{B}(\mathcal{G})$ denote the set of maximal undirected connected components of $\mathcal{G}(V \cup H)$. Given a set of variables $A \subset (V \cup H)$, the post-intervention distribution where we intervene on A is obtained via an extension of the g-formula (Pearl, 2009; Robins, 1986) for chain graphs (Lauritzen and Richardson, 2002),

$$p(V \setminus A \mid \text{do}(a)) = \prod_{B \in \mathcal{B}(\mathcal{G})} p(B \setminus A \mid \text{pa}_\mathcal{G}(B), B \cap A)\Big|_{A=a}.$$

While these post-intervention distributions are always identified when there are no hidden variables ($H = \emptyset$), this is not the case in general.

Under the assumptions of Figure 1(b), we can partition H into distinct exogenous sets H_L, H_A, H_Y (any of which could be empty if the respective layer contains undirected edges instead of bidirected edges) that have outgoing edges to vertices in L, A, Y respectively. Then, by the CG g-formula,

$$\begin{aligned} p(L, Y \mid \text{do}(a)) = \sum_{H_L, H_A, H_Y} \big(&p(H_L) \times p(L \mid H_L) \\ &\times p(H_A) \times p(H_Y) \times p(Y \mid H_Y, A = a, L)\big). \end{aligned}$$

By evaluating the sum over H_A and gathering like terms we can simplify the right hand side to,

$$\Big(\sum_{H_L} p(H_L, L)\Big)\Big(\sum_{H_Y} p(H_Y) \times p(Y \mid A = a, L)\Big).$$

Finally, since $H_Y \perp\!\!\!\perp A, L$ by applying s-separation in $\mathcal{G}(V \cup H)$ we can further simplify to,

$$\Big(\sum_{H_L} p(H_L, L)\Big)\Big(\sum_{H_Y} p(H_Y, Y \mid A = a, L)\Big).$$

This gives the following identifying formula in terms of just observed variables,

$$p(L, Y \mid \text{do}(a)) = p(L) \times p(Y \mid A = a, L).$$

The result follows as,

$$\mathbb{E}[Y_i \mid \text{do}(a)] = \sum_{L,Y} p(L, Y \mid \text{do}(a)) \times Y_i = \sum_{L,Y} p(L) \times p(Y \mid A = a, L) \times Y_i.$$

■

Appendix F. Proof of Different Identifying Functionals for $\mathbb{E}[Y_i \mid \text{do}(Y_{-i} = y_{-i})]$

Proof First consider the case when the edges in the Y layer are undirected. In this case $H_Y = \emptyset$, while H_L and H_A may or may not be empty. Applying the chain graph g-formula gives us,

$$p(L, A, Y_i \mid \text{do}(y_{-i})) = \sum_{H_L, H_A} p(H_L) \times p(L \mid H_L) \times p(H_A) \times p(A \mid H_A, L) \times p(Y_i \mid y_{-i}, A, L).$$

Since $H_A \perp\!\!\!\perp L$ in $\mathcal{G}(V \cup H)$ we can simplify the above and gather like terms to get,

$$\left(\sum_{H_L} p(H_L, L) \right) \left(\sum_{H_A} p(H_A, A \mid L) \right) \times p(Y \mid y_{-i}, A, L).$$

This gives the following identifying formula in terms of just observed variables,

$$p(L, A, Y_i \mid \text{do}(y_{-i}) = p(L) \times p(A \mid L) \times p(Y \mid y_{-i}, A, L).$$

The result follows as,

$$\mathbb{E}[Y_i \mid \text{do}(y_{-i})] = \sum_{L,A,Y_i} p(L, A, Y_i \mid \text{do}(y_{-i})) \times Y_i = \sum_{L,A,Y_i} p(L, A) \times p(Y \mid y_{-i}, A, L) \times Y_i.$$

Now we turn to the case when the edges in the Y layer are bidirected. Applying the chain graph g-formula in this setting gives us,

$$p(L, A, Y_i \mid \text{do}(y_{-i})) = \sum_{H_L, H_A, H_Y} p(H_L) \times p(L \mid H_L) \times p(H_A) \times p(A \mid H_A, L) \times p(H_Y) \times p(Y_i \mid H_Y, A, L).$$

