

情报理论与实践
Information Studies: Theory & Application
ISSN 1000-7490, CN 11-1762/G3

《情报理论与实践》网络首发论文

题目：基于 CCM 的突发事件网络舆情传播建模方法研究
作者：张军，王学金，李鹏，庄云蓓
网络首发日期：2022-01-18
引用格式：张军，王学金，李鹏，庄云蓓. 基于 CCM 的突发事件网络舆情传播建模方法研究[J/OL]. 情报理论与实践.
<https://kns.cnki.net/kcms/detail/11.1762.g3.20220118.0841.002.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

●张 军, 王学金, 李 鹏, 庄云蓓 (山东理工大学管理学院, 山东 淄博 255000)

基于 CCM 的突发事件网络舆情传播建模方法研究

Research on Modeling Method of Network Public Opinion Propagation of Emergency Based on CCM

摘要: [目的/意义]在数据驱动下开展群体行为计算, 设计舆情传播建模方法, 为网络舆情管理创新提供支持。[方法/过程]结合群体行为理论和收敛交叉映射算法, 提出一种基于数据驱动的突发事件网络舆情传播建模方法。首先从群体结构、群体规范和群体过程视角提取舆情系统特征, 之后根据最邻近方法和时间序列收敛判断法识别特征之间的因果关系, 构建舆情传播模型。该方法将突发事件网络舆情中的社会强化机制计算化, 通过系统特征的因果关系强弱和差异分析, 揭示网络舆情事件的演化机理。[结果/结论]基于新浪微博平台中“双黄连事件”“白银越野赛事件”“十堰爆炸事件”和“青海地震事件”等 4 组舆情数据的实证结果发现, 各事件中的群体交互持续能力和聚集度都维持在较高的水平, 即突发事件网络舆情是典型的群体事件。此外, 不同类型的舆情事件系统特征的因果关系强弱程度不同, 因果关系越弱舆情演化不确定性越大。

关键词: 数据驱动; 突发事件; 建模方法; 收敛交叉映射算法; 因果关系

Abstract: [Purpose/significance] With data-driven, it is necessary to carry out group behavior computing, and design public opinion communication modeling method, in order to realize the innovation of network public opinion management. [Method/process] Combined with group behavior theory and convergent cross mapping algorithm, this paper proposes a data-driven modeling method for the dissemination of emergency network public opinions. Firstly, the features of public opinion system are extracted from the perspectives of group structure, group norms and group process. Then, the causal relationship between features is identified according to the nearest neighbor method and time series convergence judgment method. Finally, the public opinion propagation model is built with the strength and difference analysis of causal relationship between those system features. [Result/conclusion] Based on SinaWeibo platform, the empirical results on "Shuanghuanglian ", "Silver cross-country race ", "Shiyan explosion " and "Qinghai earthquake" reveal that the persistence and aggregation level of group interaction in each emergency is high, in other words, emergency network public opinion is a typical group event. Besides, different emergency public opinion events have different degree of causality, the weaker the causality relationship, the greater the evolve uncertainty of the public opinion event.

Keywords: data driven; emergency; modeling method; convergent cross mapping algorithm; causality

0 引言

当前, 我国正处于百年不遇大变局的关键时期, 社会经济、价值观念等正经历着全面转型和急速变化, 自然灾害、事故灾害、社会安全和公共卫生事件等各类突发事件进入高发、频发阶段。互联网和移动通信技术的普遍应用使社交媒体成为社会突发事件的第二现场, 公众在各类社交媒体平台上发布消息、表达观点, 并利用在线社会网络进行观点交互和消息扩散, 构成了突发事件的网络舆情。由于不同类型的突发事件网络舆情的引爆方式、发酵方式、传播结果各有区别^[1], 构建传播模型并挖掘它们的演化机理及差异, 对网络舆情的控制和引导至关重要。把握不同突发事件舆情传播规律, 可以为政府和组织制定合理应对策略, 为防范和化解突发事件舆情风险赢得先机。然而, 当前主流的网络舆情传播建模方法多是基于传染病模型、意见动力学模型等数理分析方法, 传播规律解析多依赖实例分析和仿真技术, 难以满足网络舆情治理的复杂需求。因此, 在大数据时代背景下应充分利用突发事件在网络上的数字痕迹, 在数据驱动研究范式下开展舆情传播建模和分析, 创新网络舆情治理的管理和决策范式^[2]。

