

基于斯塔克尔伯格博弈的在线社交网络扭曲信息干预算法



袁得崙^{1,2} 陈世聪¹ 高 见^{1,2} 王小娟³

1 中国人民公安大学警务信息工程与网络安全学院 北京 100038

2 安全防范与风险评估公安部重点实验室 北京 100038

3 北京邮电大学电子工程学院 北京 100876

(yuandeyu@ppsuc.edu.cn)

摘 要 在新冠肺炎疫情期间,社交媒体以前所未有的速度向全世界传播消息。然而,扭曲信息隐藏在海量社交数据中,对国家安全、社会稳定提出了前所未有的挑战。目前的干预措施大多是在对关键节点和关键链路进行控制的基础之上,即删帖和封号,往往效果不佳且容易产生副作用。基于扭曲信息的定义和分析,打破传统思维的限定,在信息蔓延过程中通过发布辟谣信息来干扰扭曲信息的演化过程。借助斯塔克尔伯格博弈理论,文中通过设置奖励来鼓励更多的社交网络用户参与信息对冲过程,从而阻止扭曲信息的爆发效应。基于所提出的斯塔克尔伯格博弈模型,分析了斯塔克尔伯格博弈均衡解的存在性和唯一性,并从理论上推导出斯塔克尔伯格博弈的闭式均衡解,提出了基于最优策略的扭曲信息干预算法。实际网络中的仿真实验表明,相比传统的基于网络结构的免疫策略以及其他基于博弈论的干预算法,所提算法最高可将扭曲信息的传播范围分别降低41%和9%,因此能够有效抑制扭曲信息的传播。

关键词: 在线社交网络;扭曲信息;斯塔克尔伯格博弈;信息传播;逆向干预

中图法分类号 TP309

Intervention Algorithm for Distorted Information in Online Social Networks Based on Stackelberg Game

YUAN De-yu^{1,2}, CHEN Shi-cong¹, GAO Jian^{1,2} and WANG Xiao-juan³

1 Department of Police Information Engineering and Cyber Security, People's Public Security University of China, Beijing 100038, China

2 Key Laboratory of Safety Precautions and Risk Assessment, Ministry of Public Security, Beijing 100038, China

3 School of Electronic and Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract During the 2019-nCoV epidemic, social media spread news around the world at an unprecedented rate. Distorted information is hidden in massive social data, which presents unprecedented challenges to national security and social stability. Most of the current intervention strategies are based on the control of key nodes and key links, that is, deleting tweets and blocking accounts, which are often ineffective and prone to side effects. Based on the definition and analysis of distorted information, this paper breaks the limitation of traditional thinking and disturbs the evolution of distorted information by publishing clarifications during the spread of distorted information. With the help of Stackelberg game theory, more users are encouraged to participate in the information hedging process by setting up rewards, thereby hindering the explosive effect of distorted information. Based on the established Stackelberg game, the existence and uniqueness of the equilibrium solution are analyzed, and the closed equilibrium solution and information intervention algorithm is proposed. Simulation experiments in the actual network show that the proposed algorithm reduces the spread of distorted information by up to 41% and 9% compared to the traditional immune strategy based on network structure and other intervention algorithms based on game theory, thus can effectively suppress the spread of distorted information.

Keywords Online social networks, Distorted information, Stackelberg game, Information dissemination, Reverse intervention

到稿日期:2020-04-20 返修日期:2020-07-29 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61771072);中国人民公安大学专项项目(2020JWCX01);警务物联网应用技术公安部重点实验室开放课题

This work was supported by the National Natural Science Foundation of China(61771072), Special Project of People's Public Security University of China(2020JWCX01) and Open Project of the Key Laboratory of the Police Internet of Things Application Technology (Ministry of Public Security).

通信作者:高见(gaojian@ppsuc.edu.cn)

1 引言

当前,世界各国仍在同新冠肺炎疫情作斗争,社交媒体以前所未有的速度向全世界传播信息,在带给人们希望的同时也加剧了恐慌和种族歧视。互联网上的在线社交媒体(如 Twitter、Facebook、微信、微博、MySpace、抖音、YouTube 等)构成了在线社交网络(Online Social Network, OSN)。OSN 不仅折射着现实社会,同时也深刻地影响着现实社会。2020 年 2 月 2 日,世界卫生组织将新型冠状病毒疫情称为大规模的信息疫情。OSN 使虚假信息以前所未有的速度迅速传播,造成了不确定性,加剧了人们的焦虑和种族歧视。尽管世卫组织尝试通过与 Twitter、Facebook、腾讯、TikTok 合作来遏制虚假信息,但是由于虚假信息泛滥,即使不遗余力也无法清除所有的杂音,反过来却为仇中势力创造了温床,种族主义和诽谤事件在 TikTok 和 Facebook 上激增。

目前“假新闻”“谣言”等术语过分关注“真实”与“虚假”,显得过于简单化。复杂的社交网络变迁使人们更容易受到有害信息和阴谋的影响,而扭曲信息则是其中的典型,且明显带有“误导”的影子。扭曲信息隐藏在海量社交数据中,使得网络成为一把双刃剑,虚拟的环境对国家安全、社会稳定提出了前所未有的挑战。“谣言一张嘴,辟谣跑断腿”,扭曲信息快速蔓延进而破坏现实格局所带来的“信息安全”问题是学术界和产业界亟需解决的问题。因此,通过干预控制来引导网络信息传播秩序具有重要的理论和现实意义。

