文章编号：1006-3080(2025)03-0391-09 DOI：10.14135/j.cnki.1006-3080.20240823004

一类存在恶意观点的社会网络舆论演化与防御问题研究

丁文杰， 杨 文
（华东理工大学能源化工过程智能制造教育部重点实验室，上海 200237）

摘要：针对社会网络中的一类舆论安全问题，提出包含嫌疑人、防御者和普通个体的社会网络观点演化模型，将嫌疑人和防御者建模为带外源输入的网络节点，嫌疑人向其他个体传播恶意观点，防御者向其他个体传播积极观点，通过计算群体期望观点到恶意观点和积极观点的距离来判断公共舆论是否安全，据此确定了维护公共舆论安全的积极观点应满足的参数条件，并进一步分析社会群体期望观点收敛的充分条件，设计低成本的有效积极观点的计算方法。

关键词：复杂网络；公共舆论；安全；社会网络；一致性；观点动力学

中图分类号：TP301 **文献标志码：**A

社会舆论深刻地影响个体的行为，并影响时代文化的形成，进而引起社会发展方向的剧烈变化。为抑制部分激烈的社会舆论对社会发展方向的影响，亟需研究舆论传播和调控机制。观点动力学理论主要聚焦于社会网络中观点的共识及调控[1]。公共舆论调控中的一个重要挑战是个体由于受到复杂情感的影响，其意见和行为的变化难以量化表达。社会网络中的个体相互连接，并交换观点或信息，进而推动个体观点的更新及群体观点的演化，这类似于信息物理系统[2-4](Cyber-Physical System, CPS)的通信，CPS 中的部分研究结果也用于观点演化模型的分析[5-7]。

社会网络中的观点演化过程可概括为观点演化模型，包含离散观点演化模型和连续观点演化模型。在离散观点演化模型中，个体观点只有有限备选，观点可由离散变量表示，包括 Ising 模型[8]，多数规则模型[9]，投票者模型[10]，Sznajd 模型[11] 等。现实中个体观点可能处在赞成和反对之间，这时需要用连续变量来概括观点倾向于赞成或反对的程度，具体包含 DeGroot模型[12]，Friedkin-Johnsen(FJ)模型[13] 等。

DeGroot 模型[12] 探究了群体观点的共识现象，FJ 模型[13] 进一步考察了个体偏见对观点演化的影响，这使得群体观点无法达成全面共识。Altafini[14] 分析了对抗环境中的共识问题，并证明了群体观点在一种社会网络(Signed Networks)[15] 上能达到共识。Borkar 等[16] 在 DeGroot 模型基础上研究了带社会规划者的观点演化模型。He 等[17] 在 FJ 模型基础上研究了对抗关系下，包含顽固个体的多个相关观点的演化模型。Parsegov 等[18] 考虑了一个多话题观点演化问题，用耦合矩阵 $\boldsymbol{C}$ 解释多个话题的观点间的相互作用。

有界信任模型是一类特殊的连续观点演化模型，它研究个体支持相近观点、排斥相异观点的社交现象，经典模型有 Deffuant 模型[19] 和 Hegselmann-Krause(HK)模型[20]。在 Deffuant 模型中，个体随机选择与另一个体交换观点，如果两方的观点相近，则相互靠近对方的观点，否则暂停一次观点更新。在 HK 模型中，个体与所有相近的观点交互，并基于所有相近观点更新自身观点。

本文结合 DeGroot 和 FJ 观点模型，提出一种社会网络舆论安全视角下的观点演化模型，将舆论破坏者和守护者分别建模为攻击者和政府，攻击者通过嫌疑人个体将恶意观点传播到社会网络中，并引导公共舆论向消极方向发展；政府通过防御者个体

收稿日期：2024-08-23
基金项目：国家重点研发计划项目 (2023YFF1204805)；国家自然科学基金 (62336005, 62122026)
作者简介：丁文杰(1995—)，男，安徽人，博士生，主要研究方向为状态估计和网络安全。E-mail: dingwenjie95@gmail.com
通信联系人：杨 文，E-mail: weny@ecust.edu.cn
引用本文：丁文杰，杨 文. 一类存在恶意观点的社会网络舆论演化与防御问题研究 [J]. 华东理工大学学报(自然科学版)，2025, 51(3): 391-399.
Citation: DING Wenjie, YANG Wen. Evolution and Defense of Public Opinion in a Class of Social Networks with Malicious Opinions[J]. Journal of East China University of Science and Technology, 2025, 51(3): 391-399.

传播积极观点并引导公共舆论远离恶意观点。基于此,本文分析并论证了群体期望观点收敛的充分条件,并设计了计算低成本的积极观点的方法。

1 问题描述

1.1 社会网络

一个社会网络可以建模为一个加权有向图 $\mathcal{G}=(\mathcal{V},\mathcal{E},\boldsymbol{W})$,其中 $\mathcal{V}=\{v_1,\cdots,v_n\}$ 是节点集合,社会网络中的节点就是个体,n 是个体的数量,$\mathcal{E}\subset\mathcal{V}\times\mathcal{V}$ 是边集合,边 $(v_i,v_j)\in\mathcal{E}$ 表示节点 v_i 和 v_j 相连,即个体 i 和个体 j 会交流观点。集合 $N_i=\{v_j\in\mathcal{V}:(v_i,v_j)\in\mathcal{E}\}$ 是节点 v_i 的邻居节点集,且 $v_i\notin N_i$。非负的权矩阵 $\boldsymbol{W}$ 表示图中的节点邻接关系,w_{ij} 是矩阵 $\boldsymbol{W}$ 中第 i 行、第 j 列的元素,如果 $(v_i,v_j)\in\mathcal{E}$,那么 $w_{ij}>0$;否则 $w_{ij}=0$。如果权矩阵 $\boldsymbol{W}$ 满足 $\Sigma_j w_{ij}=1$,$i\in\{1,2,\cdots,n\}$,$j\in\{1,2,\cdots,n\}$,它就是一个行随机矩阵。

借助社会网络的图模型可以有效表达不同个体之间交换观点的过程。由于不同个体对不同思想、观点的接受能力不同,个体 i 和个体 j 受对方观点影响的权重是不同的,甚至可能出现 $w_{ij}>0$,而 $w_{ji}=0$ 的情况,因此不同个体之间交换观点的连接关系由有向图表达。

1.2 DeGroot 模型

将数量为 n 的个体集合 $\{1,2,\cdots,n\}$ 记为 Ω,在 DeGroot 模型中,每个个体按如下方式更新观点:

$$\boldsymbol{x}_i(k+1)=\Sigma_{j\in N_i\cup\{i\}}w_{ij}\boldsymbol{x}_j(k),\ i\in\Omega \tag{1}$$

其中,$\boldsymbol{x}_i(k)$ 是个体 i 在 k 时刻的观点,w_{ij} 是个体 j 对个体 i 观点影响的权重。在模型 (1) 中,当个体观点 $\boldsymbol{x}_i(k)$ 是标量时(针对单个话题的观点可由标量表达),根据平均一致性协议,所有标量观点都收敛到初始标量值的平均值;当 $\boldsymbol{x}_i(k)$ 是向量时(针对多个话题的观点可由向量表达),由于不同维度的观点之间不会相互影响,所有维度的观点都会收敛到这个维度上所有初始值的平均值,即所有观点会收敛到观点向量初始值的平均值。

