

# Diffusion in Experiments

Ye Wang  
New York University

Peking University, Dec 28 2018

# 个人介绍

- ▶ 2007-2011 复旦大学数学系

# 个人介绍

- ▶ 2007-2011 复旦大学数学系
- ▶ 2011-2014 北京大学中国经济研究中心

# 个人介绍

- ▶ 2007-2011 复旦大学数学系
- ▶ 2011-2014 北京大学中国经济研究中心
- ▶ 2014-2015 威斯康星大学麦迪逊分校经济系博士

# 个人介绍

- ▶ 2007-2011 复旦大学数学系
- ▶ 2011-2014 北京大学中国经济研究中心
- ▶ 2014-2015 威斯康星大学麦迪逊分校经济系博士  
(太冷, 转学了)

# 个人介绍

- ▶ 2007-2011 复旦大学数学系
- ▶ 2011-2014 北京大学中国经济研究中心
- ▶ 2014-2015 威斯康星大学麦迪逊分校经济系博士  
(太冷, 转学了)
- ▶ 2015至今 纽约大学政治系博士

# 个人介绍

- ▶ 2007-2011 复旦大学数学系
- ▶ 2011-2014 北京大学中国经济研究中心
- ▶ 2014-2015 威斯康星大学麦迪逊分校经济系博士  
(太冷, 转学了)
- ▶ 2015至今 纽约大学政治系博士
- ▶ 领域: 政治学研究方法, 当代威权主义和政治转型

# 个人介绍

- ▶ 2007-2011 复旦大学数学系
- ▶ 2011-2014 北京大学中国经济研究中心
- ▶ 2014-2015 威斯康星大学麦迪逊分校经济系博士  
(太冷, 转学了)
- ▶ 2015至今 纽约大学政治系博士
- ▶ 领域: 政治学研究方法, 当代威权主义和政治转型
- ▶ 2014年至今, 政见(CNPolitics)撰稿人

# 概要

- ▶ 为什么要研究实验中的扩散?
- ▶ 实验研究的基本假设
- ▶ 情况一: 固定网络和随机处理
- ▶ 情况二: 随机网络和固定处理

为什么要研究实验中的扩散?

# 为什么要研究实验中的扩散?

两个子问题:

- ▶ 为什么要做实验?
- ▶ 为什么我们关心扩散?

# 为什么要做实验?

- ▶ 实验是当今社会科学中的“黄金标准”
- ▶ 终极目标: 识别因果关系

# 为什么要做实验？

- ▶ 实验是当今社会科学中的“黄金标准”
- ▶ 终极目标：识别因果关系

经济增长到底能不能带来民主？

给非洲人民发帐篷能不能降低疟疾发病率？

教育会不会提升政治参与？

# 为什么要做实验?

- ▶ 理想实验: 将感兴趣的群体随机分成两个各方面都相似的组, 改变其中一组的X, 观察两组的Y是否出现差异

# 为什么要做实验?

- ▶ 理想实验: 将感兴趣的群体随机分成两个各方面都相似的组, 改变其中一组的X, 观察两组的Y是否出现差异
- ▶ 实验的好处: 内部效度(internal validity)非常高, 能够保证得到因果关系

# 为什么要做实验?

- ▶ 理想实验: 将感兴趣的群体随机分成两个各方面都相似的组, 改变其中一组的X, 观察两组的Y是否出现差异
- ▶ 实验的好处: 内部效度(internal validity)非常高, 能够保证得到因果关系
- ▶ 实验方法的风行改变了我们思考研究设计的方式

# 为什么要做实验?

- ▶ 理想实验: 将感兴趣的群体随机分成两个各方面都相似的组, 改变其中一组的X, 观察两组的Y是否出现差异
- ▶ 实验的好处: 内部效度(internal validity)非常高, 能够保证得到因果关系
- ▶ 实验方法的风行改变了我们思考研究设计的方式
- ▶ 相比之下

# 为什么要做实验?

- ▶ 理想实验: 将感兴趣的群体随机分成两个各方面都相似的组, 改变其中一组的X, 观察两组的Y是否出现差异
- ▶ 实验的好处: 内部效度(internal validity)非常高, 能够保证得到因果关系
- ▶ 实验方法的风行改变了我们思考研究设计的方式
- ▶ 相比之下

案例研究: 可以揭示内在机制, 但样本量小, 可信度低

# 为什么要做实验?

- ▶ 理想实验: 将感兴趣的群体随机分成两个各方面都相似的组, 改变其中一组的X, 观察两组的Y是否出现差异
- ▶ 实验的好处: 内部效度(internal validity)非常高, 能够保证得到因果关系
- ▶ 实验方法的风行改变了我们思考研究设计的方式
- ▶ 相比之下

案例研究: 可以揭示内在机制, 但样本量小, 可信度低

观察性研究: 代表性强, 但无法排除混淆变量

# 为什么要做实验?

- ▶ 理想实验: 将感兴趣的群体随机分成两个各方面都相似的组, 改变其中一组的X, 观察两组的Y是否出现差异
- ▶ 实验的好处: 内部效度(internal validity)非常高, 能够保证得到因果关系
- ▶ 实验方法的风行改变了我们思考研究设计的方式
- ▶ 相比之下

案例研究: 可以揭示内在机制, 但样本量小, 可信度低

观察性研究: 代表性强, 但无法排除混淆变量

规范性研究: 需要坚实的因果关系作为基础

# 为什么要做实验?

- ▶ 实验的缺点:

# 为什么要做实验?

- ▶ 实验的缺点: 也很多!

