

面向舆情治理的信息管理机制演化博弈模型

刘淑婷^{1,2}, 马颖华^{1,2}, 陈秀真^{1,2}

(1. 上海交通大学网络安全技术研究院, 上海 200240; 2. 上海市信息安全综合管理技术重点实验室, 上海 200240)

摘要: 在线社交网络平台中存在的冲突信息加剧了群体观念对立, 需要科学合理的管理机制来实现舆情治理。提出面向舆情治理的信息管理机制演化博弈模型, 通过建立动态奖惩策略, 自动化地实现冲突抑制, 激活网络自我净化能力。通过引入演化博弈理论, 构建基于累积前景理论的社交网络用户信息冲突演化博弈模型, 使用数值模拟的方法研究 9 个约束条件下的 18 种理论应用情景的演化结果, 分析了双方行为互动演化规律, 得出约束条件五中博弈双方理性对待信息冲突的结论。通过建立博弈分析的系统动力学模型, 对动力学演变进行分析, 并在满足约束条件五的前提下对模型参数进行了详细的灵敏度分析, 并为动态信息管理机制提供了实用化的策略建议。所提模型首次对利益驱动、从众心理等多种因素在舆情演化中的相互作用进行建模, 演化分析结果与谣言优势传播规律相符合, 证实了所提模型的合理性。所提模型为社交网络中恶意信息控制系统的开发和设计提供了重要借鉴, 为舆情治理的研究实践提供了新的思路。

关键词: 演化博弈; 信息管理; 信誉; 社交网络; 舆情治理

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2023086

Evolutionary game model of information management mechanism for public opinion governance

LIU Shuting^{1,2}, MA Yinghua^{1,2}, CHEN Xiuzhen^{1,2}

1. Institute of Cyber Science and Technology, Shanghai Jiao Tong University, Shanghai 200240, China

2. Shanghai Municipal Key Lab of Integrated Management Technology for Information Security, Shanghai 200240, China

Abstract: Conflicting information in online social networks, often stemming from opposing views, can lead to rumors and the fragmentation of society. A reputation-based dynamic information management mechanism was proposed, adopting a dynamic reward and punishment strategy to decrease unnecessary information conflict so as to activate self-purification of network. By utilizing evolutionary game theory, a cumulative-prospect-based evolutionary game

收稿日期: 2022-11-21; 修回日期: 2023-04-23

通信作者: 陈秀真, chenxz@sjtu.edu.cn

基金项目: 国家自然科学基金(U2003206); 上海市科委“科技创新行动计划”(22511101202)

Foundation Items: The National Natural Science Foundation of China (U2003206), Science and Technology Innovation Program of Science and Technology Commission of Shanghai Municipality (22511101202)

引用格式: 刘淑婷, 马颖华, 陈秀真. 面向舆情治理的信息管理机制演化博弈模型[J]. 网络与信息安全学报, 2023, 9(6): 102-115.

Citation Format: LIU S T, MA Y H, CHEN X Z. Evolutionary game model of information management mechanism for public opinion governance[J]. Chinese Journal of Network and Information Security, 2023, 9(6): 102-115.

model of social network information conflict was constructed to validate the efficacy of the proposed mechanism. Numerical simulations were conducted to analyze the evolutionary trends of 18 different theoretical application scenarios under 9 constraints. The results demonstrate that, under constraint 5, both sides of the game tend to exhibit relatively rational behavior towards conflicts. Furthermore, dynamic evolution analysis and sensitivity analysis of the model parameters were performed while satisfying constraint 5, providing practical strategic suggestions for the dynamic information management mechanism. This model represents a novel attempt to capture the interaction of multiple factors, such as interest-driven behavior and conformity psychology, in the evolution of public opinion. Moreover, the model successfully simulates well-known patterns of rumor dissemination, confirming its validity. The proposed model offers valuable insights for the development and design of malicious information control systems in social networks and provides new ideas for the research and implementation of public opinion governance.

Keywords: evolutionary game, information management, reputation, social network, public opinion governance

0 引言

新兴社交媒体使信息传播速度更快、范围更广，为公众提供便捷的发声平台，并重塑了网络舆情格局。对立竞争的舆情相关信息间的冲突作用可能加剧群体观点对立，引起舆情高峰，推动舆情危机演化成谣言^[1]，误导公众舆论，甚至扰乱正常社会秩序，危及社会安全和稳定。因此，探索社交网络舆情治理引导手段，有助于维护健康清朗的网络空间环境，对社会长治久安具有重要意义。

舆情研究可分为相互关联的3类：舆情分析、舆情传播、舆情引导。其中，舆情分析对舆情信息进行特征提取和智能分类；舆情传播借助传播动力学等理论构建舆情传播模型，预测传播路径和传播规模；舆情引导对舆情传播进行实用性的控制，降低负面舆情的有害影响^[2]。

舆情引导研究工作中对负面舆情的处理分为两类：“事前预防型”和“事后补救型”。“事后补救型”的研究较多，然而因为网络舆情监测和传播模型准确率有限，实施效果不明显。在非技术领域，尽管法律提供了对负面舆情发布传播者的惩罚依据和指导，但成本较高，且现有的大多数引导策略是从宏观角度出发的，缺乏科学性和多样性，难以落地^[3]。

本文从“事前预防”的角度出发，提出面向舆情治理的信息管理机制演化博弈模型，通过演化博弈建模分析验证了信息管理机制的有效性；通过对争议信息实施公开的冲突抑制策略，可引导社交网络用户理性发布和传播信息，及时降低争议信息传播范围，减少舆情危机的产生和传播，

充分发挥网络自我净化能力，协同治理网络环境。

1 相关工作

本文研究主要涉及负面舆情的分析处理，目前已有工作从“事前预防”和“事后补救”两个角度开展，具体工作如下。

“事前预防型”舆情引导倾向于减少谣言的产生和传播，将舆情问题解决在成灾之前。一些研究集中在改善信息管理机制，通过引入信誉来激励和规制用户的行为^[4]。文献[5]提出了利用虚拟信息信用来进行每个对等信息交换，避免谣言的快速、密集传播。一些研究则集中在运用系统动力学和传播学模型，对新机制下的舆情信息传播动态进行仿真，探索网络拓扑结构、遗忘机制、反馈机制、从众效应、权威效应、信息的模糊性与吸引力特征、噪声环境、系统开放程度、传播环境差异等相关因素对传播过程的影响，从而提出舆论引导建议，反过来指导信息管理机制的完善^[6-9]。但是，系统动力学传播模型没有考虑用户的决策自主能力和从众的特点，预测精度较低，且无法体现用户交互作用对决策行为的影响。还有一些研究采用博弈理论构建社交网络信息传播决策模型^[10]，如文献[11]应用博弈论研究在线社交网络中商品广播过程，把信息传播的过程看作用户在交互中博弈争取最大利益的过程，验证了博弈论在信息冲突分析中的可行性。但是大部分博弈研究采用期望效用理论^[12]（假设人类完全理性和策略执行力固定），偏离现实，容易产生悖论。

“事后补救型”舆情引导措施有遏制传播和证

伪澄清,前者指在有较高影响力的用户或社区桥梁进行谣言等有害信息的屏蔽^[13-15],即定位负面舆情传播者,阻断传播路径;后者是控制某些节点传播真相来减轻信息污染^[16-18],即选取澄清者,掌握网络话语权。“事后补救型”舆情引导措施需要及时且高精度的舆情监测和谣言识别作为前提条件。抑制的效果取决于遏制集和澄清集的选择,澄清效果受许多社会因素影响,澄清节点对于一部分用户是权威的,对其他用户则可能不是,因此,抑制措施的具体实施存在较大困难。