在 DeGroot 模型中,个体的观点只受这个网络中的其他个体或自身的观点影响。考虑到不同社会的观念差异及社会群体为维护内部社会稳定有将群体观点扩散到其他社会的动机,一个社会内部的个体观点可能受到外源输入的影响。基于 FJ 模型,根据个体观点是否受外源输入的影响及外源输入的属性,可将个体分成 3 类:普通个体、防御者和嫌疑人,其中外源输入可分为受社会群体内部组织(简化为政府)影响和社会群体外部组织(简化为攻击者)影响的两类。个体观点只受网络中已有的观点影响的个体称为普通个体,个体观点会受到网络中已有的观点和政府主导的积极观点影响的个体称为防御者,个体观点会受到网络中已有的观点和攻击者主导的恶意观点影响的个体称为嫌疑人。

基于 DeGroot 模型 (1),普通个体的观点更新方程为:

$$\boldsymbol{x}_i(k+1)=\Sigma_{j\in N_i\cup\{i\}}w_{ij}\boldsymbol{C}\boldsymbol{x}_j(k),\ i\in\{\Omega/(J\cup D)\} \tag{2}$$

其中,$\boldsymbol{x}_i(k)=(\boldsymbol{x}_{i1}(k),\cdots,\boldsymbol{x}_{im}(k))^{\mathrm{T}}$ 是 k 时刻个体 i 关于 m 个话题的观点,行随机矩阵 $\boldsymbol{C}$ 表示不同的话题观点之间的耦合关系,J 是嫌疑人集合,D 是防御者集合。

1.3 攻击模型

攻击者影响公共舆论的方法包含两步,筛选可用的个体(即嫌疑人),将恶意观点渗透给嫌疑人并借助嫌疑人将恶意观点传播给其他个体。受攻击者影响的嫌疑人有两种观点更新模式:(1) 嫌疑人的观点来自攻击者的恶意观点;(2) 嫌疑人的观点来自网络已有的观点。考虑到个体观点容易受到极端情绪影响,短暂受到恶意观点影响的个体不会被识别为嫌疑人。频繁传播恶意观点的个体容易被识别为嫌疑人,为保障嫌疑人的隐匿性,嫌疑人需要交替地根据两种模式更新观点,且会将当前观点传播给相邻个体。

记 $\boldsymbol{u}\in\mathbb{R}^m$ 是攻击者关于 m 个话题的恶意观点,$\tilde{\boldsymbol{u}}\in\mathbb{R}^m$ 是政府关于 m 个话题的积极观点。嫌疑人的观点更新方程为:

$$\boldsymbol{x}_i(k+1)=(1-\boldsymbol{h}_i(k))\Sigma_{j\in N_i\cup\{i\}}w_{ij}\boldsymbol{C}\boldsymbol{x}_j(k)+\boldsymbol{h}_i(k)\boldsymbol{u},\ i\in J \tag{3}$$

其中,$\boldsymbol{h}_i(k)\in\{0,1\}$ 是嫌疑人选择更新模式的二值变量,$\boldsymbol{u}=(u_1,\cdots,u_m)^{\mathrm{T}}$ 是攻击者渗透给嫌疑人的恶意观点,$\boldsymbol{u}_l$ 是攻击者对第 l 个话题的恶意观点,$l\in\{1,2,\cdots,m\}$。当 $\boldsymbol{h}_i(k)=1$ 时,嫌疑人 i 的观点来自攻击者的恶意观点,且嫌疑人将传播恶意观点;当 $\boldsymbol{h}_i(k)=0$ 时,嫌疑人 i 的观点来自社会网络中的已有观点。当嫌疑人频繁传播恶意观点时,身份就会暴露。为调控嫌疑人传播恶意观点的频率,攻击者为嫌疑人 i 设定调控变量 $\boldsymbol{p}_i\in[0,1]$。在 k 时刻,嫌疑人 i 根据如下公式确定 $\boldsymbol{h}_i(k)$:

$$\boldsymbol{h}_i(k)=\begin{cases}1, & s>\boldsymbol{p}_i\\ 0, & \text{Otherwise}\end{cases} \tag{4}$$

其中,s 表示嫌疑人 i 根据 0~1 之间的均匀分布生成

的随机数, 且变量 $\boldsymbol{h}_i(k)$ 满足 $\text{Prob}(\boldsymbol{h}_i(k)=1)=1-\boldsymbol{p}_i$ 和 $\text{Prob}(\boldsymbol{h}_i(k)=0)=\boldsymbol{p}_i$。

观点更新式 (3) 中有随机变量 $\boldsymbol{h}_i(k)$, 可通过在方程两边同时取期望获得嫌疑人期望观点(观点的期望)的更新方程:

$$E[\boldsymbol{x}_i(k+1)]=E[(1-\boldsymbol{h}_i(k)\Sigma_{j\in N_i\cup\{i\}}w_{ij}\boldsymbol{C}\times E[\boldsymbol{x}_j(k)]+ E[\boldsymbol{h}_i(k)]\boldsymbol{u},\ i\in J \tag{5}$$

其中, $\boldsymbol{h}_i(k)$ 和 $\boldsymbol{x}_j(k)$ 相互独立, $E[\boldsymbol{x}_j(k)]=[E[\boldsymbol{x}_{j1}(k)],\cdots, E[\boldsymbol{x}_{jm}(k)]]^{\mathrm{T}}$。进一步简化方程 (5) 可得

$$E[\boldsymbol{x}_i(k+1)]=\boldsymbol{p}_i\Sigma_{j\in N_i\cup\{i\}}\boldsymbol{w}_{ij}\boldsymbol{C}\times E[\boldsymbol{x}_j(k)]+ (1-\boldsymbol{p}_i)\boldsymbol{u},\ i\in J \tag{6}$$

结合图 1 中案例, 进一步解释嫌疑人和普通个体在社会网络中的观点演化过程。个体 1 是嫌疑人, 在 k 时刻, 生成随机数并计算 $\boldsymbol{h}_i(k)$, 随后根据式 (3) 更新观点。个体 2 和个体 3 是两个普通个体, 根据式 (2) 更新观点。

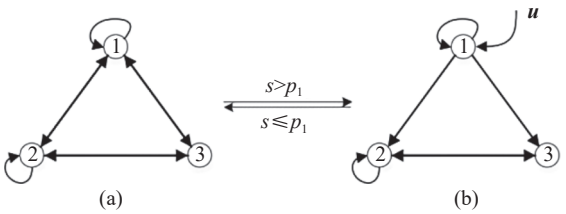


图 1　嫌疑人和普通个体交互

Fig. 1　A suspect interacts with masses

2　观点演化分析

2.1　防御模型

为抵御嫌疑人传播的恶意观点, 政府可效仿攻击者, 构建防御机制调控公共舆论。首先政府筛选出防御者, 然后防御者根据两种模式更新观点: (1) 防御者的观点来自网络已有的观点; (2) 防御者的观点来自政府的积极观点。初始阶段, 防御者按照第 1 种模式更新观点, 当防御者感知到恶意观点在传播时, 它会按照第 2 种模式更新观点。为帮助防御者感知恶意观点是否在传播, 进行如下定义:

定义 1　危险距离和安全距离

$$\text{记}\quad D_{ij}(k)=\Sigma_{r\in N_i\cup\{i\}}(\boldsymbol{x}_{ij}(k)-\boldsymbol{u}_j)^2,\ i\in D \tag{7}$$

