Appendix to: Intervening on Network Ties

Eli Sherman
Department of Computer Science
Johns Hopkins University
Baltimore, MD

Ilya Shpitser
Department of Computer Science
Johns Hopkins University
Baltimore, MD

APPENDIX

A: Additional Network Intervention Examples

Housing Vouchers

In urban development economics, housing vouchers have been proposed as a means of inducing families to move from areas with less opportunity for upward socioeconomic mobility to areas of with greater opportunity [2]. As pointed out in [3], oftentimes these proposals ignore the social network-related implications of such a policy (e.g. decision is impacted by talking to neighbors). In Fig. 1 we consider the specific impacts of transplanting a family from one neighborhood to the other.

Each unit i in Fig. 1 has a variable C_i representing the unit's demographic information, and an outcome Y_i which represents the socioeconomic outcome targeted by the housing voucher, such as annual income. In Fig 1 (a) we see the pre-intervention network, where unit 3 is neighbors with unit 2, while after the intervention (Fig. 1 (b)) unit 3 has been moved to a new neighborhood and is now neighbors with unit 4 instead. This example represents both a severance and a connection intervention.

Influencer Networks

Understanding social influence in networks is of interest to a wide variety of fields. Researchers who study infectious diseases and intravenous drug abuse often attempt to identify major influencers that impact many people. More recently, determining and leveraging knowledge of influence has become the focus of the algorithmic marketing community. If one can identify the strongest influencer in a network (not necessarily the individual with the most connections), then understanding the effects of removing that individual from the network (e.g. by arresting a drug-kingpin) might be useful for policymakers.

In Fig. 2, we represent this phenomenon via an undirected graph where each unit, represented by Y_i, is internally a DAG and the undirected edges between units encode symmetric directed relationships similar to those in our other examples. Fig. 2 (b) depicts the result of a hypothetical intervention on Fig. 2 (a) in which unit 3 is effectively *removed* from the network by severing all connections with it's friends. As a side-effect unit 4 is also effectively removed from the network.

B: Proofs

We first prove two utility results on factorizations of joint densities following Chen [1]. The first is a simple lemma which is needed to prove the corollary that follows. We will use Corollary 1 in each of the results that follow.

Let $\mathbf{a}$ be a set of fixed values. Then for a two-variable conditional density $f(Y_1, Y_2 \mid \mathbf{a})$, we have:

$$\frac{f(Y_1 \mid Y_2^0, \mathbf{a}) OR(Y_1, Y_2 \mid \mathbf{a}) f(Y_2 \mid Y_1^0, \mathbf{a})}{\sum_{Y_1, Y_2} f(Y_1 \mid Y_2^0, \mathbf{a}) OR(Y_1, Y_2 \mid \mathbf{a}) f(Y_2 \mid Y_1^0, \mathbf{a})},$$

where the odds ratio is given by

$$\begin{aligned} OR(Y_1, Y_2 \mid \mathbf{a}) &= \frac{f(Y_1 \mid Y_2, \mathbf{a}) f(Y_1^0 \mid Y_2^0, \mathbf{a})}{f(Y_1 \mid Y_2^0, \mathbf{a}) f(Y_1^0 \mid Y_2, \mathbf{a})} \\ &= \frac{f(Y_1, Y_2 \mid \mathbf{a}) f(Y_1^0, Y_2^0 \mid \mathbf{a})}{f(Y_1, Y_2^0 \mid \mathbf{a}) f(Y_1^0, Y_2 \mid \mathbf{a})}. \end{aligned}$$

and Y_1^0 and Y_2^0 signify reference values of Y_1 and Y_2.

Lemma 1

$$\frac{f(Y \mid Z, X)}{f(Y_0 \mid Z, X)} = \frac{f(Y \mid X) OR(Z, Y \mid X)}{f(Y_0 \mid X) \mathbb{E}[OR(Z, Y \mid X) \mid Y_0, X]}$$

Figure 1: (a) A DAG representing hypothetical connections between family units in two separate neighborhoods; (b) the DAG resulting from moving unit 3 from the first neighborhood to the second neighborhood via both a severance and connection intervention.

Figure 2: (a) An undirected graph representing connections in an influence network between 6 agents; (b) the undirected graph resulting from intervening on the network in (a) such that unit 3 is removed from the network.

Proof:

$$
\begin{aligned}
\frac{f(Y \mid Z, X)}{f(Y_0 \mid Z, X)} &=^1 \frac{f(Y \mid X) f(Z \mid Y, X)/f(Z \mid X)}{f(Y_0 \mid X) f(Z \mid Y_0, X)/f(Z \mid X)} \\
&=^2 \frac{f(Y \mid X) f(Z \mid Y, X)}{f(Y_0 \mid X) f(Z \mid Y_0, X)} \\
&=^3 \frac{f(Y \mid X) \frac{f(Z|Y_0,X) OR(Z,Y|X)}{\sum_Z f(Z|Y_0,X) OR(Z,Y|X)}}{f(Y_0 \mid X) f(Z \mid Y_0, X)} \\
&=^4 \frac{f(Y \mid X) OR(Z, Y \mid X)}{f(Y_0 \mid X) \mathbb{E}[OR(Z, Y \mid X) \mid Y_0, X]},
\end{aligned}
$$

where equality 1 is by Bayes rule, 2 by cancellation, 3 by the Chen factorization of a conditional density, and 4 by definition. □

Corollary 1 *$f(Y_1, Y_2 \mid \mathbf{a}) = \frac{\mathbf{X}}{\sum_{Y_1,Y_2} \mathbf{X}}$ where $\mathbf{X} =$*

$$
\frac{f(Y_1 \mid \mathbf{a})\ OR(Y_1, Y_2 \mid \mathbf{a}) OR(Y_1^0, Y_2|\mathbf{a}) OR(Y_1, Y_2^0|\mathbf{a})\ f(Y_2 \mid \mathbf{a})}{\mathbb{E}[OR(Y_1, Y_2^0 \mid \mathbf{a}) \mid Y_1^0, \mathbf{a}] \mathbb{E}[OR(Y_1^0, Y_2 \mid \mathbf{a}) \mid Y_2^0, \mathbf{a}]}
$$

