网络群体认知计算建模研究综述

尹公主，张宏莉，田逸艺，田泽庶，孟辰，何嘉豪，黄圣鹏

（哈尔滨工业大学网络空间安全学院，黑龙江 哈尔滨 150001）

摘　要：现有网络群体认知计算建模研究散布于不同子领域，缺乏统一的建模框架，为此，基于计算机科学视角，使用统一符号体系回顾近期的研究成果，将相关研究归纳为3条技术主线：群体认知形成计算建模，聚焦个体认知如何涌现为群体层面认知；群体认知演化计算建模，刻画群体认知在时间维度上的演变模式；群体认知行为计算模拟，借助多智能体系统模拟并生成群体认知行为。在此基础上，系统梳理了相关技术的发展脉络，总结当前面临的关键挑战，并对未来研究方向进行展望。

关键词：群体认知建模；认知行为模拟；网络行为分析；群体决策；认知安全

中图分类号：TP391

文献标志码：A

DOI：10.11959/j.issn.1000−436x.2026034

Survey on computational modeling of network group cognition

Yin Gongzhu, Zhang Hongli, Tian Yiyi, Tian Zeshu, Meng Chen, He Jiahao, Huang Shengpeng

School of Cyberspace Science, Harbin Institute of Technology, Harbin 150001, China

Abstract: The current research on computational group cognition was scattered across several subfields and lacked a unified modeling framework. Therefore, based on the computer-science perspective, and used a unified notation system to review recent work. The literature was organized into three technical lines. First, group-cognition formation modeling examined how individual cognition gives rise to group-level cognitive states. Then, group-cognition evolution modeling described how group cognition changes over time and characterized typical patterns. Finally, group-cognition behavior simulation used multi-agent systems to reproduce and generate group cognitive behavior. On this basis, the development of key methods was traced, and identified current challenges and essential directions for future research.

Keywords: group cognition modeling, cognitive behavior simulation, online behavior analysis, group decision making, cognitive security

0　引言

网络的广泛普及使人类社会进入了一个以“数字互动”为核心的全新阶段。人们在网络中生成、传播并接收海量信息，点赞、评论、转发和关注构成了巨大的在线社会网络。网络不再只是信息交流的媒介，而是塑造公众认知的复杂生态系统。在这一系统中，多层次的网络群体不断涌现，包括明确成员边界的网络群组、高度互动的社区，以及以个体为中心的自我中心网络等[1]。在这些多层次群体交互中，个体的观点与情绪通过传播与共鸣不断积累和反馈，逐渐形成群体层面的认知，即由多主体互动涌现出的共享信念、观点态度和决策[2-3]。

当前，群体认知已成为影响公共舆论、社会治

收稿日期：2025−11−23；**修回日期**：2026−01−29

通信作者：张宏莉，zhanghongli@hit.edu.cn

基金项目：国家重点研发计划基金资助项目（No.2016QY03D0501）

Foundation Item：The National Key Research and Development Program of China (No.2016QY03D0501)

理和国家安全的重要因素，具有深远的应用背景与研究意义。例如，“剑桥分析”事件揭示了平台数据，结合心理定向投放可以大规模重塑选民认知；社交媒体驱动的集体情绪曾引发散户抱团等金融现象；虚假信息通过群体模仿与情感共振被放大，影响社会稳定[4]。与此同时，算法推荐通过内容排序与社交关系推荐等机制也在重塑网络社群结构与群体认知。此外，在认知战环境下，如何防御外部对网络群体认知的恶意入侵与诱导，已成为网络空间安全领域的关键议题[5]。

这些现实挑战驱动着学术界对网络群体认知形成与演化机制进行深入研究[6]。例如，社会心理学从经典理论（如从众效应、群体极化等）出发，揭示了群体观点趋同或分化的内在机制[7]；网络科学和统计物理则通过信息传播模型和相变分析刻画宏观认知演化模式[8-9]。然而，这些经典理论往往以定性解释或理想化假设为主，难以精准量化“在何种条件下出现何种效应”，且难以直接应对现实网络数据呈现的大规模、高维度、多源异质与快速演化等特征[10]。

近年来，以数据挖掘和人工智能为代表的计算科学方法为此提供了新的路径。尽管相关工作未必直接以“群体认知”命名，但其研究对象与核心目标在机理上一致：以个体为基本单元、以交互网络为载体，通过建模“个体–个体及个体–群体”的信息交换与相互影响过程，分析并预测群体认知层面的可观测现象。换言之，这些方法把传统研究中难以直接度量的“群体认知形成与演化过程”转化为数据驱动的可学习、可计算和可检验的模型组件，从而为群体认知的量化研究提供支撑。

然而，现有研究往往分散在社交网络分析、复杂网络、多智能体等不同子领域。虽然这些子领域均能为群体认知建模提供关键的技术组件，但目前尚缺乏一个统一的计算框架对其进行逻辑整合与系统梳理。例如，群体推荐研究[11]侧重于静态的“群体偏好聚合”，观点动力学研究侧重于观点在群体中的“时序演化”[12]，而多智能体系统[13]则侧重于微观交互机制驱动下的生成式“群体行为模拟”。

基于上述背景，本文从计算机科学视角出发，围绕网络中的群体认知问题，采用下列方法对文献进行检索。

1) 检索关键词。group recommendation、group/collective consensus、group decision、group polarization、group opinion dynamics 和 multi-agent social simulation。

2) 检索范围。近年（重点为2020—2025年，也包括更早的高引用文章）来自计算机领域顶级期刊（TKDE、TOIS、TPAMI、Knowledge-Based Systems、ACM Computing Surveys 等）、会议（KDD、WWW、NeurIPS、ICML、ICLR、ACL、EMNLP、NAACL、SIGIR、CHI、CIKM、AAAI、IJCAI、ICDE、WSDM、ICWSM、CSCW等），及交叉学科期刊（Nature Human Behaviour、PNAS、Nature Communications等）中与群体认知、集体行为紧密相关的代表性研究成果。

基于上述检索逻辑，本文从中筛选出与群体认知相关的研究，将其整理为3条相互关联、递进而又各具侧重的技术主线。1) 群体认知形成计算建模：在给定时间戳上，个体认知状态如何通过多主体交互与聚合机制涌现为群体层面的认知表征并引导群体行为，典型任务包括群体推荐、群体决策建模等。2) 群体认知演化计算建模：关注群体认知在时间维度上的动力学演化过程，利用基于方程的观点动力学模型和数据驱动的时序模型刻画共识形成、极化、碎片化等动态群体认知现象。3) 群体认知行为计算模拟：通过构建基于多智能体的模拟环境，模拟生成群体认知行为，为算法设计与政策评估提供“计算实验室”。图1展示了近5年群体认知计算建模相关文献的文章数量变化。

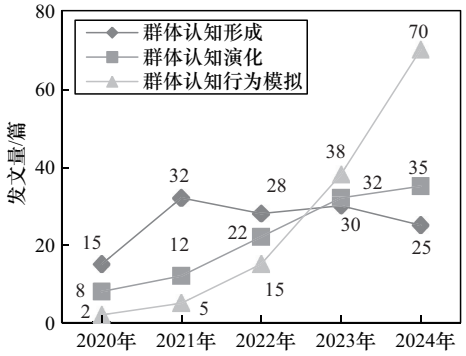


图1　近5年群体认知计算建模相关文献的文章数量变化

结合近5年群体认知计算建模相关文献的文章数量变化，本文观察到了以下趋势：群体认知形成发文量较为稳定，显示出其作为领域基石的重要性；群体认知演化发文量呈现稳步上升趋势，表明

学术界对群体认知动态过程的关注度持续增加；群体认知行为模拟近两年发文量呈现爆发式增长，这一跃升主要得益于大语言模型（large language model, LLM）等生成式人工智能技术的突破，使高保真和强拟人的大规模群体行为模拟成为可能，并逐渐占据研究前沿。

为了直观呈现本文的逻辑脉络与领域全貌，图2构建了一体化的网络群体认知计算建模研究体系框架。基于此框架，本文将归纳现有的计算建模方法，深入剖析当前面临的核心挑战。希望通过本文的系统梳理，能厘清从“静态聚合”到“动态演化”再到“生成式模拟”的技术演进逻辑，为后续研究提供有价值的参考与指引。

1　群体认知关键概念及定义

为构建严谨的网络群体认知分析框架，本文从计算机科学的视角出发，在统一符号体系下构建了一个层次化的概念体系，使用的主要符号及其含义如表1所示。

1.1　网络群体

在社会科学中，“群体”涵盖从家庭、社区到政党、民族等各种规模的社会集合，通常强调成员互动、共同目标与心理认同等要素[3]。为了在大规模网络数据上开展定量分析，需要将这一概念转化为在图结构中可操作的对象。

表1　主要符号及其含义

符号	含义
$G=(V,\mathcal{E})$	网络图
i	网络个体
$g=(V_g,\mathcal{E}_g)$	网络群体
$V_g \subseteq V$	群体g的个体节点集合
$\mathcal{E}_g \subseteq \mathcal{E}$	群体g的内部交互结构
$C_i(t),C_g(t)$	个体i/群体g在t的认知系统
$E_i(t),E_g(t)$	作用于个体i/群体g的外部信息流与交互
$O_i(t)/O_g(t)$	个体i/群体g的对外可观测行为（如发言、决策、转发等）
$H_g(t)$	群体g的认知宏观状态
π_i,π_g	认知系统的对外输出映射函数
$\mathcal{F}_{\text{form}}$	群体认知形成函数
$\mathcal{F}_{\text{evo}}$	群体认知演化函数

定义1　网络图。网络可抽象为一个图$G=(V,\mathcal{E})$，其中节点集合$V=\{v_1,v_2,\cdots,v_N\}$表示用户账号等行动者，边集合$\mathcal{E}\subseteq V\times V$表示行动者之间的特定关系（如关注、好友）或交互行为（如转发、评论）。根据应用场景，G可以是有向/无向图、带权/无权图。

定义2　网络群体。给定网络图$G=(V,\mathcal{E})$，一个网络群体定义为其任意非空子图，记为$g=$

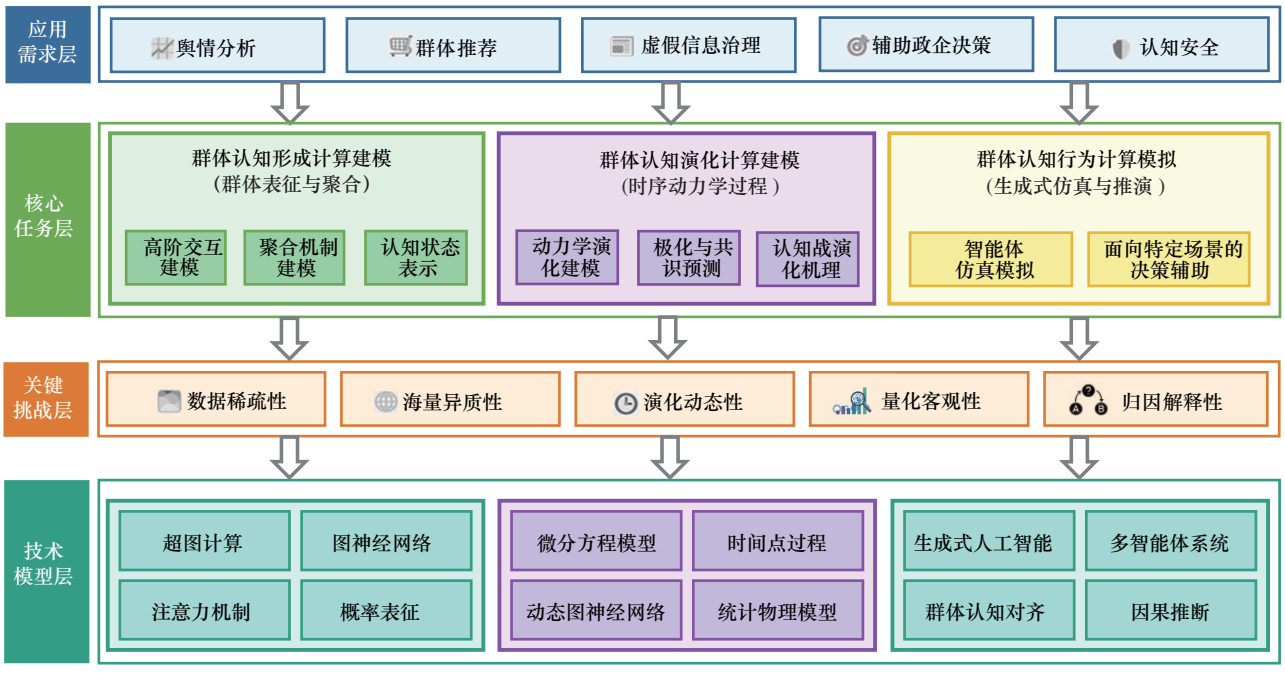


图2　网络群体认知计算建模研究体系框架

$(V_g,\mathcal{E}_g)$，其中 $V_g \subseteq V$，$\mathcal{E}_g \subseteq \mathcal{E}$。

根据其形成方式和识别方法的不同，网络群体在计算视角下主要分为以下 3 种类型。

1) 显式群体。由用户基于共同兴趣、身份或目标主动创建或加入的客观存在的群组，具有明确的成员边界和身份标识。典型的例子包括微信群、Facebook 小组、豆瓣小组或 Reddit 的子版块。由于其成员边界清晰、主题明确，显式群体是研究特定话题下群体话语形成、情感演化和认知同步的理想实验对象。

