Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23)
Sister Conferences Best Papers Track

Learning Causal Effects on Hypergraphs (Extended Abstract)*

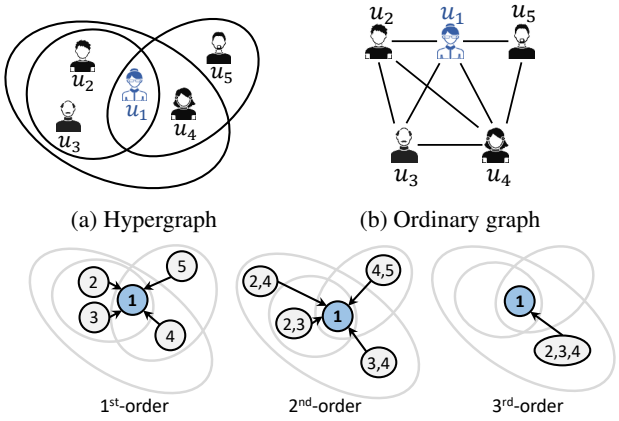
Jing Ma[1], **Mengting Wan**[2], **Longqi Yang**[2], **Jundong Li**[1], **Brent Hecht**[2], **Jaime Teevan**[2]
[1]University of Virginia, Charlottesville, VA, USA
[2]Microsoft, Redmond, WA, USA
{jm3mr, jundong}@virginia.edu, {mengting.wan, longqi.yang, brent.hecht, teevan}@microsoft.com

Abstract

Hypergraphs provide an effective abstraction for modeling multi-way group interactions among nodes, where each hyperedge can connect any number of nodes. Different from most existing studies which leverage *statistical dependencies*, we study hypergraphs from the perspective of *causality*. Specifically, we focus on the problem of individual treatment effect (ITE) estimation on hypergraphs, aiming to estimate how much an intervention (e.g., wearing face covering) would causally affect an outcome (e.g., COVID-19 infection) of each individual node. Existing works on ITE estimation either assume that the outcome of one individual should not be influenced by the treatment of other individuals (i.e., no *interference*), or assume the interference only exists between connected individuals in an ordinary graph. We argue that these assumptions can be unrealistic on real-world hypergraphs, where higher-order interference can affect the ITE estimations due to group interactions. We investigate high-order interference modeling, and propose a new causality learning framework powered by hypergraph neural networks. Extensive experiments on real-world hypergraphs verify the superiority of our framework over existing baselines.

1 Introduction

Group interactions among individuals exist in many scenarios, e.g., massive gathering events. Although the conventional pairwise graph definition covers various applications (e.g., physical contact networks or social networks [Mitchell, 1974]), it fails to capture the complete information of group interactions (where each interaction may involve more than two individuals) [Bai *et al.*, 2021; Feng *et al.*, 2019]. *Hypergraphs* can be introduced to address this limitation. Consider a hypergraph example that individuals are connected via in-person social events, each event can be represented as a *hyperedge* (Fig. 1a). Each hyperedge can connect any number of individuals, unlike an ordinary edge connecting exactly two nodes (Fig. 1b).



(a) Hypergraph (b) Ordinary graph

(c) First, second, and third-order interferences with u_1.

Figure 1: (a) An illustrative example of group interactions on a hypergraph, where each circle represents a hyperedge (group); (b) An ordinary graph projected from this hypergraph; (c) Illustration of interferences with u_1 from its neighbors on the hypergraph. Note interference on (b) is pairwise (first-order only) while higher-order interference exists on the original hypergraph (a).

While many studies have been devoted to utilizing such a generalized hypergraph structure to facilitate machine learning tasks [Bai *et al.*, 2021; Feng *et al.*, 2019], the majority were still executed at the statistical correlation level. A critical limitation here is the lack of *causality*, which is particularly important for understanding the impact of a policy intervention (e.g., wearing face covering) on an outcome of interest (e.g., COVID-19 infection). For individuals connected as in Fig. 1a, one may ask "how would each individual's face covering practice (treatment) *causally* influence their infection risk (outcome)?" This causal problem is particularly hard on hypergraphs, since the outcome of each individual is not only affected by their own treatment, but also influenced by the treatment of other individuals on the hypergraph (e.g., face covering practice of other individuals who may physically contact the target individual through a gathering event).

