arXiv:2302.04997v1 [stat.ME] 10 Feb 2023

基于模型的回归调整与用于网络干扰的无模型协变量

Kevin Han，斯坦福大学统计系；Johan Ugander，斯坦福大学管理科学与工程系

2023年2月13日

关键词：干扰下的因果推断、稳定单元处理值假设、A/B测试、回归调整、社会网络分析

摘要

在网络干扰下估计总体平均处理效应（GATE）时，个体与处理之间的关系可能大不相同，这取决于其网络邻域的结构、干扰机制的结构，以及处理在其邻域中的分配方式。在本研究中，我们提出一种序贯程序，用于生成和选择基于图和处理的协变量，以便通过回归调整估计 GATE。我们证明，采用这一程序能够同时实现较低偏差并显著降低方差。为解决特征生成与选择过程带来的推断难题，我们提出一种基于区块自助法构造置信区间的方法。我们通过若干采用伯努利设计的半合成实验展示，该选择程序及其后续估计量在均方根误差方面表现良好，且优于利用已知底层干扰结构进行回归调整的预言机估计量。我们将该方法应用于一个有强有力干扰证据的真实世界实验数据集，并证明，即使事先不了解干扰过程，该方法也能较为准确地估计 GATE。

1 引言

在标准实验中，研究者通常假设一个个体的分配不会影响另一个个体的响应；这通常称为无干扰假设（Cox 1958，第 2 章），或稳定单元处理值假设（SUTVA）（Rubin 1974）。然而，当实验单位彼此互动时，SUTVA 往往难以成立。在许多应用中都发现了 SUTVA 的违背，包括政治学（Sinclair 等人 2012）、教育学（Hong 和 Raudenbush 2006, Rosenbaum 2007）、经济学（Sobel 2006，Manski，2013）和公共卫生（Halloran 和 Struchiner 1995）。近年来，开发具有社交或市场互动产品的科技公司，已制定方法来管理产品实验中显著的干扰（Eckles 等人 2017，Pouget-Abadie 等 2019，Karrer 等人 2021）。实践中，研究者会寻找能够限制干扰范围的底层结构，并在假设该结构的基础上估计因果效应。Aronow 和 Samii（2017）提出使用干扰机制的低维表示，并据此估计因果效应。在无干扰文献中，回归调整已被证明在理论上（Lin 2013）和实践中（Deng 等 2013）均有效。Chin（2019）在假设结果变量服从线性模型的条件下，研究了存在干扰时的回归调整，并根据实验数据估计模型参数。线性模型假设并不少见，也曾在实验设计（Harshaw 等人 2022）和干扰检测（Pouget-Abadie 等 2019）中得到研究。此外，也有研究提出新设计来应对干扰带来的复杂问题。例如，Ugander 等人（2013）以及 Ugander 和 Yin（2020）研究了（随机化）整群随机设计，通过在簇级而非个体级进行随机化，有效地处理干扰。

本文提出一种估计总体平均处理效应的方法，通过回归调整进行估计，而无需像 Chin（2019）那样假设真实的特征集合。我们以无模型的方式，根据观测到的实验数据生成用于调整的特征。本文首先介绍问题设定的预备知识，并通过计量经济学中经典的线性均值模型来阐释我们的方法。接着，我们提出一种通用方法，根据观测到的实验数据生成无模型协变量。最后，我们展示如何利用无模型协变量估计总体平均处理效应并进行推断。最后，我们通过模拟、一个实证应用和讨论作结。

2 设定

考虑一个针对 n个体开展的随机实验，其中存在一个简单无向图$G=(V,\mathcal{E})$，用于描述 n 个体之间的社交网络互动。图 G对应一个对称矩阵 $A\in\mathbb{R}^n$，使得 $A_{ij}=1$ 在 $(i,j)\in\mathcal{E}$ 时成立，否则为零。令 $\mathcal{N}_i^{(k)}$ 表示每个节点 $i\in V$周围的 k跳邻域。当 $k=1$ 时，我们省略上标，并令 d_i 表示每个节点的度（或等价地，$d_i=|\mathcal{N}_i|$）。我们用 W_i 表示随机分配，用 $x_i\in\mathcal{X}$ 表示个体 i的处理前协变量。我们

假设实验总体就是我们关注的总体，因此将处理前协变量视为固定值。我们只考虑二元处理，但需要指出，扩展到非二元处理也很直接。全文中，我们用带有相应下标的小写字母表示随机变量的实现值以及非随机量。

我们采用 Rubin 因果模型（Rubin 1974，Holland 1986，Imbens & Rubin 2015）。对于每个个体 i，我们将其与潜在结果 $Y_i(w) \in \mathbb{R}$ for $w \in \{0,1\}^n$ 联系起来。我们关注以下因果估计量，并将其称为总体平均处理效应（GATE）：

$$\tau = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[Y_i(\mathbf{1}) - Y_i(\mathbf{0})]. \tag{1}$$

这里，1 表示 n维全 1 向量，0 也同理。在一些研究中，GATE 估计量也称为总处理效应（TTE）（Yu et al. 2022），用于衡量干预对实验单位的总体效应。在稳定单元处理值假设（SUTVA）下，其他个体的分配不会影响某个个体的响应，因此每个个体只有两个潜在结果，即 $Y_i(0)$ 和 $Y_i(1)$。在 SUTVA 下，GATE 就是平均处理效应（ATE）。当网络中存在干扰时，每个个体最多可能有 2^n 个不同的潜在结果。在没有进一步假设的情况下，不可能观察到某个个体 i 的 $Y_i(1)$，同时又观察到任何其他个体 j的 $Y_j(0)$。

在本文中，我们采用回归视角，并假设存在两个函数 f_0 和 f_1，使得对于每个个体 i 和每个分配向量 $w \in \{0,1\}^n$，

$$Y_i(w) = w_i f_1(i, w, x_i, G) + (1 - w_i) f_0(i, w, x_i, G) + \epsilon_i, \tag{2}$$

其中 ϵ_i 均为外生变量，即 $\mathbb{E}[\epsilon_i|w] = 0$。函数 f_0 和 f_1 均以节点标签 i、分配向量 w、协变量向量 x_i 和图 G为输入。这种方法采用暴露映射（Aronow 和 Samii 2017）作为函数，将分配向量w 和 x_i 映射为特定的暴露值；因此，如果两个分配向量 w 和 w' 为某个个体带来相同的暴露值，那么它们对应的潜在结果值也相同。由于潜在结果只取决于暴露值，我们可以将其视为暴露值的函数，并将潜在结果改写为式（2）的形式。根据式（2），由于函数 f_1 和 f_0 在所有个体间共享，我们可以利用处理组单位估计 f_1，并利用对照组单位估计 f_0。假设 f_0 和 f_1 分别是 f_0 和 f_1 的两个估计值，那么 GATE 的一个自然估计量为

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} [\hat{f}_1(i, \mathbf{1}, x_i, G) - \hat{f}_0(i, \mathbf{0}, x_i, G)].$$

遗憾的是，如果不对函数 f_0 和 f_1[1]的结构作出进一步假设，就无法估计 GATE。为了说明我们对 f_0和 f_1所作的结构假设，我们来看下面的例子。

[1]Basse 和 Airoldi（2018）从推断视角进行了讨论。

示例 1（线性均值模型）。考虑以下结构模型（Manski 1993、Moffit 2001、Bramoull´e 等，2009）

$$\mathbf{y} = \alpha\mathbf{1} + \beta\tilde{A}\mathbf{y} + \gamma\mathbf{w} + \delta\tilde{A}\mathbf{w} + \boldsymbol{\epsilon}, \quad \mathbb{E}[\boldsymbol{\epsilon}|\mathbf{w}] = 0, \tag{3}$$

其中，y 是 $n \times 1$ 结果向量，A 是度归一化邻接矩阵，即$\tilde{A}_{ij} = A_{ij}/d_i$；w 是分配向量，而（$\alpha$、$\beta$、$\gamma$、$\delta$）是参数。Bramoull´e 等人（2009）证明，在系数和图 G满足某些温和条件时，我们可以将上述模型改写为

$$\mathbf{y} = \alpha/(1-\beta)\mathbf{1} + \gamma\mathbf{w} + (\gamma\beta + \delta)\sum_{j=0}^{\infty}\beta^j\tilde{A}^{j+1}\mathbf{w} + \sum_{j=0}^{\infty}\beta^j\tilde{A}^{j+1}\boldsymbol{\epsilon}. \tag{4}$$

注意，现在结果变量不仅关于分配向量 w 呈线性，也关于 $\{\tilde{A}^{j+1}\mathbf{w}\}_{j=0}^{\infty}$呈线性。令$f_0(i, w, x_i, G) = f_1(i, w, x_i, G) = \alpha/(1-\beta) + \gamma w_i + (\gamma\beta + \delta)\sum_{j=0}^{\infty}\beta^j\tilde{A}^{j+1}w$，并注意到 $\mathbb{E}[\sum_{j=0}^{\infty}\beta^j\tilde{A}^{j+1}\boldsymbol{\epsilon}|w] = 0$。因此，线性均值模型（3）可以写成（2）的形式。

虽然在此例中，线性模型是无限维的，但式（4）的线性结构促使我们为 f_0 和 f_1都考虑线性模型。为使其形式化，我们给出以下定义：

定义 2.1（线性干扰）。若存在函数 g：$[n] \times \{0,1\}^n \times \mathcal{X} \times \mathcal{G} \to \mathbb{R}^K$and $\theta_0 \in \mathbb{R}^K, \theta_1 \in \mathbb{R}^K$，使得 $f_0(i, w, x_i, G) = \theta_0^T g(i, w, x_i, G)$ 且 $f_1(i, w, x_i, G) = \theta_1^T g(i, w, x_i, G)$，则称模型 $\mathcal{Y} = \{Y_i(w)$：$w \in \{0,1\}^n, i \in [n]\}$具有线性干扰。我们将 g 的每个坐标函数 g_j 称为干扰特征。

尽管线性干扰形式简单，但从图的角度来看，可以证明图上的卷积能够通过线性展开得到良好近似（Hammond 等，2011）。这种线性干扰假设并不少见（Deng 等，2013；Pouget-Abadie 等，2019；Chin，2019）。Chin（2019）展示了在获得预言机g 后如何进行推断，而 Pouget-Abadie 等（2019）则给出了在线性干扰假设下检测网络干扰的检验程序。此外，由于我们关注的是估计函数 f_0 和 f_1 在（仅） $w = \mathbf{0}, \mathbf{1}$ 上的质量，因此实际上是在尝试实现泛化。简单模型通常具有良好的泛化能力（Bousquet 等，2004；von Luxburg 和 Sch¨olkopf，2011），因此，在线性干扰下，即使 g 可能任意复杂，也能使推断更具可信度，同时不失灵活性。

在继续之前，我们可以对（2）稍作简化。注意，

$$\begin{aligned} Y_i(w) &= w_i f_1(i, w, x_i, G) + (1 - w_i) f_0(i, w, x_i, G) + \epsilon_i \\ &= w_i f_1(i, w^{(i\to 1)}, x_i, G) + (1 - w_i) f_0(i, w^{(i\to 0)}, x_i, G) + \epsilon_i \\ &= w_i \tilde{f}_1(i, w^{(-i)}, x_i, G) + (1 - w_i)\tilde{f}_0(i, w^{(-i)}, x_i, G) + \epsilon_i, \end{aligned} \tag{5}$$

其中，$w^{(i\to t)}$ 表示一个 n–维向量，它将 w_i 替换为 t，而 $\tilde{f}_t$ 仅是i、$w^{(-i)}$、x_i 和 G 的函数。因此，不失一般性，我们假设g 的定义域，以及 f_0 和 f_1 的定义域，均为 $[n] \times \{0,1\}^{n-1} \times \mathcal{X} \times \mathcal{G}$。

从这里开始，为了表述简洁，我们将在讨论中省略处理前协变量 x_i。对于纳入处理前协变量的情形，若其推广方式不明显，我们会予以讨论。因此， g 仅是节点标签 i、分配向量 w和图 G 的函数。

我们关注满足以下均匀性假设的设计：假设 2.2（均匀性）。我们假设 W_i's 相互独立，并且对于某个 $0 < p_i < 1$， $\forall i, \mathbb{P}(W_i = 1) = p_i$。

