doi :10.3969/j.issn.1671-1122.2021.12.008

基于信息行为的社交网络节点信誉评估模型研究

熊建英

（江西警察学院科技与信息安全系，南昌 330003）

摘　要：为促进社交网络用户自律与自治，提高社交网络的可信度，文章提出一种基于用户信息行为监督反馈的动态信誉评估方法。该方法在节点综合信誉评估中融合身份信誉与行为信誉机制，在动态更新中设置新节点考核期与分阶段信誉更新机制。通过节点信息披露、网络特征计算节点身份信誉；根据监督节点在社交网的信息发布、转发，对信息行为的自我纠正与不良信息阻断行为给与奖惩，形成行为信誉。实验结果表明，相比传统信任评估机制，对信息行为设置奖惩引导不仅可以提高信誉评估准确度，还可以抑制不良信息的交互。

关键词：社交网络；节点信誉；信息行为；信誉奖惩；网络治理

中图分类号：TP309　**文献标志码：**A　**文章编号：**1671-1122（2021）12-0051-09

中文引用格式：熊建英 . 基于信息行为的社交网络节点信誉评估模型研究 [J]. 信息网络安全，2021，21（12）：51–59.

英文引用格式：XIONG Jianying. Reputation Evaluation Model in Social Network Based on Information Behavior[J]. Netinfo Security, 2021, 21(12): 51–59.

Reputation Evaluation Model in Social Network Based on Information Behavior

XIONG Jianying

（*Security Technology Department, Jiangxi Police Institute, Nanchang* 330003, *China*）

Abstract: In order to promote self-discipline and autonomy of social network users and improve the credibility of social network, a dynamic reputation evaluation method based on user information behavior supervision and feedback is studied. The comprehensive reputation contains identity and behavior reputation, set evaluation period for a new node and update mechanism in different stage. Identity reputation is calculated by information disclosure and network characteristics; behavior reputation is calculated by information release and forwarding, and rewards or punishments will be given to self correction of information behavior or blocking of bad information. The simulation results show that compared with the traditional trust evaluation mechanism, setting rewards and punishments guidance can improve the accuracy of reputation evaluation. Reputation incentive can also inhibit the

收稿日期：2021-09-23

基金项目：江西省教育厅科技计划 [GJJ212201]

作者简介：熊建英（1982—），女，江西，副教授，博士，主要研究方向为信息安全、情报研判。

通信作者：熊建英 special8212@sohu.com

interaction of bad information.

Key words: social network; reputation; information behavior; reputation rewards and punishments; network governance

0 引言

社交网络是一种基于“内容型”的关系网络，随着社交网络应用的普及与规模的扩张，社交内容已经成为网民获取信息的重要渠道，社交媒体的概念也应运而生。社交媒体中每个节点都是内容的产生者、传播者和接受者，内容的传播形态也形成点对点、点对面、面对面的无序无限传播态势，这使得社交媒体治理比其他网络更为困难。如果只依靠平台治理信息，无疑是一个巨大的挑战。然而社交媒体中每个节点又可以是内容的发布者、监督者和阻断者，如果从网络节点自治角度入手，则可以实现从安全问题源头进行控制，构建良好的社交网生态。与社会网络关系类似，在开放的社交网络环境下，节点在社区中的信誉、节点间的信任都是一种重要网络人际关系。随着《网络安全法》实施，对网络实名认证、网络诚信建设都提出要求，这也使得社交网络的信誉越来越受到用户的重视。综合节点在社交网络的信息披露及其信息行为进行信誉评估，则可以对节点的行为产生约束，达到更好的行为自律。模型构建时，通过设置社交行为奖惩反馈可以对节点信誉进行动态调整，在保障自律的同时，也有利于引导节点的正向信息行为，对社交媒体中质量低下、不良不实信息传播形成他律，更好地实现群防群控的自治机制。