Simplifying this functional based on independences $H_A \perp\!\!\!\perp L$ and $H_Y \perp\!\!\!\perp A, L$ gives us,

$$p(L, A, Y_i \mid \text{do}(y_{-i})) = p(L) \times p(A \mid L) \times p(Y_i \mid A, L) = p(L, A, Y_i).$$

The result follows as,

$$\mathbb{E}[Y_i \mid \text{do}(y_{-i})] = \sum_{L,A,Y_i} p(L, A, Y_i \mid \text{do}(y_{-i})) \times Y_i = \sum_{Y_i} p(Y_i) \times Y_i.$$

■

Appendix G. Proof of Lemma 4

The proof for this is very similar to the proof in Appendix D and involves checking some relatively simple s-separation conditions compared to those in Appendix D as we never condition on variables in colliders or collider sections. Thus, we omit further details of the proof.

Appendix H. Gibbs Samplers for L and Y

The Gibbs samplers for drawing samples from $p(L)$ and $p(Y \mid A, L)$ when the edges in the L and Y layer are undirected are described in Algorithms 3 and 4. In these algorithms, $< i$ denotes all units in $\mathcal{F}$ that precede unit i in a total ordering of all units and $> i$ denotes all units that come after unit i. In our experiments we set the number of draws M from the Gibbs sampler to be $0.3 \times N$, the thinning interval T to be 3, and the burn-in period m^* to be 200.

Algorithm 3 Gibbs sampler L

Inputs: SG $\mathcal{G}$; Parameters $\theta_{L_i - L_j}$
Initialize an empty list L_{draws} of size M corresponding to number of draws we want
Initialize a random vector $L^{(0)}$
for $m = 0, 1, \ldots, T \times M + m^*$ **do** // m^* is the burn-in period; T is the thinning interval
    for $i = 1, \ldots, N$ **do**
        $L_i^{(m+1)} \leftarrow$ draw from $p(L_i \mid L_{<i}^{(m+1)}, L_{>i}^{(m)}; \theta_{L_i - L_j})$
    if $(m > m^*)$ and $(m \bmod T = 0)$ **then**
        Append $L^{(m)}$ to L_{draws}
return L_{draws}

Algorithm 4 Gibbs sampler Y

Inputs: SG $\mathcal{G}$; Parameters for $\theta_{Y_i - Y_j}$; Previously sampled L_{draws}, a size M list; Treatment assignment vector a.
Initialize Y_{draws}, an empty list of size M
for $m = 1, \ldots, M$ **do**
    $L^{(m)} \leftarrow$ the m-th element of L_{draws}
    Initialize a random vector $Y^{(0)}$;
    for $m' = 0, 1 \ldots, m^*$ **do** // m^* is the burn-in period.
        for $i = 1, \ldots, N$ **do**
            $Y_i^{(m'+1)} \leftarrow$ draw from $p(Y_i \mid Y_{<i}^{(m'+1)}, Y_{>i}^{(m')}, A = a, L^{(m)}; \theta_{Y_i - Y_j})$
    Append $Y^{(m^*)}$ to Y_{draws}
return Y_{draws}

Appendix I. Experiment Details

I.1. Data Generating Process (DGP) for Evaluating the Likelihood Ratio Tests

In this section, we outline the DGPs for the two configurations of $\mathcal{G}$: one with undirected edges in all three layers and the other with bidirected edges in all layers; and directed edges in both configurations follow the set up in the partially determined SG $\mathcal{G}_{\circ\!-\!\circ}$ in Figure 1 (b).

For the configuration where all layers contain undirected edges, we generate L, A, and Y as vectors of binary values using the following Gibbs factors:

$$
\begin{aligned}
L_i &\sim \text{Bernoulli}(\text{expit}(-0.1 * \sum_{j \in \mathcal{F}_i^{(1)}} L_j));\\
A_i &\sim \text{Bernoulli}(\text{expit}(0.8 * L_i - 0.1 * \sum_{j \in \mathcal{F}_i^{(1)}} L_j - 0.1 * \sum_{j \in \mathcal{F}_i^{(1)}} A_j));\\
Y_i &\sim \text{Bernoulli}(\text{expit}(0.8 * L_i + 1.7 * A_i - 0.1 * \sum_{j \in \mathcal{F}_i^{(1)}} L_j - 0.1 * \sum_{j \in \mathcal{F}_i^{(1)}} A_j - 0.1 * \sum_{j \in \mathcal{F}_i^{(1)}} Y_j)).
\end{aligned}
$$

Using these Gibbs factors, we generate values layers by layers in topological order. First, we perform Gibbs sampling to generate L. Next, we fix the L vector and use it as inputs to generate A. Finally, L and A are used as inputs to generate Y. For the Gibbs sampling process at each layer, we use a burn-in period equaling 200 multiplied by the number of units in the network $\mathcal{F}$.