其中, $D_{ij}(k)$ 为防御者 i 感知的关于第 j 个话题的观点到恶意观点的危险距离。相似地, 记防御者 i 感知的关于第 j 个话题的观点到积极观点的安全距离为:

$$S_{ij}(k)=\Sigma_{r\in N_i\cup\{i\}}(\boldsymbol{x}_{ij}(k)-\tilde{\boldsymbol{u}}_j)^2,\ i\in D \tag{8}$$

政府需要先确定恶意观点 $\boldsymbol{u}$, 在筛选出防御者后, 计算积极观点 $\tilde{\boldsymbol{u}}$, 将 $\boldsymbol{u}$ 和 $\tilde{\boldsymbol{u}}$ 共享给防御者。在每个时刻, 防御者 i 都会计算 $D_{ij}(k)$ 和 $S_{ij}(k)$, 当 $D_{ij}(k)\leqslant S_{ij}(k)$ 时, 防御者 i 认为第 j 个话题的观点演化已经受到恶意观点影响, 并将根据第 2 种模式更新第 j 个话题的观点。此时, 防御者 i 自身关于第 j 个话题的观点被设置为 $\tilde{\boldsymbol{u}}_j$, 且通过传播观点 $\tilde{\boldsymbol{u}}_j$ 来进一步缩短感知到的安全距离 $S_{ij}(k)$, 以达到调控公共舆论的目的。

记防御者 i 关于第 j 个话题的观点更新模式的切换时间为 $\boldsymbol{\tau}_{ij}$, $\boldsymbol{\tau}_{ij}$ 是使 $D_{ij}(k)\leqslant S_{ij}(k)$ 成立的最小 k。结合 $\boldsymbol{\tau}_{ij}$, 防御者 i 的观点更新方程为:

$$\begin{cases}\boldsymbol{x}_{ij}(k+1)=\boldsymbol{e}_j^{\mathrm{T}}(\Sigma_{l\in N_i\cup\{i\}}w_{il}\boldsymbol{C}\boldsymbol{x}_l(k)),\ k\leqslant\boldsymbol{\tau}_{ij}\\ \boldsymbol{x}_{ij}(k+1)=\tilde{\boldsymbol{u}}_j,\ k>\boldsymbol{\tau}_{ij}\\ i\in D, j\in\{1,\cdots,m\}\end{cases} \tag{9}$$

其中, $\tilde{\boldsymbol{u}}_j$ 是政府针对第 j 个话题设置的积极观点, $\boldsymbol{e}_j$ 是一个 m 维基向量, 其第 j 个元素为 1, 其他元素为 0。防御者期望观点的更新方程为:

$$\begin{cases}E[\boldsymbol{x}_i(k+1)]=\boldsymbol{e}_j^{\mathrm{T}}(\Sigma_{l\in N_i\cup\{i\}}w_{il}\boldsymbol{C}\times E[x_l(k)]),\ k\leqslant\boldsymbol{\tau}_{ij},\\ E[\boldsymbol{x}_i(k+1)]=\tilde{\boldsymbol{u}}_j, k>\boldsymbol{\tau}_{ij}\\ i\in D, j\in\{1,\cdots,m\}\end{cases} \tag{10}$$

防御者和嫌疑人观点更新的相似之处在于两者都包含两种模式, 且在一种模式下, 观点都来自网络中的已有观点, 在另一种模式下, 观点都来自外源输入。两者观点更新的不同之处在于, 防御者根据每个话题的观点演变是否受到恶意观点影响来切换更新模式, 每次只调整单个观点的更新模式, 且防御者的更新模式切换更稳定。一旦防御者切换更新模式, 就能更稳定地向网络传播积极观点。基于以上分析, 政府通过防御者调控公共舆情, 能更有针对性地调控受到恶意观点影响的话题观点, 并保护其他话题观点的自由。

2.2　公共舆论分析

为分析包含 c 个嫌疑人、d 个防御者和($n-c-d$)个普通个体的社会网络公共舆情演化特性, 定义社会群体观点向量 $\boldsymbol{x}(k)=[\boldsymbol{x}_1(k)^{\mathrm{T}},\boldsymbol{x}_2(k)^{\mathrm{T}},\cdots,\boldsymbol{x}_n(k)^{\mathrm{T}}]^{\mathrm{T}}$ (个体的排列顺序为: 嫌疑人、防御者、普通个体), 对应的社会群体期望观点为 $E[x(k)]=[E[\boldsymbol{x}_1(k)]^{\mathrm{T}},\cdots, E[\boldsymbol{x}_n(k)]^{\mathrm{T}}]^{\mathrm{T}}$。记 c 个嫌疑人的切换概率构成的向量为 $\boldsymbol{p}=[p_1,\cdots,p_c]^{\mathrm{T}}$, 普通个体期望观点的更新方程为:

$$E[\boldsymbol{x}_i(k+1)]=\Sigma_{j\in N_i\cup\{i\}}\boldsymbol{w}_{ij}\boldsymbol{C}\times E[\boldsymbol{x}_j(k)],\quad i\in\{\Omega/(J\cup D)\} \tag{11}$$

结合嫌疑人、防御者和普通个体的期望观点分别更新方程 (6)、(10) 和 (11), 得到社会群体期望观点的更新方程:

$$E[\boldsymbol{x}(k+1)]=\boldsymbol{Q}(\boldsymbol{\tau})\{\boldsymbol{P}(\boldsymbol{p})\cdot\boldsymbol{W}\otimes\boldsymbol{C}\}\times E[\boldsymbol{x}(k)]+\boldsymbol{U}(\boldsymbol{\tau},\boldsymbol{p}),\quad k=0,1,2,\cdots \tag{12}$$

其中，$\boldsymbol{W}\otimes\boldsymbol{C}$ 表示矩阵 $\boldsymbol{W}$ 和矩阵 $\boldsymbol{C}$ 的克罗内克积，

$$\boldsymbol{Q}(\boldsymbol{\tau})=\mathrm{diag}\{1,\cdots,1,1-\delta(\boldsymbol{\tau}_{11}),\cdots,1-\delta(\boldsymbol{\tau}_{dm}),1,\cdots,1\}_{mn\times mn},$$

$$\boldsymbol{P}(\boldsymbol{p})=\mathrm{diag}\{\boldsymbol{p}_1\boldsymbol{I}_m,\cdots,\boldsymbol{p}_c\boldsymbol{I}_m,\boldsymbol{I}_m,\cdots,\boldsymbol{I}_m\}_{mn\times mn},$$

$$\boldsymbol{U}(\boldsymbol{\tau},\boldsymbol{p})=\begin{bmatrix}(1-\boldsymbol{p}_1)\boldsymbol{u}\\ \vdots\\ (1-\boldsymbol{p}_c)\boldsymbol{u}\\ \delta(\boldsymbol{\tau}_{11})\tilde{\boldsymbol{u}}_1\\ \vdots\\ \delta(\boldsymbol{\tau}_{dm})\tilde{\boldsymbol{u}}_m\\ 0\\ \vdots\\ 0\end{bmatrix}_{mn\times 1}$$

$$\delta(\boldsymbol{\tau}_{ij})=\begin{cases}1, & k>\boldsymbol{\tau}_{ij}\\ 0, & \text{Otherwise}\end{cases}$$

同理，结合式 (2)、(3) 和 (9)，可得社会群体观点的更新方程：

$$\boldsymbol{x}(k+1)=\boldsymbol{Q}(\boldsymbol{\tau})\{\boldsymbol{H}(\boldsymbol{h}(k))\cdot\boldsymbol{W}\otimes\boldsymbol{C}\}\boldsymbol{x}(k)+\boldsymbol{U}(\boldsymbol{\tau},\boldsymbol{h}(k)),\quad k=0,1,2,\cdots \tag{13}$$