Proof: We have the following for $f(Y_1, Y_2 \mid \mathbf{a})$ (from Chen [1]):

$$
\begin{aligned}
&\frac{f(Y_1 \mid Y_2^0, \mathbf{a}) OR(Y_1, Y_2 \mid \mathbf{a}) f(Y_2 \mid Y_1^0, \mathbf{a})}{\sum_{Y_1,Y_2} f(Y_1 \mid Y_2^0, \mathbf{a}) OR(Y_1, Y_2 \mid \mathbf{a}) f(Y_2 \mid Y_1^0, \mathbf{a})} \\
&=^1 \frac{\frac{f(Y_1|Y_2^0,\mathbf{a})}{f(Y_1^0|Y_2^0,\mathbf{a})} OR(Y_1, Y_2 \mid \mathbf{a}) \frac{f(Y_2|Y_1^0,\mathbf{a})}{f(Y_2^0|Y_1^0,\mathbf{a})}}{\sum_{Y_1,Y_2} \frac{f(Y_1|Y_2^0,\mathbf{a})}{f(Y_1^0|Y_2^0,\mathbf{a})} OR(Y_1, Y_2 \mid \mathbf{a}) \frac{f(Y_2|Y_1^0,\mathbf{a})}{f(Y_2^0|Y_1^0,\mathbf{a})}} \\
&=^2 \left[\frac{f(Y_1 \mid \mathbf{a})}{f(Y_1^0 \mid \mathbf{a})} \frac{OR(Y_1, Y_2^0 \mid \mathbf{a})}{\mathbb{E}[OR(Y_1, Y_2^0 \mid \mathbf{a}) \mid Y_1^0, \mathbf{a}]} \times OR(Y_1, Y_2 \mid \mathbf{a}) \times \frac{OR(Y_1^0, Y_2 \mid \mathbf{a})}{\mathbb{E}[OR(Y_1^0, Y_2 \mid \mathbf{a}) \mid Y_2^0, \mathbf{a}]} \frac{f(Y_2 \mid \mathbf{a})}{f(Y_2^0 \mid \mathbf{a})} \right] \\
&\times \left[\sum_{Y_1,Y_2} \left[\frac{f(Y_1 \mid \mathbf{a})}{f(Y_1^0 \mid \mathbf{a})} \frac{OR(Y_1, Y_2^0 \mid \mathbf{a})}{\mathbb{E}[OR(Y_1, Y_2^0 \mid \mathbf{a}) \mid Y_1^0, \mathbf{a}]} \times OR(Y_1, Y_2 \mid \mathbf{a}) \times \frac{OR(Y_1^0, Y_2 \mid \mathbf{a})}{\mathbb{E}[OR(Y_1^0, Y_2 \mid \mathbf{a}) \mid Y_2^0, \mathbf{a}]} \frac{f(Y_2 \mid \mathbf{a})}{f(Y_2^0 \mid \mathbf{a})} \right] \right]^{-1} \\
&=^3 \frac{\mathbf{X}}{\sum_{Y_1,Y_2} \mathbf{X}}.
\end{aligned}
$$

Equality 1 holds by probability rules since Y_1^0 and Y_2^0 are fixed. Equality 2 holds by application of Lemma 1. Equality 3 holds by reverse application of the identity for Y_1^0 and Y_2^0. □

We now prove the main results of the paper. Each applies Corollary 1 to show increasingly general results pertaining to the KL-divergence of a distribution $\tilde{p}$ with additional independence constraints relative to an observational distribution p.

Theorem 1 *Let $\mathbf{V}$ be a set of random variables with $p(\mathbf{V})$ corresponding to a DAG $\mathcal{G}$. Let $A \in \mathbf{V}$. Let $\mathcal{P}(\mathbf{V})$ be the set of probability distributions that factorize according to $\mathcal{G}$. Then*

$$p(A) \prod_{V \in \mathbf{V} \setminus A} p(V | \operatorname{pa}_{\mathcal{G}}(V)) = \arg\min_{\tilde{p} \in \mathcal{P}(\mathbf{V})} D_{KL}(p || \tilde{p}) \quad \text{s.t. } A \perp\!\!\!\perp \operatorname{pa}_{\mathcal{G}}(A)$$

Proof: Applying Corollary 1, we can express the KL-divergence of $\tilde{p}(A, \operatorname{pa}_{\mathcal{G}}(A))$ from $p(A, \operatorname{pa}_{\mathcal{G}})$ as proportional to:

$$\begin{aligned}
&\log \frac{p(A, \operatorname{pa}_{\mathcal{G}}(A))}{\tilde{p}(A, \operatorname{pa}_{\mathcal{G}}(A))} \\
&= \log \frac{\left[\frac{p(A)\frac{OR_{num}}{OR_{den}} p(\operatorname{pa}_{\mathcal{G}}(A))}{\sum_{A, \operatorname{pa}_{\mathcal{G}}(A)} p(A) \frac{OR_{num}}{OR_{den}} p(\operatorname{pa}_{\mathcal{G}}(A))}\right]}{\left[\frac{\tilde{p}(A)\frac{\widetilde{OR}_{num}}{\widetilde{OR}_{den}} \tilde{p}(\operatorname{pa}_{\mathcal{G}}(A))}{\sum_{A, \operatorname{pa}_{\mathcal{G}}(A)} \tilde{p}(A) \frac{\widetilde{OR}_{num}}{\widetilde{OR}_{den}} \tilde{p}(\operatorname{pa}_{\mathcal{G}}(A))}\right]} \\
&= \log \frac{p(A)}{\tilde{p}(A)} + \log \frac{p(\operatorname{pa}_{\mathcal{G}}(A))}{\tilde{p}(\operatorname{pa}_{\mathcal{G}}(A))} \\
&\log \frac{OR_{num}}{OR_{den}} - \log \frac{\widetilde{OR}_{num}}{\widetilde{OR}_{den}} \\
&+ \log \sum_{A, \operatorname{pa}_{\mathcal{G}}(A)} \tilde{p}(A) \frac{\widetilde{OR}_{num}}{\widetilde{OR}_{den}} \tilde{p}(\operatorname{pa}_{\mathcal{G}}(A)) \\
&- \log \sum_{A, \operatorname{pa}_{\mathcal{G}}(A)} p(A) \frac{OR_{num}}{OR_{den}} p(\operatorname{pa}_{\mathcal{G}}(A))
\end{aligned}$$