我们作出这一假设，是为了遵循在网络实验中采用伯努利随机化的惯例，例如 Karrer 等人（2021）。作为替代方案，考虑网络干扰的设计（例如图簇随机化）所得到的估计值可能会遭受较大的方差（Ugander 和 Yin 2020）。下文中，我们假设 W_i's 是服从伯努利(p) 分布且具有 $0 < p < 1$ 的独立同分布随机变量，即我们使用来自伯努利设计实验的数据。

如果我们事先知道函数 g，Chin（2019）提供了完整的解决方案。然而，如果我们不知道函数 g，那么就会面临三个重大挑战，而本文将逐一应对这些挑战。第一，我们应如何构造 g，使得所构造的函数能够逼近真实函数？第二，假设我们有许多候选函数，那么应如何在其中进行选择？第三，即使我们已经很好地回答了前两个问题，又应如何进行推断？我们将在下一节讨论前两个挑战，并在之后讨论第三个挑战。

3 无模型协变量

现在，根据 (5)，定义 2.1 中的函数 g 以节点标签 i、$w^{(-i)}$ 和 G 为输入，并输出一个 K维向量。该函数 g 本质上所做的，是基于 $w^{(-i)}$ 和 G 为每个个体 i生成 K 协变量。本节将介绍一种生成和选择无模型协变量的序贯程序。我们的方法大致如下：仅根据图结构和分配向量生成丰富的候选特征，再根据观测结果从中进行选择。我们先在下方给出算法 1 中的程序，然后更详细地说明各个步骤。我们将该程序称为 ReFeX-LASSO，因为它以图挖掘技术 ReFeX（Henderson 等人（2011））生成候选特征，并使用 LASSO（Tibshirani，1996）选择特征。

ReFeX（递归特征提取）最初旨在为图挖掘任务生成特征，也可以被视为一种递归算法：它从图中每个节点的基础特征开始，并根据对邻居节点特征进行聚合的结果，迭代地（i）添加和（ii）剪枝特征。ReFeX 可以被视为近年来基于图卷积网络（图卷积网络，GCNs）的图表示学习方法的早期简单先驱（Hamilton 等，2017；Kipf 和 Welling，2017）。我们采用 ReFeX 算法中的特征生成步骤，但将原算法中的特征剪枝部分替换为 LASSO。这一修改使我们能够更准确地刻画算法在任意给定步骤可用的特征。

算法 1 ReFeX-LASSO 输入：图 $G = (V, \mathcal{E})$、分配向量 $w \in \{0,1\}^n$、最大迭代次数 T。输出：协变量集合 S。1: 初始化 $S = \{\}$、活跃特征集 $A = \{\}$。2: 对于每个节点/单位 i，构造 m个基础特征，并将 m 个基础特征添加到 A。3: 对于 $t = 1$ 到 T 执行4: 使用 LASSO 将 y 对 w 以及来自 S 和 A 的特征进行回归，其中不对来自 S的特征施加惩罚。5: 如果未选定 A 中的任何特征，则返回 S。否则，将从 A中选定的特征添加到 S。6: 通过在一跳邻域中的邻居上对 A 中的特征进行聚合，递归地构造特征。7: 删除 A 中的旧特征，并将这些新特征添加到 A。8: 结束循环9: 返回 S。

ReFeX 包含两项要素——基础特征和聚合函数。给定 w、 $\{x_i\}_{i=1}^n$和 G，基础特征是指只需查看每个节点的一跳邻域即可构造的特征。只要满足这种局部查找约束，基础特征可以是任意的。基础特征可以是纯粹的图特征，例如度、中心性、聚类系数等；也可以是处理前协变量 x_i。通常，我们还希望基础特征不仅依赖函数 g 的一个输入，还依赖由两个输入 g计算得出的特征。例如，处理组邻居数量这类特征就同时取决于分配向量 w 和图 G。再如，所有邻居的平均特征值，取决于处理前协变量和 G。使用 ReFeX 时，基础特征由分析人员选择。聚合函数以相邻节点的特征作为输入，并输出单个值。因此，一个聚合函数本质上是在邻居的特征值样本上计算统计量。聚合函数同样可以是任意的，并由分析人员选择。常见的例子包括最小值、最大值、总和、均值和方差（Henderson 等人（2011））。

现在我们可以介绍 ReFeX-LASSO 算法了。ReFeX-LASSO 算法从两个空特征集开始，即目标集 S 和活跃特征集 A。第一个集合 S 存储选定特征，而 S 中的特征将用于调整 GATE 估计值。活跃特征集 A 包含上一步递归添加但尚未选定的特征。在该过程开始时，我们为每个个体 i 构造基础特征。获得一组基础特征后，每次我们都使用 LASSO 将结果向量 y 对集合 S 和集合 A 中的特征进行回归。LASSO 正则化参数可通过交叉验证选取，因此算法不需要额外的超参数。注意，我们不会对 S 中的特征施加惩罚，因为它们已经被选定，应该予以保留。这一步背后的直觉是，通常来说，之后生成的特征（从图中更远处获取信息）不应比之前选定的特征具有更强的预测能力

。接下来，根据新选定特征的数量，我们要么终止构造并返回当前的 S，要么将这些选定特征添加到 S 并继续递归构造。然后，我们需要生成新特征并将其添加到 A。为此，我们现在对所有邻近个体的旧特征进行聚合。最后，我们将这些特征添加到 A，并删除 A 中所有旧特征。

算法 1 中的最大迭代次数限制了我们能够从图中提取信息的距离。尽管每一步仅对一跳邻域中的邻居执行聚合，但通过反复执行聚合，我们能够构造出对 k 跳邻域具有信息性的特征。为说明这一点，我们给出一个示例。

示例 2（ReFeX 与多跳信息）。假设我们在 ReFeX-LASSO 中使用的某个基础特征是处理组邻居比例，

$$\rho_i = \frac{1}{d_i} \sum_{j \in \mathcal{N}_i} w_j,$$

并且假设我们仅限于均值聚合，即考察每个个体的邻居，并使用均值函数聚合它们的处理组邻居比例。我们将这一新特征称为 $\tilde{\rho}_i$。于是有

$$\begin{aligned}
\tilde{\rho}_i &= \frac{1}{d_i} \sum_{j \in \mathcal{N}_i} \rho_j \\
&= \frac{1}{d_i} \sum_{j \in \mathcal{N}_i} \frac{1}{d_j} \sum_{k \in \mathcal{N}_j} w_k \\
&= \sum_{j=1}^{n} \frac{A_{ij}}{d_i} \sum_{k=1}^{n} \frac{A_{jk}}{d_j} w_k \\
&= \sum_{j=1}^{n} \tilde{A}_{ij} \sum_{k=1}^{n} \tilde{A}_{jk} w_k \\
&= [\tilde{A}^2 w]_i,
\end{aligned}$$

其中，$\tilde{A}$ 和 A 与公式 (3) 中线性均值模型示例所定义的相同。注意，当且仅当 A_{ij}、A_{jk} 和 w_k 均为 1 时，被加项才为 1。换言之，如果忽略归一化项，该总和实质上表示 G 中从个体 i 出发并到达一个已处理个体的长度为 2 的路径数量。考虑归一化项后，它接近于从个体 i 出发的所有长度为 2 的路径中，此类路径所占的比例。显然，该特征能够反映个体 i 的 2 跳邻域信息。

上面的例子展示了递归的强大之处。它使我们能够获取关于更大邻域的信息，而无需实际查阅更大邻域中的所有个体。事实上，ReFeX-LASSO 中的 ReFeX 组件在计算复杂度方面

非常高效（Henderson 等人（2011）），因此这套流程非常适合在网络干扰普遍存在的在线平台上开展大规模实验。我们算法的另一项优势是，生成的所有协变量都与模型无关，或不依赖模型——我们不会根据任何特定的响应模型（或图模型）生成协变量。由于聚合函数不受限制，即使迭代预算不高，ReFeX 也能快速生成大量特征 T。除了刚才看到的接受处理的邻居所占比例外，我们还可以将求和作为聚合函数，从而获得每个个体的处理组邻居数量。一般而言，使用更复杂的聚合函数会产生更复杂的特征。因此，递归步骤能为每个个体提供丰富的特征。

经过少量修改可以看出，我们过程中的所有剪枝步骤都可以集中起来在事前完成，即在运行实验并观测结果之前完成。然后，在实验结束后，我们利用观测结果，从所生成的全部协变量中选定协变量。该方法具有一定优势，因此为完整起见，我们在下面的算法 2 中给出这一修改版的 ReFeX-LASSO，并将其称为 post-ReFeX-LASSO。

算法 2 post-ReFeX-LASSO

输入：图 $G = (V, \mathcal{E})$、分配向量 $w \in \{0,1\}^n$、最大迭代次数 T。输出：协变量集合 S。1: 初始化 $S = \{\}$。2: 对于每个节点/单位 i，构造 m基础特征，并将 m 基础特征添加到 S 中。3: 对于 $t = 1$ 到 T 执行4: 通过在一跳邻域中的邻居上，对 S 中上一轮迭代添加的特征进行聚合，递归构造特征。5: 将新构造的特征添加到 S。6: 结束循环7: 使用 LASSO 对 y 关于 w 以及来自 S 的特征进行回归。8: 保留 S 中选定的特征，并移除 S 中的其他特征。9: 返回 S。

post-ReFeX-LASSO 的一项操作优势在于，算法的两个部分——特征生成与选择——可以分开进行。然而，我们在实践中发现，post-ReFeX-LASSO 得出的估计值方差较大。我们对方差增大的解释有两点。首先，由于 ReFeX 生成的特征数量可能很多，将生成步骤与选择步骤分开似乎会使选择步骤不稳定。其次，post-ReFeX-LASSO 过程中生成的许多特征彼此相关，同时纳入所有这些特征会使所选特征的不确定性更大。因此，这会导致估计值的方差增大；只要操作上可行，我们便建议在所有使用情形中优先采用 ReFeX-LASSO，而非 post-ReFeX-LASSO。

4 使用无模型协变量进行推断

在上一节中，我们给出了一个序贯程序，该程序输出一组协变量 S，可用于估计 GATEs 时进行回归调整。本节专门讨论基于无模型协变量的推断。我们首先讨论如何使用 ReFeX-LASSO 或 post-ReFeX-LASSO 得到的无模型协变量进行回归调整。随后，我们展示 ReFeX-LASSO 的一项选择性质。然后，我们给出使用无模型协变量的 GATE 回归调整估计量的理论性质，并介绍一种为 τ 构造置信区间的简单方法。

4.1 估计

令 u_i^1、$\cdots$、u_i^K 表示 ReFeX-LASSO 或 post-ReFeX-LASSO 为个体 i 返回的 K 协变量，并令 $u_i = \left[u_i^1, \cdots, u_i^K\right]^T \in \mathbb{R}^K$ 表示个体 i 的完整特征向量。进一步令 $\hat{}g$ 表示一个函数，该函数将 (i、w、x_i、G) 映射为每个个体的 u_i。最后，我们用 n_c 表示对照组单位的数量，用 n_t 表示满足 $n_c + n_t = n$ 的处理组单位数量。

为估计 GATE，我们使用 u_i's 在对照组和处理组单位上分别拟合两个线性模型。理想情况下，我们希望存在向量 β_0、β_1，使得 $\beta_0^T u_i$ 和 $\beta_1^T u_i$ 能够很好地近似 f_0 和 f_1。具体而言，我们首先仅使用来自对照组的、具有 $\hat{}$ 观测值的数据运行普通最小二乘法，并得到 β_0。随后，我们再次运行普通最小二乘法，但这次仅使用来自处理组的观测值，并得到 β_1。同时，特征 u_i 是处理分配为 w 时收集到响应的所有特征。为了估计 GATE，我们关注的不是当时的 u_i 下的响应，而是当 $w = \mathbf{0}$ 或 $w = \mathbf{1}$ 时 u_i 本应产生的响应。因此，我们将 0 和 1 分别传入 $\hat{}g$，以获得全局对照和全局处理下的特征向量 u_i^{gc} 和 u_i^{gt}。

将系数估计值 $\hat{}\beta_1$ 和 $\hat{}\beta_0$ 与向量 u_i^{gc} 和 u_i^{gt} 结合起来，我们对 GATE 的估计值便是

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_1^T u_i^{gt} - \hat{\beta}_0^T u_i^{gc}). \tag{6}$$