2) 隐式群体。并非由用户直接声明，而是通过计算方法从网络中挖掘得到（如社区发现将网络划分为“内连密集、外连稀疏”的节点子集），或依据研究目标按共同社会属性/心理认同构建的集合。隐式群体揭示了信息可能高效传播的内生结构，是刻画观点极化、舆论阵营形成等宏观现象的关键载体。

3) 自我中心网络。以某一节点为中心，由其及直接邻居构成的局部子网（如个人“朋友圈”）。该视角关注个体如何嵌入其近邻环境及其对行为与认知的影响。

综上所述，“网络群体”在计算视角下并不是一个单一的概念，其具体定义和识别方法取决于研究目标[14]。“在线群体”和“社群”皆属于其研究子集。研究过程中可以根据任务需要选取显式群体、隐式群体或自我中心网络作为分析单位，并在统一的图模型框架下进行量化研究和对比。

在此基础上，本文对群体认知计算框架中的各关键概念进行定义，并在图 3 中直观展现了群体认知计算框架中各关键概念之间的联系。

1.2　计算视角下的个体认知

群体认知的涌现以个体为基础[15]。传统计算认知科学通常将个体视为一个信息处理系统[16]，其观测行为由内部状态和外部刺激共同决定。内部状态在环境输入下持续更新，并通过决策机制外显为外部行为与表达[17]。例如，在典型群体认知计算框架贝叶斯认知模型中[18]，人类行为被表示为一个条件概率分布 $\boldsymbol{P}$（输出|信息流刺激，认知系统），其中认知系统的状态共同决定行为输出。传统的计算认知科学长期致力于模拟个体层面的简单生物性认知活动[19]，如感知、记忆、语言理解和决策行为。在社交网络场景中，个体认知系统需进一步扩展以模拟更复杂的社会认知活动[20-21]。该系统包含一系列内部状态变量，即人格、信念、价值观等稳定认知状态，以及情绪、注意等非稳定认知状态[22]。当系统接收外部刺激（如社交网络中的信息流、邻居的互动行为等）时，个体将在当前情境下产生相应的认知输出，如点赞、评论、转发等可观测行为输出[23]。

定义 3　个体认知系统。对于任意一个体 i，其

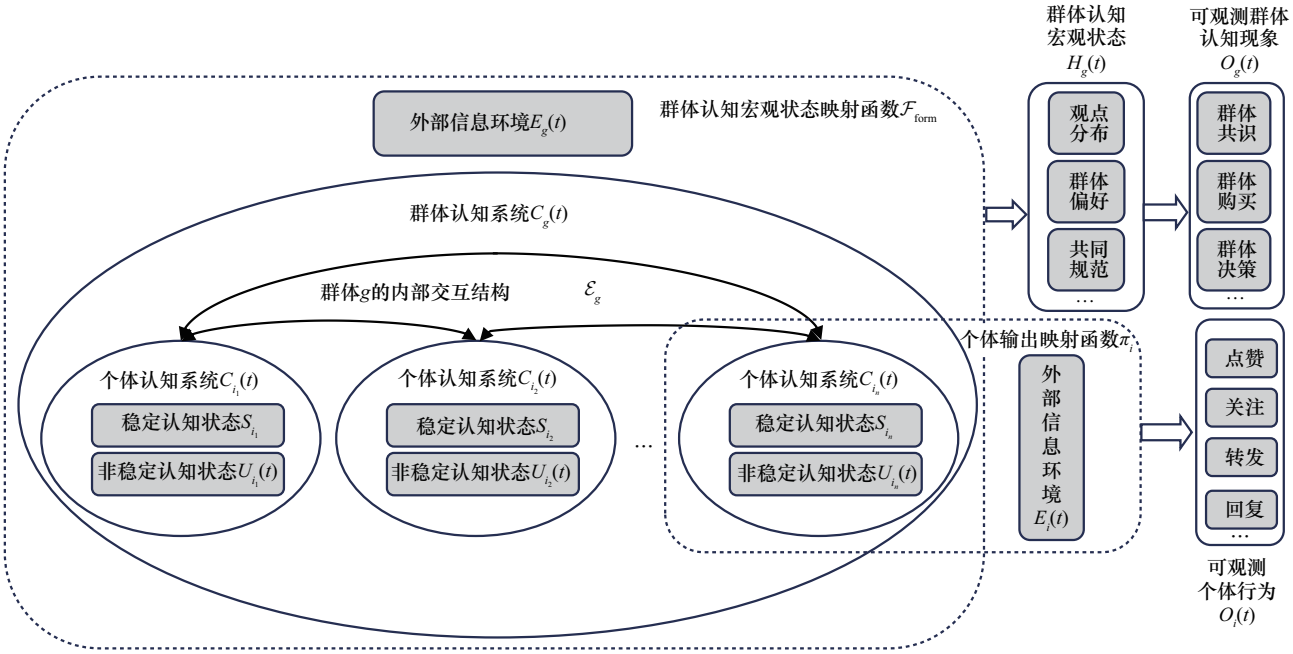


图 3　群体认知计算框架中各关键概念之间的联系

在时刻t的认知系统可表示为

$$C_i(t)=(S_i,U_i(t)) \tag{1}$$

其中，S_i为稳定认知状态，其刻画在较长时间尺度上基本不变或缓慢变化的特质（如人格、长期偏好、价值取向、信念、观点等），时间尺度长、更新稀疏；$U_i(t)$为非稳定认知状态，其刻画在短时间尺度上易波动的认知状态，如当前情绪、瞬时注意焦点、工作负荷、短时记忆痕迹等。

定义 4　外部信息环境。在时刻t，作用于个体i的外部信息环境表示为$E_i(t)$，包含时间线上呈现的信息内容、邻居节点的可见行为轨迹（如点赞、转发和评论）、平台排序与推荐信号、线下环境线索等。

在实际情况中，认知系统不可被直接观测，只能通过接收外部刺激后状态映射为可观测输出的方式被动观测[24]。

定义 5　个体认知输出与行为生成。在计算认知模型中，个体的可观测行为被视为认知系统对外部环境的输出。形式上，引入输出映射函数为

$$O_i(t)=\pi_i(C_i(t),E_i(t)) \tag{2}$$

其中，$O_i(t)$表示个体在时刻t的可观测输出，如转发某条内容、发表某言论等。

1.3　计算视角下的网络群体认知

在定义了网络结构中的分析单元（群体）和构成单元（个体认知系统）之后，本节首先给出“网络群体认知”的社会学理论基础与形式化计算框架。从社会心理学与复杂系统理论出发，群体认知具有以下 3 个关键理论。

1) 分布式认知。该理论由 Hutchins[25]提出，认为认知单元可以超越个体大脑，扩展至由个体、工具和环境构成的系统。社交网络中的帖子、评论结构和推荐算法共同参与了信息加工，因而群体认知应被视为“个体–网络环境”耦合系统的产物[26]。

2) 集体智能。在满足多样性、独立性、去中心化与有效聚合机制等条件时，群体在问题求解和决策方面可能优于任何单个成员。在线众包平台、维基协作、开源创新社区等都是集体智能的体现[27]。

3) 涌现。复杂系统理论强调，微观个体遵循简单规则相互作用，能够在宏观产生全新和无法由个体属性直接预测的模式[28]。观点极化、舆论共识等群体认知现象，正是个体间简单的社会影响规则在网络中迭代作用后涌现出的宏观格局。

这些理论共同指向一个结论：对群体认知的计算建模，必须超越个体认知的简单聚合，转向对“个体–交互–结构”三位一体的系统建模，并显式关注宏观集群现象的涌现和演化。

在网络中，群体认知系统并非个体认知的简单线性加总，而是个体通过信息交流、情绪共鸣与社会互动而涌现出的宏观行为背后的分布式、动态演化的复杂驱动系统[2,6]。

定义 6　群体认知系统。群体g在t时刻的认知系统定义为$C_g(t)=(\{C_i(t)\}_{i\in V_g},\mathcal{E}_g)$，表示由其内部个体和交互结构构成。

在此基础上，参照群体认知相关理论[2,14]，本文进一步将“群体认知宏观状态”界定为：在社交网络场景中，由多个个体通过持续交互、观点传播与相互影响所涌现出的一种共享认知结构，可形式化为群体在给定问题或信息刺激下形成的集体反应函数。此体系为后续群体认知的形成机理建模、演化过程分析及群体认知引导的行为建模提供了基础。

定义 7　群体认知宏观状态与形成函数。群体g在t时刻的认知宏观状态记为$H_g(t)$，它压缩了群体在特定议题或任务维度上的整体认知特征（如观点分布、情绪分布、群体偏好、共同规范等），令认知宏观状态为

$$H_g(t)=\mathcal{F}_{\text{form}}\left(C_g(t),E_g(t)\right) \tag{3}$$

在具体模型中，$\mathcal{F}_{\text{form}}$可以是显式的规则型聚合函数（如加权平均、投票）、可学习的神经网络（如图神经网络（graph neural network, GNN）、超图网络）或概率图模型等。

定义 8　群体认知的可观测输出。群体g在t时刻的可观测输出由群体输出映射函数表示为$O_g(t)=\pi_g(H_g(t))$，其中$O_g(t)$表示群体推荐系统中群体购买、观点演化中群体共识、极化现象等。

定义 9　群体认知演化函数。群体认知在时间上的演化可以用群体认知演化函数描述为$H_g(t+1)=\mathcal{F}_{\text{evo}}\left(H_g(t)\right)$，在不同建模范式下，$\mathcal{F}_{\text{evo}}$可以具体化为经典观点动力学中的差分/微分方程、神经常微分方程（ordinary differential equation, ODE）/时序 GNN 等数据驱动模型、基于智能体模拟中的显式迭代规则等。

2　群体认知形成计算建模

根据群体认知的形式化计算框架，群体认知形成可函数化表示为 $H_g(t)=\mathcal{F}_{\text{form}}\left(C_g(t),E_g\right)$，由群体宏观认知状态生成可观测输出的过程可表示为 $O_g(t)=\pi_g(H_g(t))$。在群体决策、群体推荐等群体认知现象的建模过程中，其在计算上有3个需要解决的核心问题：1) 如何建模多元主体间通过高阶交互形成群体认知系统 $C_g(t)=(\{C_i(t)\}_{i\in V_g},\mathcal{E}_g)$ 的过程；2) 如何设计“聚合算法”$\mathcal{F}_{\text{form}}$ 来生成群体宏观认知状态；3) 形成的群体认知状态 $H_g(t)$ 本身如何表征。本节围绕这3个问题展开，对其所涉及的基础模型与前沿方法进行分析、对比与归纳。

2.1　群体间高阶交互建模

传统的社会网络分析主要基于二元图，然而，群体认知（如共识、极化）的涌现本质上是“多主体”同时交互的结果[28]。例如，一次群聊、一场集体讨论或一次共同创作，这些都不是由简单的二元交互累加而成的。

二元图范式在试图表征这些群体高阶交互时存在严重的信息损失[29]。例如，它将一个 N 人的群体交互“降维”为 $\frac{N(N-1)}{2}$ 条成对的边（即一个“团”），或者引入一个虚拟的“群体节点”。前者无法区分“N 个独立的双边对话”和“1个 N 人会议”；后者则割裂了群体的“内生性”，使群体成为一个外在的、与个体相分离的实体。因此，为精准捕获群体认知形成的结构基础，计算建模转向了能够显式建模“多对多”交互的高阶结构。本节对各群体间高阶交互建模方法总结如图4所示。

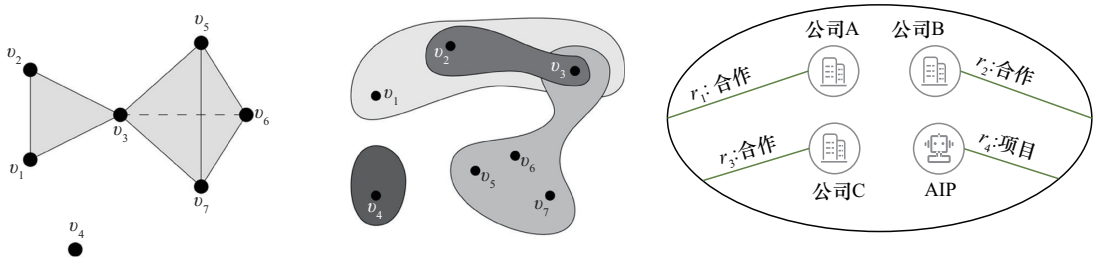
1) 基于超图的高阶交互建模方法

超图理论是目前解决高阶交互建模最主流且最具扩展性的数学工具。一个超图可表示为 $H=(V,E)$，其中 V 是节点集（代表个体用户），E 是超边集。与二元图不同，一条超边 $e\in E$ 可以包含任意数量的节点（$|e|\geqslant 2$）。这种“集合”的特性使其天然地对应现实世界中的“社交群组”“共同关注话题的集合”或“协同互动的团队”。

在群体认知形成的计算建模中，超图神经网络（hypergraph neural network, HyperGNN）[30]的应用标志着从结构特征工程向端到端表示学习的跨越。早期的超图学习主要侧重于谱分析[31]，利用拉普拉斯矩阵的特征分解来挖掘群体结构，但这种方法难以处理大规模动态网络。较新的HyperGNN范式则引入了非线性的消息传递机制[32]，通过“节点→超边→节点”的两阶段聚合，实现了组内信息的全域交互。