# 为什么要做实验?

- ▶ 实验的缺点: 也很多!
- ▶ 贵, 而且越来越贵

# 为什么要做实验?

- ▶ 实验的缺点: 也很多!
- ▶ 贵, 而且越来越贵
- ▶ 能研究的问题十分有限

# 为什么要做实验?

- ▶ 实验的缺点: 也很多!
- ▶ 贵, 而且越来越贵
- ▶ 能研究的问题十分有限
- ▶ 结论的外部效度(external validity)低

# 为什么要做实验?

- ▶ 实验的缺点: 也很多!
- ▶ 贵, 而且越来越贵
- ▶ 能研究的问题十分有限
- ▶ 结论的外部效度(external validity)低
- ▶ 未必能反映真实的机制

# 为什么要关心扩散?

- ▶ 很多时候干扰研究结果

# 为什么要关心扩散?

- ▶ 很多时候干扰研究结果 (把处理组的药吃了)

# 为什么要关心扩散?

- ▶ 很多时候干扰研究结果 (把处理组的药吃了)
- ▶ 本身也很有意义

# 为什么要关心扩散?

- ▶ 很多时候干扰研究结果 (把处理组的药吃了)
- ▶ 本身也很有意义

打疫苗的正外部性有多强?

吸烟的习惯会不会传染?

跟竞赛生住一个宿舍对GPA有帮助吗?

# 为什么要关心扩散?

- ▶ 很多时候干扰研究结果 (把处理组的药吃了)
- ▶ 本身也很有意义

打疫苗的正外部性有多强?

吸烟的习惯会不会传染?

跟竞赛生住一个宿舍对GPA有帮助吗?

- ▶ 核心问题: 怎么定义扩散? 怎么用实验估计其效应大小?

## 实验研究的基本假设

# Rubin模型

发明人: 隔壁的Donald Rubin教授



# Rubin模型

发明人: 隔壁的Donald Rubin教授



Not!

# Rubin模型

发明人: 隔壁的Donald Rubin教授



Not!

历史可以追溯到Neyman (1923)

# Rubin模型 (基本概念)

- ▶ 潜在结果 (potential outcome):

$$Y_i \text{ (健康状况)} = \begin{cases} Y_i(1) & \text{if } D_i = 1 \text{ (吃了药)} \\ Y_i(0) & \text{if } D_i = 0 \text{ (没吃药)} \end{cases}$$

- ▶ 处理效应 (treatment effect):

$$\tau_i = Y_i(1) - Y_i(0)$$

- ▶ 平均处理效应 (ATE):

$$E[\tau_i] = E[Y_i(1)] - E[Y_i(0)]$$

## “因果推断的基本问题”

对于任意个体 $i$ , 我们不可能同时观察到 $Y_i(1)$ 和 $Y_i(0)$  (Holland, 1986)

## “因果推断的基本问题”

对于任意个体 $i$ , 我们不可能同时观察到 $Y_i(1)$ 和 $Y_i(0)$  (Holland, 1986)  
怎么办?

## “因果推断的基本问题”

对于任意个体 $i$ , 我们不可能同时观察到 $Y_i(1)$ 和 $Y_i(0)$  (Holland, 1986)

怎么办?

- ▶ 科学方案 (施加假设)
- ▶ 统计学方案 (增大样本, 开展实验)

# 统计学方案

假定处理 $D$ 是“外生”的, 独立于潜在结果 $Y_i(1)$ 和 $Y_i(0)$ , 那么

$$\begin{aligned} E[\tau_i] &= E[Y_i(1)] - E[Y_i(0)] \\ &= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] \\ &= E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \end{aligned}$$

在实际操作中,

$$\widehat{ATE} = \frac{1}{N_1} \sum_{D_i=1} Y_i - \frac{1}{N_0} \sum_{D_i=0} Y_i$$

扩散改变了什么？



扩散改变了什么？



## 扩散改变了什么？

潜在因果框架不再成立:  $Y_i = Y_i(1)$  if  $D_i = 0$

# 扩散改变了什么？

潜在因果框架不再成立:  $Y_i = Y_i(1)$  if  $D_i = 0$

不再有“SUTVA” (Stable Unit Treatment Value Assumption)

# 扩散改变了什么？

潜在因果框架不再成立:  $Y_i = Y_i(1)$  if  $D_i = 0$

不再有“SUTVA” (Stable Unit Treatment Value Assumption)

但无需紧张... 我们只需要定义一个新的变量

# 扩散改变了什么？

潜在因果框架不再成立:  $Y_i = Y_i(1)$  if  $D_i = 0$

不再有“SUTVA” (Stable Unit Treatment Value Assumption)