与以往研究相比较,本文工作的特点在于:
① 提出基于信誉的动态信息管理机制,分析了该机制对舆情信息引起的用户观点冲突的影响;
② 采用演化博弈的方法,其有限理性和动态演化的前提比较符合社交网络用户的行为规律和决策特征;
③ 引入累积前景理论^[19]和从众效应^[20]对用户决策过程的心理价值进行量化;
④ 提出公开的面向舆情治理的定量化信息管理策略建议。

2 社交网络动态信息管理演化博弈模型

2.1 基于信誉的动态信息管理机制

借鉴商业社会中的信誉机制^[21-23],本文在社交网络信息发布平台引入动态信息信誉管理方法。人们在发布或转发信息时须以信誉担保,为信息冲突引入利益冲突,促使人们增强决策权衡,激活网络自我净化的能力,抑制不实信息的传播。基于信誉的动态信息管理机制如图1所示。

1) 信息发布机制:用户采用信誉值做担保进行发布操作,若信息在一段时间内未引起争议,即未被举报,则发布者根据奖惩策略获得一定的信誉值奖励,反之,则获得信誉值消减的惩罚。

2) 信息转发机制:对信息持积极态度的用户(支持者)往往会转发消息促进传播,或选择潜水(如点赞没有直接传播信息),若该信息在一段时间内未引起争议,则支持者阵营内的用户获得信誉值奖励,反之,则获得信誉值消减的惩罚。

3) 信息举报机制:对信息持批评态度的用户(反对者)或者举报该消息,或者潜水(如点踩归到潜水一类)。若信息在一段时间引起大量举报,造成信息冲突,则反对者阵营内的用户获得信誉值奖励,反之,则获得信誉值消减的惩罚。

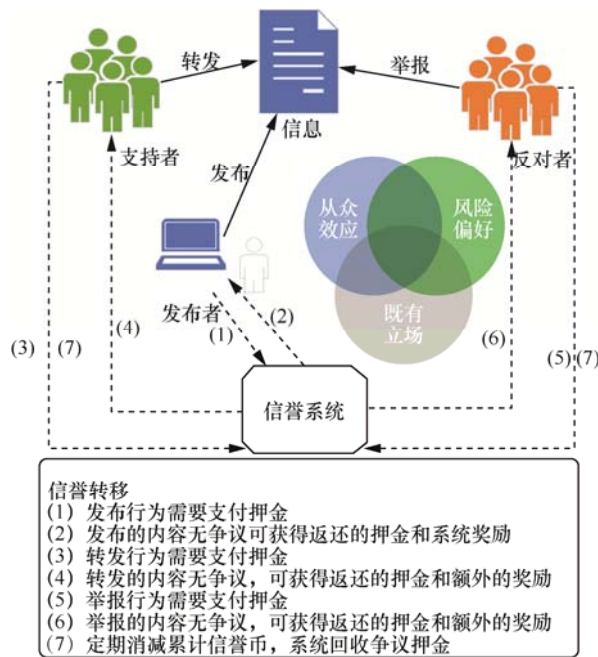


图1 基于信誉的动态信息管理机制
Figure 1 Reputation-based dynamic information management mechanism

社交网络平台上的用户在“回声室效应”^[24]的影响下自发形成“群体圈子”。在上述的动态信息管理机制的背景下,将“群体圈子”简化为支持者和反对者两个群体。这两个群体具有有限理性,立场对立,群体间用户(转发、举报、潜水)存在竞争关系。

在社交网络中,支持者可以获得系统内信誉值奖励和系统外收益,作为其基础收益。反对者举报一般只能获得系统给予的信誉值奖励作为其基础收益。大量转发和举报引发信息冲突,支持者和反对者都受到系统的信誉值惩罚作为争议损失。系统对支持者和反对者的激励和惩罚促成了两个群体之间博弈关系的形成。

在支持者和反对者博弈时,支持者之间并不是相互独立的。支持者的策略选择不仅考虑反对者的策略,也受到同时段其他支持者策略选择和信息热度的影响。同样,反对者也是如此。决策者选择什么行为与可获得的损益相关,随着两个群体的动态变化,这一心理过程可借助演化博弈模型加以解释。本文考虑在两个有限理性的大群体(支持者和反对者)中随机配对进行博弈的演化问题,建立社交网络信息冲突演化博弈模型。

2.2 社交网络中信息冲突的基本假设与分析

为了下文描述和分析方便，这里先给出支持者和反对者策略相关的一些变量、集合的定义说明。

B ：支持者的策略集（转发 B_1 和潜水 B_2 ）。

O ：反对者的策略集（举报 O_1 和潜水 O_2 ）。

S_{ij} ：支持者和反对者的混合策略，下标的 i ， j 分别表示支持者选择了策略 B_i ，反对者选择了策略 O_j 。

在真实社交网络中，某个消息的支持者群体和反对者群体是动态的，与消息内容以及接收消息人群的反应有关，从消息发出到淡出所有人的视野，才能最终确定两个群体的范围。在建模的情况下，这些群体相关的特征都转化为系统模拟的多个参数，来仿真消息的各类情况，模型中的相关变量说明如表 1 所示。

表 1 模型中的相关变量说明
Table 1 Description of variables of interest in model

变量	含义	说明
支持者	x	支持者选择转发信息的概率 $0 \leq x \leq 1$
	T_1	支持者的信誉值 $T_1 > 0$
	K_1	支持者传播信息的基础收益 $K_1 > 0$
	C_1	支持者争议基金 $C_1 > 0$
反对者	y	反对者选择举报信息的概率 $0 \leq y \leq 1$
	T_2	反对者的信誉值 $T_2 > 0$
	K_2	反对者反馈信息的基础收益 $K_2 > 0$
	C_2	反对者争议基金 $C_2 > 0$

基于上述假设和分析，每个博弈方都有 2 个纯策略选项，共可形成 4 种混合策略的组合。

1) 混合策略 S_{11} ，即支持者选择转发和反对者选择举报。此时发生信息争议冲突，支持者在获得基础收益 K_1 的同时还需支付争议损失 L_1 ，反对者在获得基础收益 K_2 的同时需支付争议损失 L_2 ，计算公式如下。

$$w = w_1 = w_2 = (|x\theta_1 - y\theta_2| - 1)(x\theta_1 + y\theta_2) \quad (1)$$

$$\begin{cases} \theta_1 = T_1 / (T_1 + T_2) \\ \theta_2 = T_2 / (T_1 + T_2) \end{cases} \quad (2)$$

$$\begin{cases} L_1 = w_1 C_1 \\ L_2 = w_2 C_2 \end{cases} \quad (3)$$

其中，信息争议系数 w 指的是信息冲突的程度，考虑博弈双方的参与程度与分歧程度，采用文献[23]的加权聚合计算的方法。为了便于后续描述， w_1 和 w_2 分别用于支持者和反对者相关的计算。

2) 混合策略 S_{12} ，即支持者选择转发，反对者选择潜水。此时未发生信息争议冲突，支持者可获得基础收益 K_1 ，而反对者由于未积极反馈，无基础收益和争议损失。

3) 混合策略 S_{21} ，即反对者选择举报，而支持者选择潜水。此时未发生信息争议冲突，反对者可获得基础收益 K_2 ，而支持者无基础收益和争议损失。

4) 混合策略 S_{22} ，即支持者和反对者都选择潜水。此时未发生信息争议冲突，两方都无基础收益和争议损失。

基于上述分析，可得到社交网络信息冲突博弈收益矩阵如图 2 所示。

	反对者	举报 (y)	潜水
支持者	转发 (x)	$K_1 + w_1 C_1, K_2 + w_2 C_2$	$K_1, 0$
	潜水	$0, K_2$	$0, 0$