For the configuration where all layers contain bidirected edges, we generate L, A, and Y as vectors of binary values by first explicitly generating the hidden variables H_{ij}^L, H_{ij}^A, and H_{ij}^Y corresponding to the bidirected edges $L_i \leftrightarrow L_j$, $A_i \leftrightarrow A_j$, and $Y_i \leftrightarrow Y_j$, respectively, for each connected pair (i, j) in $\mathcal{F}$. After L, A, and Y are generated, we discard the H values to make them unobserved. The parameters of this DGP are as follows:

$$
\begin{aligned}
&H_{ij}^L, H_{ij}^A, H_{ij}^Y \sim \text{Normal}(0, 1);\\
&L_i \sim \text{Bernoulli}(\text{expit}(5 * \sum_{k \in \mathcal{F}_i^{(1)}} H_{ik}^L));\\
&A_i \sim \text{Bernoulli}(\text{expit}(0.2 * L_i + 0.1 * \sum_{k \in \mathcal{F}_i^{(1)}} L_k + 5 * \sum_{k \in \mathcal{F}_i^{(1)}} H_{ik}^A));\\
&Y_i \sim \text{Bernoulli}(\text{expit}(0.2 * L_i - 0.3 * A_i + 0.1 * \sum_{k \in \mathcal{F}_i^{(1)}} L_k - 0.2 * \sum_{k \in \mathcal{F}_i^{(1)}} A_k + 5 * \sum_{k \in \mathcal{F}_i^{(1)}} H_{ik}^Y));
\end{aligned}
$$

I.2. Parametric Models for Evaluating the Likelihood Ratio Tests

When conducting the likelihood ratio tests, we use the following nested parametric models. Models (9) and (10) correspond to the null and alternative models we fit when conducting the likelihood ratio test for the L layer. Similarly, (11) and (12) are the null and alternative models for the A layer, and (13) and (14) are for the Y layer.

$$p(L_i = 1 \mid L_{\mathcal{F}_i^{(1)}}; \beta) = \text{expit}(\beta_0 + \beta_1 \sum_{k \in \mathcal{F}_i^{(1)}} L_k) \tag{9}$$

$$p(L_i = 1 \mid L_{\mathcal{F}_i^{(1,2)}}; \gamma) = \text{expit}(\gamma_0 + \gamma_1 \sum_{k \in \mathcal{F}_i^{(1)}} L_k + \gamma_2 \sum_{k \in \mathcal{F}_i^{(1)}} L_k) \tag{10}$$

$$p(A_i = 1 \mid A_{\mathcal{F}_i^{(1)}}, L_{\mathcal{F}_i^{(0,1,2,3)}}; \beta) = \text{expit}(\beta_0 + \beta_1 \sum_{k \in \mathcal{F}_i^{(1)}} A_k + \tag{11}$$
$$\beta_2 L_i + \beta_3 \sum_{k \in \mathcal{F}_i^{(1)}} L_k + \beta_4 \sum_{k \in \mathcal{F}_i^{(2)}} L_k + \beta_5 \sum_{k \in \mathcal{F}_i^{(3)}} L_k)$$

$$p(A_i = 1 \mid A_{\mathcal{F}_i^{(1,2)}}, L_{\mathcal{F}_i^{(0,1,2,3)}}; \gamma) = \text{expit}(\gamma_0 + \gamma_1 \sum_{k \in \mathcal{F}_i^{(1)}} A_k + \gamma_2 \sum_{k \in \mathcal{F}_i^{(2)}} A_k + \tag{12}$$
$$\gamma_3 L_i + \gamma_4 \sum_{k \in \mathcal{F}_i^{(1)}} L_k + \gamma_5 \sum_{k \in \mathcal{F}_i^{(2)}} L_k + \gamma_6 \sum_{k \in \mathcal{F}_i^{(3)}} L_k)$$