where we apply the Chen factorization for $p(A, \operatorname{pa}_{\mathcal{G}}(A))$ and $\tilde{p}(A, \operatorname{pa}_{\mathcal{G}}(A))$ and

$$\begin{aligned}
OR_{num} &= OR(A, \operatorname{pa}_{\mathcal{G}}(A)) OR(A, \operatorname{pa}_{\mathcal{G}}(A)^0) OR(A^0, \operatorname{pa}_{\mathcal{G}}(A)) \\
OR_{den} &= E[OR(A, \operatorname{pa}_{\mathcal{G}}(A)^0) | A^0] \times \\
&\quad E[OR(A^0, \operatorname{pa}_{\mathcal{G}}(A)) | \operatorname{pa}_{\mathcal{G}}(A)^0]
\end{aligned}$$

and analogously for the $\widetilde{OR}$'s.

Suppose we pick $\tilde{p}(A, \operatorname{pa}_{\mathcal{G}}(A)) = p(A)p(\operatorname{pa}_{\mathcal{G}}(A))$. Then $A \perp\!\!\!\perp \operatorname{pa}_{\mathcal{G}}(A)$ in $\tilde{p}$ and so $\frac{\widetilde{OR}_{num}/\widetilde{OR}_{den}}{\sum_{A, \operatorname{pa}_{\mathcal{G}}(A)} \tilde{p}(A) \frac{\widetilde{OR}_{num}}{\widetilde{OR}_{den}} \tilde{p}(\operatorname{pa}_{\mathcal{G}}(A))} = 1$. Thus, the previous expression simplifies to:

$$\log \frac{OR_{num}}{OR_{den}} - \log \sum_{A, \operatorname{pa}_{\mathcal{G}}(A)} p(A) \frac{OR_{num}}{OR_{den}} p(\operatorname{pa}_{\mathcal{G}}(A)) \tag{1}$$

Suppose we instead picked some *other* $\tilde{p}(A, \operatorname{pa}_{\mathcal{G}}(A)) = \tilde{p}(A)\tilde{p}(\operatorname{pa}_{\mathcal{G}}(A))$ (i.e. one in which $A \perp\!\!\!\perp \operatorname{pa}_{\mathcal{G}}(A)$). Then the above expression would have additional non-zero terms $\log \frac{p(A)}{\tilde{p}(A)} + \log \frac{p(\operatorname{pa}_{\mathcal{G}}(A))}{\tilde{p}(\operatorname{pa}_{\mathcal{G}}(A))}$. For this alternative $\tilde{p}$ to yield a lower KL divergence than that given by Eq. 1, one of the terms, $\log \frac{p(A)}{\tilde{p}(A)}$ or $\log \frac{p(\operatorname{pa}_{\mathcal{G}}(A))}{\tilde{p}(\operatorname{pa}_{\mathcal{G}}(A))}$, must be less than 0 (since the other terms in Eq. 1 remain the same under the independence of A and $\operatorname{pa}_{\mathcal{G}}(A)$). However, if $\log \frac{p(A)}{\tilde{p}(A)} < 0$ then the KL-divergence of $\tilde{p}(A)$ from $p(A)$ is negative, which violates Gibbs' inequality. The same holds for the distributions over $\operatorname{pa}_{\mathcal{G}}(A)$. We therefore can conclude that $\tilde{p}(A, \operatorname{pa}_{\mathcal{G}}(A)) = p(A)p(\operatorname{pa}_{\mathcal{G}}(A))$ is the KL-closest distribution to $p(A, \operatorname{pa}_{\mathcal{G}}(A))$ such that $A \perp\!\!\!\perp \operatorname{pa}_{\mathcal{G}}(A)$.

Now, as a subclaim, we prove: if $\tilde{p}$ is KL-closest to p, then any conditional obtained from $\tilde{p}$ (by dividing by some (potentially conditional) distribution p^* over a variable $V \in \mathbf{B}$), is KL-closest to the corresponding conditional obtained from p.

This is a simple consequence of the formula for KL-divergence:

$$D_{KL}\left(\frac{p}{p^*} || \frac{\tilde{p}}{p^*}\right) \propto \log \frac{\frac{p}{p^*}}{\frac{\tilde{p}}{p^*}} = \log \frac{p}{\tilde{p}}$$

The KL-divergence between the two distributions does not change by conditioning.

By the above two subclaims, the KL-closest distribution $\tilde{p}(A | \operatorname{pa}_{\mathcal{G}}(A))$ to $p(A | \operatorname{pa}_{\mathcal{G}}(A))$ is $p(A)$. By the local Markov property of DAGs, the claim holds, with $\tilde{p}(\mathbf{V} \setminus A) = p(\mathbf{V} \setminus A)$. □

Theorem 2 *Let $\mathbf{V}$ be a set of random variables with $p(\mathbf{V})$ corresponding to a DAG $\mathcal{G}$. Let $A \in \mathbf{V}$ and $\mathbf{B} \subseteq \mathbf{V}$ such that $\mathbf{B} \subseteq \operatorname{pa}_{\mathcal{G}}(A)$. Let $\mathcal{P}(\mathbf{V})$ be the set of probability distributions that factorize according to $\mathcal{G}$. Then*

$$\begin{aligned}
&p(A | \operatorname{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}) \prod_{V \in \mathbf{V} \setminus A} p(V | \operatorname{pa}_{\mathcal{G}}(V)) \\
&\quad = \arg\min_{\tilde{p} \in \mathcal{P}(\mathbf{V})} D_{KL}(p || \tilde{p}) \text{ s.t. } A \perp\!\!\!\perp \mathbf{B} | \operatorname{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}
\end{aligned}$$

