

# Learning Causal Effects on Hypergraphs

Jing Ma, Mengting Wan, Longqi Yang, Jundong Li, Brent Hecht, Jaime Teevan

## 摘要

超图为节点间的多路组交互建模提供了一种有效的抽象，其中每个超边可以连接任意数量的节点。与大多数现有研究利用统计依赖关系不同，论文从因果关系的角度研究超图。具体而言，在本文中，作者重点研究了超图上的个体治疗效果估计 (ITE) 问题，旨在估计干预措施 (如戴口罩) 对每个个体节点结果 (如 COVID-19 感染) 的影响程度。现有的 ITE 估计工作要么假设一个个体的结果不受其他个体的实验分配影响，要么假设干扰只存在于普通图中连接的个体之间。作者认为，由于群体相互作用的存在，这些假设在现实世界的超图上是现实的。在这项工作中，作者研究了高阶干扰建模，并提出了一种新的由超图神经网络驱动的因果关系学习框架。

**关键词：**因果网络；超图；社交图谱

## 1 引言

个人之间的群体互动存在于广泛的场景中，例如大型聚会活动，日常的群聊，以及微软团队上的职场互动。虽然传统的成对图定义涵盖了大量的应用 (例如，人与人之间的物理接触网络或社交网络)，但它无法捕获这些群体交互的完整信息。因此可以引入超图的概念来解决这个限制。考虑一个超图例子，个体通过面对面的社会事件联系在一起，每个聚集事件都可以表示为超边。每个超边可以连接任意数量的个体，而普通边只连接两个节点。

虽然许多研究致力于利用这种广义超图结构来促进机器学习任务，但大多数研究仍然在统计相关水平上执行，例如，通过捕获人口统计信息、个人群体聚集历史和感染结果之间的相关性来预测每个个体的 COVID-19 感染风险。这里的一个关键限制是缺乏对干预和相关结果的因果关系分析。因果推断任务需要通过保持所有其他可能因素不变，除了感兴趣的变量，来构建同一个人的反事实状态。这在超图数据上是一个特别困难的问题，因为每个个体的结果不仅受到其自身的混杂因素 (例如，个人的健康状况和疫苗接种状况) 的影响，而且还受到超图上其他个体的干扰 (例如，可能与目标个体进行身体接触的其他个体是否戴口罩会通过聚集事件影响个体的结果)。

由此，作者引入超图的概念来解决这个问题。以图 1(a) 所示的超图为例，每个人通过面对面的社交活动建立连接，而每个把人聚集起来的活动可以被表示为一个超边。每个超边可以连接任意数量的人/节点，而普通图的边，如图 1(b) 所示，则只能连接两人/节点。

