arXiv:2601.07267v1 [stat.ME] 12 Jan 2026

连接作为处理：网络中边干预的因果推断

Shuli Chen* Jie Hu† Zhichao Jiang‡

2026年1月13日

摘要

因果推断传统上侧重于个体层级的干预。然而，在许多应用中，核心问题涉及个体之间连接的因果效应，例如交通连接、社会关系或协作联系。我们为网络中的边干预构建了一个因果框架，其中处理对应于边的存在或缺失。该框架在针对网络结构的随机干预下定义因果估计目标，并在边分配满足无混杂假设的前提下，提出一种逆概率加权估计量。我们使用指数随机图模型（一类广泛应用的网络模型）估计边概率。我们证明了所提出估计量的一致性和渐近正态性。最后，我们将该方法应用于中国的交通网络，以估计铁路连接对区域经济发展的因果影响。**关键词：** 基于设计的框架；干扰；逆概率加权；随机图模型；随机干预

*中山大学数学学院，中国广东省广州市 510275。电子邮箱：chen- shli27@mail2.sysu.edu.cn

†厦门大学数学科学学院，中国福建省厦门市

‡中山大学数学学院，中国广东省广州市，邮编 510275。电子邮箱：jiangzhch7@mail.sysu.edu.cn

1 引言

推动因果推断发展的一个关键因素，是人们不断努力使其能够适用于日益多样的处理形式。早期方法主要关注二元处理，因为这为发展和形式化因果推断的基础思想提供了一个简单的场景（例如，Rosenbaum 和 Rubin，1983；Imbens 和 Rubin，2015；Ding，2024）。后续发展将这一框架扩展到更复杂的处理类型，包括类别型、连续型、函数型以及点过程数据（例如，Imbens，2000；Imai 和 Van Dyk，2004；Kennedy 等人，2017；Li 和 Li，2019；Zhang 等人，2021；Papadogeorgou 等人，2022；Jiang 等人，2023；Tan 等人，2025）。尽管取得了这些进展，大多数因果推断研究仍然聚焦于分配到个体层级的处理。然而，在许多实际情境中，干预的主要对象并非个体本身，而是个体之间的关系。例如，Sun 和 Taylor（2020）考察了突出共同属性如何影响在线社交网络中的友谊形成；Gonzalez-Navarro 和 Quintana-Domeque（2016）则研究了铺设街道对家庭消费的影响。这些研究考察了针对连接的干预，但没有提供用于定义或识别边层级效应的正式因果框架。

我们为无向网络中的边干预构建了一个基于设计的因果推断框架，其中处理由个体之间边的存在与否定义。所有个体对的处理共同构成一个处理矩阵，即网络的邻接矩阵。我们将潜在结果定义为网络邻接矩阵的函数，因此某个个体的结果不仅可能取决于与其直接相连的边，还可能取决于网络中的其他边。这一表述推广了干扰文献中使用的标准记法；该记法通常允许某个个体的结果取决于其他个体的处理分配（例如，Manski, 1993；Sobel, 2006；Hudgens and Halloran, 2008；Aronow and Samii, 2017；Sävje, 2024）。为使推断可行，我们引入局部干扰假设，将某个个体的潜在结果对邻域内边的依赖限制在一定范围内。这将网络环境中先前提出的局部干扰假设扩展到边级干预情境（Toulis and Kao, 2013；Sussman and Airoldi, 2017；Karwa and Airoldi, 2018；Jagadeesan et al., 2020；Awan et al., 2020；Forastiere et al., 2021, 2022；Leung, 2022；Belloni et al., 2025；Gao and Ding, 2025；Owusu, 2025；Wang et al., 2025）。

边干预在定义和识别因果效应方面带来了独特挑战。首先，每个个体邻域内的边集可能各不相同，因此难以在单一且固定的处理条件下定义平均潜在结局。其次，通常采用的重叠

假设常因可能的处理取值数量众多而遭到违反。为应对这些挑战，我们在随机干预下定义因果估计目标；这种方法能够通过低维参数来指定高维处理分配机制（例如，Kennedy，2019年；Papadogeorgou 等人，2019年；Barkley 等人，2020年；Díaz 和 Hejazi，2020年；Papadogeorgou 等人，2022年；Díaz 等人，2023；Wu 等人，2024年；Zhou 等人，2024年；Lee 等人，2025年）。

我们在针对边干预的广义可忽略性假设下，确立了所提出因果估计目标的可识别性，并开发了相应的逆概率加权（IPW）估计量。为实施逆概率加权估计量，我们通过对带有样本空间约束的指数随机图模型（ERGM）进行修改来估计边概率。这些约束禁止相距较远的单位之间存在边，从而反映了在我们的应用中建立直接连接在实践中并不可行。我们还给出了估计量满足一致性和渐近正态性的条件。这些条件在局部干扰假设下，对邻域大小以及网络中边之间的依赖结构施加限制。基于这些结果，我们提出了有效的推断程序。

我们将所提出的框架应用于评估城市间铁路交通连接如何影响中国当地的经济结果。在这一应用中，处理被分配给城市对，而非单个城市。这自然形成一种网络表示，其中每条边都编码一种处理状态。现有关于交通基础设施影响的研究大多通过将处理定义在单位层级来简化这一结构，通常以二元指标表示某城市是否连接到主要交通网络（例如，Faber，2014；Banerjee 等，2020年）。这种简化掩盖了处理的关系属性，也限制了研究者捕捉连通性变化如何影响直接和间接相连城市的能力。我们的分析则将城市间连接视为边层面处理，展示了如何运用所提出的方法研究连接概率随机增加或减少时经济结果的变化。

本文其余部分安排如下。第 2 节介绍边干预的因果框架。第 3 节推导识别结果，提出逆概率加权估计量，描述用于估计边概率的修正指数随机图模型，并确立渐近性质。第 4 节报告模拟研究，第 5 节应用所提出的方法分析中国城市间铁路网络的数据。第 6 节总结全文。

2 边干预的因果框架

2.1 设定

我们考虑一个有限总体情境，其中所有潜在结果和协变量均为条件变量，并将处理分配视为唯一的随机性来源。这种基于设计的框架常用于因果推断研究（例如，Imbens and Rubin, 2015; Aronow and Samii, 2017; Abadie et al., 2020; Leung, 2022; Sävje, 2024）。我们关注无权无向网络中的边干预，该网络包含 n 个个体，索引为 $\mathcal{N} = \{1,\ldots,n\}$。处理由网络的邻接矩阵 $\boldsymbol{A} = (A_{ij})_{i,j=1}^{n}$ 表示，其中每个条目 A_{ij} 都是一个二元指标，用于表示个体 i 与 j之间是否存在连接。排除自连接，因此对于所有 $i \in \mathcal{N}$， $A_{ii} = 0$ 。所有此类二元邻接矩阵的集合记为 $\mathcal{A}$。每个个体 $i \in \mathcal{N}$都有一个观测结果 Y_i 和一个协变量向量 X_i。协变量向量 X_i 可以包含单位层面特征，以及与其和其他个体之间连接有关的特征。所有协变量的集合记为 $\boldsymbol{X} = (X_1, \ldots, X_n)$。

我们将个体 i 的潜在结果定义为 $Y_i(\boldsymbol{a})$，表示当网络结构设为 $\boldsymbol{a} \in \mathcal{A}$ 时，个体 i 将观察到的结果。此记号意味着， $Y_i(\boldsymbol{a})$ 不仅取决于与个体 i 直接相连的边，还取决于更广泛的网络结构，包括其他个体之间的边。因此，每个个体都有 $2^{n(n-1)/2}$ 个潜在结果，若无额外假设，因果推断将无法进行。

我们对潜在结果依赖网络结构的方式施加限制。具体而言，我们假设个体 i 的潜在结果仅取决于由一部分个体构成的子网络中的连接。我们将这部分个体称为排除邻域，记为 $\mathcal{N}_i \subset \mathcal{N}$，其中包括个体 i 本身。该邻域定义为其连接可能影响 Y_i 的最小个体集合。因此，$\mathcal{N}_i$ 之外的网络变化不会影响 Y_i。令 $\boldsymbol{A}_i = (A_{kl} : k, l \in \mathcal{N}_i)$ 表示 $\mathcal{N}_i$ 上子网络的邻接矩阵，并令 $\mathcal{A}_i$ 为所有此类二元邻接矩阵的集合。我们在以下假设中形式化这一限制。

假设***1*** （局部干扰）。对于每个 $i \in \mathcal{N}$， $Y_i(\boldsymbol{a}) = Y_i(\boldsymbol{a}')$ 如果 $\boldsymbol{a}_i = \boldsymbol{a}'_i$ 对于任意 $\boldsymbol{a}$， $\boldsymbol{a}' \in \mathcal{A}$。假设1意味着，个体 i 的结果仅取决于其排除邻域$\mathcal{N}_i$内的网络结构。在此假设下，我们可以将 $Y_i(\boldsymbol{a})$简化为 $Y_i(\boldsymbol{a}_i)$。排除邻域$\mathcal{N}_i$ 是预先指定的，不依赖于实际实现的网络。其定义为

基于外部信息或领域知识，确定哪些个体可能会对个体i产生影响。例如，在交通网络中，$\mathcal{N}_i$ 可能由一定地理半径范围内的城市构成。在社交网络中，它可能包括人口统计特征、兴趣或组织归属相似的个人。

限制干扰结构的另一种方法是采用暴露映射，它假设潜在结果仅通过处理取值的某个指定函数依赖于处理。这种方法便于对因果效应进行正式定义和识别，因此在涉及单位层面处理的干扰文献中得到了广泛应用。许多相关研究假设暴露映射是预先已知的（例如，Hudgens and Halloran, 2008；Manski, 2013；Aronow and Samii, 2017；Basse and Feller, 2018；Forastiere et al., 2022；Bargagli-Stoffi et al., 2025），而较新的研究则提出了适用于映射设定错误或未知情形的方法（Sävje et al., 2021；Leung, 2022；Sävje, 2024；Viviano and Rudder, 2024；Gao and Ding, 2025）。对于边干预，可以将暴露映射的思路推广为：假设潜在结果仅通过该子网络的某些汇总统计量依赖于 $\boldsymbol{a}_i$。然而，这一假设严格限制了潜在结果对该子网络结构的依赖方式，而且通常无法通过观测数据验证这些限制。因此，我们不采用这一思路，而仅依赖假设1。

没有暴露映射时，会出现一个关键难题：排除邻域的大小 $|\mathcal{N}_i|$ 可能因个体而异。因此，局部处理 $\boldsymbol{a}_i$ 的支持集因个体而异。这种异质性使得难以定义一个适用于所有个体、且基于单一固定处理条件的平均潜在结果。此外， $\boldsymbol{a}_i$可能取值众多，往往导致某些个体对某些取值的概率为零，从而违反重叠假设。在下一小节中，我们将使用随机干预来定义因果估计量。

2.2 随机干预下的因果估计量

令 δ 表示一种处理策略，它定义为针对每个个体 i，其所有可能局部处理值所构成集合上的一个概率分布。在该策略下，个体 i 接受的处理是一个随机变量 $\boldsymbol{A}_i^{\delta}$，从 δ指定的分布中抽取。量 $\mathbb{P}(\boldsymbol{A}_i^{\delta} = \boldsymbol{a}_i)$表示在 δ下，个体 i 接受特定局部处理值 $\boldsymbol{a}_i$ 的概率。我们通过以下对数优势比对该概率进行参数化：

$$\log\left\{\frac{\mathbb{P}(\boldsymbol{A}_i^{\delta}=\boldsymbol{a}_i)/\mathbb{P}(\boldsymbol{A}_i^{\delta}=\boldsymbol{0}_i)}{\mathbb{P}(\boldsymbol{A}_i=\boldsymbol{a}_i)/\mathbb{P}(\boldsymbol{A}_i=\boldsymbol{0}_i)}\right\} \quad = \quad e(\boldsymbol{a}_i)\lambda_{\delta} \tag{1}$$

对于任意 $\boldsymbol{a}_i \in \mathcal{A}_i$，其中 $\mathbf{0}_i$ 表示所有分量均为零的局部处理值，$e(\boldsymbol{a}_i) = \sum_{k,\ j\in\mathcal{N}_i,\ k<j} a_{kj}$ 表示 $\boldsymbol{a}_i$ 中的边数，而 λ_δ 是一个预先指定的标量参数，用于刻画策略 δ。方程（1）规定了处理策略 δ 下 $\boldsymbol{a}_i$ 的概率与 $\mathbf{0}_i$的概率之比，相对于观测分布下对应优势的优势比。结合归一化约束 $\sum_{\boldsymbol{a}_i\in\mathcal{A}_i} \mathbb{P}(\boldsymbol{A}_i^\delta = \boldsymbol{a}_i) = 1$，这一设定确定了 $\boldsymbol{A}_i^\delta$的完整概率分布。

参数 λ_δ 是个体 i排除邻域内边形成的全局缩放因子。考虑两个局部处理值 $\boldsymbol{a}_i$ 和 $\boldsymbol{a}_i'$，它们仅相差一个对称项，其中 $\boldsymbol{a}_i$ 包含一条边，而 $\boldsymbol{a}_i'$ 不包含该边。于是，公式 (1) 表明:

$$\log\left\{\frac{\mathbb{P}(\boldsymbol{A}_i^\delta = \boldsymbol{a}_i)/\mathbb{P}(\boldsymbol{A}_i^\delta = \boldsymbol{a}_i')}{\mathbb{P}(\boldsymbol{A}_i = \boldsymbol{a}_i)/\mathbb{P}(\boldsymbol{A}_i = \boldsymbol{a}_i')}\right\} \;=\; \{e(\boldsymbol{a}_i) - e(\boldsymbol{a}_i')\}\lambda_\delta \;=\; \lambda_\delta.$$

因此，λ_δ 衡量的是在策略 δ 下，相较于观测分布，添加单条边所导致的对数优势变化。因此，正的 λ_δ 会提高个体 i局部处理下边形成的概率，而负的 λ_δ 则会降低该概率。当 $\lambda_\delta = 0$时，处理策略δ 与观测到的边分布一致。

公式 (1) 中的设定可以推广为允许 λ_δ 因个体而异，从而能够对特定个体（例如在经济或战略上重要的城市）予以不同程度的重视。然而，这种单位特定参数化要求为每个个体分别指定一个参数，这通常不切实际，也难以论证其合理性。为便于处理，我们聚焦于一个由各个体共享的标量 λ_δ。

根据方程（1），如果 $\mathbb{P}(\boldsymbol{A}_i = \boldsymbol{a}_i) = 0$ 对于某个 $\boldsymbol{a}_i$成立，那么 $\mathbb{P}(\boldsymbol{A}_i^\delta = \boldsymbol{a}_i) = 0$ 也成立。换言之，处理策略 δ 不会为观测分布下不可能出现的任何局部处理值赋予正概率。这保证了在我们对处理策略的表述下，以下重叠条件成立。

假设*2*（重叠）。存在一个常数 $\alpha > 0$，使得 $\mathbb{P}(\boldsymbol{A}_i = \boldsymbol{a}_i) > \alpha\mathbb{P}(\boldsymbol{A}_i^\delta = \boldsymbol{a}_i)$，对所有 $i \in \boldsymbol{N}$ 和 $\boldsymbol{a}_i \in \boldsymbol{A}_i$ 均成立。

假设2弱于标准重叠假设，因为它允许某些处理取值的概率为零。它要求处理策略 δ 仅对观测分布中概率非零的局部处理值赋予正概率。这一限制缩小了 δ 的选择范围，但避免了设定在观测网络中不可能出现的反事实处理值。随机干预文献中也常采用类似的重叠假设（例如，

Papadogeorgou 等，2022年；Díaz 等，2023；Wu 等，2024年；Zhou 等，2024年）。

现在我们定义感兴趣的因果估计目标。对于每个个体 i，在处理策略 δ 下的平均潜在结果为

$$\theta_i^\delta \quad = \quad \mathbb{E}\left\{Y_i(\boldsymbol{A}_i^\delta)\right\} \quad = \quad \sum_{\boldsymbol{a}_i \in \mathcal{A}_i} Y_i(\boldsymbol{a}_i)\mathbb{P}(\boldsymbol{A}_i^\delta = \boldsymbol{a}_i),$$

它表示个体 i 在按照策略 δ分配局部处理时的预期结果。对所有个体取平均，得到总体平均潜在结果

$$\theta^\delta \quad = \quad \frac{1}{n}\sum_{i=1}^{n}\theta_i^\delta.$$

作为一个特例，我们将策略满足$\lambda_\delta = 0$时对应的平均潜在结果记为 θ^0；在该策略下，$\boldsymbol{A}_i^\delta$ 与观测到的处理 $\boldsymbol{A}_i$具有相同的分布。量 θ^δ可通过比较平均潜在结果来比较不同处理策略。例如，将 θ^δ 与 θ^0 进行比较，可以评估将局部网络连接的分布从观测分布转变为策略 δ所指定分布的因果影响。

3 识别与估计

3.1 识别

我们扩展 t 将标准无混杂假设推广至边干预情形 s.