尽管假设线性模型有所限制，但正如我们先前所讨论的，如果能够生成预测特征，那么线性模型可以很好地近似真实模型。ReFeX-LASSO 或 post-ReFeX-LASSO 有助于我们选择适合用于调整的特征，从而既降低估计值[2] 的方差，也减少忽略干扰时通常会产生的偏差。

4.2 选择性质

在深入探讨推断细节之前，我们先讨论 ReFeX-LASSO 的选择性质，并借鉴了关于序贯 LASSO（Luo和Chen（2014））的既有研究。为此，我们引入一些额外的符号。对于每次迭代 t，令 $\{u_1^t, u_2^t, \cdots, u_{i_t}^t\}$ 表示 ReFeX-LASSO 中 ReFeX 步骤生成的特征集合，并令 t 表示第 $\{u_1^t, u_2^t, \cdots, u_{i_t}^t\}$ 次迭代选定的特征集合

[2]事实上，在不存在干扰的情况下，Lin（2013）表明，进行线性调整只能提高精确度。

（注意，s_{*t} 可能包含先前迭代中选定的特征，因此这些特征不属于集合 $\{u_1^t, u_2^t, \cdots, u_{i_t}^t\}$）。此外，我们用 $\mathcal{R}(s)$ 表示由 s 中的特征张成的空间。

命题 4.1。对于 $t \geq 1$ 和任意 $j \in \{1, \cdots, i_{t+1}\}$，如果 $u_j^{t+1} \in \mathcal{R}(s_{*t})$，那么 $j \notin s_{*(t+1)}$。

第一个命题意味着两点。首先，在每次迭代中，我们都有一个满秩设计矩阵。其次，后续选择将忽略与现有特征高度相关的特征，因此也解释了为什么 post-ReFeX- LASSO 会导致具有较大方差的估计值。如果没有（非 post-）ReFeX-LASSO 的序贯程序，两个高度相关的特征可能会同时进入选择阶段。命题 4.2。我们的选择具有嵌套性，即 $s_{*1} \subseteq s_{*2} \subseteq \cdots \subseteq s_{*T}$。

第二个命题相对而言不言自明，并确保序贯程序确实产生嵌套的特征集，即通过取消对已选定特征的惩罚，我们能够将其保留在特征集中 S。尽管我们在 ReFeX- LASSO 中的选择程序与序贯 LASSO（Luo & Chen 2014）有很大不同，但上述两个命题的证明与 Luo & Chen (2014) 中的证明类似。我们的选择程序与序贯 LASSO 之间有两个关键区别。首先，我们并非在每次迭代中保留所有特征，而是舍弃先前迭代中未被选定的特征。其次，每次迭代中考虑的是新生成的特征，而不是现有特征。换言之，我们发现 Luo & Chen (2014) 中的分析对于这种程序变化仍然稳健。请注意，序贯 LASSO 可用于 post-ReFeX- LASSO（但不能用于 ReFeX-LASSO），因为对于 post-ReFeX-LASSO，我们会预先生成所有候选特征。这两个命题共同确立了 ReFeX-LASSO 中选择步骤的两个直观性质，我们预期这些性质会对我们的目的成立。其证明见附录 A。

4.3 一致性

现在，我们证明在 LASSO 一致性所需的标准假设下，post-ReFeX-LASSO 能够得到 GATE 的一致估计量。对于每个个体 i，我们将 post-ReFeX-LASSO 中 ReFeX 步骤生成的特征集合记为 $\{u_i^1, \cdots, u_i^M\}$。提及第 j 个特征向量时，我们省略下标 i，即 $u^j = [u_1^j, \cdots, u_n^j]^T$。此外，我们假设存在一个子集 $S_* \subset \{u^1, \cdots, u^M\}$，满足 $|S_*| = s$，使得 f_0 和 f_1 都是 S_* 中特征的线性函数，其系数向量分别为 β_0 和 β_1。最后，我们将估计 β_0 时的设计矩阵记为 U^0，将估计 β_1 时的设计矩阵记为 U^1。

定理 4.3. 假设存在一个常数 $C > 0$，使得

$$\max_{j=1,\cdots,M} \frac{\|u^j\|_2}{\sqrt{n}} \leq C,$$

并且，两个设计矩阵 U^0 和 U^1 在 S 上满足 (κ; 3)-受限特征值条件，则 $\hat{}\tau$ 对于 τ 是一致的。

定理 4.3 的证明见附录 A，其中主要使用了研究 LASSO ℓ_2-误差界的标准工具（Wainwright 2019）。定理 4.3 中的受限特征值（RE）条件，是证明系数向量的 ℓ_2-误差界时的一项标准假设。它限制了欧几里得空间中特定向量子集的曲率。其定义如下（Bickel 等人 2009，van de Geer & B¨uhlmann 2009，Raskutti 等人 2010）

定义 4.4. 若矩阵 X 在 S 上满足受限特征值（RE）条件，参数为（$\kappa; \alpha$），则

$$\frac{1}{n}\|\mathbf{X}\Delta\|_2^2 \geq \kappa\|\Delta\|_2^2 \qquad \text{for all } \Delta \in \mathbb{C}_\alpha(S),$$

其中 $\mathbb{C}_\alpha(S) := \{\Delta \in \mathbb{R}^d \mid \|\Delta_{S^c}\|_1 \leq \alpha\|\Delta_S\|_1\}$。

在定理 4.3 的假设下，我们现在能够证明，在至少一些简单情形下（例如以下情形），采用基于 LASSO 的特征选择时具有 GATE 一致性；在该示例情形中，我们的特征生成程序会输出两个简单特征。

命题 4.5. 假设我们进行处理概率为 $0 < p < 1$ 的伯努利随机实验，并且只生成两个特征：处理组邻居比例 ρ_i和处理组邻居数量 ν_i。此外，假设图 G 由大小为 $3 \leq m_c \leq M$ 的互不相交的团组成（m_c 是第 c 个簇的大小），其中 $M \geq 3$为某个正数常量。若真实的 f_0 和 f_1 仅关于 ρ_i呈线性，则 $\hat{\tau}$ 是 τ的一致估计量。

对 m_c 的下界是出于可识别性的考虑，因为当所有簇的大小均为 2 时，ρ_i和 ν_i 本质上相同，最终会得到完全重复的特征。还要注意，当所有 m_c均相等时，会出现完全共线性，因此在这种情况下，我们不会考虑区分这两个特征。虽然上述结果仅适用于简单情形，但其本身具有重要意义。在实践中，根据处理组邻居比例进行调整，并将所得估计值报告为 GATE 的估计值，并不少见（Saint-Jacques 等人，2019；Karrer 等人，2021）。上述命题表明，当我们只想区分处理组邻居比例和处理组邻居数量这两个协变量时，LASSO 是一种便捷的工具。

4.4 通过区块自助法构造置信区间

研究者通常不仅关注 GATE 的点估计，还希望了解估计值所包含的不确定性，例如通过置信区间来了解。ReFeX-LASSO 为 GATE 估计中的回归调整带来了灵活性，但没有免费的午餐，它也增加了推断的难度，即难以为 τ 构造置信区间。首先，与假设存在预言机模型的 Chin（2019）不同，这里的真实模型未知。其次，Chin（2019）中构造的特征并未使用观测结果。对于 ReFeX-LASSO，尽管由 ReFeX 构造的所有特征都不使用结果变量，但我们对协变量的选择取决于实际观测到的结果。因此，ReFeX-LASSO 得到的估计量没有明确的方差表达式。此外，由于我们的最终估计值取决于实际选定的协变量，我们需要采用某种类似于 Lee 等人（2016）所用选择后推断的技术。Lee 等人（2016）考虑的是在系数被 LASSO 选中的条件下，对系数构造置信

区间。然而，我们关注的是 τ 的置信区间，而非系数的置信区间；我们的估计值 $\hat{\tau}$ 是根据估计系数和选定的协变量计算得出的。由于这些复杂因素相互交织，我们无法直接套用任何已知的、适用于这一情形的推断结果。

让我们考虑一下当前所面临的推断问题的性质。一般而言，估计值的随机性不仅来自潜在结果的随机性，也来自分配向量的随机性。为了构造置信区间，我们需要量化这两种随机性来源如何影响对 GATE 的估计值。注意，由于我们知道分配向量的分布，因此给定特征的分布实际上也是已知的。我们尚未得到良好刻画的是：由分配向量的随机性所导致的特征选择程序的随机性。换言之，我们需要理解随机分配如何影响特征选择程序。

为了解决这一复杂问题，我们介绍一种基于区块自助法构造置信区间的方法。理想情况下，如果我们能够无限次地进行实验，就可以运行 2^n 次实验，并计算 GATE 的 2^n 个估计值。随后，便可以很容易地得到 τ 的置信区间。显然，我们面临的困难是：如何利用单个样本来近似样本的随机性。为此，我们采用区块自助法（Efron 1979；Efron & Tibshirani 1994；Cameron 等，2008）。采用这一方法的直觉是，个体的特征会根据 G 的特定图结构而相关，因此，通过对簇进行抽样（我们预计这些簇之间相对不连通），可以使自助抽样样本看起来类似于原始样本。另一方面，重新抽样个体将会失败，因为这种做法无法复现数据中的潜在相关结构。尽管我们不提供理论保证，但我们将展示，在实践中覆盖率良好，所得置信区间的宽度也较为合理。顺便指出，Kojevnikov（2021）的近期结果表明，对于某些高度程式化的网络过程，存在一种能够提供理论保证的区块自助法版本。

例3。考虑我们的社交网络 G由 C 个大小为 m 的互不相交的团 $\mathcal{C}_1, \cdots, \mathcal{C}_C$ 组成的情形。每个团内的个体彼此完全连接。此设置可视为 Basse 和 Feller（2018）研究的家庭实验的一个特例。在这种情况下，自然的做法是对所有 C 个团进行有放回抽样，从而获得自助抽样样本。对于满足某些技术假设的网络依赖过程，根据 Kojevnikov（2021）中的论证，这种抽样过程是正确的。假设我们有一个满足 Kojevnikov（2021）中假设的网络依赖过程 $\{Y_n, G_n\}$。为了使区块自助法具有一致性，即得到水平一致的置信区间，Kojevnikov（2021）中的假设 4.1 必须成立。简要采用该假设中的记号，很容易验证，在我们的情形下，$\delta_n(s_n) = m$, $\Delta_n(s_n, 2) = 0$，并且对于 $\forall s_n \geq \max_c \text{diam}(\mathcal{C}_c)$，$D_n(s_n) = m$，因为我们的图由大小相等且互不重叠的区块组成 m。此外，

$$\omega_n(i,j) = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are in the same cluster,} \\ 0 & \text{otherwise.} \end{cases}$$

并且对于所有 $j \in [n]$, $\omega_n(j) = 1$。根据这些取值，我们立即可以看出，只要满足 $m = o(n)$，Kojevnikov（2021）中的假设 4.1 就成立。由于使区块自助法具有一致性所需的其余假设仅涉及网络依赖过程本身，因此我们可以得出结论：对于 Kojevnikov（2021）中给出的网络依赖过程，在这个玩具模型中区块自助法是有效的。

我们在此介绍区块自助法的两个版本：一个用于采用 post-ReFeX-LASSO 的回归调整，另一个用于采用 ReFeX-LASSO 的回归调整。在给出这两种区块自助法的具体步骤之前，我们先介绍区块自助法中的关键要素——一种随机图聚类算法。我们的区块自助法需要将图划分为若干个簇。我们使用的通用算法是k 跳最大聚类（Ugander 和 Yin，2020），它是对 CKR 划分算法（Calinescu 等，2005）的简单改编。具体内容见算法 3。该算法根据随机初始条件对图进行随机聚类。当 $k = 1$ 时，该算法的计算开销较小，因为我们只需查看节点的直接邻居。此外，它会返回邻域型簇。作为与前述示例的联系，值得指出的是，如果我们的图由互不相交的全连接簇组成，那么 1 跳最大聚类就能将这些簇原样作为最终输出。一般而言，当 $k > 1$ 时，我们会得到更大的簇，这些簇以较少的节点为中心。

算法 **3** k-跳最大图聚类

输入：图 $G = (V, E)$。
输出：图聚类 $\mathcal{C}_1, \cdots, \mathcal{C}_c$。

对 $i \in V$ 执行
    $X_i \leftarrow \mathcal{U}(0, 1)$;
结束循环
对于 $i \in V$ 执行
    $i \leftarrow$argmax($[X_j$ 对于 $j \in B_k(i)]$);
结束循环
返回 $\mathcal{C}_1, \cdots, \mathcal{C}_c$.