在具体的应用场景中，ConsRec模型深入探索了超图在群体推荐中的共识建模能力[33]。该研究指出，群体决策不仅仅是成员偏好的加权，还受群体作为独立实体“固有偏好”的影响。ConsRec模型利用超图卷积层显式建模了群体成员间的交互，在聚合成员偏好的同时，通过超边嵌入捕捉群体的整体特征，从而在数学上将“个体意愿”与“群体规范”进行了统一建模。类似地，Xu等[34]进一步引入了对比学习框架，在超图结构上执行成员偏好与群体共识的对齐操作。该研究针对群体认知中常见的“长尾沉默”问题（即少数人的强需求被多数人的弱需求淹没），提出了一种通过超边内的对齐损失函数来平衡多数人与少数人利益的计算机制，从而更真实地反映了群体认知的协商过程。

此外，超图的动态性建模也是当前的研究热点。在群体活动识别任务中，动态注意力超图卷积网络（dynamic attention hypergraph convolutional network, DAHGCN）[35]挑战了静态超图的假设。在真实的社交互动中，群体的构成与交互强度是随时间流动的。DAHGCN引入了动态注意力机制，



(a) 单纯复形　　(b) 超图　　(c) n元关系知识图谱

图4　群体间高阶交互建模方法

根据节点在当前时间的交互特征调整超边权重与结构。

2) 其他高阶交互建模方法

① 单纯复形。相比超图，单纯复形[36]是一个更具拓扑结构的高阶模型。它要求高阶交互（如3人组，即一个2-simplex）必须包含其所有的子结构（即3个2人组，即1-simplex）。这种“闭包特性”使其特别适用于建模由内而外的层次结构。例如，Chen等[37]提出的单纯复形传染模型指出，简单的二元接触（1-simplex）产生的是线性的感染概率，而高阶的群体接触（2-simplex及以上）能产生非线性的“协同增强”效应。在社会学中，一个4人组的集体认同，往往是建立在组内所有2人和3人子群体认同基础上的。在共识形成或传染病动力学中，群体压力和高阶影响机制可以通过单纯复形的拓扑结构[38]得到更精确的刻画。

② n元关系知识图谱（knowledge graph, KG）。在更细粒度的认知建模中，群体内部往往存在明确的“角色分工”和“权力结构”。例如，在科研合作网络中，一篇论文的作者构成一个群体，但“第一作者”“通信作者”与“其他作者”对群体认知产出的贡献权重和责任是截然不同的。超图和单纯复形通常将节点视为同质的，难以处理这种语义异质性。n元关系知识图谱[39]和超关系KG[40]提供了另一种高阶建模思路。与超图的“无序集合”不同，n元关系KG事实包含一个关系r和多个实体，并且可以为实体分配在交互中的“角色”。例如，“群体”在超图中是无序超边，但n元关系KG中有（relation=coauthor; role1=first_author; role2=advisor; …）。n元关系KG已被前沿工作用于事件建模[41]和社交网络分析[42]。在群体认知建模中，这种富含语义的结构有望进一步捕捉群体内部的“角色”和“结构”（如意见领袖、附和者），这是超图和单纯复形无法显式表征的。

表2对群体间高阶交互建模常用方法进行了详细对比分析，对比了不同建模框架在“建模单元、典型应用、优点与不足”上的差异。从二元图到超图、单纯复形再到n元关系KG，表达能力逐步增强，能从“两两交互”扩展到“多主体交互”“层级拓扑结构”乃至“事件内部语义角色”。但与此同时，模型构建与推理实现也更复杂、计算成本更高，工程工具链的成熟度往往更弱，因此需要在表达精度与可落地性之间权衡。

2.2　群体认知聚合机制

群体认知聚合机制决定了个体认知如何在交互网络中融合并涌现为群体层输出。随着深度学习技术的发展，聚合机制已从早期基于社会选择理论的启发式规则演进为数据驱动的、具有高度自适应性的神经聚合模型。

1) 基于规则的聚合方法

早期的群体认知建模（尤其是在群体推荐中）依赖于固定的、启发式的聚合函数，这类方法可被称为“分数聚合”。这些聚合函数通常是研究者基于对社会动力学的先验假设而“手动设计”的。典型方法如平均策略、最不满意策略、最大影响力策略等[43]。这类方法的优点是简单和可解释性强，其核心缺陷是预设了一种群体动力学模式，忽略了群体的异质性和情境依赖性。例如，在一个政治讨论群中，聚合可能遵循“多数决”；而在一个技术互助群中，聚合可能遵循“权威决”。静态规则无法捕捉这种动态的交互规律。

2) 基于学习的聚合方法

为克服规则型聚合方法的局限性，研究转向使用“偏好聚合”的学习型方法。这类方法不再预设固定的聚合规则，而是从数据中学习聚合权重和函数。

注意力机制。注意力机制是实现差异化聚合的核心。它允许模型根据数据学习群体中每个成员对最终群体决策的不同权重（即影响力）。例如，代表性工作MoSAN[43]使用注意力机制来获取用户相

表2　群体间高阶交互建模常用方法对比

方法类别	建模单元	典型应用	优点	不足
二元图	边	传统共识模型	构建简单、算法成熟	仅刻画两两交互
超图	超边	群体共识；群体推荐	建模多主体交互、算法较成熟	无层级及语义关系
单纯复形	单形	群体意见拓扑分析	适用于建模层次结构	灵活性低、计算代价大
n元关系KG/超关系KG	n元关系	多元事件建模	可表达关系内部语义角色	构建和推理复杂、缺少成熟应用工具

对于所有其他成员的偏好嵌入，动态计算成员的相对影响权重。GroupIM[44]和SIGR[45]进一步引入了上下文变量。例如，在不同的议题下，意见领袖是切换的。通过引入“议题嵌入”作为查询向量，注意力机制实现了对群体认知重心的动态捕捉。

超图神经网络聚合。超图神经网络的消息传递范式本身就是一种复杂的、非线性的邻居聚合机制。在ConsRec模型[33]中，HyperGNN的卷积操作即是聚合算法的具体实现。它聚合了超边（群体）内的所有成员（节点）信息，以生成群体表征。超图卷积在数学上实现了“信息的一致性平滑”，但也带来了“过平滑”的风险，即随着层数加深，所有成员的表征趋于一致，丢失了个性化特征，这在模拟群体趋同效应时是合理的，但在需要保留不同意见时则成为弊端。

异质性聚合。为解决群体的异质性聚合问题，Duan等[46]提出了DGA-GNN“动态分组聚合”方法，并指出在聚合时应考虑邻居节点的“可区分性”。该方法并不直接聚合所有邻居的信息，而是设计了一个反馈驱动的动态分组模块，首先根据节点的特征距离和历史行为模式，将邻居节点划分为若干潜在的子群（如“支持派”“反对派”或“中立派”）。随后，模型在子群内部执行强聚合，在子群之间执行交互融合。在舆情极化分析中，这种方法能够避免正负样本直接平均导致的特征抵消，从而建模认知冲突。类似地，基于超图的ConfGR[47]模型将“从众心理”显式编码进自监督目标，以构建更符合社会规律的偏好聚合模型。

2.3 群体认知状态表示

经过高阶交互与聚合计算后，群体认知的宏观状态$H_g(t)$需要在向量空间中进行形式化表征。这一表征不仅是后续行为映射（如群体购买、集体决策）的输入，更是对群体认知本质的一种数学抽象。如表3所示，随着研究的深入[48-55]，早期的单一向量表征已难以满足对群体认知复杂性的描述需求，表征范式正经历从确定性点向量转向概率分布乃至量子态的变革。

1) 单一向量表征

单一向量表征是最传统和最简单的表征方式。通过2.2节的聚合算法，群体认知（如偏好、观点）被表征为嵌入空间中的一个单一向量。然而，单一向量一般是有限长度下的“有损压缩”。它能在一定程度上表征群体“共识”，但难以表征群体的“不确定性”“多样性”或“冲突”。

2) 多视角表征

为解决单一向量表征的局限性，研究者提出了多视角表征。Wu等[33]提出的ConsRec是多视角表征的典范，其明确指出群体共识是一个复杂的概念，不能仅靠聚合个体信息来捕获。ConsRec将群体偏好分为成员个体偏好聚合向量、项目特征聚合向量和群体层固有偏好向量3个部分，最终通过一个“自适应融合组件”将这3个不同来源的多视角表征向量融合，得到最终的群体认知表征。Yin等[48]在建模群体与群体交互时认知偏好动态性，也用多个视图向量来表征群体，并指出在不同交互中各视图下权重分布是不一样的，需要根据历史数据对典型的权重分布进行学习。

3) 概率分布表征

为表征群体认知的不确定性，一些方法主张在建模时，应从“单点平均值”转向使用“概率分布”来表示[49]。在群体认知建模中，这意味着$H_g(t)$不应是一个向量v，而应是一个概率分布$P(v)$，如一个高斯分布$\mathcal{N}(\mu,\Sigma)$。其中，均值向量可以表征群体的“平均认知”或“共识点”，协方差矩阵则捕获了群体的“认知多样性”或“不确定性”。这种概率方法已在用户建模中有所应用[50]，将其从个体层面扩展到群体层面是群体认知表征的重要发展方向。

表3 常见群体认知状态计算表示方法对比

表征范式	数学形式（举例）	核心思想	代表文献
单一向量表征	$v \in \mathbb{R}^d$	将群体视为个体特征的加权平均或聚合点	文献[43]
多视角表征	$\mathcal{X} \in \mathbb{R}^{I \times J \times K}$	基于多视图或多向量表征	文献[33,46]
概率分布表征	$P(x) \sim \mathcal{N}(\mu,\Sigma)$	不确定性建模：均值代表共识，方差代表多样性	文献[50]
模糊表征	$\mu_A(x) \in [0,1]$	用隶属度或语言变量描述“非黑即白”之外的中间态	文献[51]
量子认知表征	态矢量	利用概率幅的干涉效应解释非理性、非交换的决策逻辑	文献[55]

4) 模糊表征

为处理语义层面的模糊性，模糊表征方法被广泛采用。例如，Jia 等[51]在社交网络群体决策（social network group decision-making, SNGDM）框架中引入语言术语集与三支决策机制，将成员意见建模为模糊集或概率语言信息，从而显式量化群体决策过程中的犹豫度与歧义性。

5) 量子认知表征

社会心理学与认知科学的研究表明，人类群体的决策行为往往违背经典概率论（如全概率公式），表现出“顺序效应”“合取谬误”和“干扰效应”。例如，先讨论议题 A 再讨论议题 B，与反之顺序讨论，往往会产生截然不同的群体共识。经典的贝叶斯网络难以解释这种非交换性。为解决这个问题，近期研究已开始将“量子贝叶斯网络（quantum-like Bayesian network, QLBN）”[52]应用于 SNGDM[53]。量子认知模型不再试图表征一个静态的群体认知状态（如一个向量或一个分布），而是表征一个动态的、受上下文干扰的认知“过程”或“波函数”。在群体决策中，这意味着在“观测”（如投票、发言）之前，群体认知是不确定且易受干扰的。QLBN 中的干扰项显式地建模了测量这一行为（如民意调查、信息流刺激）如何导致“波函数坍缩”[54]，使群体认知从不确定的叠加态变为一个确定的决策结果。Li 等[55]进一步结合社会互动模式，利用量子算子模拟了外部信息流（作为“测量”操作）如何导致群体认知的“波函数坍缩”，从而形成确定的决策。这种模型在解释复杂的群体博弈、非理性舆论反转以及顺序敏感的群体决策时，展现出了超越经典概率模型的拟合能力。

3　群体认知演化计算建模

第 2 节探讨了群体认知“形成”的计算建模，其核心是聚合机制，即如何将离散的个体认知状态融合，在特定时刻“涌现”为群体的宏观认知表征。然而，群体认知是一个动态过程，它在持续的社会互动和外部刺激下不断演变。例如，一个群体可能从最初的观点分歧演化为高度共识，也可能在争议性议题的冲击下从温和一致走向极端极化。

群体认知演化计算建模的核心挑战在于构建一个函数 $\mathcal{F}_{\text{evo}}$ 以刻画群体宏观认知状态 $H_g(t)$ 的演化规律。该函数不仅需要描述个体层面的微观影响机制（如社会同质性、从众心理），还需要捕捉宏观层面的涌现现象（如共识达成、回声室形成、社会撕裂）。随着研究范式的变迁，现有计算建模方法经历了从“基于理论的方程建模”到“基于数据的深度学习”，再到“理论与数据融合的混合增强”的演进历程。本节将对基于方程的动力学模型、基于数据驱动的深度学习模型和融合社会学知识的混合模型这三类方法进行详尽梳理与分析。

3.1　基于方程的群体认知演化模型

基于方程的模型（equation-based model, EBM）是群体认知动力学研究的理论基石。在此范式中，研究者基于社会学、心理学或统计物理学的经典理论，手动设计个体间的微观交互规则（如观点平均、信任选择）或宏观系统变量间的关系。随后，通过数学推导（如均值场近似）或数值模拟来推导群体演化过程和终态。

1) 经典观点动力学模型

在 EBM 中，最成熟的分支是观点动力学模型，它将对认知状态中的“观点”进行分析（通常是一个连续或离散的标量），并专注于模拟其演化过程。

DeGroot 模型。这是最经典的连续观点模型。在 DeGroot 模型中[56]，假设存在 n 个个体，每个个体 i 在时刻 t 的观点用标量 $x_i(t)$ 表示。个体根据固定权重矩阵 $\boldsymbol{W}=(w_{ij})$ 对邻居（包括自身）的观点进行加权平均，得到下一时刻的观点。其更新式为

$$x_i(t+1)=\sum_{j=1}^{n} w_{ij}x_j(t),\quad i=1,2,\cdots,n \tag{4}$$

这是一个线性动力系统，其收敛性可以通过马尔可夫链和矩阵理论进行严格分析。当影响力网络 $\boldsymbol{W}$ 对应的图是强连通且非周期的，DeGroot 模型必然收敛到全局共识。DeGroot 模型的线性“平均”特性是其最大的优点，也是其最大的局限，它无法解释极化或持续的观点分歧。在实际社会中，个体通常具有自身偏好、认知局限、信任选择性或随时间变化的权重，这些都超出了该模型的表征能力。

