

观点的涟漪：基于社会判断理论的大语言模型智能体说服传播模拟

摘要：大语言模型对于社会现象模拟具有重要潜力。本研究结合社会判断理论（Social Judgment Theory, SJT）与观点动力学，利用大语言模型驱动的智能体模拟方法，探究社交网络中群体说服传播态度演化规律及其影响因素。研究借助 AgentSociety 社会模拟平台，构建基于 SJT 的智能体模型，通过连续观点交互机制模拟个体在不同争议程度议题中的态度变化，并分析网络结构、宏观社会意见及微观心理因素的作用。实验结果表明：（1）群体态度在互动中呈现显著趋同趋势，最终形成一致支持或中立意见；（2）微观层面，负面情绪是驱动态度变化的关键因素，个体状态与积极情绪影响较弱；（3）网络拓扑结构显著影响态度演化速度，BA 无标度网络因枢纽节点存在加速意见收敛，而 WS 小世界网络因局部聚集性延缓传播进程；（4）宏观社会意见塑造群体观点演化方向。本研究将传播学理论与 LLM 仿真相结合，提出多层级说服传播模型，揭示了网络结构与情绪机制在观点演化中的作用，为政府舆情治理和舆论引导提供了理论依据。

关键词：社会判断理论；观点动力学；大语言模型；说服传播；情绪效应

一、引言

从“您认为每个人都应该养成定期锻炼的习惯吗？”和“您如何看待高校扩招？”到“您是否支持以成绩为标准划分重点班和普通班？”。这些不同争议程度话题促使人们采取不同修辞策略，进而说服他人。探究说服发生机制成为一个重要议题。

说服是什么？简单来说，说服就是说服者使用一些信息去影响接收者态度或者观点，以此来实现自身目标和期望。人类社会中存在各种各样说服过程。但是说服成功与否往往受到接收者个人经验、已有知识和新获得信息的影响。个体对信息的加工最终呈现为个人观点，这些观点作为新信息加入群体观点互动，最终经过一系列复杂交互，群体状态从无序走向有序^[1]。说服就是群体信息交流和观点互动过程。

社会个体之间的相互影响和作用关系构成社会网络。传统 ABM 方法需要人

工设定交互规则，采用虚拟数据进行试验，难以捕捉规则之外的变化。通过海量数据训练的大语言模型（LLM）能够结合真实世界环境，打破人为规则的束缚，实现更加真实的类人交流。

探究社交网络舆论规律，实现国家和政府对公众的正确引导和说服，需要坚实的传播系统。本文结合社会判断理论^[3]，对社交网络说服过程进行建模分析；基于 AgentSociety 社会模拟平台，构建基于说服传播智能体，考察说服传播演化过程及规律。通过对社会现象高度抽象，探究社会网络中群体观点和行为演化，揭示其背后蕴含的一般规律，为社会决策和政府治理提供理论依据。

二、文献综述

（一）说服传播理论发展

传播学中许多学者都将说服研究历史上溯至古希腊时期。亚里士多德兼顾柏拉图和智者之间的争论，综合两种思想流派优点，将修辞学看作理解说服力的工具，这些思想的影响力延续至今，对说服乃至整个传播学研究产生巨大影响。

亚里士多德并未给出说服的完整定义。后来的传播学者们在长期研究过程中认为，说服是一个象征性过程，在这个过程中，传者试图通过传递信息来改变受众对某个问题的态度和行为^[4]。该定义强调受众接收、认知、处理信息以及最终行为，表明说服包含四个基本过程。第一，说服者需要明确自身需求和目标；第二，明确能够实现目标的人群，即合适的传播受众；第三，了解受众特点，构建传播信息；第四，将信息传递给受众。

在现代历史中，说服研究起源于一战和二战期间的战时宣传研究以及二战之后研究的制度化。这些研究开创了传播学和心理学跨学科研究先河，为后续研究奠定了方法程序基础。其中最著名的就是卡尔·霍夫兰在耶鲁大学主导的一系列说服传播研究。

传播学者在说服研究方面取得了长足的进步。霍夫兰的 Yale 研究提出一系列识别说服传播影响因素和过程的方法和程序，为之后研究打下坚实基础。但是未能揭示说服发生的具体过程。Osgood 和 Tannenbaum 的一致性理论（Congruity Theory^[5]）认为人们喜欢现实世界与自身认知的一致与和谐，不一致想法带来的压力会成为态度改变的动力。认知失调理论（Cognitive Dissonance Theory）则同

时考虑多种认知，将它们之间的关系划分为一致、不一致和无关三种，认为人们可以通过改变现有认知、增加新认知和改变认知重要性排序来减少失调，恢复自身和谐状态^[6]。信息整合理论（Information Integration Theory）^[7]认为新态度的产生往往是对原有态度和新信息混合、组合以及整合的结果。Petty 和 Cacioppo 提出的精细加工可能性模型（Elaboration Likelihood Model, ELM）认为，存在中心和边缘两条说服“路线”^[8]。

与以往理论不同，由 Sherif 和 Hovland 所提出的社会判断理论（Social Judgment Theory^[3]）解释了说服发生的过程和内在机制。他们认为受众在接受信息的瞬间，会将其与自身当前态度进行比较，从而对信息进行感知和评价，确定它在态度量表中的位置，最终决定是否以及在多大程度上被说服。社会判断理论不仅详细说明了观点交互规则，并且指出了感知观点和认知偏差概念。

（二）说服传播与基于大语言模型仿真研究

说服传播研究发展至今，学者们已经从多种角度进行了探讨。本小节将从影响因素，模型拓展方面对现有文献进行梳理，同时介绍基于大语言模型仿真方法的应用研究。

（1）说服传播影响因素研究

信源方面。信源指发出信息的源头。传统研究主要将源头看作和受众同样的人类。Wagenknecht 等研究了信源匿名性如何影响在线讨论中的沟通说服力^[9]。具体地，他们发现感知匿名（perceived anonymity）对感知社会存在（perceived social presence）具有边际和负面影响，并通过感知社会存在对感知用户可信度产生影响。感知用户可信度和个人参与同时受到感知社会存在影响，二者又对沟通说服力（communication persuasiveness）具有相反影响。随着各种互联网广告平台的兴起，信源概念被拓展到信息发布场所。Drossos 等研究了说服技巧如何促使人们进行网络购物^[10]。他们的实验结果表明，社会证明（social proof）和稀缺性说服技巧会影响用户行为并促进购物者点击“添加到购物车”按钮。王烨娣等以互联网个性推荐广告为背景，探讨了互联网平台可信度、调节聚焦和隐私显著性对广告最终效果的影响^[11]。

