

信念系统动力学作为单层神经网络的网络

Yujian Fu

*Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA and
Center for the Study of Complex Systems, University of Michigan, Ann Arbor, Michigan 48109, USA*

随着政治极化和虚假信息传播问题变得严重，社交网络上的信念传播成为一个重要的探索问题。在理解群体共识或极化如何在一个群体中发生方面已经取得了一些算法上的突破。本文提出了一种改进的 Friedkin-Johnsen 模型，试图通过将每个个体视为单层神经网络来解释个体固有的顽固性以及可能的反效果。这个神经网络处理特定陈述的一组证据，输入是每项证据的信心水平，而对该陈述的信念则是该神经网络的输出。

在这篇论文中，我们重申了麦迪逊解决派系弊端的方案的重要性，并发现当理解结构两极化时，一个具有巨量组件的网络比两个社区的网络更能减少信念分布的方差，但这样也会产生更多的社会压力。我们还发现，当形成社区结构时，信念分布的方差对个人信心水平变得不那么敏感。该模型可以应用于由错误信息和相互冲突的经济利益引起的政治和历史问题，也可以应用于人格理论和行为心理学。

I. 介绍

已经做出了贡献以理解信念极化和群体共识的潜在机制。数学社会学中的一些示例包括 Dandekar 等人引入的偏差同化，或 Friedkin 和 Johnsen 引入到 DeGroot 模型 [6][7][8] 中的固执。

受 Friedkin 的逻辑约束模型和 Madalina 在心理学 [3][4] 中应用 BENDING 模型的启发，引入了一个更丰富的认知结构来解释个体信念的形成和维持。逻辑相关的信念陈述往往是相互关联的，证据在信念形成中的作用与感知规范一样重要。

对于同一个想法，不同的人可能有不同的理解方式。两个相信同一件事的人在经历了同一事件并以相同的方式更新他们的证据后，可能会以不同的方式更新他们的信念。例如，考虑这句话：我想学习物理。有些人可能决定学物理是因为他们喜欢比别人更快地解决难题，有些人则可能因为喜欢发现和美感而信仰它。这两种人都能在物理课上取得成功，但在面对长期研究问题时，只有其中一种人才能坚持到底。

本文提出了一种信念传播的替代模型，其中个体对某一陈述的信念被分解为两个组成部分：自我推理和社会规范。自我推理是对支持个人同意特定陈述的一组证据的加权和，而社会规范则是对个体邻居信念的加权和。自我推理上的权重与 Friedkin-Johnsen 模型中的顽固值具有相同的解释。为了本文的目的，我们将称 Friedkin-Johnsen 模型中的顽固值为自信度，因为它个人对自己理解结构的信心水平。通过引入证

据和理解结构作为信念的附加底层结构，我们能够重现逻辑约束在多个陈述上的效果，以及它们共用的一组证据，并且数值上实现 BENDING 模型。

该模型还试图通过引入对立证据来解释信念传播中的反效果。例如，有些人喜欢巧克力是因为它甜，但有些人喜欢巧克力是因为它苦。但是苦和甜是相反的概念，却支持同一个陈述——巧克力很好。当个体接收到与其理解结构中考虑的某条证据相矛盾的证据时，这不会增加此人对该信念的确信度，反而会降低该信念。

该模型还将个体所受的社会压力衡量为他们的最终信念与自我推理之间的差异。在每个时间步，证据上的信心水平根据 DeGroot 模型在社交网络中传播，并通过证据和邻居信念的加权和来计算信念。

巧克力偏好的示例说明了基于纯粹生物学偏好形成的理解结构是如何根据持有者的观点来看是合乎逻辑的，但在其他人看来可能并非如此。不同的理解结构的形成可以由多种不同因素促成，包括文化、社会经济地位、教育水平、童年经历等。每个人以自己独特的方式了解世界，每个人也可能对自己的偏见一无所知。理解结构不是固定不变的。它会随着时间的推移而改变，基于经验，有时甚至是随机地变化。在本文中，我们对基于人们在某一固定时间点上的理解结构形成的信念感兴趣。

为了使理论具有普遍性，本文保持陈述的任意性，并采用一组任意的证据。通过选择不同的陈述，该理论有可能在政治科学、历史、发展心理学、人格理论和

行为心理学等领域中得到应用，解释理解结构作为对不同行动动机的权重。

在不同的人口规模、网络几何结构、自我信心水平和理解结构的假设下，发现了信念分布和社会压力的结果。

II. 方法

模型的具体实现见下文各节。

A. 模型环境

我们用两个对象初始化模型环境：

1. **网络**：智能体交互的社会网络被定义为一个埃爾德什-勒伊随机图，其中任意两个节点之间的边的概率等于 $\frac{k}{n}$ ，其中 n 是人口规模， $k < n$ 是连接性指数，描述了一个个体平均连接的人数。网络的邻接矩阵 $\{w_{ij}\}$ 被制作成双随机矩阵以确保一个节点的所有边权重之和为 1。网络可以通过两种不同的选项生成：

