arXiv:2602.04230v1 [stat.ME] 4 Feb 2026

在真实实验中验证因果消息传递与网络感知方法

Albert Tan* Sadegh Shirani† James Nordlund* Mohsen Bayati*,†

摘要

在存在网络干扰的情况下估计总处理效应，通常需要了解底层交互结构。然而，在许多实际情境中，网络数据要么不可获取、不完整，要么测量误差很大。我们证明，因果消息传递利用结果数据中的时间结构而非网络拓扑，可以恢复与网络感知方法相当的总处理效应。我们将因果消息传递应用于两项大规模现场实验，并以一种近期开发的、需要网络知识的二部图方法作为基准。尽管无法获取交互网络，因果消息传递在所有指标上的效应估计值方向均与网络感知方法一致，且在主要决策指标上具有相同的统计显著性。我们的发现验证了因果消息传递的基本前提：估计溢出效应时，结果的时间变化可以有效替代对网络的观测。这具有重要的实践意义：面对网络数据收集成本高、数据专有或不可靠的情境，从业者可以转而利用实验数据的时间动态。

1 引言

随机实验是各领域循证决策的基础，涵盖从技术服务到公共卫生项目等诸多领域。对此类实验的经典分析依赖于稳定单元处理值假设（SUTVA），该假设认为每个单位的结果仅取决于其自身的处理分配（Cox, 1958）。然而，在单位通过社会联系、共享资源或竞争动态相互作用的网络环境中，这一假设并不成立。一个单位的结果不仅可能取决于其自身接受的处理，还可能取决于相连单元接受的处理，这种现象称为 网络干扰 或 溢出效应 (Halloran 和 Struchiner, 1995; Sobel, 2006; Hudgens 和 Halloran, 2008)。

针对干扰的方法论应对大致沿着两条路径展开。第一条路径利用对网络结构的明确了解，开发以网络暴露为条件或对其进行调整的估计量（Aronow 和 Samii, 2017; Leung, 2022）。二分实验设计的最新进展体现了这一思路：当处理单元（例如服务提供者）和结果单元（例如客户）构成双边结构时，方法可以利用观测到的交互图来定义暴露映射，并构建干扰感知估计量（Pouget-Abadie 等人, 2019; Doudchenko 等人, 2020; Zigler 和 Papadogeorgou, 2021; Harshaw 等人, 2023; Tan 等人, 2025）。当交互结构可观测时，这些网络感知方法表现出色，但它们需要的数据往往无法获取、具有专有性，或存在较大的测量误差。

第二条路径针对的是网络未知或仅能部分观测的情形。近期研究提出了一些方法，以时间变异或基于设计的变异取代网络知识（Johari 等人, 2022; Cortez 等人, 2022; Munro 等人, 2021）。其中，因果消息传递 （CMP）提供了一个基于严谨原理的框架，利用了一项关键洞见：支配处理效应在网络中传播方式的机制随时间呈现结构不变性（Shirani and Bayati, 2024; Shirani 等, 2025; Shiraniand Bayati, 2025）。通过观察结果随时间段的变化，CMP 无需观测个体网络连接，便能学习干扰的总体动态。

* 亚马逊 † 斯坦福大学

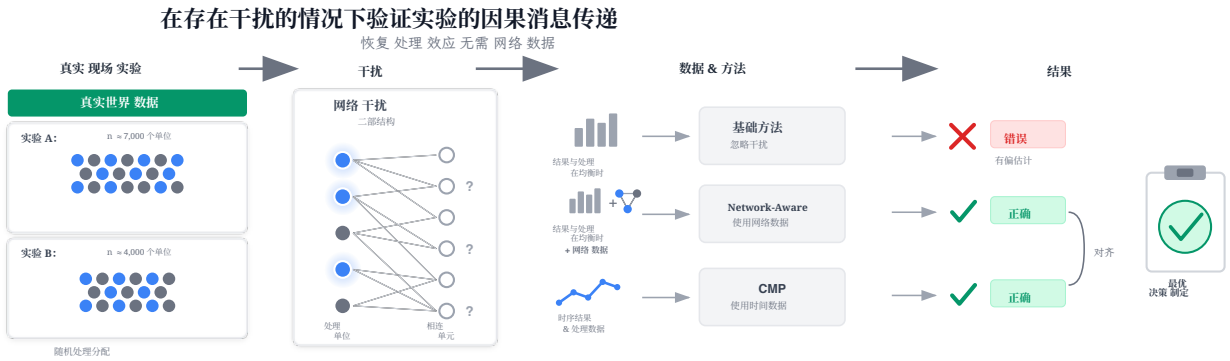

图1：我们的实证验证框架概览。我们将三种估计方法应用于两项真实现场实验（实验A： $n \approx 7{,}000$；实验B： $n \approx 4{,}000$），这些实验在二部图设置中开展，处理单元与相连单元相互作用，从而产生网络干扰。该 基本方法 忽略干扰并产生有偏估计。该 *Network-Aware* 方法利用观测到的网络数据来校正溢出效应。因果消息传递（*CMP*） 仅使用结果的时间动态，而不使用任何网络信息，但其估计值在主要决策指标上与Network-Aware在方向和显著性上均一致。

本文通过将 CMP 与一种网络感知基准进行实证验证，弥合了这两种方法路径之间的差距。我们将 CMP 应用于两个大型现场实验；这两个实验此前曾采用Tan 等人 （2025）的二部图方法进行分析，该方法明确纳入了处理单元与结果单元之间的交互结构。我们的主要发现是，尽管 CMP 不使用任何网络信息，却能得出与网络感知方法的结果高度一致的总处理效应估计，并正确恢复领域知识所预测的干扰偏差方向。