但无需紧张... 我们只需要定义一个新的变量

$W_i$ : 每个个体实际接受的处理, 或者说“treatment exposure”

# 扩散改变了什么?

潜在因果框架不再成立:  $Y_i = Y_i(1)$  if  $D_i = 0$

不再有“SUTVA” (Stable Unit Treatment Value Assumption)

但无需紧张... 我们只需要定义一个新的变量

$W_i$ : 每个个体实际接受的处理, 或者说“treatment exposure”

很多时候  $W_i$  是一个连续变量, 所以我们感兴趣的估计量变成了

$$ATE(w, w') = E[Y_i(w)] - E[Y_i(w')]$$

余下的问题: 1. 怎么测量  $W_i$  (我们只知道  $D_i$ ) ? 2. 独立性假设是否仍然成立 ?

# 双管齐下！

我们需要更强的假设和更精巧的实验设计：

# 双管齐下！

我们需要更强的假设和更精巧的实验设计：

- ▶ 扩散常常局限在一定范围之内 (班级内部)
- ▶ 一般沿着社交网络的边 (edge) 传播
- ▶ 距离出发点越远, 扩散效应就越小

# 双管齐下！

我们需要更强的假设和更精巧的实验设计：

- ▶ 扩散常常局限在一定范围之内 (班级内部)
- ▶ 一般沿着社交网络的边 (edge) 传播
- ▶ 距离出发点越远, 扩散效应就越小

这些假设可以帮我们估计实际的 $W_i$

# 双管齐下！

我们需要更强的假设和更精巧的实验设计：

- ▶ 扩散常常局限在一定范围之内 (班级内部)
- ▶ 一般沿着社交网络的边 (edge) 传播
- ▶ 距离出发点越远, 扩散效应就越小

这些假设可以帮我们估计实际的  $W_i$

e.g. 班级平均接受的处理水平

# 双管齐下！

实验设计: 如何在假设的基础上识别出我们感兴趣的因果效应?

# 双管齐下！

实验设计: 如何在假设的基础上识别出我们感兴趣的因果效应?  
某种药物的直接效果是多大? 通过扩散导致的间接效果有多强?

# 双管齐下！

实验设计: 如何在假设的基础上识别出我们感兴趣的因果效应?

某种药物的直接效果是多大? 通过扩散导致的间接效果有多强?

两种设计思路:

- ▶ 固定网络 + 随机处理
- ▶ 随机网络 + 固定处理

## 固定网络和随机处理

## 例子A: 推销保险

怎么才能劝说更多的大学行政人员购买养老保险?

## 例子A: 推销保险

怎么才能劝说更多的大学行政人员购买养老保险?

直接的想法: 给每位行政人员发放宣传材料

## 例子A: 推销保险

怎么才能劝说更多的大学行政人员购买养老保险?

直接的想法: 给每位行政人员发放宣传材料

更有效率的做法: 只劝说一部分人, 靠社交网络的传播来带动其他人

## 例子A: 推销保险

怎么才能劝说更多的大学行政人员购买养老保险?

直接的想法: 给每位行政人员发放宣传材料

更有效率的做法: 只劝说一部分人, 靠社交网络的传播来带动其他人

- ▶ Duflo and Saez (2003): 在MIT进行的一项随机实验
- ▶ 随机抽取了全校三分之二的系作为处理组, 余下三分之一为控制组
- ▶ 在每个位于处理组的系中, 二分之一的雇员被选中接受处理 (宣讲会邀请函)

## 例子A: 推销保险

最简单的社交网络: 假定行政人员只与本系人员交流, 扩散只发生在各系内部

## 例子A: 推销保险

最简单的社交网络: 假定行政人员只与本系人员交流, 扩散只发生在各系内部

三个不同组别: 处理组中接受处理的个体, 处理组中未接受处理的个体, 以及控制组中的个体

## 例子A: 推销保险

最简单的社交网络: 假定行政人员只与本系人员交流, 扩散只发生在各系内部

三个不同组别: 处理组中接受处理的个体, 处理组中未接受处理的个体, 以及控制组中的个体

组2 - 组3: 处理的扩散效应

组1 - 组3: 处理的直接效应 + 扩散效应

组1 - 组2: 处理的直接效应

## 例子A: 推销保险

最简单的社交网络: 假定行政人员只与本系人员交流, 扩散只发生在各系内部

三个不同组别: 处理组中接受处理的个体, 处理组中未接受处理的个体, 以及控制组中的个体

组2 - 组3: 处理的扩散效应

组1 - 组3: 处理的直接效应 + 扩散效应

组1 - 组2: 处理的直接效应

思考题: 实验设计如何帮助我们识别了扩散效应?

## 理论A: 直接效应 VS. 间接效应

怎么在Rubin模型的框架下描述例子A中的实验?

## 理论A: 直接效应 VS. 间接效应

怎么在Rubin模型的框架下描述例子A中的实验?