- a. 一个巨大的组件
- b. 两个社区

2. **证据池**：针对特定陈述的证据池由长度为 $2m$ 的 m 个陈述加上那些陈述的 m 个否定构成。个体理解结构将由从 m 个正/负证据对中选择的 m 个非零权重构建而成，其余 m 个权重为零。理解结构的定义见下文的代理属性部分。

在初始化这两个主要对象之后，模型随后读取预生成的文件，这些文件存储了用于初始化代理属性的信息以设置仿真。

B. 代理属性

在模拟过程中，每个智能体在其记忆中存储以下五条信息：

1. **理解结构**：对于个体 p ，其对某一特定陈述的理解结构被定义为与支持该陈述的证据 e_i 对应的权重集 $0 \leq u_i(p) \leq 1$ ，这些证据满足 $\sum_j^{2m} u_j(p) = 1$ 。理解结构可以通过两种不同的选项构建：

- a. 随机生成的
- b. 以概率 p 倾向于证据池的正方，以概率 $1-p$ 倾向于证据池的负方。其中 p 被定义为极化指数。

2. **证据的置信水平**：对于个体 p ，对证据 $0 \leq b_i(p) \leq 1$ 的置信水平集表示对该特定证据为真的确定程度。其中 1 表示最大程度的同意而 0 表示最大程度的不同意。当初始化证据的置信水平时，也可以通过两种不同的选项来构建它：

- a. 随机生成的
- b. 正向证据具有初始信心 a ，负向证据具有初始信心 $1-a$ 。

3. **自信心水平**：对于个体 p ，自信心水平 $0 \leq c(p) \leq 1$ 是自我推理对信念的权重贡献。 $1-c(p)$ 因此是社会规范对信念的权重贡献。
4. **信念**：对于个体 p ，对某一特定陈述的信念 $X(p)$ 计算如下：

$$X(p) = c(p) \sum_i b_i(p) u_i(p) + (1 - c(p)) \sum_{k \in N(p)} X(k) w_{kp}$$

其中 $N(p)$ 是个体 p 的邻居集合。 w_{kp} 是连接人 k 和 p 之间的边的权重。

5. **社会压力**：对于个体 p ，特定陈述的社会压力 $P(p)$ 按照以下方式计算：

$$P(p) = \left| X(p) - \sum_i b_i(p) u_i(p) \right|$$

C. 证据传播

在每一步中，一个人将使用每个邻居信心水平的加权和来更新其对证据的信心水平。例如，在时间点 t ， $b_k(i)$ 和 $b_l(i)$ 由个人 i 传递给个人 p 。如果 $u_k(p) = u_l(p) = 0$ ，则 $b_k(p), b_l(p)$ 没有变化。如果其中至少有一个不是，不失一般性假设 $u_k(p) \neq 0$ ，则