图 2 社交网络信息冲突博弈收益矩阵

Figure 2 Payoff matrix of social network information conflict game

2.3 演化博弈模型

用户基于个体效益最大化（即对系统当前状态进行理性判断）做出决策。系统状态随着时间不断演化，收益函数会不断变化，从而导致支持者和反对者的行为策略不断动态调整。支持者选择转发信息的概率 x 的动态变化速度 $\frac{\partial x}{\partial t}$ 和反对者选择举报信息的概率 y 的动态变化速度 $\frac{\partial y}{\partial t}$ 可以如下的复制动态方程^[25]来表示。

$$\frac{\partial x}{\partial t} = x(U_r - U) = x(1-x)(U_r - U_d) \quad (4)$$

$$\frac{\partial y}{\partial t} = y(N_r - N) = y(1-y)(N_r - N_d) \quad (5)$$

其中 U_r 为采取转发策略（ B_1 ）的支持者的期望收益， U_d 为采取潜水策略（ B_2 ）的支持者的期望收益， $U = xU_r + (1-x)U_d$ 为所有支持者的平均期望收益。 N_r 为采取举报策略（ O_1 ）的反对者的

期望收益, N_d 为采取潜水策略 (O_2) 的反对者的期望收益, $N = yN_r + (1-y)N_d$ 为所有反对者的平均期望收益。根据 2.2 节的收益矩阵, 可知

$$\begin{cases} U_r = y(K_1 + w_1 C_1) + (1-y)(K_1) \\ U_d = y(0) + (1-y)(0) \end{cases} \quad (6)$$

$$\begin{cases} N_r = x(K_2 + w_2 C_2) + (1-x)(K_2) \\ N_d = x(0) + (1-x)(0) \end{cases} \quad (7)$$

在本博弈模型中, 参与者的纯策略选项的结果是可能性的而非确定性的, 即支持者和反对者所做的决策都是风险决策。考虑到社交网络用户的有限理性特征以及复杂心理机制的影响, 本文重点考虑风险偏好和从众效应这两种心理机制。

在社交网络谣言扩散实验中发现个体行为存在参考依赖特性, 面临“收益”时是风险追求的, 面临“损失”时是风险规避的, 这说明社交网络用户的风险态度与累积前景理论的描述一致, 因此本文引入累积前景理论重构信息博弈模型, 对决策概率与收益函数进行修正, 以求更客观真实地描述支持者和反对者的决策行为。

累积前景理论框架由 3 部分 (参考点、价值函数和决策权重函数) 组成, 从“收益”和“亏损”的角度描述社交网络用户个体的风险态度^[19]。

假设 S 是有限的状态集合, 其子集 A 表示事件, X 则是结果集合, 不确定前景 f 是 S 到 X 的一个映射 $f: S \rightarrow X$, 为每一个事件分配一个结果, 前景序列用 $f = (x_i, A_i), i=1, 2, \dots, n$ 表示, 若事件 A_i 发生, 则产生结果 x_i , 对所有序列 f 按结果递增排序, 可得到 $x_1 \leq x_2 \leq \dots \leq x_j \leq \dots \leq x_n$, 选定 x_j 作为参考点, 可得到价值函数如式(8)所示, 其中 α 和 β 为风险态度系数, 衡量结果远离参考点 x_j 的敏感性递减程度, α 和 β 越大表示用户对风险越敏感, λ 为损失厌恶系数, 反映个体对于损失更加敏感的事实, 大量的实验结果表明, 取 $\alpha = \beta = 0.88$, $\lambda = 2.25$ 较为合适, 累积前景理论相关变量说明如表 2 所示。

$$v(x_i) = \begin{cases} (x_i - x_j)^\alpha, & x_i \geq x_j \\ -\lambda(x_j - x_i)^\beta, & x_i < x_j \end{cases} \quad (8)$$

如果 $x_i \geq x_j$, 则表明结果为收益性, 决策者此时的得失为正前景价值 $V(f^+)$, 否则为负前景

价值 $V(f^-)$, 其计算式如下。

$$\begin{cases} V(f^+) = \sum_{i=1}^n \pi_i^+ v(x_i) \\ V(f^-) = \sum_{i=1}^j \pi_i^- v(x_i) \end{cases} \quad (9)$$

其中, π_i^+ 、 π_i^- 是决策权重函数, 分别对应收益性、损失性结果, 其计算式如下。

表 2 累积前景理论相关变量说明
Table 2 Description of variables in cumulative prospect theory

变量	含义	说明
α	收益状态下风险敏感度	$0 < \alpha < 1$
β	亏损状态下风险敏感度	$0 < \beta < 1$
λ	损失规避程度	$\lambda > 1$
γ	收益状态下决策函数的弯曲程度	$0 \leq \gamma \leq 1$
δ	亏损状态下决策函数的弯曲程度	$0 \leq \delta \leq 1$

$$\pi_i^+ = w^+(P(A_i \cup \dots \cup A_n)) - w^+(P(A_{i+1} \cup \dots \cup A_n)), j \leq i \leq n \quad (10)$$

$$\pi_i^- = w^-(P(A_i \cup \dots \cup A_j)) - w^-(P(A_i \cup \dots \cup A_{i+1})), 1 \leq i \leq j \quad (11)$$

其中, $P(A_i)$ 是事件 A_i 发生的理想概率, 在本模型中, 事件 A_i 指的是支持者做出转发或潜水决策, 反对者做出举报或者潜水决策的事件。 w^+ 、 w^- 对应收益性、损失性结果的容量函数, 如式(12)所示。

$$\begin{cases} w^+(p) = p^\gamma / (p^\gamma + (1-p)^\gamma)^{1/\gamma} \\ w^-(p) = p^\delta / (p^\delta + (1-p)^\delta)^{1/\delta} \end{cases} \quad (12)$$

其中, γ 表示收益状态下决策函数的弯曲程度, δ 表示亏损状态下决策函数的弯曲程度, 对应的值越小, 则权重函数弯曲程度越大, 一般取 $\gamma = 0.61$, $\delta = 0.69$ 。

根据以上描述, 本文设定支持者的价值函数参考点为 $x_j = K_1$, 则可得到累积前景理论修正后的支持者的期望收益 U_r 和 U_d 如下。

$$\begin{cases} U_r = \pi_{11}^-(y)v(K_1 + w_1 C_1) + \pi_{12}^+(1-y)v(K_1) \\ U_d = \pi_{21}^-(y)v(0) + \pi_{22}^-(1-y)v(0) \end{cases} \quad (13)$$

同理, 设反对者的价值函数参考点为 $x_j = K_2$, 可得修正后的反对者的期望收益 N_r 和 N_d 如下。

$$\begin{cases} N_r = \pi_{11}^-(x)v(K_2 + w_2 C_2) + \pi_{21}^+(1-x)v(K_2) \\ N_d = \pi_{12}^-(x)v(0) + \pi_{22}^-(1-x)v(0) \end{cases} \quad (14)$$

以上风险偏好的分析是假设决策者完全不知道其他人的行为，其决策只与对方的状态有关。但在现实社交网络中，不确定信息环境下从众效应伴随着用户行为决策过程。决策者通过观察他人行为来调整自身策略，即决策者的期望收益和上面的计算发生偏差，需引入从众效应量化方程^[20]对支持者和反对者的期望收益进行修正。在此用 ε 表示决策者的从众效应，用 μ 表示其重要程度，原始期望收益为 E ，修正后的收益为 E' 。

$$E' = (1 - \mu)E + \mu\varepsilon \quad (15)$$

一般反对者举报行为不可见，但信息的转发热度是可见的。设支持者中转发人数比例为 P_{B1} ，潜水比例为 P_{B2} ， $P_{B1} + P_{B2} = 1$ ，随着 P_{B1} 的增加，选择转发的决策者的期望收益会增加，因此支持者的

$$F_1(x, y) = x(1-x)(-\lambda)(1-2\mu_1+2\mu_1P_{B1}) \left\{ \frac{y^\delta}{[y^\delta + (1-y)^\delta]^{\frac{1}{\delta}}} \left[((-w_1)C_1)^\beta - K_1^\beta \right] - \frac{[y^\delta + (1-y)^\delta]^{\frac{1}{\delta}} - y^\delta}{[y^\delta + (1-y)^\delta]^{\frac{1}{\delta}}} K_1^\beta \right\} \quad (17)$$

$$F_2(x, y) = y(1-y)(-\lambda) \left\{ \frac{x^\delta}{[x^\delta + (1-x)^\delta]^{\frac{1}{\delta}}} \left[((-w_2)C_2)^\beta - K_2^\beta \right] - \frac{[x^\delta + (1-x)^\delta]^{\frac{1}{\delta}} - x^\delta}{[x^\delta + (1-x)^\delta]^{\frac{1}{\delta}}} K_2^\beta \right\} \quad (18)$$