We focus on learning causal effects on hypergraphs. Specifically, we aim to estimate the individual treatment effect (ITE) under hypergraph interference from observational data. Classic ITE estimation relies on the Stable Unit Treat-

*This is an extended abstract of a paper [Ma *et al.*, 2022] that won the Best Paper Award at KDD 2022.

ment Value (SUTVA) assumption [Fisher, 1936; Splawa-Neyman *et al.*, 1990] that there is no interference [Tchetgen and VanderWeele, 2012; Hudgens and Halloran, 2008] (i.e., spillover effect) among individuals/instances (also referred to as *units* in causal inference literature). That means the outcomes for any instance are not affected by the treatment of other instances. This assumption is often impractical in the real world, especially on graphs where the interference is ubiquitous [Ahluwalia *et al.*, 2001; Yilmaz *et al.*, 2002]. Most existing studies of interference [Aronow and Samii, 2017; Basse and Feller, 2018; Imai *et al.*, 2020; Kohavi *et al.*, 2013; Tchetgen and VanderWeele, 2012; Ugander *et al.*, 2013; Yuan *et al.*, 2021; Ma and Tresp, 2021] assume the interference only exists in a pairwise way on ordinary graphs. This pairwise interference notion is insufficient to characterize the high-order interference on hypergraphs. As shown in Fig. 1c, within a gathering event (hyperedge) between u_1, u_2 and u_3, an individual's (u_1) infection outcome can be affected by the *first-order* interference from other individuals ($u_2 \rightarrow u_1$ and $u_3 \rightarrow u_1$) as well as the *high-order* interference from the interactions among other individuals (the interaction between u_2 and u_3 may also act on influencing the exposure of the virus to u_1; consequently, u_1's infection risk can be affected by this second-order interaction effect, i.e., $u_2 \times u_3 \rightarrow u_1$). This demands techniques capable of modeling high-order interference, but to the best of our knowledge, very little work has been done in this area.

We propose a novel framework— **C**ausal **I**nference under **S**pillover Effects in **Hyper**graphs (***HyperSCI***)—to model high-order interference and promote ITE estimation performance. At a high level, this framework controls for the confounders and models high-order interference based on representation learning, then estimates the outcomes based on the learned representations. We evaluate **HyperSCI** through extensive experiments under different scenarios of high-order interference, and provide in-depth analysis of how our framework acts on different nodes and hyperedges.

2 Problem Definition

Definition 1. *(Hypergraph) A **hypergraph** $\mathcal{H} = \{\mathcal{V}, \mathcal{E}\}$ consists of a set of n nodes $\mathcal{V} = \{v_i\}_{i=1}^n$ and a set of m hyperedges $\mathcal{E} = \{\mathbf{e}_k\}_{k=1}^m$. Each hyperedge can connect any number of nodes.*

In the studied problem, the given observational data is denoted by $\{\mathbf{X}, \mathcal{H}, \mathbf{T}, \mathbf{Y}\}$. Here, $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$, $\mathbf{T} = \{t_i\}_{i=1}^n$ and $\mathbf{Y} = \{y_i\}_{i=1}^n$ represent node features, treatment assignments, and observed outcomes, respectively. $\mathbf{H} = \{h_{i,e}\} \in \mathbb{R}^{n \times m}$ is an incidence matrix for hypergraph $\mathcal{H}$. Here, $h_{i,e} = 1$ if node i is in hyperedge e, otherwise $h_{i,e} = 0$. The treatment assignment for each node is binary (i.e., $t_i \in \{0, 1\}$). We use H, X, T to denote the random variables for the hypergraph structure, features, and treatment for any node.