$$b_k(p) = (1 - w_{ip}) b_k(p) + w_{ip} b_k(i)$$

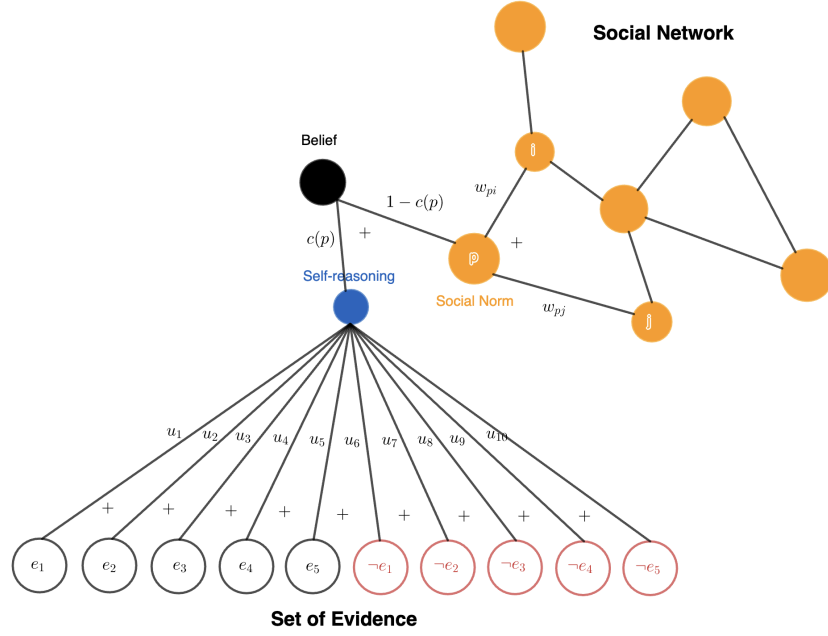


图 1. 模型框架中关于代理 p 的信念形成和传播的示意图。信念被分解为社会规范，这是邻居感知到的信念的加权总和以及自我推理，后者是一组正/负证据对的加权总和。

如果两者都不是，则

$$b_k(p) = (1 - w_{ip})b_k(p) + w_{ip}b_k(i)$$

$$b_l(p) = (1 - w_{ip})b_k(p) + w_{ip}b_l(i)$$

如果 e_k 和 e_l 是**相反证据**，并且 $b_k(i)$ 由人 i 传达给人 p ，那么我们不是根据邻居的信念的确信水平来调整新的确信水平，而是基于其补集进行调整。除了更新置信水平 $b_k(p)$ 外，我们还需要根据

$$b_l(p) = (1 - w_{ip})b_k(p) + w_{ip}(1 - b_k(i))$$

更新 $b_l(p)$

以及社会压力图。还展示了信念分布的标准差与其他变量之间的关系。

种群大小通常设置为 $n = 400$ ，正证据的数量通常设置为 $m = 5$ ，自信水平设置为 0.5，连接性设置为 $k = 10$ ，以允许大多数节点位于 Erdős – Rényi 随机图中的一个巨型组件内。对于每次模拟我们运行 40 个时间步骤，这确保了所考虑情况的收敛。

III. 实验

实验包括三次试验，第一次试验是在具有巨大组件的网络上进行随机初始理解结构和初始信心水平；第二次试验是在具有社区结构的网络上进行极化理解结构和初始信心水平；第三次试验是在具有巨大组件的网络上进行极化理解结构和初始信心水平。在每次试验中，都会绘制网络可视化、初始和最终信念分布

A. 随机初始化

为了随机初始理解结构和证据的初始置信度水平，该模型为每个证据生成一个 0 到 1 之间的随机浮点数作为置信度水平，同样地，对于以极化指数 $p = 0.5$ 选定的理解结构的每个非零元素也是如此。然后将理解结构的求和归一化为 1。