Hudgens and Halloran (2008) 给我们提供了答案

- ▶ 假设有 $N$ 个互相之间没有联系的组(例子中的各个系)
- ▶ 两步实验: 先确定各个组中接受处理的比例, 再进行组内分配
- ▶ 对于某组 $i$ 中的任意个体 $j$ ,  $Y_{ij} = Y_{ij}(\mathbf{d})$ , 这里 $\mathbf{d} = (d_1, d_2, \dots, d_{n_i})$
- ▶ Trick: 保持自己的treatment status不变, 对其他人的treatment status取期望
- ▶ 直接效应:  $\tau_{ij}^D = Y_{ij}(d_{ij} = 1, \mathbf{d}_{(j)i}) - Y_{ij}(d_{ij} = 0, \mathbf{d}_{(j)i})$
- ▶ 间接效应:  $\tau_{ij}^I = Y_{ij}(d_{ij} = 0, \mathbf{d}_{(j)i}) - Y_{ij}(d_{ij} = 0, \mathbf{d}'_{(j)i})$

## 例子B: 打击暴力行为

美国有一个因暴力行为而臭名昭著的群体 —

## 例子B: 打击暴力行为

美国有一个因暴力行为而臭名昭著的群体 — 中学生

## 例子B: 打击暴力行为

美国有一个因暴力行为而臭名昭著的群体 — 中学生  
如何在新泽西的中学里减少暴力行为发生的概率？

## 例子B: 打击暴力行为

美国有一个因暴力行为而臭名昭著的群体 — 中学生  
如何在新泽西的中学里减少暴力行为发生的概率?  
可以使用之前的设计吗 (两步随机实验)?

## 例子B: 打击暴力行为

美国有一个因暴力行为而臭名昭著的群体 — 中学生  
如何在新泽西的中学里减少暴力行为发生的概率?

可以使用之前的设计吗 (两步随机实验)?

不太行, 因为很难假设扩散局限在班级内部

## 理论B: Exposure Mapping

跟例子A不同, 这里并没有明确的组别

## 理论B: Exposure Mapping

跟例子A不同, 这里并没有明确的组别

每个个体得到的处理都可能沿着社交网络向各个方向扩散

## 理论B: Exposure Mapping

跟例子A不同, 这里并没有明确的组别

每个个体得到的处理都可能沿着社交网络向各个方向扩散

无法根据组别来区分直接效应和间接效应

## 理论B: Exposure Mapping

跟例子A不同, 这里并没有明确的组别

每个个体得到的处理都可能沿着社交网络向各个方向扩散

无法根据组别来区分直接效应和间接效应

Aronow and Samii (2017): 定义从处理分配 $D_i$ 到实际处理 $W_i$ 的映射 $f$

## 理论B: Exposure Mapping

跟例子A不同, 这里并没有明确的组别

每个个体得到的处理都可能沿着社交网络向各个方向扩散

无法根据组别来区分直接效应和间接效应

Aronow and Samii (2017): 定义从处理分配 $D_i$ 到实际处理 $W_i$ 的映射 $f$

根据映射 $f$ 可以计算每个 $W_i$ 出现的概率, 进而将 $D_i$ 的外生性转化为 $W_i$ 的外生性

## 例子B: 打击暴力行为

Paluck, Shepherd, Aronow (2016)

## 例子B: 打击暴力行为

Paluck, Shepherd, Aronow (2016)

- ▶ 首先调查学校内部的朋友网络

## 例子B: 打击暴力行为

Paluck, Shepherd, Aronow (2016)

- ▶ 首先调查学校内部的朋友网络
- ▶ 在网络内部随机选取节点接受处理

## 例子B: 打击暴力行为

Paluck, Shepherd, Aronow (2016)

- ▶ 首先调查学校内部的朋友网络
- ▶ 在网络内部随机选取节点接受处理
- ▶ 假定扩散局限于直接朋友之间

## 例子B: 打击暴力行为

Paluck, Shepherd, Aronow (2016)

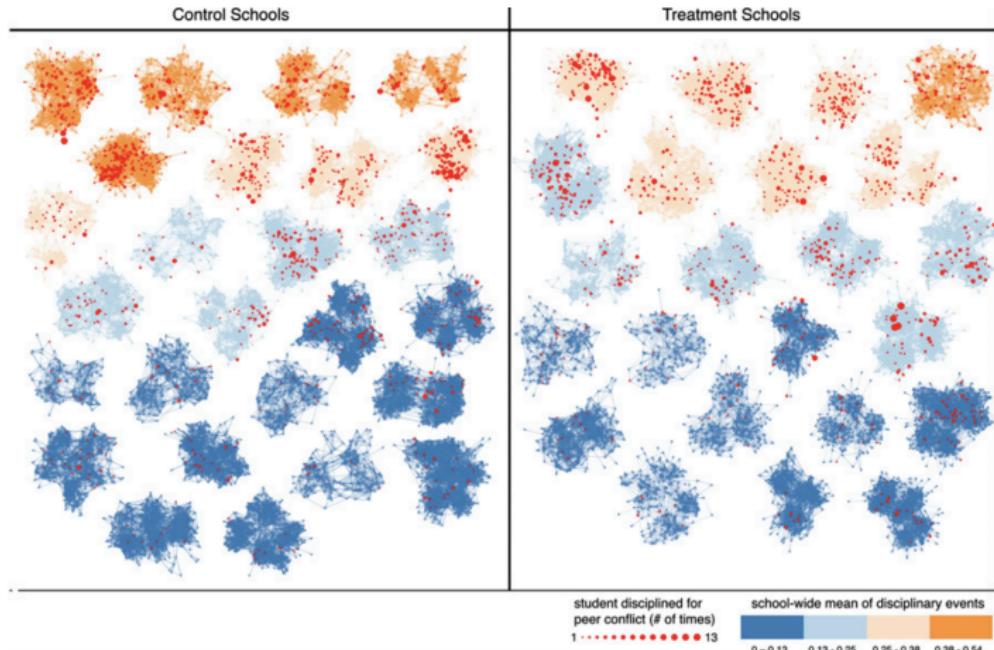
- ▶ 首先调查学校内部的朋友网络
- ▶ 在网络内部随机选取节点接受处理
- ▶ 假定扩散局限于直接朋友之间
- ▶ 根据身边朋友接受处理的比例计算treatment exposure