本文为国家自然科学基金项目“社交媒体上观点、信息和行为的耦合动力学机制研究”(项目编号: G71801145)和山东省社会科学规划研究项目“数字时代社会突发事件网络舆情演化及治理研究”(项目编号: 21CTQJ06)的研究成果。

突发事件网络舆情传播研究涉及了传播学、管理学、信息科学等多个学科,如何利用各类社交媒体平台上的网络舆情数据,从中挖掘传播规律并构建传播模型一直是各个领域专家们关注的研究热点。伴随现实世界中突发事件的发展,公众会在各类社交媒体平台上发布消息、评论和转发消息,网络舆情传播的复杂性和动态性日益加剧。它不仅颠覆了传播学领域的两级传播、大众传播等经典理论,还促使研究者们寻求更严谨的方式对复杂社会问题进行抽象和表示^[3]。一方面,以海量社交媒体数据为基础,学者们尝试将社会学与传播学交叉的研究设想从概念化转变为可计算化。如利用知乎用户的点赞行为数据可以获取“群体认同”和“知晓”的程度^[4],通过文本挖掘分析微博用户交互内容构建线上用户画像、关键用户识别^[5]。但总体而言多学科视角下的计算研究处于起步阶段^[6],尤其是突发事件舆情中的群体行为计算研究还可进一步深入。另一方面,建模技术路线从利用静态截面数据和数学模型,逐渐转变为关注社会过程研究。如Watts团队结合了阈值模型与在线社会关系网络结构对消息扩散过程建模^[7],并在最近的一项工作中揭示了如何利用传播网络动态特性识别新冠肺炎虚假新闻^[8]。Porter的团队提出了利用时变网络对疾病传播、消息传播等时间特性进行表示^[9]。此外,由于网络舆情相关研究也是一类典型的复杂管理问题,以社交媒体数据为基础厘清道德、情感、群体、社会环境等因素之间的因果关系,成为网络舆情建模亟待解决的关键问题^[10],而这方面的研究还不多见。

由此,本文提出基于CCM的突发事件网络舆情传播建模方法,该方法不仅可以对底层传播数据进行解析和特征提取,还可以利用因果推断算法挖掘各个特征之间关系,自动生成可解释的传播模型。该方法重点解决以下两个问题:①如何利用突发事件网络舆情的传播记录,对突发事件网络舆情背后的群体行为相关概念开展计算,也即对舆情演化的“原因”进行量化;②如何设计算法识别和量化它们之间的因果关系,并生成突发事件网络舆情传播模型。

1 文献综述

1.1 突发事件网络舆情传播建模方法

常见的突发事件网络舆情动力学模型构建方法可以分为三类。①引入疾病传染等相关领域的动力学模型,刻画突发事件网络舆情演化规律。网络舆情传播与疾病传播有很大的区别,如消息会随时间衰减、个体的记忆历史会影响消息传播。因此,相关研究分别对SIR、SEIR等模型做出了改进和扩展^[11-13]。②系统动力学建模方法,将舆情传播问题抽象为复杂的系统性问题,从整体的角度刻画网络舆情演化的流程。建模初期需要对系统动力中的构成要素进行筛选,如邓建高^[14]、兰月新^[15-16]等总结了20余种影响因素,它极大地依赖研究者的理论分析能力和研究经验。③Agent仿真建模方法是研究复杂适应系统随时间演化的一种方法。其主要观点是在微观层面将个体信息传播行为用参数和假设进行描述,运用计算机仿真模拟可能发生的现象,根据仿真结果在宏观层面解释传播规律^[17-18]。然而,模型单次模拟运行得出的结论往往与真实的传播数据差别很大,要想复现实际的舆情传播结果,需要做大量的重复性实验排除随机的干扰。

1.2 突发事件网络舆情传播中的因果

综合上述建模方法可以看出,除了找到合适的抽象方法对传播进行表示以外,建模的目的还在于解释各种因素对传播结果的作用。并且随着互联网的兴起和实时在线数据的易得,如何从数据中找到各种因素与传播结果之间的相关性,再从中发现因果关系,成为现阶段研究中的关键问题。例如:已有的研究成果很容易就能发现政务媒体回应能力(回帖数)与传播规模之间的相关性、网民情绪与突发公共事件的信息传播倾向有相关性^[19-20],但是不能帮助我们确定政务媒体回帖及网民情绪是突发事件网络舆情传播的原因。因此若将某些突发事件网络舆情中的治理方式迁移到另外一个突发事件网络舆情中,其作用不显著^[21]。换句话说,网络舆情事件的相关性结果只能部分的解释真实舆情演化的规律^[22],并不能揭示舆情演化的原因。