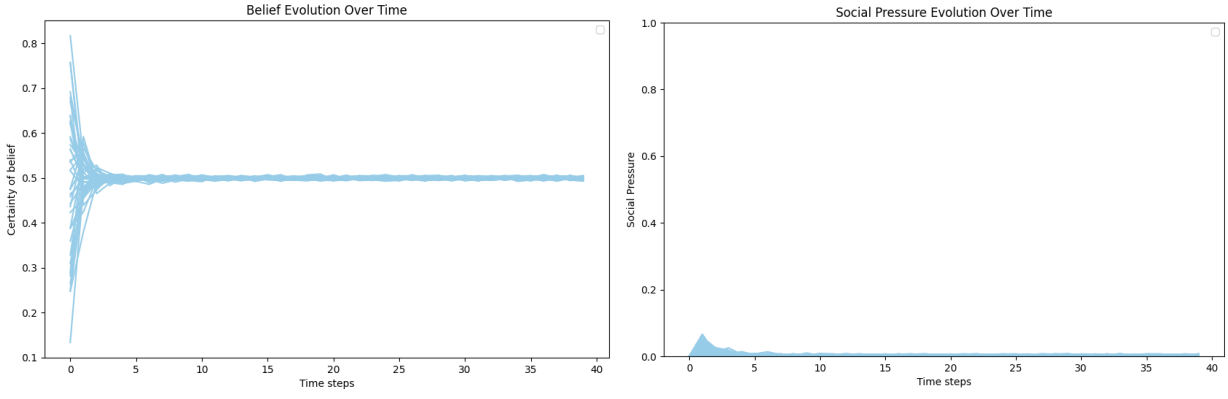


图 2. 前 40 个智能体的信念（左）和社会压力（右）随时间的变化，基于随机生成的理解结构和初始信心水平。

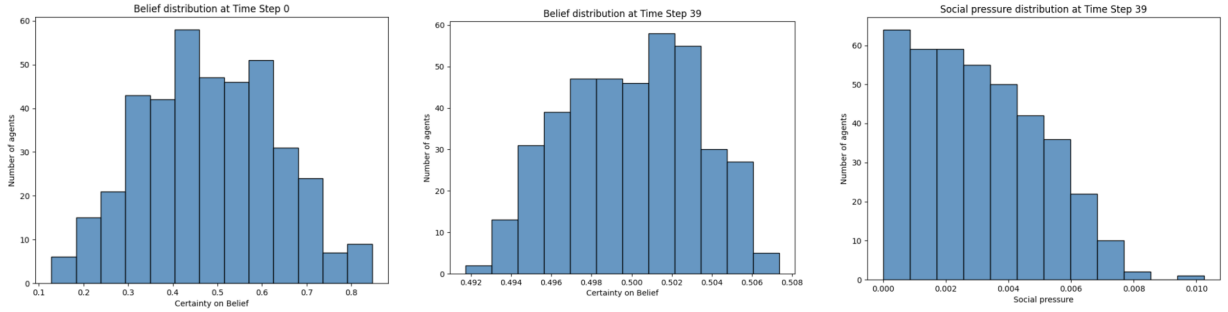


图 3. 初始和最终信念分布及社会压力分布



图 4. 初始和最终信念网络可视化对于随机生成的理解结构和初始信心水平。

模型随后在每个时间步生成信念分布、网络可视化和社会压力。在网络可视化中，节点的大小代表代理对声明的一致性水平。结果总结如下。

观察信念分布的初始和最终网络可视化，我们可以看到经过 40 步后，人群的信念已经达到了共识。

通过观察图 3 中的三个分布，即使人们的理解结构非常不同，但由于证据信息的广泛扩散以及社会规范的影响，最终信念分布收敛到 0.492 和 0.508 之间的

范围。总体而言，人口的社会压力较小，并呈下降趋势，最大值仅为 0.01。

信念分布会在 5 个时间步内收敛到社区中的一个平衡状态，如图 2 所示。社会压力全部从 0 开始，在初始时达到峰值，然后过了一段时间后减小。请注意，随着信念确定性的增加，社会压力的强度也会增大。这一现象在接下来的两个实验中非常重要。

生成信念分布和网络可视化之后，分别在图 13 中绘制了信念分布的标准差与连接性指数以及种群规模的关系。

B. 偏振初始化

1. 具有社区结构

为了生成理解的极化结构和证据的初始信心水平，我们使用与随机初始化相同的方法，但采用极化指数 $p = 0.8$ 。初始置信水平是根据一半的人群在第一个 5 证据上的置信水平为 0.2，第二个 5 证据上为 0.8；另一半人群在第一个 5 证据上的置信水平为 0.8，第二