信息方面。对信息本身的研究吸引了大量学者。首先是对信息内容的探究，

Tal 和 Wansink 研究了科学信息对说服效果的影响^[12]。他们的研究表明图表和化学符号可以增强说服力，并且对科学有更大信仰的人更容易受图表影响。Dragicevic 和 Jansen 重新审视了这个结果^[13]，提出图表对理解的贡献大于对说服的贡献，并且带有图表的文本有时甚至比单独的文本具有更低说服力。除了文本，有学者使用机器学习方法对喜马拉雅 FM 上声音信息进行分析，探究了声音带来的温暖感、能力和权威感以及辅助文本信息如何影响受众购买意愿^[14]。

其他学者则从**信息框架**角度对说服效果进行分析。通过对不同类型的内容信息进行分析，Sung 和 Lee 发现持有负面态度的个体较少受到他人态度影响，并且这些群体更容易接受相对客观的信息（同时展示事件的好处和坏处）^[15]。Nera 等通过对照实验，从实证角度验证了阴谋论叙事对现实世界观念的影响^[16]。借助三种不同的说服理论框架，Scannell 等分析了 Twitter 用户对 COVID-19 疫苗的情绪以及其说服技巧，发现人类和机器人在反对和支持疫苗时都会采用不同的叙事框架，进而提出了解决错误和虚假信息的应对框架^[17]。刘楠等研究不同文本框架及时间距离对受众说服效果的影响。在“现在”情景中，消极目标框架通过启动更强的回避动机，从而产生更强的说服效果；在“未来”情景中，积极目标框架通过激发更强的趋近动机，从而具有更大说服力^[18]。

受众方面。结合社会判断理论，Aiello 等使用访谈和多种实验混合的方法，探究了受众在不同阶段的感知温暖如何影响其展示个人信息的意愿，进而促进在线消费进程^[19]。Shu 等结合社会判断理论以及 Cialdini 的说服模型^[20]，研究了促进信任的信息在多大程度上可以改变消费者对使用在线信用卡的态度^[21]。他们发现，对于具有消极态度（拒绝使用在线信用卡）的消费者而言，Authority-plus-contrast-plus-scarcity 比 Authority-plus-contrast 和 Authority-only 更能促使消费者转变为积极态度，不同的消息来源（国内和国外）对消费者态度改变的影响没有显著差异。

社会文化方面。随着说服研究的发展，研究者开始将视角从微观层面转移到宏观层面。通过对荷兰不同民族人群的研究，Li 等发现文化背景在说服传播中起到了调节作用。具体地，自我说服和直接说服分别对拥有自由主义和集体主义文化背景人群更有效^[22]。在国内，杨莉明和徐智通过分析中国的中老年群体对防疫政策接受程度，发现了垂直集体主义价值观对防疫行为的解释力要高于其余影响

因素，揭示了文化观念对说服效果的影响力^[23]。

上述研究大多考虑单一因素对说服传播的影响，考虑过程难免显得单一。因此，有学者开始将不同影响因素相互联系，同时考虑多种因素的作用。通过增加说服情境感知互动性变量，曾丽红和黄蝶探究了信源、信息以及受众心理微观因素对受众购买意愿的影响^[24]。有学者从数量、内容和信源三个维度对新浪微博用户信息接触行为进行分析，总结了接触数量和说服效果的关系，提出了信息的“接触—采纳”曲线；发现接触反对信息对疫苗采纳结果没有显著影响，专业型和同质型信源对受众有不同的说服效果^[25]。通过采用双面阐释或者唤起受众情感价值的说服策略，可以更加有效劝说受众^[26]。同时，说服信息的框架能够对受众的态度和意愿有显著影响，社会规范和信源可信度在不同的信息框架中起到不同的调节作用^[27]。社会关系、个人感知以及风险披露也会对说服效果产生影响^[28]。

（2）说服传播模型拓展和理论验证研究

上述研究均为说服理论和模型应用研究，此处对拓展研究进行梳理。Bergier 和 Faucher 基于 ELM 模型，提出了一个信息处理模型^[29]。该模型将消息吸引程度，受众处理信息的能力、意愿、动机以及精细化程度纳入考虑范围，得出信息说服力。如果信息说服力高于下限，该信息将被接受，否则被拒绝。Hall 和 Oppenheimer 基于社会判断理论模型开发了错误分析模型，并通过经验数据对该模型进行了验证^[30]。

除此之外，有学者通过实证研究验证了理论模型。Moura 等对博客和 Facebook 页面上与糖相关的帖子和评论进行专题分析，并对专家进行深度访谈，以此来研究父母对糖的认知和态度^[31]。他们发现在处理关于禁欲主义和享乐主义矛盾时，丹麦和法国父母表现出三种主要行为：拒绝，温和以及开放，分别对应社会判断理论中的拒绝域，中立域和接受域。

（3）基于大语言模型仿真研究

近年来，大语言模型作为自然语言处理领域的重要突破，正被越来越多地用于仿真研究中。在传统社会行为仿真中，智能体通常由规则驱动、方程或机器学习算法构建^[32-34]。这类智能体虽能完成基本任务，但由于缺乏对语言语境和复杂认知过程的理解能力，其行为难以逼近真实的人类表现^[35]。

通过获取海量网络知识，大语言模型展现出接近人类智能水平的潜力，为构建“类人智能体”提供了新的路径^[36]。例如有学者基于大语言模型构建智能体，用于宏观经济仿真。结果表明，由大语言模型赋能的智能体，能够做出更为逼真的决策，所呈现的宏观经济现象也比现有的基于规则或基于学习的智能体更为合理^[37]。Gao 等利用大语言模型构建了一个社会网络仿真系统，通过模拟智能体的情绪、态度与互动行为，实现了信息、情绪和态度传播等群体级社会现象的再现^[38]。另外，Park 等通过大语言模型驱动的生成式智能体，实现了对人类行为（如计划、记忆和反思）的类人模拟，并在互动环境中展现出可信的个体与群体行为，从而用于社会行为仿真^[39]。Piao 等通过整合大语言模型驱动的智能体与复杂环境，模拟了超过一万名智能体的五百万次互动，并用于研究社会极化、虚假信息传播、基本收入政策与自然灾害等社会议题^[35]。