Proof: We adapt the argument from Thm. 1. We can express the KL-divergence of $\tilde{p}(A, \operatorname{pa}_{\mathcal{G}}(A))$ from

$p(A, \mathrm{pa}_{\mathcal{G}})$ as proportional to:

$$\log \frac{p(A, \mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}) p(\mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})}{\tilde{p}(A, \mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}) \tilde{p}(\mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})}$$

$$= \log \frac{\left[\frac{p(A| \mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B}) \frac{OR_{num}}{OR_{den}} p(\mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B})}{\sum_{A,\mathbf{B}} p(A| \mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B}) \frac{OR_{num}}{OR_{den}} p(\mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B})}\right]}{\left[\frac{\tilde{p}(A| \mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B}) \frac{\widetilde{OR}_{num}}{\widetilde{OR}_{den}} \tilde{p}(\mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B})}{\sum_{A,\mathbf{B}} \tilde{p}(A| \mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B}) \frac{\widetilde{OR}_{num}}{\widetilde{OR}_{den}} \tilde{p}(\mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B})}\right]}$$

$$+ \log \frac{p(\mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})}{\tilde{p}(\mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})}$$

$$= \log \frac{p(A| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})}{\tilde{p}(A| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})} + \log \frac{p(\mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})}{\tilde{p}(\mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})}$$

$$+ \log \frac{OR_{num}}{OR_{den}} - \log \frac{\widetilde{OR}_{num}}{\widetilde{OR}_{den}}$$

$$+ \log \sum_{A,\mathbf{B}} \tilde{p}(A| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}) \frac{\widetilde{OR}_{num}}{\widetilde{OR}_{den}} \tilde{p}(\mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})$$

$$- \log \sum_{A,\mathbf{B}} p(A| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}) \frac{OR_{num}}{OR_{den}} p(\mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})$$

$$+ \log \frac{p(\mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})}{\tilde{p}(\mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})}$$

where we apply the Chen factorization for $p(A, \mathrm{pa}_{\mathcal{G}}(A))$ and $\tilde{p}(A, \mathrm{pa}_{\mathcal{G}}(A))$ and

$$\begin{aligned} OR_{num} &= OR(A, \mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}) \\ &\quad \times OR(A^0, \mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}) \\ &\quad \times OR(A, \mathbf{B}^0| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}) \\ OR_{den} &= E[OR(A, \mathbf{B}^0| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})|A^0, \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}] \\ &\quad \times E[OR(A^0, \mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})|\mathbf{B}^0, \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}] \end{aligned}$$

and analogously for the $\widetilde{OR}$'s.

Suppose we pick $\tilde{p}(A, \mathrm{pa}_{\mathcal{G}}(A)) = p(A| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}) p(\mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}) p(\mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})$. Then $A \perp\!\!\!\perp \mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}$ in $\tilde{p}$ and so $\frac{\widetilde{OR}_{num}/\widetilde{OR}_{den}}{\sum_{A,\mathbf{B}} \tilde{p}(A| \mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B}) \frac{\widetilde{OR}_{num}}{\widetilde{OR}_{den}} \tilde{p}(\mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B})} = 1$. Thus, the previous expression simplifies to:

$$\log \frac{OR_{num}}{OR_{den}} - \log \sum_{A,\mathbf{B}} p(A| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}) \frac{OR_{num}}{OR_{den}} p(\mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}) \tag{2}$$

Suppose we instead picked some *other* $\tilde{p}(A, \mathrm{pa}_{\mathcal{G}}(A)) = \tilde{p}(A)\tilde{p}(\mathrm{pa}_{\mathcal{G}}(A))$ (i.e. one in which $A \perp\!\!\!\perp \mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}$). Then the above expression would have additional non-zero terms $\log \frac{p(A| \mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B})}{\tilde{p}(A| \mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B})} + \log \frac{p(\mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B})}{\tilde{p}(\mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B})} + \log \frac{p(\mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B})}{\tilde{p}(\mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B})}$. For this alternative $\tilde{p}$ to yield a lower KL divergence than that given by Eq. 2, one of the terms in the above sum must be less than 0 (since the other terms in Eq. 2 remain the same under the conditional independence of A and $\mathbf{B}$). However, if $\log \frac{p(A| \mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B})}{\tilde{p}(A| \mathrm{pa}_{\mathcal{G}}(A)\setminus\mathbf{B})} < 0$ then the KL-divergence of $\tilde{p}(A| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})$ from $p(A| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})$ is negative, which violates Gibbs' inequality. The same holds for the distributions over $\mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}$ and $\mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}$. We therefore can conclude that $\tilde{p}(A, \mathrm{pa}_{\mathcal{G}}(A)) = p(A| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}) p(\mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}) p(\mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})$ is the KL-closest distribution to $p(A, \mathrm{pa}_{\mathcal{G}}(A))$ such that $A \perp\!\!\!\perp \mathbf{B}| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B}$.

By the above argument and application of the conditioning argument in Thm. 1, the KL-closest distribution $\tilde{p}(A| \mathrm{pa}_{\mathcal{G}}(A))$ to $p(A| \mathrm{pa}_{\mathcal{G}}(A))$ satisfying the necessary independence constraint is $p(A| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathbf{B})$. By the local Markov property, the result is immediate since if chose $\tilde{p} = p$ for variables $\mathbf{V} \setminus \{A, \mathrm{pa}_{\mathcal{G}}(A)\}$.

□

Theorem 3 *Let* $\mathbf{V}$ *be a set of random variables with* $p(\mathbf{V})$ *corresponding to a DAG* $\mathcal{G}$*. Let* $\mathbf{A} \in \mathbf{V}$ *and for each* $A \in \mathbf{A}$ *define* $\mathrm{In}(A) \subseteq \mathrm{pa}_{\mathcal{G}}(A)$*, the set of parents of* A *whose edges into* A *we wish to remove. Let* $\mathcal{P}(\mathbf{V})$ *be the set of probability distributions that factorize according to* $\mathcal{G}$*. Then*

$$\begin{aligned} &\prod_{A \in \mathbf{A}} p(A| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathrm{In}(A)) \prod_{V \in \mathbf{V} \setminus \mathbf{A}} p(V| \mathrm{pa}_{\mathcal{G}}(V)) \\ &\quad = \arg\min_{\tilde{p} \in \mathcal{P}(\mathbf{V})} D_{KL}(p||\tilde{p}) \\ &\quad\quad \textit{s.t. } A \perp\!\!\!\perp \mathrm{In}(A)| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathrm{In}(A) \;\; \forall A \in \mathbf{A} \end{aligned}$$

Proof: We prove the claim inductively. When $|\mathbf{A}| = 1$, the claim holds trivially by application of Thm 2.

