

Inference and influence of network structure using snapshot social behavior without network data

Antonia Godoy-Lorite and Nick S. Jones

摘要

Population behavior, like voting and vaccination, depends on the structure of social networks. This structure can differ depending on behavior type and is typically hidden. However, we do often have behavioral data, albeit only snapshots taken at one time point. We present a method jointly inferring a model for both network structure and human behavior using only snapshot population-level behavioral data. This exploits the simplicity of a few parameter model, geometric sociodemographic network model, and a spin-based model of behavior. We illustrate, for the European Union referendum and two London mayoral elections, how the model offers both prediction and the interpretation of the homophilic inclinations of the population. Beyond extracting behavior-specific network structure from behavioral datasets, our approach yields a framework linking inequalities and social preferences to behavioral outcomes. We illustrate potential network-sensitive policies: How changes to income inequality, social temperature, and homophilic preferences might have reduced polarization in a recent election.

关键词: inference of network structure; Blau's theory; Ising model; SRGG;

1 引言

受到社交网络结构的影响，个人行为会有不同的表现。^[1]通过复现和改变网络结构，可以加深我们对社会网络的理解，阐明社会不平等变化如何影响社会两级分化。虽然我们有大量高质量的社会科学数据集，但是将个人特定行为与其个人具体情况（年龄、性别、收入等）联系起来的社交网络实现成本非常高，甚至难以清楚地描述出这种特定行为的社交网络。

目前科技公司拥有海量社交网络结构数据，如 Twitter、Facebook 等^{[2][3]}。但这些数据具有商业敏感性，其中个人隐私问题无法解决，并且不能确认这些数据集是真实世界表述。以上原因使得这些数据集难以访问和共享。

这篇论文提出一种方法，从没有网络结构的社会行为快照数据再现出网络的结构，并提高对行为过程及行为与社会不平等之间的联系的理解^[4]。由于其简单性和可解释性，可将 KBI 模型推广至其他的行为或者态度研究过程中。该方法还产生了一个将不平等和社会偏好与行为结果联系起来的框架。并说明了潜在的网络敏感政策：收入不平等、社会温度和同质化偏好的变化可能会减少选举中的两极分化。

2 相关工作

2.1 Blau 社交结构几何观理论

Peter Blau 在其社交结构理论中提出一个假设，将每个人表示为社会人口参数（如年龄、性别和收入水平）的高维空间中的一个点。在该空间中，因为点在空间中聚集并在参数之间建立关联。将社

会人口学参数嵌入空间意味着可以计算某种距离度量，例如根据两个人在空间中的距离来计算他们的相似程度。该高维空间被称为 Blau 空间^[5]。

3.2 社会科学中的自旋模型

Ising 模型源于描述物质的磁铁性质，该模型被证明成功地解释了磁体的磁性相变。自旋模型的核心是相互作用的思想，其中每个顶点都具有特定状态，同时结合了相邻顶点状态的影响。将自旋模型应用于社会科学的一个例子是网络上的经典选民模型。选民模型涉及简化的 Ising 状态和交互，其中状态根据相邻顶点的状态对齐/翻转。