综上所述，本研究关注观点表达主体之间的联系与互动，同时考虑外部因素，从微观、中观和宏观层面探讨说服效果产生、演化规律及其作用机制。具体研究问题如下：

问题 1：群体态度如何交互？

问题 2：态度交互结果以及影响因素是什么？

问题 3：宏观社会意见如何影响群体观点演化？

三、研究设计

（一）理论基础——社会判断理论

社会判断理论（Social Judgment Theory, SJT）^[3]由 Sherif 和 Hovland 所提出，用于研究说服发生的过程和内在机制。该理论认为受众在接受信息的瞬间，会将其与自身当前态度进行比较，从而对信息进行感知和评价，确定它在态度量表中的位置，最终决定是否以及在多大程度上被说服。具体而言，该理论有以下假设：

假设 1：个体观点可以被划分成三个区域，包括接受域（latitude of acceptance），中立域（latitude of non-commitment）和拒绝域（latitude of rejection）。如图 1 所示，图中锚点表示接收者自身观点位置，锚点两侧是其接受域，接受域两侧是两个中立域，观点轴最外侧为拒绝域。

假设 2：只有当说服者观点落在接收者接受域和中立域时，这些观点才会有

说服效果，并且距离接收者观点越远，其观点变动幅度越大，说服效果越好。但是当说服者观点距离接收者观点过远，以至于落入其拒绝域中时，则没有说服效果。

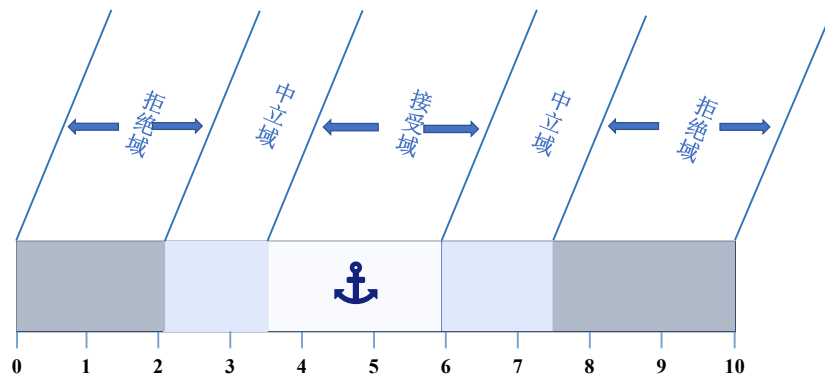


图 1. 受众观点区域划分

假设 3：个体对他人观点的感知和判断存在两种知觉扭曲效应——同化（assimilation）和对比（contrast）。同化效应会导致接收者感知的说服者观点比实际更接近自身观点，即当说服者观点落在接收者接受域内时，二者观点比实际更加相似。相反，对比效应则会导致落入拒绝域的观点相互远离。

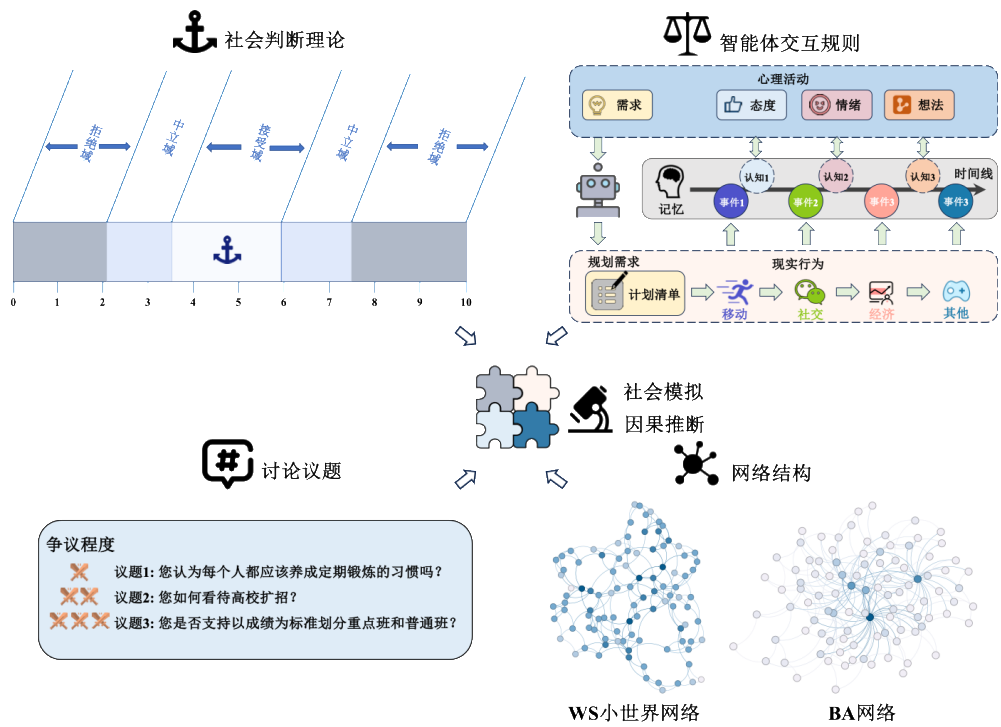


图 2. 整体研究框架

本研究将说服传播过程看作是社会公众通过互动产生观点和态度变化趋势、过程以及结果，基于结合社会判断理论，构建说服传播理论模型，借助 AgentSociety 平台进行 LLM 社会模拟仿真，最后结合因果推断方法，考察个体微观因素对态度演化的影响。具体研究框架如图 2 所示。

（二）实验设置

本研究根据智能体意见锚点间的距离来规定劝说落入哪一邻域，设定接受域 $U_\alpha \in (0, 2]$ ，若 $0 < |A_{it} - A_{jt}| \leq 2$ ，则观点落入接受域；设定中立域 $U_n \in (2, 4]$ ，若 $2 < |A_{it} - A_{jt}| \leq 4$ ，则观点落入中立域；设定中立域 $U_r \in (4, +\infty)$ ，若 $|A_{it} - A_{jt}| > 4$ ，则落入拒绝域。符号定义见表 1。