Suppose $|\mathbf{A}| > 1$. Impose a reverse topological ordering $\prec$ on $\mathbf{V}$ (e.g. variables have higher indexes in the ordering than their parents). This ordering assumption is not necessary to prove the claim, however it helps simplify the argument.

Suppose that for some $\mathbf{A}' \subset \mathbf{A}$, where every $A \in \mathbf{A}'$ precedes every $A^* \in \mathbf{A} \setminus \mathbf{A}'$ in $\prec$, we know

$$\begin{aligned} \tilde{p}(\mathbf{V}) = &\prod_{A \in \mathbf{A}'} p(A| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathrm{In}(A)) \\ &\times \prod_{V \in \mathbf{V} \setminus \mathbf{A}'} p(V| \mathrm{pa}_{\mathcal{G}}(V)) \end{aligned} \tag{3}$$

is the KL-closest distribution to $p(\mathbf{V})$ which satisfies $A \perp\!\!\!\perp \mathrm{In}(A)| \mathrm{pa}_{\mathcal{G}}(A) \setminus \mathrm{In}(A)$ for all $A \in \mathbf{A}'$. Then it suffices to show for some $A^* \in \mathbf{A} \setminus \mathbf{A}'$ that

$$\tilde{p}(\mathbf{V}) = \prod_{A \in (\mathbf{A}' \cup A^*)} p(A | \operatorname{pa}_{\mathcal{G}}(A) \setminus \operatorname{In}(A)) \times \prod_{V \in \mathbf{V} \setminus (\mathbf{A}' \cup A^*)} p(V | \operatorname{pa}_{\mathcal{G}}(V))$$

is the KL-closest distribution to $p(\mathbf{V})$ that satisfies $A \perp\!\!\!\perp \operatorname{In}(A) | \operatorname{pa}_{\mathcal{G}}(A) \setminus \operatorname{In}(A)$ for all $A \in \mathbf{A}' \cup A^*$.

We can factorize p (and analogously $\tilde{p}$) by chain rule:

$$p(\mathbf{V}) = p(A^*, \operatorname{In}(A^*) | \operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*)) \times p(\operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*)) \times p(\mathbf{V} \setminus (A^* \cup \operatorname{pa}_{\mathcal{G}}(A^*)))$$

By application of Cor. 1, we can re-write the first term as $\frac{X_p}{Y_p}$, where:

$$X_p = p(A^* | \operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*)) \times \frac{OR_{num}}{OR_{den}} \times p(\operatorname{In}(A^*) | \operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*))$$

and $Y_p = \sum_{A^*, \operatorname{In}(A^*)} X_p$, and analogously for $X_{\tilde{p}}$ and $Y_{\tilde{p}}$. Similar to previous arguments, for notational simplicity, we use the shorthands $OR_{num} =$

$$OR(A^\star, \operatorname{In}(A^\star) | \operatorname{pa}_{\mathcal{G}}(A^\star \setminus \operatorname{In}(A^\star), \mathbf{V} \setminus (A^\star \cup \operatorname{pa}_{\mathcal{G}}(A^\star))) \times OR(A^{\star 0}, \operatorname{In}(A^\star) | \operatorname{pa}_{\mathcal{G}}(A^\star \setminus \operatorname{In}(A^\star), \mathbf{V} \setminus (A^\star \cup \operatorname{pa}_{\mathcal{G}}(A^\star))) \times OR(A^\star, \operatorname{In}(A^\star)^0 | \operatorname{pa}_{\mathcal{G}}(A^\star \setminus \operatorname{In}(A^\star), \mathbf{V} \setminus (A^\star \cup \operatorname{pa}_{\mathcal{G}}(A^\star)))$$

and $OR_{den} =$

$$E\Big[OR(A^\star, \operatorname{In}(A^\star)^0) | A^{\star 0}, \operatorname{pa}_{\mathcal{G}}(A^\star \setminus \operatorname{In}(A^\star), \mathbf{V} \setminus (A^\star \cup \operatorname{pa}_{\mathcal{G}}(A^\star))\Big] \times E\Big[OR(A^{\star 0}, \operatorname{In}(A^\star)) | \operatorname{In}(A^\star)^0, \operatorname{pa}_{\mathcal{G}}(A^\star \setminus \operatorname{In}(A^\star), \mathbf{V} \setminus (A^\star \cup \operatorname{pa}_{\mathcal{G}}(A^\star))\Big]$$

(analogously $\widetilde{OR}_{num}$ and $\widetilde{OR}_{den}$).

As before, we can express the KL-divergence from p to $\tilde{p}$ as proportional to:

$$\log \frac{p(A^*, \operatorname{pa}_{\mathcal{G}}(A^*)) p(\mathbf{V} \setminus (A^* \cup \operatorname{pa}_{\mathcal{G}}(A^*)))}{\tilde{p}(A^*, \operatorname{pa}_{\mathcal{G}}(A^*)) \tilde{p}(\mathbf{V} \setminus (A^* \cup \operatorname{pa}_{\mathcal{G}}(A^*)))}$$
$$= \Big[\log \frac{p(A^* | \operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*))}{\tilde{p}(A^* | \operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*))} + \log \frac{p(\operatorname{In}(A^*) | \operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*))}{\tilde{p}(\operatorname{In}(A^*) | \operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*))}$$
$$+ \log \frac{OR_{num}}{OR_{den}} - \log \sum_{A^*, \operatorname{In}(A^*)} \Big[p(A^* | \operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*)) \times \frac{OR_{num}}{OR_{den}} p(\operatorname{In}(A^*) | \operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*))\Big]$$
$$- \frac{\widetilde{OR}_{num}}{\widetilde{OR}_{den}} + \log \sum_{A^*, \operatorname{In}(A^*)} \Big[\tilde{p}(A^* | \operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*)) \times \frac{\widetilde{OR}_{num}}{\widetilde{OR}_{den}} \tilde{p}(\operatorname{In}(A^*) | \operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*))\Big]$$
$$+ \log \frac{p(\operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*))}{\tilde{p}(\operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*))} + \log \frac{p(\mathbf{V} \setminus (A^* \cup \operatorname{pa}_{\mathcal{G}}(A^*)))}{\tilde{p}(\mathbf{V} \setminus (A^* \cup \operatorname{pa}_{\mathcal{G}}(A^*)))}\Big]$$