我们首先介绍 post-ReFeX-LASSO 的区块自助法流程，如算法 4 所示。对于 post-ReFeX-LASSO，自助法流程更简单，因为特征生成与选择部分彼此分离。不同于通常的自助法（从个体中有放回地随机抽样），这里我们从图聚类算法得到的簇中有放回地随机抽样。其直觉是，个体的特征 u_i 会根据 G 的特定图结构而相关，因此，通过对簇进行抽样（我们预期这些簇之间相对独立），能够使自助抽样样本“看起来像”原始样本。不过需要特别注意的是，尽管自助抽样样本的期望样本量为 n，但如果簇的大小不一致，自助抽样样本最终的样本量可能会大得多或小得多。因此，我们运行图聚类算法 l 次，并针对每次聚类运行 B次区块自助法重复抽样。在算法 4 中，我们使用 $k = T^* + 1$ 进行 k-跳最大聚类，其中 T^* 是

特征仍被选中的迭代次数（因为如果在第 $(T^* + 1)$ 次迭代中没有特征被选中，那么干扰应该发生在 $(T^* + 1)$-跳邻域内）。

算法 4 post-ReFeX-LASSO 的区块自助法

输入：图 $G = (V, \mathcal{E})$、分配向量 $w \in \{0, 1\}^n$、自助样本数量 B。
输出：τ 的置信区间。

1: 收集运行 LASSO 前在 S 中生成的分配 w_i、特征 u_i^1、$\cdots$、u_i^M，每个个体 i 的结果变量 y_i。记录在该迭代中生成的某个特征被选中时的最大迭代次数 {v3}。
2: 使用 $k = T^* + 1$ 对 n 个个体进行 C 簇划分，采用 k 跳最大簇算法 $\mathcal{C}_1, \cdots, \mathcal{C}_C$.
3: 对于 $b = 1$ 至 B，执行
4: 　从 $\mathcal{C}_1$、$\cdots$、$\mathcal{C}_C$ 中有放回地抽取 C 个簇。
5: 　使用抽样簇中的个体构造第 b 个自助抽样样本（w^b、$u^{1,b}$、$\cdots$、$u^{M,b}$、y^b）簇。
6: 　使用 LASSO 将 y 对 w 以及 M 个特征进行回归。
7: 　使用选定特征和自助抽样样本计算估计值 $\hat{\tau}^b$。
8: 结束循环
9: 将第 2–8 行重复 ℓ 次，总共获得 $\ell \cdot B$ 个自助法估计值。
10: 计算第 $\alpha/2$ 个分位数 $q^*_{\alpha/2}$ 以及第 $(1 - \alpha/2)$ 个分位数 $q^*_{1-\alpha/2}$ 样本的所有自助法估计值 $\hat{\tau}^1, \ldots, \hat{\tau}^{\ell B}$。
11: 将 $\left[q^*_{\alpha/2}, q^*_{1-\alpha/2}\right]$ 作为 τ 的 $(1 - \alpha) \times 100\%$ 置信区间。

接下来，我们介绍使用 ReFeX-LASSO 的区块自助法版本，详见算法 5。请注意，我们不能直接使用相同的算法，因为它会同时进行特征生成和特征选择。与算法 4 相比，T^* 现在表示 ReFeX-LASSO 的停止时间。同时，与算法 4 类似，自助抽样样本仅用于 ReFeX-LASSO 的特征选择步骤。也就是说，在每次迭代中，我们仍使用相同的图 G 生成特征，但随后使用这些特征的自助抽样样本进行选择。使用原始图的直觉依据是，我们将图视为固定的，因此所有特征的相关结构均由该图诱导。因此，我们不会将所有抽样簇拼接在一起形成新图，以便在下一次迭代中生成特征。另一方面，如果我们确实认为图是由某种随机过程生成的，那么也可以通过将所有抽样簇拼接在一起，根据抽样个体重构图。

在上述两种算法中，我们使用了一种易于实现的随机图聚类算法。当然，这并非可供选择的唯一图聚类算法。顺便指出，实践者可以使用许多具有不同特性的图聚类算法（Nishimura & Ugander 2013, Spielman & Teng 2013, Awadelkarim & Ugander 2020, Shi & Chen 2020）。

我们以讨论如何恰当地选择簇的大小结束本节。

Alg算法 5 区块自助法 p 用于 ReFeX-LASSO

输入：图 $G=(V,\mathcal{E})$、分配向量 $w\in\{0,1\}^n$、自助样本数量 B。输出： τ的置信区间。1：收集每个个体 i的分配 w_i 和结果变量 y_i 。记录 ReFeX-LASSO 的停止时间 T^*。2：使用带有 $k=T^*+1$的 k跳最大聚类，将 n个体划分为 C个簇 $\mathcal{C}_1$、$\cdots$、$\mathcal{C}_C$。3：对于 $b=1$ 到 B，执行：4：从 $\mathcal{C}_1$、$\cdots$、$\mathcal{C}_C$ 中有放回地抽取 C 个簇。5：使用抽样簇中的个体构造第 b个自助抽样样本。6：使用原始样本进行特征生成，并使用自助抽样样本进行特征选择，重新运行 ReFeX-LASSO。7：使用上一步返回的协变量以及自助抽样样本，获得 τ的估计值 $\hat{}\tau^b$。8：循环结束。9：将第 2—8 行重复 ℓ 次，共获得 $\ell\cdot B$ 个自助法估计值。10：计算所有自助法估计值样本 $\hat{}\tau^1$、$\cdots$、$\hat{}\tau^{\ell B}$的 $\alpha/2$分位数 $q^*_{\alpha/2}$ 和（$1-\alpha/2$）分位数 $q^*_{1-\alpha/2}$。11：返回 $\left[q^*_{\alpha/2}、\ q^*_{1-\alpha/2}\right]$，将其作为 τ的（$1-\alpha$）$\times 100$%置信区间。

我们考虑三种情形，并通过 Kojevnikov（2021）中的启发式方法说明它们为何可能失效。尽管我们研究的问题与 Kojevnikov（2021）中的问题并不相同，而且我们的设定更为复杂，但我们并不认为弱于 Kojevnikov（2021）中条件的假设足以使我们的覆盖率达到良好水平。因此，我们认为，若要使区块自助法具有一致性，至少应满足 Kojevnikov（2021）中的假设。

我们考虑的第一个情形是有 $O(n)$ 个大小不恒定的簇。此时，区块大小的二阶绝对中心矩可能不趋于零（即 $n\to\infty$），但平均区块大小为 $O(1)$。这意味着，除非各簇的大小相对均匀，否则将违反 Kojevnikov（2021）中的假设 4.1。作为第二种情形，考虑有 $O(1)$ 个簇的情况。此时，最大区块大小必定为 $O(n)$ 阶，而平均区块大小至多为 $O(n)$，因此必然违反 Kojevnikov（2021）中的假设 4.1。一般而言，我们不希望簇过多或过少。最后，考虑这样一种情形：有 $\sqrt{n}-1$ 个大小为 $\sqrt{n}$ 的簇，以及 $\sqrt{n}$ 个大小为 1 的簇。此时，平均区块大小为 $O(n^{1/2})$ 阶，而区块大小的二阶绝对中心矩并非低阶量，这意味着该比率不会随 $n\to 0$ 趋于零，因此再次违反假设 4.1。这个最后的例子表明，簇大小并非只需避免簇过大或过小、过少或过多；相反，我们可以看到，不能存在大小量级不同的两组簇。总之，建议采用数量适中、大小大致处于同一量级的簇。如果网络不太密集，算法 4 和 5 是不错的默认选择。

5 模拟实验

本节中，我们通过模拟，在理论依据不足时为我们的方法提供经验性指导，并提供其有效性的经验证据。我们使用真实社交网络的 Facebook 100 数据集（Traud 等，2012）。该数据集中的网络是涵盖一百所学院和大学的完整在线友谊网络，采集自 Facebook 在2005年9月某一天的快照。模拟中，我们使用规模适中的斯沃斯莫尔学院学生网络。我们提取斯沃斯莫尔网络中最大的连通分量，得到一个包含 1,657 个节点和 61,049 条边的社交网络。该网络的直径为 6，平均两两距离为 2.32。由于该网络相当密集，当干扰较强时，很难估计 GATE。我们使用该网络来说明，即使对于这样的网络，仍然能够通过（后续）ReFeX-LASSO 获得相对较好的估计值。

我们采用成功概率为0.5的伯努利设计生成分配向量，并根据若干模型生成结果变量，这些模型中的网络干扰程度各不相同；模型汇总见表1和表2。我们将在第5.2节详细讨论这些结果模型。我们的模拟可视为半合成实验——我们使用真实的社交网络，但根据指定的模型生成结果。

第5.1节介绍了我们在模拟中进行比较的基准估计量。第5.2节讨论了用于生成不同干扰程度下结果的结果模型。第5.3节将使用无模型协变量的回归调整估计量与第5.1节介绍的实践中常用估计量进行比较，并表明该估计量在均方根误差方面表现良好。第5.4节考察通过区块自助法构造的置信区间的经验表现，并讨论该过程中的一些实际问题。

5.1 GATE 估计

构建无模型协变量的最终目标是将其用于 GATE 估计。我们首先考察使用无模型协变量的回归调整估计量的实证表现。具体而言，我们将其与实践中常用的两类估计量进行比较：(i) 均值差估计量；(ii) 在网络暴露模型下的 H´ajek 估计量（Manski，2013）。均值差估计量计算处理组单位的平均结果变量与对照组单位的平均结果变量之差：

$$\hat{\tau}^{DM} = \frac{1}{\sum_{i=1}^{n} W_i} \sum_{i=1}^{n} Y_i W_i - \frac{1}{\sum_{i=1}^{n} (1 - W_i)} \sum_{i=1}^{n} Y_i (1 - W_i).$$

显然，该估计量忽略了干扰，因此在干扰显著时会产生较大的偏差。

ATE 的基本 H´ajek 估计量定义为

$$\hat{\tau}^{\text{Hájek}} = \frac{\sum_{i=1}^{n} Y_i W_i / \mathbb{P}(W_i = 1)}{\sum_{i=1}^{n} \mathbb{I}(W_i = 1)/\mathbb{P}(W_i = 1)} - \frac{\sum_{i=1}^{n} Y_i (1 - W_i)/\mathbb{P}(W_i = 0)}{\sum_{i=1}^{n} \mathbb{I}(W_i = 0)/\mathbb{P}(W_i = 0)}.$$

这里我们将考虑一种能够处理干扰的 H´ajek 估计量。Manski（2013）研究了干扰下潜在结果分布的识别。一个具体的例子是，个体的结果变量仅取决于其自身的分配以及其邻居的分配分布。Ugander 等人（2013）进一步研究了分数暴露模型，该模型假设：如果个体接受处理，且其邻居中有 $q > 0.5$ 比例接受处理，那么该个体的结果变量就等于与分配向量 **1** 相对应的潜在结果。类似地，在这一暴露模型中，如果个体未接受处理，且其接受处理的邻居比例至多为 $1 - q$，那么该个体的结果变量就等于与分配向量 **0** 相对应的潜在结果。形式化地，$\forall w, w' \in \{0,1\}^n$，该分数暴露模型假设：

$$w_i = 1, \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} w_j \geq q \implies Y_i(w) = Y_i(\mathbf{1}),$$

and

$$w_i = 0, \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} w_j \leq 1 - q \implies Y_i(w) = Y_i(\mathbf{0}).$$

随后，我们可以使用 H´ajek 估计量，在伯努利设计下校正满足这些条件的概率。具体而言，我们定义事件 $E_i^{1\,q} = \{W_i = 1, \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} w_j \geq q\}$ $E_i^{0\,1-q} = \{W_i = 0, \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} w_j \leq 1 - q\}$ 以及。相应的分数暴露模型下的 H´ajek 估计量为

$$\hat{\tau}_{q,1-q}^{\text{Hájek}} = \frac{\sum_{i=1}^{n} Y_i \mathbb{I}(E_i^{1,q})/\mathbb{P}(E_i^{1,q})}{\sum_{i=1}^{n} \mathbb{I}(E_i^{1,q})/\mathbb{P}(E_i^{1,q})} - \frac{\sum_{i=1}^{n} Y_i \mathbb{I}(E_i^{0,1-q})/\mathbb{P}(E_i^{0,1-q})}{\sum_{i=1}^{n} \mathbb{I}(E_i^{0,1-q})/\mathbb{P}(E_i^{0,1-q})}. \tag{7}$$