3 本文方法

3.1 本文方法概述

输入为行为数据和社会人口统计学数据，经过贝叶斯近似计算得出模型参数和网络结构，方法如图 1 所示：

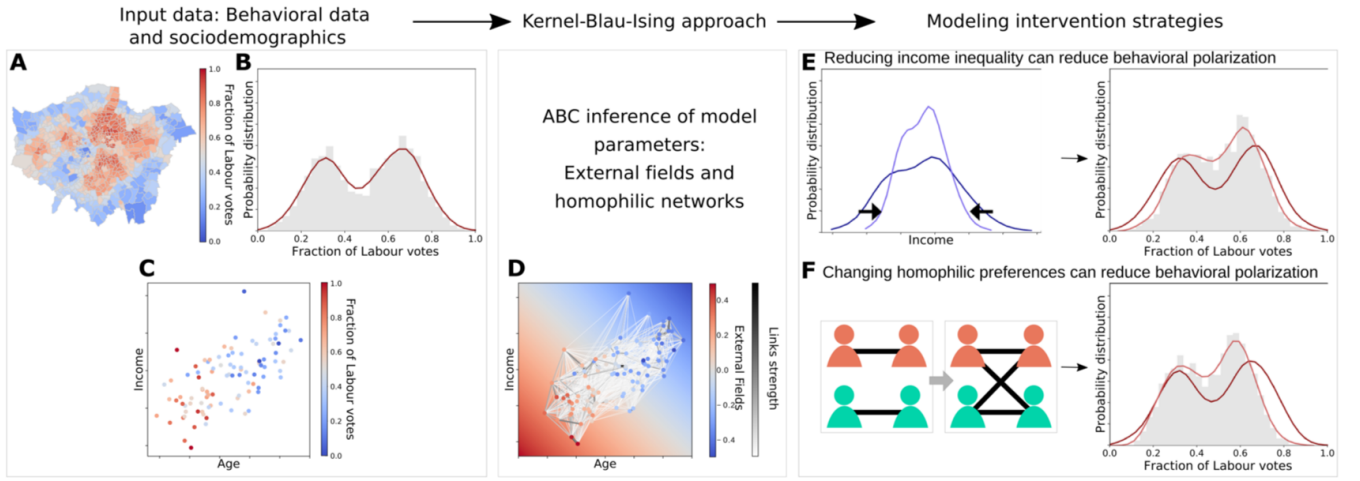


图 1: 方法示意图

3.2 网络模型

3.2.1 邻接矩阵

A 是邻接矩阵，当点 i 与点 j 相连接时 $A_{ij} = 1$ 。节点对之间连接概率服从伯努利分布

$$P(A_{ij} = 1 | z_i, z_j, \theta) = \rho(z_i, z_j, \theta) \quad (1)$$

3.2.2 连接核

使用 sigmoid 函数作为连接核 ρ

$$\rho(z_i, z_j, \theta) = \frac{1}{1 + \exp(d_{ij})} \quad (2)$$

3.2.3 节点距离

$$d_{ij} = \theta_0 + \sum_{k=1}^K \theta_k |z_{ik} - z_{jk}| \quad (3)$$

d_{ij} 是 Blau 空间中节点对之间的距离。 θ_0 是一个偏差项，解释了整体连接密度。 θ_k 是 Blau 空间第 k 维度的连接系数，它影响该维度中的距离对整体距离的重要程度， θ_k 越大该维度的同质性越强。

3.3 行为模型

3.3.1 Ising 模型

从 Ising 模型可以生成自旋方向，个人的自旋方向取决于外部域和与其相连接的人的自旋方向。每个个人 i 的行为可由 $\sigma_i = [-1, 1]$ 表示，例如 -1 代表 i 给保守党投票，1 代表 i 给工党投票。整个人群的自旋可表示为 $\sigma \in [-1, 1]^N$ 。

3.3.2 External Fields

外部场为 Blau 空间每个维度中的线性场，其中 Blau 空间 k 的每个维度中的线性系数为 h_k ，EFs 与个体自选相作用表示为 $h \cdot z_i = \sum_k h_k z_{ik}$ 。自旋仅根据它们的坐标与 EFs 相互作用，因此它们倾向于与 EFs 对齐。

3.3.3 Hamiltonian function

自旋结构的能量由哈密顿函数给出， $H(\sigma, h, J, A) = -\sum_i (\sum_k h_k z_{ik}) \sigma_i - J \sum_{ij} A_{ij} \sigma_i \sigma_j$ ， J 是连接强度， A_{ij} 是邻接矩阵。

3.3.4 Boltzmann distribution

$$p(\sigma | \beta, h, J, A) = \frac{e^{-\beta H(\sigma, h, J, A)}}{Z(\beta, h, J, A)}$$

3.3.5 Normalization constant

$$Z(\beta, h, J, A) = \sum_{\sigma' \in \Omega} e^{-\beta H(\sigma', h, J, A)}$$

4 复现细节

4.1 模型参数推断

在 Ising 模型中，需要计算 2^N 个项，即使是小系统也难以计算出函数 Z 。并且 Z 的值取决于模型参数，所以需要为每个可能的参数集重新计算。在此情况下，很难计算出参数的精确值。为了计算出模型参数，所以使用近似贝叶斯计算求得近似值。

Procedure 1 approximate Bayesian computation rejection algorithm

```

for  $i = 1$  to  $N$  do
  repeat
    Draw independent proposal  $\theta' \sim \pi(\theta)$ 
    Generate  $A' \sim \rho(z, \theta')$ 
    Generate  $\sigma' \sim p(A', \theta')$ 
  until  $\|\eta(\sigma', \sigma)\| < \varepsilon$ ;
  Set  $\theta_i \leftarrow \theta'$ 
end