假设***3*** （无混杂）。 $\mathbb{P}(\boldsymbol{A} \mid \boldsymbol{X}, \{Y_i(\boldsymbol{a}) : \boldsymbol{a} \in \boldsymbol{A}\ i \in \boldsymbol{N}\}) = \mathbb{P}(\boldsymbol{A} \mid \boldsymbol{X})$。

假设3指出，给定协变量 $\boldsymbol{X}$后，邻接矩阵$\boldsymbol{A}$ 的分布不依赖于潜在结果。这要求 $\boldsymbol{X}$ 包含同时影响网络结构和结果的所有协变量。如果从 $\boldsymbol{X}$中遗漏了相关混杂因素，则该假设将不成立，从而导致估计产生偏差。在这种情况下，需要进行敏感性分析，以评估研究结果的稳健性。

以下定理确立了 θ^δ的非参数识别。

定理***1*** 在假设*1*、*2*和*3*下，θ^δ 可由以下内容识别

$$\theta^\delta \quad = \quad \mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{P}(\boldsymbol{A}_i^\delta)}{\mathbb{P}(\boldsymbol{A}_i)}Y_i\right\}.$$

定理1表明，θ^δ 等于观测结果的重新加权平均的期望，其中权重 $\mathbb{P}(\boldsymbol{A}_i^\delta)/\mathbb{P}(\boldsymbol{A}_i)$ 表示在策略 $\boldsymbol{A}_i$下观测到局部处理的相对似然

相较于观测分布。假设2保证，对于所有满足 $\mathbb{P}(\boldsymbol{A}_i = \boldsymbol{a}_i) > 0$ 的 $\boldsymbol{a}_i$，这些权重都有明确定义。

定理1启发我们得到以下针对 θ^δ的无偏估计量：

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{P}(\boldsymbol{A}_i^\delta)}{\mathbb{P}(\boldsymbol{A}_i)}Y_i.$$

然而，这个估计量无法直接计算，因为 $\mathbb{P}(\boldsymbol{A}_i)$和 $\mathbb{P}(\boldsymbol{A}_i^\delta)$均未知。因此，我们先估计这些概率，再将其代入（2），得到以下 Horvitz-Thompson 逆概率加权（IPW）估计量：

$$\hat{\theta}_1^\delta \quad = \quad \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\mathbb{P}}(\boldsymbol{A}_i^\delta)}{\hat{\mathbb{P}}(\boldsymbol{A}_i)}Y_i.$$

在有限样本中，估计权重$\hat{}\mathbb{P}(\boldsymbol{A}_i^\delta)/\hat{}\mathbb{P}(\boldsymbol{A}_i)$可能会表现出较大的变异性，尤其是在$\hat{}\mathbb{P}(\boldsymbol{A}_i)$接近零时。为缓解这一问题，我们可以使用以下Hájek IPW估计量：

$$\hat{\theta}_2^\delta \quad = \quad \sum_{i=1}^{n}\frac{\hat{\mathbb{P}}(\boldsymbol{A}_i^\delta)}{\hat{\mathbb{P}}(\boldsymbol{A}_i)}Y_i \Bigg/ \sum_{i=1}^{n}\frac{\hat{\mathbb{P}}(\boldsymbol{A}_i^\delta)}{\hat{\mathbb{P}}(\boldsymbol{A}_i)}.$$

该估计量会对$\hat{}\theta_1^\delta$中的权重进行归一化，从而减小极端值的影响，并提高有限样本稳定性。

与传统的单位层面处理情境相比，在边干预的背景下估计这些处理概率更具挑战性。我们将在下一小节介绍我们的估计方法。

3.2 基于网络建模的估计

我们使用指数随机图模型（ERGM）估计处理概率 $\mathbb{P}(\boldsymbol{A}_i)$ 和 $\mathbb{P}(\boldsymbol{A}_i^\delta)$。指数随机图模型（ERGM）是广泛用于网络数据的一类统计模型（例如 Holland and Leinhardt, 1981；Frank and Strauss, 1986）。在标准形式下，指数随机图模型（ERGM）为邻接矩阵 $\boldsymbol{A}$ 指定一个指数族分布：

$$\mathbb{P}_\eta(\boldsymbol{A} = \boldsymbol{a}) \quad = \quad \frac{1}{\sum_{\boldsymbol{a}' \in \mathcal{A}} \exp\{\eta^\top g(\boldsymbol{a}', \boldsymbol{x})\}} \exp\{\eta^T g(\boldsymbol{a}, \boldsymbol{x})\} \tag{2}$$

对于所有 $\boldsymbol{a} \in \mathcal{A}$，其中 $g(\boldsymbol{a}, \boldsymbol{x}) = (g_1(\boldsymbol{a}, \boldsymbol{x}) \ldots, g_d(\boldsymbol{a}, \boldsymbol{x}))^\top$ 是网络统计量向量，$\eta \in \mathbb{R}^d$是相应的参数向量。网络统计量 $g(\boldsymbol{a}, \boldsymbol{x})$ 捕捉网络的结构特征和依赖协变量的特征，例如边的总数、节点度，以及三角形或 k-星等高阶模式。给定邻接矩阵 $\boldsymbol{a}$，令 $\boldsymbol{a}^{ij+}$ 和 $\boldsymbol{a}^{ij-}$ 分别表示将 (i, j) 项设为 1 和 0、同时保持所有其他项不变所得的邻接矩阵

。指数随机图模型（2）蕴含了以下关于单位 i 和单位 j之间一条边的条件对数优势表达式：

$$\log\left\{\frac{\mathbb{P}\left(A_{ij}=1 \mid A_{uv}=a_{uv}\forall\ (u,v)\neq(i,j)\right)}{1-\mathbb{P}\left(A_{ij}=1 \mid A_{uv}=a_{uv}\forall\ (u,v)\neq(i,j)\right)}\right\} \quad = \quad \eta^T\{g(\boldsymbol{a}^{ij+},\boldsymbol{x})-g(\boldsymbol{a}^{ij-},\boldsymbol{x})\}. \tag{3}$$

该表达式表明，形成一条边的对数优势取决于切换该边时网络统计量的变化。因此，网络统计量的选择直接决定了网络中边之间的依赖类型，而参数向量 η 则量化了这些依赖关系的方向和强度。

指数随机图模型的一个常见特点是，对每个可能的二元邻接矩阵都赋予严格大于零的概率，这意味着任意一对个体之间都有非零概率形成连接。在我们的应用中，这一性质使标准指数随机图模型不适合用于道路网络建模，因为某些城市对相距过远，无法在两小时出行时间范围内实现可行的直接连接。

我们对标准指数随机图模型进行调整，将 $\boldsymbol{A}$ 的样本空间限制为这样的邻接矩阵：仅允许指定邻域内的个体之间存在边。对于每个个体 i，令$\mathcal{U}_i$ 表示有资格与个体 i形成边的个体集合。根据定义，若个体 $j\in\mathcal{U}_i$，则 $i\in\mathcal{U}_j$。集合$\{\mathcal{U}_i, i\in\mathcal{N}\}$ 定义了受限样本空间$\mathcal{A}_{\mathcal{U}}=\{(a_{ij}):\ j\in\mathcal{U}_i\ \forall(i,j)\}$，其中包含所有符合以下要求的二元邻接矩阵：边只能在 i 与$\mathcal{U}_i$ 中的个体之间形成，适用于所有 i。由此定义受约束指数随机图模型：

$$\mathbb{P}_\eta(\boldsymbol{A}=\boldsymbol{a}) \quad = \quad \frac{\mathbf{1}(\boldsymbol{a}\in\mathcal{A}_{\mathcal{U}})\exp\{\eta^T g(\boldsymbol{a},\boldsymbol{x})\}}{\sum_{\boldsymbol{a}'\in\mathcal{A}}\mathbf{1}(\boldsymbol{a}'\in\mathcal{A}_{\mathcal{U}})\exp\{\eta^T g(\boldsymbol{a}',\boldsymbol{x})\}}, \tag{4}$$

其中，$\mathbf{1}(\cdot)$为示性函数。该表述保留了$\mathcal{A}_{\mathcal{U}}$ 中网络的指数族结构，同时为所有其他网络赋予零概率。邻域$\mathcal{U}_i$ 应根据实质性知识或关于可行连接的外部约束事先指定。在实践中，这类限制可能由交通网络中的地理距离，或社交网络中的制度或技术限制所决定。我们强调，$\mathcal{U}_i$ 表示一个预先指定的可行性邻域，用于限制网络支撑集；它不同于排除邻域$\mathcal{N}_i$，后者规定个体 i 的结果可能如何依赖于网络中的边。

我们使用马尔可夫链蒙特卡洛最大似然估计方法估计参数 η 。该方法最初由Geyer和Thompson（1992）针对相关数据的复杂指数族模型提出，随后被应用于指数随机图模型的估计（例如，Snijders等人，2002年；Hunter等人，2008年）。对于我们的受约束指数随机图模型，我们修改了

Metropolis–Hastings 步骤中的提议分布，使所有提议的网络都满足约束，从而确保该链始终完全处于受限空间内。

接下来，我们计算基于受约束指数随机图模型的估计量。令 $\boldsymbol{a}_{-i}$ 表示从 $\boldsymbol{a}$ 中移除排除邻域 $\mathcal{N}_i$ 内所有边后得到的子邻接矩阵，并令 $\mathcal{A}_{-i}$ 为 $\boldsymbol{a}_{-i}$ 的所有可能取值所构成的集合。于是，完整邻接矩阵可表示为 $\boldsymbol{a} = (\boldsymbol{a}_i, \boldsymbol{a}_{-i})$。根据式 (4) 中的受约束指数随机图模型，局部处理的观测分布为

$$\mathbb{P}(\boldsymbol{A}_i = \boldsymbol{a}_i) \;=\; \sum_{\boldsymbol{a}_{-i}\in\mathcal{A}_{-i}} \mathbb{P}\left(\boldsymbol{A} = (\boldsymbol{a}_i, \boldsymbol{a}_{-i})\right) \;=\; \frac{\sum_{\boldsymbol{a}_{-i}\in\mathcal{A}_{-i}} \mathbf{1}((\boldsymbol{a}_i, \boldsymbol{a}_{-i}) \in \mathcal{A}_{\mathcal{U}}) \exp\{\eta^T g(\boldsymbol{a}_i, \boldsymbol{a}_{-i}, \boldsymbol{x})\}}{\sum_{\boldsymbol{a}'\in\mathcal{A}} \mathbf{1}(\boldsymbol{a}' \in \mathcal{A}_{\mathcal{U}}) \exp\{\eta^T g(\boldsymbol{a}', \boldsymbol{x})\}},$$

它对 $\mathcal{A}_{-i}$ 中 $\boldsymbol{a}_{-i}$ 的所有可能取值进行了边缘化。根据式 (1) 中定义的随机干预，处理策略 δ 下局部处理的概率为

$$\mathbb{P}(\boldsymbol{A}_i^{\delta} = \boldsymbol{a}_i) \;=\; \frac{\sum_{\boldsymbol{a}_{-i}\in\mathcal{A}_{-i}} \mathbf{1}((\boldsymbol{a}_i, \boldsymbol{a}_{-i}) \in \mathcal{A}_{\mathcal{U}}) \exp\{\eta^T g(\boldsymbol{a}_i, \boldsymbol{a}_{-i}, \boldsymbol{x}) + e(\boldsymbol{a}_i)\lambda_\delta\}}{\sum_{\boldsymbol{a}'\in\mathcal{A}} \mathbf{1}(\boldsymbol{a}' \in \mathcal{A}_{\mathcal{U}}) \exp\{\eta^T g(\boldsymbol{a}', \boldsymbol{x}) + e(\boldsymbol{a}'_i)\lambda_\delta\}}.$$

将指数随机图模型估计值 $\hat{}\eta$ 代入 $\mathbb{P}(\boldsymbol{A}_i = \boldsymbol{a}_i)$ 和 $\mathbb{P}(\boldsymbol{A}_i^{\delta} = \boldsymbol{a}_i)$ 的表达式，可得

$$\begin{aligned}
\mathbb{P}_{\hat{\eta}}(\boldsymbol{A}_i = \boldsymbol{a}_i) \;&=\; \frac{\sum_{\boldsymbol{a}_{-i}\in\mathcal{A}_{-i}} \mathbf{1}((\boldsymbol{a}_i, \boldsymbol{a}_{-i}) \in \mathcal{A}_{\mathcal{U}}) \exp\{\hat{\eta}^T g(\boldsymbol{a}_i, \boldsymbol{a}_{-i}, \boldsymbol{x})\}}{\sum_{\boldsymbol{a}'\in\mathcal{A}} \mathbf{1}(\boldsymbol{a}' \in \mathcal{A}_{\mathcal{U}}) \exp\{\hat{\eta}^T g(\boldsymbol{a}', \boldsymbol{x})\}}, \\
\mathbb{P}_{\hat{\eta}}(\boldsymbol{A}_i^{\delta} = \boldsymbol{a}_i) \;&=\; \frac{\sum_{\boldsymbol{a}_{-i}\in\mathcal{A}_{-i}} \mathbf{1}((\boldsymbol{a}_i, \boldsymbol{a}_{-i}) \in \mathcal{A}_{\mathcal{U}}) \exp\{\hat{\eta}^T g(\boldsymbol{a}_i, \boldsymbol{a}_{-i}, \boldsymbol{x}) + e(\boldsymbol{a}_i)\lambda_\delta\}}{\sum_{\boldsymbol{a}'\in\mathcal{A}} \mathbf{1}(\boldsymbol{a}' \in \mathcal{A}_{\mathcal{U}}) \exp\{\hat{\eta}^T g(\boldsymbol{a}', \boldsymbol{x}) + e(\boldsymbol{a}'_i)\lambda_\delta\}},
\end{aligned}$$

我们使用下标 $\hat{}\eta$ 来强调，这些概率是在估计得到的指数随机图模型参数下计算的。这些概率给出了以下逆概率加权估计量：

$$\hat{\theta}_1^{\delta} \;=\; \frac{1}{n}\sum_{i=1}^{n} \frac{\exp\{e(\boldsymbol{A}_i)\lambda_\delta\}}{\sum_{\boldsymbol{a}'_i\in\mathcal{A}_i} \exp\{e(\boldsymbol{a}'_i)\lambda_\delta\}\mathbb{P}_{\hat{\eta}}(\boldsymbol{A}_i = \boldsymbol{a}'_i)} Y_i, \tag{5}$$

$$\hat{\theta}_2^{\delta} \;=\; \sum_{i=1}^{n} \frac{\exp\{e(\boldsymbol{A}_i)\lambda_\delta\} Y_i}{\sum_{\boldsymbol{a}'_i\in\mathcal{A}_i} \exp\{e(\boldsymbol{a}'_i)\lambda_\delta\}\mathbb{P}_{\hat{\eta}}(\boldsymbol{A}_i = \boldsymbol{a}'_i)} \Bigg/ \sum_{i=1}^{n} \frac{\exp\{e(\boldsymbol{A}_i)\lambda_\delta\}}{\sum_{\boldsymbol{a}'_i\in\mathcal{A}_i} \exp\{e(\boldsymbol{a}'_i)\lambda_\delta\}\mathbb{P}_{\hat{\eta}}(\boldsymbol{A}_i = \boldsymbol{a}'_i)}. \tag{6}$$