表 1. 模型符号定义

符号	意义
A_{it}	个体 <i>i</i> 在 <i>t</i> 时刻的态度
U_α	个体接受域宽度
U_r	个体中立域宽度
$ \cdot $	态度差值

大语言模型训练依赖于大规模社会语料，其生成内容在一定程度上反映了社会整体的语义分布与舆论偏向【StereoSet: Measuring stereotypical bias in pretrained language models】。本研究选择了“您认为每个人都应该养成定期锻炼的习惯吗？”（基本无争议）、“您如何看待高校扩招？”（争议程度中等）与“您是否支持以成绩为标准划分重点班和普通班？”（争议程度高）三个争议度不同的主题进行模拟，考察宏观社会意见对说服传播的影响。其次，本研究构建 WS 小世界网络和 BA 无标度网络，网络参数如表 2，考察中观网络结构的影响。

表 2. 两种社交网络基本参数

网络	智能体数量	社交关系总数	平均聚类系数
BA	30	56	0.216
WS	30	60	0.159

最后，本文采用双向固定效应，探究影响态度变化的微观因素。具体公式如下：

$$Attitude_{it} = \beta Individual_status_{it} + \delta Emotion_{it} + \alpha_i + \gamma_t + \varepsilon_{it} \quad (1)$$

其中， $Attitude_{it}$ 表示第 i 个智能体在第 t 个时间步的态度值。 $Individual_status_{it}$ 包括饥饿、疲惫、安全感、社交多种个体状态变量。 $Emotion_{it}$ 包括悲伤、愉悦、害怕、厌恶、愤怒和惊讶多种情绪状态变量。 α_i 为个体固定效应，控制个体自身特有且不随时间变化的因素。 γ_t 是时间固定效应，控制所有个体共同经历的时间趋势和外部冲击。 ε_{it} 是误差项，捕捉无法解释的随机误差或其他微小干扰因素。

四、研究结果

（一）态度交互结果

如图 3 所示，在不同网络拓扑结构中，智能体在经过模拟交互后，态度均发生了演变，最终都出现了收敛，在同一话题下均呈现出支持的态度。这体现出群体在网络交互过程中，意见逐渐趋同，原本分散的态度，通过网络连接互动，形成了一致的群体意见，展现出网络结构里群体意见演变从多元到统一的过程。

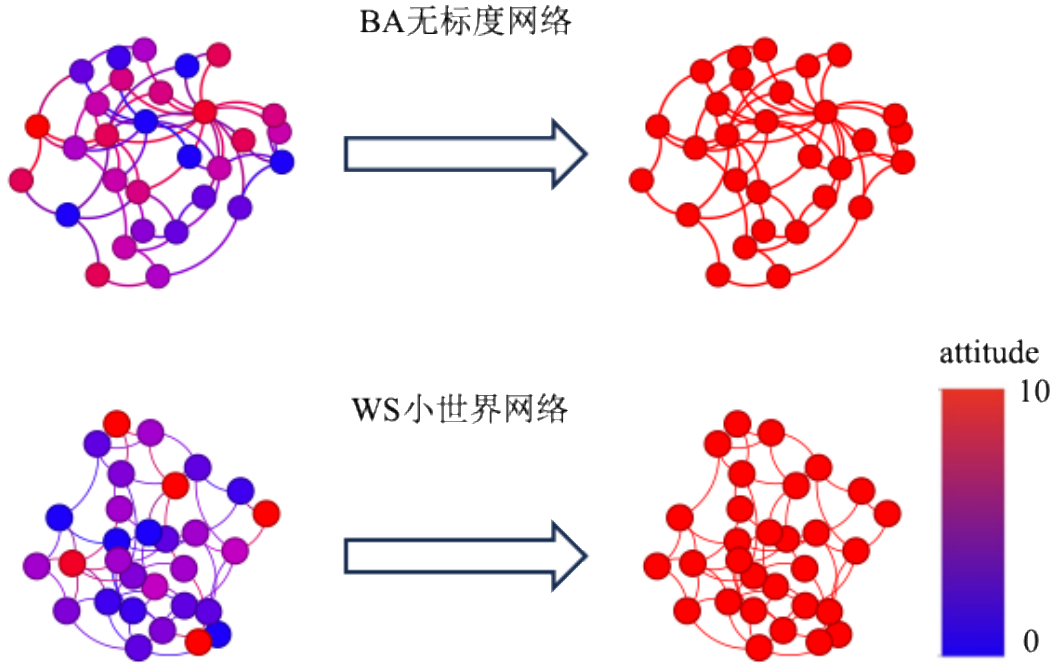


图 3. 群体意见演变结果

劝说内容能够在一定程度上反映出个体的态度。图 4(a)、(b)和(c)分别是根据智能体在“您认为每个人都应该养成定期锻炼的习惯吗？”、“您如何看待高校扩招？”和“您是否支持以成绩为标准划分重点班和普通班？”三个议题中的劝说内容绘制出的词云图。



图 4. 智能体在不同议题下的劝说内容词云图。(a) “您认为每个人都应该养成定期锻炼的习惯吗？”；(b) “您如何看待高校扩招？”；(c) “您是否支持以成绩为标准划分重点班和普通班？”。

在子图(a)中，词云直观地呈现出智能体在该议题下态度高度一致，普遍支持锻炼的积极价值。诸如 **mental**、**wellbeing** 等词语凸显了智能体对于锻炼有助于改善心理状态、促进身心健康的广泛共识；**boost**、**improve** 等词进一步反映出锻炼在提升精力和生活质量方面所获得的认可。此外，从 **essential** 与 **crucial** 等词中可以看出，智能体普遍将定期锻炼视为维系健康、拥抱积极生活方式的重要习惯。

在子图(b)中，词云显示出智能体普遍支持高校扩招。在个体层面，**provide**、**enhance** 与 **opportunities** 等关键词表明智能体认为扩招有助于拓宽个体的发展路径，提升教育可及性与机会平等；在社会层面，**contribute**、**workforce**、**societal**、**economic**、**growth** 与 **innovation** 等词汇则体现了智能体在高校扩招对社会经济发展和创新潜力方面达成了积极共识。