Definition 2. *(Potential outcome) The **potential outcome** [Rubin, 1980] of the unit i (denoted by y_i^1 or y_i^0) is defined as the outcome which would be realized for unit i under treatment $t_i = 1$ or $t_i = 0$. These potential outcomes can be obtained via a transformation function $Y_i^{T_i} = \Phi_Y(T_i, X_i, T_{-i}, X_{-i}, H)$. Here, Φ_Y is a (non-deterministic) function, i.e., $y_i^{t_i} = \Phi_Y(t_i, \mathbf{x}_i, \mathbf{T}_{-i}, \mathbf{X}_{-i}, \mathbf{H})$, where $(\cdot)_{-i}$ denotes all other nodes on $\mathcal{H}$ except i.*

Our aim is to estimate ITE in a hypergraph. The ITE in the studied problem is defined as follows:

Definition 3. *For each node i on the hypergraph $\mathcal{H}$, the **individual treatment effect** (ITE) $\tau(\mathbf{x}_i, \mathbf{T}_{-i}, \mathbf{X}_{-i}, \mathbf{H})$ is defined by the difference between potential outcomes corresponding to $t_i = 1$ and $t_i = 0$:*

$$\begin{aligned}&\tau(\mathbf{x}_i, \mathbf{T}_{-i}, \mathbf{X}_{-i}, \mathbf{H})\\ =&\mathbb{E}[Y_i^1 - Y_i^0 | X_i = \mathbf{x}_i, T_{-i} = \mathbf{T}_{-i}, X_{-i} = \mathbf{X}_{-i}, H = \mathbf{H}]\\ =&\mathbb{E}[\Phi_Y(1, \mathbf{x}_i, \mathbf{T}_{-i}, \mathbf{X}_{-i}, \mathbf{H}) - \Phi_Y(0, \mathbf{x}_i, \mathbf{T}_{-i}, \mathbf{X}_{-i}, \mathbf{H})].\end{aligned} \tag{1}$$

Here, we give our main assumptions and theoretical analysis for this study:

Assumption 1. *(Expressiveness of summary function) There is an expressive summary function $\mathbf{o}_i = \texttt{SMR}(\mathbf{H}, \mathbf{T}_{-i}, \mathbf{X}_{-i})$ to summarize the environmental information for each node. For any node i, any values of H, X_{-i}, and T_{-i}, if the output $\mathbf{o}_i$ is determined, then the value of the potential outcomes y_i^1 and y_i^0 with feature $\mathbf{x}_i$ are also determined.*

Assumption 2. *(Unconfoundedness) For any node i, given the node features, the potential outcomes are independent with the treatment assignment and summary of neighbors, i.e., $Y_i^1, Y_i^0 \perp\!\!\!\perp T_i, O_i | X_i$.*

Theorem 1. *(Identifiability) The defined ITE can be identifiable from observational data under the above assumptions.*

We refer readers to [Ma *et al.*, 2022] for a full derivation.

3 Proposed Method

Fig. 2 shows an overview of our framework HyperSCI, which contains three main components: confounder representation learning, interference modeling, and outcome prediction.

3.1 Confounder Representation Learning

HyperSCI learns representations of confounders by mapping the node features $\mathbf{x}_i$ into a latent space with a multilayer perceptron (MLP) module, i.e., $\mathbf{z}_i = \text{MLP}(\mathbf{x}_i)$. The confounder representations for all the nodes are denoted by $\mathbf{Z} = \{\mathbf{z}_i\}_{i=1}^n$. Similar as [Shalit *et al.*, 2017], a Wasserstein-1 distance-based representation balancing method is used to minimize the distance between the representation distributions of the treatment group and control group, aiming to improve the treatment effect estimation performance.

3.2 Interference Modeling

An interference modeling module is developed to model the high-order interference among nodes in the hypergraph. More specifically, a function $\Psi(\cdot)$ is learned via a hypergraph neural network module to obtain the interference representations ($\mathbf{p}_i$) for each node i, i.e., $\mathbf{p}_i = \Psi(\mathbf{Z}, \mathbf{H}, \mathbf{T}_{-i}, t_i)$. The illustration of this module is shown in Fig. 3. This module is implemented based on a hypergraph convolutional network [Bai *et al.*, 2021; Yadati *et al.*, 2019] as well as a hypergraph attention mechanism [Bai *et al.*, 2021; Zhang *et al.*, 2019; Ding *et al.*, 2020].