## 例子B: 打击暴力行为

Paluck, Shepherd, Aronow (2016)

- ▶ 首先调查学校内部的朋友网络
- ▶ 在网络内部随机选取节点接受处理
- ▶ 假定扩散局限于直接朋友之间
- ▶ 根据身边朋友接受处理的比例计算treatment exposure
- ▶ 用Aronow and Samii (2017)提供的方法给出估计值

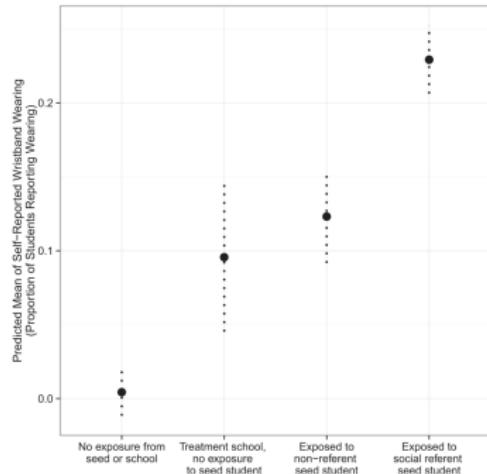
## 例子B: 打击暴力行为



## 例子B: 打击暴力行为

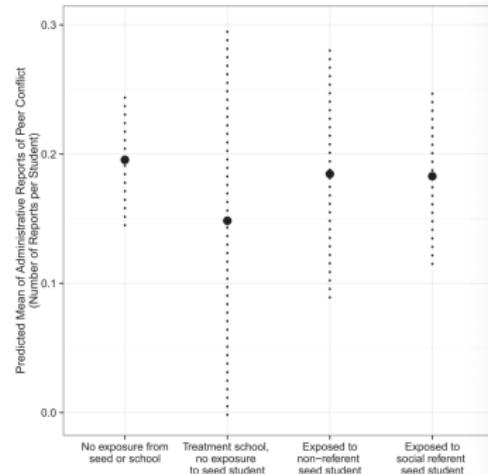
C

Wearing anti-conflict wristband



D

Peer conflict (disciplinary records)



## 例子C: 打虫药

Miguel and Kremer (2004): 打虫药的效果能否扩散到周边地区？

## 例子C: 打虫药

Miguel and Kremer (2004): 打虫药的效果能否扩散到周边地区?  
一项在肯尼亚的随机实验

## 例子C: 打虫药

Miguel and Kremer (2004): 打虫药的效果能否扩散到周边地区?

一项在肯尼亚的随机实验

- ▶ 50所学校中25所被选为处理组, 学生得到了打虫药

## 例子C: 打虫药

Miguel and Kremer (2004): 打虫药的效果能否扩散到周边地区?

一项在肯尼亚的随机实验

- ▶ 50所学校中25所被选为处理组, 学生得到了打虫药
- ▶ 我们知道每所学校的地理坐标

## 例子C: 打虫药

Miguel and Kremer (2004): 打虫药的效果能否扩散到周边地区?

一项在肯尼亚的随机实验

- ▶ 50所学校中25所被选为处理组, 学生得到了打虫药
- ▶ 我们知道每所学校的地理坐标
- ▶ 距离处理组学校较近的控制组学校能否从中受益?