3 演化博弈模型分析

3.1 均衡点求解

根据演化博弈理论，对上述信息博弈演化模型进行稳定性分析，当支持者选择转发信息概率 x 和反对者选择举报信息概率 y 不再发生变化时，双方改变策略的收益无法超过原策略收益，系统达到均衡状态。令复制动态系统中的每个方程都等于0，如式(19)所示。

$$E = \begin{pmatrix} F_1(x, y) \\ F_2(x, y) \end{pmatrix} = f(E, t) = 0 \quad (19)$$

上述二阶非线性微分方程组的解为 $(0, 0)$ ， $(0, 1)$ ， $(1, 0)$ ， $(1, 1)$ 和 (x^*, y^*) ， $x^* \in [0, 1]$ ， $y^* \in [0, 1]$ ，即系统有5个均衡点，分别为沉默点 $F_1 = (0, 0)$ 、举报点 $F_2 = (0, 1)$ 、转发点 $F_3 = (1, 0)$ 、冲突点 $F_4 = (1, 1)$ 和中心点 $F_5 = (x^*, y^*)$ 。

但是并非所有均衡点都是演化稳定点，因为

从众效应 ε_i 定义为转发和潜水的期望收益差 $(U_r - U_d)$ 和 P_{Bi} 的增函数，重要程度为 μ_i ，修正后的支持者期望收益如式(16)所示，从众效应相关变量说明如表3所示。

$$\begin{cases} U_r' = (1 - \mu_1)U_r + \mu_1P_{B1}(U_r - U_d) \\ U_d' = (1 - \mu_1)U_d + \mu_1P_{B2}(U_r - U_d) \end{cases} \quad (16)$$

表3 从众效应相关变量说明
Table 3 Description of variables in conformity effect

变量	含义	说明
μ_i	从众效应权重	$0 \leq \mu_i \leq 1$
P_{Bi}	信息转发热度	$0 \leq P_{Bi} \leq 1$

$$\text{令 } F_1(x, y) = \frac{\partial x}{\partial t}, F_2(x, y) = \frac{\partial y}{\partial t}, \text{ 将式(6)~式(16)}$$

代入式(4)、式(5)中，得式(17)、式(18)，即最终的混合策略信息博弈演化模型。

演化稳定点还必须具备抵抗有限理性造成的误差或偏差的能力，即扰动后能恢复到稳定的状态。Friedman^[26]指出，局部渐近稳定性方法可用于雅可比矩阵以确定均衡点是否稳定。判别规则如下：假设雅可比矩阵 J 的行列式为 $\det(J)$ ，迹为 $\text{tr}(J)$ ，当 $\det(J) > 0$ ， $\text{tr}(J) < 0$ 时，均衡点是稳定的；当 $\det(J) > 0$ ， $\text{tr}(J) > 0$ 时，均衡点是不稳定的；当 $\det(J) < 0$ ， $\text{tr}(J)$ 为任意值时，均衡点为鞍点。

混合策略信息博弈演化模型的雅克比矩阵 J 及其行列式 $\det(J)$ 和迹 $\text{tr}(J)$ 的计算如式(20)~式(25)所示。

$$\det(J) = \begin{bmatrix} (1-2x)J_{11} + x(1-x)J_{11}' \\ (1-2y)J_{22} + y(1-y)J_{22}' \end{bmatrix} - \quad (20)$$

$$\begin{aligned} & xy(1-x)(1-y)J_{12}J_{21} \\ \text{tr}(J) &= (1-2x)J_{11} + x(1-x)J_{11}' + \\ & (1-2y)J_{22} + y(1-y)J_{22}' \end{aligned} \quad (21)$$

$$\mathbf{J} = \begin{pmatrix} \frac{\partial F_1(x, y)}{\partial x} & \frac{\partial F_1(x, y)}{\partial y} \\ \frac{\partial F_2(x, y)}{\partial x} & \frac{\partial F_2(x, y)}{\partial y} \end{pmatrix} = \begin{pmatrix} (1-2x)J_{11} + x(1-x)\frac{\partial J_{11}}{\partial x} & x(1-x)J_{12} \\ y(1-y)J_{21} & (1-2y)J_{22} + y(1-y)\frac{\partial J_{22}}{\partial y} \end{pmatrix} \quad (22)$$

$$J_{11} = (-\lambda)(1-2\mu_1 + 2\mu_1 P_{B1}) \left\{ \frac{y^\delta}{[y^\delta + (1-y)^\delta]^{\frac{1}{\delta}}} [((-w_1)C_1)^\beta - K_1^\beta] - \frac{[y^\delta + (1-y)^\delta]^{\frac{1}{\delta}} - y^\delta}{[y^\delta + (1-y)^\delta]^{\frac{1}{\delta}}} K_1^\beta \right\} \quad (23)$$

$$J_{22} = (-\lambda) \left\{ \frac{x^\delta}{[x^\delta + (1-x)^\delta]^{\frac{1}{\delta}}} [((-w_2)C_2)^\beta - K_2^\beta] - \frac{[x^\delta + (1-x)^\delta]^{\frac{1}{\delta}} - x^\delta}{[x^\delta + (1-x)^\delta]^{\frac{1}{\delta}}} K_2^\beta \right\} \quad (24)$$

$$\begin{aligned} J_{11}' &= \frac{\partial J_{11}}{\partial x}, \\ J_{12} &= \frac{\partial J_{11}}{\partial y}, \\ J_{21} &= \frac{\partial J_{22}}{\partial x}, \\ J_{22}' &= \frac{\partial J_{22}}{\partial y} \end{aligned} \quad (25)$$

将 5 组均衡点代入, 不同均衡点对应雅克比矩阵的行列式和迹结果如表 4 所示。

表 4 不同均衡点对应雅克比矩阵的行列式和迹
Table 4 Rank and trace of jacobi matrix corresponding to different equilibrium points

均衡解	$\det(\mathbf{J})$	$\text{tr}(\mathbf{J})$
F_1	$\lambda^2 E_1 (K_1 + C_1)^\beta (K_2 + C_2)^\beta$	$\lambda [E_1 (K_1 + C_1)^\beta + (K_2 + C_2)^\beta]$
F_2	$\lambda^2 E_1 D_1 H_2$	$-\lambda (E_1 D_1 + H_2)$
F_3	$\lambda^2 E_1 D_2 H_1$	$-\lambda (E_1 H_1 + D_2)$
F_4	$\lambda^2 E_1 I_1 I_2$	$\lambda (E_1 I_1 + I_2)$
F_5	$x^* y^* (1-x^*)(1-y^*) \cdot (J_{11}' _{x=x^*} J_{22}' _{y=y^*} - J_{12} _{y=y^*} J_{21} _{x=x^*})$	$x^* (1-x^*) J_{22}' _{y=y^*} + y^* (1-y^*) J_{11}' _{x=x^*}$