若将估计得到的指数随机图模型参数 $\hat{}\eta$ 视为固定值，则式 (5) 和 (6) 中权重的分母可写为

$$\sum_{\boldsymbol{a}'_i} \exp\{e(\boldsymbol{a}'_i)\lambda_\delta\}\mathbb{P}_{\hat{\eta}}(\boldsymbol{A}_i = \boldsymbol{a}'_i) \;=\; \mathbb{E}_{\hat{\eta}}\left\{\exp\{e(\boldsymbol{A}_i)\lambda_\delta\}\right\}, \tag{7}$$

其中，$\mathbb{E}_{\hat{\eta}}$ 表示相对于估计分布 $\mathbb{P}_{\hat{\eta}}(\boldsymbol{A}_i)$计算的期望。实际操作中，我们首先在受约束指数随机图模型中估计$\hat{}\eta$，然后从拟合模型中进行蒙特卡洛模拟，以近似计算式（7）中的期望。

3.3 渐近性质

现在，我们考察所提出估计量的渐近性质。由于我们的逆概率加权估计量依赖于估计得到的指数随机图模型参数$\hat{\eta}$，其渐近行为取决于指数随机图模型估计量的性质。然而，由于边之间存在复杂的依赖结构，要建立指数随机图模型的渐近理论颇具挑战性。尽管在假设边相互独立的条件下已有一些研究结果，但指数随机图模型估计量的一般渐近行为仍未得到充分理解。有关这些挑战的进一步讨论，参见 Schweinberger and Stewart（2020）和 Schweinberger 等人（2020）。

为使渐近分析易于处理，在研究估计量的渐近行为时，我们将估计参数 $\hat{\eta}$ 视为固定值。这实际上忽略了指数随机图模型中的估计不确定性，相当于将数据视为来自一项条件随机实验，其中处理由参数为 $\hat{\eta}$ 的受约束指数随机图模型生成。在匹配和协变量平衡文献中，以及在难以量化估计处理概率的不确定性的研究中（例如，Hainmueller，2012；Shook-Sa and Hudgens，2022；Papadogeorgou 等人，2025年），类似的简化很常见。我们将通过模拟研究评估这一简化的影响。

渐近分析聚焦于 $\hat{\theta}_1^{\delta}$；关于 $\hat{\theta}_2^{\delta}$ 的结果类似。根据（5）和（7），我们可以写出

$$\hat{\theta}_1^{\delta} \quad = \quad \frac{1}{n}\sum_{i=1}^{n} \frac{\exp\{e(\boldsymbol{A}_i)\lambda_{\delta}\}}{\mathbb{E}_{\hat{\eta}}\left\{\exp\{e(\boldsymbol{A}_i)\lambda_{\delta}\}\right\}} Y_i(\boldsymbol{A}_i). \tag{8}$$

式（8）中的加项并不独立，因为对于 $i \neq j$， $\boldsymbol{A}_i$ 和 $\boldsymbol{A}_j$ 可能存在相关性。如果排除邻域$\mathcal{N}_i$ 和 $\mathcal{N}j$ 相交，那么 $\boldsymbol{A}_i$ 和 $\boldsymbol{A}_j$ 会共享一些边，从而在二者之间产生相关性。网络因果推断文献已充分认识到这一依赖来源，通常通过限制排除邻域的大小来加以处理。然而，对于边干预而言，此类限制通常并不足够，因为边本身可能存在相关性。这与先前假设不同：先前研究假设不同单位之间的处理相互独立（例如 Leung, 2022；Gao and Ding, 2025）。因此，在我们的设定中， $\boldsymbol{A}_i$之间的依赖结构更加复杂。为处理这一额外复杂性，我们对边之间的依赖关系作出以下假设。

假设 ***4*** （局部依赖）对于每个个体 $i \in \boldsymbol{N}$，都存在一个子集 $\boldsymbol{R}_i \subset \boldsymbol{N}$，使得$\boldsymbol{A}_i \perp\!\!\!\perp \boldsymbol{A}_j$，对于所有 $j \notin \boldsymbol{R}_i$。

假设 4 表明，局部处理 $\boldsymbol{A}_i$ 与集合$\mathcal{R}_i$之外的其他局部处理相互独立。区分集合$\mathcal{N}_i$ 和$\mathcal{R}_i$非常重要。集合$\mathcal{N}_i$ 决定哪些边会直接影响个体 i的潜在结果，而$\mathcal{R}_i$ 则通过指定其局部处理可能影响 $\boldsymbol{A}_i$分布的个体集合，刻画网络的相关结构。

$\mathcal{R}_i$ 的结构既取决于$\mathcal{N}_i$ ，也取决于指数随机图模型所诱导的边依赖结构。在边相互独立的伯努利随机图模型下，只要$\mathcal{N}_i$ 与$\mathcal{N}j$ 共享的个体少于两个， $\boldsymbol{A}_i$与 $\boldsymbol{A}_j$ 就相互独立。在这种情况下，我们可以定义$\mathcal{R}_i = \{j : \ |\mathcal{N}_i \cap \mathcal{N}_j| \geq 2\}$。然而，在边之间存在全局依赖关系的更一般的指数随机图模型下（Frank and Strauss，1986；Snijders等人，2006），这样的集合$\mathcal{R}_i$ 可能并不存在。为确保存在有效的局部依赖结构，必须将边依赖限制在预先指定的子图上，如Schweinberger and Handcock（2015）提出的框架所示。在模拟和实证分析中，我们重点研究具有样本空间约束的伯努利随机图模型。附录中给出了具有局部依赖的更复杂指数随机图模型下的其他结果。

以下定理在$\mathcal{R}_i$满足某个条件时，证明了$\hat{}\theta_1^\delta$ 的一致性。

定理 *2* 假设假设 *1* 至 *4* 成立，且结果有界。若 $\sum_{i=1}^n |R_i| = o_{\mathbb{P}}(n^2)$，则 $\hat{}\theta_1^\delta$ 依概率收敛于 θ^δ 。

有界结果这一假设可以放宽为一个较弱的条件，只要求结果之间的协方差 Y_i 与 Y_j 一致有界。为简便起见，我们在此采用较强的条件，而该条件在我们的应用中成立。条件 $\sum_{i=1}^n |\mathcal{R}_i| = o_{\mathbb{P}}(n^2)$ 限制了人口中依赖关系的总体强度。它要求个体之间的依赖关系足够稀疏：尽管每个个体的局部处理 $\boldsymbol{A}_i$ 可能取决于 $\mathcal{R}_i$ 中其他个体的处理，但此类依赖关系的总数必须以低于 n^2 的速度增长。如果每个 $\mathcal{R}_i$ 都包含人口中固定比例的个体，从而导致边之间存在稠密依赖，那么当 $\sum_{i=1}^n |\mathcal{R}_i| = O_{\mathbb{P}}(n^2)$ 时，该条件便不成立。所述条件排除了此类依赖结构，而是要求随着n 增加，大多数个体仅与人口中趋于零的一小部分个体发生相互作用。

以下定理确立了$\hat{}\theta_1^\delta$的渐近正态性。

定理*3* 假设假设 *1* 至 *4* 成立、结果有界，且定理 *2* 中的条件

定理 *2* 成立。如果

$$\max\left[n^{-2}\left\{\mathrm{var}(\hat{\theta}_1^{\delta})\right\}^{-3/2}\max_{i\in\mathcal{N}}|\mathcal{R}_i|^2, n^{-3/2}\left\{\mathrm{var}(\hat{\theta}_1^{\delta})\right\}^{-1}\max_{i\in\mathcal{N}}|\mathcal{R}_i|^{3/2}\right]\to 0,$$

then

$$\frac{\hat{\theta}_1^{\delta}-\theta^{\delta}}{\sqrt{\mathrm{var}(\hat{\theta}_1^{\delta})}}\xrightarrow{d}\mathcal{N}(0,1).$$

定理3对 $\max_{i\in\mathcal{N}}|\mathcal{R}_i|$ 施加了速率条件，该量表示与给定个体的局部处理相关的其他个体的最大数量。与定理 2 中一致性所需的条件相比，该条件要求网络依赖结构具有更强的稀疏性。具体而言，定理 2 允许少数个体拥有较大的依赖邻域，而定理3则限制所有个体的最大依赖程度。这确保没有任何单一个体会对估计量的变异性产生不成比例的影响，这对于确立渐近正态性至关重要。此类条件在网络因果推断文献中很常见；在这类研究中，对于聚合相关观测的估计量，会确立其渐近正态性（Kojevnikov 等人，2021；Leung, 2022；Gao and Ding, 2025）。

接下来我们考虑方差估计。由于式（8）中的加项彼此相关，标准自助法无效。因此，我们采用 Kojevnikov（2021）提出的修正的依赖型野生自助法，该方法构造的自助扰动具有相关结构，能够反映原始网络数据中的依赖关系。

我们生成自助样本如下

$$Y_i^{*} \;=\; \hat{\theta}_1^{\delta}+\left[\frac{\exp\{e(\boldsymbol{A}_i)\lambda_{\delta}\}}{\mathbb{E}_{\hat{\eta}}\left\{\exp\{e(\boldsymbol{A}_i)\lambda_{\delta}\}\right\}}Y_i(\boldsymbol{A}_i)-\hat{\theta}_1^{\delta}\right]W_i$$

对所有 i，其中 (W_1、.....、W_n) 服从均值为零、协方差矩阵为 Ω 的多元正态分布。矩阵 Ω 刻画了局部处理之间的依赖关系：

$$\Omega \;=\; (\omega_{ij}),\quad \omega_{ij} \;=\; \frac{|\mathcal{R}_i\cap\mathcal{R}_j|}{n^{-1}\sum_{i=1}^{n}|\mathcal{R}_i|}.$$

对于每个自助样本，我们计算 $\theta_1^{*} \;=\; n^{-1}\sum_{i=1}^{n}Y_i^{*}$，并使用自助重复抽样的样本方差来估计方差。Kojevnikov（2021）表明，该自助法方差估计量可以直接根据原始样本计算如下：

$$\frac{1}{n^2}\sum_{i,j\in\mathcal{N}}\omega_{ij}\left[\frac{\exp\{e(\boldsymbol{A}_i)\lambda_{\delta}\}}{\mathbb{E}_{\hat{\eta}}\left\{\exp\{e(\boldsymbol{A}_i)\lambda_{\delta}\}\right\}}Y_i(\boldsymbol{A}_i)-\hat{\theta}_1^{\delta}\right]\left[\frac{\exp\{e(\boldsymbol{A}_j)\lambda_{\delta}\}}{\mathbb{E}_{\hat{\eta}}\left\{\exp\{e(\boldsymbol{A}_j)\lambda_{\delta}\}\right\}}Y_j(\boldsymbol{A}_j)-\hat{\theta}_1^{\delta}\right]. \tag{9}$$

因此，我们直接应用（9）进行方差估计。根据 Kojevnikov（2021）的结果，该估计量具有渐近方差一致性。

4 模拟

我们进行模拟，以评估所提方法的有限样本表现。我们评估点估计量和方差估计量的准确性、自助法方差估计量的有效性，以及基于估计量渐近正态性的置信区间覆盖率。

我们首先从标准正态分布中生成两个独立的协变量 X_{i1} 和 X_{i2}。遵循 Leung（2022年）以及 Gao and Ding（2025年）的方法，我们为每个个体在二维欧几里得空间中分配一个位置，其中两个坐标均独立地从 $[0,1]$上的均匀分布中抽取。我们计算个体 i 与 j 之间的欧几里得距离 d_{ij}，并使用 l个最近邻规则定义排除邻域$\mathcal{N}_i$：$\mathcal{N}_i$ 由个体 i 及其 $l-1$ 个最近邻组成。式（4）中的受约束样本空间$\mathcal{U}_i$ 定义为$\mathcal{U}_i = \{j : d_{ij} < 0.2\}$。

我们根据以下线性模型生成潜在结果：

$$Y_i(\boldsymbol{a}_i) \quad = \quad 1 + 2X_{i1} + 1.5X_{i2} + \sum_{j \in \mathcal{N}_i} a_{ij}(X_{j1} + X_{j2}) + \epsilon_i,$$

其中，$\epsilon_i \sim N(0,1)$。该设定确保每个个体的潜在结果仅取决于$\mathcal{N}_i$内的边，从而满足假设1。

我们固定模拟中的潜在结果，并从式（4）的受约束指数随机图模型中生成具有三个网络统计量的边干预 $\boldsymbol{A}$

$$g_1(\boldsymbol{a},\boldsymbol{x}) = e(\boldsymbol{a}), \quad g_2(\boldsymbol{a},\boldsymbol{x}) = \sum_{1 \le i < j \le n} a_{ij}(x_{i1} + x_{j1}), \quad g_3(\boldsymbol{a},\boldsymbol{x}) = \sum_{1 \le i < j \le n} a_{ij}(x_{i2} + x_{j2}),$$

并将相应的参数向量设为 $\eta = (-1.5,\ 0.5,\ -0.5)^\top$。统计量 g_1 对应于边的总数，而 g_2 和 g_3 则纳入了两个协变量。在这一设定下，给定协变量后，各边相互独立。在附录中，我们进一步探讨了允许边之间存在依赖关系的更复杂指数随机图模型设定。最后，我们根据生成的边干预计算观测结果以及每个个体的潜在结果。

我们考虑三种样本量 $n = 300$、1000 和 5000，并将排除邻域大小设为 $l = 3$、5 和 7。为涵盖不同的干预强度，我们考察随机干预参数的五个取值：$\lambda_\delta = -\log(2)$、$-\log(1.5)$、0、$\log(1.5)$ 和 $\log(2)$。这些取值分别对应于增加一条边的优势比 0.5、0.67、1、1.5 和 2。我们报告 $\hat{\theta}_2^\delta$ 的结果，并在附录中呈现 $\hat{\theta}_1^\delta$ 的结果。

我们首先评估点估计的表现。图1展示了Hájek估计量$\hat{\theta}_2^\delta$的偏差和均方根误差。当 λ_δ 沿任一方向偏离零时，这两项指标均会增加，反映出在更极端的随机干预下变异性更大

。排除邻域越大，偏差和均方根误差越高，这是由于局部处理之间的依赖性增强所致。正如预期，随着样本量增加，偏差和均方根误差均会下降，表明估计量精度得到提高，也支持所提出方法的渐近性质。

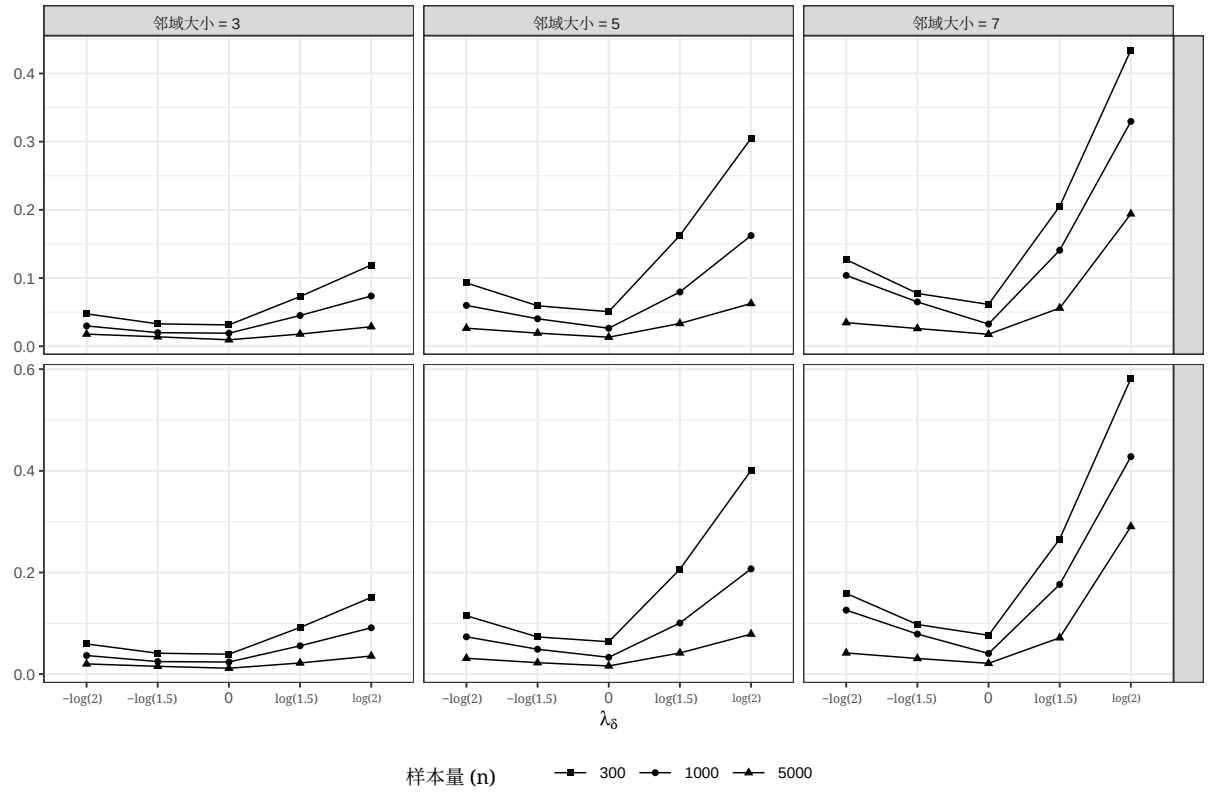


图1：Hájek估计量 $\hat{}\theta_2^{\delta}$ 的偏差和均方根误差。第一行显示偏差，第二行显示均方根误差。每列表示不同的排除邻域大小 $l \in \{3,5,7\}$。