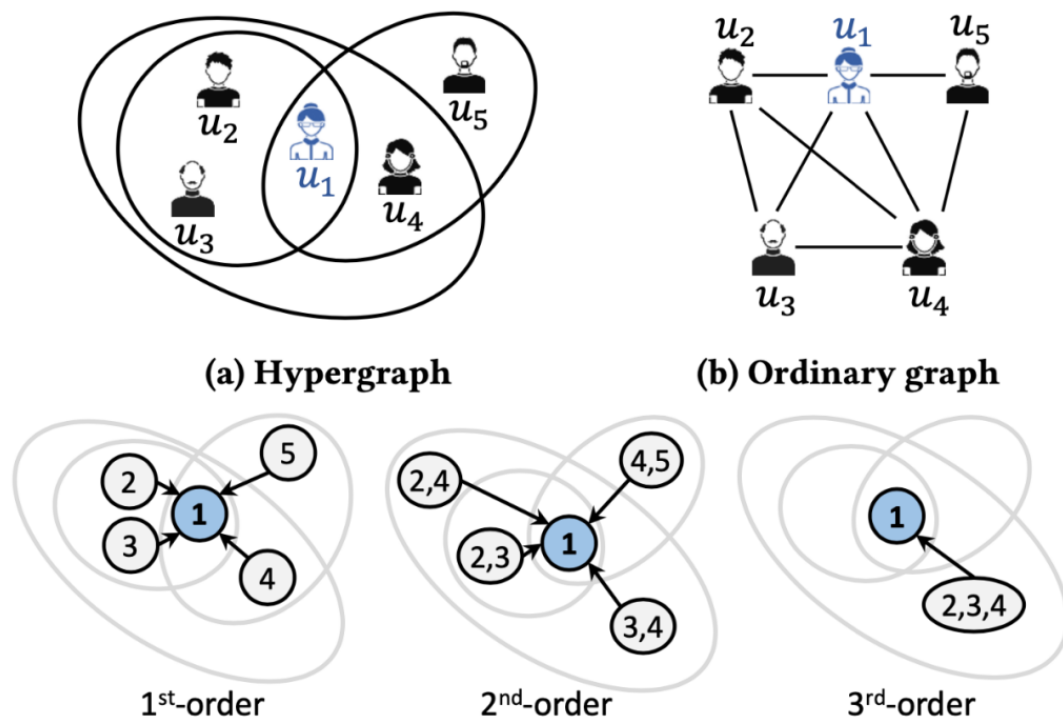


图 1: 超图与普通图

## 2 相关工作

### 2.1 网络干扰下的因果关系研究

已经有许多因果研究<sup>[1-7]</sup>，这些研究解决了网络干扰的存在。工作主要包括以下几类:(i) 干扰下的随机分配策略<sup>[1-2,4,8]</sup>。这些工作集中在干扰下的实验研究。在一些研究中，假设每个组内存在强干扰，而不同组间不存在干扰;(ii) 干扰下的观测数据的因果效应估计<sup>[5,9-10]</sup>。其中，Rakesh 等提出了一种关联因果变分自编码器 (LCVA) 框架，用于在对实例之间存在干扰的情况下估计治疗对结果的因果影响。不同于这些侧重于两两溢出效应的工作，Ma 等<sup>[5]</sup>考虑了网络结构中的溢出效应，并提出了基于图神经网络 (GNN)<sup>[11]</sup>的网络干扰下因果效应估计框架。但这些工作仍然局限于个体对或普通图，缺乏对高阶干扰的考虑。另一个研究方向是二部因果推断<sup>[12]</sup>。传统上，二部因果推断涉及两种类型的单位：介入/结果单位。介入单位分配治疗，并从结果单位观察结果。虽然这种设置与我们的不同，但考虑到每个超图 (和普通图) 都对应一个节点-超边二部，因此两种建模方法 (二部和超图) 在概念上是相似的。尽管如此，我们认为超图在许多情况下是一个更合适的框架，因为:i) 超图不需要将边实例化为额外的节点，也不需要区别对待这两种节点，因此计算效率更高;ii) 超图在泛化到新的超边时具有更方便和高效的潜力，而二部图既需要为新的超边生成新节点，也需要为它们关联的新边生成新节点。

### 2.2 超图算法和神经网络

为了处理下游任务的超图结构，一系列工作通过采用复杂的多路交互的抽象表示来简化超图结构<sup>[13]</sup>。其他作品直接处理原始超图结构<sup>[14-17]</sup>。近年来，关于超图神经网络的研究工作大量开展<sup>[18]</sup>。Feng 等提出了超图神经网络 (HGNN) 框架，在超图结构中编码高阶数据相关性。为表示学习设计了一种超边卷积运算。Bai 等<sup>[18]</sup>引入了两个端到端的可训练算子超图卷积和超图注意来学习超图中的节点表示。Yadati 等开发了一种基于自注意的超图神经网络 Hyper-SAGNN，该网络适用于具有可变超边大小的同构或异构超图。Jiang 等提出了一种动态超图神经网络 (DHGNN)，可以动态更新每一层的超图