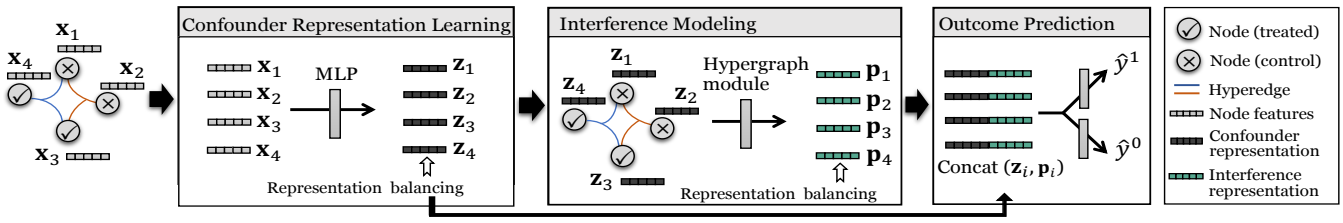


Figure 2: An illustration of HyperSCI, including confounder representation learning, interference modeling, and outcome prediction.

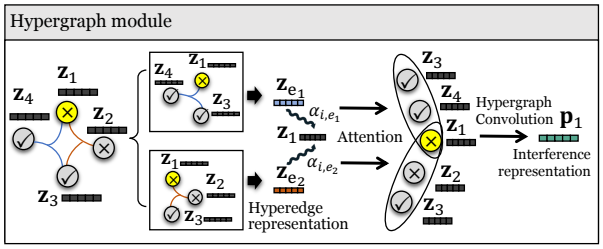


Figure 3: An illustration of the hypergraph module in HyperSCI. Here node v_1 (highlighted in yellow) is taken as an example.

To learn the interference representations for each node, the treatment and confounder representations are propagated through the hypergraph structure. A vanilla Laplacian matrix for the given hypergraph $\mathcal{H}$ can be calculated as:

$$\mathbf{L} = \mathbf{D}^{-1/2}\mathbf{H}\mathbf{B}^{-1}\mathbf{H}^{\top}\mathbf{D}^{-1/2}, \tag{2}$$

where $\mathbf{D} \in \mathbb{R}^{n\times n}$ is a diagonal matrix in which each element stands for the node degree (i.e., $\sum_{e=1}^{m} h_{i,e}$). $\mathbf{B} \in \mathbb{R}^{m\times m}$ is a diagonal matrix in which each element corresponds to the size of each hyperedge ($\sum_{i=1}^{n} h_{i,e}$). The hypergraph convolution operation is defined as:

$$\mathbf{P}^{(l+1)} = \text{LeakyReLU}\left(\mathbf{L}\mathbf{P}^{(l)}\mathbf{W}^{(l+1)}\right), \tag{3}$$

where $\mathbf{P}^{(l)}$ denotes the representations in the l-th layer of the hypergraph module. The input of the first layer is the confounder representation masked by the treatment assignment, i.e., $\mathbf{p}_i^{(0)} = t_i * \mathbf{z}_i$. Here, $*$ is element-wise multiplication. $\mathbf{W}^{(l+1)} \in \mathbb{R}^{d^{(l)}\times d^{(l+1)}}$ represents the parameter matrix in the (l+1)-th layer of the hypergraph module, where $d^{(l)}$ and $d^{(l+1)}$ are the dimensionality of the l-th layer and (l+1)-th layer, respectively.

While the hypergraph convolution layer allows for interference modeling through hyperedges, it lacks flexibility to consider the varying significance of interference on different nodes via different hyperedges. To address this, a hypergraph attention mechanism [Bai *et al.*, 2021; Ding *et al.*, 2020; Zhang *et al.*, 2019] is utilized to capture the intrinsic relationship between nodes and hyperedges. Specifically, the attention weights are learned for each node and its corresponding hyperedges, which allows for a better understanding of how certain individuals, such as those participating in group events, may have a greater influence on or be influenced by others in these groups within the context of a hypergraph, as seen in the COVID-19 example. Specifically, the attention score between a node i and a hyperedge e is calculated as:

$$\alpha_{i,e} = \frac{\exp(\sigma(\text{sim}(\mathbf{z}_i\mathbf{W}_a, \mathbf{z}_e\mathbf{W}_a)))}{\sum_{k\in\mathcal{E}_i}\exp(\sigma(\text{sim}(\mathbf{z}_i\mathbf{W}_a, \mathbf{z}_k\mathbf{W}_a)))}, \tag{4}$$

where $\sigma(\cdot)$ is an activation function, $\mathcal{E}_i$ is the set of hyperedges which contain the node i. $\mathbf{z}_e$ is the representation for each hyperedge e, obtained by aggregating across the representations of its associated nodes. $\mathbf{W}_a$ denotes a parameter matrix to compute the node-hyperedge attention. $\text{sim}(\cdot)$ denotes a similarity function, which can be implemented as:

$$\text{sim}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{a}^{\top}[\mathbf{x}_i, \mathbf{x}_j]. \tag{5}$$