该估计量通过考虑直接邻居的分配来处理干扰。如果我们仍然假设局部干扰，即只有个体的直接邻居会影响其响应，但希望设定完全不受任何假设约束，那么可以选择$q = 1$（注意，在这种情况下，H´ajek 估计量是一致的）。在我们的情形中，每个个体通常都有相当多的邻居；在独立伯努利分配下，我们预计不会观察到很多邻居全部接受处理或全部未接受处理的个体。作为偏差-方差折衷，我们选择 $q = 0.8$。

最后，我们还将我们的（post-）ReFeX-LASSO 回归调整估计量与两种针对特定特征进行调整的线性回归调整估计量进行比较。稍后在介绍第5.3节的模拟结果时，我们会详细描述这两种估计量。对于 post-ReFeX-LASSO 和 ReFeX-LASSO，我们选择 $T = 2$，并将基础特征设为处理组邻居比例、处理组邻居数量、邻域中连接已处理单元与对照单元的边所占比例，以及邻域中连接两个已处理单元的边所占比例。对于（post-）ReFeX-LASSO 中的聚合函数，我们使用均值和方差。

5.2 结果模型

这里我们介绍模拟研究中使用的结果模型。我们沿用命题 4.5 中的记号，使用 ρ_i 表示个体 i 的已处理直接邻居比例，并使用 ν_i 表示已处理直接邻居的数量。

我们首先考虑线性干扰下的估计。第一个模型是关于处理组邻居数量和处理组邻居比例的线性模型。Pouget-Abadie 等（2019）和 Chin（2019）也考虑了此类模型。具体而言，

$$f_0(w, G) = \alpha_0 + \xi_0 \rho_i + \gamma_0 \nu_i \tag{8}$$

and

$$f_1(w, G) = \alpha_1 + \xi_1 \rho_i + \gamma_1 \nu_i. \tag{9}$$

差值 $\alpha_1 - \alpha_0$ 可视为处理的主要效应，而系数（ξ_w, γ_w）分别表示 $w = 0$、1 如何影响个体对处理和对照的响应。特别是，如果 $\xi_w = \gamma_w = 0$ ，则不存在干扰，我们回到了在稳定单元处理值假设下进行 ATE 估计的常规设置。请注意，对于该模型，干扰不会超出一跳邻域，因此估计问题要容易得多。我们将这种响应模型称为简单线性干扰。

在引言中讨论线性均值模型的基础上，我们还考虑一种响应模型，其中干扰会传播到 k 跳邻域，适用于 $k \geq 2$。该模型可视为截断线性均值模型；我们不再将模型求和至无穷，而是在 $j = J$ 处截断， $J > 1$ 为某个数值。

模型类型	(α_0, α_1)	(ξ_0, ξ_1)	(γ_0, γ_1)
模型 0	(0, 2)	(0, 0)	(0, 0)
模型 1	(0, 2)	(1, 1.5)	(0.005, 0.0025)
模型2	(0, 2)	(1, 2)	(0.005, 0.01)

表1：模拟实验中使用的简单线性干扰结果模型（(8) 和 (9)）的参数。

总体而言，我们考虑以下线性干扰模型配置。表1和表2汇总了我们在模拟中考虑的模型配置。请注意，模型 0 不存在干扰。对于所有模型，误差项均相互独立，且服从方差为 1 的正态分布。这些结果模型中的真实 GATE（通过精确计算或基于斯沃斯莫尔网络进行蒙特卡洛估计）分别为 2、3.69、4.74、15、20、15 和 35。

除了线性干扰外，我们还考察了一个稍复杂的情形，即线性干扰假设不成立的情形。具体而言，我们考虑在 ρ_i和 ν_i上呈非线性的 f_0 和 f_1 。我们使用的非线性函数为 S 形，因此很难用任何线性模型 f_0 来近似

模型类型 α		β	γ	J
模型 3	1	5	2	2
模型 4	1	5	3	2
模型 5	1	5	1	3
模型 6	1	5	2	3

表2：模拟实验中使用的截断的均值线性结果模型的参数。

。报告模拟结果时，我们将蒙特卡洛估计 9.55 作为真实的 GATE。这里的目的是说明，即使存在违反线性干扰假设的非线性 f_0 和f_1 ，我们的方法仍能得到性能合理的估计量。这也呼应了我们之前的讨论。在进行 GATE 估计时，我们总是在观测/训练数据范围之外预测某个数据点，因此简单模型也可能相当可靠。

估计量	$\hat{\tau}^{\mathrm{DM}}$	$\hat{\tau}^{\text{H´ajek}}_{0.8,0.2}$	$\hat{\tau}_{\mathrm{frac}}$	$\hat{\tau}_{\mathrm{num}}$	post-ReFeX-LASSO	ReFeX-LASSO	$\hat{\tau}_{\mathrm{oracle}}$
模型 0	0.05	0.76	0.24	0.07	0.50	0.32	0.05
模型 1	1.53	1.02	0.36	1.22	1.54	0.70	0.25
模型2	2.06	1.41	0.47	1.49	1.49	0.59	0.24
模型 3	10.02	3.84	0.37	9.86	1.08	0.93	0.37
模型 4	15.02	5.60	0.56	14.72	1.68	1.59	0.56
模型 5	9.92	6.61	4.53	9.82	1.38	1.73	1.98
模型 6	29.67	22.31	18.22	29.46	2.42	2.47	4.45

表3：假设干扰为线性（简单线性干扰和截断线性均值模型）结果模型时，GATE 估计量的均方根误差。

5.3 模拟结果

我们研究了这些不同模型下各估计量的偏差和均方根误差（RMSE）。表3和表4展示了线性干扰下几种不同估计量的均方根误差和偏差。在这两张表中，我们展示了两种回归调整估计量的结果。$\hat{\tau}_{\mathrm{frac}}$ 是对处理组邻居比例进行调整的回归调整估计量，$\hat{\tau}_{\mathrm{num}}$ 则对处理组邻居数量进行调整。这两种估计量也在 Chin（2019）中进行了研究。我们还展示了预言机调整估计量

[3]我们在附录B中记录了这个模型。

估计量	$\hat{\tau}^{\mathrm{DM}}$	$\hat{\tau}^{\mathrm{H\acute{a}jek}}_{0.8,0.2}$	$\hat{\tau}_{\mathrm{frac}}$	$\hat{\tau}_{\mathrm{num}}$	post-ReFeX-LASSO	ReFeX-LASSO	$\hat{\tau}_{\mathrm{oracle}}$
模型 0	0.004	0.120	0.021	0.009	0.106	0.042	0.004
模型 1	-1.53	-0.68	-0.25	-1.22	-0.72	-0.004	-0.05
模型2	-2.06	-1.16	-0.39	-1.48	-0.56	-0.29	0.008
模型 3	-10.02	-3.67	-0.01	-9.85	0.28	0.19	-0.01
模型 4	-15.02	-5.46	0.003	-14.72	0.36	0.31	0.03
模型 5	-9.92	-6.55	-4.52	-9.82	-0.01	-0.24	0.68
模型 6	-29.67	-22.28	-18.21	-29.45	-0.21	-0.31	2.77

表4：假设干扰为线性干扰（简单线性干扰和截断线性均值模型）结果模型时，GATE 估计量的经验偏差。

$\hat{\tau}_{\mathrm{oracle}}$ 作为参照，它标示出在完全了解响应模型的情况下我们所能达到的最佳表现。请注意，在某些情况下，其他估计量可能优于预言机，因为预言机调整估计量仅表示我们使用了预言机控制协变量。协变量不可避免地具有随机性，而且我们并未对所有可能的分配向量取平均。此外，对于截断线性均值模型，真实协变量之间高度相关，导致预言机调整估计量的方差较大。最后，$\hat{\tau}^{\mathrm{DM}}$ 和$\hat{\tau}^{\mathrm{H\acute{a}jek}}_{0.8,0.2}$分别指简单的均值差估计量，以及前文提到的、在方程（7）中令 $q = 0.8$时的 H´ajek 估计量。

首先，考察模型 0 的结果，即不存在干扰的情形，可以发现，与 H´ajek 估计量相比，post-ReFeX-LASSO 和 ReFeX-LASSO 的表现都更好。其次，对于模型 1 和模型2，真实的干扰机制是简单的线性干扰。从表3和表4的前两行可以看出，如果未能考虑某个特征，偏差和/或均方根误差可能会很大。此外，ReFeX-LASSO 优于 post-ReFeX-LASSO，偏差和均方根误差显著更低，因为在这种情况下，ReFeX-LASSO 能够在第一次迭代后停止考虑更多特征。对于模型 3–6，底层模型是截断线性均值模型，它们之间唯一的区别是停止次数 J。对于具有 $J = 2$ 的模型（模型 3 和 4），干扰仍是局部的，即仅发生在个体的直接邻居范围内；而对于具有 $J = 3$ 的模型（模型 5 和 6），考虑来自两跳邻居的信息至关重要。我们的模拟结果验证了这一判断。我们发现，$\hat{\tau}_{\mathrm{frac}}$ 在模型 3 和 4 中表现良好，但在模型 5 和 6 中表现很差。对于这些更具挑战性的响应模型，post-ReFeX-LASSO 和 ReFeX-LASSO 得到的估计量都具有相对较小的偏差和均方根误差。

接下来讨论非线性模型，下面的表 5 展示了我们在该模型下的结果。在这种情况下，$\hat{\tau}_{\mathrm{frac}}$ 和 $\hat{\tau}_{\mathrm{num}}$ 表示与线性情形中相同的回归调整估计量。与均值差和 H´ajek 相比，ReFeX-LASSO 得到的估计量表现要好得多。此外，通过比较 $\hat{\tau}_{\mathrm{frac}}$、$\hat{\tau}_{\mathrm{num}}$ 和 ReFeX-LASSO，我们看到，与线性干扰情形一样，即使我们恰好调整了某个重要特征，但未能将所有相关特征纳入考虑，也会导致估计量

偏差较大、方差较大，或两者兼有。换言之，ReFeX-LASSO 有助于选择要调整的特征集，从而大幅降低偏差或方差。

估计量	$\hat{\tau}^{\mathrm{DM}}$	$\hat{\tau}^{\mathrm{H\acute{}ajek}}_{0.8,0.2}$	$\hat{\tau}_{\mathrm{frac}}$	$\hat{\tau}_{\mathrm{num}}$	post-ReFeX-LASSO	ReFeX-LASSO
Bias	-5.54	-2.72	-1.56	-2.72	1.29	1.33
RMSE	5.55	3.73	1.92	2.73	5.68	2.75

表 5：假设结果模型存在非线性干扰时，GATE 估计量的均方根误差和经验偏差（附录B）。

从这些模拟中我们得出，ReFeX-LASSO 能够识别出用于回归调整的重要特征，从而在许多模型设定下产生表现相对良好的估计量。我们还看到，ReFeX-LASSO 通常（尽管并非总是如此）明显优于 post-ReFeX-LASSO。这是因为我们按顺序选择特征，从而降低了方差。相比之下，Chin（2019）针对某些网络特征所考虑的标准回归调整估计量（在我们的模拟中为 $\hat{\tau}_{\mathrm{frac}}$ 和 $\hat{\tau}_{\mathrm{num}}$）如果未能选择正确的特征，可能会偏差很大。最后，基于暴露映射的估计量，例如分数暴露 H´ajek 估计量，如果存在与此类估计量所假设的暴露模型大不相同的干扰，也可能表现得相当糟糕。

5.4 GATE 的置信区间

在第4.4节中，我们介绍了一种通过区块自助法为 τ 构造置信区间的方法，并给出了基于图的区块构造的明确算法。现在，我们评估通过区块自助法得到的置信区间的经验覆盖率。本节始终关注 τ的90%置信区间。不同于上一节使用斯沃斯莫尔学院网络，我们采用Cai等人（2015）的农户网络；与斯沃斯莫尔学院网络相比，该网络规模更大且更稀疏。事实上，斯沃斯莫尔网络中2-hop 邻域的平均规模为1092.65，三跳邻域的平均规模为1622.27。因此，如果我们认为干扰超出了一跳邻域，那么在这样一个稠密图上，自助法将无法取得良好效果，因为很难构造出尊重原始样本结构的自助抽样样本[4]。另一方面，Cai 等人（2015）的农户网络密度较低，其中2-hop 邻域的平均规模为23.95，三跳邻域的平均规模为41.49。我们将在第6节中更详细地介绍该网络的背景和具体细节。一般而言，如果网络过于稠密，无法生成彼此充分隔离且均衡的簇，那么自助法就会失效。需要注意的是，农户网络本身具有一种自然的聚类方式，即根据每位农户所在的村庄进行聚类