当突发事件发生时社交媒体将数量巨大的用户“链接”起来,组成一个规模巨大的复杂动态系统^[23]。大规模用户在网上展示事件进程、发布求助信息,甚至组成在线社团进行紧急应对,是典型的集体行动。已有研究结果显示,在突发事件发生时,事件发生地的目击者、公众号及利益相关者会发布相关消息,成为消息源头并引领事件进展的原因之一^[24]。他们与自己的在线邻居在社会强化、社会影响等社会规范作用下进行互动和联系,构成了特殊的社会交互结构^[25],成为舆情演化的动因。随着研究深入,Homel等的研究成果还发现用户间交互的先后次序、交互持续的时间等动态关系的作用不可忽视^[26],并提出了时序网络

刻画社交媒体上“什么人在什么时间对谁说了什么”。时序网络上的节点或连边具有时间属性，可以抽象和表示个体间在线交互发生的次序、局部沟通持续的时间，能动态反应群体结构和群体过程的动态变化。由此，本文基于时序网络理论，对网络舆情系统中的群体行为特征开展计算，厘清舆情状态是在什么“原因的影响”下产生，为舆情传播建模提供基础。

1.3 收敛交叉映射原理

2007 年吉姆·格雷发表了《第四范式：数据密集型科学发现》，推动了各学科向大数据研究的第四研究范式转型。由此，通过深度学习、注意力网络等各类机器学习算法构建模型成为研究焦点^[27]。因果推断是这场数据科学革命中的一个重要分支，即从现有数据中以归纳的方式形成某种假设，然后再用算法进行演绎和验证的方式进行研究^[28]。这种方法不仅揭示了数据中蕴含的相关关系，还能从相关性中推导出因果关系。在这一领域，比较有代表性的算法有 George 等在 2012 年提出的收敛交叉映射法(Convergent Cross Mapping, CCM)，它可以识别复杂动态系统中多种元素之间的因果关系^[29]。CCM 理论基础来源于相空间重构理论：若两个变量之间存在因果联系，则两个变量在任何子空间的投影序列都会存在一定联系^[30]。即在复杂动态系统中，如果两个时间序列变量 X 和 Y 是动态耦合的，那么在 X 、 Y 之间必定存在共同且随时间变化的吸引子流形 M (Attractor Manifold)。利用变量 X 和 Y 的滞后坐标，可以重构 X 和 Y 的影子流形 M_X 和 M_Y (Shadow Manifold)。由于吸引子流形 M 中动态耦合的变量 X 和 Y ，其 M_X 附近上的点与 M_Y 上附近的点在时间上是相互对应的，所以可以通过计算 M_X 和 M_Y 在时间上相对应的精确程度，来判断 X 和 Y 之间的因果关系。最终 X 和 Y 的关系可以分为单向因果、双向因果、伪因果(相关性)及无因果关系。CCM 理论的应用比较广泛，涉及了交通拥堵与大气污染^[31]、经济政策与企业行为^[32]、全球股票市场投资风险^[33]等社会经济系统中的因果关系。因此，只要能够收集复杂动态系统中的运行参数，就可以将这些时间序列重构为若干子序列。在重构的时间序列上，挖掘系统参数的邻近性、相关性、收敛性等特征，可以用于判断参数之间的因果联系是否存在、关系强度等系统状态。最后将这些计算结果输出，就自动生成复杂系统的模型。在此背景下，舆情传播研究可以从对各类舆情事件的统计分析研究向通过推理和归纳获取传播规律进行转变。

综上所述，目前对突发事件网络舆情演化过程中的相关性统计分析多、因果性探索少，且数据驱动的建模研究还有待深入。本文从群体行为理论出发，考虑突发事件网络舆情演化中的动态交互关系结构，提出了相应的特征概念及其计算方法。改进收敛交叉映射算法分析它们之间的因果关系，并构建传播模型。本文提出的舆情传播模型是在数据驱动下由因果推断算法完成，解决了传统建模方法的经验依赖等局限性。