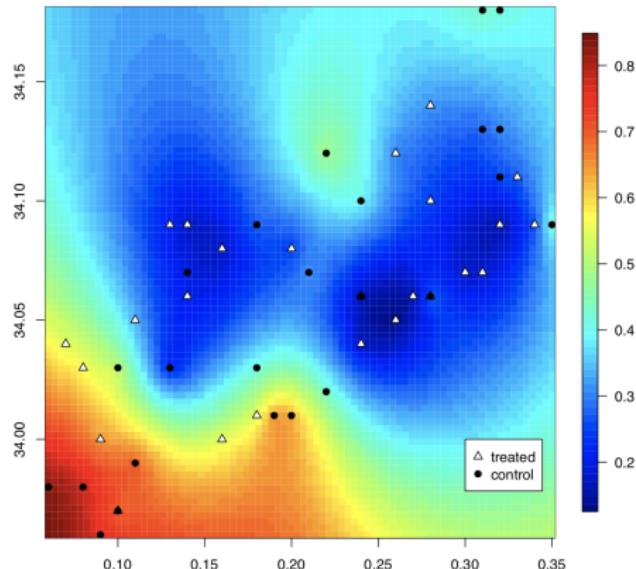
## 理论C: MIR曲线

直接的想法: 估计任意处理组学校方圆五公里内的寄生虫发病率

## 理论C: MIR曲线

直接的想法: 估计任意处理组学校方圆五公里内的寄生虫发病率

问题: 各个学校的五公里范围会互相重合

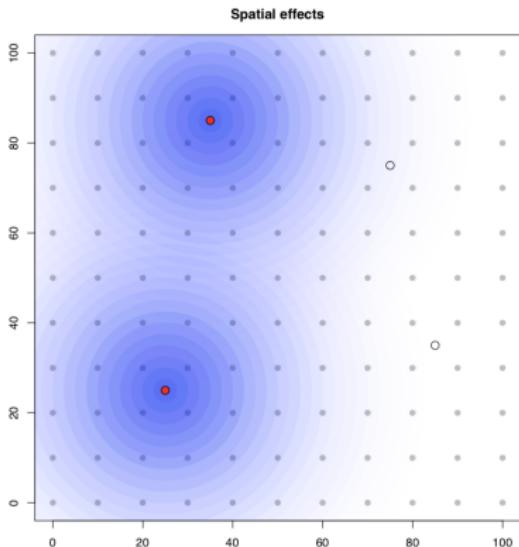


## 理论C: MIR曲线

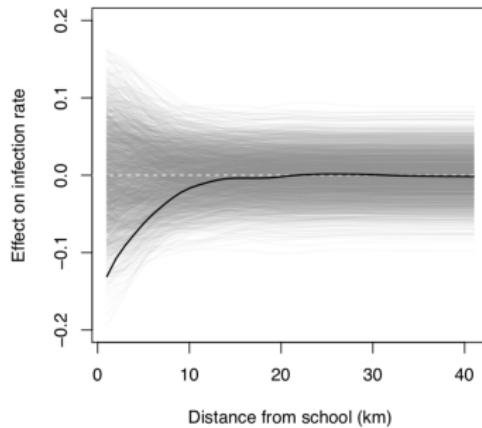
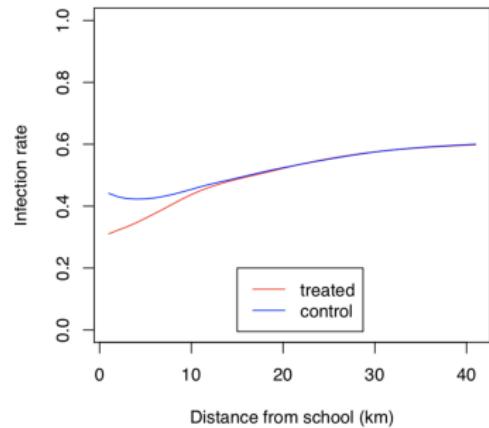
Aronow, Samii and Wang (2019): 估计Marginalized Individualistic Response (MIR)曲线

## 理论C: MIR曲线

Aronow, Samii and Wang (2019): 估计Marginalized Individualistic Response (MIR)曲线



## 例子C: 打虫药



## 随机网络和固定处理

## 例子D: 胖子

你变胖的原因会不会是.... 身边的朋友都太胖了?

## 例子D: 胖子

你变胖的原因会不会是.... 身边的朋友都太胖了?

Fowler和Christakis的著名结论....

## 例子D: 胖子

你变胖的原因会不会是.... 身边的朋友都太胖了?

Fowler和Christakis的著名结论....

但后来的学者用同样的数据和方法发现 —

## 例子D: 胖子

你变胖的原因会不会是.... 身边的朋友都太胖了?

Fowler和Christakis的著名结论....

但后来的学者用同样的数据和方法发现 — 高个子的朋友们也会变得更高....



## 例子D: 胖子

你变胖的原因会不会是.... 身边的朋友都太胖了?

Fowler和Christakis的著名结论....

但后来的学者用同样的数据和方法发现 — 高个子的朋友们也会变得更高....



这是一个“同侪效应(peer effect)”的经典例子: 具有不同特质的个体, 被同一个网络连接起来, 他们之间的互动会产生哪些影响?

## 理论D: 估计同侪效应的风险

- ▶ Homophily: 相似的人会扎堆 (胖子)

## 理论D: 估计同侪效应的风险

- ▶ Homophily: 相似的人会扎堆 (胖子)
- ▶ Common shocks: 火锅毁所有

## 理论D: 估计同侪效应的风险

- ▶ Homophily: 相似的人会扎堆 (胖子)
- ▶ Common shocks: 火锅毁所有
- ▶ 用谁做解释变量? 结果还是特质? 要不要leave one out?

## 理论D: 估计同侪效应的风险

- ▶ Homophily: 相似的人会扎堆 (胖子)
- ▶ Common shocks: 火锅毁所有
- ▶ 用谁做解释变量? 结果还是特质? 要不要leave one out?
- ▶ 用自己的GPA对宿舍平均GPA做回归, 系数是多少?