接下来，我们评估式（9）中定义的方差估计量的表现。我们将估计方差与Hájek估计量 $\hat{}\theta_2^{\delta}$ 的真实抽样方差进行比较；后者通过 1,000 次蒙特卡洛重复近似得到。我们考虑两种真实方差。第一种纳入了网络模型估计带来的不确定性。每次重复中，我们根据真实指数随机图模型生成边干预，并在计算估计量之前重新估计模型参数。因此，所得抽样方差同时反映结果变异和模型估计的不确定性。第二种则以真实网络模型参数为条件。每次重复中，我们仍根据真实指数随机图模型生成边干预，并使用已知的真实参数值计算估计量，而不重新估计模型。因此，该方差将指数随机图模型参数视为固定值，不包含模型估计所带来的不确定性。

图2展示了结果。在所有设定下，估计的标准误均高于两种真实标准误，表明方差估计偏保守。这种保守性会随样本量和邻域大小的增加而减弱。一种可能的解释是，方差估计中采用的相关结构可能高估了边之间实际的依赖关系。此外，两种真实标准误通常彼此接近，其中忽略模型不确定性的版本略大。这一经验模式支持我们的方差估计量具有有效性。要对这一现象获得更深入的理论理解，还需要运用超出本文范围的精细概率工具。

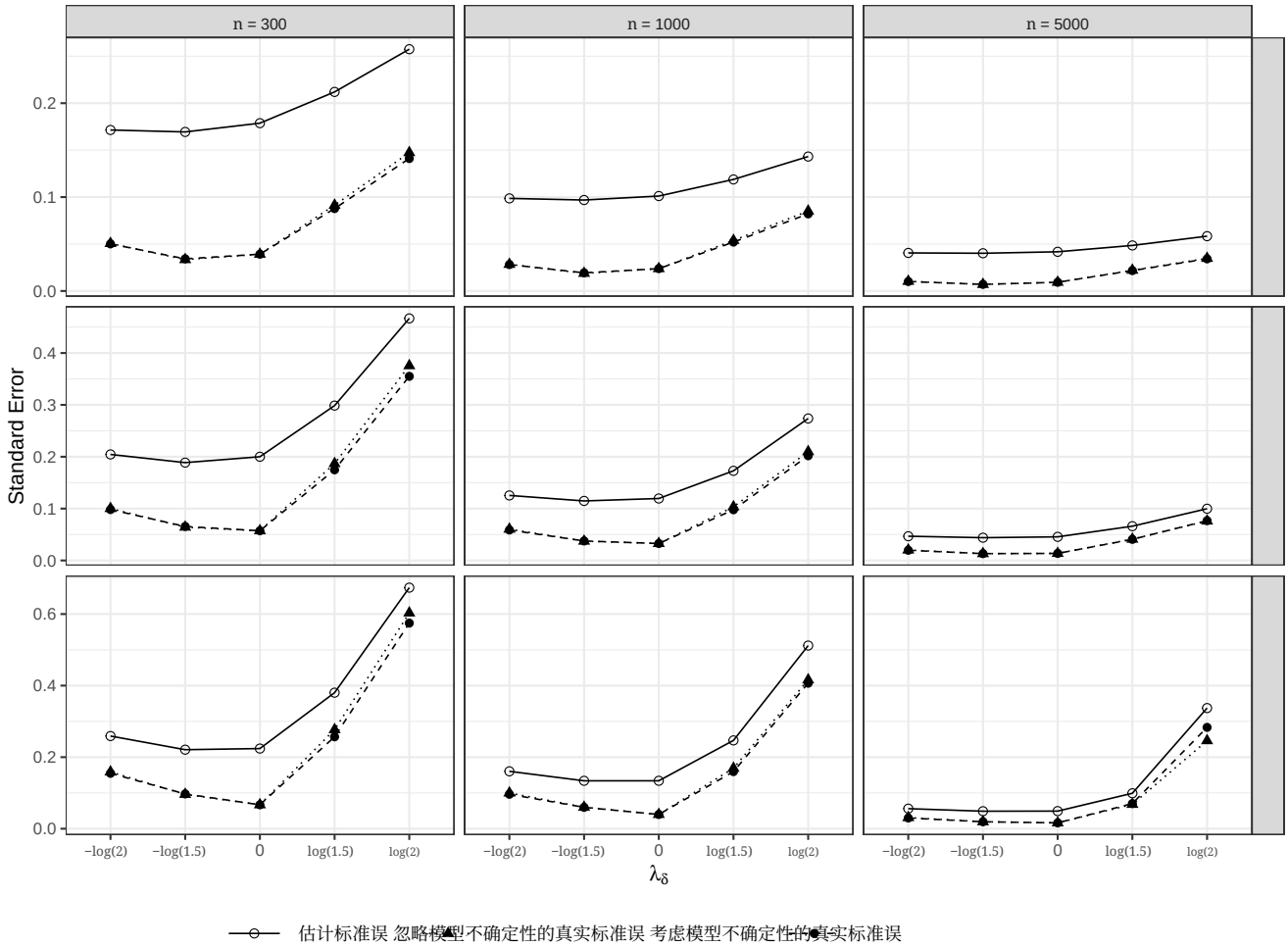


图2：$\hat{\theta}_2^{\delta}$ 的估计标准误与真实标准误（SE）的比较。每个面板显示 λ_δ 取不同值时的标准误。列对应样本量（$n = 300$，1000，5000），行对应排除邻域大小（$l = 3$，5，7）。实线表示估计标准误。虚线表示考虑 ERGM 估计不确定性后的真实标准误。点线表示假设拟合的 ERGM 为真实模型时的真实标准误。

我们在附录中进行了更多模拟。首先，我们使用与正文相同的数据生成过程，考察 $\hat{\theta}_1^{\delta}$ 的表现。其次，我们通过从含有依赖边的指数随机图模型中生成边干预，考虑更复杂的网络模型。为确保满足假设 4，我们在指数随机图模型设定中纳入了 Schweinberger and Handcock（2015）提出的局部依赖结构。

5 实证分析

5.1 数据与设定

现在，我们将所提出的框架应用于评估城际铁路连接对中国经济的影响。Ma 和 Tang（2024）整理的主要数据集记录了1994年至2017年中国279个地级市之间以小时计的铁路旅行时间。为确保分析的一致性，我们从样本中剔除四座城市。莱芜因行政区划调整而被剔除；拉萨、乌鲁木齐和克拉玛依则因地理隔离程度极高，与网络中的其他城市不具可比性而被排除。最终数据集因此包含275个地级市。对于每一城市对，数据集均提供最短可行铁路旅行时间，该时间通过在栅格网格上运行快速行进算法计算得出。网格中的每个像素均根据建设年代、基础设施类别和地形特征赋予相应的设计速度。随后，该算法模拟穿越网格的最优出行路径，以得出城市对之间的最短旅行时间。

为便于分析，我们做了几项简化。首先，我们采用两小时阈值，将城市对之间的旅行时间二分化。如果两座城市之间的最短铁路旅行时间少于两小时，就视为两座城市相连。这一选择兼顾了交通效率和通勤行为。已有研究表明，当铁路旅行时间低于两小时时，铁路会成为主要出行方式（Jorritsma，2009年；Kroes 和 Savelberg，2019年）；通勤研究则指出，这一时长的行程仍适合日常通勤（He 等，2016；De Vos，2025年）。其次，我们聚焦于数据中最近一年的铁路网络，即2017年铁路网络。这些简化构成了一个包含275个节点和647条边的无向无权网络。图3(a)展示了该网络的度分布，其偏斜程度很高。大多数城市只有少数几条直接连接，而少数城市则是重要的交通枢纽。这种格局反映了中国铁路基础设施扩张的选择性和不均衡性；扩张优先考虑了经济和战略上重要的地区。图3(b)将该网络映射到地理底图上，其中节点代表城市，边表示旅行时间在两小时以内的铁路连接。

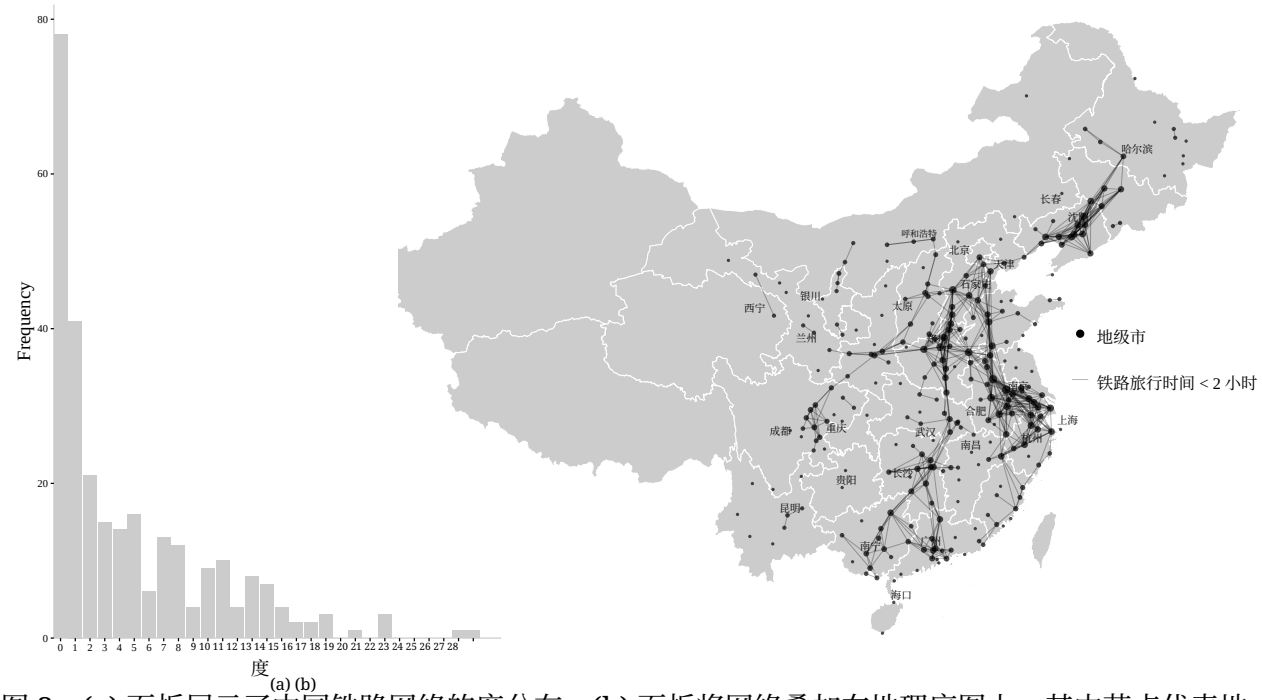


图 3：(a) 面板展示了中国铁路网络的度分布。(b) 面板将网络叠加在地理底图上，其中节点代表地级市，边表示旅行时间少于两小时的铁路连接。节点大小与直接连接的数量成正比。

我们的目标是评估2017年的铁路连接如何影响2018年的第三产业增加值，并以此作为经济发展水平的代理指标。此前的研究考察了高速铁路发展对中国区域增长各方面的影响（例如，Wang 等，2020年；Wang 和 Dong，2022年）。这些研究通常在城市层面界定处理，使用二元指标表示某座城市是否连接到铁路网络，并且往往依赖结构性结果模型。这种以个体为单位的框架忽略了连通性本质上的关系属性，而依赖结果建模则会带来模型设定错误的风险。

5.2 估计与结果

我们通过构建一个明确考虑边干预的因果框架来应对这些局限。在接下来的章节中，我们将定义边干预的因果估计目标，确立识别条件，并提出能够处理边级处理复杂性的估计程序。

我们考察不同处理策略下平均潜在结果的变化，重点关注给定随机干预与观测网络之间的对比 $\theta^{\delta} - \theta^{0}$。我们考虑

干预参数的四个取值： $\lambda_\delta = -\log(2)\ -\log(1.5), \log(1.5), \log(2)$。负值表示降低本地铁路连接可能性的策略，正值则表示提高这种可能性的策略。

我们拟合式（4）中的受约束指数随机图模型，以估计处理概率。网络统计量包括仅网络项和协变量相关项。仅网络项刻画边的总数。对于每个协变量 X_{ip}，我们纳入两个协变量相关项：一个刻画相连城市之间的总体水平，$\sum_{1\le i<j\le n} a_{ij}(x_{ip} + x_{jp})$；另一个衡量各边之间的相似性。对于连续协变量，相似性项的形式为 $\sum_{1\le i<j\le n} a_{ij}|x_{ip} - x_{jp}|$；对于类别协变量，其形式为 $\sum_{1\le i<j\le n} a_{ij}\mathbf{1}(x_i = x_j)$。我们从《中国城市统计年鉴》2016年版中收集了一组丰富的协变量，包括人均国内生产总值、地方财政收入、年平均人口、公路货运量、民航可达性和行政级别。此外，我们还纳入了来自中国地表起伏度数据集的城市级地形起伏数据（You等（2018）），以刻画基础设施扩张所面临的地理约束。

我们通过根据实际距离限制可能连接的集合，对指数随机图模型施加地理约束。稍微滥用一下记号，我们定义$\mathcal{U}_i = \{j : d_{ij} < d_u\}$，其中 d_{ij} 表示城市 i与 j之间的实际距离， d_u 表示距离阈值。这一约束反映了现实中的交通限制：距离过远、无法在两小时内通过铁路抵达的城市对，不太可能形成直接连接。鉴于中国高速铁路的最高运行速度通常为每小时400公里，我们将 $d_u = 800$ 公里设为基准阈值。为评估结果的稳健性，我们在附录中通过改变假定的出行时间进行敏感性分析。具体而言，我们考虑对应于1.5小时和2.5小时出行时间的阈值，分别为 $d_u = 600$ 公里和 $d_u = 1000$ 公里。

我们将城市 i 的排除邻域$\mathcal{N}_i$ 定义为其 l 个最近邻（包括其自身）组成的集合。分析中考虑四种可能的取值 $l \in \{3, 4, 5, 6\}$。这一设定假定每个城市的结果仅取决于其与这些邻居之间的连接，以及这些邻居彼此之间形成的连接。

图4报告了基于Hájek估计量 $\hat{\theta}_2^\delta$ 得出的估计结果。在不同干预水平下，点估计总体呈上升趋势，这与本地铁路连通性改善和第三产业增加值之间可能存在正相关的判断一致。然而，相应的方差估计较大，导致所有干预水平均未达到统计显著性。我们还观察到，估计方差会随着

排除邻域规模的扩大而增大。这一趋势符合预期，因为更大的排除邻域会加剧跨单元依赖，并有效减少可用于估计的独立信息量。基于其他出行时间阈值的补充结果见附录，定性趋势依然相似。