2 突发事件网络舆情建模方法研究

2.1 研究方法

本文按照数据获取、特征计算、关系发现、模型构建的顺序开展突发事件网络舆情建模方法研究，具体过程如图 1 所示。通过爬虫收集新浪微博某个突发事件的网络舆情数据。根据用户原创、转发、评论等交互行为构建时序网络，表示舆情系统及其动态过程。基于群体行为理论提出舆情系统特征定义，并根据时序网络理论构建计算法则。基于收敛交叉映射算法，设计突发事件网络舆情建模流程。发现和解释不同类型突发事件的网络舆情传播模型。

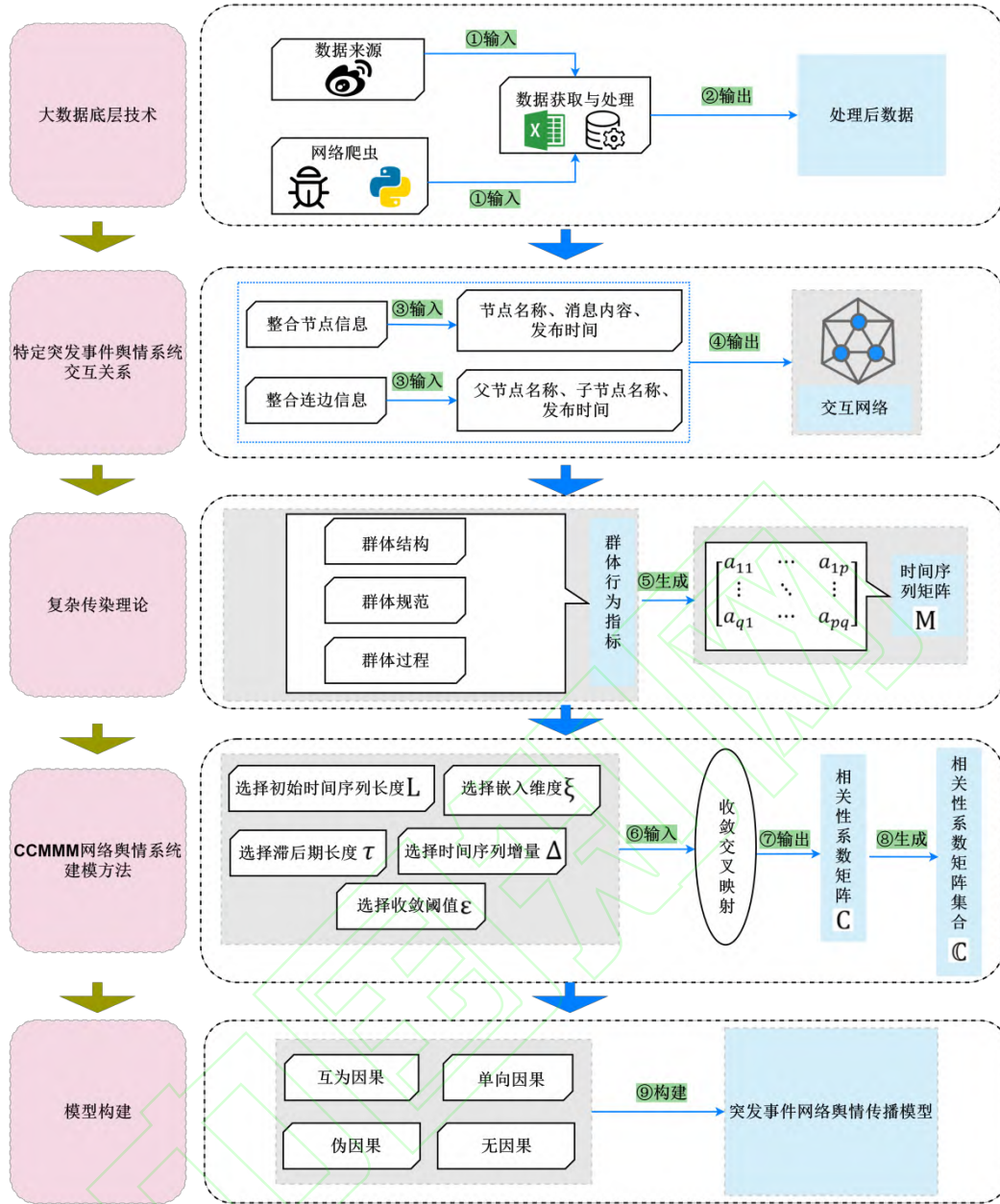


图 1 突发事件舆情传播建模流程

Fig.1 The process of modeling network public opinion propagation of emergency

2.2 突发事件网络舆情系统

某个突发事件发生时，大规模在线用户会在虚拟空间中进行群体性沟通，组成一个特定的网络舆情系统。系统中的个体可以直接转发或复制在线邻居的消息，可以使用@直接与他人对话，还可以通过#符号跟踪消息。根据个体间的交互行为，可以用时序网络 $G(V, E, T)$ 表示突发事件舆情系统^[34]。其中， $T = \{[t_0, t_1], [t_1, t_2], \dots\}$ 是网络舆情持续的时间； V 是时序网络中的节点集，表示参与舆情传播的在线个体集合； E 是时序网络中的边集，表示个体之间的交互关系集合， $\forall e \in E, e < v_i, v_j, t_l >$ 表示个体 v_i 在 t_l 时刻转发或评论 v_j 的事件消息， $t_0 \leq t_l \leq t_m$ 。基于时序网络刻画突发事件网络舆情系统，可以清晰地量化个体行为、群体互动行为及时序关系结构。

2.3 突发事件网络舆情系统特征计算