Suppose we let $\tilde{p}(A^* | \operatorname{pa}_{\mathcal{G}}(A^*)) = p(A^* | \operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*))$ and $\tilde{p}(\operatorname{In}(A^*) | \operatorname{pa}_{\mathcal{G}}(A^* \setminus \operatorname{In}(A^*))) = p(\operatorname{In}(A^*) | \operatorname{pa}_{\mathcal{G}}(A^* \setminus \operatorname{In}(A^*)))$. Then, similar to the previous theorems, we induce conditional independence between $A^\star$ and $\operatorname{In}(A^\star)$ given $A^\star$'s other parents $\operatorname{pa}_{\mathcal{G}}(A^\star) \setminus \operatorname{In}(A^\star)$. In turn, the above expression simplifies to the following:

$$= \log \frac{OR_{num}}{OR_{den}} - \log \sum_{A^*, \operatorname{In}(A^*)} \Big[p(A^* | \operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*)) \times \frac{OR_{num}}{OR_{den}} p(\operatorname{In}(A^*) | \operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*))\Big]$$
$$+ \log \frac{p(\operatorname{pa}_{\mathcal{G}}(A^\star) \setminus \operatorname{In}(A^\star)}{\tilde{p}(\operatorname{pa}_{\mathcal{G}}(A^\star) \setminus \operatorname{In}(A^\star)} + \log \frac{p(\mathbf{V} \setminus (A^* \cup \operatorname{pa}_{\mathcal{G}}(A^*)))}{\tilde{p}(\mathbf{V} \setminus (A^* \cup \operatorname{pa}_{\mathcal{G}}(A^*)))}$$

Under the assumption of a topological ordering $\prec$, choosing this choice of $\tilde{p}$ does not affect whether the constraint $A \perp\!\!\!\perp \operatorname{In}(A) | \operatorname{pa}_{\mathcal{G}}(A)$ for $A \in \mathbf{A}' \cup A^*$ holds. This is because of the assumption made in Eq. 3. As a consequence, the ratio of terms with respect to $\operatorname{pa}_{\mathcal{G}}(A^*) \setminus \operatorname{In}(A^*)$ cancels in equality 1 above, leaving us with:

$$= \log \frac{OR_{num}}{OR_{den}} - \log \sum_{A^*, \mathrm{In}(A^*)} \Big[p(A^* | \,\mathrm{pa}_{\mathcal{G}}(A^*) \setminus \mathrm{In}(A^*)) \times \frac{OR_{num}}{OR_{den}} p(\mathrm{In}(A^*) | \,\mathrm{pa}_{\mathcal{G}}(A^*) \setminus \mathrm{In}(A^*)) \Big] + \log \frac{p(\mathbf{V} \setminus (A^* \cup \mathrm{pa}_{\mathcal{G}}(A^*)))}{\tilde{p}(\mathbf{V} \setminus (A^* \cup \mathrm{pa}_{\mathcal{G}}(A^*)))}$$

Now suppose we wish find a p^* that yields a lower KL-divergence, corresponding to decrease the quantity in the above expression by changing $\tilde{p}$ for one or more of the terms. By application of the argument in Thm. 2, changing $\tilde{p}(A^* | \,\mathrm{pa}_{\mathcal{G}}(A^*))$ to a function other than $p(A^* | \,\mathrm{pa}_{\mathcal{G}}(A^*) \setminus \mathrm{In}(A))$ would necessarily violate Gibbs' inequality. So, we must consider changing $\tilde{p}(V | \,\mathrm{pa}_{\mathcal{G}}(V))$ for some $V \in \mathbf{V} \setminus (A^* \cup \mathrm{pa}_{\mathcal{G}}(A^*))$.

If $V \in \mathrm{nd}_{\mathcal{G}}(A^*)$ then, by the local Markov property of DAGs, $A^* \perp\!\!\!\perp V | \,\mathrm{pa}_{\mathcal{G}}(A^*) \setminus \mathrm{In}(A^*)$ and so choosing $\tilde{p}(V | \,\mathrm{pa}_{\mathcal{G}}(V)) = p(V | \,\mathrm{pa}_{\mathcal{G}}(V))$ will maintain the necessary independence constraints and ensure that the term for V has 0 contribution to the KL quantity above. That is $\log p(V | \,\mathrm{pa}_{\mathcal{G}}(V)) - \log \tilde{p}(V | \,\mathrm{pa}_{\mathcal{G}}(V)) = 0$. Changing $\tilde{p}(V | \,\mathrm{pa}_{\mathcal{G}}(V))$ will therefore not move $\tilde{p}$ closer to p.

If, on the other hand, $V \in \mathrm{de}_{\mathcal{G}}(A^*)$, then $\mathrm{In}(V)$ is either empty or non-empty. If $\mathrm{In}(V) = \emptyset$ then by the same argument as for $V \in \mathrm{nd}_{\mathcal{G}}(A^*)$, choosing $\tilde{p}(V | \,\mathrm{pa}_{\mathcal{G}}(V)) = p(V | \,\mathrm{pa}_{\mathcal{G}}(V))$ will ensure that the necessary constraints hold and that V's contribution the KL-divergence expression will be 0. If $\mathrm{In}(V) \neq \emptyset$, then $\tilde{p}(V | \,\mathrm{pa}_{\mathcal{G}}(V))$ was already set to be $p(V | \,\mathrm{pa}_{\mathcal{G}}(V) \setminus \mathrm{In}(V))$ by the assumption in Eq. 3. By the argument in Thm. 2, changing this setting of $\tilde{p}$ would violate Gibbs' inequality.