```

4.2 程序运行流程

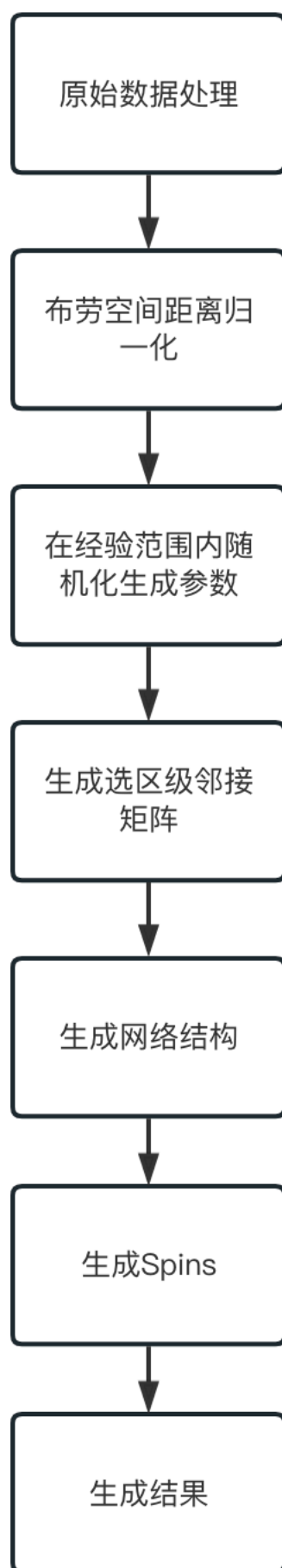


图 2: 程序流程图

4.3 实验环境搭建

因为有大量运算，实验环境需要安装 NumPy 用于计算。为了代码的高效运行，使用 PyPy 进行编译，使用的版本为 2.7。

5 实验结果分析

年长的选民更喜欢投给保守党和选择脱欧; 男性投票给保守党和选择脱欧比女性多; 高收入选民更喜欢保守党而不是工党。2016 年伦敦市长选举和脱欧公投只相隔 49 天，但是模型参数有一定差异，因此推断社会变化不是由于社会连接结构的突然改变，而是表明与市长选举相比，欧盟公投中是不同的社会关系在发挥作用。

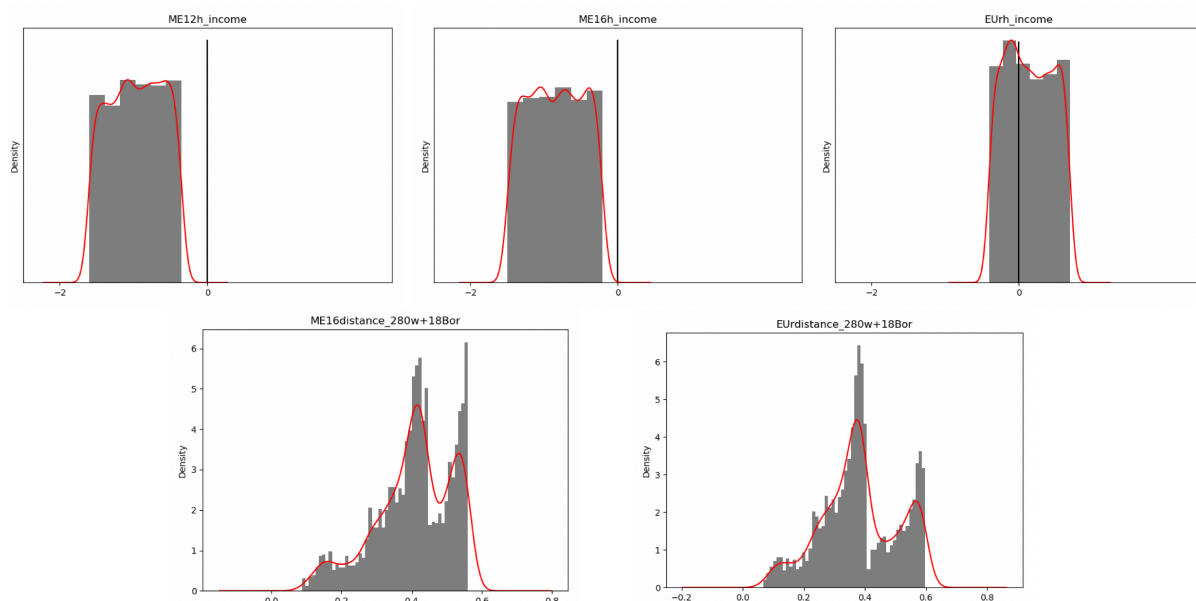


图 3: 实验结果分析图

6 总结与展望

该研究提出的 KBI 方法具有参数可解释性，仅使用人口行为“快照”数据，参数有效地反应了包含网络效应的社会行为。该方法可以精确地再现人口水平的行为和已知的同质依附倾向。通过有效地推断 Blau 空间中与行为相关的社交关系，该研究不仅能够提高人们对行为过程及其与不平等和社会偏好之间联系的理解，还能够提高对未观察到的社会结果的可预测性。该方法的局限性也显而易见，由于所有数据都被视为事件的快照，因此忽略了社会行为的时间序列信息，往往历史行为也具有重要作用。并且论文中对参数的分析是描述性的而且是非常主观的。KBI 方法相对于简单回归分析的优势在于其包含了网络效应，通过公开的普查数据结合更高维度的 Ising 模型可以创造出更为复杂的网络模型。

参考文献

- [1] J. ZHANG D C. Social Networks and Health: New Developments in Diffusion, Online and Offline[J]. Annual Review of Sociology, 2019, 45: 91-109.
- [2] D. LAZER L A, A. Pentland. Computational Social Science[J]. SCIENCE, 2009, 323: 721-723.
- [3] M. KOSINSKI S D G, S. C. Matz. Facebook as a research tool for the social sciences: Opportunities,

challenges, ethical considerations, and practical guidelines[J]. *American Psychologist*, 2015, 70: 543-556.

- [4] R. JOHNSTON R S, K. Jones. Party support and the neighbourhood effect: Spatial polarisation of the British electorate, 1991–2001[J]. *Political Geography*, 2004, 23: 367-402.
- [5] T. HOFFMANN N S J. Inference of a universal social scale and segregation measures using social connectivity kernels[J]. *Royal Society*, 2020, 17: 171.