[4]我们发现，当干扰是局部的（即仅限于直接邻居之间）时，区块自助法在斯沃斯莫尔网络上的覆盖率仍接近名义覆盖率。

；也就是说，每个村庄都可以视为网络中的一个簇。因此，在这里的模拟中，我们还展示了使用基于村庄的“神谕聚类”构造区块自助法置信区间的结果，该方法采用ReFeX-LASSO。最后，由于我们使用的是更稀疏的网络（这使得干扰更易于管理），我们考虑了线性模型的两组不同参数，使处理组邻居数量产生的效应更大（从而使GATE 估计更加困难）。表6展示了这些参数的取值；它们大致基于模型2（因此命名为2a和2b）

我们首先评估这种自助法的有效性。假设干扰是线性的，并考虑模型3-6以及模型2a和2b。我们固定 $\ell = 3$, $B = 100$ ，并通过重复整个过程100次来计算覆盖率。表 7和表8显示

模型类型	(α_0, α_1)	(ξ_0, ξ_1)	(γ_0, γ_1)
模型2a	(0, 2)	(1, 3)	(0.01, 0.025)
模型 2b	(0, 2)	(1, 3)	(0.05, 0.15)

表6：模拟实验中使用的简单线性干扰模型（（8）和（9））的其他参数。

post-ReFeX-LASSO 和 ReFeX-LASSO 的区块自助法所构造置信区间的覆盖率和平均长度。为展示使用区块自助法以及考虑分配向量随机性的必要性，我们还列出了使用朴素自助法构造置信区间的结果，其中我们仅对每个个体进行有放回抽样。

模型	post-ReFeX-LASSO	ReFeX-LASSO	朴素自助法	使用预言机聚类的自助法
模型2a	93%	92%	94%	92%
模型 2b	92%	96%	93%	95%
模型 3	90%	90%	83%	91%
模型 4	88%	87%	80%	91%
模型 5	91%	93%	84%	92%
模型 6	90%	91%	67%	93%

表 7：在线性干扰（简单线性干扰和截断线性均值模型）结果模型下，不同 GATE 自助法 90 % 置信区间的覆盖率。

从结果中可以看出，我们的区块自助法对 ReFeX-LASSO 的覆盖率接近名义覆盖率，对 post-ReFeX-LASSO 的覆盖率略差，但仍接近名义覆盖率。然而，朴素自助法无法提供具有名义覆盖率的置信区间。事实上，在某些情况下，基于朴素自助法的置信区间可能会导致非常差的覆盖率。如果使用与网络相关的预言机聚类，我们也能得到较好的置信区间。在网络中存在明显自然簇的情形下，

模型	post-ReFeX-LASSO	ReFeX-LASSO	朴素自助法	使用预言机聚类的自助法
模型2a	0.245	0.220	0.235	0.228
模型 2b	0.435	0.384	0.390	0.382
模型 3	0.403	0.380	0.330	0.414
模型 4	0.569	0.534	0.437	0.593
模型 5	0.552	0.549	0.431	0.567
模型 6	1.316	1.316	0.751	1.412

表 8：采用线性干扰（简单线性干扰和截断线性均值模型）结果模型时，GATE 的 90% 置信区间平均长度。

这些簇可以作为区块自助法的一个不错的默认选择。此外，如表 8 所示，ReFeX-LASSO 的区块自助法置信区间和 post-ReFeX-LASSO 的区块自助法置信区间长度都较为合理。

我们以一项模拟结束本节，说明在区块自助法程序中自适应选择用于 k-hop 最大聚类的 k 为何重要，以及将图划分为仅两个簇为何无法获得正确的覆盖率。为此，我们考虑使用 2-hop 最大和 3-hop 最大聚类将个体划分为簇，也考虑在不考虑底层图结构的情况下，将个体独立同分布地随机划分为五个簇。我们选择考察 2-hop 最大和 3-hop 最大，是因为在模拟中我们发现，在大多数情况下，ReFeX-LASSO 在选取了关于 2 跳邻域的特征后便会停止。对于 Cai 网络，2-hop 最大聚类和 3-hop 聚类平均分别产生 267 个和 269 个簇。我们选择将它们与五簇划分进行比较，因为五远少于使用 k-hop 最大聚类可能得到的簇数。我们使用 ReFeX-LASSO，为模型 6 用这些新簇重新运行区块自助法程序。表 9 显示了置信区间的覆盖率。可以看到，与表 7 中基于自适应 k-hop 最大聚类的区块自助法所达到的 91% 覆盖率相比，这三种聚类方法均未能达到名义覆盖率。尤其是，完全忽略图结构（“五个簇”）会导致置信区间的覆盖率极差。

模型	2-hop 最大	3-hop 最大	五个簇
模型 6	84%	89%	45%

表 9：以模型 6 作为真实结果模型时，采用不同图聚类算法计算 GATE 的区块自助法 90% 置信区间的覆盖率。

6 真实数据示例

本节中，我们希望将所提出的方法应用于一个已知存在干扰的真实实验；在这种情况下，均值差等简单估计量会给出较差的 GATE 估计值。我们考察 Cai 等人（2015）所述干预的数据。他们设计了一项随机实验，以研究社交网络在中国农村保险采用中的作用。具体而言，他们向随机抽取的一部分农民提供了有关某种保险产品的密集信息宣讲。Cai 等人（2015）发现，保险知识的传播推动了产品采用中的网络效应。因此，这些数据非常适合我们的研究目的，因为我们可以确定稳定单元处理值假设遭到违反，不应依赖简单的均值差估计值来估计 GATE。此外，尽管我们知道网络效应确实存在，但要像 Aronow 和 Samii（2017）那样定义一个精确的暴露模型并不容易。因此，Chin（2019）所做的分析存在局限，因为其中只考虑了四个预先指定的特征，回归调整估计量因而隐含地假设了某种暴露模型。我们重新考察这一实验，并使用我们的方法估计 GATE。

在 Cai 等人（2015）的原始现场实验中，密集信息说明会分两轮举行，因此设有四个不同的处理组。为便于分析，我们按照 Chin（2019）的做法，将两轮密集信息说明会视为同一个处理组，从而把原始现场实验简化为二元随机实验。与 Cai 等人（2015）的做法相同，我们将天气保险采用的二元指示变量设为结果变量，并排除处理或响应信息缺失的村民，以及网络信息缺失的村民。我们还将所有村庄合并为一个社交网络，并用 G 表示这个社交网络。总之，该网络包含 4,382 个节点和 17,069 条边。我们在第5.4节中使用的也是这个网络。

我们方法的第一步是生成无模型协变量。我们使用与前文模拟部分完全相同的一组基础特征——处理组邻居比例、处理组邻居数量、邻域中连接已处理单元与对照单元的边所占比例，以及邻域中连接两个已处理单元的边所占比例。随后，我们使用 ReFeX-LASSO 生成一组协变量，采用均值和方差聚合函数（同样与前文模拟部分一致），并通过使用线性模型对这些协变量进行调整来估计 GATE。我们将区块自助法得到的标准误估计值与 Chin（2019）计算的结果进行比较。

表 10 展示了所得的 GATE 估计值，其中 $\hat{\tau}_{\text{chin}}$ 是 Chin（2019）中的估计量，该估计量针对四个协变量进行调整：处理组邻居比例、处理组邻居数量、2-hop 邻域中的处理组邻居比例，以及 2-hop 邻域中的处理组邻居数量。同时，$\hat{\tau}_{\text{num}}$ 仅针对处理组邻居数量进行调整，而 $\hat{\tau}_{\text{refex-lasso}}$ 是基于 ReFeX-LASSO 的调整估计量。DM 指均值差估计量。H´ajek 1hop 假设 1-hop 邻域采用分数暴露模型，而 H´ajek 2hop 假设 2-hop 邻域采用分数暴露模型，即我们使用公式（7），但改为考虑两跳邻居。其直观解释是，有时并非直接邻居、而是直接邻居的邻居的个体也会产生影响，而

估计量	估计值	标准误
DM	0.078	——
H´ajek 1跳（$q = 0.75$）	0.163	——
H´ajek 2跳（$q = 0.75$）	0.167	——
$\hat{\tau}_{\text{chin}}$	0.122	0.056
$\hat{\tau}_{\text{num}}$	0.178	0.027
$\hat{\tau}_{\text{ReFeX-LASSO}}$	0.178	0.043

表 10：Cai 等人（2015）中，保险采用的全局平均处理效应的不同估计量的估计值和标准误。

通过针对 2-hop 邻域采用分数暴露模型，我们能够在暴露模型中将这些个体纳入考虑。我们注意到，$\hat{\tau}_{\text{num}}$ 和 $\hat{\tau}_{\text{refex-lasso}}$ 给出了相同的估计值；事实上，ReFeX-LASSO 选出的唯一协变量就是处理组邻居数量。与 $\hat{\tau}_{\text{chin}}$ 相比，$\hat{\tau}_{\text{refex-lasso}}$ 的标准误更小，效应估计值更大。最后，尽管 $\hat{\tau}_{\text{num}}$ 和 $\hat{\tau}_{\text{refex-lasso}}$ 给出的估计值几乎相同（精确到小数点后三位），但我们看到前者的标准误更小。原因有两个。首先，自助法通常较为保守。其次，由于涉及随机选择程序，ReFeX 估计值的方差应当更大。

7 讨论

本文提出了一种在存在网络干扰时，对全局平均处理效应（GATE）进行估计和推断的方法。我们开发了一套程序，无需预先指定暴露映射或结果模型即可估计 GATE。我们还提出了一种使用区块自助法构建 GATE 置信区间的方法。我们通过模拟和一个真实数据示例评估了该方法。

本文尚未探讨许多值得进一步研究的方向。首先，我们的结果仅适用于满足均匀性假设的设计（例如伯努利设计）；这当然有所局限，但也提供了一个有用的基准。我们尤其希望探究如何将本研究扩展到违反均匀性假设的设计，例如整群随机设计。由于我们用于调整的协变量可能与处理分配相关，这一问题颇具挑战性。其次，尽管模拟结果表明区块自助法在实践中表现良好，但除简单的玩具设定外，我们尚无正式的理论结果。第三，除线性调整外，我们也可能希望采用完全非线性的模型，利用 ReFeX-LASSO 特征生成与选择过程返回的协变量来估计结果。

致谢

本研究部分得到了 ARO MURI 奖项 #W911NF-20-1-0252 和 NSF CAREER 奖项 #2143176 的支持。

参考文献

Aronow, P. M. 与 Samii, C.（2017），‘在一般干扰下估计平均因果效应：社会网络实验的应用’，《应用统计年鉴》11(4)，1912–1947。Awadelkarim, A. 与 Ugander, J.（2020），‘用于平衡图划分的优先级重流算法’，载于《第26届 ACM SIGKDD知识发现与数据挖掘国际会议论文集》，KDD ’20，计算机协会，纽约，纽约州，美国，第1877–1887页。Basse, G. 与 Feller, A.（2018），‘存在干扰时的两阶段实验分析’，《美国统计学会会刊》113(521)，41–55。Basse, G. W. 与 Airoldi, E. M.（2018），‘任意干扰和网络干扰下基于设计的因果推断与 A/B测试的局限性’，《社会学方法论》48(1)，136–151。Bickel, P. J.、Ritov, Y. 与 Tsybakov, A. B.（2009），‘LASSO 与 Dantzig选择器的同时分析’，《统计年鉴》37(4)，1705–1732。Bousquet, O.、Boucheron, S. 与 Lugosi, G.（2004），《统计学习理论导论》，施普林格·柏林·海德堡，柏林、海德堡，第169–207页。Bramoull´e, Y.、Djebbari, H. 与 Fortin, B.（2009），‘通过社交网络识别同伴效应’，《计量经济学杂志》150(1)，41–55。Cai, J.、De Janvry, A. 与 Sadoulet, E.（2015），‘社交网络与投保决策’，《美国经济学杂志：应用经济学》7(2)，81–108。Calinescu, G.、Karloff, H. 与 Rabani, Y.（2005），‘0-扩展问题的近似算法’，《SIAM计算杂志》34(2)，358–372。Cameron, A. C.、Gelbach, J. B. 与 Miller, D. L.（2008），‘基于自助法的聚类误差推断改进’，《经济学与统计学评论》90(3)，414–427。Chin, A.（2019），‘通过回归调整估计存在干扰的实验中的总体处理效应’，《因果推断杂志》7(2)，20180026。Cox, D. R.（1958），《实验设计》，纽约，Wiley。Deng, A.、Xu, Y.、Kohavi, R. 与 Walker, T.（2013），‘利用实验前数据提高在线对照实验的敏感性’，载于《第六届 ACM网络搜索与数据挖掘国际会议论文集》，WSDM ’13，计算机协会，纽约，纽约州，美国，第123–132页。Eckles, D.、Karrer, B. 与 Ugander, J.（2017），‘网络中的实验设计与分析：减少干扰造成的偏差’，《因果推断杂志》5(1)，20150021。