Here, $\mathbf{a}$ is a weight vector, $[\cdot,\cdot]$ is a concatenation operation. The attention scores are used to model the different significance of interference. More specifically, the original incidence matrix $\mathbf{H}$ of the hypergraph in Eq. 2 is replaced with an attention-involved matrix $\tilde{\mathbf{H}} = \{\tilde{h}_{i,e}\}$, where $\tilde{h}_{i,e} = \alpha_{i,e}h_{i,e}$.

3.3 Outcome Prediction

Based on the confounder representations and the interference representations, the potential outcomes are predicted by:

$$\hat{y}_i^1 = f_1([\mathbf{z}_i, \mathbf{p}_i]),\ \hat{y}_i^0 = f_0([\mathbf{z}_i, \mathbf{p}_i]), \tag{6}$$

where $f_1(\cdot)$ and $f_0(\cdot)$ are learnable functions which are trained to predict potential outcomes for treatment assignment 1 and 0, respectively. The ITE for each node i is then estimated by: $\hat{\tau}_i = \hat{y}_i^1 - \hat{y}_i^0$. The prediction for the observed outcome is obtained by $\hat{y}_i = \hat{y}_i^{t_i}$. The final loss function is:

$$\mathcal{L} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \alpha\mathcal{L}_b + \lambda\|\mathbf{\Theta}\|^2, \tag{7}$$

where the first term is the outcome prediction loss, which can be implemented by standard mean squared error. $\mathcal{L}_b$ is the representation balancing loss, as introduced in Section 3.1. $\mathbf{\Theta}$ denotes all the model parameters. α and λ are hyperparameters that control the weights for representation balancing and model regularization, respectively.

4 Experimental Evaluation

It is often very hard to obtain the ground-truth counterfactuals as only one of the two potential outcomes can be obtained in the observational data. Hence, we follow a standard practice to evaluate our framework and the alternative approaches on

Data	Method	Linear		Quadratic	
		$\sqrt{\epsilon_{PEHE}}$	ϵ_{ATE}	$\sqrt{\epsilon_{PEHE}}$	ϵ_{ATE}
CT	LR	25.41	9.11	38.22	20.28
	CEVAE	22.88	8.29	35.28	18.22
	CFR	24.04	7.17	32.24	17.28
	Netdeconf	10.22	4.29	21.23	11.39
	GNN-HSIC	7.42	2.06	16.28	7.28
	GCN-HSIC	7.28	2.08	14.23	6.27
	HyperSCI	**3.45**	**1.39**	**9.20**	**2.24**
GR	LR	23.01	13.42	48.56	31.19
	CEVAE	22.69	12.49	45.21	29.22
	CFR	20.30	13.21	41.72	26.28
	Netdeconf	18.39	12.20	35.18	21.20
	GNN-HSIC	17.20	12.18	27.22	16.87
	GCN-HSIC	16.01	12.06	25.42	16.28
	HyperSCI	**15.68**	**11.81**	**19.23**	**13.33**
MS	LR	22.80	21.41	414.17	192.80
	CEVAE	19.36	8.63	315.01	188.47
	CFR	25.23	18.28	392.56	189.75
	Netdeconf	11.11	9.22	241.02	147.29
	GNN-HSIC	9.38	6.91	114.28	81.21
	GCN-HSIC	8.27	6.60	109.57	77.75
	HyperSCI	**5.13**	**4.46**	**81.08**	**74.41**

Table 1: ITE estimation performance. "CT", "GR" and "MS" refer to Contact, GoodReads, and Microsoft Teams datasets, respectively.