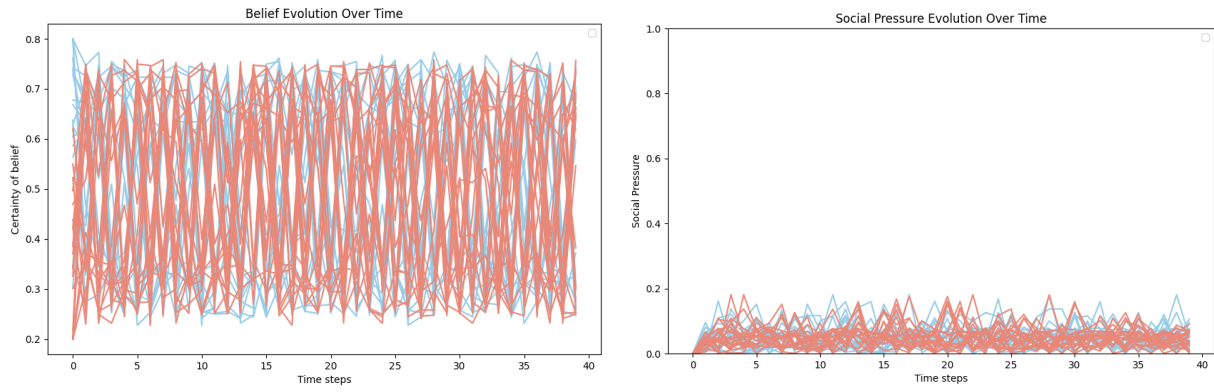


图 5. 信念（左）和社会压力（右）随时间在具有两个社区的网络中理解的极化结构和初始信心水平下，各自群体前 20 个和后 20 个代理的变化。

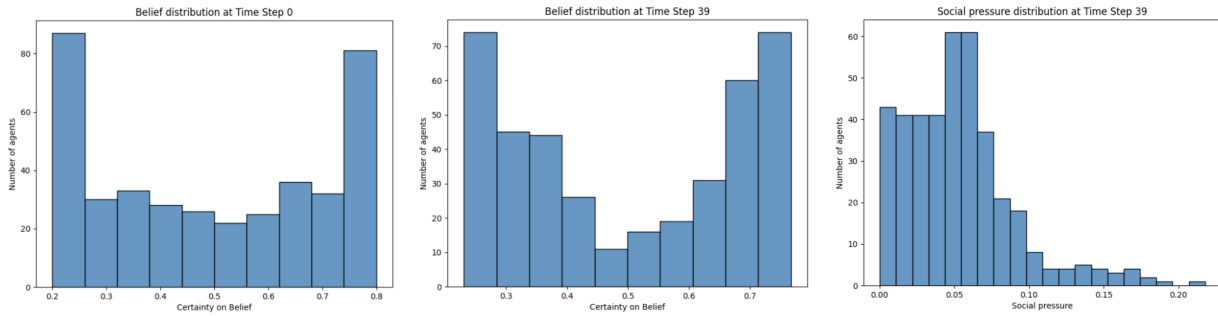


图 6. 初始和最终的信念分布及社会压力分布。

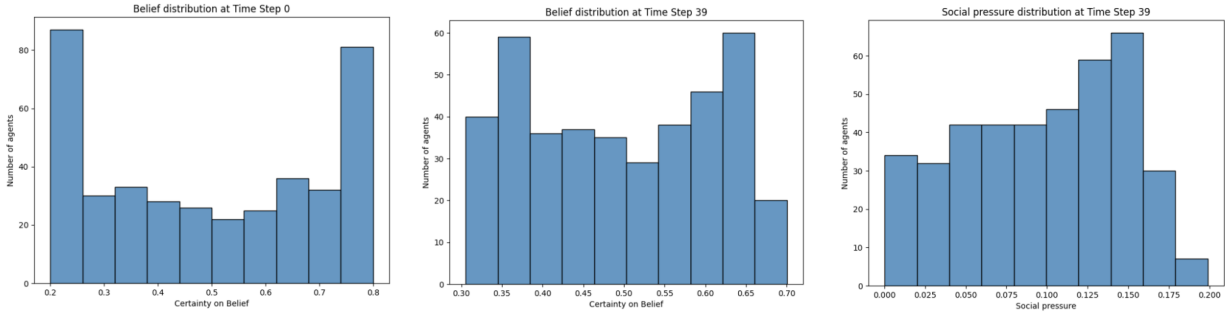


图 7. 初始和最终信念分布以及理解极化结构内的社会压力分布及巨型组件内的初始信心水平。

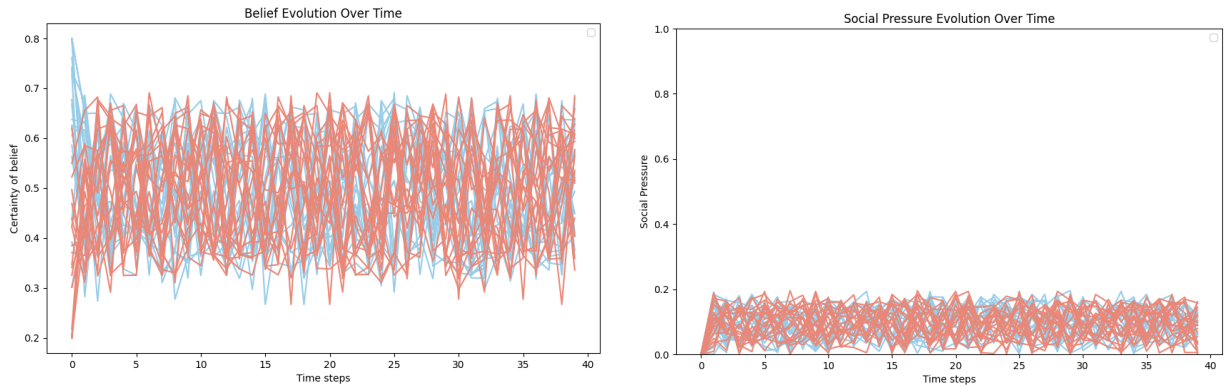


图 8. 信念（左）和社交压力（右）在大型组件中来自每个极化组的前 20 个和最后 20 个代理随时间的变化。

个 5 证据上为 0.2 生成的。一个具有两个明显等大小社区结构的网络被构建。

模型随后生成了每个时间步的信念分布、网络可视化和社会压力。在网络可视化中，节点大小表示代理与陈述的一致水平，就像之前一样，红色和蓝色用于分别表示两个极化群体。我们还展示了社会压力网络可视化，其理念相同。

通过观察图 7 中的信念分布网络可视化，我们可以看到最初蓝色社区比红色社区更倾向于相信该陈述，其中蓝色和红色分别代表倾向证据正反两面的社区。在分布收敛后，每个社区的信念动态开始波动，而整体信念分布保持不变。在每个社区中，相信该陈述的人数与不相信的人数大致相等。这一点也可以从图 5 和图 6 中的信念演化和社会压力演化图表中反映出来。适度收敛之后，个体的信念开始与社会压力不同步地振荡。这与我们对社会压力是信念变化驱动因素之一的理解一致。