类似于疫情防控中的封城和隔离措施,目前绝大部分针对扭曲信息的干预措施都是建立在对关键节点和关键链路进行控制的基础之上,也就是删帖和封号。然而,扭曲信息的传播网络通常是去中心化的,因此无法找到一个核心节点来一劳永逸地解决问题。无标度特性使得在线社交网络更具嵌套性,而嵌套性的网络在遭到外部打击时能迅速复原。删帖、封号反而会产生更大的报复性副作用,在不同平台之间,当一个平台将某一主题的扭曲信息取缔后,该主题对应的用户可以迁移到其他主题上,往往还会使用其他的文字或“梗”对扭曲信息进行“加密”,这使得新信息更难被发现。如果只是根据常识,而没有注意到扭曲信息传播的特征,简单粗暴地屏蔽只会使得扭曲信息如野草一样烧不尽。

对于扭曲信息的发布者来说,选择合理的策略可以使信息快速蔓延,达到攻击的目的。然而,在现实中系统是开放的,从理论角度来看,网络监管者也可以在信息蔓延过程中通过加注扰动变量来实施干扰;从实际角度来说,可以通过发布辟谣信息、将相互矛盾的扭曲信息同时曝光给用户等方式,使得扭曲信息和人为扰动形成不同的信息对冲,从而阻止信息的快速蔓延直至消亡。

本文将打破删帖、封号这种传统思维的限定,通过发布辟谣消息的方式来扰动扭曲信息的演化过程。同时,为了借助在线社交网络中广大用户的力量,本文将问题建模为斯塔克尔伯格博弈模型,网络监管者通过设置奖励来鼓励广大用户参与信息对冲与逆向干预过程,从而阻止扭曲信息的爆发效应。

2 相关工作

2.1 面向 OSN 的扭曲信息控制策略

当前,针对 OSN 的控制策略已有一定的研究,Wang 等^[1]通过分析用户的网络地位、社会强化效应以及用户感知价值等影响信息传播控制的关键因素,构建了基于用户相对权重的网络舆情传播控制模型,并分析了 OSN 中舆情传播与控制的干预时机选择问题。Zuo 等^[2]分析了社交网络中有关首次代币发行(Initial Coin Offering, ICO)的舆情传播特征,通过引入外部舆论场对于舆情传播的影响,提出了小世界网络中舆情的传播与控制模型。仿真实验显示,在舆情爆发初期,通过技术手段来控制传播范围可以达到较好的控制效果。Peng 等^[3]基于影响力模型提出了一种利用大数据进行干预的方法。该方法采用广度优先搜索策略来构造影响力传播树,并通过遍历传播树来衡量个体的间接影响力。通过选取网络中最具影响力的前 K 个节点,提出了一种新的免疫算法。Shang 等^[4]提出了一种基于社区结构和阈值的动态节点免疫模型,该模型将携带有害信息的节点视为网络中的新节点,并通过使用社区结构信息和节点之间的相似度函数来确定节点之间的传播概率。在 8 个真实网络上的仿真实验表明,该方法能够实现更好的免疫效果。

已有的控制策略大多是在建立在对关键节点和关键链路进行评估的基础之上的删帖和封号^[5-6]。如 Yang 等^[7]借助影响力最小化目标函数来寻找 K 个最优阻塞节点,最终通过阻塞 K 个未感染节点来最小化最终受感染节点的规模。也有研究人员分析了网络抗毁性对社交网络控制的影响,如文献^[8]基于渗流理论和随机动力学方法,提出了一种预测社交网络系统变化的新方法,并分析了社交网络密度对渗透阈值的影响。

从网络结构的角度看,删帖和封号的确可以达到阻碍扭曲信息传播的目的,并为管理决策提供一定的政策建议^[9]。然而,近期发表在 *Nature* 上的一篇文章表明,在扭曲信息传播的小世界网络中,关键词屏蔽、封号等传统监管方式不仅效果不佳,而且可能引发报复性的反作用^[10]。因此,亟需创新性的解决方案,如在封号和删帖的基础上发布辟谣消息,或者通过曝光相互矛盾的扭曲信息的方式来扰动扭曲信息的演化过程。

2.2 针对攻防博弈的逆向干预理论

在社交网络中,信息最初从一组种子用户传播,然后以连接关系从感染用户传播到未感染用户。扭曲信息与其他信息一起传播,并且持相反意见的信息会相互竞争。在现实世界中,收到辟谣信息的用户将不再接受扭曲信息。然而,如果扭曲信息捷足先登,可能会立即形成爆发性的舆情态势,因此,后到达的辟谣信息所起的作用将大打折扣。例如,当受到 Twitter 上的猪流感信息影响时,人们可能在收到世界卫生组织的澄清消息之前就误打了疫苗。在这个框架的推动下,相关研究提出了基于博弈扩散模型的扭曲信息逆向控制问题。如 Xiao 等^[11]在充分考虑谣言传播特征的复杂性以及表示学习在数据特征提取中的优势的基础上,分析了谣言和辟谣信息的群体行为模式。借助进化博弈理论,Xiao 等分析了谣言