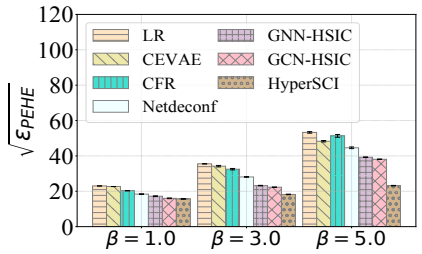


Figure 4: ITE estimation performance under different values of β in linear setting on GoodReads.

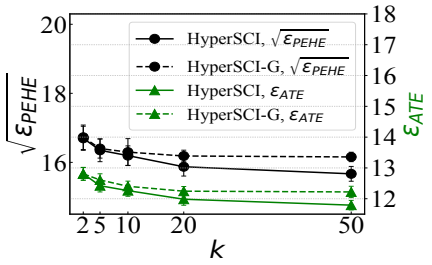


Figure 5: ITE estimation performance of HyperSCI / HyperSCI-G on GoodReads with hyperedge size no more than k.

three semi-synthetic datasets Contact, GoodReads, and Microsoft Teams. We leverage as much real-world information as possible in the simulated environment. Our datasets are all based on real-world hypergraphs, and we retain the treatment allocations as well as node features (covariates) if they are available. We simulate the outcome generation process to assess the true ITEs, which allow us to evaluate the performance of ITE estimation. Full details of the datasets, baselines, and experimental settings can be found in [Ma *et al.*, 2022].

4.1 ITE Estimation Performance

The performance of ITE estimation in hypergraph is shown in Table 1. HyperSCI outperforms all the baselines under different (linear and quadratic) simulation settings of the outcome generation function. As for the reasons, HyperSCI can leverage the hypergraph structure to model the high-order interference. In this way, it mitigates the influence of the interference on ITE estimation performance. Among baselines, some of them consider the pairwise network interference (GCN-HSIC and GNN-HSIC [Ma and Tresp, 2021]), or use the graph structure to infer the hidden confounders (Netdeconf [Guo *et al.*, 2020]). These methods perform better than those baselines (LR, CEVAE [Louizos *et al.*, 2017], CFR [Shalit *et al.*, 2017]) which cannot handle graph information. Furthermore, in the simulation, the hyperparameter β controls the level of hypergraph spillover effect in the outcome simulation. The ITE estimation results under different values of β are shown in Fig. 4. When β increases, the outcome is more strongly affected by interference, and larger performance gains can be observed from HyperSCI compared with the baselines.

4.2 A Closer Look at High-Order Interference

We further take a closer look at high-order interference. We investigate how HyperSCI responds to hyperedges with different sizes. More specifically, we remove the hyperedges with size larger than k, denote the modified hypergraph as $\mathcal{H}^{(k)}$, and vary the value of k. In Fig 5, we compare the ITE estimation performance of HyperSCI with its variant on the projected ordinary graph **HyperSCI-G**. We observe that: 1) When $k = 2$ (hyperedge size ≤ 2), the performance of HyperSCI-G is close to HyperSCI. Because when $k = 2$, graph convolution can be regarded as a special case of hypergraph convolution with small differences in the graph Laplacian matrix (as illustrated in [Bai *et al.*, 2021]). Empirically this leads to a minor performance difference between HyperSCI-G and HyperSCI; 2) When k increases, the performance of ITE estimation from both methods are gradually improved, but such an improvement becomes less significant when k is larger. Besides, we notice HyperSCI consistently outperforms HyperSCI-G and such a difference becomes larger as k increases, indicating its efficacy on modeling high-order interference especially on large hyperedges.

5 Conclusion

We study an important research problem of ITE estimation with the existence of high-order interference on hypergraphs. We identify and analyze the influence of high-order interference in causal effect estimation. To address this problem, we propose a novel framework **HyperSCI**, which estimates the ITEs based on representation learning. More specifically, HyperSCI learns the representation of confounders, models the high-order interference with a hypergraph neural network module, then predicts the potential outcomes for each instance with the learned representations. Extensive experiments conducted on semi-synthetic data based on real-world hypergraphs consistently validate the effectiveness of HyperSCI in ITE estimation under different interference scenarios.