1.1 贡献

我们的贡献如下：

1. *CMP*的实证验证。我们证明，CMP能够在真实实验数据上恢复与网络感知方法相当的总处理效应。这提供了这些方法在现实世界现场实验上的首次直接比较。2. 无网络数据的偏差校正。 我们表明，对于经济学理论可提供方向性预测的指标，CMP能够正确识别干扰偏差的方向。与网络感知基准一样，CMP扭转了由朴素均值差估计得出的误导性结论。3. 实践指南。 我们的结果表明，面临网络数据成本高昂、专有或不可靠情形的从业者可以将CMP作为一种可行的替代方案。实验结果的时间动态包含足以针对干扰效应进行调整的信息。

1.2 相关工作

理解和估计处理与干预的效应，是许多领域的一项基础问题，近年来受到越来越多的关注（Halloran和Struchiner，1995；Sobel，2006；Hudgens和Halloran，2008；Tchetgen Tchetgen和VanderWeele，2012；Manski，2013；Aronow和Samii，2017；Saveskiet al.，2017；Athey等人，2018；Viviano，2020；Hu etal.，2022；Han et al.，2022；Leung，2022；Agarwal等人，2022；Belloni等人，2022；Li和Wager，2022a、b；Farias等人，2022；Candogan等人，2024；Ni等人，2023；Viviano等人，2023；Harshaw等人，2023；Imbens，2024；Eichhorn等人，2024；Johari等人，2024；Larsen等人，2024；Peng等人，2025）。

当前研究将干扰文献的多个方向联系起来，涵盖从网络感知设计，到网络结构未知的情境，再到市场实验。具体而言，Ugander等人(2013) 和 Eckles 等人(2017) 已提出网络感知实验设计，而二部干扰情境则在（Pouget-Abadie 等人，2019；Doudchenko 等人，2020；Zigler 和 Papadogeorgou，2021；Brennan 等人，2022；Harshaw 等人，2023；Tan 等人，2025）中受到广泛关注。

在服务实验的背景下，越来越多的研究考察了多种经济环境与实验设计中的干扰效应（Blake 和 Coey，2014；Holtz 等人，2020；Wager 和 Xu，2021；Munro 等人，2021；Johari 等人，2022；Li 等人，2022；Bajari 等人，2023；Farias 等人，2023；Zhu 等人，2024；Wu 等人，2024；Bright 等人，2025；Zhang 等人，2025；Johari 等人，2025）。利用时间观测与结构来应对干扰也已被证明颇具前景（Farias等人，2022；Hu 和 Wager，2022；Bojinov 等人，2023；Li 等人，2023；Ni等人，2023；Han 等人，2024；Jia 等人，2025）。

对于网络结构未知的情形，现有方法包括平均场型方法（Wager 和 Xu，2021；Johari 等人，2022；Munro 等人，2021）、基于设计的实验策略（Cortez 等人，2022；Hu 和 Wager，2022），以及无需明确网络信息的总效应估计量（Yu 等人，2022）。本文采用了这一研究方向中近期提出的一种方法，尤其是 Shirani and Bayati （ 2024 ）提出并由 Bayati 等人（ 2024 ）扩展至更高阶动态情形的因果消息传递（CMP）框架。尽管 CMP 的这些早期版本在有足够时间变异或横截面变异时表现良好，但在短期实验或结果—处理动态复杂的环境中，其有效性可能受限。为应对这一挑战，Shirani 等人（2025）提出了一种保分布网络自助法，可从单次实验中构建多个代表性子群体，从而增加估计的有效样本量；本文采用了这一方法。近期，Shirani and Bayati（2025）则从暴露映射视角，为基于演化的方法提供了公理化基础。本文的贡献在于使用真实实验数据，对二部网络感知设计与网络盲因果消息传递进行直接实证比较。

2 方法概述

本节概述我们比较的两种方法：网络感知二部方法和因果消息传递。不过，详细描述请参阅原始论文。具体而言，请参阅（Tan 等人，2025）以及（Shirani and Bayati，2024；Shirani 等，2025）。两者针对相同的估计目标，即全面部署下的总处理效应，但在信息需求方面存在根本差异。

2.1 估计目标：总处理效应

考虑一个包含 N 个单位的实验，这些单位的索引为 $i = 1, \ldots\ldots, N$，并在时间段 $t = 0$、1..... T 内进行观测。令 $W_t^i \in \{0,1\}$ 表示单位 i 在时间 t 的处理分配，并令 $Y_t^i(\mathbf{W})$ 表示单位 i 在时间 t 、采用完整处理分配矩阵 $\mathbf{W}$ 时的潜在结果，其中 $\mathbf{W}$ 是一个 $N \times T$ 矩阵，其条目 $(i,\ t)$ 为 W_t^i。

开展实验时，从业者需要决定处理分配如何随时间变化。一种方法是在整个实验期间固定每个单位的处理状态：实验开始时将单位分配至处理组或对照组，实验人员则观察结果，直至其稳定在反映处理完整效应的均衡状态。另一种方法允许处理分配随时间变化，单位可能在不同时间段之间切换处理状态和对照状态。后一类中一种尤为常见的设计是分阶段推广：所有单位最初都处于对照状态，在实验期间的某个时点，单位转为处理状态，并从此持续接受处理。本文主要关注符合分阶段推出设计的情形。因此，结果和处理自然都表示为 $N \times T$ 矩阵，以捕捉每个单位状态和响应随时间的变化。