$$B = T_2 / (T_1 + T_2) \quad (26)$$

$$E_1 = 1 - 2\mu_1 + 2\mu_1 P_{B1} \quad (27)$$

$$D_1 = (-B(B-1)C_1)^\beta - K_1^\beta \quad (28)$$

$$H_1 = K_1^\beta, H_2 = K_2^\beta \quad (29)$$

$$I_1 = (-|1-2B|-1)C_1)^\beta - K_1^\beta \quad (30)$$

$$D_2 = (-B(B-1)C_2)^\beta - K_2^\beta \quad (31)$$

$$I_2 = (-|1-2B|-1)C_2)^\beta - K_2^\beta \quad (32)$$

3.2 均衡点的稳定性分析

由表 4 可知, $\det(\mathbf{J})$ 和 $\text{tr}(\mathbf{J})$ 的正负主要由系统收益矩阵的变量取值决定, 其中 $H_1 > 0$, $H_2 > 0$, 因此均衡点的稳定性主要由 E_1 、 D_1 、 D_2 、 I_1 、 I_2 确定。当 $E_1 > 0$, 即 $\mu_1 < 0.5$ 时, 表示支持者从众弱相关, 当 $\mu_1 > 0.5$, $P_{B1} > 1 - 1/2\mu_1$ 时, 表示支持者从众强相关, 且信息转发存在热度的情况。当 $E_1 < 0$, 即 $\mu_1 > 0.5$ 时, $P_{B1} < 1 - 1/2\mu_1$, 表示支持者从众强相关, 且信息转发热度较低。 D_1 和 I_1 分别反映在反对者选择举报信息概率高涨时, 支持者选择转发信息的最小冒险成本和最大冒险成本, D_2 和 I_2 分别反映在支持者选择转发信息概率高涨时, 反对者选择举报信息的最小冒险成本和最大冒险成本, 这 5 组判别条件可组合成 18 种不同的信息流通情景, 不同信息流通情况下均衡点的稳定性分析结果如表 5 所示。

根据表 5 可知, 在基于信誉的动态信息管理机制的背景下, 支持者和反对者按照对信息的认知权衡自身收益, 做出决策, 在不同约束条件和信息情景下, 系统的演化结果不同。观察偶数序号的情景可知, 在支持者强从众但信息较冷门的情况下, 支持者对于利益的敏感度下降, 倾向于模仿其他支持者决策, 反对者容易占据主动, 最终形成支持者避让反对者或者两者冲突的结果。

观察奇数序号情景可知, 在支持者从众弱相关、支持者从众强相关且信息具有一定热度的情况下, 支持者容易进行风险追逐, 对于利益的敏感度上升。以约束条件五情况下为例, 系统最终收敛到 2 个中心点 $F_5 = (1, y^*)$ 和 $F_5 = (x^*, 1)$, 博弈演化过程如图 3 所示。

表 5 不同信息流通情景下均衡点的稳定性分析结果
Table 5 Stability analysis results of equilibrium points under different information cases

约束条件编号	约束条件内容	从众	情景	沉默点	举报点	转发点	冲突点	中心点
一	$D_1 > 0, I_1 > 0$	$E_1 > 0$	1		√	√		
	$D_2 > 0, I_2 > 0$	$E_1 < 0$	2					√
二	$D_1 > 0, I_1 > 0$	$E_1 > 0$	3		√			
	$D_2 < 0, I_2 > 0$	$E_1 < 0$	4					√
三	$D_1 > 0, I_1 > 0$	$E_1 > 0$	5		√			
	$D_2 < 0, I_2 < 0$	$E_1 < 0$	6				√	
四	$D_1 < 0, I_1 > 0$	$E_1 > 0$	7			√		
	$D_2 > 0, I_2 > 0$	$E_1 < 0$	8		√			
五	$D_1 < 0, I_1 > 0$	$E_1 > 0$	9					√
	$D_2 < 0, I_2 > 0$	$E_1 < 0$	10		√			
六	$D_1 < 0, I_1 > 0$	$E_1 > 0$	11					√
	$D_2 < 0, I_2 < 0$	$E_1 < 0$	12		√		√	
七	$D_1 < 0, I_1 < 0$	$E_1 > 0$	13			√		
	$D_2 > 0, I_2 > 0$	$E_1 < 0$	14		√			
八	$D_1 < 0, I_1 < 0$	$E_1 > 0$	15					√
	$D_2 < 0, I_2 > 0$	$E_1 < 0$	16		√			
九	$D_1 < 0, I_1 < 0$	$E_1 > 0$	17				√	
	$D_2 < 0, I_2 < 0$	$E_1 < 0$	18		√			

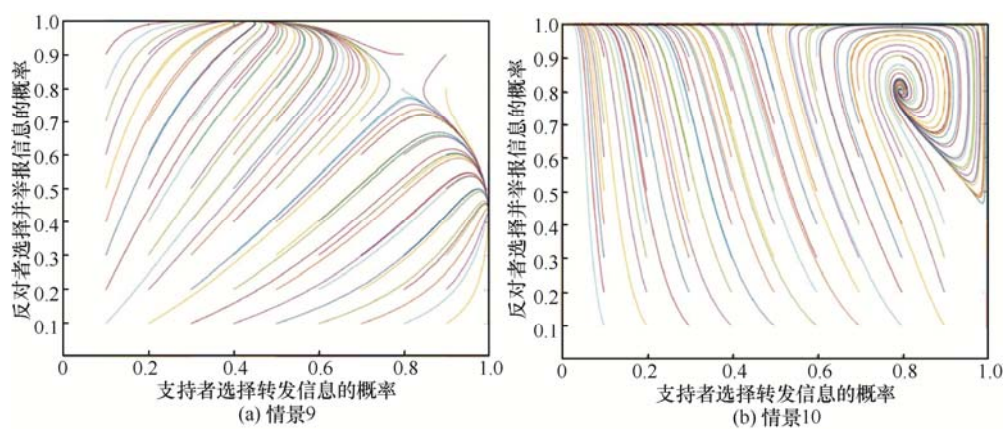


图 3 约束条件五情况下的博弈演化过程
Figure 3 Game evolution process under constraint five

当初始策略条件下支持者强于反对者时，说明信息共鸣者较多，此时收敛到 $F_5 = (1, y^*)$ ，可通过冲突惩罚降低反对者基于利益出发的举报概率，一定程度上避免恶意举报。

而当初始策略条件下支持者弱于反对者时，说明信息大概率存在争议，此时收敛到 $F_5 = (x^*, 1)$ ，支持者在冲突惩罚的措施下会降低转发概率，减小争议信息传播范围，避免冲突扩大。

因此约束条件五可为实际谣言抑制措施中的参数设计提供参考。

4 实验结果分析

为了验证本文所提混合策略的信息博弈演化模型，本文基于 Vensim 软件，将模型中参数实例化，建立系统动力学模型，主要由 4 个流位、2 个流率、8 个中间变量和 11 个外部变量构成。4 个

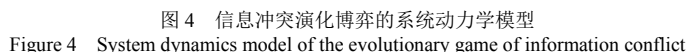
议基金、信誉值、从众效应。下面将固定初始策略 (x, y) 为 $(0.5, 0.5)$ ，进一步分析上述因素对信息冲突抑制策略实施的影响，以下分析均在满足约束条件五的情况下进行。

4.2.1 基础收益对演化结果的影响

对于社交网络用户来说,能否从信息中获益至关重要,对于支持者和反对者来说,基础收益 K_1 和基础收益 K_2 是他们在做出非潜水决策时的预估获利。根据式(33)和式(34)可知,支持者选择转发信息的概率 x 关于 t 的变化率与 $1-2\mu_1+2\mu_1P_{B1}$ 有关。当 $1-2\mu_1+2\mu_1P_{B1} > 0$, 即 $\mu_1 < 0.5$ 时,表示支持者从众弱相关, $\mu_1 > 0.5$, $P_{B1} > 1-1/2\mu_1$ 时,表示支持者从众强相关,且信息转发存在热度的情况下, x 关于 t 的变化率随着 K_1 增大而增大。当 $1-2\mu_1+2\mu_1P_{B1} < 0$, 即 $\mu_1 > 0.5$ 时, $P_{B1} < 1-1/2\mu_1$, 表示支持者从众强相关,但是在信息转发缺乏热度的情况下, x 关于 t 的变化率随着 K_1 的增大而减小,而反对者选择举报信息的概率 y 关于 t 的变化率随着 K_2 增大而增大。