Efron, B. (1979), ‘自助法：再论刀切法’, 《统计年鉴》7(1), 1 – 26. Efron, B. & Tibshirani, R. (1994), 《自助法导论》, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis. Halloran, M. E. & Struchiner, C. J. (1995), ‘传染病中的因果推断’, Epi- demiology 6(2), 142–151. Hamilton, W., Ying, Z. & Leskovec, J. (2017), 大型图上的归纳式表示学习, 载于 I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan & R. Garnett 编, ‘神经信息处理系统进展’, 第 30 卷, Curran Associates, Inc. Hammond, D. K., Vandergheynst, P. & Gribonval, R. (2011), ‘基于谱图论的图小波’, Applied and Computational Harmonic Analysis 30(2), 129–150. Harshaw, C., S¨avje, F., Eisenstat, D., Mirrokni, V. & Pouget-Abadie, J. (2022), 线性暴露—响应模型下二部实验的设计与分析, 载于 ‘第 23 届 ACM 经济学与计算会议论文集’, EC ’22, 计算机协会, 纽约, NY, USA, p. 606. Henderson, K., Gallagher, B., Li, L., Akoglu, L., Eliassi-Rad, T., Tong, H. & Faloutsos, C. (2011), 你认识谁很重要：利用递归结构特征进行图挖掘, 载于 ‘第 17 届 ACM SIGKDD知识发现与数据挖掘国际会议论文集’, KDD ’11, 计算机协会, 纽约, NY, USA, pp. 663–671. Holland, P. W. (1986), ‘统计学与因果推断’, 《美国统计学会会刊》81(396), 945–960. Hong, G. & Raudenbush, S. W. (2006), ‘评估幼儿园留级政策’, 《美国统计学会会刊》101(475), 901–910. Imbens, G. W. & Rubin, D. B. (2015), 《统计学、社会科学与生物医学科学中的因果推断：导论》, 剑桥大学出版社. Karrer, B., Shi, L., Bhole, M., Goldman, M., Palmer, T., Gelman, C., Konutgan, M. & Sun, F. (2021), 大规模网络实验, 载于 ‘第 27 届 ACM SIGKDD知识发现与数据挖掘会议论文集’, KDD ’21, 计算机协会, 纽约, NY, USA, pp. 3106–3116. Kipf, T. N. & Welling, M. (2017), 基于图卷积网络的半监督分类, 载于 ‘国际学习表征会议’. Kojevnikov, D. (2021), ‘网络依赖过程的自助法’, arXiv 预印本 arXiv:2101.12312 .

Lee, J. D., Sun, D. L., Sun, Y. & Taylor, J. E. (2016), ‘精确后选择推断及其在 LASSO 中的应用’, 《统计年鉴》44(3), 907 – 927. Lin, W. (2013), ‘关于实验数据回归调整的不可知论札记：重新审视 Freedman 的批评’, 《应用统计年鉴》7(1), 295 – 318. Luo, S. & Chen, Z. (2014), ‘超高维特征空间中特征选择的序贯 LASSO 与扩展贝叶斯信息准则’, 《美国统计学会会刊》109(507), 1229–1240. Manski, C. F. (1993), ‘内生社会效应的识别：反射问题’, 《经济研究评论》60(3), 531–542. Manski, C. F. (2013), ‘社会互动下治疗反应的识别’, 《计量经济学杂志》 16(1), S1–S23. Moffit, R. A. (2001), 政策干预、低水平均衡与社会互动, 载于 ‘社会动力学’, 麻省理工学院出版社. Nishimura, J. & Ugander, J. (2013), 重流式图划分：用于高级均衡的简单通用算法, 载于 ‘第 19 届 ACM SIGKDD知识发现与数据挖掘国际会议论文集’, KDD ’13, 计算机协会, 纽约, NY, USA, p. 1106–1114. Pouget-Abadie, J., Saint-Jacques, G., Saveski, M., Duan, W., Ghosh, S., Xu, Y. & Airoldi, E. M. (2019), ‘在实验平台上检验任意干扰’, 《生物统计学》106(4), 929–940. Raskutti, G., Wainwright, M. J. & Yu, B. (2010), ‘相关高斯设计的受限特征值性质’, 《机器学习研究杂志》11(78), 2241–2259. Rosenbaum, P. R. (2007), ‘随机实验中的个体间干扰’, 《美国统计学会会刊》102(477), 191–200. Rubin, D. B. (1974), ‘估计随机化与非随机化研究中的处理因果效应’, 《教育心理学杂志》66(5), 688. Saint-Jacques, G., Varshney, M., Simpson, J. & Xu, Y. (2019), ‘利用自我中心簇测量 LinkedIn 上的网络效应’. Shi, L. & Chen, B. (2020), ‘图聚类算法的比较与基准评测’, arXiv 预印本 arXiv:2005.04806 . Sinclair, B., McConnell, M. & Green, D. P. (2012), ‘检测溢出效应：多层实验的设计与分析’, 《美国政治学杂志》56(4), 1055–1069.

Sobel, M. E. (2006), ‘住房流动性随机研究揭示了什么？’, 《美国统计学会会刊》 101(476), 1398–1407. Spielman, D. A. & Teng, S.-H. (2013), ‘针对大规模图的局部聚类算法及其在近线性时间图划分中的应用’, SIAM计算杂志 42(1), 1–26. Tibshirani, R. (1996), ‘通过 LASSO 进行回归收缩与选择’, 英国皇家统计学会会刊. B辑（方法论） 58(1), 267–288. Traud, A. L., Mucha, P. J. & Porter, M. A. (2012), ‘Facebook网络的社会结构’, 物理学A：统计力学及其应用 391(16), 4165–4180. Ugander, J., Karrer, B., Backstrom, L. & Kleinberg, J. M. (2013), 图簇随机化：网络对多个宇宙的暴露，载于 I. S. Dhillon, Y. Koren, R. Ghani, T. E. Senator, P. Bradley, R. Parekh, J. He, R. L. Grossman & R. Uthurusamy 编，《第19届 ACM SIGKDD知识发现与数据挖掘国际会议, KDD 2013，美国伊利诺伊州芝加哥，2013年8月11—14日》，ACM，第329–337页。 Ugander, J. & Yin, H. (2020), ‘随机图簇随机化’, arXiv预印本 arXiv:2009.02297。van de Geer, S. A. & B¨uhlmann, P. (2009), ‘关于证明 LASSO 预言机结果所用的条件’, 电子统计学杂志 3(none), 1360–1392。von Luxburg, U. & Sch¨olkopf, B. (2011),《统计学习理论：模型、概念与结果》，第10卷，Elsevier North Holland，阿姆斯特丹，荷兰，第651–706页。Wainwright, M. J. (2019),《高维统计：非渐近视角》，剑桥统计与概率数学丛书，剑桥大学出版社。Yu, C. L., Airoldi, E. M., Borgs, C. & Chayes, J. T. (2022), ‘在网络结构未知的随机实验中估计总处理效应’, 美国国家科学院院刊 119(44), e2208975119。

证明

一旦我们认识到，只要纳入惩罚项的特征与已选择的特征不重叠，就可以直接采用 Luo & Chen (2014) 中的证明，那么命题 4.1 和命题 4.2 的证明就与 Luo & Chen (2014) 中命题 1 和命题 2 的证明完全相同。也就是说，尽管我们的序贯选择程序不同于 Luo & Chen (2014) 中的程序，但我们具备使这两个命题成立的相同性质。

命题 4.1 的证明。我们用 $X(s)$ 表示在 s 中包含特征的设计矩阵，即若 $|s| = m$，则 $X(s)$ 是一个 $n \times m$ 矩阵。在第 $(t+1)$ 次迭代中，$-th$ β 将是一个 $(|s_{*t}| + i_{t+1})$ 维向量，我们用 $\beta(s)$ 表示一个 $|s|$ 维向量，其坐标仅包含 β 中属于 s 的坐标。最后，我们用 A_{t+1} 表示集合 $\{u_1^{t+1}, u_2^{t+1}, \cdots, u_{i_{t+1}}^{t+1}\}$。

首先注意到，由于 $u_j^{t+1} \in \mathcal{R}(s_{*t})$，存在 $\exists v \in \mathbb{R}^{|s_{*t}|}$ 使得 $u_j^{t+1} = X(s_{*t})v$。现在我们考虑第 $(t+1)$ 次迭代时的目标函数 l_{t+1}。

$$\begin{aligned} l_{t+1} &= \|y - X(s_{*t})(\beta(s_{*t}) + \beta(\{j\})v) - X(A_{t+1}/\{j\})\beta(A_{t+1}/\{j\})\|_2^2 \\ &\quad + \lambda\left(|\beta(\{j\})| + \|\beta(A_{t+1}/\{j\})\|_1\right) \\ &= \|y - X(s_{*t})\tilde{\beta}(s_{*t}) - X(A_{t+1}/\{j\})\beta(A_{t+1}/\{j\})\|_2^2 \\ &\quad + \lambda\left(|\beta(\{j\})| + \|\beta(A_{t+1}/\{j\})\|_1\right) \\ &\geq \|y - X(s_{*t})\tilde{\beta}(s_{*t}) - X(A_{t+1}/\{j\})\beta(A_{t+1}/\{j\})\|_2^2 \\ &\quad + \lambda\|\beta(A_{t+1}/\{j\})\|_1 \end{aligned}$$

因此，当 l_{t+1} 被最小化时，$\beta(\{j\})$ 必须为 0，且 $j \notin s_{*(t+1)}$。 □

命题 4.2 的证明。再次考虑第 $(t+1)$ 次迭代时的目标函数。

$$l_{t+1} = \|y - X(s_{*t})\beta(s_{*t}) - X(A_{t+1})\beta(A_{t+1})\|_2^2 + \lambda\|\beta(A_{t+1})\|_1.$$

对 l_{t+1} 关于 $\beta(s_{*t})$ 求导，我们有

$$\frac{\partial l_{t+1}}{\partial \beta(s_{*t})} = -2X^T(s_{*t})y + 2X^T(s_{*t})X(s_{*t})\beta(s_{*t}) + 2X^T(s_{*t})X(A_{t+1})\beta(A_{t+1}).$$

令上述导数为零，我们得到

$$\hat{\beta}(s_{*t}) = [X^T(s_{*t})X(s_{*t})]^{-1}X^T(s_{*t})[y - X(A_{t+1})\beta(A_{t+1})]. \tag{10}$$

将（10）代入目标函数，我们得到

$$\begin{aligned} l_{t+1} &= \|y - X(s_{*t})\beta(s_{*t}) - X(A_{t+1})\beta(A_{t+1})\|_2^2 + \lambda\|\beta(A_{t+1})\|_1 \\ &= \|y - X(s_{*t})[X^T(s_{*t})X(s_{*t})]^{-1}X^T(s_{*t})[y - X(A_{t+1})\beta(A_{t+1})] - X(A_{t+1})\beta(A_{t+1})\|_2^2 \\ &\quad + \lambda\|\beta(A_{t+1})\|_1 \\ &= \|(I - X(s_{*t})[X^T(s_{*t})X(s_{*t})]^{-1}X^T(s_{*t}))y \\ &\quad - (I - X(s_{*t})[X^T(s_{*t})X(s_{*t})]^{-1}X^T(s_{*t}))X(A_{t+1})\beta(A_{t+1})\|_2^2 \\ &\quad + \lambda\|\beta(A_{t+1})\|_1. \end{aligned}$$