结构。

### 3 本文方法

#### 3.1 本文方法概述

此部分对本文将要复现的工作进行概述，HyperSCI 框架主要分为三部分，控制混杂因素，构建高阶干扰模型，预测潜在结果，如图 2所示：

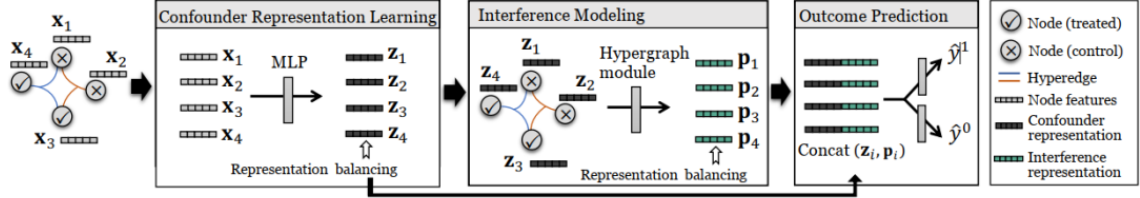


图 2: HyperSCI 框架示意图

#### 3.2 控制混杂因素

作者先通过多层感知器 MLP 模块将节点特征编码到一个隐空间，即  $z_i = MLP(x_i)$ 。这样就可以得到一组表征  $Z = \{z_i\}_{i=1}^n$ 。Z 可以捕捉到所有潜在的混杂因素，所以模型可以通过控制  $z_i$  来减轻混杂偏倚。

由于混杂因素表征 Z 的分布在对照组和实验组可能存在差异，从而对因果效应估计的结果造成偏差。为了使这种分布的差异最小化，作者决定在损失函数中加入差异惩罚项来平衡表征 Z。差异惩罚项可以用任何计算两个分布间距离指标来计算。本文采用 Wasserstein-1 距离作为计算对照组和实验组间表征分布的差异惩罚项。

#### 3.3 高阶干扰建模

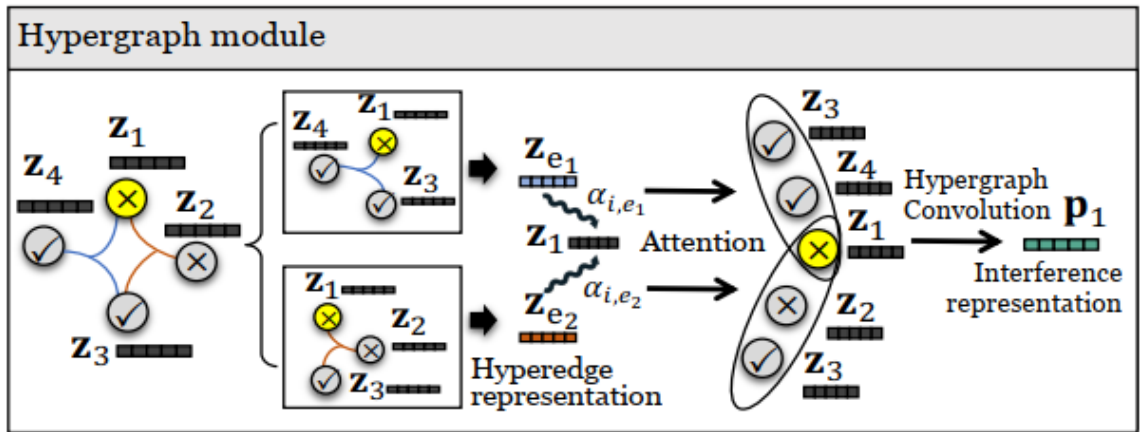


图 3: 超图卷积网络和超图注意力

在高级干扰模块中，作者把混杂因素表征 Z、实验分配 T 和超图上的关系信息  $H = h_{i,e}$  作为输入，来捕捉个体的高阶干涉。具体来说，作者通过超图来学习一个转换函数  $Psi$  来生成每个节点的干涉表征，即  $p_i = \Psi(Z, H, T_{-i}, t_i)$ 。如 3所示，该模块由超图卷积网络和超图注意力机制构成。其中，卷积算子形成了超边干涉的骨架，而注意力算子通过衡量节点对超边的贡献度来增强这个骨架。

为了学习超图上每个节点的干涉表征，作者用超图卷积层来传播实验分配和混杂因素表征。首先为超图 H 引入一个拉普拉斯矩阵，如下公式所示：

$$L = D^{-1/2}HB^{-1}H^TD^{-1/2}$$

其中,  $D \in R^{n \times n}$  是一个对角矩阵,  $D$  中的每个元素表示节点的度  $\sum_{e=1}^m h_{i,e}$ 。  $B \in R^{m \times m}$  也是一个对角矩阵,  $B$  中的每个元素表示每条超边的大小  $\sum_{i=1}^n h_{i,e}$ 。