总处理效应 (TTE)比较完全处理下的平均结果与

无处理下的结果：

$$\mathrm{TTE} = \frac{1}{N}\sum_{i=1}^{N}\left[\mathbb{E}[Y_T^i(\mathbf{1})] - \mathbb{E}[Y_T^i(\mathbf{0})]\right], \tag{1}$$

其中，$\mathbf{1}$ 表示全处理分配，$\mathbf{0}$ 表示所有时间段的全对照分配。该估计目标同时涵盖处理对已处理单元的直接效应，以及通过网络并随时间传播的间接溢出效应。

在二部图情境中，处理施加于图的一侧的单元（例如服务提供者），而这些处理单元会与另一侧的单元互动；本文将后者称为相连单元。我们指出，Tan 等人(2025)，以及更广泛的二分实验文献，将这些相连单元称为“结果单元”；我们刻意避免使用这一术语，以免产生混淆，原因将在下文解释。在二分实验设计中，结果通常在边级上观测，也就是说，在处理单元与相连单元的每次互动中观测。例如，服务提供者为用户提供服务时，可以记录该次互动的结果，例如提供的服务数量。有时，这些边级结果可以聚合到二部图的任一侧；（Tan 等人（Tan 等人，2025）将此称为线性可加边假设。可以将每个处理单元与其服务的所有相连单元之间的边结果求和，从而聚合到处理侧；也可以将每个相连单元与为其提供服务的所有处理单元之间的边结果求和，从而聚合到相连单元侧。本文关注聚合到处理侧的结果：每当我们提及结果 Y_t^i时，指的都是处理单元 i在时间段 t的聚合结果，其计算方式是将该处理单元 i 与之互动的所有相连单元的边级结果相加。正因采用了这一聚合约定，我们避免将右侧单元称为“结果单元”，因为这样会混淆单元与结果本身，并模糊聚合结构。

关于哪些处理单元符合干预条件，还存在一个进一步的区分。Tan 等人(2025)强调，处理单元通常分为两类：符合处理条件的单元，以及因运营或设计限制而不符合条件的单元。他们的主要贡献之一是将这一结构形式化，并提出了两个相应的估计目标。主要总处理效应(PTTE)衡量符合条件的处理单元接受干预时所受到的影响，而次要总处理效应(STTE)则刻画不符合条件的处理单元的结果可能如何受到溢出效应的影响，尽管这些单元本身无法接受处理。本文只关注符合条件的单元：我们的索引$i = 1, \ldots, N$仅指符合干预条件的处理单元（但我们允许存在其他不符合条件的单元）。因此，我们研究的总处理效应对应于 Tan 等人(2025)所称的 PTTE，这也正是他们在真实实验分析中报告的估计目标。我们采取同样的关注重点。探讨如何使用因果消息传递等时间方法估计 STTE，是一个有趣的研究方向，但超出了本文的范围。

2.2 Network-Aware基准：二部图方法论

Tan 等人提出的网络感知方法 (2025)并不直接处理时间数据，而是使用单位层面跨时间聚合的结果——具体而言，即每个单位处理后与处理前的平均结果之差。这种聚合结果结构正是我们所说的基于稳定单元处理值假设的“基本方法”的输入。处理状态由一个指示变量表示 $W^{(j)} \in \{0, 1\}$，用于每个处理单元 j，表示该单位被分配到处理组还是对照组。

该方法利用连接处理单元与相连单元的观测二部图。对于每个处理单元 j，该方法定义了两个暴露变量。直接暴露 用于描述单位 j 自身是否接受了处理，以及它在图中是否存在连接：

$$E_j^{\mathrm{Dir}}(\mathbf{W}) = W^{(j)} \cdot |\mathcal{C}_j|, \tag{2}$$

其中，$\mathcal{C}_j$ 表示与处理单元 j相连的单元集合。间接暴露反映其他处理单元通过相连单元产生的溢出效应：

$$E_j^{\mathrm{Ind}}(\mathbf{W}) = \sum_{c \in \mathcal{C}_j}\sum_{k \neq j} W^{(k)} \cdot \mathbb{I}(k \in \mathcal{T}_c), \tag{3}$$

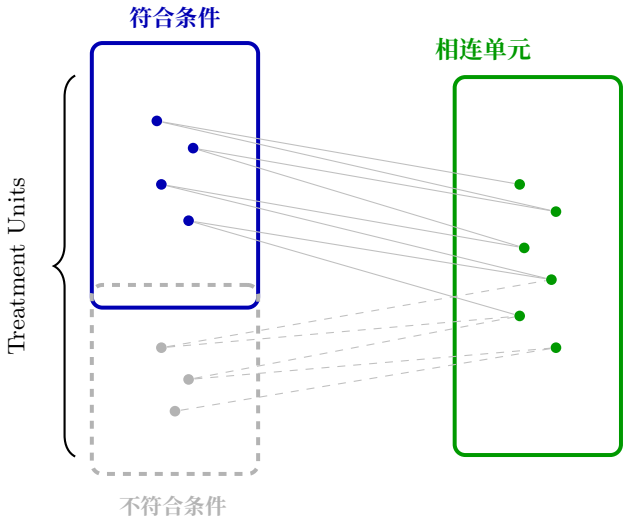

图2：实验设置中的二部结构。处理单元（左侧）分为符合条件的单元（可分配接受处理）和不符合条件的单元（在整个实验过程中始终处于对照状态）。相连单元（右侧）通过代表服务关系的边与处理单元互动。结果通过对边级结果求和，在处理单位层级进行聚合。实线边连接符合条件的处理单元与相连单元；虚线边连接不符合条件的处理单元。本文的分析聚焦于符合条件的处理单元，索引为 $i = 1$、.....、 N。