和辟谣信息之间的冲突和共生关系,得出了谣言和辟谣信息对用户影响程度的结构特征表达。Tong 等^[12]提出了一种基于博弈论的随机逼近算法,该算法在运行时间方面优于现有方法。Srivastava 等^[13]提出了一种基于网络的主动虚假新闻纠正招募机制(Recruitment for Active Fake News Correction, RecAnt),该机制通过寻找给定数量的个体并激励他们来进行新闻检查和传递真实新闻,从而达到抑制虚假新闻传播的目的。Yu 等^[14]设计了一种地理社交网络中重叠种子的广告博弈决策机制。Shan 等^[15]在分析转发双方收益的基础上,提出了一种基于博弈论的 OSN 转发控制机制,该方法能够有效地阻止转发者的非诚信转发行为。

通过发布辟谣消息或者曝光相互矛盾的扭曲信息的方式来扰动扭曲信息,同样需要分析消息之间的博弈过程以及消息的扩散过程^[16]。已有学者基于博弈论提出了信息扩散模型。如 Lu 等^[17]研究了选举网络中 K 名候选人的竞争影响传播过程。他们提出了一种新的平均模型,分析了在生成随机块模型(Stochastic Block Model, SBM)的随机网络中,竞争者需要在选民中安置多少种子用户才能赢得最终的选举。Liu 等^[18]结合实际谣言传播过程中存在的从众效应,借助博弈论建立了谣言传播模型。Li 等^[19]通过分析涉及某一特定主题的多个消息在扩散过程中的相互作用以及竞争关系,提出了一种能够反映内部竞争机制的信息扩散模型。Kaligotla 等^[20]基于主体的建模框架,分析了竞争性谣言传播中的主体行为,并尝试通过互动来影响同伴对于谣言的看法。此外,在扭曲信息逆向干预中,监管者旨在最小化扭曲信息的传播范围(即最小化扭曲信息激活的节点数量)。针对此类问题,He 等^[21]提出了基于博弈论的 LT 模型的 $(1-1/e)$ 近似算法。然而以上研究均未考虑监管者的作用。在现实中,网络监管者通过实施一定的措施可以产生更加积极的级联效果。

3 传播模型与问题描述

3.1 扭曲信息的定义

为了理解和研究信息生态系统的复杂性,本文首先给出扭曲信息的定义。

定义 1 扭曲信息是指在 OSN 中有意或者无意传播的能够制造混乱、产生隔阂的内容,包括错误信息(Misinformation)、虚假信息(Disinformation)以及恶意信息(Malinformation) 3 种类型。

错误信息是指有误导性的错误内容。虚假信息是故意捏造并有意造成损害的内容。恶意信息则为有意造成政治、经济影响的极端言论。

3.2 信息传播模型

用有向图 $G=(V, E)$ 表示在线社交网络,其中 V 为网络中的节点集合, E 为网络中边的集合。由边连接的两个节点称为邻居(即存在关注关系)。在某一时刻,在线社交网络中的攻击者发布扭曲信息 m ,网络中的其他节点收到该消息 m 并选择是否进行转发(感染)。

现有的社区网络中的信息传播模型可以分为以下两类:流行病传染模型(如 SI, SIR 和 SIS 模型)以及影响力扩散模型(如 IC, LT 模型)。本文采用影响力扩散模型,即独立级联

(Independence Cascade, IC)和线性阈值(Linear Threshold, LT)模型。

(1)IC 模型:一个感染节点 v 以独立的概率 $p_{v,w}$ 感染其邻居节点 w ,并且一个感染节点只有一次机会感染其邻居节点。

(2)LT 模型:未感染节点 w 是否采纳扭曲信息取决于其所有邻居节点 $v \in N(w)$ 的影响权重 $b_{v,w}$ 之和 $\sum_{v \in N(w)} b_{v,w}$ 。在初始状态,每个节点独立地选择一个阈值 $\theta_v \in [0, 1]$ 。在 t 时刻,前一时刻感染的节点保持感染状态,且当未感染节点 w 的所有感染邻居节点的权重之和大于或等于 θ_v ,即 $\sum_{v \in N(w)} b_{v,w} \geq \theta_v$ 时,节点 v 被感染。

给定在线社交网络和信息传播模型(本文选择 LT 模型)。根据扭曲信息 m 的传播态势,网络监管者的目标是通过发布辟谣信息来控制扭曲信息的传播范围,以便在 T 时间窗口之后将整个网络上的扭曲信息感染率控制在一定范围内。假设辟谣信息也是通过 LT 模型在网络中传播,节点 v 对于辟谣信息的接受阈值为 $\sigma_v \in [0, 1]$ 。由于辟谣信息的可信性通常高于扭曲信息,收到辟谣信息的用户将大概率不再接受扭曲信息,因此本文假设 σ_v 远小于 θ_v 。