这样一来, 超图卷积算子的定义就如下所示:

$$p^{(l+1)} = \text{LeakyReLU}(LP^{(l)}W^{(l+1)})$$

其中,  $p^{(l)}$  代表第  $l$  层的表征。我们用经过实验分配后的混杂因素表征作为干涉表征的第一层, 即  $p_i^{(0)} = t_i * z_i$ ,  $*$  表示对应元素相乘。  $W^{(l+1)} \in R^{d^{(l)} \times d^{(l+1)}}$  是第  $(l+1)$  层的参数矩阵, 其中  $d^{(l)}$  和  $d^{(l+1)}$  分别指的是第  $(l)$  层和第  $(l+1)$  层干涉表征的维度。

尽管上述卷积层可以通过超边传播干涉, 但它没有提供太多的灵活性来说明不同节点在不同超边下干涉的意义。在 COVID-19 的例子中, 从直觉上来说, 那些活跃参加聚会活动的个体更有可能感染或被这些群体中的其他人所传染。因此, 为了更好捕捉到超图上节点和超边之间这种内在的关, 作者利用超图注意力机制来学习每个节点及包含该节点的超边的权重。

具体来说, 我们给每条超边  $e$  计算其表征  $x_e$ 。该表征是通过聚合其有关联的节点  $N_e$  得到, 即:  $z_e = \text{AGG}(z_i | i \in N_e)$ , 其中  $\text{AGG}(\cdot)$  可以是任何聚合函数。对于每个节点  $i$  及其相关的超边  $e$ , 节点  $i$  和超边  $e$  之间的注意力得分的计算如公式所示:

$$\alpha_{i,e} = \frac{\exp(\sigma(\text{sim}(z_i W_a, x_e W_a)))}{\sum_{k \in \varepsilon_i} \exp(\sigma(\text{sim}(z_i W_a, z_k W_a)))}$$

其中,  $\sigma(\cdot)$  是一个非线性激活函数,  $\varepsilon_i$  表示与节点  $i$  相关的超边集合。我们用  $W_a$  表示一个用来计算节点-超边间注意力的参数矩阵,  $\text{sim}(\cdot)$  是一个相似度函数, 它的实现如公式所示:

$$\text{sim}(x_i, x_j) = a^T[x_i || x_j]$$

作者用注意力分数来模拟不同程度的干涉。具体来说, 作者用一个增强矩阵  $\bar{H} = \bar{h}_{i,e}$  来取代原始关联矩阵  $H$ , 其中  $\bar{h}_{i,e} = \alpha_{i,e} h_{i,e}$ 。这样一来, 在同一超边内不同节点的干涉就可以被赋予不同的权重, 来表示对建模干涉不同程度的贡献。作者将卷积层最后一层的表征定义为  $P = \{p_i\}_{i=1}^n$ , 并希望它能捕捉到每个节点的高阶干涉。

### 3.4 潜在结果预测

在得到混杂因素表征  $z_i$  和干涉表征  $p_i$  后, 对潜在结果进行建模, 如公式所示:

$$\hat{y}_i^1 = f_1([z_i || p_i]), \hat{y}_i^0 = f_0([z_i || p_i])$$

其中,  $f_1(\cdot)$  和  $f_0(\cdot)$  是可学习的函数, 用来预测潜在结果即  $t=1$  和  $t=0$ 。在本文中,  $f_1(\cdot)$  和  $f_0(\cdot)$  是两个  $MLP$ 。作者通过  $\hat{y}_i = \hat{y}_i^{t_i}$  得到对观察结果的预测。通过最小化下列损失函数来优化模型, 损失函数如公式所示:

$$L = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \alpha L_b + \lambda ||\theta||^2$$

其中, 损失函数的第一项是标准的均方误差,  $L_b$  是表征平衡损失,  $\theta$  表示这个神经网络模型的参数。  $\alpha$  和  $\lambda$  是两个超参数, 用来控制表征平衡损失和参数正则项的权重。每个实例  $i$  的 ITE 可以被估计为:  $\hat{\tau}_i = \hat{y}_i^1 - \hat{y}_i^0$