1 相关工作

社交网络中无障碍信息传播及存在的信息内容真实性、有效性保障等问题，导致社交媒体安全问题愈发凸显。对此，国家相关部门以及网络平台出台了相关条例、制度和规范[1]。但在复杂的社交网络环境中，用户不诚信和恶意行为是不可避免的。通过信息技术手段设计相应算法是社交网治理的重要方面，其中信誉机制是保障社交网络安全常用的解决方案[2]。信息行为可以是所有与信息源、信息获取、利用、传播等相关人类行为，也是一种社会行为。用户主体在网络社会中充当着不同的角色，信息的交互仍然是人与人之间的交互。在社交网络中发布与传播虚假信息、不良信息、不文明评论、恶意诽谤、恶语攻击等行为极易被网络群体扩散，社交网络中内容产生的便捷性要求用户对信息行为进行自我判断与纠正[3]。聂勇浩[4]等人借鉴社会学、心理学、行为学研究用户信息行为，揭示用户内在人格特质、动机、认知等，以及信任、文化等外在社会属性都对信息行为有显著影响。不同研究领域对信息行为的类型也有不同划分，如从知识管理视角（博客、知识解答、添加标签、标注用户评论等）、商务视角（商品评论、评分、推荐等）、传播学视角（信息贡献、搜索获取、转帖分享、交流评论等）[5]，社交网络谣言传播识别中，一般将用户信息分解为信息产生、传递、判断和影响几个阶段。网络节点的影响力与社交关系的强度相关，主要表现在用户关注数、粉丝数、交互频率、转发数、评论数等[6]。此外，张大勇[7]等人在社交网用户信息贡献行为研究中指出，用户自我信息披露、社会化搜索核查等行为反映了用户自身的责任意识和自觉性。张会平[8]等人研究了基于计划行为和威慑理论，提出社交平台增加惩罚机制也可以促进网民进行主动判断，降低网络谣言传播的可能性。

信任管理与信誉机制已经在计算机网络、通信、电子商务、云计算等领域被广泛研究和应用。社交网络中信任度量一般可以通过直接信任、间接信任等方式来计算，但信任更偏向的是节点之间的、局部的信赖关系。信任可以理解为一方对另一方可靠性的认可，反映的是相互之间的信赖关系。信誉则被认为是节点在整个虚拟社区全局的信任度量，类似于社会信誉，是一种累积的社交资产[9]。对于社交网信任与信誉评估已有不少研究。例如，于真[10]等人根据用户社交网消息扩散的拓扑结构，综合网络中推荐节点的信任给出用户最终可信度。赵丽华[11]等人从权威性和真实性

两方面对微博社区用户进行信任评估。张继东[12]等人提出通过对主观感知的虚拟属性进行模糊化度量，依据交互时间衰减评估移动社交网用户交互信任。张宁[13]等人在提高移动群治感知网络用户参与度方面，融入信誉模型激励用户参与，不仅提升了网络任务处理效率，同时也降低了处理成本。

基于信息行为的社交网节点信誉评估研究主要建立在社交网信息行为研究与社交网络信任与信誉研究基础之上。这些领域研究已产生较为丰富的研究成果，但从社交媒体治理角度来看，仍缺乏融合信息行为及行为奖惩反馈的信誉评估模型研究。社交网节点信誉为用户在虚拟社会的行为表现提供了一个用户画像，有利于促进用户为维护自身形象，激发其对网络信息行为的责任心与自觉性。因此，信誉不仅是用户社交行为的一个累积表现，还可以作为监督与引导用户行为的一种手段。

本文提出的信誉评估模型融合了用户信息披露、网络特征、历史信息行为数据，设置不同信息行为的信誉动态更新规则，以实现利用信誉对用户行为进行监督与约束的目的。通过模拟诚信节点与恶意节点的信息发布、转发等行为对信誉的动态更新，表明本文模型可以更准确评估节点信誉；同时通过模拟信息发布质量、自我阻断、核查判断等行为增加奖惩因子修正信誉，表明本文模型可以降低不良信息传播速度，提高不良信息核查效率。

2 信誉评估模型构建

2.1 信誉模型工作原理

在对节点信誉评估时，获取信息行为数据可以从3方面进行考虑：用户身份信誉是为了实现用户身份鉴别，保障社交网节点是真实有效的；节点行为信誉是对历史信息交互行为的综合评价；信誉奖惩是用于对节点信息行为进行反馈，以引导节点调整信息行为，也是对节点行为的正向期望。社交网节点信誉（RP）是基于社交网络中用户身份、信息行为及行为反馈的综合可信度评估，也是对社交网具有有效身份的节点、预期信息行为的一种评估。

社交网络中节点相互关注关系、信息传播关系都可以将节点形成一种网络关系。基于信息行为的信誉模型以信息源的信息披露作与社交网中位置状态为节点身份信誉；以用户之间信息交互行为作为节点行为信誉；以信息交互产生的反馈行为作为信誉奖惩调节机制，综合产生节点信誉，节点信誉框架如图1所示。