根据文献[27], 设定模型中亏损状态下决策函数的弯曲程度 δ 为 0.689 1、风险态度系数 β 为 0.879 8、损失厌恶系数参数 λ 为 2.251 4。

从动态演化图(图3)中可以发现,信息的支持者和反对者进行信息反馈(转发或者举报)的选择主要取决于对方的选择和预期收益以及对风险的偏好程度。根据方程式(17)和式(18)可以发现,影响参与方最终策略的因素有基础收益、争



$$\frac{\partial F_1(x, y)}{\partial (K_1^\beta)} = \lambda x(1-x)(1-2\mu_1 + 2\mu_1 P_{B1}) \quad (33)$$

$$\frac{\partial F_2(x, y)}{\partial (K_2^\beta)} = \lambda y(1-y) \geq 0 \quad (34)$$

1) 设参数为 $\mu_1 = 0.3$, $P_{B1} = 0.6$, $C_1 = C_2 = 4$, $T_1 = T_2 = 5$, $K_2 = 2$, 控制 K_1 分别为 1.5、2.5、3.5, 支持者基础收益变化的仿真结果如图 5(a) 所示。

2) 设参数为 $\mu_1 = 0.3$, $P_{B1} = 0.6$, $C_1 = C_2 = 4$, $T_1 = T_2 = 5$, $K_1 = 2$, 控制 K_2 分别为 1.5、2.5、3.5, 反对者基础收益变化的仿真结果如图 5(b) 所示。

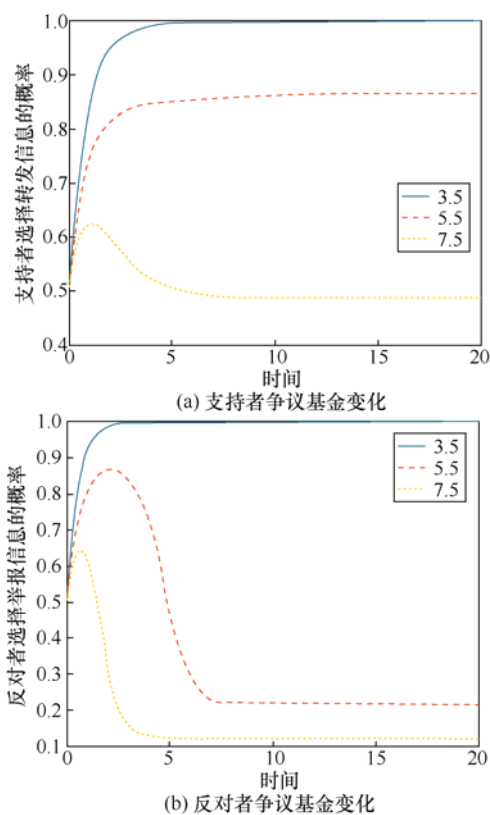


图5 基础收益对演化结果的影响
Figure 5 Influence of basic income on evolution results

由图 5 可知：在其他参数固定的前提下，基础收益减小时，博弈方会选择潜水策略的概率增大；基础收益增大时，博弈方选择信息反馈策略（转发或举报）概率上升，且收敛速度加快。

结果显示，本文模型可以仿真实际社交网络中特定信息在传播上更具优势的现象^[28]。因此，如果想要更好地对抗谣言等信息，需要增大反对者的收益激励。

4.2.2 争议基金对演化结果的影响

对于支持者和反对者来说，争议基金 C_1 和争议基金 C_2 是在做出非潜水决策时的最大预估冒险成本。根据式(35)和式(36)可知，支持者选择转发信息的概率 x 关于 t 的变化率与 $(1-2\mu_1 + 2\mu_1 P_{B1})$ 有关。当 $(1-2\mu_1 + 2\mu_1 P_{B1}) > 0$ ，即 $\mu_1 < 0.5$ 时，表示支持者从众弱相关； $\mu_1 > 0.5$ ， $P_{B1} > 1-1/2\mu_1$ 时，表示支持者从众强相关，且信息转发存在热度的情况下。 x 关于 t 的变化率随着 C_1 增大而减小；当 $(1-2\mu_1 + 2\mu_1 P_{B1}) < 0$ ，即 $\mu_1 > 0.5$ 时， $P_{B1} < 1-1/2\mu_1$ ，支持者从众强相关，但是在信息转发缺乏热度的情况下， x 关于 t 的变化率随着 C_1 增大而增大，而反对者选择举报信息的概率 y 关于 t 的变化率随着 C_2 增大而减小。

$$\frac{\partial F_1(x, y)}{\partial (C_1^\beta)} = x(1-x)(-\lambda)(1-2\mu_1 + 2\mu_1 P_{B1}) \cdot \frac{y^\delta}{[y^\delta + (1-y)^\delta]^{\frac{1}{\delta}}} (-w_1)^\beta \quad (35)$$

$$\frac{\partial F_2(x, y)}{\partial (C_2^\beta)} = y(1-y)(-\lambda) \cdot \frac{x^\delta}{[x^\delta + (1-x)^\delta]^{\frac{1}{\delta}}} (-w_2)^\beta \leq 0 \quad (36)$$

1) 设置参数 $\mu_1 = 0.3$, $P_{B1} = 0.6$, $K_1 = K_2 = 2$, $T_1 = T_2 = 5$, $C_2 = 4$, 控制 C_1 分别为 3.5、5.5、7.5, 支持者争议基金变化的仿真结果如图 6(a) 所示。

2) 设置参数为 $\mu_1 = 0.3$, $P_{B1} = 0.6$, $K_1 = K_2 = 2$, $T_1 = T_2 = 5$, $C_1 = 4$, 控制 C_2 分别为 3.5、5.5、7.5, 反对者争议基金变化的仿真结果如图 6(b) 所示。

由图 6 可知：其他参数固定的前提下，争议基金较小时，博弈方会选择信息反馈策略（转发或者举报）。在争议基金增大的情况中，当冲突较小时，冲突带来的损失低于收益，博弈方选择信息反馈策略的概率上升；当冲突变大时，损失高于收益，博弈方选择信息反馈策略的概率有所下降。争议基金继续增大，使得小冲突时的损失高于收益，则博弈方快速选择潜水策略，即随着争议基金的增大，博弈方选择信息反馈策略（转发或者举报）的概率下降，且收敛速度减缓。

结果显示，本文模型中争议基金的设置能够

对高冲突信息的传播进行抑制。即在对负面舆情的处理中应增大支持者的争议基金，同时减小反对者的争议基金。

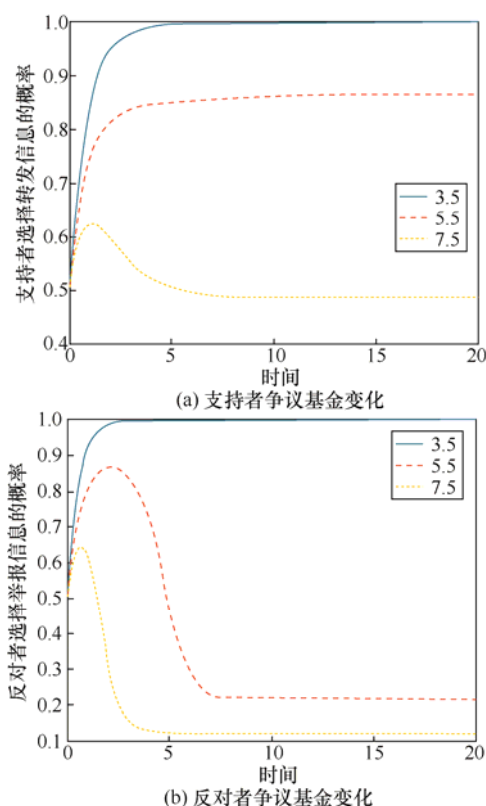


图6 争议基金对演化结果的影响
Figure 6 The impact of dispute funds on evolutionary results

4.2.3 从众效应对演化结果的影响

从众效应的高低将导致支持者和反对者选择转发、举报概率变化趋势的改变。为了探究从众效应的重要程度 μ_1 和转发热度 P_{B1} 这两个参数对博弈双方决策概率变化的影响情况，变更 μ_1 取值为0.3和0.7，分别表示支持者从众弱相关和强相关，变更 P_{B1} 取值为0.1、0.5、0.9。