Friedkin-Johnsen（FJ）模型。FJ 模型[57]是对 DeGroot 模型的关键扩展，它引入了心理学上更为现实的假设——个体的“固执”或“对初始信念的坚守”。在该模型中，个体更新观点时，不仅会受到邻居的影响，还会以一定的权重锚定在自己最初的观点上。其数学形式为

$$x_i(t+1)=\lambda_i\sum_{j=1}^{N}W_{ij}x_j(t)+(1-\lambda_i)x_i(0) \quad (5)$$

其中，λ_i是个体i的“易感性”。当$\lambda_i=1$时，模型退化为DeGroot；当$\lambda_i=0$时，个体是完全顽固的。它能有效解释为何在持续的社会互动之后，群体往往无法达成完全共识，而是形成一个稳定的、存在持续分歧的意见格局。缺点在于其仍是线性模型，忽略了互动中的非线性效果、随机扰动与时间变异性。此外，它仍然是一个“同化性”模型，无法解释观点相互“远离”的“极化”现象。

有界信任模型。有界信任模型引入“信任阈值”这一关键的非线性机制。最著名的代表是Hegselmann-Krause（HK）模型[58]。HK模型在DeGroot模型“平均”机制的基础上引入“信任范围”或“接受区间”这一约束。具体地，个体i在时刻t仅对那些与其观点差异不超过某个信任阈值ε的邻居（包括自身）加以信任与采纳，再对这些邻居的观点取平均值来更新自身观点。形式上定义信任邻域为

$$N_i(t)=\{\,j\,\big|\,|x_j(t)-x_i(t)|\leqslant\varepsilon\,\} \quad (6)$$

对应的更新规则为

$$x_i(t+1)=\frac{1}{|N_i(t)|}\sum_{j\in N_i(t)}x_j(t) \quad (7)$$

信任阈值ε的存在使模型变为高度非线性的，因为影响力权重（0或1）取决于观点状态本身。这导致模型难以解析求解，必须依赖大规模计算机模拟。在HK模型中，随着ε的不同设定，系统可能演化为完全共识，也可能分裂为多个意见簇。值得注意的是，HK模型仍然是一个“同化性”模型（即观点只会相互靠近或保持不变），它无法解释观点簇之间相互“推开”的双峰极化现象。文献[58]进一步将此模型扩展到了更复杂的场景，探讨了在多层或时变网络结构中，顽固主体的存在如何通过级联效应锁定局部群体的认知状态。该研究表明，少数高度顽固的关键节点在特定网络拓扑下，足以决定整个群体的认知走向，这为理解网络“沉默螺旋”现象提供了数学工具。

排斥性影响模型。这类模型在有界信任模型（吸引）的基础上，增加了“排斥”机制。例如，Wang等[59]提出，当观点差异过大时，个体不但不被吸引，反而会向相反方向移动，以增大彼此的差异。这种“吸引–排斥”的共同作用被认为是产生双峰观点分布（即“极化”）的必要计算机制。它解释了为什么社会在面对争议性议题时，会分裂成两个相互“推开”、不断远离的阵营[60-61]。

Nettasinghe等[62]进一步为计算社会科学领域提供了模拟“情感极化”的专用动力学模型。与模拟连续观点的模型不同，他们构建了一个离散选择模型（如是否戴口罩），其动力学显式地由“内群体之爱”和“外群体之恨”这两个核心参数驱动。该研究进一步推导了模型的均值场近似，建立了一套描述群体状态演化的ODE系统。通过调节参数内群体同化和外群体排斥的相对强度，该模型能够完整复现从群体共识到剧烈党派极化的演变过程，为理解情感极化提供了坚实的动力学理论基础。

2) 群体认知演化动力学模型

经典观点动力学模型的核心是建模群体内“微观”的个体认知以得到群体认知的总体演化情况。EBM的另一条重要技术路线是借鉴系统动力学思想，尤其在认知安全领域，直接对宏观的群体认知状态（如群体知识总量）进行建模。

基于Lotka-Volterra的认知入侵模型。刘佳豪[63]等借鉴生态学中的“捕食者–猎物”模型，将人工智能内容生成（artificial intelligence generated content, AIGC）环境下的认知博弈形式化为非线性动力系统。模型定义了两个核心对抗变量：“目标群体信息感知量”$x(t)$（猎物）和“认知入侵主体信息释放量”$y(t)$（捕食者），并定义了对应的动力学方程。该模型揭示了在AIGC的信息对抗下，群体认知演化呈现周期性特征，并量化了“吃一堑，长一智”的知识增长规律。

知识增长与对抗博弈模型。黄凤翔等[64]认为在认知战环境下，模型的核心变量不再是个体观点，而是宏观的“群体知识量”$x(t)$。研究基于“认知操纵”和“认知觉醒”两个核心对抗机制，构建了群体知识量演化的数学模型。该模型的动力学方程为

$$\frac{\mathrm{d}x}{\mathrm{d}t}=rx\left(1-\frac{x}{K}\right)-(r_1-r_0)cx \quad (8)$$

其中，$rx\left(1-\frac{x}{K}\right)$表示群体知识的Logistic增长项，代表在没有对抗的自然状态下，知识量的增长和饱和；$(r_1-r_0)cx$表示对抗博弈项，c是认知战信息量，r_1是认知操纵系数，r_0是认知觉醒系数。当认

知操纵大于认知觉醒时，群体知识量减少；反之，则群体可实现“从战争中凝聚智慧”，知识量增加。

3.2　基于数据驱动的群体认知演化模型

基于方程的群体认知演化模型的最大优点在于其可解释性和理论洞察力。然而，其核心弱点在于，无论是微观规则还是宏观方程，都是研究者手动预设的。这种“理论驱动”的范式在面对大规模、异质和复杂的真实网络数据时，显得过于僵化和简化，难以捕捉真实世界中复杂的非线性交互关系。近年来的前沿进展试图扭转这一范式，转向从大规模网络数据中利用机器学习（特别是深度学习），从大规模、高维度的时序数据中学习复杂的、非线性的和异构的演化函数。

1) 基于GNN的观点动力学

GNN的消息传递机制为学习观点动力学提供了天然的框架。GNN的核心操作“聚合邻居信息并更新中心节点表示”在数学上与DeGroot模型的“对邻居观点进行加权平均”是同构的。因此，GNN可以被视为一个极其强大的、非线性的和参数可学习的广义DeGroot模型。

然而，将标准GNN直接用于模拟长期观点演化面临一个核心挑战，即“过度平滑”。GNN通过不断聚合邻居信息，在层数加深（即时间步增加）后，所有节点的表示会不可避免地收敛到同一个值。这一特性对于模拟“共识”是有效的，但对于模拟“极化”或“稳定的分歧”却是灾难性的，因为它会抹除所有群体结构。

为解决过度平滑问题，Li等[65]提出了UniGO框架。UniGO的动机是平衡GNN对动力学平衡现象的学习与避免过度平滑之间的矛盾。其核心机制是引入“粗化–精化”架构：首先，在粗化阶段使用图池化方法将原始的大规模图压缩为一个更小规模的“骨架图”，原始节点被聚合成“超节点”；然后，GNN只在这个小规模的骨架图上运行，模拟超节点之间的动力学演化；最后，将超节点学习到的状态表示映射回原始的各个节点。通过这种多粒度建模，UniGO成功地在捕捉宏观演化趋势的同时，保留了微观层面的观点差异，能够同时拟合共识与分歧共存的复杂平衡态。

2) 基于神经ODE的连续时间动力学建模

GNN通常将演化建模为离散的时间步（$t \to t+1$）。然而，现实世界中的社会互动是异步和连续发生的。神经ODE[66]提供了一个更自然的框架，它不再学习离散的转移函数$X(t+1)=f(X(t))$，而是学习连续时间的导数$\frac{\mathrm{d}X}{\mathrm{d}t}=f(X,t)$，其中$f$是一个（图）神经网络。

Zang等[67]提出了NDCN（neural dynamics on complex network）模型，是GNN与ODE结合的开创性工作之一。具体而言，NDCN模型将GNN嵌入ODE的导数函数中，即用一个图神经网络$g_\theta(x(t),G)$来参数化系统在任意时刻t的瞬时变化率，即$\frac{\mathrm{d}x(t)}{\mathrm{d}t}=g_\theta(x(t),G)$。NDCN模型的前向传播通过ODE求解器（如Runge-Kutta或Dopri5）在连续时间轴上积分完成。这使NDCN模型能够自然地处理不规则时间间隔的观测数据，并能预测任意未来时刻的认知状态。实验表明，NDCN模型在预测网络热度扩散和谣言传播动力学方面显著优于离散的循环神经网络变体。

针对认知极化现象，Duan等[68]提出的Bi-Dynamic Graph ODE指出，仅用单一的潜在向量表示观点状态不足以刻画观点演化中复杂的矛盾心理。为模拟观点的吸引和排斥，该研究设计了一个“双重观点编码器”，将个体认知状态显式解耦为“正面情感空间”（x_{pos}）和“负面情感空间”（x_{neg}）两个独立的潜在状态。模型构建了两个相互耦合的神经ODE，分别控制正负情感的演化轨迹，并通过交叉注意力机制模拟两种情感之间的对抗与转化。这种解耦设计使模型能够精准复现“爱恨交织”的复杂心理变化以及由此引发的群体极化过程，是目前数据驱动极化建模的前沿工作之一。

3.3　融合社会学知识的混合模型

纯粹的数据驱动模型虽然强大，但往往面临“黑盒”不可解释、数据稀疏时过拟合等问题。而纯理论模型虽然逻辑自洽，却难以拟合真实数据。因此，融合理论与数据的混合模型成为前沿研究方向。

基于社会学原理的神经网络（sociologically-informed neural network, SINN）[69]是这一研究方向的代表性工作。SINN是一种混合建模范式，它的核心思想借鉴于“物理信息神经网络”（physics-informed neural network, PINN）[70]。PINN通过将物理定律作为偏微分方程（partial differential equation, PDE）残差项加入损失函数，来约束神经网络

的学习。SINN将这一思想移植到了社会科学，具体而言，SINN的损失函数由两部分构成，如式(9)所示。

$$L_{\text{total}} = L_{\text{data}} + \lambda L_{\text{theory}} \tag{9}$$

数据驱动损失L_{data}通过训练一个神经网络（如神经ODE）来拟合观测到的社交媒体观点数据；理论约束项L_{theory}衡量模型学习到的动力学轨迹与经典社会学模型（如FJ模型或舆论衰减定律）推导出的轨迹之间的偏差。SINN通过L_{theory}这一“社会学先验”，强制数据驱动的神经网络在拟合数据的同时，优先选择那些符合基本社会动力学法则的解。这不仅赋予了神经网络参数明确的社会学含义（如可以从学习到的权重中反推网络中的“顽固系数”），还极大地提高了模型在小样本场景下的泛化能力，即利用先验理论知识来填补数据的缺失。

表4从核心驱动力、可解释性、数据依赖度等角度对本节所介绍的群体认知演化计算建模不同类别方法进行了对比分析，并对常用群体认知演化计算模型评价指标进行了总结。

3.4 群体认知演化计算模型评价指标

群体认知演化计算模型的评估比静态预测更为复杂，它不仅要求模型对特定任务结果预测准确，还要求模型在宏观分布和演化过程上符合现实情况。基于现有文献，本文将评价指标体系分为3个维度：特定任务下预测结果准确性、终态宏观分布拟合度和动力学演化过程指标。

1) 特定任务下预测结果准确性

误差度量指标，如均方误差（mean square error, MSE）及平均绝对误差（mean absolute error, MAE），常用于观点值、情绪强度等连续变量的预测，能有效衡量预测轨迹与真实观测值的偏离程度，广泛应用于观点动力学拟合任务。

分类指标（Accuracy/F1分数）。适用于立场分类或符号预测（如人际关系的敌友演化）。在类别不平衡场景下（如极端少数派检测），F1分数能提供更客观的性能评价。

2) 终态宏观分布拟合度

分布距离（散度/Wasserstein距离）。用于度量预测与真实分布的差异。KL/JS散度衡量信息损失，越小表示模型越能准确捕捉真实分布特征。Wasserstein距离则能量化两个分布在数值空间上的“运输成本”，因其能有效捕捉意见均值漂移和分散度变化，在连续意见演化模型中被广泛采用[71]。

极化指数。用于评估群体分裂或极化现象。常用的极化指数包括基于方差的极化度量[72]，方差越大表示群体分歧越严重。此外，还有网络分歧指数（network divergence index, NDI）、全局分歧指数（global divergence index, GDI）、双峰系数（bi-modality coefficient, BC）、广义欧氏距离[73]等。

ER（Esteban-Ray）指数。源自经济学，被引入观点动力学以衡量“聚类下的对立”[74]。它同时考虑了群组内部的认同感和群组之间的疏离感。

结构相似度。针对共演化模型，需评估生成的网络拓扑或聚类结构与真实结构的相似性。常用指标包括Jaccard系数（衡量边重合率）以及归一化互信息（normalized mutual information, NMI）或兰德指数（rand index, RI）（衡量群体分群的一致性）。在复杂交互场景下，还可采用对话结构相似度来评估讨论动态的拟真性[75]。

3) 动力学演化过程指标

共识时间。指系统演化至稳态（如观点方差低于阈值）所需的时间步数[76]。在DeGroot等经典线性模型中，共识时间被证明与网络拉普拉斯矩阵的谱隙呈负相关，谱隙越小，群体“固执”度越高，达成共识越慢。