在子图(c)中，词云反映了智能体在“是否应以成绩为标准划分重点班与普通班”这一议题上出现了分歧。支持方频繁使用 **resource**、**effectively**、**tailored** 与 **elite** 等词，表明他们认为按成绩分班有助于教育资源的精准配置与个性化教学，从而提升整体教育效果；而反对方常用 **confidence**、**growth** 与 **potentially** 等词，则体现出其对分班可能带来的学生自信心受损与个人成长受限的担忧。

（二）影响态度变化的因素

1. 宏观社会意见显著影响态度演变方向

为探讨宏观社会意见对群体态度演化的影响，本研究选取三个社会争议程度不同的议题，结合 WS 小世界网络和 BA 无标度网络开展社会模拟。图 5 显示，在不同议题与网络结构下的态度演变存在差异：在“您认为每个人都应该养成定期锻炼的习惯吗？”以及“您如何看待高校扩招？”两个议题中，两种网络群体态度均呈现趋同，最终形成较为一致支持意见（图 a-d 和图 b-e）；在“您是否支持以成绩为标准划分重点班和普通班？”议题中，尽管智能体态度同样表现出一定程度的趋同，但相比前两个议题，最终态度更多集中于中立区域（图 c-f）。

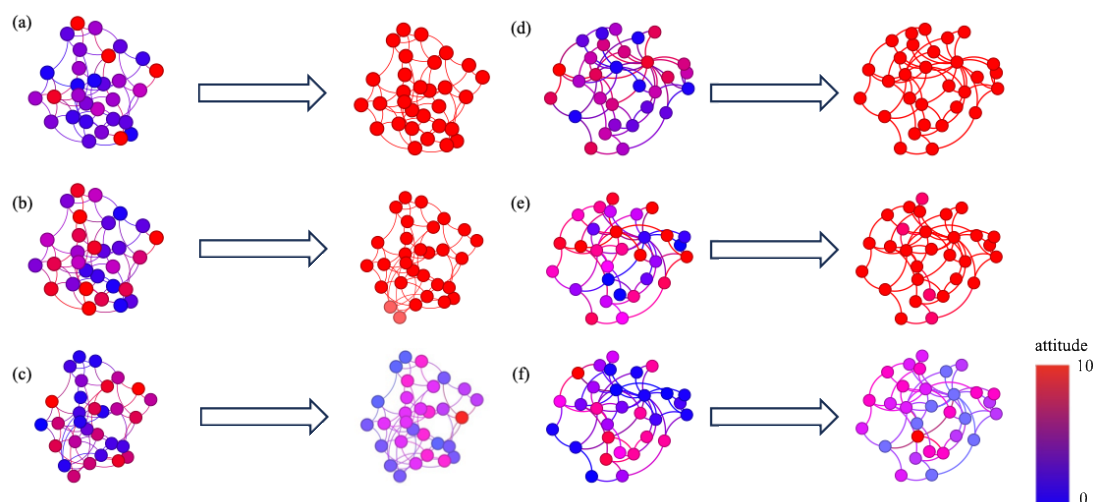


图 5. 智能体在不同议题、不同网络拓扑结构的态度演变结果。(a) “您认为每个人都应该养成定期锻炼的习惯吗？”， WS 小世界网络；(b) “您如何看待高校扩招？”， WS 小世界网络；(c) “您是否支持以成绩为标准划分重点班和普通班？”， WS 小世界网络；(d) “您认为每个人都应该养成定期锻炼的习惯吗？”， BA 无标度网络；(e) “您如何看待高校扩招？”， BA 无标度网络；(f) “您是否支持以成绩为标准划分重点班和普通班？”， BA 无标度网络。

话题争议程度根植于宏观社会意见，个体立场分歧并非孤立生成，可能受宏观舆论结构影响与塑造。大语言模型训练依赖大规模社会语料，其生成内容在一定程度上反映社会整体语义分布与舆论偏向。因此，本研究以具有不同争议程度

的话题作为宏观社会意见代理变量，以捕捉社会舆论环境对个体态度演变潜在影响。研究发现表明，宏观社会意见对群体态度演变路径产生重要影响。议题本身争议性决定了群体观点演化走向。具体而言，话题争议程度不仅调节意见传播路径与速度，也影响群体能否达成一致意见的可能性。

2. 网络拓扑结构显著影响态度演变速度

网络拓扑结构在很大程度上决定信息传播路径与效率，进而对群体态度演变速度产生显著影响。具体而言，BA 无标度网络中枢纽节点的存在使得信息能够在短时间内通过高连接度个体迅速扩散至整个网络，加快群体意见收敛过程；而 WS 小世界网络则依赖局部聚集与少量随机连接的结合，在一定程度上延缓了信息全网传播速度，使得个体态度更新具有阶段性与局部性。研究结果表明，尽管两种拓扑结构最终导向相似意见收敛方向，但在态度转变时间维度上表现出显著差异。在 BA 网络中，态度变化更迅速，系统更早达到稳定状态；而 WS 网络则呈现出更为缓慢的演化轨迹，图 6（a）和（b）所示。

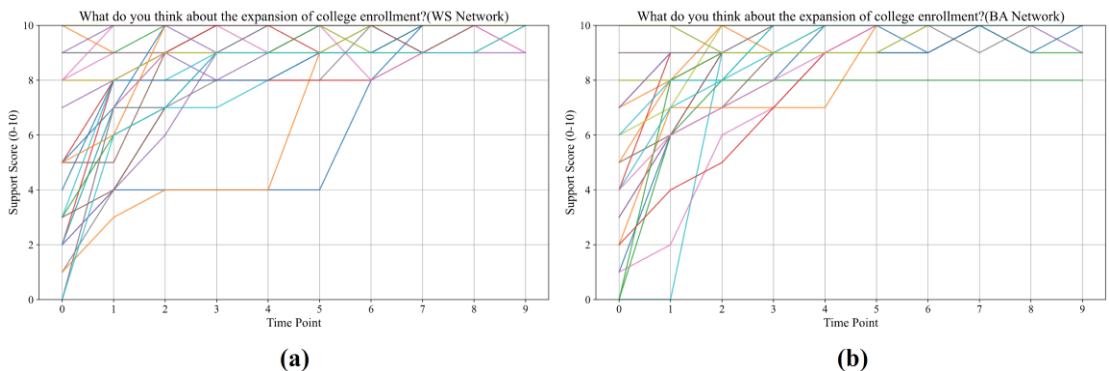


图 6. 智能体态度演变历程

3. 负面情绪是影响态度变化的重要微观因素

如表 3 所示，个体状态（如饥饿、疲惫、安全感、社交）并非影响态度变化主要因素；正向情绪对不同议题和网络结构中的态度影响不稳定，也并非驱动态度变化的核心因素。相比之下，负向情绪在多数模型中表现出显著且稳定的影响力。