3.3 问题描述

扭曲信息的逆向干预与社交网络影响力扩散模型密切相关。扭曲信息被视为与辟谣信息一起传播,并且持有相反意见的信息会相互竞争。在现实世界中,收到辟谣信息的用户将不再接受扭曲信息。为了防止人们被扭曲信息误导,一种自然的方法是引入辟谣消息,使其在扭曲信息到来之前尽快传播到未感染用户。一旦检测到扭曲信息,网络监管者(如社交平台)就可以通过鼓励更多的用户参与(转发辟谣信息)来产生具有竞争性的积极级联效果,从而最小化被扭曲信息感染的用户数量。

因此,本文将在在线社交网络中扭曲信息的干预问题建模为斯塔克尔伯格博弈问题。在斯塔克尔伯格博弈模型中,存在一个领导者和若干追随者,并假设所有的参与者都是理性和自私的,即博弈的参与者只关心自身效益的最大化。

在本文的斯塔克尔伯格博弈模型中,舆情的监管者作为领导者率先行动,网络中的参与用户作为追随者根据领导者的策略随后做出决策。在斯塔克尔伯格博弈的第一阶段,舆情监管者作为领导者,根据当前扭曲信息所造成的影响来评估损失,并根据用户对辟谣信息的参与情况来决定总奖励 W ,以最大化自身效益。在斯塔克尔伯格博弈的第二阶段,网络中的用户作为追随者,根据总奖励情况决定对辟谣过程的参与度来获得奖励。基于以上分析,斯塔克尔伯格博弈每个阶段的策略都会影响其他阶段的决策,领导者做决定时需要同时考虑该决策可能对追随者造成的影响。下面将使用后向推导的方法详细描述该博弈过程。

假设舆情监管者通过向网络中的参与用户支付奖励的方式来达到借助用户协助对扭曲信息进行干预的目的。舆情监管者根据用户的投入情况(通过用户对辟谣信息的转发次数、投入时间等来衡量)将奖励分成不同份额分配给各个参与用户,投入更多的用户将获得更多的报酬。本文假设参与用户获得的报酬与其投入占有所有参与用户投入的总和的比例成正

比。因此,参与用户 n 获得的回报 w_n 为总报酬 W 乘以用户 n 的投入占有所有用户投入总和的比例,即:

$$w_n = \frac{c_n}{\sum_{n=1}^N c_n} W \quad (1)$$

各参与用户以合理的投入下最大化其收益为目标。因此,参与用户 n 的效用函数可以被定义为获得的报酬减去为获得报酬所投入的成本,如式(2)所示:

$$U_n = w_n - a_n c_n \quad (2)$$

其中, a_n 相当于用户 n 单位投入所消耗的收益。因此,参与用户 n 通过求解以下的效用最优问题来求得其最优策略。

最优问题 1:

$$\max_{c_n \geq 0} \sum_{n=1}^N c_n \frac{c_n}{\sum_{n=1}^N c_n} W - a_n c_n \quad (3)$$

定义舆情监管者的效用函数为用户参与所带来的干预效益(减小的扭曲信息传播范围) $f(c_n)$ 减去所支出的奖励:

$$U_M = \sum_{n=1}^N x_n f(c_n) - W \quad (4)$$

其中, x_n 为用户 n 单位投入所带来的干预增益, x_n 与用户的中心性与可信度密切相关。在参与用户通过转发辟谣消息、发布对冲信息来抑制扭曲信息传播的过程中,随着参与用户投入的增加,干预效果的提升起初相对较快,之后将逐渐放缓(例如,在参与用户转发辟谣消息的过程中,随着投入的增加,未接收过辟谣消息的邻居节点比例将会越来越低,已接收过辟谣消息的邻居节点对于该类信息逐渐失去兴趣),这非常符合对数函数的特点。因此,本文选择对数函数作为 f 函数。相应地,舆情监管者的效用函数则变为:

$$U_M = \sum_{n=1}^N x_n \log(c_n) - W \quad (5)$$

对于舆情监管者来说,其目标是通过为参与用户分配奖励来最大化其自身效用。同时,舆情监管者也需要考虑扭曲信息爆发可能造成的严重的经济、政治损失,因此监管者将不计代价以确保扭曲信息的传播始终在爆发阈值之下。因此,舆情监管者的效用最优问题如下。

最优问题 2:

$$\max_W \sum_{n=1}^N x_n \log(c_n(W)) - W \quad (6)$$

最优问题 1 和 2 组成了本文的斯塔克尔伯格博弈模型。斯塔克尔伯格博弈的目标是找到斯塔克尔伯格博弈均衡解(Stackerberg Equilibrium, SE),在该均衡点,无论是领导者还是追随者都不会单方面地改变策略从而背离斯塔克尔伯格博弈均衡点。本文给出 SE 的定义如下。

定义 2 假设 W^* 为最优问题 2 的最优解, $\mathbf{c} = \{c_n^*, n = 1, 2, \dots, N\}$ 为最优问题 1 中每个参与用户的最优解。对于任意的 $(W, \mathbf{c})$, 其中 $W \geq 0, \mathbf{c} \geq 0$, 当且仅当满足以下不等式时, $(W^*, \mathbf{c}^*)$ 为斯塔克尔伯格博弈的 SE 解。

$$U_M(W^*, \mathbf{c}^*) \geq U_M(W, \mathbf{c}^*) \quad (7)$$

$$U_n(c_n^*, \mathbf{c}_{-n}^*, W^*) \geq U_n(c_n, \mathbf{c}_{-n}^*, W^*), \forall n \quad (8)$$