突发事件网络舆情传播是典型的群体行为涌现的结果。本文结合群体结构、群体规范和群体过程等方面对突发事件网络舆情系统特征进行刻画。

1) 网络舆情系统中的群体结构计算。当突发事件发生时, 在线群体会在共同社会规范作用下进行沟通, 以消除自己或他人对事件的不确定性和风险。根据个体参与行为模式的差别可以说明群体组织方式及群体结构^[35]。由此, 观察突发事件网络舆情传播中的个体参与行为并将他们分成三类, 具体如图 2 所示。

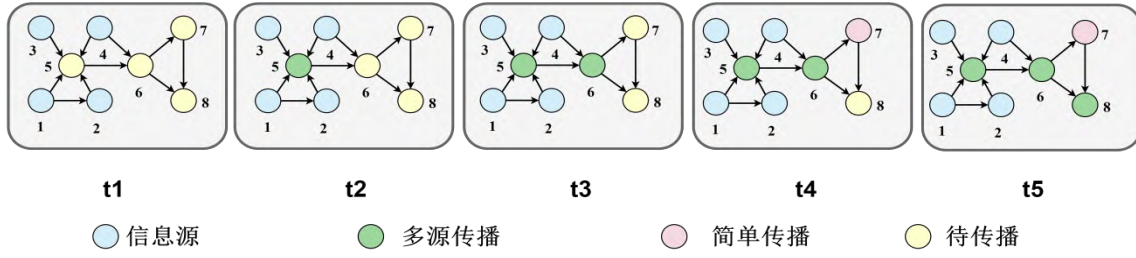


图 2 不同参与行为相互交织的舆情传播过程

Fig.2 The process of public opinion propagation constructed by different behaviors

第一类是信息源。这些个体从舆情系统外获取事件内容并在线发布, 给舆情系统提供新的内容或者观点, 节点 1、节点 2、节点 3 和节点 4 就是信息源。第二类是简单传播个体。如节点 7 在 t_3 看到一位邻居推送消息即可发生传播行为, 他们产生传播行为只需看到消息即可。第三类是多源传播个体。如节点 5、节点 6 在 t_1 、 t_2 时刻看到多个邻居已转发消息之后再参与传播。这意味着部分个体或组织要传播一条消息, 必须从至少两个以上不同的信息源处获得同一个消息, 确认消息的真实性和合法性^[36]。网络舆情传播结果由信息源规模、简单传播规模和多源传播规模三种不同的行为结果组成, 它们的具体定义和计量方式如下:

定义 1: 信息源规模是指某个特定时间段 \mathcal{T} 内, 在突发事件网络舆情系统中主动提供消息、视频等的个体数量, 记为 $O(\mathcal{T})$, $\mathcal{T} = [t_i, t_j]$ 且 $\mathcal{T} \in \mathcal{T}$ 。其中, 信息源是从舆情系统外获取事件内容并在线发布, 给舆情系统提供了新的内容或者观点。