表3. 影响态度交互结果的微观因素

	分班议题		高校扩招议题		锻炼议题	
	Attitude (BA)	Attitude (WS)	Attitude (BA)	Attitude (WS)	Attitude (BA)	Attitude (WS)
Individual status						
hungry	-0.010 (0.155)	-0.064 (0.145)	-0.117** (0.056)	0.017 (0.096)	-0.053 (0.040)	-0.020 (0.042)
tired	-0.059 (0.254)	0.151 (0.136)	0.020 (0.081)	0.115 (0.070)	0.018 (0.051)	-0.047 (0.062)
safe	-1.020* (0.502)	0.153 (0.436)	-0.153 (0.103)	0.113 (0.098)	-0.084** (0.039)	0.020 (0.044)
social	0.586 (1.846)	0.575 (1.949)	-0.209 (0.378)	-1.387 (0.948)	-0.221 (0.741)	-0.607 (0.670)
Positive emotion						
surprise	0.572 (0.416)	0.708 (0.458)	-0.028 (0.017)	0.006 (0.024)	0.171 (0.102)	-0.000 (0.025)
joy	0.155 (0.186)	-0.041 (0.215)	0.138** (0.055)	0.039 (0.087)	0.025 (0.056)	-0.045 (0.062)
Negative emotion						
sadness	0.986* (0.500)	-1.249** (0.532)	-0.605 (0.499)	-0.354 (0.643)	0.181*** (0.062)	1.665*** (0.504)
fear	-0.851** (0.328)	-1.479** (0.548)	0.643** (0.292)	-0.047 (0.269)	0.584*** (0.160)	0.084 (0.365)
disgust	-1.227*** (0.339)	0.121 (0.415)	0.721 (0.540)	-0.044 (1.106)	-0.071 (0.092)	-2.941*** (0.725)
anger	0.841** (0.314)	1.505* (0.739)	-1.175*** (0.280)	-1.591 (1.376)	-0.847*** (0.211)	1.662*** (0.511)
Constant	4.508*** (0.589)	6.526*** (0.999)	8.880*** (0.480)	9.630*** (0.845)	9.515*** (0.242)	10.118*** (0.472)
Observations	1,260	1,260	1,260	1,260	1,260	1,260
R-squared	0.648	0.594	0.579	0.622	0.770	0.785
Agentid FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes

注释：回归使用个体层面稳健标准误。*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ 。Agentid FE是指 agent 个体固定效应。Time FE是指时间固定效应。

五、结论与讨论

（一）结论

说服研究涉及宏观（社会）、中观（群体）和微观（个体）三个层面，需要

能够联系三个层面的理论和方法工具。基于 LLM 和社会判断理论，本文探究并分析了宏观、中观和微观三个层面因素对公众说服传播效果的影响。研究结论如下：

结论一：LLM 社会模拟能够结合传播学理论做出观点演化的有效仿真，能够挖掘说服传播一般规律和影响因素，具有现实指导和预测功能。

本文首先将智能体对特定议题的态度界定为对某一对象所持的情感、行为和认知成分的综合评价[Empirical validation of affect, behavior, and cognition as distinct components of attitude]，将其量化为 0 至 10 的数值区间；随后，基于社会判断理论构建智能体模型，梳理其在各态度区间的观点交互机制，并通过编写交互内容生成提示词以实现群体间信息互动；最后，依据智能体记忆模块中存储的情感、行为与认知信息，定期更新群体态度和状态，使智能体能够基于更新后的态度参与后续群体交互。

基于上述设计，研究结果表明，基于 LLM 社会模拟不仅能够捕捉不同议题中观点演化动态轨迹，还能有效验证传播学理论在多主体互动环境下的适用性与解释力。更重要的是，社会模拟能够揭示说服传播过程中影响说服效果的关键因素，如宏观社会意见、中观群体网络拓扑结构和微观心理状态，并凝练出一定的普遍规律。这一发现不仅证明了模型有效性，也显示出其在现实社会中具有重要指导与预测功能，例如辅助公共议题舆情研判、优化政策传播路径、提升组织沟通策略等，从而为未来传播研究与实践提供了可操作参考框架。

结论二：宏观社会意见通过影响智能体在不同讨论议题中的立场，致使最终态度演化方向发生改变。

虽然群体观点最终均呈现出显著趋同趋势，表现为态度收敛并形成相对一致意见，但对于不同争议程度的议题而言，收敛方向并不相同。在争议程度相对低的锻炼与高校扩招议题中，群体观点一致赞成；在争议程度相对高的分班议题中，群体观点最终聚集于中立区。

这一现象表明，宏观层面社会意见通过大模型预训练知识为智能体初始立场提供了参照框架，并且塑造互动过程中的态度边界与接纳区间，影响群体意见的最终走向。当社会议题争议程度较低时，主流意见能够顺畅地向个体渗透并在群体内扩散，促成一致的积极态度；而在争议程度较高的议题中，个体间的立场差

异和信息冲突会削弱单一意见凝聚力，使群体最终趋于折中，形成中立或模糊的共识。

结论三：中观群体网络拓扑结构通过影响态度演化的速度，致使最终态度收敛所需时间存在差异。

实验结果表明，BA 无标度网络的态度收敛所需时间小于 WS 小世界网络。BA 无标度网络中的枢纽节点可能加快了意见传播和群体态度收敛；而 WS 小世界网络因路径随机性与局部聚集性并存，使得态度演化更具阶段性与时延性。