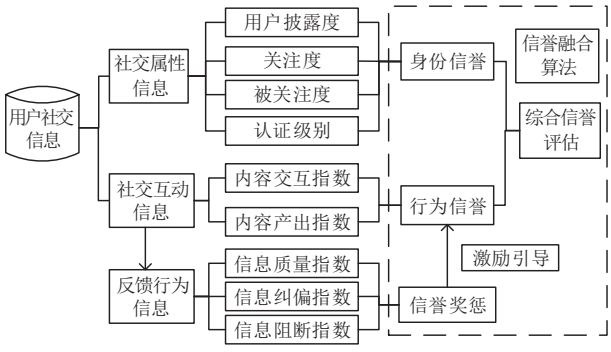

图 1 节点信誉框架

2.2 身份信誉

节点的身份信誉IR主要由用户信息的披露性、节点社交网络中位置状态以及平台对用户身份的认证所决定。

1）用户披露度UDI。社交平台需要用户填写的个人信息一般包括姓名、性别、生日、所在地、从事行业、教育程度等。用户通过披露个人信息维持发展人际关系，同时也对社交网络信任关系的建立产生影响[4]。从3个维度对披露指数进行评估，包括披露信息的完整性、合理性和权威性。完整性表示资料越完整，越能表示用户对自己虚拟身份的重视，也越能为自己的信息行为负责。权威性是指具备信息判断、识别能力基础，会影响其他社交节点对其话语产生价值认可的指标，如学历、职业、职称指标。在计算完整性与权威性的同时，需要验证用户披露信息的真实程度。社交平台不会对所有披露信息进行认证，因此，系统可进一步根据常识性规则进行合理性概率推断，如命名规则、时间范围、信息冲突等。

在身份信誉计算过程中，UI_i表示节点i信息披露的完整性；$f\chi(i)_j$表示节点i的第j项内容是否填报，0代

表未填报，1代表已填报；w_j表示该项内容的必要性权值。

P_j表示该项内容的合理性，通过经验常识，构造正则表达式或常用规则库对各项内容进行判别，默认合理性为1，如果命中不合理评判规则进行扣分。选择部分属性构建不合理取值判定规则如表1所示。

表 1 属性不合理性判断规则示例

属性	不合理评判规则
姓名	国籍为中国时，含非中文字符：0 国籍为中国时，长度小于 2：0 汉族条件下，长度大于 4：0.5 汉族条件下，第一个字符不能匹配姓氏库样本：0.5
年龄	年龄小于 6 或年龄大于 120：0 在职状态：年龄低于 18 或高于 80：0 学生状态：年龄低于 6 或高于 45：0.5
毕业学校	校名库中无学校信息：0.5
所属组织	组织库中无组织信息：0.5

UI_i计算如公式（1）所示，若属性未设置合理性验证，则P_j为1；否则，按规则设置取值。

$$UI_i = \sum_{j=1}^{n}(fy(i)_j\, w_j P_j) \tag{1}$$

UA_i表示节点i信息披露的权威性，设置权威属性列表$a=\{a_1,a_2,\cdots,a_n\}$，属性值量化为af_j。参照社交网络用户信息设置选项与取值内容，wf_j表示该项内容的权威性权值，对常见的权威影响因素进行离散化标注如表2所示。

表 2 用户权威性属性指标示例

属性	属性内涵	离散规则
行业	从业类型	专业性职业 1，学生 0.5，非专业性职业 0
组织	组织规模	高规模 1，中小规模 0.5，无组织 0
学历	知识水平	研究生以上 1，大中专 0.5，高中及以下 0
职称	从业资历	高级 1，中级 0.5，初级 0

UA_i计算如公式（2）所示，权威属性只计算权威属性列表a中选项，并根据属性合理性进行调整。其中，$j \in a$。

$$UA_i = \sum_{j=1}^{n}(af_j wf_j P_j) \tag{2}$$

2）关注度。社会行为存在聚类现象，假设节点i关注其他节点的数量为n，i的社交关注度使用所关注的n个节点的平均信誉值作为计算依据，同时为了减少i通过关注少量节点获取高平均值的误差，用节点数量作为强化系数s_1。设经验值N作为社区平均关注规模，s_1为n与N的比值，也代表i的相对关注规模。当n大于N时，直接将s_1记为1。关注度UF_i计算方法如公式（3）所示。

$$UF_i = \frac{1}{n}\sum_{j=1}^{n} RP_{i\to j} s_1 \tag{3}$$

3）社交被关注度。节点拥有的粉丝节点越多表明信息扩散作用越明显，粉丝数量是社交网络中其他节点对其信誉的一种投票认可。如果节点拥有的粉丝节点数量为m，则i的社交被关注度用m个粉丝节点的平均信誉值来计算，设经验值M作为平均粉丝规模，同理设置加强系数s_2。社交被关注度UN_i计算方法如公式（4）所示，其中$s_2=m/M$，当$m \geqslant M$时，$s_2=1$。