By the above argument, as well as the argument given in Thm. 1 that states that applying conditioning to two distributions doesn't affect their KL-divergence, we have shown that

$$\prod_{A \in (\mathbf{A}' \cup A^*)} p(A | \,\mathrm{pa}_{\mathcal{G}}(A) \setminus \mathrm{In}(A)) \prod_{V \in \mathbf{V} \setminus (\mathbf{A}' \cup A^*)} p(V | \,\mathrm{pa}_{\mathcal{G}}(V))$$

is the KL-closest distribution to p which satisfies $A \perp\!\!\!\perp \mathrm{In}(A) | \,\mathrm{pa}_{\mathcal{G}}(A) \setminus \mathrm{In}(A)$ for all $A \in \mathbf{A}' \cup A^*$. By induction, the claim of the theorem for $\mathbf{A}$ follows immediately. □

Theorem 4 *Let $\mathbf{V}$ be a set of random variables with $p(\mathbf{V})$ corresponding to a DAG $\mathcal{G}$. Let $\mathbf{A} \subseteq \mathbf{V}$ and assume that for some $\mathbf{a}$ we have $p(\mathbf{A} = \mathbf{a}) > 0$. Let $\mathcal{P}(\mathbf{V})$ be the set of probability distributions that factorize according to $\mathcal{G}$. Then*

$$\prod_{V \in \mathbf{V} \setminus \mathbf{A}} p(V | \,\mathrm{pa}_{\mathcal{G}}(V))|_{\mathbf{A}=\mathbf{a}} = \operatorname*{arg\,min}_{\tilde{p} \in \mathcal{P}(\mathbf{V})} D_{KL}(p || \tilde{p})$$
$$\text{s.t. } \tilde{p}(A_i | \,\mathrm{nd}_{\mathcal{G}}(A_i)) = I(A_i = a_i) \, \forall i \in [|\mathbf{A}|]$$

where $[|\mathbf{A}|] = \{1, \ldots, |\mathbf{A}|\}$.

Proof: This is a simple consequence of Thm. 3. For each $A \in \mathbf{A}$, if we let $\mathrm{In}(A) = \mathrm{pa}_{\mathcal{G}}$, then by the local Markov property, we have

$$\prod_{A \in \mathbf{A}} p(A) \prod_{V \in \mathbf{V} \setminus \mathbf{A}} p(V | \,\mathrm{pa}_{\mathcal{G}}(V))$$

is the KL-closest distribution to $p(\mathbf{V})$ that satisfies $A \perp\!\!\!\perp \mathrm{nd}_{\mathcal{G}}(A)$ for all $A \in \mathbf{A}$. Now, by previous arguments, replacing each $p(A)$ with $I(A_i = a_i)$ maintains the KL-closeness of $\tilde{p}$ since $\tilde{p}(\mathbf{V} \setminus \mathbf{A}) = p(\mathbf{V} \setminus \mathbf{A})$ and now the required constraint holds. Since we are replacing each $p(A)$ with an indicator, this is equivalent to just evaluating $p(\mathbf{V} \setminus \mathbf{A})$ with $\mathbf{A} = \mathbf{a}$:

$$\prod_{V \in \mathbf{V} \setminus \mathbf{A}} p(V | \,\mathrm{pa}_{\mathcal{G}}(V))|_{\mathbf{A}=\mathbf{a}}$$

□

We extend the above result to the case of edge interventions. To simplify our argument, we formulate this theorem in terms of extended graphs which are inspired by [7] and requires the following additional background notation [5]:

For a set of variables $\mathbf{A}$ and a set of edges α out of $\mathbf{A}$, define for each $A_i \in \mathbf{A}$, the synthetic nodes $A_i^{ch} = \{A_i^j | V_j \in \mathrm{ch}_{\mathcal{G}}(A_i)\}$. That is, for each $V_j \in \mathrm{ch}_{\mathcal{G}}(A_i)$, we create a synthetic node A_i^j. Let $\mathbf{A}^{ch} = \bigcup_{A_i \in \mathbf{A}}$.

Define the extended graph of $\mathcal{G}(\mathbf{V})$, denoted $\mathcal{G}^e(\mathbf{V} \cup \mathbf{A}^{ch})$, as the graph obtained by adding the synthetic A_i^j's to $\mathcal{G}$ with edges $A_i \to A_i^j \to V_j$ if and only if $A_i \to V_j$ appears in $\mathcal{G}$. The relationship for each edge of type $A_i \to A_i^j$ as assumed to be deterministic. Following [5], $\mathcal{G}^e(\mathbf{V} \cup \mathbf{A}^{ch})$ is a valid DAG under the structural equation model assumption.

Theorem 1 *Let $\mathbf{V}$ be a set of random variables with $p(\mathbf{V})$ corresponding to a DAG $\mathcal{G}$. Let α be a set of edges in $\mathcal{G}$ and let $\mathbf{A}_\alpha = \{A | (AB)_\to \in \alpha\} \subseteq \mathbf{V}$. For the corresponding $\mathbf{A}^{\mathrm{ch}}$ and $\mathcal{G}^e(\mathbf{V} \cup \mathbf{A}^{\mathrm{ch}})$, if we let $\mathcal{P}^e(\mathbf{V})$ be the set of probability distributions that factorize according to $\mathcal{G}^e$ and assume for some $\mathbf{a}^{\mathrm{ch}}$, $p(\mathbf{A}^{\mathrm{ch}} = \mathbf{a}^{\mathrm{ch}}) > 0$ then,*

$$\prod_{V \in \mathbf{V}} p^e(V | \,\mathrm{pa}_{\mathcal{G}^e}(V)) = \operatorname*{arg\,min}_{\tilde{p}^e \in \mathcal{P}^e(\mathbf{V})} D_{KL}(p^e || \tilde{p}^e) \text{ s.t.}$$
$$\tilde{p}^e(A_i | \,\mathrm{nd}_{\mathcal{G}^e}(A_i)) = I(A_i = a_i) \text{ for } i = \{1, \ldots, |\mathbf{A}^{\mathrm{ch}}|\}$$