值得注意的是,在斯塔克尔伯格博弈中,必须首先确定追随者(参与用户)的最优策略,随后领导者(舆情监管者)根据追随者的最优策略确定自己的最优响应。因此,在接下来的分析中,本文将使用后向推导的方法来确定斯塔克尔伯格博弈的均衡解。

4 对斯塔克尔伯格博弈均衡解的分析

4.1 非合作时间投入选择博弈

舆情监管者根据扭曲信息的传播情况来确定总奖励 W , 参与用户通过参与辟谣过程来获得奖励,以最大化其效用,因此各个参与用户之间形成一个非合作投入选择博弈(Non-Cooperative Input Selective Game, NISG) $G = [F, \{\mathbf{C}_n\}, \{U_n(\cdot)\}]$, 其中 F 是参与用户集合, $\mathbf{C}_n$ 是参与用户 n 的策略空间, $U_n(\cdot)$ 是参与用户 n 的效用函数。 U_n 可以通过式(1)和式(2)计算求得:

$$U_n = \frac{c_n}{\sum_{n=1}^N c_n} W - a_n c_n \quad (9)$$

求 U_n 关于 c_n 的一阶和二阶导数,可以得到:

$$\frac{\partial U_n}{\partial c_n} = \frac{\sum_{i \neq n} c_i W}{(c_n + \sum_{i \neq n} c_i)^2} - a_n \quad (10)$$

$$\frac{\partial^2 U_n}{\partial c_n^2} = \frac{-2 \sum_{i \neq n} c_i W}{(c_n + \sum_{i \neq n} c_i)^3} \quad (11)$$

由式(10)可知, U_n 为凸函数(U_n 的二阶导数总是为负)。因此,为了求得最大值,当其一阶导数等于 0 时,可得对应的最优解为:

$$c_n^* = \sqrt{\frac{\sum_{i \neq n} c_i W}{a_n}} - \sum_{i \neq n} c_i \quad (12)$$

由于各个用户制定其最优策略时均要考虑其他用户的决策,当所有用户都没有单方面改变自身策略的动机时, NISG 达到纳什均衡(Nash Equilibrium, NE)。本文提出定理 1 来分析 NISG 的 NE 的唯一性。

定理 1(NE 的存在性和唯一性) NISG 存在唯一的 NE 解,且对于用户 n ,其最优策略可以由下式确定:

$$c_n^* = \left(1 - \frac{a_n(N-1)}{\sum_{n=1}^N a_n}\right) \frac{W(N-1)}{\sum_{n=1}^N a_n} \quad (13)$$

证明:为了得到 NE 解,每个用户采取式(11)给出的最优

响应策略 $c_n^* = \sqrt{\frac{\sum_{i \neq n} c_i W}{a_n}} - \sum_{i \neq n} c_i$ 。因此可以得到:

$$c_n^* = A^* (1 - \frac{a_n}{W} A^*) \quad (14)$$

其中, $A^* = \sum_{n=1}^N c_n^*$ 。对 N 个用户的最优响应策略求和,不难看出 $A^* = (N-1)W / \sum_{n=1}^N a_n$ 。将其代入式(13),可以得到:

$$c_n^* = \left(1 - \frac{a_n(N-1)}{\sum_{n=1}^N a_n}\right) \frac{W(N-1)}{\sum_{n=1}^N a_n} \quad (15)$$

根据博弈理论^[22],如果最优响应函数是正的、单调且有界的,那么 NE 是唯一的,很容易证明 c_n^* 满足上述要求。因此, NISG 存在唯一的 NE 解。

4.2 监管者总奖励决定策略

根据第一阶段博弈中各参与用户的最优策略,斯塔克尔伯格博弈的领导者(网络监管者)确定其最优策略 W 以最大化其效用。

将式(14)中的 c_n^* 代入最优化问题 2 中,可以得到:

$$\max_W \sum_{n=1}^N x_n \log \left(\left(1 - \frac{a_n(N-1)}{\sum_{n=1}^N a_n} \right) \frac{W(N-1)}{\sum_{n=1}^N a_n} \right) - W \quad (16)$$

对于以上函数求解关于 W 的一阶和二阶导数,可以得到:

$$\frac{\partial U_M}{\partial W} = \sum_{n=1}^N \frac{x_n}{W} - 1 \quad (17)$$

$$\frac{\partial^2 U_M}{\partial W^2} = - \sum_{n=1}^N \frac{x_n}{W^2} \quad (18)$$