表4 群体认知演化计算建模不同类别方法对比分析

方法	核心驱动力	可解释性	数据依赖度	动态适应性	高阶关系处理能力	计算复杂度
基于方程的方法（EBM）	规则驱动：基于社会学或物理学假设手动设计微观交互规则	高：参数具有明确物理意义（如固执度、信任阈值）	低：不需要训练数据，依赖参数设定	弱：通常假设网络结构或规则是静态或简单的	弱：难以刻画3人以上的复杂群组交互	低：通常有解析解或数值模拟方法
基于数据驱动的方法（GNN/ODE）	数据驱动：通过神经网络从海量观测数据中拟合演化函数	低：通常为“黑盒”或半可解释（注意力/归因分析）	高：依赖大规模、高质量的现实时序数据	强：可处理时变网络及非线性动态数据	易：可通过高阶图及超图神经网络处理高阶交互	高：需要大量计算资源进行反向传播训练
融合社会学知识的混合模型	双轮驱动：数据拟合为主，理论约束为辅	中：通过设计理论正则化项，赋予部分参数物理意义	中：理论先验可弥补数据的稀疏性	强：兼顾复杂动态数据与先验约束拟合	易：可通过高阶图及超图神经网络处理高阶交互	高：引入物理约束项增加了优化难度

4　群体认知行为计算模拟

前文分别从群体认知形成的聚合涌现建模与演化的动力学建模进行了综述。这些“自顶向下”的描述性模型在拟合宏观数据方面效果显著，但受限于真实数据难以获取、难以在受控环境中进行反事实推演（如若平台引入某种新算法，群体认知将如何演变）等。本节聚焦于一种“自底向上”的“生成式”方法—基于智能体的建模与模拟（agent-based modeling, ABM）。ABM 通过构建一个“计算实验室”为遵循特定规则的微观智能体（Agent）赋予认知与行为模型，并模拟它们在社会网络中交互后，如何在宏观层面涌现出复杂的群体认知现象[77]。本节将遵循智能体认知内核复杂度不断提升的技术演进脉络，系统性地梳理与分析。

4.1　基于经典智能体的群体认知模拟

在经典 ABM 中，智能体是研究者用于在模拟社会环境中执行特定行为的计算实体。其核心在于“认知内核”的设计，这决定了智能体的行为逻辑与拟人化程度。表 5 对智能体认知内核设计方法进行了总结对比[78-94]。

1) 智能体认知内核设计

基于规则的认知内核设计。在早期大量的经典 ABM 中，智能体的行为规则是由研究者基于社会学或心理学理论“手动设计”的启发式规则。在此框架下，个体智能体具有内部状态（如情绪、信念、规范遵从程度等）和基于局部观测的决策规则[77]，这些规则大多基于社会学和认知心理学理论，如社会影响理论、从众理论、认知失调理论等[78]。这类模型的优势在于可解释性强和计算成本低，但其拟人性和适应性严重受限于预设规则的完备性。智能体只能执行研究者预先编码的动作，无法适应未曾预见的新情境。

基于数据驱动的认知内核设计。为提升智能体的拟人性和在复杂社会环境中的适应性，研究者转向使用机器学习构建智能体的“认知内核”。智能体不再是固定规则的执行者，而是能够在复杂社会环境中学习最优策略的学习者。这一转变体现了从“模拟已知理论”到“发现未知动机”的演进。典型的技术如下。① 多智能体强化学习（multi-agent reinforcement learning, MARL）。MARL 是构建适应性智能体的核心框架[79]。在群体认知模拟中[80]，MARL 可用于建模智能体如何在复杂的社会交互（如合作、竞争、协调）中，通过最大化自身或集体的奖励函数来学习更拟人化的行为策略（如选择何时发言、相信谁、传播什么信息）。② 逆强化学习（inverse reinforcement learning, IRL）。MARL 的一个核心挑战是奖励函数需要研究者手动设计[81]，从而重新引入了主观性。IRL 提供[82]了一种“数据驱动”的解决方案。IRL 的核心思想是：给定来自真实人类的大规模观测数据（如言论内容、信息传播路径、点赞转发行为），IRL 可以反向推断出最能解释这些行为的奖励函数[83]。这对于构建拟人化的智能体至关重要，因为它使智能体的动机本身是从数据中习得的，而不是研究者手动预设的。③ 参数与规则的数据驱动学习。除了学习行为策略，计算机领域的研究者还致力于利用数据学习经典模型中的微观参数，Min 等[84]提出的 MAS-FOD 框架是一个典型代表。在该框架中，个体的关键行为参数（如影响力权重、易感性阈值）不再由人工设定，而是基于真实社交网络数据进行训练和拟合。Vargas-Pérez 等[85]提出了一种基于 GNN 的“元建模”方法，通过训练一个 GNN 来学习从宏观网络结构与初始观点分布到最终观点格局（如共识或极化）之间的非线性映射。该方法实际上是用神经

表 5　　智能体认知内核设计方法总结对比

技术范式	核心驱动机制	认知特征	优势	局限
基于规则[78]	预设的社会心理学规则	非智能化：状态变量简单更新，无语义理解	计算高效，支持百万级节点，可解释性强	规则僵化，无法适应新情境，缺乏语义
基于强化学习[80]	奖励最大化、策略梯度	适应性强：通过试错学习最优交互策略	展现出复杂的适应性与涌现性	训练不稳定，奖励函数需人工设计
基于逆强化学习[83]	从观测数据反推奖励函数	拟人化：动机源于真实数据挖掘	消除人工假设偏差，动机更真实	计算复杂度高，对数据质量敏感
数据驱动的参数学习[84-85]	神经网络拟合、元建模	参数化：微观参数由宏观数据反推	连接了理论模型与真实数据	模型可能是黑盒，缺乏过程解释性

网络替代了传统ABM烦琐的蒙特卡罗模拟过程，不仅能够利用宏观数据反推微观参数，还有助于反事实推演。

2) 典型应用与关键机制建模

经典ABM被广泛应用于模拟多种群体认知现象。

情绪传染。情绪作为一种关键的非稳定认知状态，其在群体中的传播是群体认知形成的重要组成部分。Haeringen等[86]系统地回顾了如何使用ABM来模拟人群中的情绪传染。这类模型通常将情绪传播视为一种“流行病学”过程。例如，Bosse等[87]提出的计算模型，明确建模了情绪吸收与放大，并引入人际关系强度（如信任度、社交关系）作为调节因子，展示了宏观集体情绪（如惊恐、愤怒）如何涌现。Fan等[88]的ABM研究发现，不同的情绪具有不同的网络传播动力学：“愤怒”情绪倾向于通过“弱连接”进行传播，这使其能够跨越紧密的社群边界，在全网范围引发大规模的集体愤怒或极化；而“喜悦”等积极情绪则更倾向于在“强连接”的内部社群中传播。

社会规范与语言涌现。ABM也被广泛用于模拟个体对规范的学习和演化过程。例如，Ajmeri等[89]模拟了智能体通过对违规行为的合理解释来促进社会规范的稳健涌现。Agrawal等[90]进一步将遗传算法与强化学习结合，使智能体使用遗传算法来演化和产生新的显式规范，并通过强化学习来学习和评估这些社会规范的价值，实现了规范体系的自主优化。此外，研究发现更底层的认知功能也可以通过ABM涌现。Mordatch等[91]研究展示，在没有预设语言的情况下，基于MARL的多个智能体为了完成协作任务（如导航到指定地点），自主演化出了一套具有词汇和句法结构的“组合式语言”。

观点动态演化模拟。ABM在群体认知研究中最核心的应用是模拟观点动态，即群体共识、分歧与极化的形成过程。相关研究工作指出[92-93]，与宏观的EBM相比，ABM的优势在于能精细刻画个体的异质性以及复杂的微观交互机制。在建模个体异质性方面，研究者致力于将更丰富的心理学和角色特征注入智能体内核。Li等[78]引入“认知失调”理论扩展ABM模型。其基本思想是，当智能体感知到自身观点与邻居或自身行为产生不一致时，会产生心理张力，进而驱动其修正观点以缓解内在冲突。为建模角色异质性，研究者引入了“知情智能体”[94]，即立场固定不变的“顽固个体”。这类智能体可用于模拟社交网络中的“意见领袖”“水军”或坚定的“党派人士”。模拟结果表明，即便只有极少数此类个体存在，也可能显著改变群体共识的收敛方向与速度，起到“牵引”群体观点的作用。

4.2 基于LLM多智能体的群体认知模拟

经典ABM在形式化和模拟特定社会机制方面取得了巨大成功。但其核心局限在于语义的缺失，智能体处理的观点只是一个数值，规范只是一个符号。它们无法理解观点的内容，也无法在开放的、未曾预设的情境中进行推理。2023年以来，LLM的爆发式增长为ABM带来了根本性的范式迁移，诞生了一批基于大语言模型的多智能体系统（LLM-based multi-agent system, LLM-MAS）相关研究。LLM本身是在海量文本和代码上预训练的、拥有丰富世界知识和复杂推理能力的模型。以LLM作为智能体的“认知内核”，带来了革命性的优势。① 语义理解与生成：智能体能用自然语言进行交互[95]，使观点不再是数值，而是具体的论点[96-97]。② 情境推理与常识：智能体拥有了“常识”[98]和“心智理论”[99-100]的初步能力，能理解复杂的社会情境。③ 行为的涌现性：智能体的复杂行为（如合作、欺骗、从众）可以从LLM的内部推理中涌现，而不需要研究者手动编码[101]。

1) 技术组件

构建一个LLM-MAS群体认知模拟系统，需要认知架构对齐、多智能体协作框架和模拟环境与宏观对齐3个层面的技术组件协同工作。这构成了从“拟人个体”到“拟人群体”的技术全景。

认知架构对齐。为LLM智能体（LLM-Agent）构建一个拟人的、具有认知能力的“大脑”是首要任务。斯坦福大学提出的“生成式智能体”[95]（常被称为“小镇”模拟）是该领域的奠基之作。其核心贡献是为LLM-Agent设计了一个“记忆–反思–规划”的认知架构。该架构不仅是“大脑”，更赋予了其“记忆流”，使其能够记忆、反思和规划长期行为，从而在模拟环境中展现出一致且可信的社会行为。为了使智能体能够参与群体认知形成，仅具备自我认知还不够，还必须能够理解他人的意图、信念与知识状态，即具备心智理论（theory of mind, ToM）能力。Li等[99]深入探讨了LLM-Agent

在协作中的ToM能力。研究发现，LLM（如GPT-4）具备了初步的ToM能力（即建模他人信念和意图），而这种能力是实现高效协作、避免系统性失败的关键。近年来，多个研究致力于探索及进一步提升LLM的ToM能力[100-105]。最后，要构建可信的社会智能体，还需对其“类人社会智能”进行系统评测。近期提出的SocialEval[106]则从协作、竞争、情绪推断、社会规范遵从等维度对LLM社会能力进行了全面评价，为未来构建具备稳健心智能力的多智能体系统提供了重要参考。

多智能体协作框架。为协调多个LLM-Agent，研究者开发了通用的协作框架。AgentVerse[101]是其中的代表，它提出了一个通用的LLM-MAS协作流程，包括“专家招募”“协作决策”“执行”和“评估”等阶段。该框架不局限于特定任务，而是作为一个“元平台”，用于动态组织多个LLM-Agent，以探索协作中涌现的社会行为。此外，在框架设计层面，为解决不同智能体架构难以集成的问题，Hassouna等[107]提出了LLM-Agent-UMF，这是一个统一建模框架，旨在通过定义“核心智能体”组件，实现对多种主动或被动智能体架构的无缝集成。

模拟环境与宏观对齐。高保真的LLM-MAS不仅要做到“微观拟人化”，更要实现“宏观真实性”，即模拟涌现的群体统计规律（如情绪分布、观点极化度）必须与真实大数据观测一致。在通用环境上，AdaSociety[108]提供了具有显式社会结构（如网络关系）和动态事件的多智能体环境，智能体可以在其中进行决策和交互。针对特定社会现象的拟合，研究者构建了专用的模拟环境与宏观对齐方法。例如，Zhang等[109]提出了SocialAlign框架，该框架的核心是实现“微观–宏观”双层对齐，不仅要在微观层面生成“个性化”的评论（拟合个体用户偏好），更要确保在宏观层面生成的“群体情绪分布”与真实社交媒体（如微博）观测到的分布相一致。PopALM[110]则聚焦于“流行度”（如点赞数）的对齐，使用强化学习和课程学习训练LLM，使其生成的响应能更准确地预测真实世界中的“流行度”。这些工作使LLM-MAS向可用于预测的社会数字孪生迈出了关键一步。

2) 典型群体认知现象的前沿模拟

基于上述技术组件，研究者开始复现和探索复杂的群体认知现象，并得出一系列深刻发现。

群体影响。LLM-Agent的“拟人性”是否足以对人类产生真实的社会影响？Song等[111]通过“人–机”交互实验给出了肯定的答案。研究者通过“人–机”交互实验，对比人类用户分别与“单个”或“多个”（如3个或5个）持一致观点的LLM-Agent进行互动。核心发现是：当人类用户与“多个”LLM-Agent互动时，相比于仅与“单个”LLM-Agent（即使总信息量相同）互动，人类会感知到更大的“社会压力”，从而导致更大幅度的观点转变（即“从众”）。这一发现表明LLM-Agent的“拟人性”已经跨过了一个阈值，其“群体性”本身，而非信息量，就足以触发人类在真实社会群体中所依赖的、自动化的“从众”启发式。