References

[Ahluwalia *et al.*, 2001] Rohini Ahluwalia, H Rao Unnava, and Robert E Burnkrant. The moderating role of commitment on the spillover effect of marketing communications. *Journal of Marketing research*, 38(4):458–470, 2001.

[Aronow and Samii, 2017] Peter M Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4):1912–1947, 2017.

[Bai *et al.*, 2021] Song Bai, Feihu Zhang, and Philip HS Torr. Hypergraph convolution and hypergraph attention. *Pattern Recognition*, 110:107637, 2021.

[Basse and Feller, 2018] Guillaume Basse and Avi Feller. Analyzing two-stage experiments in the presence of interference. *Journal of the American Statistical Association*, 113(521):41–55, 2018.

[Ding *et al.*, 2020] Kaize Ding, Jianling Wang, Jundong Li, Dingcheng Li, and Huan Liu. Be more with less: Hypergraph attention networks for inductive text classification. *arXiv preprint arXiv:2011.00387*, 2020.

[Feng *et al.*, 2019] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3558–3565, 2019.

[Fisher, 1936] Ronald Aylmer Fisher. Design of experiments. *Br Med J*, 1(3923):554–554, 1936.

[Guo *et al.*, 2020] Ruocheng Guo, Jundong Li, and Huan Liu. Learning individual causal effects from networked observational data. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 232–240, 2020.

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

[Imai *et al.*, 2020] Kosuke Imai, Zhichao Jiang, and Anup Malani. Causal inference with interference and noncompliance in two-stage randomized experiments. *Journal of the American Statistical Association*, pages 1–13, 2020.

[Kohavi *et al.*, 2013] Ron Kohavi, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, and Nils Pohlmann. Online controlled experiments at large scale. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1168–1176, 2013.

[Louizos *et al.*, 2017] Christos Louizos, Uri Shalit, Joris M Mooij, David Sontag, Richard Zemel, and Max Welling. Causal effect inference with deep latent-variable models. In *Advances in Neural Information Processing Systems*, 2017.

[Ma and Tresp, 2021] Yunpu Ma and Volker Tresp. Causal inference under networked interference and intervention policy enhancement. In *International Conference on Artificial Intelligence and Statistics*, pages 3700–3708. PMLR, 2021.

[Ma *et al.*, 2022] Jing Ma, Mengting Wan, Longqi Yang, Jundong Li, Brent Hecht, and Jaime Teevan. Learning causal effects on hypergraphs. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2022.

[Mitchell, 1974] J Clyde Mitchell. Social networks. *Annual review of anthropology*, 3(1):279–299, 1974.

[Rubin, 1980] Donald B Rubin. Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, 75(371):591–593, 1980.

[Shalit *et al.*, 2017] Uri Shalit, Fredrik D Johansson, and David Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In *International Conference on Machine Learning*, 2017.

[Splawa-Neyman *et al.*, 1990] Jerzy Splawa-Neyman, Dorota M Dabrowska, and TP Speed. On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, pages 465–472, 1990.

[Tchetgen and VanderWeele, 2012] Eric J Tchetgen Tchetgen and Tyler J VanderWeele. On causal inference in the presence of interference. *Statistical methods in medical research*, 21(1):55–75, 2012.

[Ugander *et al.*, 2013] Johan Ugander, Brian Karrer, Lars Backstrom, and Jon Kleinberg. Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 329–337, 2013.

[Yadati *et al.*, 2019] Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha Talukdar. Hypergcn: A new method for training graph convolutional networks on hypergraphs. *Advances in neural information processing systems*, 32, 2019.

[Yilmaz *et al.*, 2002] Serdar Yilmaz, Kingley E Haynes, and Mustafa Dinc. Geographic and network neighbors: Spillover effects of telecommunications infrastructure. *Journal of Regional Science*, 42(2):339–360, 2002.

[Yuan *et al.*, 2021] Yuan Yuan, Kristen Altenburger, and Farshad Kooti. Causal network motifs: Identifying heterogeneous spillover effects in a/b tests. In *Proceedings of the Web Conference 2021*, pages 3359–3370, 2021.

[Zhang *et al.*, 2019] Ruochi Zhang, Yuesong Zou, and Jian Ma. Hyper-sagnn: a self-attention based graph neural network for hypergraphs. *arXiv preprint arXiv:1911.02613*, 2019.