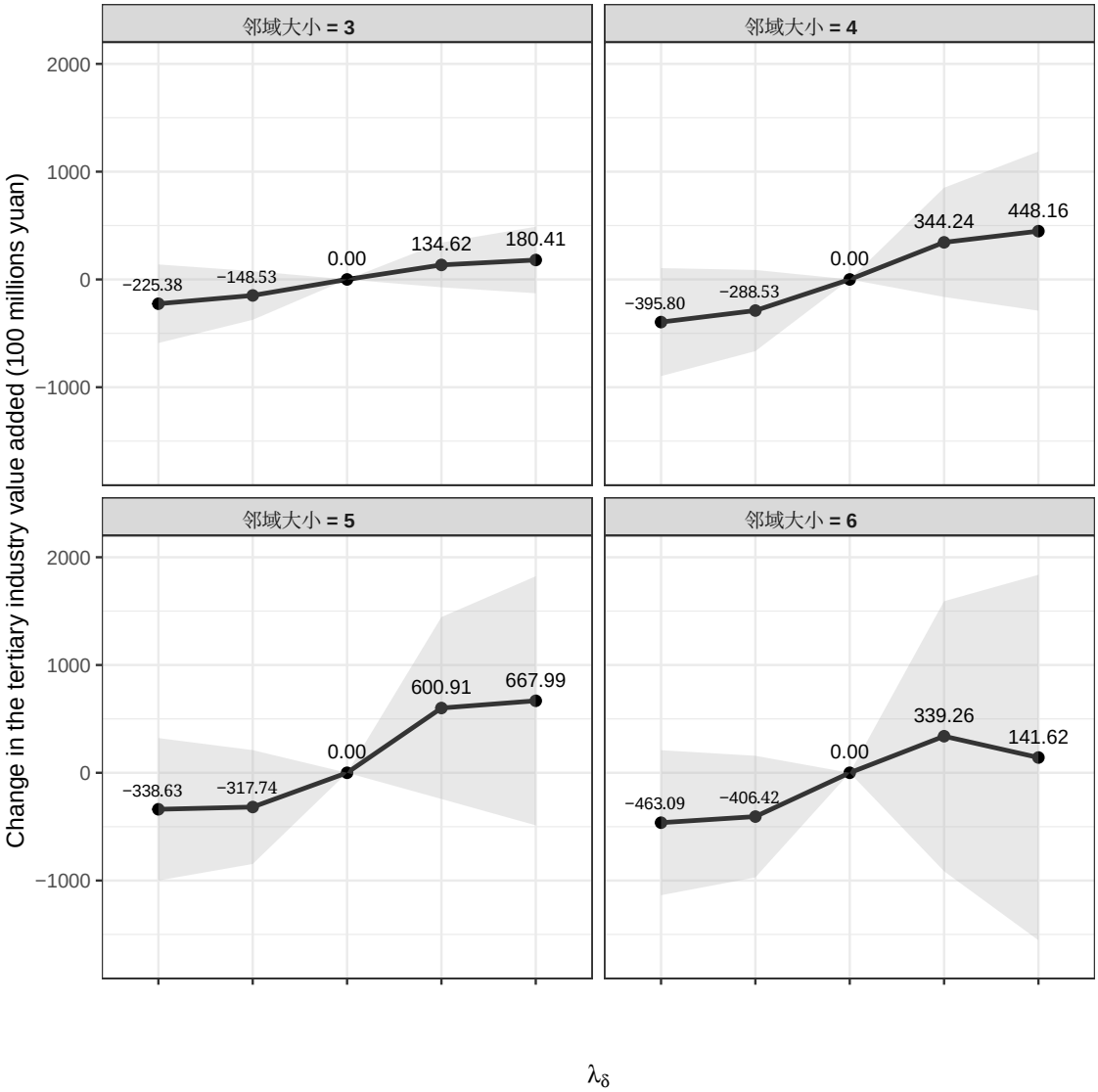


图4：在2小时出行时间阈值下，中国铁路网络对第三产业增加值的影响。每个面板展示了固定排除邻域大小（$l \in 3$、4、5、6）时的结果。曲线显示了 $\theta^{\delta} - \theta^{0}$ 在不同干预强度（λ_{δ}）下的估计值，阴影区域表示95%逐点置信区间。

这些发现可能由多种因素造成。首先，鉴于数据的相关结构，样本量可能过小，难以得出精确的估计结果。其次，各城市的基线条件差异很大。一些城市可能会从连通性改善中显著受益，而另一些城市则可能由于其现有的经济地位、地理位置或产业构成而几乎不会受到影响；

第三，基础设施改善带来的经济影响通常会随着时间推移逐渐显现。由于分析仅关注干预一年后观察到的结果，因此可能无法充分捕捉更长期的结构性变化，尤其是在服务业中，因为该领域的转型往往进展较为缓慢。

6 结论性评述

边干预带来了独特的挑战，包括复杂的干扰模式、处理分配中的重叠有限，以及需要灵活地对处理概率进行建模。我们提出了一种识别和估计边干预因果效应的方法。我们的方法利用局部干扰、随机干预和指数随机图模型来应对这些挑战。在边的依赖结构满足适当条件时，所提出的逆概率加权估计量是一致的，并且渐近正态。

实证分析凸显了网络环境下边级因果推断的可行性及其内在局限性。尽管点估计表明，提高局部铁路连通性可能会产生适度的正向影响，但估计结果并不精确，在统计上也无法与零区分。这一结果反映了网络结构所引致的强依赖，以及由此对有效样本量造成的限制。因此，实证结果应被解读为展示该框架在现实环境中的实施方式和推断表现，而不应被视为提供了精确的政策效应估计。

我们的工作与群体形成领域的文献相关，该领域考察与其他个体组建群体如何影响个体结果（例如，Li 等，2019年；Basse 等，2024年）。群体形成可以被理解为与其他个体建立连接。然而，与我们的环境不同，群体形成的连接彼此互斥，而该领域以往的研究通常聚焦于基于关于干扰的暴露映射假设开展的随机化实验。

我们将连接定义为一个指示变量，用于表示两座城市之间的铁路旅行时间是否少于两小时。这种二分法可能会模糊因果效应的解释。新连接的出现可能源于新铁路的建设、列车速度的提高，或是替代路线的发展使现有行程时间缩短。未来的研究可以将我们的框架扩展到加权边干预，以区分这些机制并提供更细致的解释。

另一个拓展方向是考察边连接的时间动态。铁路网络会随时间演变，因此研究铁路运输的动态演变如何

影响结果颇有意义。要回答这一问题，需要将处理从单个邻接矩阵扩展为一系列随时间观测到的邻接矩阵。我们将这一重要方向留待未来研究。

参考文献

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge（2020年）。回归分析中基于抽样的不确定性与基于设计的不确定性。《计量经济学》88(1)，265–296。

Aronow, P. M. and C. Samii（2017年）。一般干扰下平均因果效应的估计及其在社交网络实验中的应用。《应用统计年鉴》11(4)，1912–1947。

Awan, U., M. Morucci, V. Orlandi, S. Roy, C. Rudin, and A. Volfovsky（2020年）。网络干扰下处理效应估计的近乎精确匹配。载于第二十三届国际人工智能与统计学会议论文集，第108卷，第3252–3262页。PMLR。

Banerjee, A., E. Duflo, and N. Qian（2020年）。论道路：交通基础设施的可及性与中国的经济增长。《发展经济学杂志》145，102442。

Bargagli-Stoffi, F. J., C. Tortú 和 L. Forastiere（2025年）。聚类网络干扰下的异质处理效应和溢出效应。《应用统计年鉴》*19*(1), 28–55。

Barkley, B. G., M. G. Hudgens, J. D. Clemens, M. Ali 和 M. E. Emch（2020年）。利用观察性研究进行聚类干扰下的因果推断，并应用于一项霍乱疫苗研究。《应用统计年鉴》*14*(3), 1432 – 1448。

Basse, G., P. Ding, A. Feller 和 P. Toulis（2024年）。群体形成实验中同伴效应的随机化检验。《计量经济学》*92*(2), 567–590。

Basse, G. 和 A. Feller（2018年）。存在干扰时分析两阶段实验。《美国统计学会杂志》*113*(521), 41–55。

Belloni, A., F. Fang 和 A. Volfovsky（2025年）。网络干扰下因果推断的邻域自适应估计量。*arXiv* 预印本 *arXiv:2212.03683*。

De Vos, J.（2025年）。什么样的通勤令人愉悦：时长接近理想值，还是远低于可忍受的最长时长?《旅行行为与社会》*41*，101095。

Díaz, I. 和 N. S. Hejazi（2020）。随机干预的因果中介分析。《皇家统计学会会刊*B*辑：统计方法学》*82*(3), 661–683。

Díaz, I., N. Williams, K. L. Hoffman 和 E. J. Schenck（2023）。基于纵向修正治疗策略的非参数因果效应。《美国统计学会杂志》*118*(542), 846–857。Ding, P.（2024）。因果推断入门。Chapman and Hall/CRC。Faber, B.（2014）。贸易一体化、市场规模与工业化：中国国家干线公路系统的证据。经济研究评论 *81*(3), 1046–1070。Forastiere, L., E. M. Airoldi 和 F. Mealli（2021）。网络观察性研究中处理效应与干扰效应的识别与估计。《美国统计学会杂志》*116*(534), 901–918。Forastiere, L., F. Mealli, A. Wu 和 E. M. Airoldi（2022）。利用贝叶斯广义倾向评分估计网络干扰下的因果效应。机器学习研究期刊 *23*(289), 1–61。Frank, O. 和 D. Strauss（1986）。马尔可夫图。《美国统计学会杂志》*81*(395), 832–842。Gao, M. 和 P. Ding（2025）。网络实验中的因果推断：基于回归的分析与基于设计的性质。*arXiv* 预印本 *arXiv:2309.07476*。Geyer, C. J. 和 E. A. Thompson （1992）。相关数据的受约束蒙特卡罗最大似然。英国皇家统计学会会刊：*B* 辑（方法论）*54*(3), 657–683。Gonzalez-Navarro, M. 和 C. Quintana-Domeque（2016）。为穷人铺路：基础设施效应的实验分析。经济与统计评论 *98*(2), 254–267。Hainmueller, J.（2012）。用于因果效应估计的熵平衡：一种在观察性研究中生成平衡样本的多变量重新加权方法。政治分析 *20*(1), 25–46。He, M., S. Zhao 和 M. He（2016）。通勤时间的容忍阈值：来自中国昆明的证据。交通地理学杂志 *57*, 1–7。Holland, P. W. 和 S. Leinhardt（1981）。有向图的指数族概率分布。《美国统计学会杂志》*76*(373), 33–50。Hudgens, M. G. 和 M. E. Halloran（2008）。迈向干扰情形下的因果推断。《美国统计学会杂志》*103*(482), 832–842。

Hunter, D. R., S. M. Goodreau 和 M. S. Handcock（2008）。社会网络模型的拟合优度。《美国统计学会杂志》*103*(481), 248–258。Hunter, D. R., M. S. Handcock, C. T. Butts, S. M. Goodreau 和 M. Morris（2008）。ergm：用于拟合、模拟和诊断网络指数族模型的软件包。统计软件杂志 *24*(3), 1–29。Imai, K. 和 D. A. Van Dyk（2004）。一般处理方案下的因果推断：推广倾向得分。《美国统计学会杂志》*99*(467), 854–866。Imbens, G. W.（2000）。倾向得分在剂量-反应函数估计中的作用。*Biometrika 87*(3), 706–710。Imbens, G. W. 和 D. B. Rubin（2015）。《统计、社会与生物医学科学中的因果推断：导论》。剑桥大学出版社。Jagadeesan, R., N. S. Pillai 和 A. Volfovsky（2020）。用于估计存在干扰的网络中处理效应的设计。《统计年鉴》*48*(2), 679–712。Jiang, Z., S. Chen 和 P. Ding（2023）。点过程的工具变量法：基于反卷积的广义沃尔德估计。*Biometrika 110*(4), 989–1008。Jorritsma, P.（2009）。高速列车替代航空运输的机会。*TransportBusiness 7*, 121–128。Karwa, V. 和 E. M. Airoldi（2018）。系统考察网络干扰导致模型设定错误时的经典因果推断策略。*arXiv* 预印本 *arXiv:1810.08259*。Kennedy, E. H.（2019）。基于增量倾向得分干预的非参数因果效应。《美国统计学会杂志》*114*(526), 645–656。Kennedy, E. H., Z. Ma, M. D. McHugh 和 D. S. Small（2017）。连续处理效应双重稳健估计的非参数方法。《皇家统计学会会刊*B*辑：统计方法学》*79*(4), 1229–1245。Kojevnikov, D.（2021）。网络依赖过程的自助法。*arXiv* 预印本 *arXiv:2101.12312* 。Kojevnikov, D., V. Marmer 和 K. Song（2021）。网络依赖随机变量的极限定理。计量经济学杂志 *222*(2), 882–908。

Kroes, E. 和 F. Savelberg（2019）。从航空转向高速铁路的替代：以阿姆斯特丹机场为例。《交通运输研究记录》*2673*(5), 166–174。Lee, C., D. Zeng 和 M. G. Hudgens（2025）。聚类干扰下随机政策效应的高效非参数估计。《美国统计学会杂志》*120*(549), 382–394。Leung, M. P.（2022）。近似邻域干扰下的因果推断。*Economet-rica 90*(1), 267–293。Li, F. 和 F. Li（2019）。多重处理因果推断的倾向得分加权。《医学研究中的统计方法》*28*(3), 785–804。Li, X., P. Ding, Q. Lin, D. Yang 和 J. S. Liu（2019）。同伴效应的随机化推断。《美国统计学会杂志》*114*(528), 1651–1664。Ma, L. 和 Y. Tang（2024）。中国交通网络的分布影响。《国际经济学杂志》*148*, 103873。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）。社会互动下处理反应的识别。《计量经济学杂志》*16*(1), S1–S23。Owusu, J.（2025）。网络干扰下异质性处理效应的随机化推断。*arXiv* 预印本 *arXiv:2308.00202*。Papadogeorgou, G., K. Imai, J. Lyall 和 F. Li（2022）。时空数据的因果推断：估计空袭对伊拉克叛乱暴力的影响。《皇家统计学会会刊*B*辑：统计方法学》*84*(5), 1969–1999。Papadogeorgou, G., F. Mealli 和 C. M. Zigler（2019）。针对聚类和人口层面处理分配方案的相互干扰单位因果推断。《生物统计学》*75*(3), 778–787。Papadogeorgou, G., Z. Song, G. Imbens 和 F. Mealli（2025）。干预单位与结果单位不同时的因果推断。*arXiv* 预印本 *arXiv:2507.20231*。Rosenbaum, P. R. 和 D. B. Rubin（1983）。倾向得分在观察性研究因果效应分析中的核心作用。*Biometrika 70*(1), 41–55。

Sävje, F.（2024年）。基于错误设定的暴露映射进行因果推断：区分定义与假设。*Biometrika* *111*(1), 1–15。Sävje, F., P. Aronow, and M. Hudgens（2021）。存在未知干扰时的平均处理效应。《统计年鉴》*49*(2), 673。Schweinberger, M. and M. S. Handcock（2015年）。随机图模型中的局部依赖：特征描述、性质与统计推断。《皇家统计学会会刊*B*辑：统计方法学》*77*(3), 647–676。Schweinberger, M., P. N. Krivitsky, C. T. Butts, and J. R. Stewart（2020年）。随机图的指数族模型。《统计科学》*35*(4), 627–662。Schweinberger, M. and J. Stewart（2020年）。随机图典则指数族模型与曲指数族模型的集中性和一致性结果。《统计年鉴》*48*(1), 374–396。

Shook-Sa, B. E. 和 M. G. Hudgens（2022年）。观察性研究中点暴露效应的检验效能与样本量。《生物统计学》*78*(1), 388–398。

Snijders, T. A. 等（2002年）。指数随机图模型的马尔可夫链蒙特卡洛估计。《社会结构杂志》*3*(2), 1–40。

Snijders, T. A., P. E. Pattison, G. L. Robins 和 M. S. Handcock（2006年）。指数随机图模型的新设定。《社会学方法论》*36*(1), 99–153。