## 4 复现细节

### 4.1 与已有开源代码对比

本次复现工作主要包括数据预处理、计算平衡损失以及模型的搭建。其中，平衡损失计算是使用 Wasserstein 距离实现的，这部分代码是开源的；模型搭建过程中也引用了开源的 PyTorch Geometric 包，可轻松编写和训练图神经网络。具体的工作如下所示：

由损失函数定义可知，评估平衡损失要计算实验组和对照组的联合概率分布，在开源代码中，这个模块的计算默认两组样本集的数目相同。然后，由于个体之间存在选择偏差，实验组的样本数量往往远超过对照组，导致源代码不能直接使用。针对这点，做了以下改进：分别对两组样本集数量扩展维度。即，实验组  $R1^{m \times d} \rightarrow R1^{m \times n \times d}$ ，对照组  $R0^{n \times d} \rightarrow R0^{m \times n \times d}$ 。然后通过聚合函数对计算结果进行降维，就能得到联合概率分布矩阵。此部分的优化可以确保源代码能够适用于任意数量样本分布的距离度量。

PyTorch Geometric, 简称 PYG, 是一个基于 PyTorch 的库，包括从各种已发表的论文中对图和其他不规则结构进行深度学习的各种方法，也称为几何深度学习。在搭建 HyperSCI 框架中，主要基于 PYG 包实现干扰表示学习中的超图卷积算子以及超图注意力模块。

---

**Procedure 1** Confounder Representation Learning and Interference Modeling.

---

**Input:** confounder feature  $x_i$

**Output:** representation feature  $z_i$   $p_i$

*confounder representation* :  $z_i = \text{MLP}(x_i)$

*attention score* :  $\alpha_{i,e} = \frac{\exp(\sigma(\text{sim}(z_i W_a, z_e W_a)))}{\sum_{k \in \epsilon_i} \exp(\sigma(\text{sim}(z_i W_a, z_k W_a)))}$

*update hypergraph matrix* :  $h_{i,e} = \alpha_{i,e} h_{i,e}$

*interference representation* :  $p_i = \text{LeaklyRelu}(L_i * t_i * z_i W_i)$

---

### 4.2 实验环境搭建

Pycharm 2020

windows10

pytorch1.12

python3.7

### 4.3 创新点

- 1) 解决如何在超图干涉下，构建因果关系模型的难题
- 2) 提出了一个新的框架 HyperSCI —— 通过表示学习和超图神经网络对混杂特征和高阶干扰进行建模
- 3) 通过大量的实验验证了论文所提出的框架的有效性，并深入分析了它如何作用于不同的节点和超边

## 5 实验结果分析

在本节中，在两个公开的超图数据集上测试了 HyperSCI 和四个因果推理算法，以验证 HyperSCI 的有效性。下面给出了实验结果分析：

Data	Method	Linear		Quadratic	
		$\sqrt{\epsilon_{PEHE}}$	$\epsilon_{ATE}$	$\sqrt{\epsilon_{PEHE}}$	$\epsilon_{ATE}$
Contact	LR	32.27 $\pm$ 0.57	9.32 $\pm$ 1.25	36.48 $\pm$ 0.87	21.88 $\pm$ 1.39
	CFR	24.04 $\pm$ 0.75	7.17 $\pm$ 1.43	32.24 $\pm$ 1.01	17.28 $\pm$ 0.75
	GNN-HSIC	8.28 $\pm$ 0.87	1.31 $\pm$ 1.16	25.25 $\pm$ 1.31	3.33 $\pm$ 1.03
	GCN-HSIC	8.37 $\pm$ 0.86	1.18 $\pm$ 1.04	25.25 $\pm$ 1.34	3.07 $\pm$ 1.96
	HyperSCI	<b>7.37 <math>\pm</math> 0.87</b>	<b>0.97 <math>\pm</math> 1.05</b>	<b>10.12 <math>\pm</math> 0.93</b>	<b>1.38 <math>\pm</math> 0.77</b>
Good reads	LR	22.79 $\pm$ 0.54	13.01 $\pm$ 1.21	48.57 $\pm$ 1.21	34.11 $\pm$ 1.45
	CFR	20.31 $\pm$ 0.63	12.22 $\pm$ 1.17	40.63 $\pm$ 1.17	27.43 $\pm$ 1.31
	GNN-HSIC	18.45 $\pm$ 0.61	11.99 $\pm$ 1.17	25.17 $\pm$ 1.06	15.78 $\pm$ 1.23
	GCN-HSIC	16.28 $\pm$ 0.58	12.04 $\pm$ 1.06	24.32 $\pm$ 1.11	15.38 $\pm$ 1.23
	HyperSCI	<b>14.39 <math>\pm</math> 0.41</b>	<b>11.76 <math>\pm</math> 1.04</b>	<b>19.67 <math>\pm</math> 1.01</b>	<b>13.29 <math>\pm</math> 0.99</b>