其中, U_M 也为凸函数,令其一阶导数等于 0,可得对应的最优解为:

$$W^* = \sum_{n=1}^N x_n \quad (19)$$

即监管者根据所有参与用户的中心性以及可信度来确定总奖励。现在已经求得斯塔克尔伯格博弈的均衡解。下面提出定理 2 来证明斯塔克尔伯格博弈均衡点的存在性和唯一性。

定理 2(SE 的存在性和唯一性) 式(13)和式(19)给出的最优响应(c^* , W^*)为最优化问题 1 和 2 的斯塔克尔伯格博弈的唯一均衡解。

证明:由于式(11)确定的 NISG 均衡解为 NE 解,那么根据博弈理论,斯塔克尔伯格博弈均衡解是存在的。同时,由于上述 NISG 的每个子博弈的均衡解都是唯一的,则斯塔克尔伯格均衡也是唯一的。由此证明了斯塔克尔伯格均衡的存在性和唯一性。

5 基于斯塔克尔伯格博弈的扭曲信息逆向干预算法

基于以上关于斯塔克尔伯格博弈的分析,本文提出网络监管者以及用户决策的算法。该算法由网络监管者的奖励决策算法(Award Decision Algorithm, ADA)以及参与用户的分布式投入选择算法(Distributed Input Selection Algorithm, DISA)组成。在实际网络中,应用该算法可以使监管者和各参与用户的效用达到最优。

5.1 网络监管者的最优奖励决策

假设网络监管者已经监测到某一扭曲信息并对扭曲信息的态势做出预判(此内容不在本文的研究范畴)。由于网络监管者掌握了网络中用户的中心性、可信度以及参与程度等信息,因此,监管者可以根据当前的扭曲信息传播态势以及用户的参与情况来确定总奖励 W ,其计算公式如式(19)所示。根据博弈分析,本节给出监管者的最优奖励决策算法,如算法 1 所示。

算法 1 最优奖励决策算法

输入:每个监管者定期收集相关的扭曲信息的传播态势,以及参与用户的当前策略,并根据参与者的中心性、可信度计算 x_n

输出:监管者的最优策略 W

1. 由收集到的参与者信息,通过式(19)计算最优策略 W 。

2. 监管者将最优策略 W 发送给所有参与用户。

5.2 参与用户的分布式投入选择算法

由式(15)中最优策略闭式解可知,每个参与用户在确定自己的最优策略时,需要知道所有其他参与用户的参数,而这

通常是不现实的。但参与者可以根据其他竞争者的策略来确定自己的策略。因此,本节提出分布式的投入选择迭代算法,参与者 n 根据其他参与用户的策略来确定自己的策略 c_n 。具体算法如算法 2 所示。

算法 2 分布式投入选择算法

输入: $t=0$, 初始化所有参与用户的策略 $C(0)$, 收敛精度 $\epsilon > 0$

输出:参与用户的最优策略 $c_n^{(t)}$

1. 置 $t=t+1$ 。

2. 通过下式来计算 $c_n^{(t)}$:

$$c_n^{(t)} = \sqrt{\frac{\sum_{i \neq n} c_i^{(t-1)} \frac{W}{a_n} - \sum_{i \neq n} c_i^{(t-1)}}{a_n}}$$

3. 如果 $|C(t) - C(t-1)| < \epsilon$, 结束算法, 否则重复步骤 1 和步骤 2。

6 仿真分析

为了评估本文所提算法的有效性,使用了两个的真实数据集¹⁾。第一个数据集为包含 88 234 个边和 4 039 个节点的 ego-Facebook 数据集,该网络的平均聚类系数为 0.6055。第二个数据集为包含 34 546 个节点和 421 578 个边的 cit-HepPh 数据集,该网络的平均聚类系数为 0.2848。

我们用随机生成参数来保证算法的鲁棒性。对于任意节点 $v \in \mathbb{V}$, 在 $[0, 0.5]$ 范围内随机选择权重 $b_{v,w}$, 初始选择 $\sigma_v = 0.2$, $\theta_v = 0.8$ 。为了简便起见,假设网络中有 100 个用户参与斯塔克尔伯格博弈。对于每个参与用户,随机选择归一化参数 a_n 和 x_n 。对于每一种网络和扭曲信息传播过程,进行 500 次的模拟实验以求平均值。在每次实验的开始,我们随机挑选 5 个源节点来发布扭曲信息 m , 扭曲信息基于 LT 模型在网络中进行传播。当网络中感染扭曲消息的节点达到 5% 时,网络监管者监测到扭曲信息的存在并开始执行干预措施。

首先分析 DISA 的收敛性。假设网络中有 100 个用户参与斯塔克尔伯格博弈,在所有参与用户中随机选取 3 个用户,观察策略 c_n 的变化情况。如图 1 所示,DISA 均能在 7 次迭代以内取得收敛,说明算法具有很好的收敛性。在实际中,当网络监管者监测到扭曲信息发酵从而给出其最优策略 W 时,博弈的各方能够迅速达到斯塔克尔伯格博弈均衡。

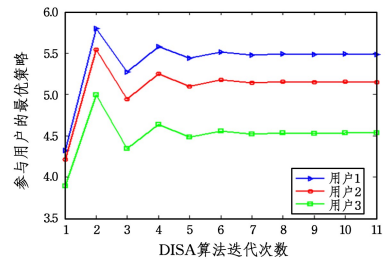


图 1 DISA 算法的收敛性

Fig. 1 Convergence of DISA

其次分析参与用户数量的变化对网络监管者以及参与用户的效用造成的影响。考虑所有参与用户的平均效用,如图 2 所示,网络监管者的效用随着参与用户数目的增加而提高,而参与用户的效用则随着用户数目的增加而降低。不难理解,对于监管者而言,随着参与用户数量的增加,将有更多的用户参与到扭曲信息的干预工作中来。而对于参与用户来