图 9. 初始和最终信念网络可视化，用于理解的极化结构及在网络中两个社区上的初始置信水平。

图 8 显示了信念分布达到平衡后网络中的社会压力分布。较大的社会压力位置似乎对连接器节点和聚类节点没有偏好。



图 10. 初始和最终社会压力网络可视化。

最初，极化存在于不同的社区中，随着时间的推移，它进入每个社区。之前的研究表明，在美国政治中确实存在党内极化 [2]。证据传播可能是原因之一。在图 6 中，社会压力分布显示出下降趋势，社会压力谱较低端的人口比例比随机生成的情况更多，也有少数人处于较高的水平。

2. 具有巨大组件

理解的极化结构和对证据的初始信心水平以与之前相同的方式构建，但这次是在具有巨大组件的网络上实现的。

在网络可视化中，图 11，我们仍然用红色和蓝色分别表示两个社区，但这次是在一个包含巨大组件的网络内部随机分布。

根据图 9 和 10，收敛效应依然存在，比随机初始弱化弱，但比具有社区结构的极化网络强。

尽管社会压力分布与具有社区结构的分布 0.2 有相似的最大值，但我们看到的是，在高压下呈上升趋势而不是下降趋势。

巨型组件内的代理具有一种更加均匀的信念分布，但它比具有社区结构的网络的社会压力要高得多。

这可能是解释人们倾向于形成社区而非一个巨大的组件以减少紧张的原因。

进一步的偏振分析如图 14 所示。

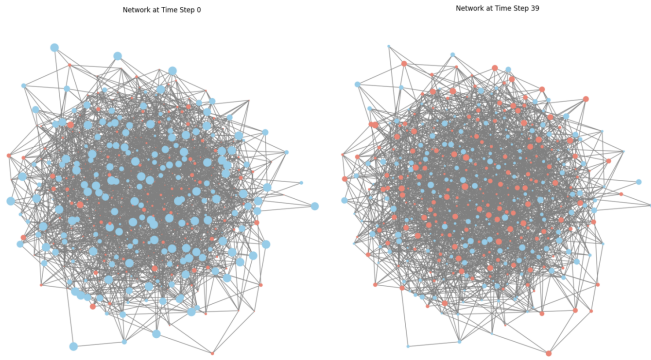


图 11. 初始和最终信念网络的可视化表示对于极化理解结构以及初始信心水平与之前相同但位于一个巨型组件内。

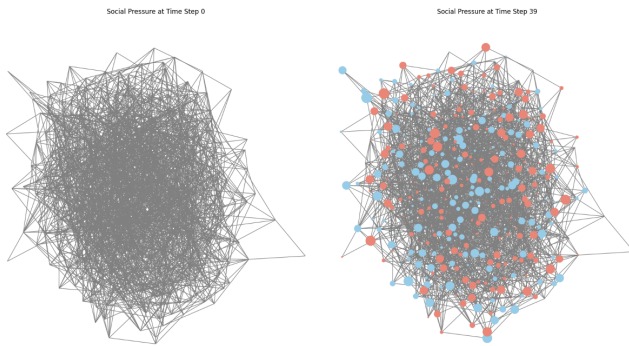


图 12. 初始和最终社会压力网络可视化。

IV. 模型参数的进一步分析

我们还研究了连接性指数、种群大小和极化指数在不同网络中如何改变均衡信念分布的变化。

根据图 13, 连通性与连接指数大于 10 的信念分布的标准差无关 (因为对于 $k < 10$, 巨型组件无法覆盖足够比例的人口, 导致孤立的代理显著增加方差), 并且随着人口规模的增加, 信念分布的标准差减小。这是一个令人惊讶的现象, 因为在不同人口规模中网络的局部结构是相同的。如预期的那样, 信念分布的标准差随信心水平的增加而增加。

为了观察不同社区结构中不同的极化程度如何影响信念分布, 我们进行了实验, 结果如图 14 所示。置信水平越高, p 或 $1 - p$ 越大, 信念分布的标准差也就越大。当自我置信度为 0 时, 会达成共识, 因为模型简化成了 Degrootian, 标准差简单地变为零。当置信水平是 1 时, 网络结构不会基于极化影响个人的决策,

因为两条曲线的形状看起来相同。在极化指数 $p = 0.5$ 下, 两种图表似乎在不同的自我置信度水平上共享相同的值, 这表明没有极化。当极化指数达到极端时, 具有社区结构的网络对信心水平的变化比具有巨大组件的网络不那么敏感。