其中， $\mathcal{T}_c$ 表示与相连单元 c相连的处理单元集合， $\mathbb{I}(\cdot)$是指示函数。处理单元 j 的结果建模为这些暴露的函数：

$$Y_j = \Psi(E_j^{\mathrm{Dir}}, E_j^{\mathrm{Ind}}, X_j) + \epsilon_j, \tag{4}$$

其中， X_j 是协变量， Ψ 通过核岭回归等机器学习方法进行估计。随后，通过比较完全处理与无处理情况下的预测结果来估计 PTTE：

$$\widehat{\mathrm{PTTE}} = \frac{1}{|\mathcal{T}|} \sum_{j \in \mathcal{T}} \left[\hat{\Psi}(E_j^{\mathrm{Dir}}(\mathbf{1}), E_j^{\mathrm{Ind}}(\mathbf{1}), X_j) - \hat{\Psi}(0, 0, X_j) \right], \tag{5}$$

其中， $\mathcal{T}$ 表示符合条件的处理单元集合。该方法的一个关键要求是能够观测到连接处理单元与相连单元的二分图结构。

2.3 Causal 消息传递：无网络数据的时间动态

因果消息传递（CMP）采用了一种根本不同的方法。CMP 不以网络结构为条件，而是利用干扰机制的时间不变性。其核心洞见是，处理效应会根据随时间保持结构稳定的规则在网络中传播，这类似于相互作用粒子系统中的不变物理定律（Shirani and Bayati，2024；Shirani 等，2025）。

CMP 对 总体层面结果分布通过状态演化方程的演化 ：

$$Y_{t+1} \overset{d}{=} f_t(Y_t, W_{t+1}), \tag{6}$$

其中， Y_t 和 W_{t+1} 分别是一维随机变量，表示时间 t 时结果和处理的极限分布，而 f_t 刻画了结果如何演化的总体动态。映射 f_t 同时纳入了直接处理效应和间接溢出效应，而无需了解个体网络连接。

估计程序分为三个步骤：

1. 特征构造： 计算每个时间段的结果和处理分布的汇总统计量（均值、高阶矩、交互项）。 2. 状态演化估计： 利用监督学习，根据结果和处理的观测时间变化估计映射 f_t 。 3. 反事实预测： 递归应用估计出的映射，预测不同处理情景下的结果（例如，完全处理与无处理）。

该t otal处理效应通过比较预测的反事实演化来估计 s:

$$\widehat{\mathrm{TTE}}^{\mathrm{CMP}} = \widehat{\mathrm{CFE}}_T(\mathbf{1}) - \widehat{\mathrm{CFE}}_T(\mathbf{0}), \tag{7}$$

其中$\widehat{\mathrm{CFE}}_T(\mathbf{w})$表示在处理分配$\mathbf{w}$下、最终时间段的估计反事实演化。

CMP的关键优势在于，它只需要随时间观测到的结果，而无需了解任何网络数据（即网络感知方法所使用的 ω_{ij}）。

2.4 方法比较

在下一节中，我们将Network-Aware与CMP方法同一种更简单的基线进行比较，我们称之为基本方法。该基本方法假设稳定单元处理值假设成立，并完全忽略干扰。如上所述，网络感知方法作用于单元层面的聚合结果，具体而言是处理后时期与处理前时期之间的一阶差分。基本方法使用相同的聚合结果结构，将时间序列折叠为每个单元的单一前后差分，但采用假设稳定单元处理值假设成立的标准双重差分估计量的机器学习增强版本。通过在估计之前对时间进行聚合，基本方法实际上舍弃了CMP用以厘清干扰的中间时间动态。网络感知方法可视为基本方法的更精细版本，即通过暴露映射纳入二分图结构。相比之下，CMP采取一种根本不同的方法，直接使用时间结果数据，并学习处理效应如何随时间传播的动态。

表 1 总结了这三种方法之间的主要差异。

表 1：基础方法、Network-Aware 与 CMP 方法比较

特征	**基础方法**	**Network-Aware**	**CMP**
假设稳定单元处理值假设	Yes	No	No
需要网络数据	No	Yes	No
需要时间数据	No	No	Yes
结果结构	聚合	聚合	时间序列
暴露定义	None	基于图	基于时间
估计策略	双重差分 + 机器学习	倾向得分 + 机器学习	状态演化 + 机器学习
识别来源	仅随机化	网络结构	时间不变性

3 实验设置

我们将所有方法应用于在服务环境中开展的两项大规模现场实验，其中处理单元通过二部结构与相连单元互动。由于数据提供方的保密限制，我们只能对这些实验作一般性描述。

3.1 实验背景

两项实验都涉及对处理侧单元的一个子集实施干预（例如，网约车服务中的司机）。二部结构自然源于交互模式：相连单元可能连接多个处理单元，而处理单元则服务多个相连单元。

重要的是，由于运营限制，只有部分处理单元符合干预资格。这形成了一个资格约束场景：随机化在符合资格的集合内进行，但不符合资格的单元仍会与结果单元交互，并可能产生溢出效应。

两种实验设置中的二部交互图都较为稀疏，处理单元的平均度相对于总体规模而言较小。这种稀疏性给网络感知估计带来了不小的挑战，因为必须依据每个单元相对较少的交互来学习暴露—响应函数。