## 理论D: 估计同侪效应的风险

- ▶ Angrist (2014): 同侪效应的估计本质上是一个2SLS (工具变量)

## 理论D: 估计同侪效应的风险

- ▶ Angrist (2014): 同侪效应的估计本质上是一个2SLS (工具变量)
- ▶ 2SLS跟OLS的结果差异可能源于扩散以外的其他原因

## 理论D: 估计同侪效应的风险

- ▶ Angrist (2014): 同侪效应的估计本质上是一个2SLS (工具变量)
- ▶ 2SLS跟OLS的结果差异可能源于扩散以外的其他原因
- ▶ 扩散的本质: 一种non-compliance, 也是mediation

## 理论D: 估计同侪效应的风险

- ▶ Angrist (2014): 同侪效应的估计本质上是一个2SLS (工具变量)
- ▶ 2SLS跟OLS的结果差异可能源于扩散以外的其他原因
- ▶ 扩散的本质: 一种non-compliance, 也是mediation
- ▶ 用简单线性模型未必能够刻画扩散现象

## 例子E: 中关村应用文理学院

在P大怎么分配宿舍对GPA的提升最大?

## 例子E: 中关村应用文理学院

在P大怎么分配宿舍对GPA的提升最大?

Li et al. (2018): 利用了宿舍分配的随机性

## 例子E: 中关村应用文理学院

在P大怎么分配宿舍对GPA的提升最大?

Li et al. (2018): 利用了宿舍分配的随机性

问题: 跟竞赛保送生住同一个宿舍对GPA有没有好处?

## 例子E: 中关村应用文理学院

在P大怎么分配宿舍对GPA的提升最大?

Li et al. (2018): 利用了宿舍分配的随机性

问题: 跟竞赛保送生住同一个宿舍对GPA有没有好处?

处理是固定的, 网络是随机的

## 理论E

关键点：室友的作用取决于你自己的特质

## 理论E

关键点：室友的作用取决于你自己的特质  
这等价于加了交叉项的线性模型

## 理论E

关键点：室友的作用取决于你自己的特质

这等价于加了交叉项的线性模型

记高考生为1, 竞赛生为2

## 理论E

关键点：室友的作用取决于你自己的特质

这等价于加了交叉项的线性模型

记高考生为1, 竞赛生为2

假定你住四人寝室, 那么你的室友组合有以下四种: (111), (112),  
(122), (222)

## 理论E

关键点：室友的作用取决于你自己的特质

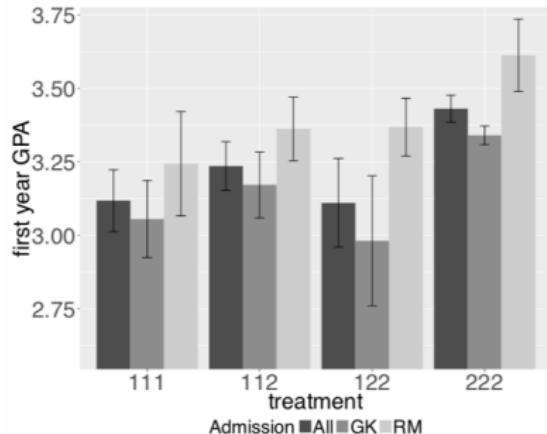
这等价于加了交叉项的线性模型

记高考生为1, 竞赛生为2

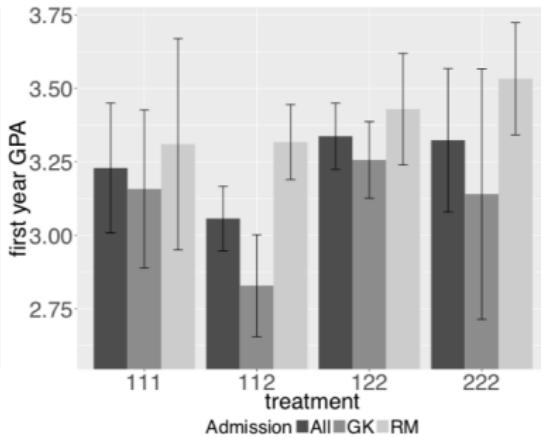
假定你住四人寝室, 那么你的室友组合有以下四种: (111), (112),  
(122), (222)