1) 控制从众重要程度为弱相关， $\mu_1 = 0.3$ ，设置参数 $K_1 = K_2 = 2$ ， $T_1 = T_2 = 5$ ， $C_1 = C_2 = 4$ ，控制 P_{B1} 分别为0.1、0.5、0.9，仿真结果如图7(a)所示。

2) 控制从众重要程度为强相关， $\mu_1 = 0.7$ ，设置参数 $K_1 = K_2 = 2$ ， $T_1 = T_2 = 5$ ， $C_1 = C_2 = 4$ ，控制 P_{B1} 分别为0.1、0.5、0.9，仿真结果如图7(b)所示。

由图7可知，在从众效应弱相关的环境下，

信息转发热度增大对支持者转发概率将起到加快收敛的作用，但变化幅度较小；而在从众效应强相关的环境下，其影响结果与弱相关一致，但变化幅度大幅增加。在从众效应影响下，双方当前状态及预估的收益或者损失不再是博弈方决策的主要依据，他们会根据同阵营内其他人（好友圈）的决策行为改变自己的行为。一旦在阵营内出现快速大量的转发行为，即信息转发热度达到心理预期，支持者就会放大自己的冒险安全性，从而做出相同的行为。并且从众效应重要程度的增大会加大对决策者的影响，缩短思考的时间，加快做出转发决策。

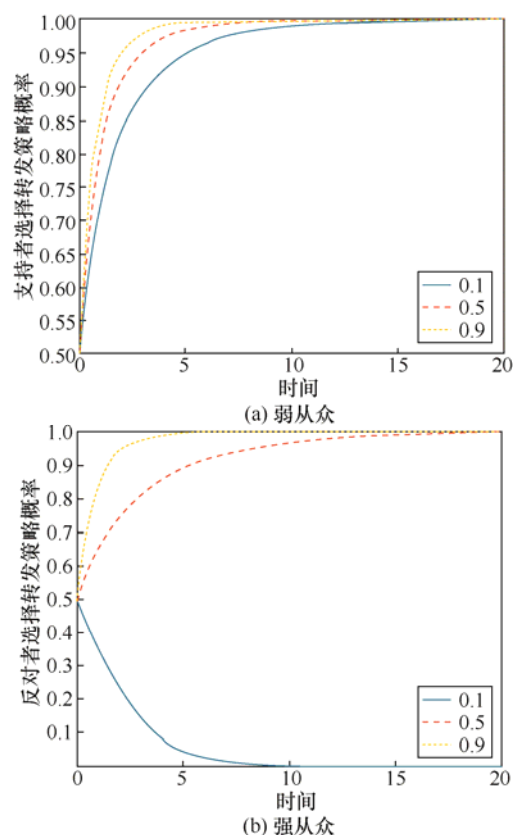


图7 从众效应对演化结果的影响
Figure 7 The influence of conformity effect on evolutionary results

结果显示，本文模型可以仿真个体对于信息的判断存在群体适应性预期的现象^[29]，如用户的从众效应和追随热门信息的特点，促成谣言等信息的“爆点”和基础数据流量。因此，在系统设计上需要在热度推高时动态加大争议基金，增强决策理性，降低群体适应性预期现象的程度。

4.3 动态信息管理机制策略建议

在基于信誉的动态信息管理机制下，冲突双方是动态变化的，信誉值为冲突双方群体的总体信誉值。为达成舆情治理的目标，可在信息争议较小时，采取宽松的冲突抑制策略，鼓励自由发言，而在信息争议较大时，采取严格的冲突抑制策略，让博弈双方认识到信息冲突中存在的问题，更理性地传播信息。

由3.2节的均衡点稳定性分析和4.2节的演化影响因子分析结果可知，基础收益与博弈方实施转发和举报策略呈正相关，而争议基金与博弈方实施转发和举报策略呈负相关，对于支持者，从众效应、信息热度的上升会促进转发概率的上升。而双方的信誉值差距较大时，对转发和举报策略起到促进作用，反之则起到抑制作用。

根据以上策略，结合基于信誉的信息管理机制，可以在社交网络中实现智能化信息管理系统，其中“信誉币”是实现用户信息调控的手段。基础收益可以实现为信誉币收益，争议基金可以实现为消息押金，具体操作如下。

假设在实际系统内，信息的转发次数为 m ，举报次数为 n ，则信息引起的冲突程度可由信息意见分歧程度 F_1 和信息影响范围 F_2 这两个部分衡量。 F_m 和 F_n 分别表示支持者和反对者参与信息冲突的程度。

$$F_1 = 1 - |m - n| / (m + n) \quad (37)$$

$$F_2 = \ln(m + n) \quad (38)$$

$$\begin{cases} F_m = m / (m + n) \\ F_n = n / (m + n) \end{cases} \quad (39)$$

其中，信息意见分歧程度 F_1 取值在0到1之间， F_1 越大，表示信息的支持者和反对者之间的分歧越大。 F_2 越大，表示信息的影响范围越大，系统可以对不同主题的消息设定不同的信息热门程度阈值 T ，则支持者在信息流通过程中选择转发可获得的信息收益为 K_1 ，反对者选择举报可获得的信息收益为 K_2 ，它们相对固定，且 K_2 必须大于 K_1 ，为了避免恶意举报，在实际系统中， K_2 可设置为 K_1 的1.5到2倍。转发或者举报信息所需要

的押金 C_1 和 C_2 可以根据收益来实时推算。

$$C_1 = e^{4(F_1 - 0.5)} \left(\frac{2F_2}{F_2 + T} \right) F_m K_1 \quad (40)$$

$$C_2 = e^{4(F_1 - 0.5)} \left(\frac{2F_2}{F_2 + T} \right) F_n K_2 \quad (41)$$

$$K_1 < K_2 \quad (42)$$

以上公式能够很好地满足约束条件五，并且在实际应用中，可通过对押金数额的调整和变化进行信息的自动化引导，达成如下功能。

1) 当 $F_1 < 0.5$ 且 $F_2 < T$ 时，信息的支持者和反对者之间的分歧较小，信息的影响范围较小。系统给予信息发布和转发者的基础收益 K_1 和所需押金 C_1 满足 $C_1 < 0.75K_1$ ，给予信息举报者的基础收益 K_2 和所需押金 C_2 满足 $C_2 < 0.75K_2$ ，保障不破坏自由发言的氛围。

2) 当 $F_1 > 0.5$ 且 $F_2 > T$ 时，信息支持者和反对者之间的分歧和信息的影响范围都较大。系统给予信息发布和转发者的基础收益 K_1 和所需押金 C_1 满足 $0.75K_1 < C_1 < e^2 K_1$ ，给予信息举报者的基础收益 K_2 和所需押金 C_2 满足 $0.75K_2 < C_2 < e^2 K_2$ ，加强信息转发和举报的理性成分。