其中：

$$\boldsymbol{H}(\boldsymbol{h}(k))=\mathrm{diag}\{(1-\boldsymbol{h}_1(k))\boldsymbol{I}_m,\cdots,(1-\boldsymbol{h}_c(k))\boldsymbol{I}_m,\boldsymbol{I}_m,\cdots,\boldsymbol{I}_m\}_{mn\times mn}$$

$$\boldsymbol{U}(\boldsymbol{\tau},\boldsymbol{h}(k))=\begin{bmatrix}\boldsymbol{h}_1(k)\boldsymbol{u}\\ \vdots\\ \boldsymbol{h}_c(k)\boldsymbol{u}\\ \delta(\boldsymbol{\tau}_{11})\tilde{\boldsymbol{u}}_1\\ \vdots\\ \delta(\boldsymbol{\tau}_{dm})\tilde{\boldsymbol{u}}_m\\ 0\\ \vdots\\ 0\end{bmatrix}_{mn\times 1}$$

基于以上定义，分析社会群体观点的收敛性。首先定义

$$\boldsymbol{A}(\boldsymbol{\tau},\boldsymbol{p})=\boldsymbol{Q}(\boldsymbol{\tau})\{\boldsymbol{P}(\boldsymbol{p})\cdot\boldsymbol{W}\otimes\boldsymbol{C}\} \tag{14}$$

$\boldsymbol{A}(\boldsymbol{\tau},\boldsymbol{p})$ 就是式 (9) 中 $E[x(k)]$ 的系数矩阵。

定义 2 正则性

如果方阵 $\boldsymbol{A}$ 满足 $\lim\limits_{k\to\infty}\boldsymbol{A}^k$ 存在且 $\boldsymbol{A}^*=\lim\limits_{k\to\infty}\boldsymbol{A}^k$，则称方阵 $\boldsymbol{A}$ 是正则的。

定理 1 记 $\boldsymbol{\tau}_{ij}$ 构成的集合为 D'，$\boldsymbol{\tau}_{\max}=\max\limits_{\boldsymbol{\tau}_{ij}\in D'}\{\boldsymbol{\tau}_{ij}\}$ 及 $E[x']\triangleq\lim\limits_{k\to\infty}E[x(k)]$。当 $k>\boldsymbol{\tau}_{\max}$ 时，如果 $\rho(\boldsymbol{A}(\boldsymbol{\tau},\ \boldsymbol{p}))<1$，则社会群体期望观点会收敛到 $E[\boldsymbol{x}']$，且 $E[\boldsymbol{x}']$ 的计算公式为：

$$E[x']=(1-\boldsymbol{A}(\boldsymbol{\tau},\boldsymbol{p}))^{-1}\boldsymbol{U}(\boldsymbol{\tau},\boldsymbol{p}) \tag{15}$$

证明：

当 $k\geqslant\boldsymbol{\tau}_{\max}$ 时，所有 $\delta(\boldsymbol{\tau}_{ij})$ 的值都为 1，$i\in D$。此时，对于嫌疑人 $i\in J$，所有 $\boldsymbol{p}_i\in[0,1]$ 都已知。这时，$\boldsymbol{Q}(\boldsymbol{\tau})$ 和 $\boldsymbol{P}(\boldsymbol{p})$ 都是常数矩阵，$\boldsymbol{A}(\boldsymbol{\tau},\boldsymbol{p})$ 是一个行次随机矩阵。根据方程 (12) 可得

$$E[\boldsymbol{x}(k)]=[\boldsymbol{A}(\boldsymbol{\tau},\boldsymbol{p})]^kE[\boldsymbol{x}(\boldsymbol{\tau}_{\max})]+\sum_{i=\boldsymbol{\tau}_{\max}}^{k}[\boldsymbol{A}(\boldsymbol{\tau},\boldsymbol{p})]^{i-\boldsymbol{\tau}_{\max}}\boldsymbol{U}(\boldsymbol{\tau},\boldsymbol{p}),\quad k>\boldsymbol{\tau}_{\max} \tag{16}$$

当 $\rho(\boldsymbol{A}(\boldsymbol{\tau},\boldsymbol{p}))<1$ 时，可得

$$\lim_{k\to\infty}E[\boldsymbol{x}(k)]=\sum_{i=\boldsymbol{\tau}_{\max}}^{\infty}[\boldsymbol{A}(\boldsymbol{\tau},\boldsymbol{p})]^{i-\boldsymbol{\tau}_{\max}}\boldsymbol{U}(\boldsymbol{\tau},\boldsymbol{p})=(1-\boldsymbol{A}(\boldsymbol{\tau},\boldsymbol{p}))^{-1}\boldsymbol{U}(\boldsymbol{\tau},\boldsymbol{p}) \tag{17}$$

即社会群体期望观点会收敛到 $E[\boldsymbol{x}']$，且 $E[x']$ 的计算如式 (15) 所示。

定理 1 分析了社会群体期望观点的收敛性，由此可以得到社会群体期望观点收敛的充分条件。在公共舆论安全问题中还需探究防御机制的有效性，为此定义社会群体观点的安全距离为 $S(k)=\|\boldsymbol{x}(k)-\boldsymbol{1}_n\otimes\tilde{\boldsymbol{u}}\|^2$，社会群体观点的危险距离为 $D(k)=\|\boldsymbol{x}(k)-\boldsymbol{1}_n\otimes\boldsymbol{u}\|^2$。当 k 趋向于 ∞ 时，如果 $E[S(k)]<E[D(k)]$ 成立，可以认为防御者采用的防御机制是有效的。进一步地，另一个重要的问题是防御者需要传播怎样的积极观点 $\tilde{\boldsymbol{u}}$ 来保障防御机制的有效性，不妨假设攻击者的恶意观点 $\boldsymbol{u}$ 总是取较小值（关于某话题的观点的数值仅仅是数学表示），积极观点 $\tilde{\boldsymbol{u}}$ 总是取较大值，恶意观点的每个维度都小于积极观点在这个维度上的取值，即 $\boldsymbol{u}<\tilde{\boldsymbol{u}}$。基于此，在如下定理中分析防御机制的有效性条件。

定理 2 假设 $\boldsymbol{u}<\tilde{\boldsymbol{u}}$，如果社会网络中观点模型 (13) 的参数满足以下线性不等式

$$\boldsymbol{\beta}<\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{\tau},\boldsymbol{p}) \tag{18}$$

其中，$\boldsymbol{\alpha}=\boldsymbol{2}[\boldsymbol{1}_n\otimes(\tilde{\boldsymbol{u}}-\boldsymbol{u})]$，$\boldsymbol{U}(\boldsymbol{\tau},\boldsymbol{p})$ 取自式 (12)，则防御策略是有效的，即 $E[S(k)]<E[D(k)]$，$k\to\infty$。此外，当防御策略有效时，防御者传播的积极观点的下界记为 $\underline{\boldsymbol{u}}$，它可以通过求解优化问题获得

$$\min_{\tilde{\boldsymbol{u}}}\{\boldsymbol{\alpha}^{\mathrm{T}}\underline{\boldsymbol{U}}-\boldsymbol{\beta}\},\ 0\leqslant\underline{\boldsymbol{u}}\leqslant\tilde{\boldsymbol{u}} \tag{19}$$