$$UN_i = \frac{1}{m}\sum_{j=1}^{m} RP_{j\to i} s_2 \tag{4}$$

4）认证指数。社交平台提供的节点认证机制主要有个人认证和官方认证形式。个人认证有兴趣认证、自媒体认证、身份认证；官方认证有政府、企业、媒体、机构、公益等。不同认证主体对节点信誉也有一定影响。根据认证状态，设置量化等级为$MA=\{$未认证–0，个体认证–0.5，官方认证–1$\}$，具体量化可根据平台认证难度与认证需求进行调整。认证指数UP_i的计算方法如公式（5）所示，其中$fmap$为认证指数映射函数，获取i节点所对应的认证量化映射表MA的值。

$$UP_i = fmap(i, MA) \tag{5}$$

2.3 行为信誉

节点的社交信誉SR由节点在社交网络中位置特征与社交网络中历史信息交互的行为所确定。

1）内容交互度。节点i每次发帖都是一次社交网络互动行为，其他节点所给出的转发、点赞、评论、踩等行为也都是一次信息接收后的反馈行为。对交互行为反馈设置为4个分值等级，其中转发为3、评论为2、点赞为1、踩为–1。这些交互节点的信誉值会影响评分的有效性，高信誉节点给出的互动对节点i信誉提

升更有价值。RP_j为给出第j个反馈的节点信誉值，fs_j为给出第j个反馈的量化评分，则i节点的内容交互度为加权累积所有给出反馈的节点信誉，内容交互度SC_i计算如公式（6）所示。

$$SC_i = \sum_{j=1}^{n} RP_j(fs_j / 3) \tag{6}$$

2）价值贡献度。节点发布的原创发布信息会对整个社区贡献价值，也可以反映节点具有主观能动性，帮助其信誉累积。设节点i发布原创信息为pa，获得的交互评分累计为fta，发布总信息为pb，获得总交互评分累计为ftb；定义原创信息相对热度sd为原创平均互动数量（fta/pa）与总体平均互动数量（ftb/pb）之比，热度所占总体热度越高，价值贡献度越高。设原创内容规模平均值NF，以原创数量nf/NF作为加强系数s_3，nf大于NF时直接记为1，则价值贡献度SO_i计算如公式（7）所示。

$$SO_i = s_3 \frac{fta}{pa} / \frac{ftb}{pb} \tag{7}$$

2.4 信誉奖惩

在社交网络中，高信誉节点会在危机情况下自觉维护网络稳定、控制负向信息传播，平台可根据节点是否有主动维护社交网络安全的信息行为对其信誉进行判断[14]。节点的自觉性还表现在对信息的谨慎转发、主动验证其真实性后分享，有较好的信息辨别能力[15]。信息传播中的“判断”是强调节点需要对发布或转发的信息进行有效甄别，如果用信誉奖惩对节点主动性进行引导，则可以更有效促进节点自治。在信誉奖惩机制中，设置内容质量、信息自我纠偏、不良信息传播阻断3个维度，具体奖惩规则如表3所示，奖惩系数也可以根据平台经验进行调整。

惩罚系数r的取值如表3所示，系统设置的惩罚情景多于奖励情景，这是因为一旦出现不良行为，节点的行为信誉下降会更快，可以促使节点更加谨慎自己的信息行为。信誉更新机制是通过对信誉行为进行实时惩罚监督，可动态调整信誉值。在调整过程中，高信誉节点的行为对社交网络具有高影响力，信誉奖惩力度与节点信誉成正比，所以增加奖惩系数后，节点信誉值RP_i的计算方法如公式（8）所示，其中RP_i'为i更新前的信誉值。

表 3 信誉奖惩规则

维度	信息行为	解释	奖惩规则	系数 r
内容核查	原创加精	对所发布原创内容，根据其点击量、转发量等设置为精华贴	每增加 1 条精华帖	+0.01
	发布不良信息	发布不良信息，受到平台对其核查	每核实发布 1 条	-0.03
	转发不良信息	转发不良信息，受到平台对其核查	每核实发布 1 条	-0.02
自我纠偏	删除不良信息	在未被平台核实之前删除，澄清	每核实曾发布 1 条	-0.01
传播阻断	有效举报	自觉核查信息，在平台对不良信息进行举报，	每核实有效 1 条	+0.02
	无效举报	未核查信息，无有效证据，违规举报	每核实无效 1 条	-0.01

$$RP_i = RP_i'(1+r) \tag{8}$$

2.5 综合信誉映射规则

1）信誉动态更新。节点的身份信誉相对稳定，初始化后更新频率会比较低，所以可以设置较长的更新期tp，如1周、1个月，这样可以降低系统计算开销，也可以防止节点通过拉粉丝或大量关注他人即时获取信誉提升。节点的行为信誉及其反馈奖惩是对节点社交行为监督的有效手段，可根据行为操作即时修正信誉，并以较短的时间窗口Δt，如1天，对综合信誉进行动态更新。