3.2 实验 A 和实验 B

实验A涉及约 7,000 个符合条件的处理单元，持续多个（$T < 100$）时间段，各实验阶段的处理概率有所不同。该干预改变了一个影响处理单元与相连单元交互方式的特征。

实验B涉及约 4,000 个符合条件的处理单元，采用了类似的多阶段设计。该干预针对处理单元与相连单元之间交互的一个相关但不同的方面。

两项实验均在所有时间段内收集了处理单位层级的结果数据，从而能够应用网络感知估计和 CMP 方法。

3.3 指标

我们评估三个预先指定的指标，分别记为 M_1、M_2 和 M_3：

- M_1：根据经济学理论和领域知识，对于实验 A 和实验 B，如果忽略溢出效应，预计会导致估计值偏高。这提供了一项事前方向性检验。
- M_2：一项次要指标，对偏差的方向没有明确预测。
- M_3：用于评估干预是否成功的主要决策指标 。该指标决定是否会大规模推出这项干预。

3.4 实施

我们实施了 Tan 等人 (2025) 提出的二部图方法，利用观测到的交互图构建暴露变量和广义倾向得分。我们使用经参数自动调优的 LightGBM 估计结果函数，并通过自助法推断计算置信区间。我们还实施了 Shirani 等 (2025) 提出的 CMP 框架，计算每个时间段的结果和处理汇总统计量。我们使用带交叉验证的岭回归估计状态演化参数，并应用保分布网络自助法生成训练样本。我们还实施了上述忽略干扰的基础方法。

4 结果

表 2展示了我们的主要结果，比较了基础方法（朴素）、Network-Aware和CMP方法在两项实验及全部三项指标上的表现。

4.1 CMP恢复Network-Aware估计值

最显著的发现是，CMP与Network-Aware估计值在两项实验及全部指标上高度一致。在所有情况下，CMP给出的效应估计值与网络感知方法具有相同的方向符号 。这种一致性尤为值得注意，因为CMP在运行时并不了解网络感知方法所明确利用的二部交互结构。

对于实验A，Network-Aware与CMP均发现：

- 对 M_1 和 M_2产生负效应，逆转了基础方法得出的正向估计值

表 2：基本方法、网络感知方法和因果消息传递在两个真实实验中的比较。对于每个指标，我们报告处理效应估计和 5% 水平下的统计显著性。偏差列展示基本方法与网络感知方法之间观察到的差异，以及根据经济学理论预期的偏差方向。 M_3 是主要决策指标。

	指标	基本方法		Network-Aware		CMP		Bias	
		ATE	Sig.	PTTE	Sig.	效应	Sig.	观察值	预期值
实验A									
	M_1	Pos.	Yes	Neg.	No	Neg.	Yes	Pos.	Pos.
	M_2	Pos.	No	Neg.	No	Neg.	Yes	Pos.	—
	M_3	Neg.	No	Pos.	Yes	Pos.	Yes	Neg.	—
实验B									
	M_1	Pos.	No	Pos.	No	Pos.	Yes	Pos.	Pos.
	M_2	Pos.	Yes	Pos.	No	Pos.	No	Pos.	—
	M_3	Pos.	No	Neg.	No	Neg.	No	Pos.	—

- 对主要决策指标 M_3产生正向且具有统计显著性的效应

对于实验B，两种方法均发现：

- 对 M_1 和 M_2的正向效应
- 对 M_3的负向（不显著）效应

在指标 M_1 上，CMP 达到了统计显著性，而 Network-Aware 未达到。一个可能的解释是数据利用方面的差异：Network-Aware 方法将时间维度压缩为每个单位的单一前后差值，而 CMP 则利用了结果的完整时间序列。然而，这两种方法在许多其他方面也存在差异，包括建模假设、估计程序以及识别变异的来源，因此很难将显著性差异归因于某一个单独因素。对这些存在根本差异的方法进行统计效率的严格比较超出了本文的范围。

虽然在主要决策指标 M_3 上，CMP 与 Network-Aware 完全一致——在两项实验中，二者的效应方向和统计显著性均相同——但在指标 M_1 的统计显著性方面，两种方法存在差异：CMP 在两项实验中均发现了显著效应，而 Network-Aware 则没有。究竟这反映了统计效率的真实差异、目标估计量的差异，还是其他方法因素，仅凭我们的比较无法确定。我们能够得出的结论是，在所研究的所有指标上，时间动态提供了足够的信号，使 CMP 得以与网络感知方法在方向性结论上保持一致。

4.2 正确恢复偏差方向

对于指标 M_1，经济学理论预测，忽略溢出效应会使估计值产生特定方向的偏差：基础方法得出的估计值应比真实的总处理效应更偏正。两项实验都证实了这一预测。重要的是，CMP 在该指标上的显著发现与理论预期一致。

在实验A中，基础方法显示出正向且显著的效应，而 Network-Aware 和 CMP 显示的都是负向效应。(基础方法 –Network-Aware) 和 (基础方法 –CMP) 的差值均为正，与预期的偏差方向一致。CMP 发现了显著的负向效应，而 Network-Aware 得出的负向效应不显著，这与关于该指标的领域知识相符。如果理论预测的真实效应确实为负，这一模式可能表明 CMP 的检验效能更高，能够检测到该效应，不过还需要进一步研究才能验证这一假设。