¹⁾ <http://snap.stanford.edu/data>

说,随着竞争者数量的增加,其获得的奖励将面临更多的竞争。因此,监管者为了获得更大的干预效果,应尽可能多地将奖励情况曝光给更多的用户。

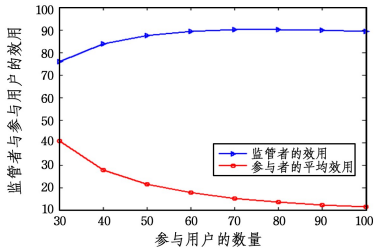


图2 效用与参与用户数量的关系
Fig. 2 Utility vs. number of users

为了说明本文所提算法对于扭曲信息的干预效果,我们选择了4种算法进行对比。其中,为了与传统的删帖和封号策略做对比,我们选择随机免疫(随机选择100个节点进行免疫)以及中心性节点免疫方法(选择具有最高入度中心性的100个节点进行免疫)进行对比。同时,选择了两种最新的基于博弈论的干预算法,即Yu等^[14]提出的LA-Greedy算法,以及Wan等^[23]最新提出的基于博弈论的干预策略(称之为Wan算法)。为了说明本文所提算法的有效性,使其与4种对比算法的参数保持一致,且假设所有节点的地理位置均相同(对标LA-Greedy算法)。本文的辟谣信息对应于Wan算法中的正向信息。当信息传播足够长的时间后,我们比较网络中感染扭曲信息的节点比例。图3给出了500次实验结果的平均值。可以看出,本文所提算法明显优于随机免疫、中心性节点免疫以及LA-Greedy 3种算法,且略好于Wan算法,说明本文所提算法取得了很好的干预效果。

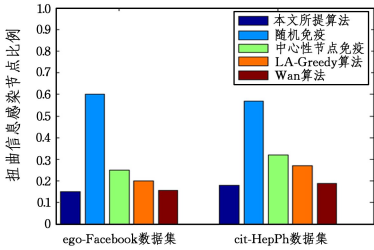


图3 不同算法的干预效果比较

Fig. 3 Comparison of intervention effects of different algorithms

此外,由于DISA采取了分布式的解决方案,本文所提算法在时间复杂度上也具有竞争力。由图4可知,本文所提算法的时间复杂度明显优于中心性节点免疫方法和LA-Greedy算法,虽然随机免疫具有最低的时间复杂度,但其干预效果最差。

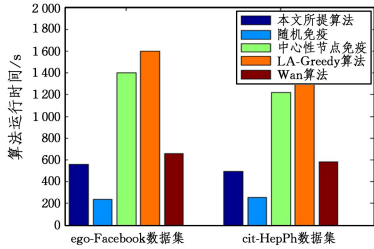


图4 不同算法的时间复杂度比较

Fig. 4 Comparison of time complexity of different algorithms

最后分析阈值 σ_v 对于本文所提算法干预效果的影响,选取 $\sigma_v \in [0, 1, 0.45]$ 。由图4可知,随着 σ_v 的增长,网络中感染扭曲信息的节点比例随之增长。这是因为 σ_v 体现了公众对辟谣信息的接受程度, σ_v 越小,公众接收到辟谣信息后越不可能再相信扭曲信息 m ,相应地,算法干预效果也越好。因此,在实际中,网络监管者要尽可能地通过官方渠道提高辟谣信息的可信度。

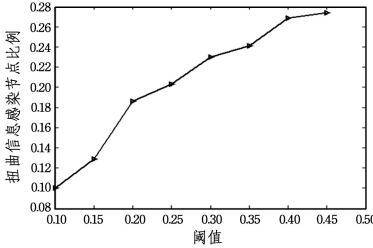


图5 阈值 σ_v 与感染节点比例的关系
Fig. 5 Relationship between threshold σ_v and proportion of infected nodes

结束语 在现实世界中,收到辟谣信息的用户通常不再关注或相信扭曲信息。然而,如果扭曲信息捷足先登,可能会立即形成爆发性的舆情态势,后到达的辟谣信息所起的作用将大打折扣。因此,本文通过设置奖励的方式鼓励更多的用户参与到辟谣信息的推广与传播中,从而与扭曲信息形成信息对冲。通过分析斯塔克尔伯格博弈均衡解的存在性和唯一性,本文推导了斯塔克尔伯格博弈的闭式均衡解,并提出了基于最优策略的扭曲信息干预算法。仿真实验表明,相比传统的基于网络结构的免疫策略以及其他基于博弈论的干预策略,本文所提算法能够有效抑制扭曲信息的传播。下一步将通过细粒度的理论分析研究用户投入(时间、转发次数)、用户中心性等对博弈过程的影响,扭曲信息传播态势与总奖励 W 之间的理论关系,以及持有相反意见的不同信息之间的竞争所构成的传播动力学。

参考文献

[1] WANG J K, YU H, WANG X H, et al. Dissemination and control model of public opinion in online social networks based on users' relative weight[J]. Systems Engineering-Theory & Practice, 2019, 39(6): 1565-1579.