其中，$\boldsymbol{\beta}=n(\tilde{\boldsymbol{u}}^{\mathrm{T}}\tilde{\boldsymbol{u}}-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u})-\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}\cdot E[x(k)]$，$\boldsymbol{A}=\boldsymbol{A}(\boldsymbol{\tau},\boldsymbol{p})$，

$$\underline{\boldsymbol{U}}=\begin{bmatrix}(1-\boldsymbol{p}_1)\boldsymbol{u}\\ \vdots\\ (1-\boldsymbol{p}_c)\boldsymbol{u}\\ \delta(\boldsymbol{\tau}_{11})\underline{\boldsymbol{u}}_1\\ \vdots\\ \delta(\boldsymbol{\tau}_{dm})\underline{\boldsymbol{u}}_m\\ 0\\ \vdots\\ 0\end{bmatrix}_{mn\times 1}$$

证明:

因为 $\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{\tau},\boldsymbol{p})>\boldsymbol{\beta}$, 所以有

$$2[\boldsymbol{1}_n\otimes(\tilde{\boldsymbol{u}}^{\mathrm{T}}-\boldsymbol{u}^{\mathrm{T}})]\boldsymbol{U}(\boldsymbol{\tau},\boldsymbol{p})>n(\tilde{\boldsymbol{u}}^{\mathrm{T}}\tilde{\boldsymbol{u}}-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u})-2[\boldsymbol{1}_n\otimes(\tilde{\boldsymbol{u}}^{\mathrm{T}}-\tilde{\boldsymbol{u}}^{\mathrm{T}})]\boldsymbol{A}\cdot E[\boldsymbol{x}(k)]$$

通过移项, 可以得到

$$2[\boldsymbol{1}_n\otimes(\tilde{\boldsymbol{u}}^{\mathrm{T}}-\boldsymbol{u}^{\mathrm{T}})][\boldsymbol{U}(\boldsymbol{\tau},\boldsymbol{p})+\boldsymbol{A}\cdot E[\boldsymbol{x}(k)]]>(\boldsymbol{1}_n\otimes\tilde{\boldsymbol{u}})^{\mathrm{T}}(\boldsymbol{1}_n\otimes\tilde{\boldsymbol{u}})-(\boldsymbol{1}_n\otimes\boldsymbol{u})^{\mathrm{T}}(\boldsymbol{1}_n\otimes\boldsymbol{u})$$

结合式 (12), 可进一步简化为:

$$2[\boldsymbol{1}_n\otimes(\tilde{\boldsymbol{u}}^{\mathrm{T}}-\boldsymbol{u}^{\mathrm{T}})]E[\boldsymbol{x}(k+1)]>(\boldsymbol{1}_n\otimes\tilde{\boldsymbol{u}})^{\mathrm{T}}(\boldsymbol{1}_n\otimes\tilde{\boldsymbol{u}})-(\boldsymbol{1}_n\otimes\boldsymbol{u})^{\mathrm{T}}(\boldsymbol{1}_n\otimes\boldsymbol{u})$$

由此, 可以得到

$$[^{E}[E[\boldsymbol{x}(k+1)]]-2\cdot\boldsymbol{1}_n\otimes\boldsymbol{u}^{\mathrm{T}}E[\boldsymbol{x}(k+1)]+(^{\boldsymbol{1}_n}(\boldsymbol{1}_n\otimes\boldsymbol{u})>$$
$$[^{E}[E[\boldsymbol{x}(k+1)]]-2\cdot\boldsymbol{1}_n\otimes\tilde{\boldsymbol{u}}^{\mathrm{T}}E[\boldsymbol{x}(k+1)]+(^{\boldsymbol{1}_n}(\boldsymbol{1}_n\otimes\tilde{\boldsymbol{u}})$$

由此可得

$$\|E[\boldsymbol{x}(k+1)]-\boldsymbol{1}_n\otimes\boldsymbol{u}\|^2>\|E[x(k+1)]-\boldsymbol{1}_n\otimes\tilde{\boldsymbol{u}}\|^2$$

进而可知防御策略是有效的, 且 $\underline{\boldsymbol{u}}$ 可通过求解以上优化问题获得。

基于以上定理, 积极观点的数值受两方面因素影响: 恶意观点的数值及嫌疑人和防御者的数量。因为普通个体的观点更新式 (2) 近似于平均一致性的状态更新, 所以普通个体某一维度的观点会受到其他个体相同维度的观点值均值影响。由于嫌疑人和防御者的观点更新方程不同于平均一致性的状态更新, 选择有效的积极观点时需要考虑权重矩阵 $\boldsymbol{W}$ 和耦合矩阵 $\boldsymbol{C}$ 的影响, 这体现于 $\boldsymbol{A}(\boldsymbol{\tau},\boldsymbol{p})$。从以上定理出发, 如果不限定积极观点 $\tilde{\boldsymbol{u}}$ 的取值范围, 积极观点的数值越大则越容易满足定理中的约束条件。因而在现实中, 政府可以通过宣传强烈的积极观点来抑制恶意观点的传播, 避免舆论危机的爆发。

3　实验分析

3.1　实验参数

首先确定社会网络的规模, 考虑一个由 $n=8$ 个成员构成的社交子网络, 假定网络拓扑为全连接图, 不同个体的观点之间相互影响的权重矩阵(先生成随机矩阵, 再做行和归一化)为:

$$\boldsymbol{W}=\begin{bmatrix}0.283\,1 & 0.122\,3 & 0.132\,3 & 0.197\,8 & 0.057\,6 & 0.079\,6 & 0.037\,1 & 0.090\,3\\ 0.045\,2 & 0.154\,2 & 0.105\,4 & 0.087\,1 & 0.111\,7 & 0.169\,5 & 0.178\,5 & 0.148\,5\\ 0.378\,7 & 0.238\,5 & 0.030\,9 & 0.004\,9 & 0.107\,2 & 0.062\,1 & 0.001\,9 & 0.175\,8\\ 0.078\,5 & 0.123\,4 & 0.012\,1 & 0.075\,7 & 0.146\,8 & 0.185\,3 & 0.173\,9 & 0.204\,4\\ 0.048\,7 & 0.227\,4 & 0.131\,6 & 0.040\,2 & 0.170\,9 & 0.133\,5 & 0.202\,6 & 0.202\,6\\ 0.050\,3 & 0.057\,3 & 0.156\,2 & 0.159\,3 & 0.150\,0 & 0.199\,7 & 0.174\,2 & 0.052\,9\\ 0.182\,4 & 0.224\,2 & 0.276\,6 & 0.092\,2 & 0.133\,4 & 0.023\,1 & 0.025\,0 & 0.043\,1\\ 0.160\,6 & 0.255\,7 & 0.044\,1 & 0.179\,3 & 0.028\,4 & 0.150\,2 & 0.135\,6 & 0.046\,2\end{bmatrix}\tag{20}$$

不同个体间交流 $m=2$ 个相关的话题, 话题耦合矩阵为 $\boldsymbol{C}=\begin{bmatrix}0.8 & 0.2\\ 0.2 & 0.8\end{bmatrix}$。

社会网络中嫌疑人和防御者的个数分别为 2 和 2, 普通个体数量为 4。嫌疑人传播恶意观点的概率为 $p_1=p_2=0.5$。防御者的初始 $\delta(\boldsymbol{\tau}_{ij})$ 为 0 ($i\in D$, $j\in\{1,2\}$), 且 $\boldsymbol{Q}(\boldsymbol{\tau})=\boldsymbol{I}_{16\times 16}$。