$$p(Y_i = 1 \mid Y_{\mathcal{F}_i^{(1)}}, A_{\mathcal{F}_i^{(0,1,2,3)}}, L_{\mathcal{F}_i^{(0,1,2,3)}}; \beta) = \text{expit}(\beta_0 + \beta_1 \sum_{k \in \mathcal{F}_i^{(1)}} Y_k + \tag{13}$$
$$\beta_2 A_i + \beta_3 \sum_{k \in \mathcal{F}_i^{(1)}} A_k + \beta_4 \sum_{k \in \mathcal{F}_i^{(2)}} A_k + \beta_5 \sum_{k \in \mathcal{F}_i^{(3)}} A_k + \beta_6 L_i + \beta_7 \sum_{k \in \mathcal{F}_i^{(1)}} L_k +$$
$$\beta_8 \sum_{k \in \mathcal{F}_i^{(2)}} L_k + \beta_9 \sum_{k \in \mathcal{F}_i^{(3)}} L_k)$$

$$p(Y_i = 1 \mid Y_{\mathcal{F}_i^{(1,2)}}, A_{\mathcal{F}_i^{(0,1,2,3)}}, L_{\mathcal{F}_i^{(0,1,2,3)}}; \gamma) = \text{expit}(\gamma_0 + \gamma_1 \sum_{k \in \mathcal{F}_i^{(1)}} Y_k + \gamma_2 \sum_{k \in \mathcal{F}_i^{(2)}} Y_k + \tag{14}$$
$$\gamma_3 A_i + \gamma_4 \sum_{k \in \mathcal{F}_i^{(1)}} A_k + \gamma_5 \sum_{k \in \mathcal{F}_i^{(2)}} A_k + \gamma_6 \sum_{k \in \mathcal{F}_i^{(3)}} A_k + \gamma_7 L_i + \gamma_8 \sum_{k \in \mathcal{F}_i^{(1)}} L_k +$$
$$\gamma_9 \sum_{k \in \mathcal{F}_i^{(2)}} L_k + \gamma_{10} \sum_{k \in \mathcal{F}_i^{(3)}} L_k)$$

I.3. DGPs for Evaluating the Causal Effect Estimation Method

We now outline the DGPs for L, A, and Y in this experiment. Recall that N denotes the number of units in $\mathcal{F}$. For layers with undirected edges, we use Gibbs sampling with a burn-in period of $m^* = 200$. We keep the sample immediately after the burn-in period as the sampled data. When deriving the ground-truth expectations $\mathbb{E}[Y \mid \text{do}(A = \mathbf{1})]$ and $\mathbb{E}[Y \mid \text{do}(A = \mathbf{0})]$, we continue

running the Gibbs sampler to obtain $M = 0.3 \times N$ draws with a thinning interval of $T = 3$ to reduce auto correlation. Thus, we average over $0.3 \times N$ samples of Y to derive the ground-truths.

When the L layer is connected by undirected edges, the distribution of L_i for all $i \in \mathcal{F}$ is given by:

$$L_i \sim \text{Bernoulli}(\text{expit}(-0.3 + 0.4 * \sum_{k \in \mathcal{F}_i^{(1)}} L_k)).$$

When the L layer is connected by bidirected edges, the distribution of the entire L vector is given by:

$$L \sim MVN(\mu, \Sigma),$$

where μ is a vector with all entries equal to 0.7, and Σ is a matrix with diagonal elements equal to 3.5. For an off-diagonal location (x, y) in Σ, the value is 0.2 if 🧍$_x$ and 🧍$_y$ are neighbors in $\mathcal{F}$ and is 0 otherwise. When the A layer is connected by undirected edges, the distribution of A_i for all $i \in \mathcal{F}$ is given by:

$$A_i \sim \text{Bernoulli}(\text{expit}(5 + 4 * L_i - 1.2 * \sum_{k \in \mathcal{F}_i^{(1)}} L_k - 2 * \sum_{k \in \mathcal{F}_i^{(1)}} A_k)).$$