这一发现说明，在现实社会传播与群体互动中，网络结构差异可能成为影响舆论扩散效率与最终共识质量的重要因素，特别是社交网络中的关键节点在信息传播与意见整合中扮演着核心角色。枢纽节点因其较高的连接度和传播能力，能够在较短时间内影响大规模群体的态度走向；而处于桥梁位置的中介节点则在不同群体之间起到跨圈层的信息传递作用，对避免信息孤岛、促进意见融合具有重要意义。因此，在公共议题治理中，不仅需要关注整体网络结构，还应识别并合理利用这些关键节点，以提高信息传播的效率和覆盖度，增强社会共识的凝聚力。

结论四：微观负向情绪对态度变化具有稳健影响。

通过双向固定效应模型对面板数据进行分析发现，个体情绪变量中，尤其是愤怒（anger）等负向情绪显著推动态度变化，而正向情绪和个体状态变量对态度变化影响不稳定或不显著。

由此观之，负面情绪是态度极化与群体动态转变的重要心理机制，其作用不仅体现在加速个体态度转变，还可能通过情绪感染效应在群体中迅速扩散，进而推动意见极化和社会分化。由此，管理群体情绪、引导负向情绪向理性表达和建设性讨论转化，成为缓解群体极化与促进社会共识的重要路径。

（二）未来展望

本研究主要关注智能体之间基于说服内容的交互对态度的影响，尚未将智能体与环境的互动，以及除说服之外的其他社交交互因素纳入考量，这在一定程度上限制了模型对真实社会情境的再现能力。未来研究可在此基础上进一步拓展模型的多模态输入与交互维度，例如将文本、图像与语音信息共同纳入智能体的感知与推理机制中，以更贴近真实传播场景下的多渠道信息环境。

此外，LLM 社会模拟具有双面性。一方面，LLM 在预训练过程中积累了整

体层面的知识，能够赋予智能体更类人的决策、沟通和适应能力[Large language models empowered agent-based modeling and simulation: a survey and perspectives]; 另一方面，这种预训练知识也可能嵌入偏见，干扰实验结果[Simulating Opinion Dynamics with Networks of LLM-based Agents; Systematic Biases in LLM Simulations of Debates]，并导致智能体在对初始化态度的理解中难以摆脱既有立场影响。因此，LLM 社会模拟需要谨慎解读在涉及特定领域时的表现。未来研究可以尝试通过结合领域知识对大模型进行微调，以提升结果有效性和针对性。综上所述，LLM 为社会模拟赋予了更强表现力与拟人化特征，但也带来了中立性、可解释性和适用性上的挑战。未来研究需要在发挥优势与规避局限之间寻求平衡。

1 参考文献

- [1] DONG Y, ZHANG H. Multiperson decision making with different preference representation structures: A direct consensus framework and its properties[J]. Knowledge-Based Systems, 2014, 58: 45-57.
- [2] 王龙, 田野, 杜金铭. 社会网络上的观念动力学[J]. 中国科学:信息科学, 2018, 48(01): 3-23.
- [3] SHERIF M, HOVLAND C I. Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change[M]. Westport, CT: Praeger, 1980.
- [4] PERLOFF R. The Dynamics of Persuasion: Communication and Attitudes in the Twenty-First Century[M]. 6 版. New York: Routledge, 2017.
- [5] OSGOOD C E, TANNENBAUM P H. The principle of congruity in the prediction of attitude change[J]. Psychological Review, 1955, 62: 42-55.
- [6] FESTINGER L. A theory of cognitive dissonance[M]. Stanford University Press, 1957: xi, 291.
- [7] ANDERSON N H. Foundations of information integration theory[M]. New York: Academic Press, 1981.
- [8] PETTY R E, CACIOPPO J T. Communication and Persuasion: Central and

- Peripheral Routes to Attitude Change[M]. New York: Springer, 1986.
- [9] WAGENKNECHT T, TEUBNER T, WEINHARDT C. A Janus-faced matter—The role of user anonymity for communication persuasiveness in online discussions[J]. *Information & Management*, 2018, 55(8): 1024-1037.
- [10] DROSSOS D, ZACHARIOUDAKIS M, DIONYSIOU G. Online Traffic Sources and Persuasion Techniques: How to Change Consumer Behavior[C/OL]//Proceedings of the 3rd International Conference on E-commerce, E-Business and E-Government. New York, NY, USA: Association for Computing Machinery, 2019: 80-84[2025-06-28]. <https://doi.org/10.1145/3340017.3342243>.
- [11] 王焯娣, 蒋玉石, 苗苗, 等. 隐私显著性、平台可信度和调节聚焦对互联网定向广告说服效果的联合效应研究[J]. *管理评论*, 2022, 34(8): 144-156.
- [12] TAL A, WANSINK B. Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy[J]. *Public Understanding of Science*, 2016, 25(1): 117-125.
- [13] DRAGICEVIC P, JANSEN Y. Blinded with Science or Informed by Charts? A Replication Study[J]. *IEEE transactions on visualization and computer graphics*, 2018, 24(1): 781-790.
- [14] 刘德文, 高维和, 闵凉宇. 声音特征和文本策略的说服效应研究[J]. *管理学报*, 2022, 19(9): 1373-1381+1408.
- [15] SUNG K H, LEE M J. Do Online Comments Influence the Public's Attitudes Toward an Organization? Effects of Online Comments Based on Individuals' Prior Attitudes[J]. *The Journal of Psychology*, 2015, 149(4): 325-338.
- [16] NERA K, PANTAZI M, KLEIN O. “These Are Just Stories, Mulder”: Exposure to Conspiracist Fiction Does Not Produce Narrative Persuasion[J/OL]. *Frontiers in Psychology*, 2018, 9[2023-02-15]. <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00684>.
- [17] SCANNELL D, DESENS L, GUADAGNO M, 等. COVID-19 Vaccine Discourse on Twitter: A Content Analysis of Persuasion Techniques, Sentiment and Mis/Disinformation[J]. *Journal of Health Communication*, 2021, 26(7): 443-