ˆ ˆ 因此，最小化 l_{t+1} 不会影响 β（s_{*t}），且 β（s_{*t}）几乎必然非零。 □

现在我们证明定理 4.3。我们将使用关于 LASSOℓ_2误差界的标准结果。回顾 Wainwright（2019）的以下结果：

引理 A.1. 假设 $y = X\theta^* + w$ ($X \in \mathbb{R}^{n\times d}$)，并考虑正则化参数 $\lambda_n \geq 2\|\frac{\mathbf{X}^T w}{n}\|_\infty$ 严格为正的拉格朗日 Lasso。进一步假设 θ^* 的支撑集是基数为 s 的子集 S，且设计矩阵在 S 上满足 (κ; 3)-受限特征值条件，则

$$\|\hat{\theta} - \theta^*\|_2 \leq \frac{3}{\kappa}\sqrt{s}\lambda_n.$$

我们可以证明，如果设计矩阵是 C–列归一化的，即

$$\max_{j=1,\cdots,d} \frac{\|X_j\|_2}{\sqrt{n}} \leq C,$$

那么，选择 $\lambda_n = 2C\sigma(\sqrt{2\log\frac{d}{n}} + \delta)$ 至少以 $1 - 2e^{-\frac{n\delta^2}{2}}$ 的概率成立。因此，我们继续进行主要证明。

证明。注意，根据无穷范数的定义，$\|\frac{\mathbf{X}^T w}{n}\|_\infty$ 对应于 d 个均值为零的高斯随机变量的绝对最大值，且每个变量的方差至多为 $\frac{C^2\sigma^2}{n}$。因此，根据高斯尾界，我们有

$$\mathbb{P}\left(\left\|\frac{\mathbf{X}^T w}{n}\right\|_\infty \geq C\sigma\left(\sqrt{\frac{2\log d}{n}} + \delta\right)\right) \leq 2e^{-\frac{n\delta^2}{2}}.$$

□

采用这一特定的 λ_n 选择后，该引理给出了上界

$$\|\hat{\theta} - \theta^*\|_2 \leq \frac{6C\sigma}{\kappa}\sqrt{s}\left(\sqrt{\frac{2\log d}{n}} + \delta\right) \tag{11}$$

且具有与 Wainwright（2019）相同的高概率。

现在我们可以证明一致性了。首先注意到

$$\begin{aligned}
|\hat{\tau} - \tau| &= \left|\frac{1}{n}\sum_{i=1}^{n}\left[(\hat{\beta}_1 - \beta_1^*)^T u_i^{gt} - (\hat{\beta}_0 - \beta_0^*)^T u_i^{gc}\right]\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\left|\left[(\hat{\beta}_1 - \beta_1^*)^T u_i^{gt} - (\hat{\beta}_0 - \beta_0^*)^T u_i^{gc}\right]\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}(\|\hat{\beta}_1 - \beta_1^*\|_2\|u_i^{gt}\|_2 + \|\hat{\beta}_0 - \beta_0^*\|_2\|u_i^{gc}\|_2) \\
&\leq C\sqrt{M}(\|\hat{\beta}_1 - \beta_1^*\|_2 + \|\hat{\beta}_0 - \beta_0^*\|_2)
\end{aligned}$$

设 n_0 为对照组单位的数量， n_1 为处理组单位的数量。根据强大数定律， $\frac{n_0}{n} \xrightarrow{a.s.} 1-p$ 和 $\frac{n_1}{n} \xrightarrow{a.s.} p$。由于设计矩阵 U^0 和U^1满足受限特征值条件，根据界（11）， $\|\hat{}\beta_1 - \beta_1^*\|_2$ 和 $\|\hat{}\beta_0 - \beta_0^*\|_2$都依概率收敛于 0。因此， $\hat{}\tau \xrightarrow{\mathbb{P}} \tau$。

命题 4.5 的证明。我们证明，对于命题 4.5 中的设定，这些设计矩阵以趋近于 1 的概率满足受限特征值条件。在证明中，设计矩阵的第一列表示处理组邻居比例，第二列表示处理组邻居数量。我们引入一个额外的记号：对于每个个体 i，我们用 m_i 表示个体 i 所属簇的大小。我们证明对照组单位 U^0的设计矩阵的情形。类似的证明也适用于 U^1。中心化后，我们用于估计 β_0 的设计矩阵为

$$\tilde{U}^0 = \begin{bmatrix} \frac{1}{n_0}\sum_{i:W_i=0}(u_i^1-\bar{u}^1)^2 & \frac{1}{n_0}\sum_{i:W_i=0}(u_i^1-\bar{u}^1)(u_i^2-\bar{u}^2) \\ \frac{1}{n_0}\sum_{i:W_i=0}(u_i^1-\bar{u}^1)(u_i^2-\bar{u}^2) & \frac{1}{n_0}\sum_{i:W_i=0}(u_i^2-\bar{u}^2)^2 \end{bmatrix}.$$

这里 $\bar{u}^1 = \frac{1}{n_0}\sum_{i:\ W_i=0} u_i^1$ 且 $\bar{u}^2 = \frac{1}{n_0}\sum_{i:\ W_i=0} u_i^2$。由于真实的 β_0 仅在第一个特征上非零， $\mathbb{C}_3(S) = \{\Delta \in \mathbb{R}^2 :\ |\Delta_2| \le 3|\Delta_1|\}$。对于这样的 Δ，我们有

$$\frac{1}{n_0}\|\tilde{U}^0\Delta\|_2^2 = \Delta_1^2\frac{1}{n_0}\sum_{i:W_i=0}(u_i^1-\bar{u}^1)^2 + 2\Delta_1\Delta_2\frac{1}{n_0}\sum_{i:W_i=0}(u_i^1-\bar{u}^1)(u_i^2-\bar{u}^2) + \Delta_2^2\frac{1}{n_0}\sum_{i:W_i=0}(u_i^2-\bar{u}^2)^2$$

注意，由于 $|\Delta_2| \le 3|\Delta_1|$, $\Delta_1\Delta_2 \ge -|\Delta_1||\Delta_2| \ge -\frac{1}{3}\Delta_2^2$。因此，

$$\begin{aligned} \frac{1}{n_0}\|\tilde{U}^0\Delta\|_2^2 \ge{}& \frac{1}{n_0}\sum_{i:W_i=0}(u_i^1-\bar{u}^1)^2\Delta_1^2 \\ &+ \left(\frac{1}{n_0}\sum_{i:W_i=0}(u_i^2-\bar{u}^2)^2 - \frac{1}{3}\frac{1}{n_0}\sum_{i:W_i=0}(u_i^1-\bar{u}^1)(u_i^2-\bar{u}^2)\right)\Delta_2^2. \end{aligned} \tag{12}$$

为简化记号，令① $= \frac{1}{n_0}\sum_{i:\ W_i=0}(u_i^1-\bar{u}^1)^2$, ② $= \frac{1}{n_0}\sum_{i:\ W_i=0}(u_i^2-\bar{u}^2)^2$ 以及 ③ $= \frac{1}{n_0}\sum_{i:\ W_i=0}(u_i^1-\bar{u}^1)(u_i^2-\bar{u}^2)\Delta_2^2$。现在，我们分别分析各项。

$$\begin{aligned} ① &= \frac{1}{n_0}\sum_{i:W_i=0}(u_i^1-\bar{u}^1)^2 \\ &= \frac{1}{n_0}\sum_{i:W_i=0}(u_i^1)^2 - (\bar{u}^1)^2 \\ &= \frac{n}{n_0}\frac{1}{n}\sum_{i=1}^{n}(1-W_i)(u_i^1)^2 - (\bar{u}^1)^2 \\ &= \frac{n}{n_0}\frac{1}{n}\sum_{i=1}^{n}(1-W_i)(u_i^1)^2 - \left(\frac{n}{n_0}\frac{1}{n}\sum_{i=1}^{n}(1-W_i)u_i^1\right)^2. \end{aligned}$$

考虑随机变量 $\{(1-W_i)(u_i^1)^2\}_{i=1}^n$ 和 $\{(1-W_i)u_i^1\}_{i=1}^n$。由于各簇互不相交，且每个簇中的个体数均有界于 M，因此协方差项之和至多为 $O(n)$，从而弱大数定律适用于这两个序列。因此，

$$\frac{1}{n}\sum_{i=1}^{n}(1-W_i)(u_i^1)^2 - \left[p(1-p)^2\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i-1} + p^2(1-p)\right] \xrightarrow{\mathbb{P}} 0.$$

类似地，

$$\frac{1}{n}\sum_{i=1}^{n}(1-W_i)u_i^1 \xrightarrow{\mathbb{P}} p(1-p).$$

注意到 $n/n_0 \xrightarrow{\mathbb{P}} 1/(1-p)$，我们得到

$$① - \left[p(1-p)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i-1}\right] \xrightarrow{\mathbb{P}} 0.$$

这里，② 可以类似地完成：

$$② - \left[p(1-p)\frac{1}{n}\sum_{i=1}^{n}(m_i-1) + p^2\frac{1}{n}\sum_{i=1}^{n}(m_i-1)^2 - p^2\left(\frac{1}{n}\sum_{i=1}^{n}(m_i-1)\right)\right] \xrightarrow{\mathbb{P}} 0.$$

对于 ③，我们有

$$\begin{aligned} ③ &= \frac{1}{n_0}\sum_{i:W_i=0}(u_i^1-\bar{u}^1)(u_i^2-\bar{u}^2) \\ &= \frac{1}{n_0}\sum_{i:W_i=0}u_i^1u_i^2 - \bar{u}^1\bar{u}^2. \end{aligned}$$

注意，我们已经证明了

$$\bar{u}^1 \xrightarrow{\mathbb{P}} p, \qquad \bar{u}^2 \xrightarrow{\mathbb{P}} p\frac{1}{n}\sum_{i=1}^{n}(m_i-1).$$

因此，$\bar{u}^1\bar{u}^2 \xrightarrow{\mathbb{P}} p^2\frac{1}{n}\sum_{i=1}^{n}(m_i-1)$。此外，

$$\frac{1}{n_0}\sum_{i:W_i=0}u_i^1u_i^2 = \frac{n}{n_0}\frac{1}{n}\sum_{i=1}^{n}(1-W_i)u_i^1u_i^2.$$

再次根据大数弱定律，

$$\frac{1}{n}\sum_{i=1}^{n}(1-W_i)u_i^1u_i^2 - \left[p(1-p)^2 + p^2(1-p)\frac{1}{n}\sum_{i=1}^{n}(m_i-1)\right] \xrightarrow{\mathbb{P}} 0.$$

因此，$3 - \left[p(1-p) + p^2 \frac{1}{n}\sum_{i=1}^{n}(m_i - 1)\right] \xrightarrow{\mathbb{P}} 0$。将这些部分合并起来，我们得到

$$\begin{aligned}
\text{RHS of (12)} - \Bigg\{ & \left[p(1-p)\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i - 1}\right]\Delta_1^2 \\
& + \left[p(1-p)\frac{1}{n}\sum_{i=1}^{n}(m_i - 1) + p^2\frac{1}{n}\sum_{i=1}^{n}(m_i - 1)^2 - p^2\left(\frac{1}{n}\sum_{i=1}^{n}(m_i - 1)\right)\right. \\
& \left. - \frac{1}{3}\left(p(1-p) + p^2\frac{1}{n}\sum_{i=1}^{n}(m_i - 1)\right)\right]\Delta_2^2 \Bigg\} \xrightarrow{\mathbb{P}} 0.
\end{aligned}$$

注意到，对于每个 i， $m_i \geq 3$和 $m_i \leq M$，我们得出结论：对于 $\kappa = \min\{\frac{p(1-p)}{M-1}, \frac{5}{3}p - \frac{1}{3}p^2\}$，

$$\frac{1}{n_0}\|\tilde{U}^0\Delta\|_2^2 \geq \kappa\|\Delta\|_2^2 \quad \text{w.p.} \quad \to 1.$$

□

B 补充材料

定义 B.1（模拟中的非线性模型）。假设分配向量为 w，则对于每个个体 i，响应为

$$y_i(w) = -5 + 2z_i w_i + 0.03\nu_i + \frac{1}{1 + 0.001\exp(-0.03\nu_i + 9)} + \frac{10}{3 + \exp(-8\rho_i + 3.2)} + \epsilon_i.$$

其中， $z_i, \epsilon_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, 1)$， ρ_i是个体 i的处理组邻居比例， ν_i是个体 i的处理组邻居数量。