Sobel, M. E.（2006年）。住房流动随机化研究揭示了什么？面对干扰的因果推断。《美国统计学会杂志》*101*(476), 1398–1407。

Sun, T. 和 S. J. Taylor（2020年）。展示共同点以促进友谊形成：一项大型随机田野实验。《定量营销与经济学》*18*(3), 237–271。Sussman, D. L. 和 E. M. Airoldi（2017年）。存在网络干扰时因果效应估计理论的要素。*arXiv* 预印本 *arXiv:1702.03578*。Tan, R., W. Huang, Z. Zhang 和 G. Yin（2025年）。函数型处理的因果效应。《机器学习研究杂志》*26*(91), 1–39。Toulis, P. 和 E. Kao（2013年）。因果同伴影响效应的估计。载于 S. Dasgupta 和 D. McAllester（编）, 《第*30*届国际机器学习会议论文集》,

《机器学习研究论文集》第28卷，亚特兰大，佐治亚州，美国，第1489–1497页。PMLR。

Viviano, D. 和 J. Rudder（2024年）。未知干扰下实验中的政策设计。*arXiv* 预印本 *arXiv:2011.08174*。

Wang, C.、W. Meng 和 X. Hou（2020年）。高速铁路对城市经济的影响：基于中国城市夜间灯光数据的研究。《交通经济学研究》*80*，100819。

Wang, Y. 和 W. Dong（2022年）。中国高速铁路如何促进地方经济：来自县级面板数据的新证据。《国际经济与金融评论》*80*，67–81。

Wang, Y.、C. Samii、H. Chang 和 P. Aronow（2025年）。未知干扰下空间实验的基于设计的推断。《应用统计年鉴》*19*(1)，744–768。

Wu, X.、K. R. Weinberger、G. A. Wellenius、F. Dominici 和 D. Braun（2024年）。评估时间序列数据中随机干预的因果效应：高温预警在预防死亡和住院方面是否有效？《生物统计学》*25*(1)，57–79。

You, Z., Z. Feng, Y. Yang, 等（2018年）。中国地表起伏度数据集（1 km）。《全球变化数据学报》*2*(2), 151–155。

Zhang, X., W. Xue, 和 Q. Wang（2021）。横断面观察性研究中函数型处理的协变量平衡函数型倾向评分。《计算统计与数据分析》*163*, 107303。

Zhou, L.、K. Imai、J. Lyall 和 G. Papadogeorgou（2024年）。时空因果推断中的异质性处理效应估计：经济援助如何调节空袭对叛乱暴力的影响。*arXiv* 预印本 *arXiv:2412.15128*。

补充材料

S1节提供了相关结果的证明。

S2节报告了更多模拟结果。

S3节进一步介绍了实证分析的细节。

S1 证明

S1.1 定理1的证明

对于每个个体 i，我们有

$$\begin{aligned}\mathbb{E}\left\{\frac{\mathbb{P}(\boldsymbol{A}_i^\delta)}{\mathbb{P}(\boldsymbol{A}_i)}Y_i\right\} &= \mathbb{E}\left\{\sum_{\boldsymbol{a}_i\in\mathcal{A}_i}\frac{\mathbb{P}(\boldsymbol{A}_i^\delta=\boldsymbol{a}_i)}{\mathbb{P}(\boldsymbol{A}_i=\boldsymbol{a}_i)}Y_i(\boldsymbol{a}_i)\mathbf{1}(\boldsymbol{A}_i=\boldsymbol{a}_i)\right\}\\ &= \sum_{\boldsymbol{a}_i\in\mathcal{A}_i}\frac{\mathbb{P}(\boldsymbol{A}_i^\delta=\boldsymbol{a}_i)}{\mathbb{P}(\boldsymbol{A}_i=\boldsymbol{a}_i)}Y_i(\boldsymbol{a}_i)\mathbb{P}(\boldsymbol{A}_i=\boldsymbol{a}_i)\\ &= \theta_i^\delta.\end{aligned}$$

由此得到定理1中的识别公式。 □

S1.2 定理 2 的证明

只需证明 $\text{var}\left(\hat{}\theta^\delta\right)\to 0$。定义

$$\tilde{Y}_i = \frac{\exp\{e(\boldsymbol{A}_i)\lambda_\delta\}}{\mathbb{E}_{\hat{\eta}}\left\{\exp\{e(\boldsymbol{A}_i)\lambda_\delta\}\right\}}Y_i(\boldsymbol{A}_i).$$

那么，$\hat{}\theta_1^\delta$可以写成 $n^{-1}\sum_{i=1}^n\tilde{Y}_i$。根据定义，若 $j\notin\mathcal{R}_i$，则有$(\boldsymbol{A}_i,\boldsymbol{A}_i^\delta)$⫫$(\boldsymbol{A}_j,\boldsymbol{A}_j^\delta)$，$\tilde{}$ $\tilde{}$这意味着 Y_i⫫Y_j。因此，我们有

$$\begin{aligned}\text{var}\left(\hat{\theta}_1^\delta\right) &= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\text{cov}(\tilde{Y}_i,\tilde{Y}_j)\\ &= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\mathbf{1}(j\in\mathcal{R}_i)\text{cov}\left(\tilde{Y}_i,\tilde{Y}_j\right)\\ &\le \frac{C}{n^2}\sum_{i=1}^n\sum_{j=1}^n\mathbf{1}(j\in\mathcal{R}_i)\\ &= \frac{C}{n^2}\sum_{i=1}^n|\mathcal{R}_i|\end{aligned}$$

对于某个常数 $C>0$。根据定理中的条件，我们得到 $\text{var}\left(\hat{}\theta^\delta\right)$ 收敛于零。

□

S1.3 定理3的证明

我们使用 Stein方法进行证明。首先介绍 Stein方法中的一些定义和引理。

定义 ***S1*** 设 W 和 Z 是两个随机变量。令 ***G*** 表示所有 ***Lipschitz*** 函数$g : \mathbb{R} \to \mathbb{R}$的集合。

(a) 柯尔莫哥洛夫–斯米尔诺夫距离定义为:

$$\mathrm{Kolm}(W, Z) = \sup_{x \in \mathbb{R}} |\mathbb{P}(W \leq x) - \mathbb{P}(Z \leq x)| .$$

*(b) Wasserstein*距离定义为:

$$\mathrm{Wass}(W, Z) = \sup_{g \in \mathcal{G}} |\mathbb{E}[g(W)] - \mathbb{E}[g(Z)]| .$$

引理 ***S1*** 假设 Z 服从标准正态分布。那么,

$$\mathrm{Kolm}(W, Z) \leq 2\sqrt{\mathrm{Wass}(W, Z)/\sqrt{2\pi}}.$$

引理 ***S2*** 假设 Z 服从标准正态分布。那么,

$$\mathrm{Wass}(W, Z) \;=\; \sup_{g \in \mathcal{G}} |\mathbb{E}[g(W)] - \mathbb{E}[g(Z)]| \;\leq\; \sup_{f \in \mathcal{F}} \left|\mathbb{E}\left[f'(W) - W f(W)\right]\right| ,$$

其中

$$\mathcal{F} \;=\; \left\{ f \in \mathcal{F} : \|f\|_\infty \leq 1, \|f'\|_\infty \leq \sqrt{\frac{2}{\pi}}, \|f''\|_\infty \leq 2 \right\}.$$

这些引理的证明是标准的, 因此省略。

接下来证明定理3。定义

$$W_i \;=\; \frac{\exp\{e(\boldsymbol{A}_i)\lambda_\delta\}}{\mathbb{E}_{\hat{\eta}}\left\{\exp\{e(\boldsymbol{A}_i)\lambda_\delta\}\right\}} Y_i(\boldsymbol{A}_i) - \theta_i^\delta, \quad \sigma^2 \;=\; \mathrm{var}\left(\sum_{i=1}^n W_i\right), \quad W \;=\; \frac{\sum_{i=1}^n W_i}{\sigma}.$$

我们可以写作

$$\frac{\hat{\theta}_1^\delta - \theta_\delta}{\sqrt{\mathrm{var}(\hat{\theta}_1^\delta)}} \;=\; W.$$

我们基于引理 S2 推导 Wass(W, Z) 的界。定义

$$W_{-i} \;=\; \frac{\sum_{j=1}^n \mathbf{1}(j \notin \mathcal{R}_i) W_j}{\sigma}.$$

根据定理 2 的证明，我们有 $W_i \perp\!\!\!\perp W_{-i}$。对于任意函数 $f \in \mathcal{F}$，我们有

$$\mathbb{E}\{Wf(W)\} \;=\; \frac{1}{\sigma}\sum_{i=1}^{n}\mathbb{E}\{W_i f(W)\} \;=\; \frac{1}{\sigma}\sum_{i=1}^{n}\mathbb{E}\left[W_i\{f(W)-f(W_{-i})\}\right],$$

其中，第二个等号来自 $\mathbb{E}\{W_i f(W_{-i})\} = \mathbb{E}(W_i)\mathbb{E}\{f(W_{-i})\} = 0$。我们还可以将 $\mathbb{E}\{Wf(W)\}$ 进一步分解为 $\mathbb{E}\{Wf(W)\} = B_1 + B_2$，其中

$$\begin{aligned} B_1 &= \frac{1}{\sigma}\sum_{i=1}^{n}\mathbb{E}\left[W_i\left\{f(W)-f(W_{-i})-(W-W_{-i})f'(W)\right\}\right], \\ B_2 &= \frac{1}{\sigma}\sum_{i=1}^{n}\mathbb{E}\left\{W_i(W-W_{-i})f'(W)\right\}. \end{aligned}$$

对 B_1应用泰勒展开，我们得到

$$B_1 \;\le\; \frac{1}{2\sigma}|f''|_\infty\sum_{i=1}^{n}\mathbb{E}|W_i(W-W_{-i})^2| \;\le\; \frac{1}{\sigma^3}\sum_{i=1}^{n}\mathbb{E}\left|W_i\left(\sum_{j\in\mathcal{R}_i}W_j\right)^2\right|.$$

对于 B_2，我们可以写成 $B_2 = \mathbb{E}\{f'(W)T\}$，其中

$$T \;=\; \frac{1}{\sigma}\sum_i W_i(W-W_{-i}).$$

因此，我们有

$$B_2 - f'(W) \;\le\; |f'|_\infty\cdot\mathbb{E}|T-1| \;\le\; \sqrt{\frac{2}{\pi}}\sqrt{\mathbb{E}(T-1)^2} \;=\; \sqrt{\frac{2}{\pi}}\sqrt{\mathrm{Var}(T)},$$

其中，最后一个等号由下式得出

$$\mathbb{E}(T) \;=\; \frac{1}{\sigma}\sum_{i=1}^{n}\mathbb{E}\left(WW_i\right) \;=\; \mathbb{E}\left(W^2\right) \;=\; 1.$$

因此，

$$\begin{aligned} &\left|\mathbb{E}\{Wf(W)\}-\mathbb{E}\{f'(W)\}\right| \\ \le\;& \frac{1}{\sigma^3}\sum_{i=1}^{n}\mathbb{E}\left|W_i\left(\sum_{j\in\mathcal{R}_i}W_j\right)^2\right| + \sqrt{\frac{2}{\pi}}\sqrt{\mathrm{Var}(T)} \\ =\;& \frac{1}{\sigma^3}\sum_{i=1}^{n}\sum_{j,k\in\mathcal{R}_i}\mathbb{E}\left|W_iW_jW_k\right| + \sqrt{\frac{2}{\pi}}\cdot\sqrt{\mathrm{var}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j\in\mathcal{R}_i}W_iW_j\right)} \\ \le\;& \frac{1}{3\sigma^3}\sum_{i}\sum_{j,k\in\mathcal{R}_i}\left(\mathbb{E}|W_i|^3+\mathbb{E}|W_j|^3+\mathbb{E}|W_k|^3\right) + \sqrt{\frac{2}{\pi}}\cdot\sqrt{\mathrm{var}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j\in\mathcal{R}_i}W_iW_j\right)}, \qquad \text{(S1)} \end{aligned}$$

其中，最后一个等号由算术-几何平均不等式得出。记 $D = \max_i|\mathcal{R}_i|$。对于 (S1) 右侧的第一项，我们有

$$\sum_{i=1}^{n}\sum_{j,k\in\mathcal{R}_i}\left(\mathbb{E}|W_i|^3+\mathbb{E}|W_j|^3+\mathbb{E}|W_k|^3\right)$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} |\mathcal{R}_i|^2 |W_i|^3 + 2\sum_{i=1}^{n}\sum_{j\in\mathcal{R}_i} |\mathcal{R}_i||W_j|^3 \\
&\leq D^2 \sum_{i=1}^{n} |W_i|^3 + 2D\sum_{i=1}^{n}\sum_{j\in\mathcal{R}_i} |W_j|^3 \\
&= D^2 \sum_{i=1}^{n} |W_i|^3 + 2D\sum_{j=1}^{n}\sum_{i=1}^{n} \mathbf{1}(j\in\mathcal{R}_i)|W_j|^3 \\
&= D^2 \sum_{i=1}^{n} |W_i|^3 + 2D\sum_{j=1}^{n}\sum_{i=1}^{n} \mathbf{1}(i\in\mathcal{R}_j)|W_j|^3 \\
&\leq 3D^2 \sum_{i=1}^{n} |W_i|^3. \qquad \text{(S2)}
\end{aligned}
$$

其中，最后一个等号由 $i \in \mathcal{R}j$ 与 $j \in \mathcal{R}_i$ 之间的等价关系得出。

对于 (S1) 右侧的第二项，我们有

$$
\mathrm{var}\left(\sum_{i=1}^{n}\sum_{j\in\mathcal{R}_i} W_iW_j\right) \leq \sum_{i=1}^{n}\sum_{j\in\mathcal{R}_i}\sum_{k=1}^{n}\sum_{l\in\mathcal{R}_k} |\mathrm{cov}(W_iW_j, W_kW_l)|.
$$

如果 k，$l \notin \mathcal{R}_i \cap \mathcal{R}j$，那么 W_iW_j 与 W_kW_l 独立。因为 $|\mathcal{R}_i \cap \mathcal{R}_j| \leq 2D$，且$\mathcal{R}_i \cap \mathcal{R}j$ 中的每个个体最多有 D 个邻居，我们有

$$
\begin{aligned}
\sum_{k=1}^{n}\sum_{l\in\mathcal{R}_k} |\mathrm{cov}(W_iW_j, W_kW_l)| &\leq 2D^2 \max_{k\in\mathcal{N}, l\in\mathcal{R}_k} |\mathrm{cov}(W_iW_j, W_kW_l)| \\
&= 2D^2 \frac{\max_{k\in\mathcal{N}, l\in\mathcal{R}_k} |\mathrm{cov}(W_iW_j, W_kW_l)|}{\mathrm{var}(W_iW_j)} \mathrm{var}(W_iW_j) \\
&\leq 2CD^2 \mathrm{var}(W_iW_j),
\end{aligned}
$$

其中常数 C 的选取满足

$$
\frac{\max_{k\in\mathcal{N}, l\in\mathcal{R}_k} |\mathrm{cov}(W_iW_j, W_kW_l)|}{\mathrm{var}(W_iW_j)} \leq C.
$$

因此，我们有

$$
\begin{aligned}
\mathrm{var}\left(\sum_{i=1}^{n}\sum_{j\in\mathcal{R}_i} W_iW_j\right) &\leq 2CD^2 \sum_{i=1}^{n}\sum_{j\in\mathcal{R}_i} \mathrm{var}(W_iW_j) \\
&\leq 2CD^2 \sum_{i=1}^{n}\sum_{j\in\mathcal{R}_i} \mathbb{E}(W_i^2W_j^2) \\
&\leq CD^2 \sum_{i=1}^{n}\sum_{j\in\mathcal{R}_i} \left\{\mathbb{E}(W_i^4) + \mathbb{E}(W_j^4)\right\} \\
&\leq CD^3 \sum_{i=1}^{n} \mathbb{E}(W_i^4) + CD^2 \sum_{i=1}^{n} \mathbf{1}(i\in\mathcal{R}_j)\mathbb{E}(W_j^4) \\
&\leq 2CD^3 \sum_{i=1}^{n} \mathbb{E}(W_i^4). \qquad \text{(S3)}
\end{aligned}
$$