When the A layer is connected by bidirected edges, we generate A by first explicitly sampling hidden variables H_{ij}^A for each connected pair (i, j) in $\mathcal{F}$ and then determining each entry of A using the following distributions (the hidden variables are discarded after A is generated):

$$\begin{aligned} H_{ij}^A &\sim \text{Normal}(2, 1); \\ A_i &\sim \text{Bernoulli}(\text{expit}(1.3 + 0.2 * \sum_{k \in \mathcal{F}_i^{(1)}} H_{ik}^A - 0.4 * L_i - 0.7 * \sum_{k \in \mathcal{F}_i^{(1)}} L_k)). \end{aligned}$$

When the Y layer is connected by undirected edges, the distribution of Y_i for all $i \in \mathcal{F}$ is given by:

$$Y_i \sim \text{Bernoulli}(\text{expit}(2 + L_i + 1.5 * A_i - 5.3 * \sum_{k \in \mathcal{F}_i^{(1)}} L_k + \sum_{k \in \mathcal{F}_i^{(1)}} A_k - 4 * \sum_{k \in \mathcal{F}_i^{(1)}} Y_k)).$$

When the Y layer is connected by bidirected edges, we first generate hidden variables H_{ij}^Y for each connected pair (i, j) in $\mathcal{F}$ and then generate Y accordingly, using the following distributions:

$$\begin{aligned} H_{ij}^Y &\sim \text{Normal}(0, 1); \\ Y_i &\sim \text{Bernoulli}(\text{expit}(-1 + 2 * \sum_{k \in \mathcal{F}_i^{(1)}} H_{ik}^Y + 0.1 * L_i + A_i - 0.3 * \sum_{k \in \mathcal{F}_i^{(1)}} L_k + \sum_{k \in \mathcal{F}_i^{(1)}} A_k)). \end{aligned}$$

I.4. Additional Experimental Results

In the main text, we have shown that our method produces consistent and unbiased causal estimates when all layers are connected by bidirected edges. In this section, we will show that our method is also consistent and unbiased for the other 6 cases not yet discussed in the main text—UBB, BUB, BBU, BUU, and UUB.

Figure 10 only has a single plot because causal effects in the UBU setup can be estimated directly using the auto-g method, as discussed in the main text. For the other plots below, we see that our methods are consistent and unbiased, as the distribution of the estimates becomes narrower as the sample size increases, and the distributions all center around the ground truth. However, naive application of the auto-g-computation method to scenarios where layer L or Y is connected by latent confounding leads to biased results. We also find that the auto-g method has some robustness against model misspecification, as its estimates in Figure 8 are centered around the ground truth. We believe this robustness is likely due to the particular DGP we have chosen, since we have demonstrated in the main text that it is theoretically incorrect to apply the auto-g-computation method when bidirected edges exist in the L layer.

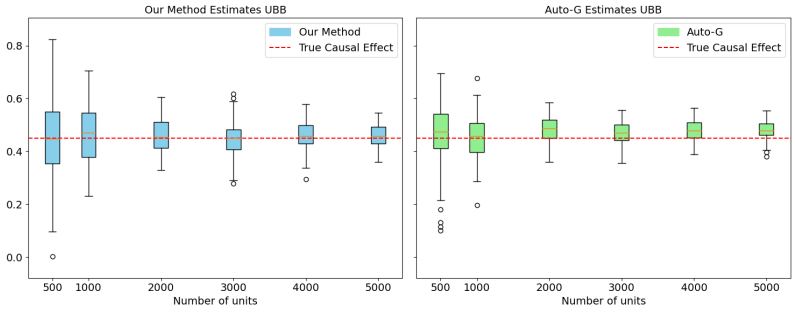

Figure 5: Network effect estimation results for the UBB setup.

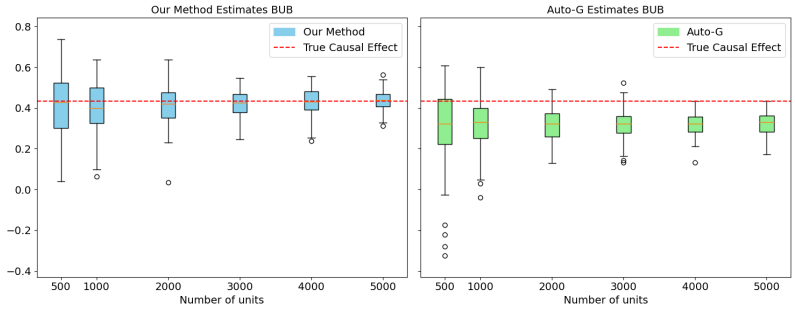

Figure 6: Network effect estimation results for the BUB setup.