2）新加入节点的行为信誉初始化与更新。新加入的节点，由于未产生信息交互行为，初始信誉过低。可在节点注册基本身份信息，设新节点考核期为TC，节点在考核期内利用行为信誉的中位值SP_{med}进行行为信誉初始化。如果考核期内信誉下降速度Sd超出阈值tr，系统给出警告并让节点进入封闭期，信誉为0。如果信誉未提升，即$\Delta RP<0$，系统将剥夺赋予的行为信誉初值；如果信誉增加，则按正常规则进行信誉动态更新。设置考核期与封闭期可以进一步规范节点行为，防止通过注册节点进行恶意操作的行为。

3）信誉融合。计算综合信誉时，身份信誉与行为

信誉中的指标可以按一定的权重进行融合，身份信誉中每个指标都是归一化的值{*UA*, *UI*, *UF*, *UN*, *UP*}，身份信誉*IR*= *UA*+*UI*+*UF*+*UN*+*UP*，加和取值范围在[0,5]之间；行为信誉*SR*=*SC*+*SO*，设这两种信誉设置权重分别为w_{ir}和w_{sr}。信誉可以随节点行为不断累积，添加行为信誉及奖惩机制后，取值范围为[0,+∞)，将社区中节点累积信誉最大值*RPM*作为参照，对信誉进行归一化处理得到综合信誉。假设节点在社区存活期为*TS*，在不同存活期阶段，综合信誉的计算方法如公式（9）所示。

$$RP_i=\begin{cases}\dfrac{(w_{ir}IR+w_{sr}SR)(1+r)}{RPM} & TS_i>TC \\ \dfrac{(w_{ir}IR+w_{sr}SP_{med})(1+r)}{RPM} & TS_i\leqslant TC\text{且}\Delta RP_i>0 \\ \dfrac{w_{ir}IR}{RPM} & TS_i\leqslant TC\text{且}Sd_i<tr\text{且}\Delta RP_i\leqslant 0 \\ 0 & TS_i\leqslant TC\text{且}Sd_i\geqslant tr\end{cases} \quad (9)$$

3 实验

3.1 评估指标

1）信誉误差

传统的社交信任模型一般利用节点公开信息、节点交互行为作为直接和间接信任计算依据。本文实验主要将针对信息行为的奖惩机制、信誉更新、新加入节点信誉管理机制对不同类型节点信誉计算的评估误差*RCE*作为实验指标[16]。*RCE*计算如公式（10）所示。

$$RCE=\frac{1}{n}\sum_{i=1}^{n}|RP_i-R_i| \quad (10)$$

RP_i为模型计算得到节点*i*的信誉值，R_i为实验中所设置节点类别应有的信誉值，$|RP_i-R_i|$则为计算偏离的误差。以不同类别中*n*个节点的平均误差*RCE*作为信誉准确度计算指标，*RCE*值越小，信誉值越能反映出社区成员真实信誉水平，模型准确性越高。

2）不良信息核实时最低交互数量

信誉奖惩机制将引导节点对不良信息传播过程中做出反馈。一般来说，奖惩系数值与举报不良信息的概率存在一定正相关性。在不同概率的举报意愿条件下，将社区中所有不良信息举报所需交互次数*NCE*作为指标，如公式（11）所示。*fm*(*i*)为确定第*i*条不良信息所需要的交互次数，*p*为不同投诉意愿的概率。*NCE*越小，信誉激励机制对不良信息抑制就越有效。

$$NCE=FN_P\left(\sum_{i=1}^{n}fm(i)\right) \quad (11)$$

3.2 实验过程

为了验证信誉计算的准确性，实验将仿真诚信与恶意两种类型节点的信息行为。实验主要验证信息行为及行为奖惩对节点信誉与行为约束的影响，对身份信誉进行统一的初始化，并设置不同类型的信息行为规则，通过程序模拟其行为计算指标差异，主要实验参数描述如下。

1）节点属性。设置节点总数为200个，恶意节点比例从0.1到0.5逐渐增加，其他节点为诚信节点。设置每个节点都随机与社区内20%节点有连接关系。

2）节点行为规则。恶意节点主要提供不良信息发布与转发、无效举报；诚信节点提供较高质量信息发布，发布不良信息后会在短时间内删除、澄清，看到不良信息会进行核查举报。由于现实中会出现偶尔的认知偏差，存在一定的反常行为，所以这两种节点的设置遵从0.1概率的反向行为。即恶意节点行为及其比例为{不良信息发布：0.9，不良信息转发：0.9，无效举报：0.9}，诚信节点行为及其比例为{优质信息发布：0.9，优质信息转发：0.9，有效举报：0.9，删除并澄清：0.9，加精比例：0.05}，平台根据举报对发布与转发的不良信息的节点进行惩罚。