观点演化。观点传播与极化是群体认知演化的核心现象之一。近年来，基于大语言模型的多智能体模拟逐渐成为研究这一现象的重要手段。多项最新研究发现，LLM-Agent并非理想的“人类模拟器”，因为它们在对齐训练中被注入了过度的“中立”与“安全”偏好。Chuang等[112]的研究表明，在默认配置下，LLM-Agent（如GPT-3.5）在观点交互过程中难以自然涌现出“极化”或“分歧”结构。只有当研究者通过“提示工程”为智能体显式注入“确认偏误”这一认知偏见后，群体才会出现多簇意见和碎片化格局，更接近传统意见动力学模型中因确认偏误而导致的极化情形。在后续工作中，Cisneros-Velarde[113]通过“公共经费分配”案例研究进一步揭示了LLM-Agent在群体认知演化中的系统性偏见。他们发现，LLM群体在多轮互动后呈现出明显的“共识倾向”“伦理考量”（即便面对带有负面含义的项目，也倾向于保留一定比例的经费）以及“谨慎、避免极端”的决策偏好。这些偏见一方面源于对齐训练对“安全”“中庸”回答的激励，另一方面也会在多智能体互动中被进一步放大，导致某些负面项目在群体讨论中仍然“残存”部分支持度。Ding等[114]则引入了更细粒度的初始条件控制。他们将个体初始心态分布划分为消极、中立、积极等类型，并构造不同的社群结构。结果显示，初始心态分布与社群结构共同决定了系统是收敛到单一共识、形成若干稳定意见簇，还是演化为高度碎片化的网络结构。为了支持更复杂的观点演化研究，Hu等[115]提出了LAIDSim框架，该

框架将LLM融入经典的信息扩散模型，使智能体能够生成细粒度的语言响应，在保留网络层级扩散结构的同时，更真实地刻画不同立场、不同话语策略在传播过程中的互动与演化。在具体系统实现层面，TrendSim[116]构建了一个面向社交媒体“热点话题”的LLM-MAS模拟环境，通过时间感知的交互机制与集中式话题发布模块来复现话题发酵过程，并显式注入投毒攻击等扰动机制，以考察恶意操控对话题极化程度和信息扩散路径的影响。实验显示，LLM的固有偏见及其对齐状态会显著改变极化模式和扩散轨迹，凸显了当前LLM-MAS在认知操控与意见污染模拟中的潜力与局限。

群体决策。Du等[117]探索了LLM的群体决策模拟场景（如“沙漠生存”游戏）。研究发现，与人类群体倾向于快速达成“社会共识”以避免冲突不同，LLM群体在决策过程中产生了更多的“分歧”和“复杂陈述”。由于缺乏“面子”“权威”等社会因素的干扰，LLM群体的“分歧”更具任务导向性，这有助于更充分地探索解空间，从而在某些任务上涌现出高于人类群体的“集体智能”。

大规模社会动力学。尽管LLM-MAS在模拟的“保真度”上取得了革命性突破，但它面临一个根本性的“保真度–可扩展性”权衡。若要实现LLM-Agent的高保真度，其计算和调用成本极其昂贵，导致模拟规模（通常几十个）远小于经典ABM（可达数百万个）。为解决这一核心冲突，研究者提出了“混合模拟”框架，这是平衡模拟“保真度”与“可扩展性”的重要探索。例如，Mou等[118]针对“社会运动”这一大规模现象，采用一种分层混合策略：对于网络中的核心节点（如意见领袖、关键传播者），使用高保真的LLM-Agent来模拟其复杂的、语义驱动的行为；而对于海量的边缘节点（普通跟随者），则使用计算高效的经典ABM（如有界信任模型）进行模拟。这种“LLM-ABM”混合范式，成功实现了在大规模网络上进行高保真社会模拟的目标。

3) 经典ABM与LLM-MAS对比分析

尽管LLM在社会及认知模拟中展现了巨大应用潜力，但其广泛应用也暴露了独特的“对齐税”问题[102]。“对齐税”指的是，由于LLM-Agent的训练偏向于最大化通用性和安全性，它们在某些情境下可能会表现出“过度保守”的行为，这种行为可能不符合群体认知系统中应有的多样性和灵活性。例如，LLM-Agent的“共识倾向”和“极端规避”偏见可能会导致其在认知模拟中产生不符合现实的偏向。这种对齐税使模拟结果可能过于“中庸”，从而错失捕捉社会行为复杂性的机会。

为了更直观地厘清经典ABM与LLM-MAS两种范式的本质区别与优劣势，表6从宏观涌现能力、交互形式等多个维度，对经典ABM与LLM-MAS在群体认知模拟中的差异进行了系统性的对比分析。

5 数据集

高质量的数据集是验证群体认知计算模型有效性的基础。尽管相比于计算机其他领域，群体认知领域尚未形成成熟统一的大规模通用基准，但不同子领域已积累了一批具有代表性的数据集。这些数据集涵盖了从静态的群体偏好记录到动态的群体观点演化，再到新兴的智能体模拟环境。

本节将现有主要数据集资源按前述3条技术主线进行分类梳理，如表7所示。

表6 经典ABM与LLM-MAS在群体认知模拟中对比分析

对比维度	经典ABM	LLM-MAS（LLM作为认知内核）
认知内核	数值规则：处理标量（观点值0.8）或离散状态（S/I/R）	语义知识：处理自然语言，具备一定推理能力
交互形式	信息扩散：简单的数值计算或状态传染	社会互动：依据预设策略进行自然语言交互
宏观涌现能力	统计物理涌现：可模拟涌现出相变、同步、幂律分布等统计规律	社会行为涌现：可模拟涌现出谣言传播、分工协作等高级社会行为
模拟保真度	结构保真：擅长复现宏观网络结构与动力学曲线	过程保真：能复现微观的心理推理过程与对话细节
可重复性	强：同参数同随机种子可复现	弱：受提示、采样与模型版本影响，需要更严格控制
可扩展性	强：可轻松模拟百万级节点	弱：受限于推理成本，需分层混合架构扩展
主要缺点	规则僵化或过简化、参数敏感	存在幻觉与对齐税，行为可能过度趋同或“政治正确”

表 7　网络群体认知计算建模研究常用数据集

类型	数据集	描述
群体认知形成计算建模	CAMRa2011[119]	群组电影推荐，数百用户及家庭组，10 万余条评分记录
	Mafengwo(S) [34]	马蜂窝群组景点签到记录，数千用户，1 万+景点，通过共同旅行日志构建群组
	WeePlaces[44]	群组地点签到记录，7 千余用户，2 万余群组（同一时间在同一地点出现的好友群体）
	Yelp (Group) [33]	基于 LBS 的商业评论数据，包含 3 万余用户、2 万余群组（通过共同活动构建的隐式群组）、110 万余次交互，适用于研究基于社交关系的群体消费决策
群体认知演化计算建模	Twitter-Nuclear/ GenDiscrim[68]	议题立场演化数据，包含用户对“核能”和“性别歧视”议题的观点序列，构建了回复/转发网络（1.5 万条边），适用于连续时间观点动力学建模
	Twitter-AGI/Twitter-HIC[109]	细粒度情绪演化数据，包含 2023—2024 年关于“通用人工智能”和“健康保险”的讨论，特点是具有高频时间戳和细粒度情绪标注，用于微观舆论追踪
	Cora-Dynamic[67]	引文网络动力学数据，虽然原始是静态图，但在神经动力学研究中被转化为连续时间任务，模拟节点特征（认知状态）随引文关系的扩散过程
	POLAR2026[120]	多语言极化检测数据，SemEval 2026 发布的基准，覆盖 7 种语言，包含关于全球事件（如气候变化）的立场与极化标注
	Reddit-Political[121]	政治意识形态演化数据，包含 2007—2024 年 Reddit 政治板块数据，用于量化分析回声室效应和左右翼观点的极化趋势
群体认知行为计算模拟	Smallville (Sandbox) [95]	生成式智能体沙盒环境，包含 25 个具有独立记忆流、反思机制的 LLM 智能体，支持涌现性社会行为（如八卦传播、选举、筹办派对）的模拟
	SocialEval[106]	LLM 社会智能评测基准，包含多个社会互动场景，评估智能体在协作、竞争、说服和规范遵从方面的能力，可用于衡量模拟真实度
	SocialWeibo[109]	模拟-现实对齐数据集，包含微博真实热点事件的评论数据，作为“Ground Truth”，用于训练 LLM 智能体使其生成的群体分布（如情绪比例）与真实人群对齐
	TrendSim[115]	面向“热点话题演化”的 LLM-MAS 模拟，引入时间机制、集中式话题发布与投毒扰动，用于研究极化/扩散在攻击下的变化
	AdaSociety[108]	带显式社会结构与自适应任务生成的多智能体环境，适合研究“社会结构–奖励–行为”的因果/机制关系
	AgentVerse[101]	多智能体协作平台与基准集合，可用于研究协作、规范、群体行为涌现

6　下一步研究工作

如图 5 所示，回顾网络群体认知计算建模的发展历程，技术演进的脉络清晰可见。从早期的规则与理论驱动，到中期的数据驱动，再到当前的生成式智能驱动，研究范式逐渐从依赖人工预设的数学解析与统计规律，转向基于深度学习的大规模非线性数据的精准拟合，进而随着 2023 年通用人工智能技术的兴起，转向对复杂社会行为的语义理解与生成能力迈进。尽管在图结构分析、动态微分方程建模到生成式智能体模拟的过程中取得了显著进展，但面对真实社会系统的极端复杂性，当前研究仍面临着严峻的技术挑战与理论瓶颈。

1) 融合高阶结构与动态演化

当前对群体认知“形成”的建模已广泛采用超图等高阶结构来捕获静态的“多对多”交互。而对“演化”的建模则更侧重于在二元图上捕捉动态过程。这两条技术路线的融合尚不充分。未来研究应着力发展“动态高阶模型”，以同时刻画群体认知演化过程中“交互结构”和“认知状态”的双重动力学。例如，不仅要模拟观点如何变化，还要模拟群体（超边）本身是如何因观点趋同或分化而形成、合并与解散的。

2) 从数据驱动到“理论–数据”融合

纯粹的数据驱动模型（如 GNN、ODE）虽然拟合能力强，但往往是“黑盒”，缺乏可解释性，且在数据稀疏时性能表现不佳。第 3 节介绍的 SINN 模型通过“社会学知识融入”的混合建模策略，为解决上述问题提供了有效思路。沿着这一方向，未来的研究重点可聚焦于如何更深层次地将经典社会学理论与深度模型相融合，以及如何设计出能显式挖掘、表征社会规律的深度学习模型，从而在提升模型可解释性与泛化能力上取得实质性突破。

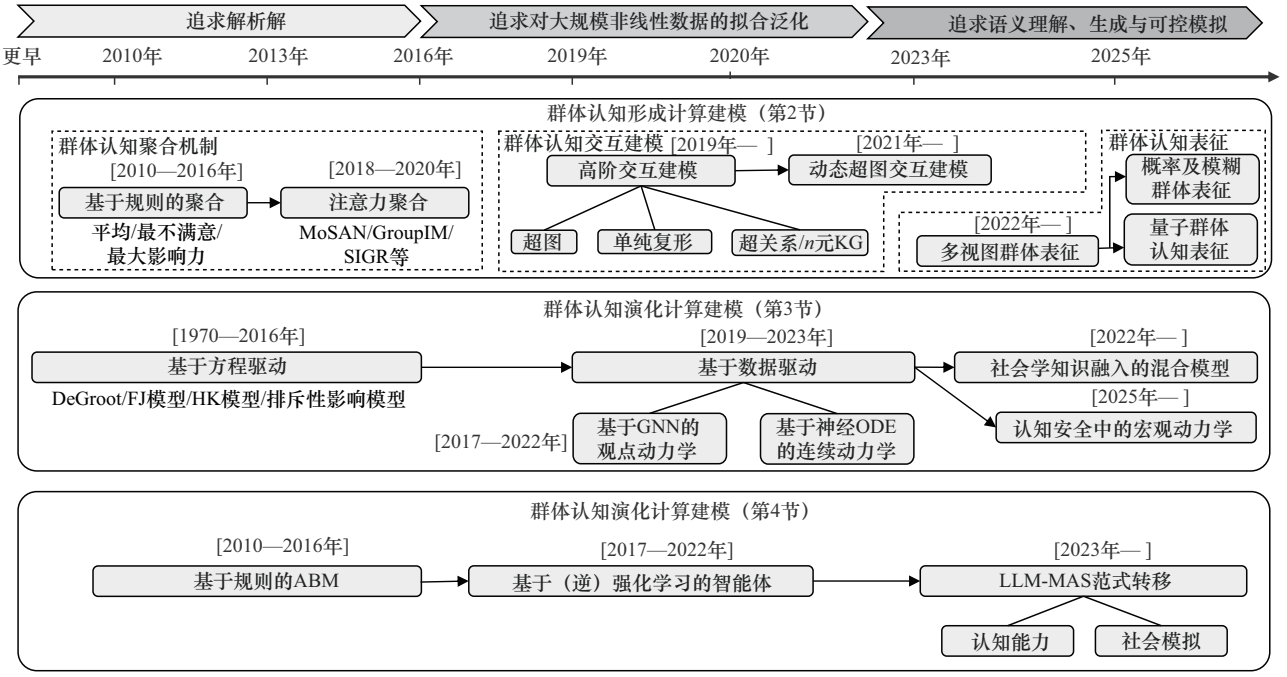


图5　网络群体认知计算建模相关技术研究时间线

3) 兼顾保真度与可扩展性的混合模拟

第4节揭示了模拟范式从经典ABM向量“可扩展性”到LLM-MAS语义“保真度”的演进。然而，LLM-MAS高昂的计算成本使其难以模拟大规模社会动力学。因此，研究平衡“保真度–可扩展性”的混合模拟框架是必然趋势。如HiSim所示[118]，未来研究可探索对于网络中哪些关键节点（如意见领袖、活跃传播者、桥接节点），需要采用高保真的LLM-Agent进行精细模拟，以捕捉其复杂的语义生成和策略行为。而对于海量的长尾节点，则自动回退到计算高效的经典ABM规则或统计模型。此外，利用LLM来“蒸馏”生成轻量级的行为规则或小模型，赋予普通智能体以较低成本的“类LLM”能力，也是值得探索的方向。这种“核心–边缘”混合建模策略将在保持宏观涌现特征真实性的同时，大幅降低计算开销。