将（S2）和（S3）代入（S1），可得

$$|\mathbb{E}\{Wf(W)\} - \mathbb{E}\{f'(W)\}| \;\leq\; \frac{D^2}{\sigma^3}\sum_{i=1}^{n}\mathbb{E}|W_i|^3 + \frac{2C}{\sqrt{\pi}\sigma^2}\cdot\sqrt{D^3\sum_{i=1}^{n}\mathbb{E}(W_i^4)}. \tag{S4}$$

由于结果有界，我们有

$$\sum_{i=1}^{n}\mathbb{E}|W_i|^3 \;=\; O_{\mathbb{P}}(n), \quad \sum_{i=1}^{n}\mathbb{E}(W_i^4) \;=\; O_{\mathbb{P}}(n). \tag{S5}$$

根据定义，$\mathrm{var}(\hat{}\theta_1^{\delta}) = \sigma^2/n^2$。因此，关于 $\max_i|\mathcal{R}_i|$ 的条件意味着

$$\frac{nD^2}{\sigma^3} \;=\; o_{\mathbb{P}}(1), \quad \frac{n^{1/2}D^{3/2}}{\sigma^2} \;=\; o_{\mathbb{P}}(1). \tag{S6}$$

将 (S5) 和 (S6) 代入 (S4)，可得 $|\mathbb{E}\{Wf(W)\} - \mathbb{E}\{f'(W)\}| = o_{\mathbb{P}}(1)$。由引理 S1 和 S2 可得 $\hat{}\theta_1^{\delta}$ 的渐近正态性。□

S2 额外模拟结果

S2.1 伯努利随机图模型下 $\hat{}\,\theta_1^{\delta}$ 的模拟结果

图 S1 展示了在与正文相同的数据生成过程中 $\hat{}\theta_1^{\delta}$ 的偏差和均方根误差（RMSE）。与 Hájek估计量 $\hat{}\theta_2^{\delta}$ 相比，随着邻域大小和干预强度增加，$\hat{}\theta_1^{\delta}$ 表现出更大的偏差和均方根误差。当这两个量都较大时，这种模式尤为明显。尽管如此，随着样本量增大，偏差和均方根误差都会减小，这与大数定律的预期一致。

图 S2 比较了针对 $\hat{}\theta_1^{\delta}$ 的估计标准误与两种真实标准误。其模式与 Hájek估计量的情况相似：估计标准误通常大于真实值。

S2.2 具有局部依赖的指数随机图模型下的模拟结果 $\mathbf{e}$

我们在允许存在依赖边的更复杂指数随机图模型下进行模拟。为确保假设 4 成立，我们采用 Schweinberger and Handcock (2015) 提出的具有局部依赖的指数随机图模型。

定义 ***S2***（局部依赖） 如果可以将个体划分为 $K \geq 2$ 个非空子集 M_1、.....、M_K，使得依赖关系仅存在于区块内部，则该随机图模型具有局部依赖。形式上， $\boldsymbol{A}$ 的概率质量函数满足

$$\Pr(\boldsymbol{A}=\boldsymbol{a}) = \prod_{k=1}^{K}\left\{\Pr\left(A_{\mathcal{M}_k}=a_{\mathcal{M}_k}\right)\prod_{l=1}^{k-1}\prod_{i\in\mathcal{M}_k,\ j\in\mathcal{M}_l}\Pr(A_{ij}=a_{ij})\right\}, \quad a\in\mathcal{A}, \tag{S7}$$

其中， $A_{\mathcal{M}_k}$ 表示由 M_k 中所有个体构成的子图的邻接矩阵。

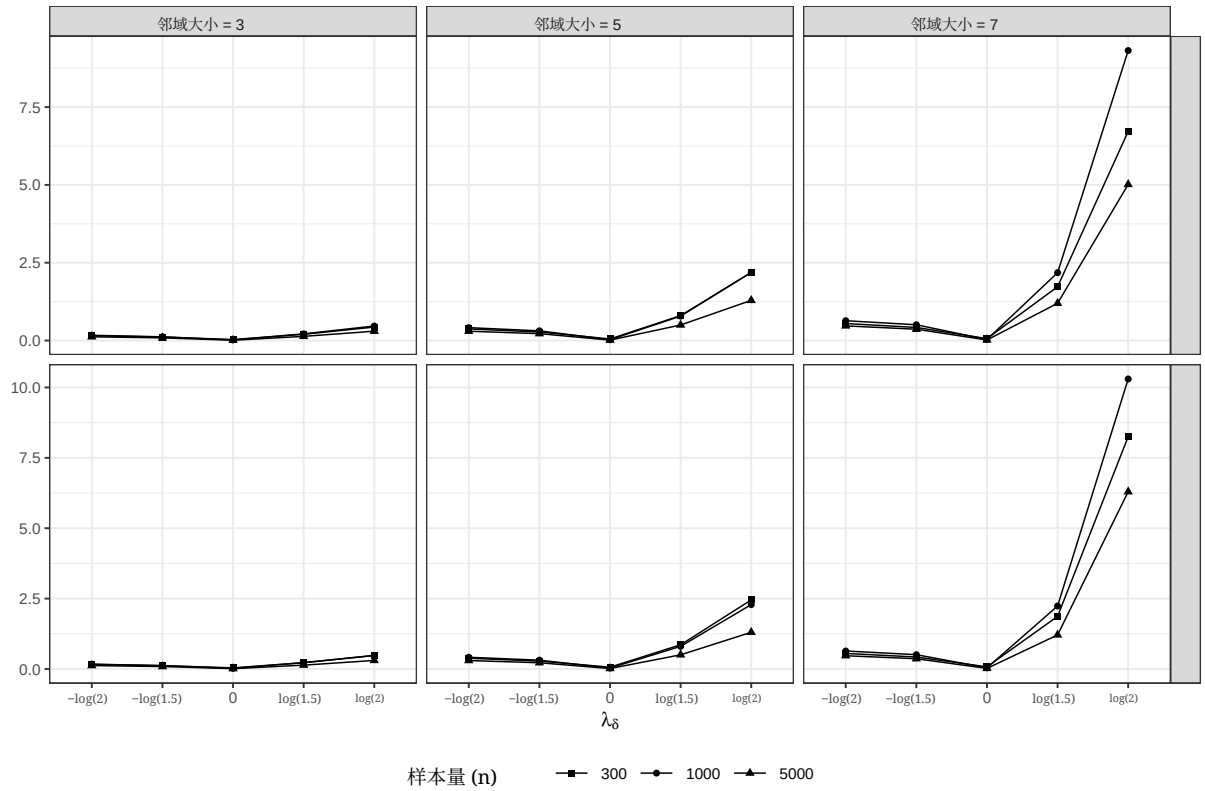


图 S1：$\hat{\theta}_1^{\delta}$ 的偏差和均方根误差。第一行显示偏差，第二行显示均方根误差。每列代表不同的排除邻域大小 $l \in \{3, 5, 7\}$。

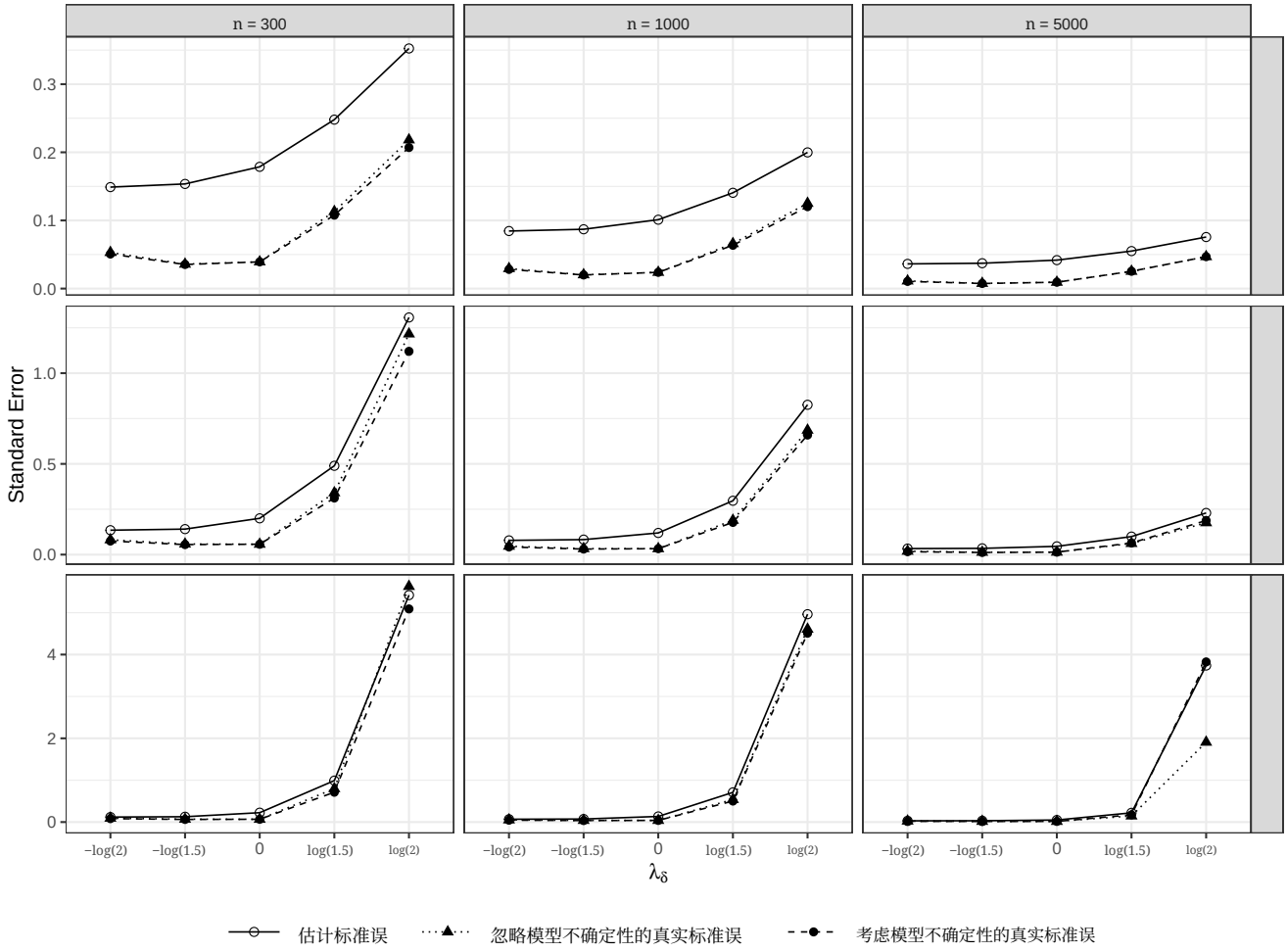


图 S2：$\hat{\theta}_1^{\delta}$ 的估计标准误（SE）与真实标准误的比较。每个面板展示了不同 λ_δ 值下的标准误。列对应不同的样本量（$n = 300$，1000，5000），行对应不同的排除邻域大小（$l = 3$，5，7）。实线表示估计的标准误。虚线表示计入 ERGM 估计不确定性后的真实标准误。点线表示假设拟合的 ERGM 为真实模型时的真实标准误。

令 $\boldsymbol{w}=(w_1、\ldots\ldots、w_n)$ 表示区块归属，其中 $w_i=k$ 表示个体 i 属于区块 $\mathcal{M}_k$。我们假设区块内和区块间子网络均服从指数族分布。具体而言，区块内子网络服从以下形式的指数随机图模型：

$$\Pr(A_{\mathcal{M}_k}=a_{\mathcal{M}_k})\propto\exp\left\{\eta_k^T g\left(a_{\mathcal{M}_k},\boldsymbol{X}_{\mathcal{M}_k}\right)\right\},$$

其中 $\boldsymbol{X}_{\mathcal{M}_k}=\{X_i:i\in\mathcal{M}_k\}$， $g(\cdot)$ 是一个包含边和协变量的区块内网络统计量向量。对于区块间的边，我们假设其服从相互独立的伯努利模型：

$$\Pr(A_{ij}=a_{ij})\propto\exp\left\{\eta_B a_{ij}+\eta_x^T a_{ij}g(x_i,x_j)\right\},\quad w_i\neq w_j.$$

结合这两个部分， $\boldsymbol{A}$ 的完整模型为

$$\Pr(\boldsymbol{A}=\boldsymbol{a})\propto\exp\left\{\sum_{k=1}^{K}\eta_k^T g\left(a_{\mathcal{M}_k},\boldsymbol{X}_{\mathcal{M}_k}\right)+\eta_B\sum_{i<j,w_i\neq w_j}a_{ij}+\eta_x^T\sum_{i<j,w_i\neq w_j}a_{ij}g(x_i,x_j)\right\}. \tag{S8}$$

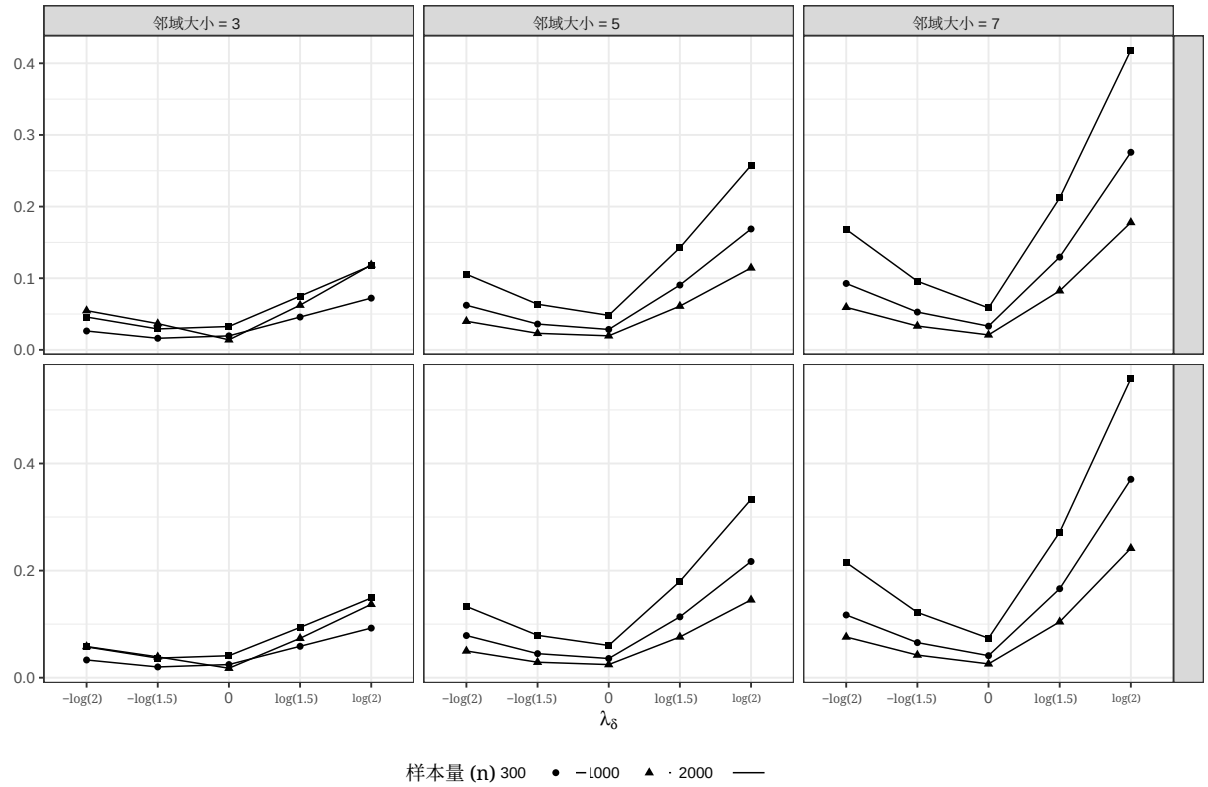
在每个区块内，我们定义四个网络统计量：

$$\begin{aligned}
g_1(\boldsymbol{a},\boldsymbol{x}) &= e(\boldsymbol{a}),\quad g_2(\boldsymbol{a},\boldsymbol{x})=\sum_{1\le i<j\le n}a_{ij}(x_{i1}+x_{j1}),\quad g_3(\boldsymbol{a},\boldsymbol{x})=\sum_{1\le i<j\le n}a_{ij}(x_{i2}+x_{j2}),\\
g_4(\boldsymbol{a},\boldsymbol{x}) &= 3\sum_{i<j<k}a_{ij}a_{ik}a_{jk}+\sum_{k=2}^{n-2}(-e^{0.3})^{-(k-1)}\sum_{i<j<k}a_{ij}\binom{L_{ij}}{k},
\end{aligned} \tag{S9}$$