在实验B中，尽管所有方法都显示出正效应，但观察到的偏差（基础方法比干扰感知估计值更偏正）仍然符合预期方向。同样，CMP在 Network-Aware 未发现显著性的情况下发现了显著性，

如果真实效应为正，这将与 CMP 具有更高的敏感性相一致，不过也不能排除其他解释。

这一模式表明，与 Network-Aware 方法一样，CMP 能够正确识别并调整因忽略溢出效应而产生的偏差。对于指标 M_1，领域知识能够给出明确的方向性预测，CMP 得出的显著结果与理论预期一致。在这些实验设置中，CMP 提取的时间信息似乎可以可靠地替代网络信息，不过，要理解基于时间的方法与基于网络的方法之间的具体权衡，仍是未来研究的重要方向。

4.3 决策相关影响

主要决策指标 M_3 揭示了考虑干扰在实践中的重要性。

在实验A中：基础方法估计处理效应为负，但未达到统计显著性。相比之下，Network-Aware 和 CMP 均发现了正向且具有统计显著性的效应。这一反转意义重大： M_3 的统计显著提升意味着主要决策指标恶化，而基础方法无法发现这一点。干扰感知方法清晰地揭示了朴素分析所掩盖的信号。重要的是，在这一对决策至关重要的指标上，CMP 和 Network-Aware 的结论完全一致：方向相同，显著性结论也相同。

在实验B中，Network-Aware 和 CMP 对 M_3 的效应估计均为负，但未达到统计显著性；而基础方法估计的效应为正，同样未达到统计显著性。尽管方向上存在分歧，但所有方法都得出了相同的实际结论：未检测到统计显著的效应。考虑干扰的方法再次得出了相同结论。

在两个实验中，CMP 与 Network-Aware 对 M_3 的结论一致，这是本文的关键验证结果。对于决定是否应大规模推行某项干预的指标，这两种方法——一种需要网络数据，另一种只需要时间数据——得出了完全相同的结论。

5 讨论与结论

CMP 与 Network-Aware 的估计值高度一致，尤其是在主要决策指标 M_3 上完全一致，这支持了因果消息传递的理论前提：结果的时间变化蕴含着有关干扰结构的信息。当处理效应沿网络传播时，结果分布会随时间变化，并呈现可预测的模式。通过学习这些模式，CMP 无需观测网络本身，就能有效重建网络干扰的总体效应。

在指标 M_1 上，CMP 达到了统计显著性，而 Network-Aware 则没有。一个合理的解释是，CMP 使用所有时间段的结果数据，有效扩大了可用于估计的样本量；但这两种方法在太多方面存在差异，因此无法对其相对效率得出明确结论。理解网络信息与时间信息之间的权衡，以及在哪些条件下一方优于另一方，是我们的实证比较提出但无法解答的重要问题。我们的结果确实表明，从业者不必将缺少网络数据视为干扰校正的根本障碍。

我们推测，当干扰机制随时间相对稳定、有足够的时间变化来识别状态演化映射，且网络结构产生的干扰模式能在结果动态中清晰显现时，就会出现观察到的一致性。未来的研究应刻画这些方法可能产生分歧的条件，并为从业者开发诊断方法。

有几项局限性值得承认。我们的验证基于两项实验，将结果推广到其他领域还需要进一步研究。两种方法都无法给出“真实”的处理效应；我们将这些方法相互比较，并与理论给出的方向性预测进行比较。CMP 需要多期结果数据，并假设干扰机制相对稳定；网络结构的快速变化可能会违背这些假设。Shirani and Bayati (2025) 提供了一种公理化处理，有助于更好地理解此类局限性。

对于面临网络数据缺失、获取成本高昂或不可靠等情况的从业者而言，CMP 提供了一种有原则的方法，能够正确判断偏差方向，并在实验A中推翻标准 SUTVA 分析可能得出的误导性

结论。更广泛地说，本研究表明，在真实数据上，Network-Aware 方法与网络盲方法在经验结果上趋于一致，这意味着时间动态与网络结构所包含的信息可能比以往认识到的联系更为紧密。

参考文献

Anish Agarwal、Sarah H Cen、Devavrat Shah 和 Christina Lee Yu。网络合成干预：网络干扰下面板数据的因果框架。*arXiv* 预印本 *arXiv:2210.11355*，2022 年。

Peter M. Aronow 和 Cyrus Samii。在一般干扰下估计平均因果效应，并应用于社交网络实验。《应用统计年鉴》，11(4):1912–1947，2017 年。doi：10.1214/16-AOAS1005。

Susan Athey、Dean Eckles 和 Guido W. Imbens。网络干扰的精确 p 值。《美国统计学会杂志》，113(521):230–240，2018 年。doi：10.1080/01621459.2016.1241178。

Patrick Bajari、Brian Burdick、Guido W. Imbens、Lorenzo Masoero、James McQueen、Thomas S. Richardson 和 Ido M. Rosen。市场平台中的实验设计。《统计科学》，38(3):458 – 476，2023 年。doi：10.1214/23-STS883。网址：https://doi.org/10.1214/23-STS883。

MohsenBayati、YuweiLuo、WilliamOverman、MohamadSadeghShiraniFaradonbeh 和 RuoxuanXiong。用于复杂干扰实验的高阶因果消息传递。《神经信息处理系统进展》，37：81836–81856，2024。

Alexandre Belloni、FeiFang 和 Alexander Volfovsky。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*，2022。

Thomas Blake 和 Dominic Coey。为什么市场平台实验比看起来更难：测试组与对照组干扰的作用。载于第十五届 *ACM*经济学与计算会议论文集，第 567–582 页，2014。