个体观点的初始值设置为:

$$[\boldsymbol{x}_1(0),\boldsymbol{x}_2(0),\cdots,\boldsymbol{x}_8(0)]=\begin{bmatrix}25 & 25 & 75 & 85 & 60 & 75 & 15 & 20\\ 25 & 15 & 55 & 45 & 35 & 50 & 20 & 70\end{bmatrix}\tag{21}$$

当政府介入时, 普通个体、嫌疑人和防御者分别根据方程 (2), (3) 和 (9) 更新观点, 其中权重矩阵 $\boldsymbol{W}$ 和话题耦合矩阵 $\boldsymbol{C}$ 已知; 当无政府介入时, 防御者根据方程 (2) 更新观点。嫌疑人根据随机变量 s 和 $\boldsymbol{p}_i$ 决定变量 $\boldsymbol{h}_i(k)$, $i\in J$。当存在防御者时, 假定政府已知恶意观点 $\boldsymbol{u}$, 并计算得到积极观点 $\tilde{\boldsymbol{u}}$, 防御者通过比较安全距离和危险距离的大小关系得到 $\boldsymbol{\tau}_{ij}$, 进而确定观点更新方式。

3.2　结果分析

首先研究政府在不采取任何防御措施情况下的观点演化, 在这种场景下, 社会网络包含 2 个嫌疑人

和 6 个普通个体，其中嫌疑人传播的恶意观点为 $[10,5]^{\mathrm{T}}$ 。所有个体的观点都收敛于嫌疑人的恶意观点，如图 2(a) 和 2(b) 所示；从图 2(c) 和 2(d) 可知，舆论观点到恶意观点的距离为 0，这意味着舆论危机的产生。在这种情况下，舆论观点到积极观点的距离的具体数值不再重要，因为舆论观点已经被恶意观点主导。为避免这种情况，政府必须引导公共舆论的发展方向。

接下来考虑政府在采取防御措施情况下的观点演化，在这种场景下，社会网络包含 2 个嫌疑人、2 个防御者和 4 个普通个体，其中防御者传播的积极观点为 $\tilde{\boldsymbol{u}} = [75, 65]^{\mathrm{T}}$ 。如图 3(a) 和 3(b) 所示，群体观点偏离了嫌疑人传播的恶意观点，却没有都收敛到积极观点，但这已经避免了舆论危机。如图 3(c) 和 3(d) 所示，在 $k = 1$ 时，安全距离首次大于危险距离，这说明在 $k = 1$ 时防御者的观点更新模式发生了切换，防御者开始传播积极观点，这对应着图 3(a) 和 3(b) 中个体 3 和个体 4 的观点更新。防御者观点更新模式切换之后，个体 3 和个体 4 的观点始终保持不变，并且普通公众的观点都向防御者传播的积极观点靠拢，这使得舆论朝着积极的方向发展。嫌疑人的观点也偏离了恶意观点，这是受到其他个体观点的影响。

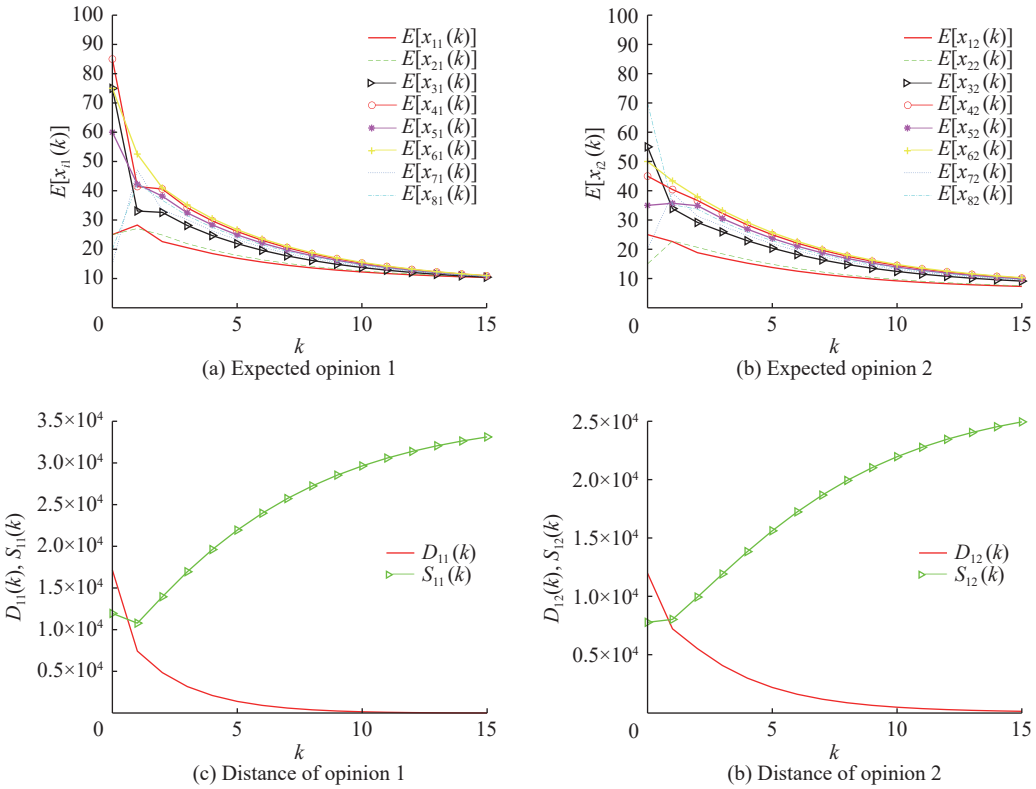
进一步分析政府引导下的观点演化过程。如图 4 所示，嫌疑人的观点会出现持续的跳变，其他个体的观点更新相对稳定，这可以作为检测社会网络中嫌疑人的一个依据。

定义群体的平均观点（ $\bar{\boldsymbol{x}}(k)$ ）为群体观点在时间轴上的平均值，即

$$\bar{\boldsymbol{x}}(k) = \frac{1}{k+1}\sum_{h=0}^{k}\boldsymbol{x}(h) \tag{22}$$

图 5 所示为个体平均观点轨迹。由图 5 可见，群体平均观点演化更加平滑，且随着 k 增加，个体的平均观点接近图 3(a) 和 3(b) 中的期望观点。

由于现实中的社会网络通常包含大量个体，考虑一个由 500 个成员组成的社会网络，其中有 20 个嫌疑人，嫌疑人传播的恶意观点为 $\boldsymbol{u} = [10,5]^{\mathrm{T}}$ ，且嫌疑人传播恶意观点的概率为 $\boldsymbol{p}_i = 0.3$ ，$i \in J$ 。当网络中没有防御者时，观点更新如图 6 所示，所有个体的观点都被恶意观点主导。将网络中 20 个普通个体替换为防御者时，防御者传播的积极观点为 $\tilde{\boldsymbol{u}} = [75, 65]^{\mathrm{T}}$ ，观点更新如图 7 所示，嫌疑人的观点偏离了恶意观点，普通个体的观点介于积极观点和嫌疑人的观点之间，防御者阻止了舆论危机的发生。