Proof: This result follows directly from Thm. 4. By re-expressing $\mathcal{G}$ as $\mathcal{G}^e$, the intervention is no longer in terms of a set of edges α but rather a set of nodes $\mathbf{A}^{ch}$. We can simply apply the result of Thm. 4 where $\mathbf{A}^{ch}$ corresponds to the set of nodes $\mathbf{A}$ for which we are inducing independence with their non-descendants. □

C Experimental Setup

The models for C, A, and Y are parametrized by τ_C, $\tau_A = [\tau_{A_0}, \tau_{A_C}, \tau_{A_{C_\mathcal{N}}}]$, and $\tau_Y = [\tau_{Y_0}, \tau_{Y_A}, \tau_{Y_C}, \tau_{Y_{A_\mathcal{N}}}, \tau_{Y_{C_\mathcal{N}}}]$, specified in Table 1. C is a 3-dimensional vector, with each component C^l drawn from a Bernoulli distribution with probability τ_{C^l}; A and Y are generated using the following parametric models:

$$\begin{aligned} p(A_i = 1 | C_i, \{C_j | j \in \mathcal{N}_i\}; \tau_A) \\ = expit\Big(\tau_{A_0} + \tau_{A_C} \cdot C_i + \sum_{j \in \mathcal{N}_i} \tau_{A_{C_\mathcal{N}}} \cdot C_j\Big) \\ p(Y_i = 1 | C_i, A_i, \{C_j, A_j | j \in \mathcal{N}_i\}; \tau_Y) \\ = expit\Big(\tau_{Y_0} + \tau_{Y_A} A_i + \tau_{Y_C} \cdot C_i \\ + \sum_{j \in \mathcal{N}_i} \big(\tau_{Y_{C_\mathcal{N}}} \cdot C_j + \tau_{Y_{A_\mathcal{N}}} \cdot A_j\big)\Big) \end{aligned}$$

Parameter	Value
τ_C	[.7, .3, .5]
τ_A	[1, 3, .15, .2, .1, .15, .15]
τ_Y	[2.5, 1.2, -1, 1.2, .2, -.13, -1, -.2, -.3]

Table 1: Parameters for data generation in simulation studies

For the first experiment, we use the Erdős-Rényi (with attachment probability $p = .05$), Barabasi-Albert (with a preferential attachment edge count of 4), and Watts-Strogatz (with nearest neighbor attachment of 4 and an edge re-wiring probability of .25) network generators.

Estimation Details

For homogeneous connections, estimating the post-intervention value of Y_i is done by simply adding the connecting unit to $\mathcal{N}_i$ for the sake of forming covariate vectors on which we perform inference.

For known policy interventions, we consider adding a weight to the terms associated with the added neighbor. This corresponds, for instance, to joining one unit gaining an addition connection on a social media service and also algorithmically promoting the content of new neighbor unit. To estimate Y_i here, we simply multiply the new neighbor's variables by the known weight and form covariate vectors as in the homogeneous case. In our simulations we use a weight of 1.2.

Finally, for unknown policy interventions, we repeat the process for known policies where a weight is added to the adjoined neighbors. Here, however, we choose the weight by maximizing a function $g(Y_i, Y_j) = min(\frac{Y_i + Y_j}{2}, .3)$. This corresponds to ensuring we satisfy a 'worst-case' scenario for the outcomes of the newly joined neighbors. We chose .3 as the floor for this function since the mean Y in our data-generating process was .395 and we wanted to simulate not making one unit better off at the expense of making the other substantially worse off. We chose optimal parameters using standard optimization software [4].

Stochastic severance interventions are estimated analogously to homogeneous connections. We remove the terms relating to the severed connection from the feature vector for predicting A_i and Y_i and perform inference using our logistic regression models. Severance interventions performed with interventional values for the severed neighbor's C and A values are estimated by simply replacing the variables in the Monte Carlo sampling procedure with the interventional values according to the g-formula [6]. For our simulations we chose the cross-unit interventional values for C_j and A_j to be 0 and 1 respectively. In either case, when a unit has no pre-intervention neighbors, the estimate of their outcome is the same in both the pre- and post-intervention worlds.

Extended Results

Network Size	Bias CI
4	(-.0082, .0051)
8	(-.0088, .0011)
16	(-.0057, .0006)
32	(-.0009, .0040)
64	(-.0009, .0016)

Table 2: 95% confidence intervals for the bias of estimates of stochastic severance task with varied network sizes.

Attachment Prob.	Bias CI
.01	(-.0008, .0017)
.05	(-.0018, .0026)
.10	(-.0043, .0003)
.15	(-.0014, .0016)
.20	(-.0011, .0004)

Table 3: 95% confidence intervals for the bias of estimates of stochastic severance task with varied Erdős-Rényi attachment probabilities.

Sample Size	Bias CI
10	(-.0278, .0178)
100	(-.0075, .0057)
1,000	(-.0007, .0028)
10,000	(-.0001, .0006)

Table 4: 95% confidence intervals for the bias of estimates of stochastic severance task with varied sample sizes.

References

[1] H. Y. Chen. A semiparametric odds ratio model for measuring association. *biometrics*, 63:413–421, 2007.

[2] R. Chetty, N. Hendren, and L. F. Katz. The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902, 2016.

[3] M. G. Hudgens and M. E. Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, 2008.

[4] E. Jones, T. Oliphant, and P. Peterson. ${$scipy$}$: open source scientific tools for ${$python$}$. 2014.

[5] D. Malinsky, I. Shpitser, and T. Richardson. A potential outcomes calculus for identifying conditional path-specific effects. In K. Chaudhuri and M. Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 3080–3088. PMLR, 16–18 Apr 2019.

[6] J. 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.

[7] J. Robins, T. Richardson, and P. Spirtes. On identification and inference for direct effects. *Epidemiology*, 2009.