Iavor Bojinov、David Simchi-Levi 和 Jinglong Zhao。切换实验的设计与分析。《管理科学》，69(7)：3759–3777，2023。

Jennifer Brennan、Vahab Mirrokni 和 Jean Pouget-Abadie。单侧二部实验的簇随机设计。载于 S. Koyejo、S. Mohamed、A. Agarwal、D. Belgrave、K. Cho 和 A. Oh 编，《神经信息处理系统进展》，第 35 卷，第 37962–37974 页。Curran Associates, Inc.，2022 年。网址https://proceedings.neurips.cc/paper_files/paper/2022/file/f7f043c3438ba9e385c51bcf50ed007e-Paper-Conference.pdf。

Ido Bright、Arthur Delarue 和 Ilan Lobel。通过影子价格减少市场干扰偏差。《管理科学》，71(8):7094–7112，2025。

Ozan Candogan、Chen Chen 和 Rad Niazadeh。相关聚类随机实验：稳健方差最小化。《管理科学》，70(6):4069–4086，2024。

Mayleen Cortez、Matthew Eichhorn 和 Christina Yu。分阶段推出设计无需网络知识，即可在存在干扰的情况下进行因果推断。载于 S. Koyejo、S. Mohamed、A. Agarwal、D. Belgrave、K. Cho 和 A. Oh 编，《神经信息处理系统进展》，第 35 卷，第 7437–7449 页。Curran Associates, Inc.，2022。网址：https://proceedings.neurips.cc/paper_files/paper/2022/file/3103b258537198475025 59bf67eb4037-Paper-Conference.pdf。

D. R. Cox。《实验设计》。Wiley，纽约，，1958年。

Nick Doudchenko、Minzhengxiong Zhang、Evgeni Drynkin、Edoardo Airoldi、Vahab Mirrokni 和 Jean Pouget-Abadie。《基于二部设计的因果推断》。*arXiv* 预印本 *arXiv:2010.02108*，2020。

Dean Eckles、Brian Karrer 和 Johan Ugander。《网络实验的设计与分析：减少干扰导致的偏差》。《因果推断杂志》，5(1):20150021，2017。doi: 10.1515/jci-2015-0021。Matthew Eichhorn、Samir Khan、Johan Ugander 和 Christina Lee Yu。《低阶结果与聚类设计：结合设计和分析以实现网络干扰下的因果推断》。arXiv 预印本 arXiv:2405.07979，2024。V Farias、A Li、T Peng 和 A Zheng。《实验中的马尔可夫干扰》。《神经信息处理系统进展》，35:535–549，2022。Vivek Farias、Hao Li、Tianyi Peng、Xinyuyang Ren、Huawei Zhang 和 Andrew Zheng。《实验中的干扰校正：抖音案例研究》。载于第 17 届 ACM推荐系统大会论文集，第 455–466 页，2023。M. Elizabeth Halloran 和 Claudio J. Struchiner。《传染病中的因果推断》。《流行病学》，6 (2):142–151，1995。doi: 10.1097/00001648-199503000-00010。Kevin Han、Shuangning Li、Jialiang Mao 和 Han Wu。《通过提高分配比例检测 A/B测试中的干扰》。arXiv 预印本 arXiv:2211.03262，2022。Kevin Han、Guillaume Basse 和 Iavor Bojinov。《面板实验中的总体干扰》。《计量经济学杂志》，238(1):105565，2024。Christopher Harshaw、Fredrik S¨avje、David Eisenstat、Vahab Mirrokni 和 Jean Pouget-Abadie。《线性暴露—响应模型下二部实验的设计与分析》。《电子统计学杂志》，17(1):464–518，2023。D Holtz、R Lobel、I Liskovich 和 S Aral。《减少在线市场定价实验中的干扰偏差》。arXiv 预印本 arXiv:2004.12489，2020。Yuchen Hu 和 Stefan Wager。《几何混合下的切换实验》。arXiv 预印本 arXiv:2209.00197，2022。Yuchen Hu、Shuangning Li 和 Stefan Wager。《干扰下的平均直接与间接因果效应》。《生物计量学》，109(4):1165–1172，2022。Michael G Hudgens 和 M Elizabeth Halloran。《迈向存在干扰时的因果推断》。《美国统计学会杂志》，103(482):832–842，2008。doi: 10.1198/016214508000000292。Guido W Imbens。《社会科学中的因果推断》。《统计学及其应用年度评论》，11，2024。Zhiwei Jia、Nathan Kallus 和 Zhiwei Steven Yu。《时空干扰下实验的聚类切换设计》。arXiv 预印本 arXiv:2312.15574，2025。URL https://arxiv.org/abs/ 2312.15574。Ramesh Johari、Hannah Li、Inessa Liskovich 和 Gabriel Y Weintraub。《双边平台上的实验设计：偏差分析》。《管理科学》，68(10):7069–7089，2022。Ramesh Johari、Hannah Li、Anushka Murthy 和 Gabriel Y Weintraub。《干扰何时重要？平台实验中的决策》。arXiv 预印本 arXiv:2410.06580，2024。Ramesh Johari、Tianyi Peng 和 Wenqian Xing。《通过截断策略梯度估计量估计非平稳条件下的处理效应》，2025。URL https://arxiv.org/abs/2506.05308。Nicholas Larsen、Jonathan Stallrich、Srijan Sengupta、Alex Deng、Ron Kohavi 和 Nathaniel T. Stevens。《在线对照实验中的统计挑战：A/B测试方法综述》。《美国统计学家》，78(2):135–149，2024。doi: 10.1080/00031305.2023.2257237。URL https://doi.org/10.1080/ 00031305.2023.2257237。