- [18] 刘楠, 安薪如, 李爱梅, 等. 现在避害,未来趋利: 目标框架和时间距离交互影响疫苗说服有效性[J]. 心理学报, 2022, 54(12): 1532-1547.
- [19] AIELLO G, DONVITO R, ACUTI D, 等. Customers' Willingness to Disclose Personal Information throughout the Customer Purchase Journey in Retailing: The Role of Perceived Warmth[J]. Journal of Retailing, 2020, 96(4): 490-506.
- [20] CIALDINI R B. Influence: the psychology of persuasion[M/OL]. Rev. ed. New York: Morrow, 1993[2022-11-13]. <http://www.gbv.de/dms/bowker/toc/9780688128166.pdf>.
- [21] SHU W, CHENG C Y. How to improve consumer attitudes toward using credit cards online: An experimental study[J]. Electronic Commerce Research and Applications, 2012, 11(4): 335-345.
- [22] LI S, VAN HALEN C, VAN BAAREN R B, 等. Self-Persuasion Increases Healthy Eating Intention Depending on Cultural Background[J]. International Journal of Environmental Research and Public Health, 2020, 17(10): 3405.
- [23] 杨莉明, 徐智. 垂直集体主义价值观在新冠防疫中的说服作用与行动影响——基于中老年群体的研究[J]. 新闻记者, 2021(9): 36-47.
- [24] 曾丽红, 黄蝶. 是什么在主导网络直播购物意愿——说服理论视域下对直播购物受众购买意愿的影响因素研究[J]. 新闻与写作, 2021(7): 50-57.
- [25] 周金连, 吴晔, 韩仪, 等. 社会化媒体信息接触对个体健康行为的说服效果研究——以 HPV 疫苗采纳为例[J]. 新闻大学, 2022(2): 1-16+117.
- [26] SHENG X, ZENG Z, ZHANG W, 等. Vlogger's persuasive strategy and consumers' purchase intention: The dual mediating role of para-social interactions and perceived value[J/OL]. Frontiers in Psychology, 2022, 13[2023-02-15]. <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.1080507>.
- [27] PENG L, JIANG H, GUO Y, 等. Effect of Information Framing on Wearing Masks During the COVID-19 Pandemic: Interaction With Social Norms and Information Credibility[J/OL]. Frontiers in Public Health, 2022, 10[2023-02-15]. <https://www.frontiersin.org/articles/10.3389/fpubh.2022.811792>.
- [28] PENG L, GUO Y, HU D. Information Framing Effect on Public's Intention to

- Receive the COVID-19 Vaccination in China[J]. *Vaccines*, 2021, 9(9): 995.
- [29] BERGIER J Y, FAUCHER C. Persuasive Communication from a Military Force to Local Civilians: A Cognitive Treatment of PsyOps Messages Based on the Elaboration Likelihood Model[J]. *Int. J. Softw. Sci. Comput. Intell.*, 2016, 8(2): 21-34.
- [30] HALL C C, OPPENHEIMER D M. Error Parsing: An alternative method of implementing social judgment theory[J]. *Judgment and Decision Making*, 2015, 10(5): 469-478.
- [31] MOURA A F, ASCHEMANN-WITZEL J. Perspectives on sugar consumption expressed on social media by French-speaking and Danish-speaking parents[J]. *Social Science & Medicine*, 2021, 270: 113636.
- [32] EPSTEIN J M, AXTELL R. *Growing Artificial Societies: Social Science from the Bottom Up*[M]. Brookings Institution Press, 1996.
- [33] HELBING D, MOLNÁR P. Social force model for pedestrian dynamics[J]. *Physical Review E*, 1995, 51(5): 4282-4286.
- [34] ZHENG S, TROTT A, SRINIVASA S, 等. The AI Economist: Taxation policy design via two-level deep multiagent reinforcement learning[J]. *Science Advances*, 2022, 8(18): eabk2607.
- [35] PIAO J, YAN Y, ZHANG J, 等. AgentSociety: Large-Scale Simulation of LLM-Driven Generative Agents Advances Understanding of Human Behaviors and Society[EB/OL]. arXiv, 2025[2025-06-25]. <http://arxiv.org/abs/2502.08691>.
- [36] WANG L, MA C, FENG X, 等. A survey on large language model based autonomous agents[J]. *Frontiers of Computer Science*, 2024, 18(6): 186345.
- [37] LIN, GAO C, LI M, 等. EconAgent: Large Language Model-Empowered Agents for Simulating Macroeconomic Activities[EB/OL]. arXiv, 2024[2025-06-25]. <http://arxiv.org/abs/2310.10436>.
- [38] GAO C, LAN X, LU Z, 等. S³: Social-network Simulation System with Large Language Model-Empowered Agents[EB/OL]. arXiv, 2025[2025-06-25]. <http://arxiv.org/abs/2307.14984>.

- [39] PARK J S, O'BRIEN J, CAI C J, 等. Generative Agents: Interactive Simulacra of Human Behavior[C/OL]//Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. San Francisco CA USA: ACM, 2023: 1-22[2025-06-25]. <https://dl.acm.org/doi/10.1145/3586183.3606763>.
- [40] ZHA Q, KOU G, ZHANG H, 等. Opinion dynamics in finance and business: a literature review and research opportunities[J]. Financial Innovation, 2021, 6(1): 44.
- [41] 邱嘉平. 因果推断实用计量方法[M/OL]. 上海财经大学出版社, 2020[2025-06-28]. <https://book.douban.com/subject/35177003/>.
- [42] 苏宏元, 王月琳. 智能传播时代网络舆论生态变化及其治理[J]. 中国编辑, 2022(12): 21-25+31.
- [43] 肖潇. 基于社会判断理论的说服传播策略研究 ——以 Twitter 平台“COVID-19 Vaccine”议题为例[D/OL]. 华中科技大学, 2025[2025-06-30]. <https://doi.org/10.27157/d.cnki.ghzku.2023.003837>.
- [44] BRECKLER S J. Empirical validation of affect, behavior, and cognition as distinct components of attitude[J]. Journal of Personality and Social Psychology, 1984, 47: 1191-1205.
- [45] GAO C, LAN X, LI N, 等. Large language models empowered agent-based modeling and simulation: a survey and perspectives[J]. Humanities and Social Sciences Communications, 2024, 11(1): 1259.
- [46] CHUANG Y S, GOYAL A, HARLALKA N, 等. Simulating Opinion Dynamics with Networks of LLM-based Agents[EB/OL]. arXiv, 2024[2025-06-28]. <http://arxiv.org/abs/2311.09618>.
- [47] TAUBENFELD A, DOVER Y, REICHART R, 等. Systematic Biases in LLM Simulations of Debates[C/OL]//Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. 2024: 251-267[2025-06-28]. <http://arxiv.org/abs/2402.04049>.