定义 2: 简单传播规模是指某个特定时间段 \mathcal{T} 内, 突发事件网络舆情系统中发生简单传播行为的个体数量, 记为 $S(\mathcal{T})$, $\mathcal{T} = [t_i, t_j]$ 且 $\mathcal{T} \in \mathcal{T}$ 。

定义 3: 多源传播规模是指某个特定时间段 \mathcal{T} 内, 突发事件网络舆情系统中发生多源传播行为的个体数量, 记为 $C(\mathcal{T})$, $\mathcal{T} = [t_i, t_j]$ 且 $\mathcal{T} \in \mathcal{T}$ 。

2) 网络舆情系统中的群体规范计算。在各类群体事件中, 所有个体都会遵守共同行为方式的总和被称为社会规范, 是多数成员共有的行为模式^[37]。在图 2 所示的传播过程中蕴含了一个重要的群体规范——社会强化。即个体第一次接触消息并不一定发生传播行为, 需要从不同邻居处接收消息, 社会强化效应达到临界点后才会传播。一般而言, 这种社会规范在现实世界中很难精确度量, 但是利用带时间标记的网络舆情传播数据可以对此进行计算, 具体如下。

定义 4: 节点的被社会强化度是指节点参与传播前, 在线邻居对其推送消息所产生的社会强化刺激。因此, 在 t 时刻之前节点 v_i 感受到的社会强化刺激可以用已转发邻居数量标记, 记为 $K_{v_i}^t$, 如公式(1)所示:

$$K_{v_i}^t = |\mathcal{E}|, \quad \mathcal{E} = \{e | \forall e < v_i, v_j, t' > \in E, t' \leq t\} \quad (1)$$

式中, $K_{v_i}^t > 1$, 则表示节点 v_i 的传播类型是多源传播; 若 $K_{v_i}^t = 1$, 则表示节点 v_i 的传播类型是简单传播; 若 $K_{v_i}^t = 0$, 则表示节点 v_i 是信息源。

定义 5: 社会强化度是指某个特定时间段 \mathcal{T} 内, 突发事件网络舆情系统社会强化的程度, 记为 $K(\mathcal{T})$, 其计算方法如公式(2)所示:

$$K(\mathcal{T}) = \max(K_{v_i}^t) t \in \mathcal{T} \quad (2)$$

3) 网络舆情系统中的群体过程计算。在社会规范作用下群体成员发生交互, 这一过程被称为群体过程。本文从两个方面对突发事件网络舆情传播中的群体过程开展计算。一是群体过程的持续能力。已有研究表明在各类事件舆情传播中, 消息质量或新颖度会随着时间衰减, 不再引起在线用户的注意力, 因此多数舆情规模都不大^[38]。而突发事件的网络舆情持续时间都比较长, 这意味着有相当规模的个体持续参与群

体互动。二是群体交互过程中的亲密程度，反应了群体之间的交互密集情况。特别地，社会强化等社会规范要发生作用也需要在线个体周围存在一个规模适度、亲密度较大的社区^[39]。这个社区给交互提供了足够数量的桥边，让个体能看到多个在线邻居对事件消息的传播行为。因此，关于群体过程的定义和计算如下。

定义 6: 持续能力是指某个特定时间段 \mathcal{T} 内在线群体互动的保持状态，记为 $DA(\mathcal{T})$ ， $\mathcal{T} = [t_i, t_j]$ 且 $\mathcal{T} \in T$ 。其计算方法如公式(3)所示，用一段时间内信息源数量变化与参与传播数量变化的比值来度量。

$$DA(\mathcal{T}) = \frac{|V_{\mathcal{T}}|}{|O_{\mathcal{T}}|} \quad (3)$$

信息源是从舆情系统外获取事件内容并在线发布，给舆情系统提供新的内容或者观点。因此， $DA(\mathcal{T})$ 的取值可以反应群体过程是否维持在较高的水平。公式(3)中 $O_{\mathcal{T}}$ 是 \mathcal{T} 时段信息源集合， $O_{\mathcal{T}}$ 是 \mathcal{T} 时段参与传播的节点集合。

定义 7: 社会亲密度是指某个特定时间段 \mathcal{T} 内，突发事件网络舆情系统中交互关系亲密程度，用这段时间内时序网络的密度来表示，其计算方法如公式(4)所示：

$$DM(\mathcal{T}) = \frac{|E_{\mathcal{T}}|}{|V_{\mathcal{T}}|(|V_{\mathcal{T}}|-1)} \quad (4)$$