(a) Expected opinion 1　(b) Expected opinion 2

(c) Distance of opinion 1　(b) Distance of opinion 2

图 2　无政府介入时个体观点演化

Fig. 2　Evolution of agent opinion without actions of the government

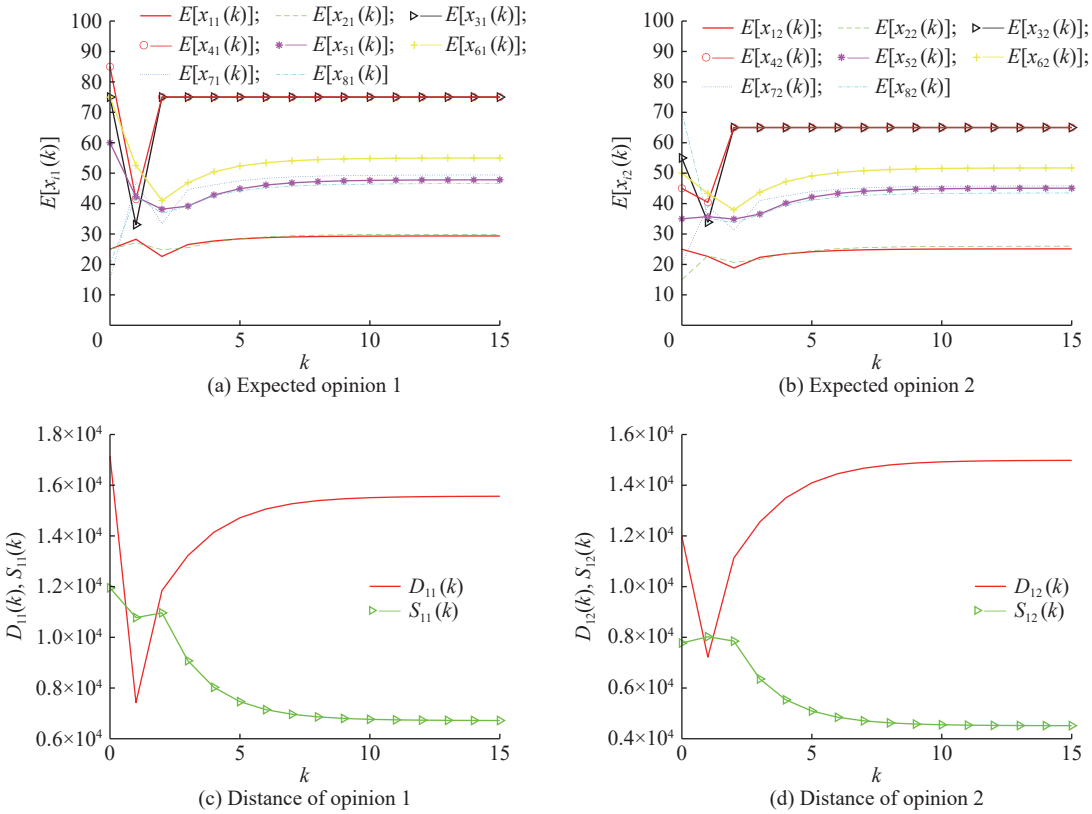


图 3　政府介入时个体观点演化

Fig. 3　Evolution of agent opinion with actions of the government

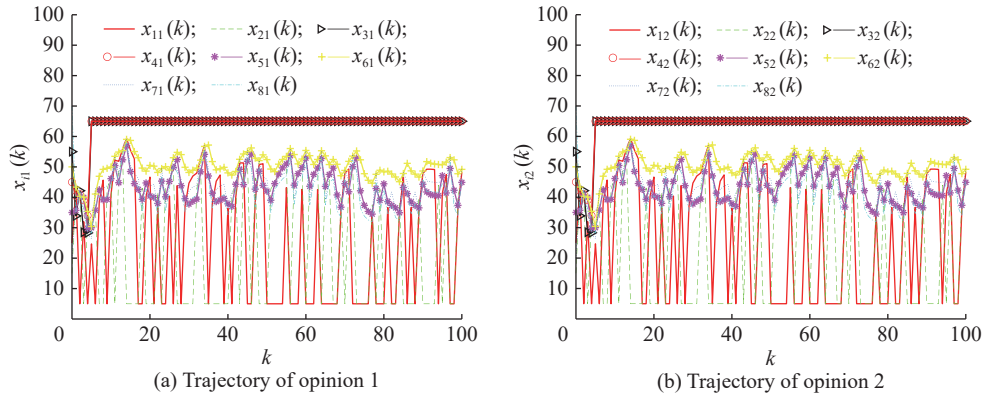


图 4　个体观点轨迹

Fig. 4　Trajectory of opinion of each agent

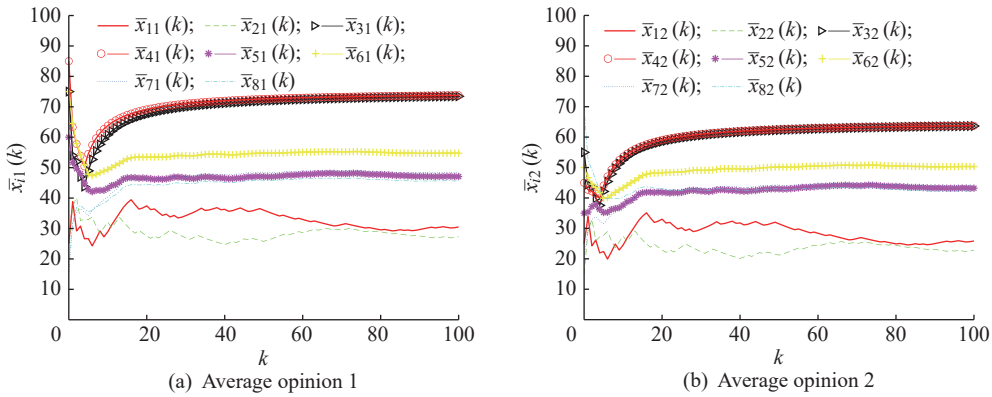
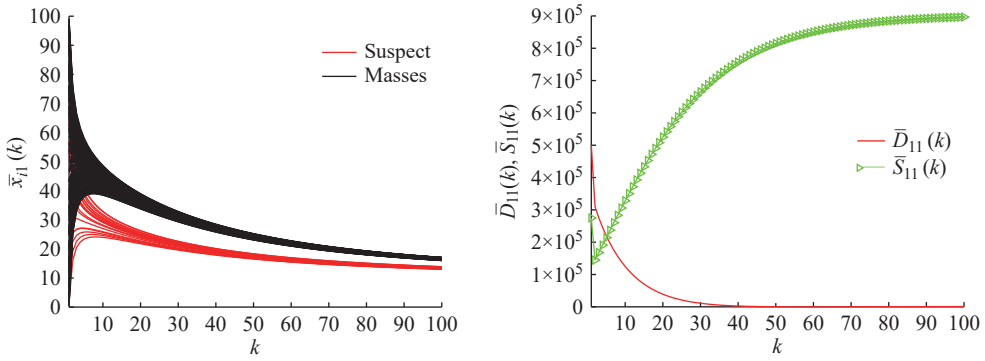