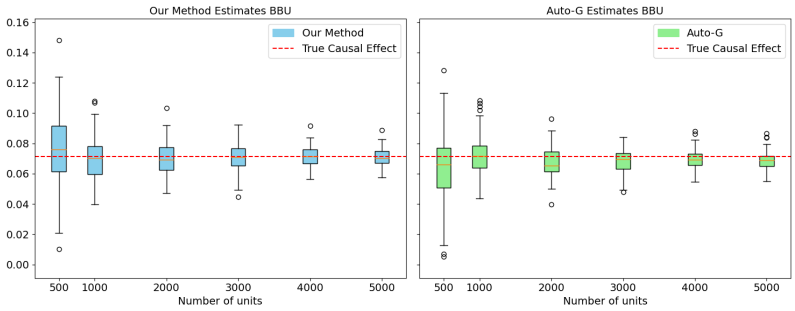

Figure 7: Network effect estimation results for the BBU setup.

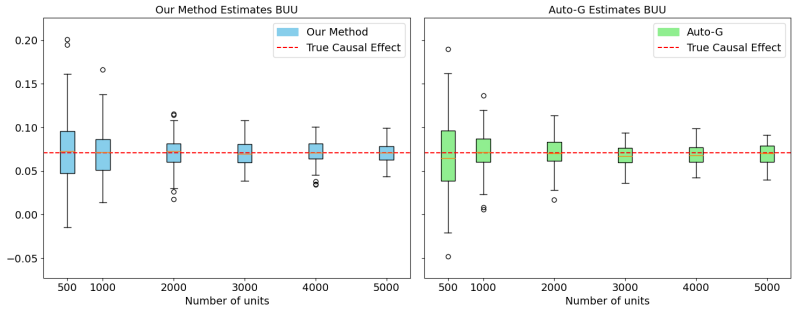

Figure 8: Network effect estimation results for the BUU setup.

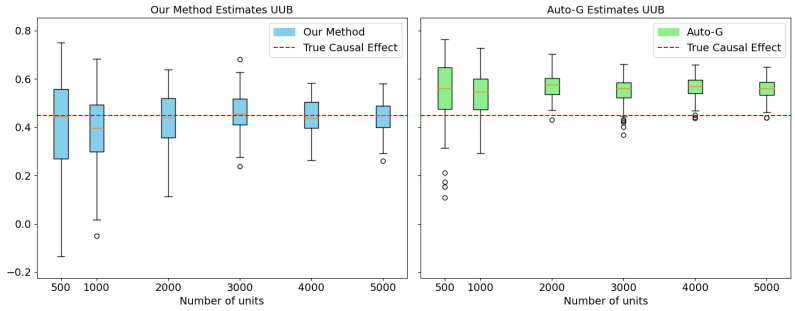

Figure 9: Network effect estimation results for the UUB setup.

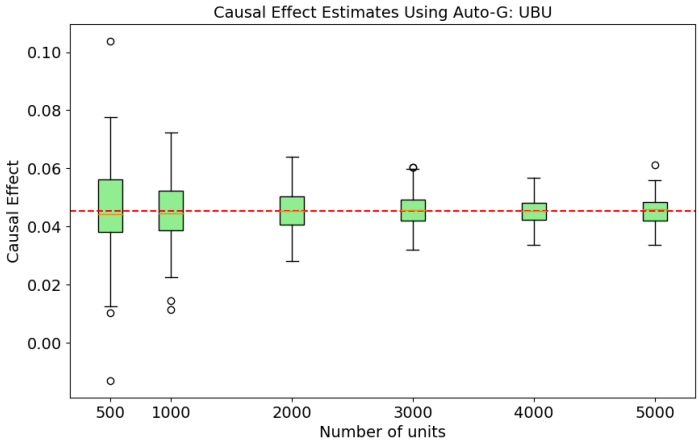

Figure 10: Consistency the auto-g-computation in the UBU setup.