3）信息交互行为规则。实验首先初始化一个社交网络，每个节点随机与20%比例的节点相连，并模拟信息行为，不同节点之间按其行为规则进行信息交互。各类节点信息发布、转发行为100次，信息互动行为500次。假定相连节点中有50%的高信誉节点对信息进行举报，则认定该信息为不良信息，对本条信息做核实标签，并对发帖或转帖人进行信誉惩罚，对举报人

进行信誉奖励。

3.3 实验结果

1）信誉计算误差

图2~图4分别展示了有信誉奖惩与无信誉奖惩两种模型下，不同类型节点在不同比例下的*RCE*值。

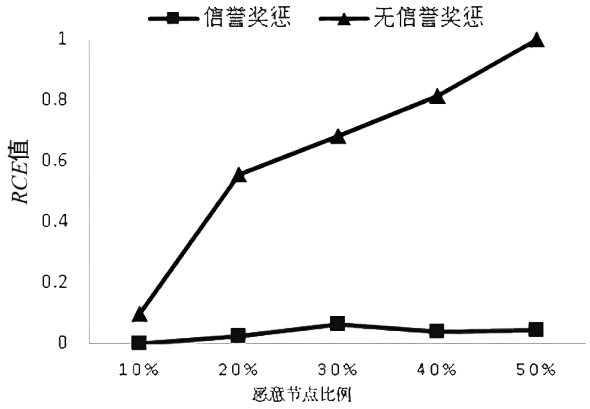

图 2 恶意节点信誉误差

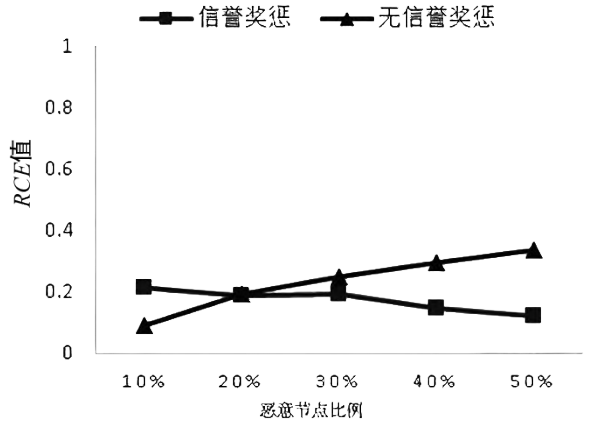

图 3 诚信节点信誉误差

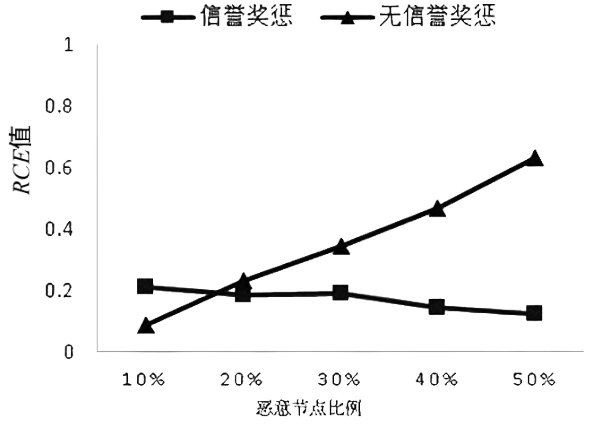

图 4 信誉计算总误差

在图2中，对于恶意节点来说，节点比例较低时，如果没有信誉奖惩，信誉主要是网络相连节点的加权集成。如果社区内恶意节点少，相互之间交互累积信誉的机会也少，则信誉误差相对也比较低；随恶意节点比例的不断增多，其信誉值越容易受到更多恶意节点交互不断更新，达到与诚信节点相平衡的数量时，系统甚至会评估其为诚信节点。对于有信誉奖惩机制的模型，由于仿真的社交网连接设定较为均衡，信息在传播中必定会被网络中诚信节点识别。诚信节点在社区里一旦浏览到节点发布或转发不良信息，会以近90%的概率进行有效投诉。假定平台可以根据投诉行为确定投诉的有效性，恶意节点的信誉将会受到一定系数惩罚，恶意节点*RCE*值则会处在10%左右随机误差范围内。

在图3中，无信誉奖惩的模型不会对诚信节点偶尔的失误信息行为进行惩罚，所以对于诚信节点来说，在恶意节点概率低时，恶意节点与诚信节点交互行为对其信誉影响不大，*RCE*最低。但随着恶意节点不断增多，与恶意节点的交互概率也会增多，会使得诚信节点*RCE*有所提升。恶意节点在社区中更容易发布与转发不良信息，与两类节点的信息交互都会增加，单独对诚信节点的交互相对有限，所以总体误差增幅并不高。而有信誉奖惩模型会激励诚信节点举报不良信息，当社交网络中恶意节点增多时，会滋生更多不良信息，也增加了诚信节点通过举报获得信誉奖励的机会。所以随着恶意节点的增加，诚信节点的*RCE*反而降低。存在少量的*RCE*，主要是实验中设定一定概率反向行为导致。