5 结束语

本文提出基于信誉的动态信息管理机制，通过建立动态奖惩策略，利用间接协作共同发现和抑制争议信息，达到激活网络自我净化能力的目标。本文的主要工作及成果如下。

1) 建立基于累积前景理论的社交网络用户信息冲突演化博弈模型，仿真结果显示：该博弈模型演化走势受到双方基础收益、争议基金、信誉值、从众效应权重和信息热门程度因素的影响。通过分析模型均衡点的稳定性，得到9个约束条件下的18个情景的演化结果，其中约束条件五可以用于引导博弈双方理性对待信息冲突，实现冲突抑制。

2) 提出信息冲突博弈的系统动力学模型，对影响因子进行敏感性分析。结果表明，双方基础收益的提高可以起到促进选择转发和举报策略概率上升的效果，双方争议基金的提高可以起到相

反的效果；信息热门程度的提升将促进支持者转发概率的上升。

3) 根据模型演化规律和信息冲突机理分析结果，给出了面向舆情治理的定量化信息管理策略建议：争议基金（信誉押金）实时计算的推荐公式，为实现该管理机制提供了理论支持。

在未来工作中，将融合相应的谣言判断算法，结合信息管理策略，实现“事前预防”和“事后补救”相结合的信息流动管理系统，改进和完善社交平台中的信息流转体系，在高效共享信息的同时有效抑制负面舆情。

参考文献：

- [1] ALLPORT G W, POSTMAN L. The psychology of rumor[R]. 1947.
- [2] 靖鲲鹏, 岑怡. 社交网络舆情下的二元信息竞争 SH2IR 模型及仿真研究[J]. 情报杂志, 2022, 41(1): 117-124.
JING K P, CENG Y. SH2IR model and simulation of binary information competition under social network public opinion[J]. Journal of Intelligence, 2022, 41(1): 117-124.
- [3] 黄炜, 余辉, 李岳峰, 等. 网络舆情事件演化的仿真实践研究[J]. 现代情报, 2017, 37(8): 65-73.
HUANG W, YU H, LI Y F, et al. Simulation practice of the network public opinion evolution[J]. Journal of Modern Information, 2017, 37(8): 65-73.
- [4] PAL S, HILL A, RABEHAJA T, et al. A blockchain-based trust management framework with verifiable interactions[R]. 2021.
- [5] CHEN Y Z, LI Q L, WANG H. Towards trusted social networks with blockchain technology[J]. Social and Information Networks, 2018, 18(1): 2796.
- [6] CHEN J, CHEN C, SONG Q, et al. Spread mechanism and control strategies of rumor propagation model considering rumor refutation and information feedback in emergency management[J]. Symmetry, 2021, 13(9): 1694.
- [7] WANG Z, CHEN A. On ISRC rumor spreading model for scale-free networks with self-purification mechanism[J]. Complexity, 2021(8): 1-9.
- [8] MO L, GUO S. Consensus of linear multi-agent systems with persistent disturbances via distributed output feedback[J]. Journal of Systems Science and Complexity, 2019, 32(3): 835-845.
- [9] ZHAO L, QIU X, WANG X, et al. Rumor spreading model considering forgetting and remembering mechanisms in inhomogeneous networks[J]. Physica A: Statistical Mechanics and its Applications, 2013, 392(4): 987-994.
- [10] ASKARIZADEH M, LADANI B T. Soft rumor control in social networks: Modeling and analysis[J]. Engineering Applications of Artificial Intelligence, 2021, 100: 104198.
- [11] SIMON S, APT K R. Choosing products in social networks[C]//8th Workshop Internet Netw Econ(WINE). 2012: 100-113.
- [12] MACHINA M J. Expected utility hypothesis[C]//Utility and Probability. 1990: 79-95.
- [13] WANG B, GE C, FU L, et al. DRIMUX: dynamic rumor influence minimization with user experience in social networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, (10): 1.
- [14] TSCHIATSCHEK S, SINGLA A, GOMEZ RODRIGUEZ M, et al. Fake news detection in social networks via crowd signals[C]//Companion Proceedings of the Web Conference 2018. 2018: 517-524.
- [15] NGUYEN N P, YAN G, THAI M T, et al. Containment of misinformation spread in online social networks[C]//Proceedings of the 4th Annual ACM Web Science Conference. 2012: 213-222.
- [16] WEN S, JIANG J, XIANG Y, et al. To shut them up or to clarify: restraining the spread of rumors in online social networks[J]. IEEE Transactions on Parallel & Distributed Systems, 2014, 25(12): 3306-3316.
- [17] SRINIVASAN S, DHINESH B. A social immunity based approach to suppress rumors in online social networks[J]. International Journal of Machine Learning and Cybernetics, 2021, 12(2): 1-16.
- [18] WU Y, HUANG H, ZHAO J, et al. Using mobile nodes to control rumors in big data based on a new rumor propagation model in vehicular social networks[J]. IEEE Access, 2018: 1.
- [19] TVERSKY A, KAHNEMAN D. Advances in prospect theory: cumulative representation of uncertainty[J]. Journal of Risk and Uncertainty, 1992, 5(4): 297-323.
- [20] SKYRMS B. Dynamics of conformist bias[J]. The Monist, 2005, 88(2): 260-269.
- [21] LAHBIB A, TOUMI K, LAOUITI A, et al. Blockchain based trust management mechanism for IoT[C]// 2019 IEEE Wireless Communications and Networking Conference (WCNC). 2019.
- [22] LI L, KURTHS J, YANG Y, et al. Prevention and trust evaluation scheme based on interpersonal relationships for large-scale peer-to-peer networks[J]. Mathematical Problems in Engineering, 2014, 2014(3).
- [23] YANG Z, YANG K, LEI L, et al. Blockchain-based decentralized trust management in vehicular networks[J]. IEEE Internet of Things Journal, 2018: 1.
- [24] CHOI D, CHUN S, OH H, et al. Rumor propagation is amplified by

- echo chambers in social media[J]. Scientific Reports, 2020, 10(1): 1-10.
- [25] GINTIS H. Game theory evolving: a problem-centered introduction to modeling strategic behavior[M]. Princeton: Princeton University Press, 2000.
- [26] FRIEDMAN D. On economic applications of evolutionary game theory[J]. Journal of Evolutionary Economics, 1998(8): 15-43.
- [27] 许论辉, 骈宇庄, 林永杰, 等. 基于累积前景理论的无信控路段人车通行冲突博弈分析[J]. 中国公路学报, 2022, 35(1): 18.
- XU L H, PIAN Y Z, LIN Y J, et al. Game analysis of conflicts between pedestrians and vehicular traffic on unsignalized road sections based on cumulative prospect theory[J]. China Journal of Highway and Transport. 2022, 35(1):18.
- [28] 吴俊明, 刘炎. 德法兼治: 网络谣言之二元治理——兼论证券市场“网络谣言”的规制[J]. 证券法苑, 2016, 18(2): 374-393.
- WU J M, LIU Y. Governing both germany and law: dual governance of internet rumors——also on the regulation of "internet rumors" in the securities market[J]. Securities Law Review, 2016, 18(2): 374-393.
- [29] 李国武. 谣言实现的社会机制及对信息的治理[J]. 社会, 2005(4): 143-155.
- LI G W. Social Mechanism of rumor realization and information governance[J]. Society, 2005(4): 143-155.

[作者简介]



刘淑婷（1999- ），女，江西赣州人，上海交通大学硕士生，主要研究方向为社交网络内容安全、区块链应用。



马颖华（1973- ），女，山东泰安人，博士，上海交通大学讲师，主要研究方向为网络内容安全、自然语言处理。



陈秀真（1977- ），女，山东聊城人，博士，上海交通大学副教授，主要研究方向为车联网信息安全、网络信息系统安全检测与评估、社交网络大数据分析。