通过引入理解的结构, 我们还可以调查麦迪逊在联邦党人文集第 10 篇中提到的解决派系祸害的方法的有效性, 即让声明成为对政府的支持程度, 证据则是公民可以关心以支持政府的问题。人们对政府的信心越普遍, 各派系之间的冲突就越少。在图 15 中, 一个人可以考虑其信仰的证据多样性的增加导致共识更有效地达成。之前的模型已经证明了这一结论 [11]。

V. 讨论

通过将理解和证据的结构概念引入 Friedkin-Johnsen 模型, 我们能够进一步探究信念传播背后的认知结构。在数值实验中, 探索了两种网络结构, 一种具有大型组件, 另一种具有两个社区。在一个极化的群体中, 对于一个具有大型组件的网络而言, 其信念分布的标准差比具有两个社区结构的网络要小, 但前者的社会压力大于后者。在一个由随机生成的理解结构和证据信心水平构成的大组件网络中, 人口规模越大, 信念分布的变化越小。连通性指数不影响信念分布的变化。

对自身理解结构的影响力也可以成为研究对象。一种更新理解结构的方法可以通过类似 Axelrod 的文化模型中的过程, 即在一定时间步长后, 每个人都会与其邻居交换一小部分他们的理解结构。另一种方法可能是改变人们理解结构最重要的方式, 即通过来自环境的压力使个人经历一个强化学习的过程。这种压力可以是社会压力, 或者可以开发出一种新的功能环境与代理的理解结构进行交互。

社交网络中的权重也可能由于认知失调而发生变化。一种允许社交网络权重根据认知失调变化的模型 [1] 已经被开发出来, 该模型可以与本模型潜在地结合使用。

当前我们正在考虑独立的证据, 这意味着接收一个证据对其他证据的确信度没有影响。模型的一个未来改进可以是每个个体视为一个多层神经网络而不是单一层次, 并且每条证据都是其自身的声明。神经

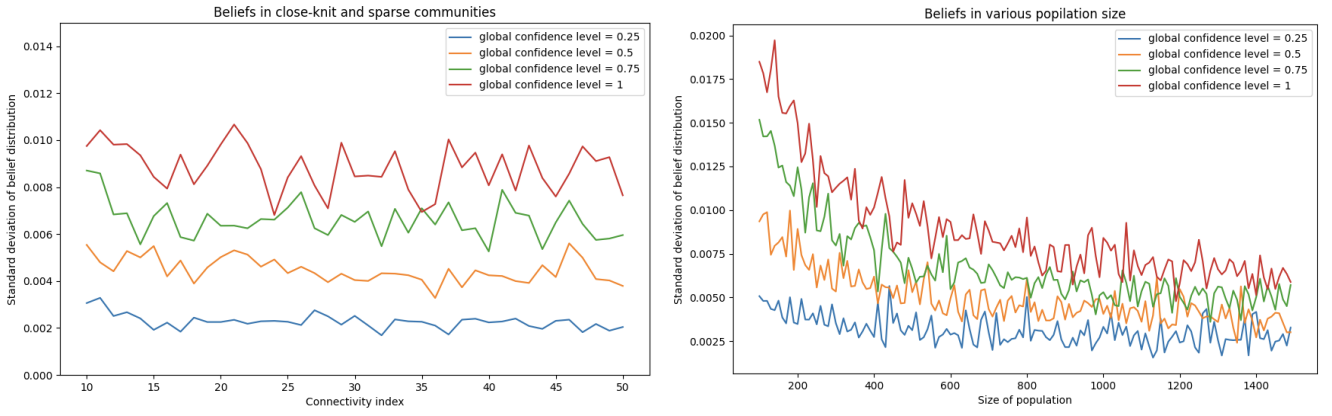


图 13. 标准偏差与连通性指数 k 的分布图（左图）以及在不同自信心水平下，结构理解的随机初始化时，标准偏差与人口规模的关系图（右图），其中连通性为 $k = 10$ ，针对 400 人的群体。