4) 可控、可信的LLM认知智能体

LLM-MAS为群体认知模拟带来了语义和推理的革命。但第4节的分析表明，LLM-Agent受其对齐训练的影响，存在固有的“共识倾向”和“极端规避”偏见，难以自然复现“极化”等真实社会现象。这导致了模拟与现实的偏差。面对这一问题，未来的核心挑战在于：既要精确量化这些偏差对建模结果的具体影响，又要构建更加“可控”的认知内核。在此基础上，利用SocialAlign等宏观对齐框架，构建SocialEval[106]等偏差量化评估方法，通过强化学习与在线学习加入更多的反馈调整机制，将成为弥合偏差、确保模拟结果“可信度”的关键路径。

5) 基于因果推断与反事实的干预评估

从方法论角度看，现有研究仍以相关性建模为主，难以回答“若采取某种干预，群体认知将如何变化”这一决策导向问题。未来有必要系统引入因果推断与反事实分析方法，以刻画平台机制、推荐策略或信息干预对群体认知演化的因果效应。具体而言，可探索将基于图结构的因果发现方法（如基于条件独立检验或结构方程模型（structural equation model, SEM）的因果图学习）、时间序列因果推断方法（如格兰杰因果分析、动态贝叶斯网络）以及反事实推演框架（如结构因果模型（structural causal model, SCM）与干预模拟）引入群体认知计算建模中。结合生成式模拟，这类方法有望实现对干预策略的事前评估与风险预演，推动群体认知研究从“拟合过去”迈向可量化、可验证的“干预未来”。

6) 面向认知安全的对抗防御

随着群体认知模型从“描述性拟合”向“生成式模拟”演进，其在认知安全领域的潜在风险愈发显著。对群体认知机制的深入理解可能被滥用于认

知操纵、舆论投毒等攻击场景，因此，亟须研究模型在恶意干预下的鲁棒性与防御机制。随着认知操纵手段的升级，如何防止虚假信息渗透和水军协同干预，已成为提升系统安全性的关键。未来研究应结合博弈论与对抗防御设计，探索构建更加安全、抗攻击的群体认知建模系统以及虚拟认知攻防靶场，以应对动态复杂的认知对抗风险，确保认知系统的可信与可控。

7　结束语

网络已成为塑造群体认知、影响社会决策的关键环境。理解其背后的计算机制是计算机与社会科学交叉领域的前沿课题。然而，相关研究分散于不同学科，缺乏统一的计算视角。

本文从计算机科学视角出发，对网络群体认知的计算建模研究进行了系统性综述，构建了一个统一的分析框架，将现有研究划分为3个既相互区别又紧密联系的核心领域。1) 群体认知形成计算建模：聚焦于“涌现”，系统梳理了如何利用超图、单纯复形等高阶结构和注意力、GNN等聚合机制，对群体共识的形成进行表征与建模。2) 群体认知演化计算建模：聚焦于“动态”，深入分析了群体观点如何从基于方程的经典动力学（如DeGroot、FJ、HK）演进到数据驱动的复杂动力学（如GNN、神经ODE）。3) 群体认知行为计算模拟：聚焦于“生成”，探讨了“自底向上”的仿真范式如何从经典ABM演进到LLM-MAS。

通过对这3个主题的技术脉络、内在联系与研究趋势的剖析，本文试图为理解和驾驭日益复杂的数字社会群体认知研究提供清晰的计算框架与路径参考，并为未来在认知安全、群体认知计算建模等方向的研究提供启示。

参考文献：

[1] Fortunato S. Community detection in graphs[J]. Physics Reports, 2010, 486(3-5): 75-174.

[2] Momennejad I, Duker A, Coman A. Bridge ties bind collective memories[J]. Nature Communications, 2019, 10: 1578.

[3] Theiner G, Allen C, Goldstone R L. Recognizing group cognition[J]. Cognitive Systems Research, 2010, 11(4): 378-395.

[4] Lu K, Zhang H L, Sun T Z, et al. GMCL: graph-enhanced multimodal contrastive learning for rumor detection[C]//Proceedings of the ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2025: 1-5.

[5] Claverie B, Cluzel F D. “Cognitive warfare”: the advent of the concept of “cognitics” in the field of warfare[J]. Cognitive Psychology, 2022, 2: 1-7.

[6] Forsyth D R. Group dynamics[M]. Calif: Wadsworth/Cengage Learning, 2010.

[7] Palermos S O. The dynamics of group cognition[J]. Minds and Machines, 2016, 26(4): 409-440.

[8] Myers D G, Lamm H. The group polarization phenomenon[J]. Psychological Bulletin, 1976, 83(4): 602-627.

[9] Castellano C, Fortunato S, Loreto V. Statistical physics of social dynamics[J]. Reviews of Modern Physics, 2009, 81(2): 591-646.

[10] Xu R X, Sun Y F, Ren M J, et al. AI for social science and social science of AI: a survey[J]. Information Processing & Management, 2024, 61(3): 103665.

[11] Dara S, Chowdary C R, Kumar C. A survey on group recommender systems[J]. Journal of Intelligent Information Systems, 2020, 54(2): 271-295.

[12] Dong Y C, Zhan M, Kou G, et al. A survey on the fusion process in opinion dynamics[J]. Information Fusion, 2018, 43: 57-65.

[13] Amirkhani A, Barshooi A H. Consensus in multi-agent systems: a review[J]. Artificial Intelligence Review, 2022, 55(5): 3897-3935.

[14] 潘理, 吴鹏, 黄丹华. 在线社交网络群体发现研究进展[J]. 电子与信息学报, 2017, 39(9): 2097-2107.
Pan L, Wu P, Huang D H. Reviews on group detection in online social networks[J]. Journal of Electronics & Information Technology, 2017, 39(9): 2097-2107.

[15] Caldeira C. Group cognition: computer support for building collaborative knowledge[J]. Journal of the American Society for Information Science and Technology, 2008, 59(9): 1531.

[16] Fu X L, Cai L H, Liu Y, et al. A computational cognition model of perception, memory, and judgment[J]. Science China Information Sciences, 2014, 57(3): 4911.

[17] Chalmers D J. A computational foundation for the study of cognition[J]. Journal of Cognitive Science, 2011, 12(4): 325-359.

[18] Griffiths T L, Chater N, Tenenbaum J B. Bayesian models of cognition[M]. Cambridge, MA, USA: The MIT Press, 2024.

[19] Binz M, Akata E, Bethge M, et al. A foundation model to predict and capture human cognition[J]. Nature, 2025, 644(8078): 1002-1009.

[20] Fiske S T, Taylor S E. Social cognition: from brains to culture[M]. London EC1Y 1SP: SAGE Publications Ltd, 2013.

[21] Lockwood P L, Wittenberg G M, Heymann G, et al. Computational modelling of social cognition and behaviour[J]. Social Cognitive and Affective Neuroscience, 2021, 16(8): 761-771.

[22] Mischel W, Shoda Y. A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure[J]. Psychological Review, 1995, 102(2): 246-268.

[23] Xu Z H, Zhang Y, Wu Y, et al. Modeling user posting behavior on social media[C]//Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2012: 545-554.

[24] Griffiths T L, Tenenbaum J B. Optimal predictions in everyday cognition[J]. Psychological Science, 2006, 17(9): 767-773.

[25] Hutchins E. Cognition in the wild[M]. Cambridge: The MIT Press, 1995.

[26] Hollan J, Hutchins E, Kirsh D. Distributed cognition: toward a new

foundation for human-computer interaction research[J]. ACM Transactions on Computer-Human Interaction, 2000, 7(2): 174-196.

[27] Nguyen V D, Tran V C, Truong H B, et al. Social networks as platforms for enhancing collective intelligence[J]. Cybernetics and Systems, 2022, 53(5): 425-442.

[28] Aisa B, Mingus B, O'Reilly R. The emergent neural modeling system[J]. Neural Networks, 2008, 21(8): 1146-1152.

[29] Lee G, Bu F C, Eliassi-Rad T, et al. A survey on hypergraph mining: patterns, tools, and generators[J]. ACM Computing Surveys, 2025, 57(8): 1-36.

[30] Feng Y F, You H X, Zhang Z Z, et al. Hypergraph neural networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 3558-3565.

[31] Antelmi A, Cordasco G, Polato M, et al. A survey on hypergraph representation learning[J]. ACM Computing Surveys, 2024, 56(1): 1-38.

[32] Huang J, Yang J. UniGNN: a unified framework for graph and hypergraph neural networks[C]//Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence. Piscataway: IEEE Press, 2021: 2563-2569.

[33] Wu X X, Xiong Y, Zhang Y, et al. ConsRec: learning consensus behind interactions for group recommendation[C]//Proceedings of the ACM Web Conference 2023. New York: ACM Press, 2023: 240-250.

[34] Xu J F, Chen Z Y, Li J Z, et al. AlignGroup: learning and aligning group consensus with member preferences for group recommendation[C]// Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2024: 2682-2691.

[35] Zhu X L, Wang D L, Li J X, et al. Dynamical attention hypergraph convolutional network for group activity recognition[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(5): 8911-8925.

[36] Majhi S, Perc M, Ghosh D. Dynamics on higher-order networks: a review[J]. Journal of the Royal Society Interface, 2022, 19(188): 20220043.

[37] Chen Y Z, Gel Y R, Marathe M V, et al. A simplicial epidemic model for COVID-19 spread analysis[J]. Proceedings of the National Academy of Sciences of the United States of America, 2023, 121: e2313171120.

[38] Wang D, Zhao Y, Leng H, et al. A social communication model based on simplicial complexes[J]. Physics Letters A, 2020, 384(35): 126895.

[39] Yin G Z, Zhang H L, Yang Y C, et al. Inductive link prediction on N-ary relational facts via semantic hypergraph reasoning[C]//Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2025: 1821-1832.

[40] Galkin M, Trivedi P, Maheshwari G, et al. Message passing for hyper-relational knowledge graphs[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg: ACL, 2020: 7346-7359.

[41] Guan S P, Cheng X Q, Bai L, et al. What is event knowledge graph: a survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(7): 7569-7589.

[42] Papanikolaou N, Vaccario G, Hormann E, et al. Consensus from group interactions: an adaptive voter model on hypergraphs[J]. Physical Review E, 2022, 105(5): 054307.

[43] Tran L V, Pham T N, Tay Y, et al. Interact and decide: medley of sub-attention networks for effective group recommendation[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2019: 255-264.

[44] Sankar A, Wu Y H, Wu Y H, et al. GroupIM: a mutual information maximization framework for neural group recommendation[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2020: 1279-1288.

[45] Yin H Z, Wang Q Y, Zheng K, et al. Social influence-based group representation learning for group recommendation[C]//Proceedings of the 2019 IEEE 35th International Conference on Data Engineering (ICDE). Piscataway: IEEE Press, 2019: 566-577.

[46] Duan M J, Zheng T Y, Gao Y, et al. DGA-GNN: dynamic grouping aggregation GNN for fraud detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(10): 11820-11828.

[47] Kou Y, Li D, Shen D R, et al. A self-supervised group recommendation model with conformity awareness[J]. Scientific Reports, 2025, 15: 35937.

[48] Yin G Z, Wang X, Zhang H L, et al. Beyond individuals: modeling mutual and multiple interactions for inductive link prediction between groups[C]//Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2023: 751-759.

[49] Ekstrand M D, Carterette B, Diaz F. Distributionally-informed recommender system evaluation[J]. ACM Transactions on Recommender Systems, 2024, 2(1): 1-27.

[50] Jiang J Y, Yang D Q, Xiao Y H, et al. Convolutional Gaussian embeddings for personalized recommendation with uncertainty[C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. 2019: 2642-2648.

[51] Jia Q L, Zhou X L, Krejcar O, et al. Fuzzy information evolution with three-way decision in social network group decision-making[J]. IEEE Transactions on Fuzzy Systems, 2025, 33(12): 4331-4344.

[52] Bruza P D, Wang Z, Busemeyer J R. Quantum cognition: a new theoretical approach to psychology[J]. Trends in Cognitive Sciences, 2015, 19(7): 383-393.

[53] Dong Y C, Zha Q B, Zhang H J, et al. Consensus reaching in social network group decision making: research paradigms and challenges[J]. Knowledge-Based Systems, 2018, 162: 3-13.

[54] Han S L, Liu X W. An extension of multi-attribute group decision making method based on quantum-like Bayesian network considering the interference of beliefs[J]. Information Fusion, 2023, 95: 143-162.

[55] Li Y Y, Liu P D, Wu Z B. Social network group consensus model considering quantum cognition-based social interaction pattern and individual utility[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2025, 55(9): 6369-6382.

[56] Degroot M H. Reaching a consensus[J]. Journal of the American Statistical Association, 1974, 69(345): 118-121.

[57] Zhou X T, Sun H X, Xu W Y, et al. Friedkin-johnsen model for opinion dynamics on signed graphs[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(12): 8313-8327.