图 4: 实验结果示意

图 4 中包含了 ITE 估计任务的结果。从这个表中我们观察到，无论使用线性还是平方合成数据进行实验，所提出的框架都优于所有基线。这些结果归因于 HyperSCI 利用超图中的交互信息来构建高阶干扰模型，从而减轻了溢出效应对 ITE 估计性能的影响。与其他基线相比，结合了对网络干扰的方法 (GCN-HSIC 和 GNN-HSIC) 的性能优于不利用交互信息的基线 (LR、CFR)。

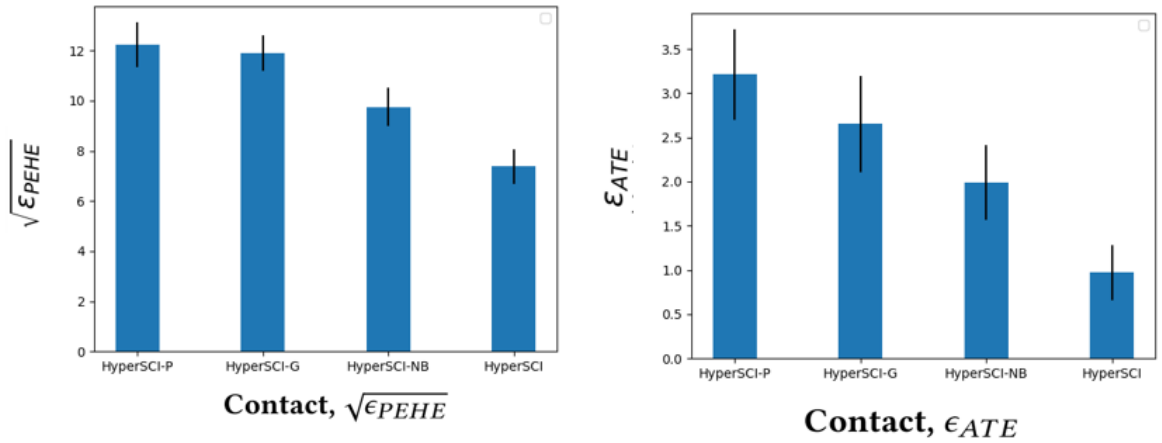


图 5: 消融实验结果示意

为了研究所提出的框架中不同组件的有效性，通过考虑以下变量来进行消融研究:1) 将所提出的模型 HyperSCI 应用于投影图 (记为 HyperSCI-p);2) 取代超图神经网络模块使用具有相同层数的图神经网络模块，然后将其应用于投影图 (HyperSCI-G)。尽管都是在投影图上计算，HyperSCI-G 用它的图神经网络模块处理普通图，而 HyperSCI-P 用它的超图神经网络模块处理超图;3) 删除了框架中的平衡技术 (HyperSCI-NB)。ITE 估计结果如图 5 所示，其中 HyperSCI-p/HyperSCI-G 与 HyperSCI 之间存在显著的性能差距，这意味着在超图上对高阶交互进行建模的有效性。在去除表示平衡模块后，ITE 估计性能下降，这表明表示平衡技术在降低 ITE 估计偏差方面是有效的。