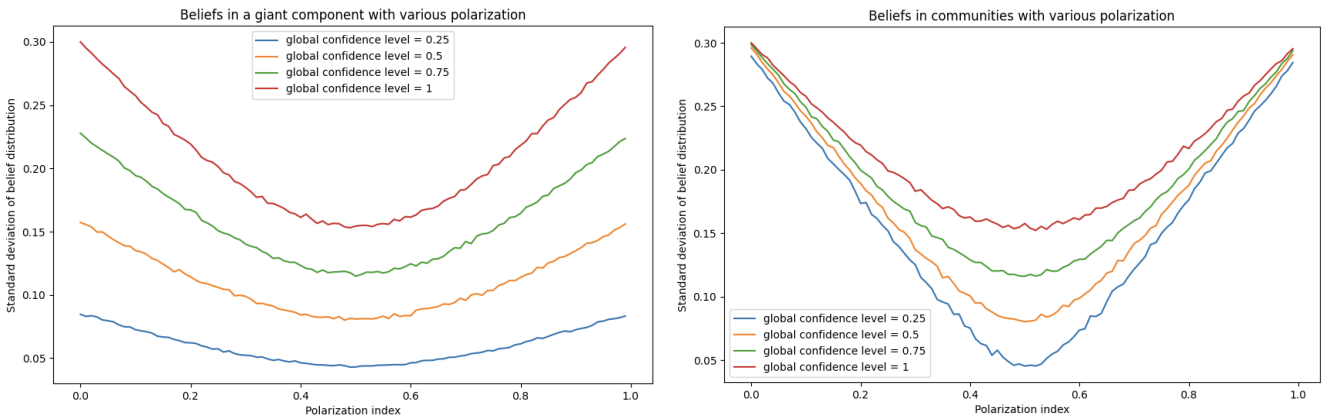


图 14. 标准偏差的信念分布与极化指数 p 的关系图，针对具有巨大组件（左图）和具有社区结构（右图）的极化群体，人口规模为 400，连接性指数 $k = 10$ 在不同自信水平上的关系如上所示。

网络越深，信念的分辨率越高。也可以实现多条声明而非单条声明的传播。

我也想感谢 mesa 包的开发者，该包用于 python 中的基于代理的模拟。[10]

ACKNOWLEDGMENTS

我要感谢密歇根大学数学系提供我在 cmplxsys 270 进行这项研究的机会。

你可以通过克隆以下仓库来访问该项目的代码。

附录 A: GitHub 仓库

<https://github.com/jeffyujiangfu/BPN.git>

- [1] M. Galesic. Integrating social and cognitive aspects of belief dynamics: towards a unifying framework. *J. R. Soc.*, 2021.
- [2] E. Groenendyk. Intraparty polarization in american politics. *Journal of Politics*, 82(4):1616–1620, 2020.

- [3] N. E. Friedkin. Network science on belief system dynamics under logic constraints. *Science*, 354(6310), 2016.
- [4] M. Vlasceanu. A network approach to investigate the dynamics of individual and collective beliefs: advances and applications of the bending model. *Perspectives on Psychological Science*, 2024.

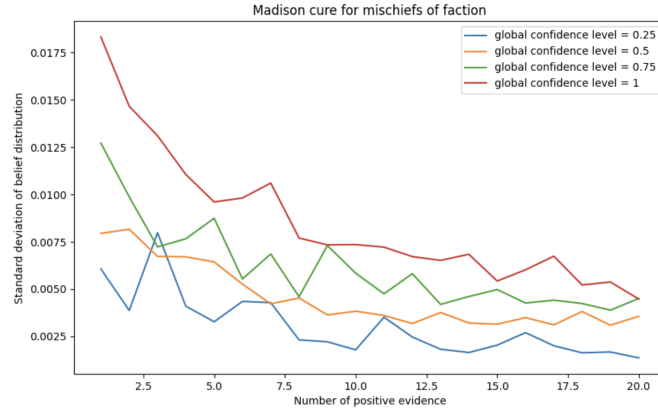


图 15. 标准偏差与正证据数量的分布图 m 对于一个 400 人的群体，在理解结构的随机初始化下，不同自我信心水平下的连接性指数 $k = 10$ 如上所示。

- [5] P. Grim. Scientific theories as bayesian nets: Structure and evidence sensitivity. *Philosophy of Science*, 2021.
- [6] P. Dandekar, A. Goel, and D. T. Lee. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, 110(15):5791–5796, 2013.
- [7] N. Friedkin and E. Johnsen. Social influence and opinions. *Journal of Mathematical Sociology - J MATH SOCIOLOGY*, 15, 1990.
- [8] N. Friedkin and E. Johnsen. Social influence networks and opinion change. *Advances in Group Processes*, 16, 1999.
- [9] D. P. W. Freeborn. Rational factionalization for agents with probabilistically related beliefs. 2024. <https://philsci-archive.pitt.edu/22907/>.
- [10] J. Kazil, D. Masad, and A. Crooks. Utilizing python for agent-based modeling: The mesa framework. In R. Thomson, H. Bisgin, C. Dancy, A. Hyder, and M. Husain, editors, *Social, Cultural, and Behavioral Modeling*, pages 308–317. Springer International Publishing, Cham, 2020.
- [11] M. Kawakatsu, Y. Lelkes, S. A. Levin, and C. E. Tarnita. Interindividual cooperation mediated by partisanship complicates madison’s cure for “mischiefs of faction”. *Proceedings of the National Academy of Sciences*, 118(50):e2102148118, 2021.