Michael P. Leung. 近似邻域干扰下的因果推断。《计量经济学》，90(1): 267–293, 2022. doi: 10.3982/ECTA17841. Hannah Li, Geng Zhao, Ramesh Johari, and Gabriel Y. Weintraub. 双边市场平台实验中的干扰、偏差与方差：平台指南。载于*2022*年 *ACM Web* 会议论文集，WWW '22, page 182–192, New York, NY, USA, 2022. 计算机协会。ISBN 9781450390965. doi: 10.1145/3485447.3512063. URLhttps://doi.org/10.1145/3485447.3512063。Shuangning Li and Stefan Wager. 微随机试验中的网络干扰。*arXiv* 预印本 *arXiv:2202.05356*，2022a. Shuangning Li and Stefan Wager. 网络干扰下处理效应估计的随机图渐近理论，2022b. URLhttps://arxiv.org/abs/2007.13302。Shuangning Li, Ramesh Johari, Xu Kuang, and Stefan Wager. 随机拥塞下的实验。*arXiv* 预印本 *arXiv:2302.12093*, 2023. Charles F Manski. 社会互动下处理响应的识别。《计量经济学期刊》，16(1):S1–S23, 2013. Evan Munro, Stefan Wager, and Kuang Xu. 市场均衡中的处理效应。*arXiv* 预印本 *arXiv:2109.11647*, 2021. Tu Ni, Iavor Bojinov, and Jinglong Zhao. 存在空间和时间干扰的面板实验设计。可见于 *SSRN 4466598*，2023. Tianyi Peng, Naimeng Ye, and Andrew Zheng. 实验中网络干扰的邻居差分法，2025. URLhttps://arxiv.org/abs/2503.02271。Jean Pouget-Abadie, Kevin Aydin, Warren Schudy, Kay Brodersen, and Vahab Mirrokni. 通过相关聚类降低二部实验中的方差。载于《神经信息处理系统进展》, pages 13309–13319, 2019. Martin Saveski, Jean Pouget-Abadie, Guillaume Saint-Jacques, Weitao Duan, Souvik Ghosh, Ya Xu, and Edoardo M Airoldi. 检测网络效应：对随机化实验进行再随机化。载于第*23*届 *ACM SIGKDD* 知识发现与数据挖掘国际会议论文集，pages 1027– 1035, 2017. Sadegh Shirani and Mohsen Bayati. 未知的一般网络干扰下实验的因果消息传递。《美国国家科学院院刊》，121(40):e2322232121, 2024. Sadegh Shirani and Mohsen Bayati. 干扰下实验的基于演化的模型。*arXiv* 预印本 *arXiv:2511.21675*，2025. Sadegh Shirani, Yuwei Luo, William Overman, Ruoxuan Xiong, and Mohsen Bayati. 一般网络干扰下能否验证反事实估计？*arXiv* 电子版，文章编号 arXiv:2502.01106, 2025年2月。doi: 10.48550/arXiv.2502.01106. Michael E Sobel. 住房流动随机研究揭示了什么？《美国统计学会杂志》，101(476):1398–1407, 2006. doi: 10.1198/016214506000000636. Albert Tan, Mohsen Bayati, James Nordlund, and Roman Istomin. 估计存在溢出效应和部分符合资格的二部实验中的总效应。*arXiv* 预印本 *arXiv:2511.11564*，2025. Eric J. Tchetgen Tchetgen and Tyler J. VanderWeele. 干扰存在时的因果推断。《医学研究统计方法》，21(1):55–75, 2012. doi: 10.1177/0962280210386779. Johan Ugander, Brian Karrer, Lars Backstrom, and Jon Kleinberg. 图聚类随机化：多个总体的网络暴露。载于第*19*届 *ACM SIGKDD* 知识发现与数据挖掘国际会议论文集, pages 329–337. ACM, 2013. doi: 10.1145/2487575.2487695.

Davide Viviano. 网络干扰下的实验设计。*arXiv* 预印本 *arXiv:2003.08421*，2020. Davide Viviano, Lihua Lei, Guido Imbens, Brian Karrer, Okke Schrijvers, and Liang Shi. 因果聚类：网络干扰下集群实验的设计。*arXiv* 预印本 *arXiv:2310.14983*，2023. Stefan Wager and Kuang Xu. 均衡中的实验。《管理科学》，67(11):6694–6715, 2021. Yifan Wu, Ramesh Johari, Vasilis Syrgkanis, and Gabriel Y. Weintraub. 前瞻性需求下的切换式价格实验，2024. URLhttps://arxiv.org/abs/2410.14904。Christina Lee Yu, Edoardo M. Airoldi, Christian Borgs, and Jennifer T. Chayes. 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), 2022. ISSN 1091-6490. Chenyu Zhang, Yuhang Wu, Zeyu Zheng, and Nian Si. 双边平台中多项逻辑模型下卖方侧实验的去偏。可见于 *SSRN 5261661*，2025. Zhihua Zhu, Zheng Cai, Liang Zheng, and Nian Si. 双边平台中由反馈回路引发干扰时的卖方侧实验。*arXiv* 预印本 *arXiv:2401.15811*，2024. Christopher M. Zigler and Georgia Papadogeorgou. 存在干扰的二部图因果推断。《统计科学》，36(1):109–123, 2021. doi: 10.1214/19-STS749.