## 6 总结与展望

这篇论文研究了超图上存在高阶干扰的个体实验效果估计的问题。通过实验，作者确定并分析了高阶干扰对因果效应估计的影响。为了解决这个问题，作者提出了一个基于表征学习来估计 ITE 的新框架 HyperSCI。

具体来说，HyperSCI 可以学习到混杂因素的特征，用超图神经网络模块建立高阶干扰模型，然后

用学到的表征预测每个实例的潜在结果。作者进行了广泛的实验来评估所提出的框架 HyperSCI，结果验证了 HyperSCI 在不同干扰场景下估计 ITE 的有效性。

最后我们再看一下论文可以优化的地方。首先，作者假设混杂特征  $X$  包含现实世界的所有混杂因素，这显示是不可能的。针对这个问题，目前的做法是使用变分编码器从已知的混杂特征  $x$  中学习整体的混杂因素，但是如何将变分编码器融入到平衡分布中，任然需要进一步的研究。其次，作者在考虑个体交互关系时，没有考虑时间因素。动态超图交互建模也是一个非常值得研究的方向。

## 参考文献

- [1] ARONOW P M, SAMII C. Estimating average causal effects under general interference, with application to a social network experiment[J]. The Annals of Applied Statistics, 2017, 11, 4: 1912-1947.
- [2] BASSE G, FELLER A. Analyzing two-stage experiments in the presence of interference[J]. Amer. Statist. Assoc, 2018, 113, 521: 41-55.
- [3] ROHIT BHATTACHARYA D M, SHPITSER I. Causal inference under interference and network uncertainty[J]. In Uncertainty in Artificial Intelligence, 2020.
- [4] KOSUKE IMAI Z J, MALANI A. Causal inference with interference and noncompliance in two-stage randomized experiments[J]. J. Amer.Statist. Assoc, 2020: 1-13.
- [5] MA Y, TRESP V. Causal Inference under Networked Interference and Intervention Policy Enhancement [J]. In International Conference on Artificial Intelligence and Statistics, 2021: 3700-3708.
- [6] TCHETGEN E J T, VANDERWEELE T J. On causal inference in the presence of interference[J]. Statistical methods in medical research, 2012, 21, 1: 55-75.
- [7] YUAN YUAN K A, KOOTI F. Causal Network Motifs: Identifying Heterogeneous Spillover Effects in A/B Tests[J]. In Proceedings of the Web Conference, 2021: 3359-3370.
- [8] FATEMI Z, ZHELEVA E. Minimizing interference and selection bias in network experiment design[J]. In International AAAI Conference on Web and Social Media, 2020.
- [9] DAVID ARBOUR D G, JENSEN D. Inferring network effects from observational data[J]. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 715-724.
- [10] ERIC J TCHETGEN TCHETGEN I R F, SHPITSER I. Auto-gcomputation of causal effects on a network[J]. J. Amer. Statist. Assoc, 2021, 116, 534: 833-844.
- [11] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv preprint, 2016.
- [12] ZIGLER C M, PAPADOGEORGOU G. Bipartite causal inference with interference[J]. Statistical science: a review journal of the Institute of Mathematical Statistics, 2021, 36, 1: 109.



- [13] YE XU D R, KLEINBAUM A M. Hyperlink prediction in hypernetworks using latent social features [J]. In International Conference on Discovery Science. Springer, 2013: 324-339.
- [14] ARYA D, WORRING M. Exploiting relational information in social networks using geometric deep learning on hypergraphs[J]. In Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval, 2018, 11, 4: 117-125.
- [15] AUSTIN R BENSON R K, TOMKINS A. Sequences of sets[J]. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018: 1148-1157.
- [16] JIN HUANG R Z, YU J X. Scalable hypergraph learning and processing[J]. In 2015 IEEE International Conference on Data Mining, 2015: 775-780.
- [17] ANKIT SHARMA J S, CHANDRA A. Predicting multi-actor collaborations using hypergraphs[J]. arXiv preprint, 2014.
- [18] ARYA D, WORRING M. Hypergraph convolution and hypergraph attention[J]. Pattern Recognition, 2021, 110: 107637.