图4的总体*RCE*中显示，在社交网中恶意节点比例较少的情况下，增加信誉奖惩并不会提高信誉评估准确度。有奖惩的机制反而使得节点出现偶然失误后再进行纠偏还会受到惩罚从而产生信誉偏离。但随着恶意节点比例不断增多，无论是诚信节点和恶意节点的*RCE*都在降低。所以对于总体误差，一旦恶意节点比例增多，在有信誉奖惩引导下，两类节点的*RCE*都

比无奖惩机制的*RCE*明显更低。

2）不同概率下不良信息核查

由于恶意节点缺乏主动维护社区治理的责任感，本文实验将其行为设置为遵从互动规则的随机举报比例。假定信誉的奖励机制会刺激诚信节点对不良信息的举报，在现实中奖励的力度与刺激举报概率关系可能涉及因素较多，如节点活跃度、认知能力、社会责任感等，本文实验暂时未进行细化，但满足一定概率进行举报。图5显示了节点以20%、40%、60%、80%、100%5种不同概率举报时的不良信息的*NCE*值。

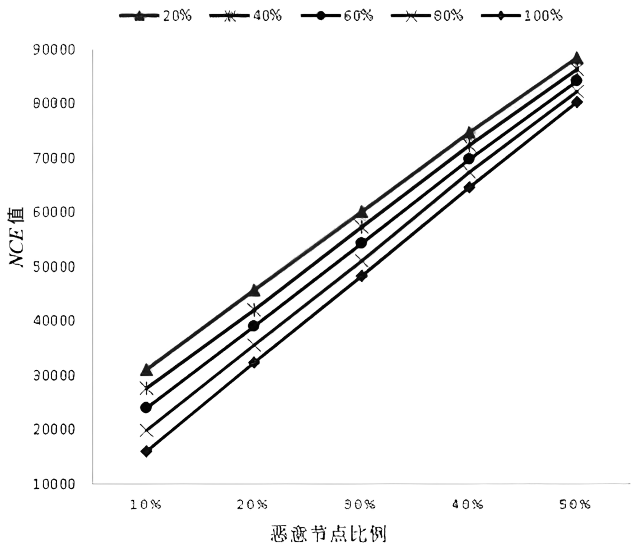

图 5 不同意愿下识别不良信息的交互次数

通过图5可得，诚信节点在5种不同举报概率条件下，通过将社区中所有发布与转发的不良信息进行识别后进行举报，平台核查依据投诉比例需要设定要求，在未被平台核查之前，信息在社交网络中传播的总次数是被交互的次数。恶意节点比例较少的情况下，社区中不良信息比例也较少，被平台识别出来之前，不良信息传播*NCE*相应较少。同理，随着恶意节点比例增多，不良信息也会增多，传播面和传播次数也越多。从5个投诉意愿比例来看，投诉意愿概率越大，不良信息在社区中*NCE*也越低，越快能被平台筛选出来。

4 结束语

社交网络中节点的信誉可以用来衡量用户行为是否可信，是治理社交网络、促进用户自律的重要手段。基于信息行为及行为奖惩引导的信誉评估模型的研究是为了鼓励用户身份信息的有效披露、引导节点信息行为。实验表明引入信誉机制后，在社区存在一定规模恶意节点时，可以更准确地对节点信誉进行评估，识别恶意节点。如果平台可以为高信誉节点赋予更多权限，信誉奖惩将引导节点及时进行自我纠正和参与群体监督，并对恶意节点不良信息行为进行抑制。实验只设置了两种节点类型，下一步将研究更多智能体及其行为规则，并对不同类型新加入节点信誉行为演化过程进行分析，进一步完善信誉模型。