[2] ZUO L M, XIA P P, HU K X, et al. Propagation model of public opinion about ICO in social network environment[J]. Computer Engineering and Design, 2019, 40(11): 3247-3253.

[3] PENG S, WANG G, ZHOU Y, et al. An Immunization Framework for Social Networks Through Big Data Based Influence Modeling[J]. IEEE Transactions on Dependable and Secure Computing, 2019, 16(6): 984-995.

[4] SHANG R H, ZHANG W T, JIAO L C, et al. Dynamic Immunization Node Model for Complex Networks Based on Community Structure and Threshold[J]. IEEE Transactions on Cybernetics, 2020(5): 1-14.

[5] XIN Y, GAO C, WANG Z, et al. Discerning influential spreaders in complex networks by accounting the spreading heterogeneity of the nodes[J]. IEEE Access, 2019, 7: 92070-92078.

[6] LIU Y, WANG X, KURTUS J. Framework of Evolutionary AI-

gorithm for Investigation of Influential Nodes in Complex Networks[J]. IEEE Transactions on Evolutionary Computation, 2019,23(6):1049-1063.

[7] YANG Y,WU C X,HE M,et al. Negative Influence Minimization Algorithm for Social Networks[J/OL]. Journal of System Simulation. <http://kns.cnki.net/kcms/detail/11.3092.V.20120120.1422.002.html>. 2019.

[8] ZHUKOV D,ZALTCMAN A G,KHVATOVA T Y. Forecasting Changes in States in Social Networks and Sentiment Security Using the Principles of Percolation Theory and Stochastic Dynamics[C]//2019 International Conference on Quality Management, Transport and Information Security, Information Technologies(IT&QM&IS). 2019:149-153.

[9] ZHENG J,PAN L. Least Cost Rumor Community Blocking optimization in Social Networks[C]// Third International Conference on Security of Smart Cities,Industrial Control System and Communications (SSIC). 2018.

[10] JOHNSON N F,LEAHY R,RESTREPO N J,et al. Hidden resilience and adaptive dynamics of the global online hate ecology[J]. Nature,2019,573:261-265.

[11] XIAO Y P,YANG Q F,SANG C Y,et al. Rumor Diffusion Model Based on Representation Learning and Anti-Rumor[J]. IEEE Transactions on Network and Service Management,2020(5):1-14.

[12] TONG G M,WU W L,GUO L,et al. An Efficient Randomized Algorithm for Rumor Blocking in Online Social Networks[J]. arXiv:1701.02368.

[13] SRIVASTAVA A,KANNAN R,CHELMIS C,et al. RecANt: Network-based Recruitment for Active Fake News Correction [C]//2019 IEEE International Conference on Big Data (Big Data). 2019:940-949.

[14] YU Y X,WANG L. An advertising game theory decision-making mechanism for overlapping seeds in geo-social network[J]. Journal of Computer Research and Development,2019,56(6):1302-1311.

[15] SHAN F F,LI H,ZHU H. Game theory based forwarding control method for social network[J]. Journal on Communications, 2018,39(3):172-180.

[16] YAO C,ZHANG Y,ZHANG X,et al. Competitive Influence Blocking in Online Social Networks:A Case Study on WeChat

[C]// 2018 24th Asia-Pacific Conference on Communications (APCC). 2018:251-256.

[17] LU P E,PENG Y H,CHANG C S,et al. Percolation Threshold for Competitive Influence in Random Networks[J/OL]. IEEE Transactions on Computational Social Systems. <https://arxiv.org/pdf/1904.05754.pdf>,2020.

[18] LIU Y Z,PAN X Z,FU W. Social Network Rumor Spreading Model Based on Bandwagon Game Between Nodes[J]. Computer Engineering,2018,44(10):309-314.

[19] LI Q,WANG Z,WU B,et al. Competition and Cooperation; Dynamical Interplay Diffusion Between Social Topic Multiple Messages in Multiplex Networks[J]. IEEE Transactions on Computational Social Systems,2019,6(3):467-478.

[20] KALIGOTLA C,YÜCESAN E,CHICK S E. An agent based model of spread of competing rumors through online interactions on social media[C]//Proceedings of the 2015 Winter Simulation Conference. 2015:3985-3996.

[21] HE Z,CAI Z,YU J,et al. Cost-efficient strategies for restraining rumor spreading in mobile social networks[J]. IEEE Trans. Veh. Technol. ,2017,66(3):2789-2800.

[22] MYERSON R B. Game theory;analysis of conflict[M]. Harvard University Press,1997.

[23] WAN P,WANG X,WANG X,et al. Intervening coupling diffusion of competitive information in online social networks[J]. IEEE Transactions on Knowledge and Data Engineering. DOI: 10.1109/TKDE. 2019.2954901.



YUAN De-yu, born in 1986, Ph.D, lecturer, Ph. D supervisor. His main research interests include cyber security and complex networks.



GAO Jian, born in 1982, Ph.D, associate professor. His main research interests include cyber security, malware and botnet.