图 5　个体平均观点轨迹

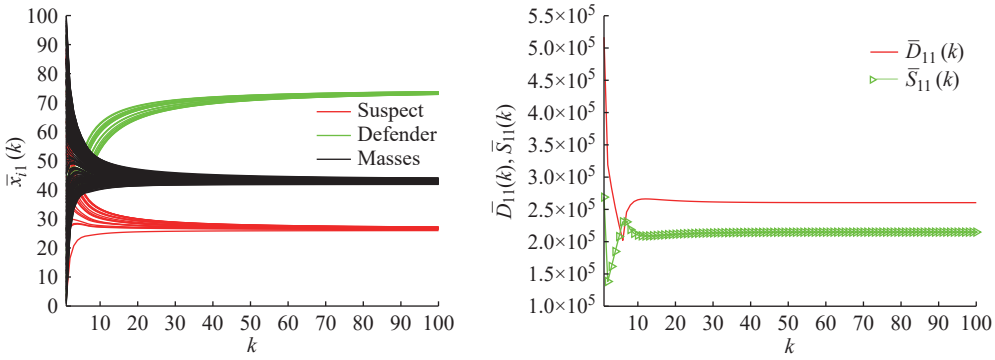
Fig. 5　Trajectory of average opinion of each agent



(a) Average opinion 1 (b) Distance of average opinion 1

图 6 无防御者时平均轨迹

Fig. 6 Trajectory of average opinion without defenders



(a) Average opinion 1 (b) Distance of average opinion 1

图 7 有防御者时平均轨迹

Fig. 7 Trajectory of average opinion with defenders

4 结 论

针对社会网络中公共舆论安全问题，提出一种带有防御者和嫌疑人的观点演化模型。为保持嫌疑人身份的隐匿性，设计嫌疑人观点随机更新策略；为阻止嫌疑人将公共舆论导向恶意观点，借助防御者向社会网络中传播积极观点。为判断公共舆论是否将出现危机，建立了结合观点的危险距离和安全距离的判据。为阻止舆论危机的发生，确定了积极观点有效阻止舆论危机的参数条件，并设计了低成本积极观点的参数计算方法。结合理论分析和实验验证，得到以下结论：

(1)当社会网络中缺少防御者时，嫌疑人通过传播恶意观点可以造成公共舆论危机，舆论危机爆发时，社会网络中所有个体都持有恶意观点；当社会网络中存在防御者时，防御者通过传播积极观点可以规避公共舆论危机的爆发，此时，可根据嫌疑人观点的异常更新模式检测出社会网络中的嫌疑人，进一步保障公共舆论的安全。

(2)嫌疑人和防御者的数量也会影响公共舆论的安全水平，嫌疑人越多时，越容易爆发公共舆论危机；防御者越多时，越能阻止公共舆论危机的爆发。

参考文献：

[1] 傅桂元. 非均匀交互观点动力学的智能体建模及分析[D]. 上海: 上海交通大学, 2015.

[2] 孔令霖, 林家骏, 周昆. 一种抗 CPS 控制层欺骗攻击的算法 [J]. 华东理工大学学报 (自然科学版), 2015, 41(2): 198-204.

[3] 汪迪, 李芳菲, 许思遥, 等. 针对信息物理系统线性欺诈攻击的水印加密策略 [J]. 华东理工大学学报 (自然科学版), 2020, 46(6): 828-832.

[4] 陈泽彬, 余昭旭, 李树刚. 网络攻击下非线性 CPSs 的事件触发自适应控制 [J]. 华东理工大学学报 (自然科学版), 2024, 50(3): 411-417.

[5] FREEMAN L. The Development of Social Network Analysis: A Study in the Sociology of Science[M]. Vancouver: Book Surge, 2004.

[6] HELBING D. QuaNtitative Sociodynamics: Stochastic Methods and Models of Social Interaction Processes[M]. Berlin: Springer, 1991.

[7] WEIDLICH W. Sociodynamics: A Systematic Approach to Mathematical Modeling in Social Sciences[M]. London:

Taylor and Francis, 2002.
[8] GRABOWSKI A, KOSINSKI R A. Ising-based model of opinion formation in a complex network of interpersonal interactions[J]. Physica A: Statistical Mechanics and Its Applications, 2006, 361(2): 651-664.
[9] GALAM S. Minority opinion spreading in random geometry[J]. European Physical Journal B, 2002, 25: 403-406.
[10] CLIFFORD P, SUDBURY A. A model for spatial conflict[J]. Biometrika, 1973, 60: 581-588.
[11] SZNAJD-WERON K, SZNAJD J. Opinion evolution in closed community[J]. International Journal of Modern Physics C, 2000, 11(6): 1157-1165.
[12] DEGROOT M H. Reaching a consensus[J]. Journal of the American Statistical Association, 1974, 69(345): 118-121.
[13] FRIEDKIN N E, JOHNSEN E C. Social influence and opinions[J]. Journal of Mathematical Sociology, 1990, 15(3/4): 193-206.
[14] ALTAFINI C. Consensus problems on networks with antagonistic interactions[J]. IEEE Transactions on Automatic Control, 2013, 58(4): 935-946.
[15] EASLEY D, KLEINBERG J. Networks Crowds and Markets: Reasoning About A Highly Connected World[M]. Cambridge: Cambridge University, 2010.
[16] BORKAR V S, REIFFERS-MASSON A. Opinion shaping in social networks using reinforcement learning[J]. IEEE Transactions on Control of Network Systems, 2022, 9(3): 1305-1316.
[17] HE G, CI Z Q, WU X T, *et al*. Opinion dynamics with antagonistic relationship and multiple interdependent topics[J]. IEEE Access, 2022, 10: 31595-31606.
[18] PARSEGOV S, PROSKURNIKOV A, TEMPO R, *et al*. A new model of opinion dynamics for social actors with multiple interdependent attitudes and prejudices[C]// Proceedings of IEEE Conference on Decision and Control (CDC). Osaka, Japan: IEEE, 2015: 3475-3480.
[19] DEFFUANT G, NEAU D, AMBLARD F, *et al*. Mixing beliefs among interacting agents[J]. Advances in Complex Systems, 2001, 3: 87-98.
[20] HEGSELMANN R, KRAUSE U. Opinion dynamics and bounded confidence models, analysis, and simulations[J]. Journal of Artificial Societies and Social Simulation, 2002, 5(3): 96-104.

Evolution and Defense of Public Opinion in a Class of Social Networks with Malicious Opinions

DING Wenjie, YANG Wen

(Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China)

Abstract: A social network opinion model is proposed to address a type of public opinion security issue in social networks, which includes suspects, defenders, and ordinary individuals. Suspects and defenders are modeled as network nodes with external inputs. Suspects spread malicious opinions to other individuals, while defenders spread positive opinion to other individuals. The safety of public opinion is judged by comparing the distance between the expected opinions of the group and the malicious and positive opinions. Based on this, the parameter conditions that should be met to maintain the safety of public opinion are determined. Furthermore, the sufficient conditions for the convergence of expected opinions of social groups are analyzed, and a calculation method for designing low-cost, effective positive opinion is designed. Based on theoretical analysis and experimental verification, the following conclusions are obtained. If there is no defender in social networks, suspects can cause public opinion crises by spreading malicious opinion. When a public opinion crisis erupts, all individuals in the social network hold malicious opinion. If there are defenders in social networks, they can prevent public opinion crisis by spreading positive opinion. In this case, suspects in the social network can be detected according to abnormal update patterns of their opinions, further ensuring the safety of public opinion. In addition, it can be found that the number of suspects and defenders also affects the security level of public opinion. The more suspects, the greater the likelihood of triggering a public opinion crisis. The more defenders, the more likely to prevent the outbreak of public opinion crisis.

Key words: complex network; public opinion; security; social network; consensus; opinion dynamics

（责任编辑：张欣）