参考文献：

[1] WEI Lulu. The Regulation Responsibility of Social Media Platforms over Its Content from the Perspective of Internet Innovation[J]. Oriental Law, 2020, 13 (1): 27–33.
魏露露 . 互联网创新视角下社交平台内容规制责任 [J]. 东方法学，2020，13（1）：27–33.
[2] JAZAIERI H, LOGLI ALLISON M, CAMPOS B, et al. Content, Structure, and Dynamics of Personal Reputation: The Role of Trust and Status Potential within Social Networks[J]. Group Processes & Intergroup Relations, 2019, 22(7): 964–983
[3] WANG Gang. Literature Review on Network Information Behavior of Domestic Based on Virtual Community[J]. Library, 2017, 45(5): 47–53.
王刚 . 基于虚拟社区的国内网络信息行为研究综述 [J]. 图书馆，2017，45（5）：47–53.
[4] NIE Yonghao, LUO Jingyue. Perceived Usefulness, Trust and Personal Information Disclosure Intention of Social Networking Site Users[J]. Documentation, Information & Knowledge, 2013, 31(5): 89–97.
聂勇浩，罗景月 . 感知有用性、信任与社交网站用户的个人信息披露意愿 [J]. 图书情报知识，2013，31（5）：89–97.
[5] ZHANG Changliang, WANG Xiwei, WANG Yameng, et al. Research on the Development Trend of User Information Behavior in Virtual Community[J]. Information Science, 2018, 36(3): 157–163.
张长亮，王晰巍，王雅梦，等 . 网络社群用户信息行为发展动态及趋势研究 [J]. 情报科学，2018，36（3）：157–163.
[6] LIU Yahui, JIN Xiaolong, SHEN Huawei, et al. A Survey on Rumor Identification over Social Media[J]. Chinese Journal of Computers, 2018, 41 (7): 108–130.
刘雅辉，靳小龙，沈华伟，等 . 社交媒体中的谣言识别研究综述 [J]. 计算机学报，2018，41（7）：108–130.
[7] ZHANG Dayong, SUN Xiaochen. Identifying the Influencing Factors of Users Information Contribution Behavior on Social Network Sites[J]. Information Science, 2018, 36 (2): 95–100.

张大勇，孙晓晨 . 社交网络用户信息贡献行为影响因素分析 [J]. 情报科学，2018，36（2）：95−100.
[8] ZHANG Huiping, GUO Xinhao, TANG Zhiwei. Impact of Sanction Mechanism on Behavioral Intention to Identify Internet Rumor[J]. Journal of Intelligence, 2016, 35 (12): 47−51.
张会平，郭昕昊，汤志伟 . 惩罚机制对网络谣言识别行为的影响研究 [J]. 情报杂志，2016，35（12）：47−51.
[9] URENA R, CHICLANA F, CARRASCO R A, et al. Leveraging Users' Trust and Reputation in Social Networks[J]. Procedia Computer Science, 2019, 162(12): 955−962.
[10] YU Zhen, ZHU Jie, SHEN Guicheng. Trust Model and Simulation of E−commerce Based on Social Networks[J]. Journal of System Simulation, 2018, 30 (8): 304−312.
于真，朱杰，申贵成 . 一种基于社交网络的电子商务信任模型与仿真 [J]. 系统仿真学报，2018，30（8）：304−312.
[11] ZHAO Lihua, YANG Yong, WEN Xi, et al. Research on the Evaluation of Micro−blog Users' Credibility Based on Public Data[J]. Journal of Tianjin University(Social Sciences), 2018, 20(3): 29−35.
赵丽华，杨勇，闻西，等 . 基于公开信息的微博用户可信性评价研究 [J]. 天津大学学报（社会科学版），2018，20（3）：29−35.
[12] ZHANG Jidong, LI Pengcheng. Research on Constructing Quantitative Model of User Trust in Mobile Social Network[J]. Information Studies: Theory & Application, 2017, 40(5): 56−60.
张继东，李鹏程 . 移动社交网络用户信任度量化模型构建研究 [J]. 情报理论与实践，2017，40（5）：56−60.
[13] ZHANG Ning, SU Hua, SUN Xuemei, et al. Crowd Sensing Network User Participation Incentive Mechanism Based on Credibility Model[J]. Computer Applications and Software, 2017, 34(5): 119−122.
张宁，苏华，孙学梅，等 . 基于信誉模型的群智感知网络用户参与激励机制 [J]. 计算机应用与软件，2017，34（5）：119−122.
[14] TENG Jie, XIA Zhijie, LUO Mengying, et al. Research on Trust Recognition of Information Subject in Network Rumor Propagation Event Based on Multi−Agent[J]. Journal of Intelligence, 2020, 39(3): 105−114.
滕婕，夏志杰，罗梦莹，等 . 基于 Multi−Agent 的网络谣言传播事件中信息主体信任识别研究 [J]. 情报杂志，2020，39（3）：105−114.
[15] DAI Bao, LIU Yezheng. Review on the Influence Factors of SNS Users' Information Behavior[J]. Information Science, 2016, 34(6): 170−176.
代宝，刘业政 . SNS 用户信息行为的影响因素研究综述 [J]. 情报科学，2016，34（6）：170−176.
[16] GIANLUCA, GIUSEPPE M L S, CellTrust: a Reputation Model for C2C Commerce[J]. Electron Commerce Research, 2008, 8(4): 193−216.