其中， $L_{ij}=\sum_k a_{ik}a_{kj}$ 是个体 i 和 j的共同邻居数。前三个统计量与正文中使用的统计量相同。第四个统计量是几何加权边共享伙伴统计量，它通过汇总三角形和 k-三角形构型来捕捉局部聚类。对于簇间部分，我们设定 $g(x_i,\ x_j)=x_i+x_j$，以捕捉跨区块的协变量总体水平。我们还通过限制样本空间$\mathcal{A}_u$、排除特定边，对网络施加结构约束，从而反映实际限制或特定领域知识。

我们按照 (S8) 和 (S9) 中的设定，从具有局部依赖的指数随机图模型中生成边干预。我们将样本随机划分为区块，每个区块平均包含 50 个个体，并采用与正文相同的受限样本空间$\mathcal{A}_u$。数据生成过程与正文中的过程相同，但排除邻域$\mathcal{N}_i$的定义除外。我们修改$\mathcal{N}_i$ 的定义，使其仅包含与个体 i 属于同一区块的个体。具体而言，$\mathcal{N}_i$ 由与个体 i 具有相同区块归属的 l 个最近邻组成。按照这一设定，排除邻域中的所有个体都属于同一区块，因此相应的依赖邻域为$\mathcal{R}_i=\{j:w_j=w_i\}$，即与 i 属于同一区块的所有个体。

我们考虑三种样本量： $n = 300$，1000，2000。对于每种情形，我们评估已知边模型时 Hájek 估计量 $\hat{\theta}_2^{\delta}$ 的表现。图S3展示了该估计量的偏差和均方根误差。结果与正文中伯努利随机图模型下得到的结果一致。



图S3：局部依赖模型下 Hájek估计量 $\hat{\theta}_2^{\delta}$ 的偏差和均方根误差。第一行显示偏差，第二行显示均方根误差。每列代表不同的排除邻域大小 $l \in \{3,5,7\}$。

S3 实证分析的其他细节

S3.1 估计结果与模型诊断

表1列出了受约束指数随机图模型的估计系数。结果表明，政府收入较高、货运交通量较大的城市，在铁路网络中更有可能彼此连接。相比之下，地形起伏度与边形成显著负相关，说明不利的地理条件会大幅阻碍铁路连通性。对于协变量差异项，大多数属性的系数均显著为负，表明网络中存在明显的同质性。具体而言，经济水平相近的城市

表1：受约束指数随机图模型的点估计，括号内为标准误。“边”表示边的总数。标记为“nodecov”的项表示每个连续协变量的$\sum_{1\leq i<j\leq n} a_{ij}(x_i + x_j)$。标记为“absdiff”的项表示连续协变量的 $\sum_{1\leq i<j\leq n} a_{ij}|x_i - x_j|$，以及类别协变量的$\sum_{1\leq i<j\leq n} a_{ij}\mathbf{1}(x_i = x_j)$。

模型项	系数	模型项	系数
边	-0.590 (0.160)***	收入绝对差	-0.130 (0.126)
节点协变量.GDP_人均_	-0.049 (0.052)	货运_交通绝对差	-0.431 (0.099)***
节点协变量.收入	0.304 (0.125)*	绝对差.人口	-0.299 (0.070)***
节点协变量.货运_交通	0.397(0.090)***	绝对差.地形	-1.670 (0.166)***
节点协变量.人口	-0.128 (0.051)*	节点匹配.机场	0.065 (0.085)
节点协变量.地形	-0.199 (0.062)**	节点匹配：首都	-0.897 (0.113)***
绝对差.GDP_人均_	-0.361 (0.067)***		

注：*** $p < 0.001$，** $p < 0.01$，* $p < 0.05$。括号内为标准误。

发展、货运活动、人口规模和地形特征显著更有可能形成直接连接。

我们按照Hunter等（2008）概述的程序，评估受约束指数随机图模型的拟合优度。具体而言，我们根据拟合模型生成100个网络，并将网络统计量和节点度的分布与观测值进行比较。

图 S4总结了结果。左图中，观测到的网络统计量接近生成的100个网络相应统计量的中位数。右图中，观测到的度分布与模拟分布没有显著偏离。综合来看，这些结果表明，该模型对观测网络的拟合效果较为合理

S3.2 敏感性分析

我们通过改变直接出行时间截断值开展敏感性分析，以评估研究发现的稳健性。除主要分析中采用的2小时阈值外，我们还考虑了1.5小时和2.5小时这两个替代截断值，分别对应 $d_u = 600$千米和 $d_u = 1000$ 千米的距离阈值。图S5和图S6展示了这些替代设定下的估计结果。在这两种情形下，定性模式总体上与主要分析中的结果一致。

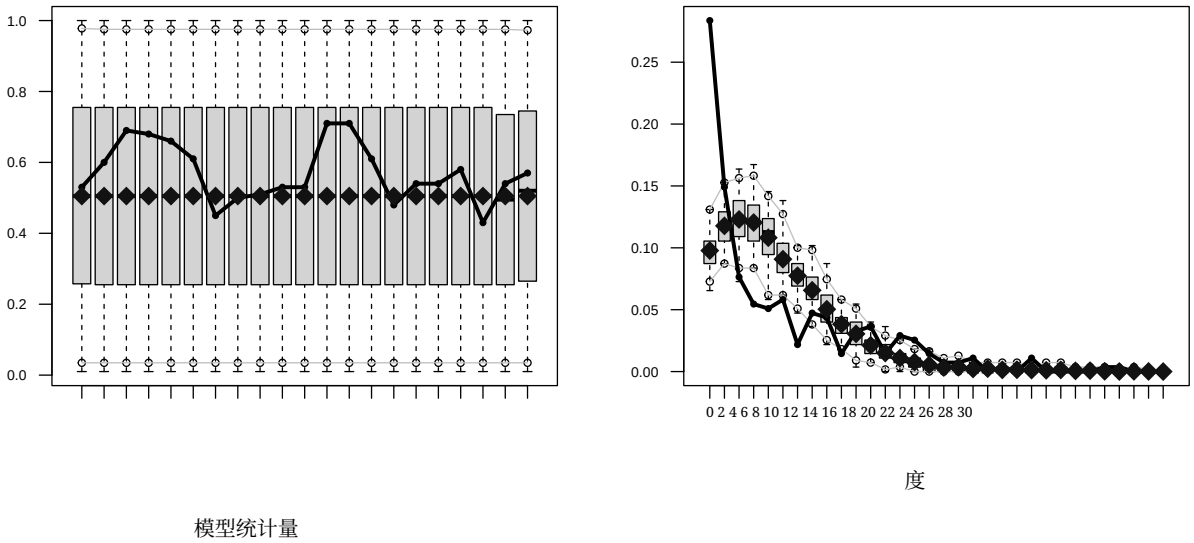


图 S4：具有局部依赖的受约束指数随机图模型的拟合优度诊断。左图比较观测网络统计量与 100 个模拟网络的统计量。灰色箱线图显示各统计量在模拟网络中的分布（标准化至区间 $[0, 1]$），黑线表示观测统计量在该分布中的百分位数。右图比较观测度分布与 100 个模拟网络的度分布，采用与左图相同的箱线图—百分位数展示方式。

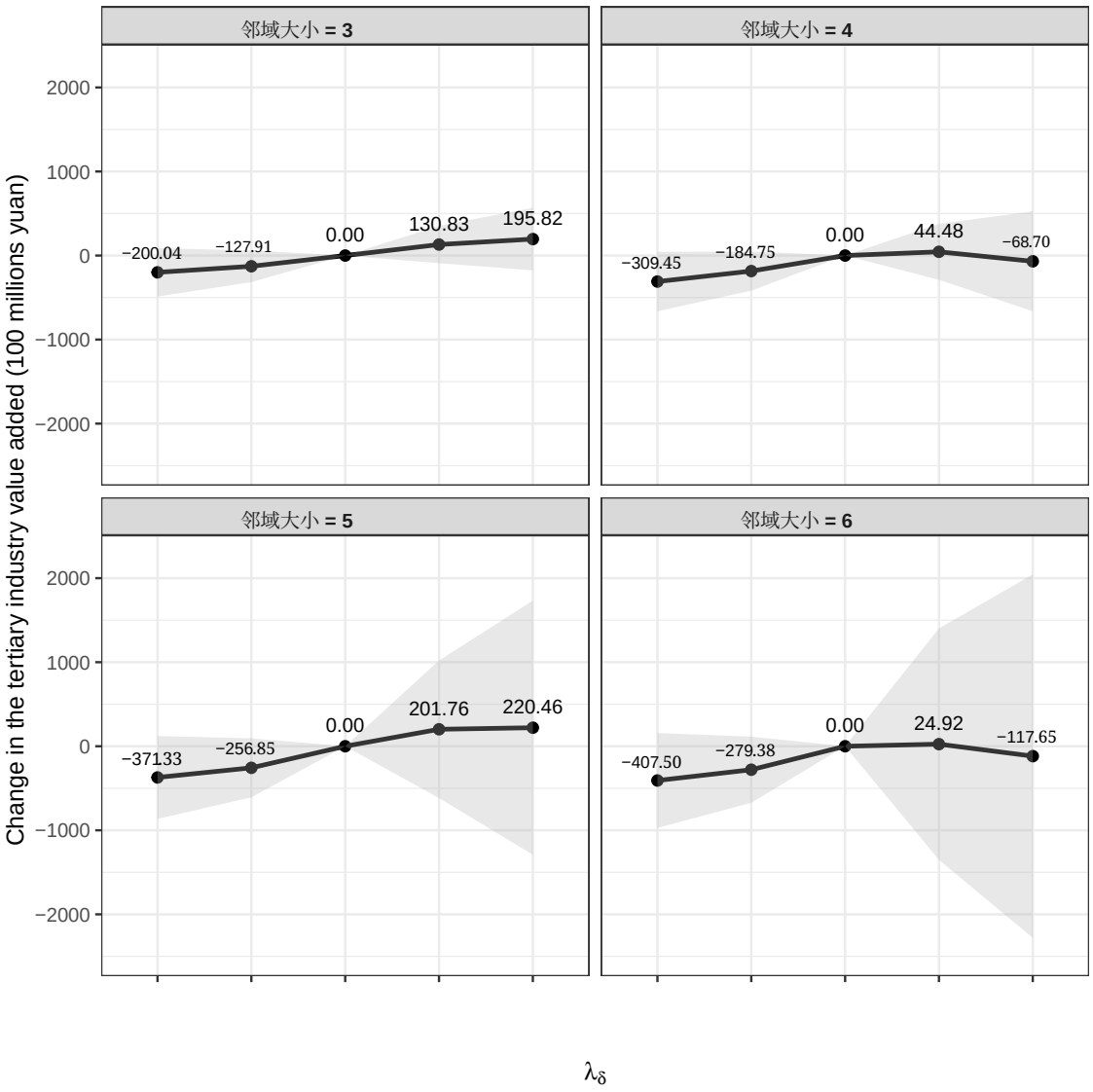


图 S5：中国铁路网络在1.5小时旅行时间阈值下对第三产业增加值的影响。每个面板展示固定排除邻域大小（$l \in 3$、4、5、6）下的结果。各条线显示 $\theta^{\delta} - \theta^{0}$ 在不同干预强度（λ_{δ}）下的估计值，阴影区域表示95%逐点置信区间。

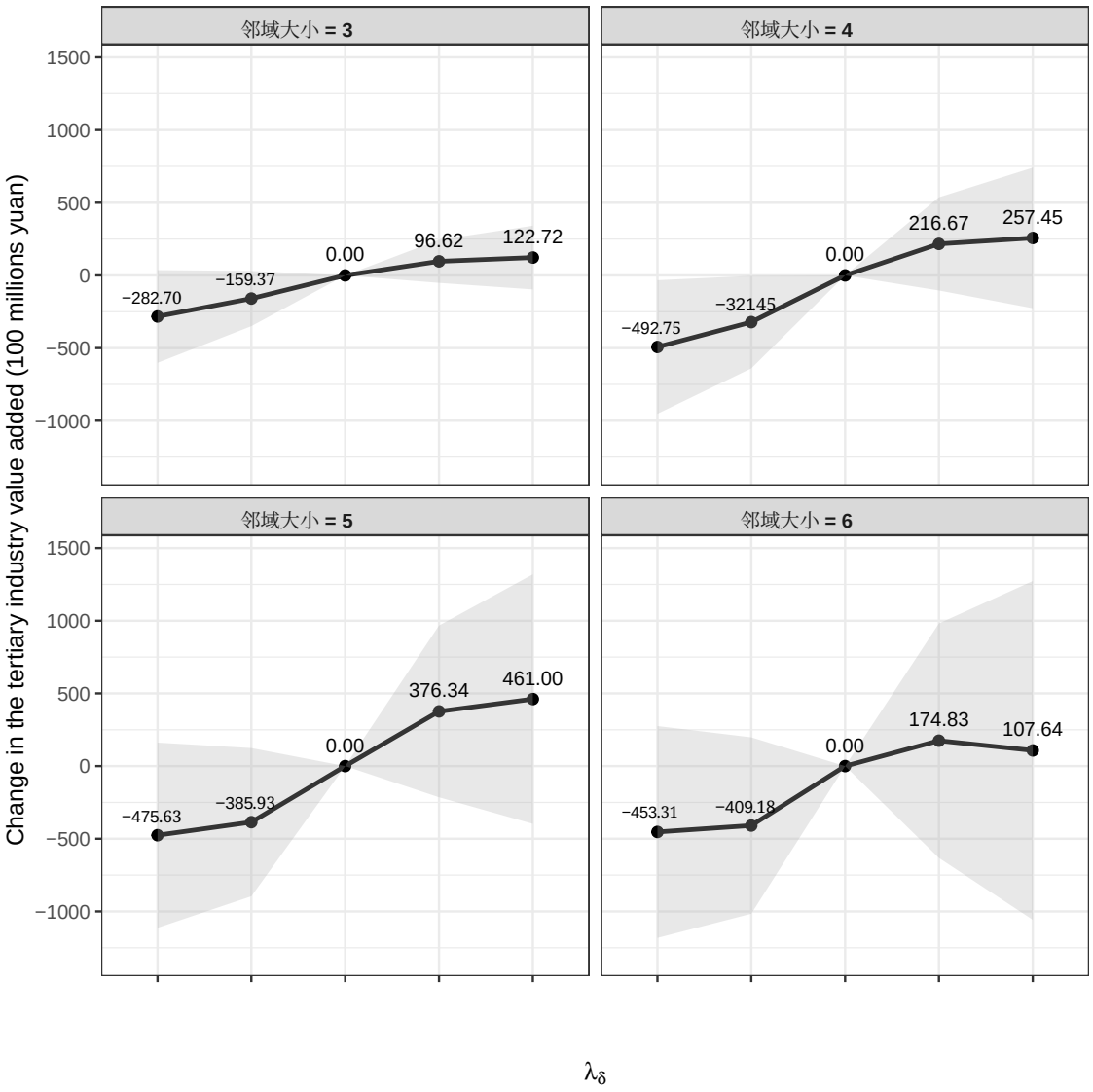


图 S6：中国铁路网络在 2.5 小时出行时间阈值下对第三产业增加值的影响。每个面板展示了固定排除邻域大小下的结果（$l \in 3$、4、5、6）。曲线表示 $\theta^{\delta} - \theta^{0}$ 在不同干预强度（λ_{δ}）下的估计值，阴影区域表示 95%逐点置信区间。