[58] Yang Y C, Dimarogonas D V, Hu X M. Opinion consensus of modified Hegselmann-Krause models[J]. Automatica, 2014, 50(2): 622-627.

[59] Wang L F, Bernardo C, Hong Y G, et al. Consensus in concatenated opinion dynamics with stubborn agents[J]. IEEE Transactions on Automatic Control, 2023, 68(7): 4008-4023.

[60] Jager W, Amblard F. Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change[J]. Computational & Mathematical Organization Theory, 2005, 10(4): 295-303.

[61] Sabin-Miller D, Abrams D M. When pull turns to shove: a continuous-time model for opinion dynamics[J]. Physical Review Research, 2020, 2(4): 043001.

[62] Nettasinghe B, Rao A, Jiang B H, et al. In-group love, out-group hate: a framework to measure affective polarization via contentious online discussions[C]//Proceedings of the ACM on Web Conference 2025. New York: ACM Press, 2025: 560-575.

[63] 刘佳豪, 夏一雪, 沈宇航, 等. AIGC环境下群体认知入侵机理与防范策略[J]. 情报杂志, 2025, 44(9): 154-163, 111.
Liu J H, Xia Y X, Shen Y H, et al. Mechanism and defense strategies of group cognitive intrusion in the context of AIGC[J]. Journal of Intelligence, 2025, 44(9): 154-163, 111.

[64] 黄凤翔, 夏一雪, 兰月新. 认知战环境下群体认知演化动力学机理研究[J]. 情报杂志, 2025, 44(8): 67-77, 42.
Huang F X, Xia Y X, Lan Y X. Research on the dynamics mechanism of group cognitive evolution in a cognitive warfare environment[J]. Journal of Intelligence, 2025, 44(8): 67-77, 42.

[65] Li H, Jiang H, Zheng Y K, et al. UniGO: a unified graph neural network for modeling opinion dynamics on graphs[C]//Proceedings of the ACM on Web Conference 2025. New York: ACM Press, 2025: 530-540.

[66] Chen R T Q, Rubanova Y, Bettencourt J, et al. Neural ordinary differential equations[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM Press, 2018: 6572-6583.

[67] Zang C X, Wang F. Neural dynamics on complex networks[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2020: 892-902.

[68] Duan B W, Deng H G, Piao J H, et al. Bi-dynamic graph ODE for opinion evolution[C]//Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1. New York: ACM Press, 2025: 260-270.

[69] Okawa M, Iwata T. Predicting opinion dynamics via sociologically-informed neural networks[C]//Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2022: 1306-1316.

[70] Cuomo S, Cola V S D, Giampaolo F, et al. Scientific machine learning through physics-informed neural networks: where we are and what's next[J]. Journal of Scientific Computing, 2022, 92(3): 88.

[71] Bakaryan T, Gu Y, Hovakimyan N, et al. Multi-population opinion dynamics model [J]. Nonlinear Dynamics, 2025, 113(1): 559-580.

[72] Dandekar P, Goel A, Lee D T. Biased assimilation, homophily, and the dynamics of polarization[J]. Proceedings of the National Academy of Sciences of the United States of America, 2013, 110(15): 5791-5796.

[73] Hohmann M, Devriendt K, Coscia M. Quantifying ideological polarization on a network using generalized Euclidean distance[J]. Science Advances, 2023, 9(9): eabq2044.

[74] Esteban J M, Ray D. On the measurement of polarization[J]. Econometrica, 1994, 62(4): 819-851.

[75] Zhang J, Hamilton W L, Danescu-Niculescu-Mizil C, et al. Community identity and user engagement in a multi-community landscape[J]. Proceedings of the International AAAI Conference on Web and Social Media, 2017, 11(1): 377-386.

[76] Olfiati-Saber R, Fax J A, Murray R M. Consensus and cooperation in networked multi-agent systems [J]. Proceedings of the IEEE, 2007, 95(1): 215-233.

[77] Varma V S, Morărescu I C, Hayel Y. Continuous time opinion dynamics of agents with multi-leveled opinions and binary actions[C]//Proceedings of the IEEE INFOCOM 2018-IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2018: 1169-1177.

[78] Li K, Liang H M, Kou G, et al. Opinion dynamics model based on the cognitive dissonance: an agent-based simulation[J]. Information Fusion, 2020, 56: 1-14.

[79] Wen M N, Kuba J G, Lin R J, et al. Multi-agent reinforcement learning is a sequence modeling problem[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New York: ACM Press, 2022: 16509-16521.

[80] Ndousse K, Eck D, Levine S, et al. Emergent social learning via multi-agent reinforcement learning[C]//International Conference on Machine Learning (ICML). New York: PMLR, 2021: 7991-8004.

[81] Abel D, Macglashan J, Littman M L. Reinforcement learning as a framework for ethical decision making[C]//Proceedings of the AAAI Conference on Artificial Intelligence Workshops. 2016:1-8.

[82] Metelli A M. Recent advancements in inverse reinforcement learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(20): 22680.

[83] Zeng Y X, Xu K, Yin Q J, et al. Inverse reinforcement learning based human behavior modeling for goal recognition in dynamic local network interdiction[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018: 646-653.

[84] Min H Y, Cao J X, Ge J W, et al. A multi-agent system for fine-grained opinion dynamics analysis in online social networks[J]. IEEE Transactions on Computational Social Systems, 2024, 11(1): 815-828.

[85] Vargas-Pérez V A, Giráldez-Cru J, Mesejo P, et al. Unveiling agents' confidence in opinion dynamics models via graph neural networks[J]. IEEE Transactions on Computational Social Systems, 2025, 12(2): 725-737.

[86] Haeringen E S V, Gerritsen C, Hindriks K V. Emotion contagion in agent-based simulations of crowds: a systematic review[J]. Autonomous Agents and Multi-Agent Systems, 2023, 37: 6.

[87] Bosse T, Duell R, Memon Z A, et al. Agent-based modeling of emotion contagion in groups[J]. Cognitive Computation, 2015, 7(1): 111-136.

[88] Fan R, Xu K, Zhao J C. An agent-based model for emotion contagion and competition in online social media[J]. Physica A: Statistical Mechanics and Its Applications, 2018, 495: 245-259.

[89] Ajmeri N, Guo H, Murukannaiah P K, et al. Robust norm emergence by revealing and reasoning about context: socially intelligent agents for enhancing privacy[C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. 2018: 28-34.

[90] Agrawal R, Ajmeri N, Singh M. Socially intelligent genetic agents for the emergence of explicit norms[C]//Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence. 2022: 10-16.

[91] Mordatch I, Abbeel P. Emergence of grounded compositional language in multi-agent populations[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1): 1495-1502

[92] Flache A, Mäs M, Feliciani T, et al. Models of social influence: towards the next frontiers[J]. Journal of Artificial Societies and Social

Simulation, 2017, 20(4): 2.

[93] Li G J, Porter M A. Bounded-confidence model of opinion dynamics with heterogeneous node-activity levels[J]. Physical Review Research, 2023, 5(2): 023179.

[94] Fan K Q, Pedrycz W. Opinion evolution influenced by informed agents[J]. Physica A: Statistical Mechanics and Its Applications, 2016, 462: 431-441.

[95] Park J S, O'Brien J, Cai C J, et al. Generative agents: interactive simulacra of human behavior[C]//Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. New York: ACM Press, 2023: 1-22.

[96] Guo T C, Chen X Y, Wang Y Q, et al. Large language model based multi-agents: a survey of progress and challenges[C]//Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence. New York: ACM Press, 2024: 8048-8057.

[97] Zhu A, Dugan L, Callison-Burch C. ReDel: a toolkit for LLM-powered recursive multi-agent systems[C]//Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Stroudsburg: ACL, 2024: 162-171.

[98] Agashe S, Fan Y, Reyna A, et al. LLM-coordination: evaluating and analyzing multi-agent coordination abilities in large language models[C]//Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2025. Stroudsburg: ACL, 2025: 8038-8057.

[99] Li H A, Chong Y, Stepputtis S, et al. Theory of mind for multi-agent collaboration via large language models[C]//Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2023: 180-192.

[100] Shi H J, Ye S Y, Fang X Y, et al. MuMA-ToM: multi-modal multi-agent theory of mind[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(2): 1510-1519.

[101] Chen W Z, Su Y S, Zuo J W, et al. Agentverse: facilitating multi-agent collaboration and exploring emergent behaviors[C]//Proceedings of the International Conference on Learning Representations (ICLR). 2024.

[102] Lin Y, Lin H Y, Xiong W, et al. Mitigating the alignment tax of RLHF[C]//Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2024: 580-606.

[103] Wu Y F, He Y H, Jia Y L, et al. Hi-ToM: a benchmark for evaluating higher-order theory of mind reasoning in large language models[C]//Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2023. Stroudsburg: ACL, 2023: 10691-10706.

[104] Wilf A, Lee S, Liang P P, et al. Think twice: perspective-taking improves large language models' theory-of-mind capabilities[C]//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: ACL, 2024: 8292-8308.

[105] Sclar M, Kumar S, West P, et al. Minding language models' (lack of) theory of mind: a plug-and-play multi-character belief tracker[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: ACL, 2023: 13960-13980.

[106] Zhou J F, Chen Y X, Shi Y H, et al. SocialEval: evaluating social intelligence of large language models[C]//Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: ACL, 2025: 30958-31012.

[107] Hassouna A B, Chaari H, Belhaj I. LLM-agent-UMF: LLM-based agent unified modeling framework for seamless design of multi active/passive core-agent architectures[J]. Information Fusion, 2026, 127: 103865.

[108] Bi M J, Feng X, Huang Y Z, et al. AdaSociety: an adaptive environment with social structures for multi-agent decision-making[C]//Proceedings of the Advances in Neural Information Processing Systems 37. 2024: 35388-35413.

[109] Zhang J H, Wan K Y, Xu L W, et al. From individuals to crowds: dual-level public response prediction in social media[C]//Proceedings of the 33rd ACM International Conference on Multimedia. New York: ACM Press, 2025: 5903-5912.

[110] Yu E X, Li J, Xu C P. PopALM: popularity-aligned language models for social media trendy response prediction[C]//Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). Kerrville: Association for Computational Linguistics, 2024: 12867-12878.

[111] Song T Q, Tan Y, Zhu Z C, et al. Multi-agents are social groups: investigating social influence of multiple agents in human-agent interactions[J]. Proceedings of the ACM on Human-Computer Interaction, 2025, 9(7): 1-33.

[112] Chuang Y S, Goyal A, Harlalka N, et al. Simulating opinion dynamics with networks of LLM-based agents[C]//Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2024. Stroudsburg: ACL, 2024: 3326-3346.

[113] Cisneros-Velarde P. Biases in opinion dynamics in multi-agent systems of large language models: a case study on funding allocation[C]//Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2025. Stroudsburg: ACL, 2025: 1889-1916.

[114] Ding G Z, Liu Z E, Li S, et al. Impact of mindset types and social community compositions on opinion dynamics: a large language model-based multi-agent simulation study[J]. Computers in Human Behavior, 2025, 172: 108730.

[115] Hu Y X, Sherpa G, Zhang L, et al. An LLM-enhanced agent-based simulation tool for information propagation[C]//Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence. New York: ACM Press, 2024: 8679-8682.

[116] Zhang Z Y, Lian J X, Ma C, et al. TrendSim: simulating trending topics in social media under poisoning attacks with LLM-based multi-agent system[C]//Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2025. Stroudsburg: ACL, 2025: 2930-2949.

[117] Du Y N, Rajivan P, Gonzalez C. Large language models for collective problem-solving: insights into group consensus decision-making[C]//Proceedings of the 46th Annual Meeting of the Cognitive Science Society. Rotterdam: Cognitive Science Society, 2024: 1040-1047.

[118] Mou X Y, Wei Z Y, Huang X J. Unveiling the truth and facilitating change: towards agent-based large-scale social movement simulation[C]//Proceedings of the Findings of the Association for Computational Linguistics ACL 2024. Stroudsburg: ACL, 2024: 4789-4809.

[119] Cao D, He X N, Miao L H, et al. Attentive group recommendation[C]//Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. New York: ACM Press, 2018: 645-654.

[120] Naseem U. SemEval-2026 task 9: detecting multilingual, multicul-

tural and multievent online polarization[C]//Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026). Stroudsburg: ACL, 2026.

[121] Waller I, Anderson A. Quantifying social organization and political polarization in online platforms[J]. Nature, 2021, 600(7888): 264-268.

[作者简介]

尹公主（1996– ），女，湖北武汉人，博士，哈尔滨工业大学助理研究员，主要研究方向为信息内容安全、关系挖掘、认知安全等。

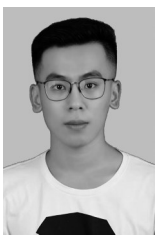
张宏莉（1973– ），女，吉林榆树人，博士，哈尔滨工业大学教授、博士生导师，主要研究方向为社交网络分析、网络与信息安全、认知安全等。

田逸艺（2002– ），女，云南昆明人，哈尔滨工业大学硕士生，主要研究方向为社交网络立场分析。

田泽庶（1997– ），男，黑龙江哈尔滨人，博士，哈尔滨工业大学副研究员，主要研究方向为知识图谱构建及应用、内容安全、认知安全等。

孟辰（2002– ），男，吉林珲春人，哈尔滨工业大学硕士生，主要研究方向为图神经网络、知识图谱、认知安全等。

何嘉豪（2003– ），男，福建莆田人，哈尔滨工业大学硕士生，主要研究方向为社交网络分析、认知安全等。



黄圣鹏（2000– ），男，江西抚州人，哈尔滨工业大学硕士生，主要研究方向为多智能体、认知安全等。