我们可以计算出不同室友组合下高考生/竞赛生的平均GPA

# 例子E: 中关村应用文理学院

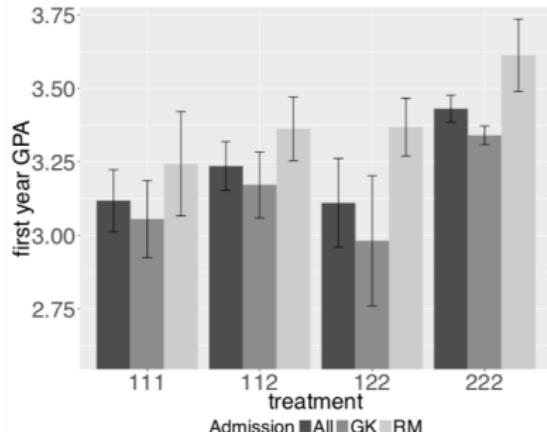


(a) Informatics

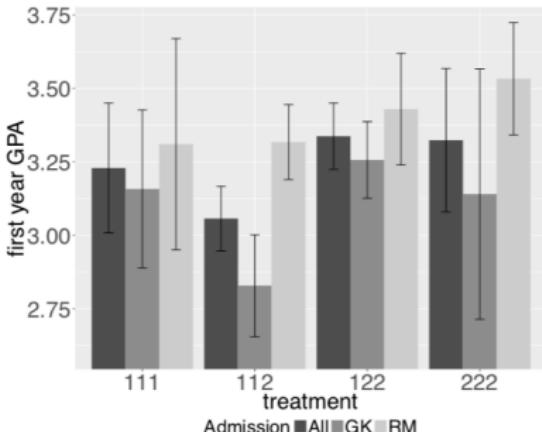


(b) Physics

# 例子E: 中关村应用文理学院



(a) Informatics



(b) Physics

结论: 保送生带带我!

# 总结

怎么设计一个可以估计扩散的实验?

## 总结

怎么设计一个可以估计扩散的实验?

在网络给定, 处理分配由研究者控制的时候

# 总结

怎么设计一个可以估计扩散的实验?

在网络给定, 处理分配由研究者控制的时候

- ▶ 如果认为扩散局限在某个组内: 两步实验 (Hudgens and Halloran, 2008)

# 总结

怎么设计一个可以估计扩散的实验?

在网络给定, 处理分配由研究者控制的时候

- ▶ 如果认为扩散局限在某个组内: 两步实验 (Hudgens and Halloran, 2008)
- ▶ 如果知道网络结构: 利用treatment exposure mapping (Aronow and Samii, 2017)

# 总结

怎么设计一个可以估计扩散的实验?

在网络给定, 处理分配由研究者控制的时候

- ▶ 如果认为扩散局限在某个组内: 两步实验 (Hudgens and Halloran, 2008)
- ▶ 如果知道网络结构: 利用treatment exposure mapping (Aronow and Samii, 2017)
- ▶ 如果扩散发生在地理空间下: 估计MIR曲线 (Aronow, Samii and Wang, 2019)

# 总结

怎么设计一个可以估计扩散的实验?

在网络给定, 处理分配由研究者控制的时候

- ▶ 如果认为扩散局限在某个组内: 两步实验 (Hudgens and Halloran, 2008)
- ▶ 如果知道网络结构: 利用treatment exposure mapping (Aronow and Samii, 2017)
- ▶ 如果扩散发生在地理空间下: 估计MIR曲线 (Aronow, Samii and Wang, 2019)

# 总结

怎么设计一个可以估计扩散的实验?

在网络给定, 处理分配由研究者控制的时候

- ▶ 如果认为扩散局限在某个组内: 两步实验 (Hudgens and Halloran, 2008)
- ▶ 如果知道网络结构: 利用treatment exposure mapping (Aronow and Samii, 2017)
- ▶ 如果扩散发生在地理空间下: 估计MIR曲线 (Aronow, Samii and Wang, 2019)

在处理给定, 网络构建由研究者控制的时候

# 总结

怎么设计一个可以估计扩散的实验?

在网络给定, 处理分配由研究者控制的时候

- ▶ 如果认为扩散局限在某个组内: 两步实验 (Hudgens and Halloran, 2008)
- ▶ 如果知道网络结构: 利用treatment exposure mapping (Aronow and Samii, 2017)
- ▶ 如果扩散发生在地理空间下: 估计MIR曲线 (Aronow, Samii and Wang, 2019)

在处理给定, 网络构建由研究者控制的时候

- ▶ 小心避免homophily和common shocks

# 总结

怎么设计一个可以估计扩散的实验?

在网络给定, 处理分配由研究者控制的时候

- ▶ 如果认为扩散局限在某个组内: 两步实验 (Hudgens and Halloran, 2008)
- ▶ 如果知道网络结构: 利用treatment exposure mapping (Aronow and Samii, 2017)
- ▶ 如果扩散发生在地理空间下: 估计MIR曲线 (Aronow, Samii and Wang, 2019)

在处理给定, 网络构建由研究者控制的时候

- ▶ 小心避免homophily和common shocks
- ▶ 不要用outcome做解释变量

# 总结

怎么设计一个可以估计扩散的实验?

在网络给定, 处理分配由研究者控制的时候

- ▶ 如果认为扩散局限在某个组内: 两步实验 (Hudgens and Halloran, 2008)
- ▶ 如果知道网络结构: 利用treatment exposure mapping (Aronow and Samii, 2017)
- ▶ 如果扩散发生在地理空间下: 估计MIR曲线 (Aronow, Samii and Wang, 2019)

在处理给定, 网络构建由研究者控制的时候

- ▶ 小心避免homophily和common shocks
- ▶ 不要用outcome做解释变量
- ▶ 效应大小取决于本身特质 (Li et al., 2018)

谢谢！