式中， $E_{\mathcal{T}}$ 是时序网络的边集， $V_{\mathcal{T}}$ 是时序网络的节点集。

4) 基于因果关系的网络舆情传播建模方法。基于数据驱动构建传播模型的关键是通过机器学习算法，对观测特征间的因果关系做出合理的判断。文章根据收敛交叉映射原理，提出 CCMMM 方法(Convergent Cross Mapping Modeling Method)，输入网络舆情系统特征取值，完成特征间因果关系识别，并输出相应传播模型。

在收敛交叉映射算法基础上，本文构建了 CCMMM 网络舆情系统建模方法。首先输入网络舆情系统特征取值时间序列矩阵 M ，循环生成不同时间尺度下的影子流矩阵 Q 。在不同的时间尺度上，对任意的特征 i 按照最邻近算法生成邻近向量集合 $Q_{i\tau}$ ，估计特征 i 和 j 之间的交叉映射估计 $\widehat{Q_{i\tau}}|Q_{j\tau}$ 、 $\widehat{Q_{j\tau}}|Q_{i\tau}$ ，重新计算估计值和真实值的相关性。综合不同时间尺度上的相关性，计算相关性的收敛性 σ_{ij} 。将满足收敛性的特征对输出，生成因果关系矩阵及动力学模型。其核心算法步骤如下：

输入：网络舆情系统特征时间序列矩阵 M 。设置初始时间序列长度 L 、嵌入维度 ξ 、滞后期 τ 、时间序列增加数量 Δ 及收敛阈值 ε 。

$$M = \begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{q1} & \cdots & a_{pq} \end{bmatrix}, a_{ik} \text{ 是第 } i \text{ 个参数在第 } k \text{ 个时刻取值, } a_{ii} = 0.$$

在第 N 步：

$$\textcircled{1} \text{ 构建长度为 } L + (N - 1)\Delta \text{ 的矩阵 } M_{L+(N-1)\Delta} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,L+(N-1)\Delta} \\ \vdots & \ddots & \vdots \\ a_{q,1} & \cdots & a_{p,L+(N-1)\Delta} \end{bmatrix}, \text{ 在此基础上，构建系统影子}$$

流矩阵 $Q_{L+(N-1)\Delta} = \begin{bmatrix} Q_{1\tau} \\ \cdots \\ Q_{p\tau} \end{bmatrix}$ ，其中 $Q_{i\tau} = \begin{bmatrix} q_{i1} \\ q_{i2} \end{bmatrix}$ ，向量 $q_{il} = \langle a_{i,l}, a_{i,(l+\tau)}, a_{i,(l+2\tau)}, \cdots, a_{i,(l+(E-1)\tau)} \rangle$ ， $Q_{i\tau}$ 的行数为

为 $\left\lfloor \frac{L+(N-1)\Delta}{\tau} \right\rfloor + 1$ 。

②将 $Q_{i\tau}$ 随机分成 $Q_{i\tau}^{train}$ 和 $Q_{i\tau}^{test}$ ，并计算其中任意两行向量的距离，找到距离最近的 $\xi + 1$ 个向量，构成集合 $\underline{Q_{i\tau}} = \begin{bmatrix} q_{i,1} \\ \cdots \\ q_{i,\xi+1} \end{bmatrix}$ 。构建第 i 个参数的权重矩阵 w_i ，其中的第 m 个向量对应的权重计算方式如公式(5)、公式(6)所示：

$$w_{i,m} = \frac{u_m}{\sum u_j} \quad j = 1, 2, \cdots, \xi + 1 \quad (5)$$

$$u_m = \exp \left\{ -d \left[q_{i1}, q_{i,m} \right] / d[q_{i1}, q] \right\} \quad (6)$$

q 是由影子流矩阵 Q_L 第一个分块矩阵中的第一个向量。

③对特征 i 和 j 的 $\xi + 1$ 个真实值进行局部加权平均, 求得两个参数之间的交叉映射估计, 表示为公式(7)、公式(8)所示:

$$\widehat{Q}_{i\tau} | Q_{j\tau} = \sum w_i Q_{i\tau} \quad (7)$$

$$\widehat{Q}_{j\tau} | Q_{i\tau} = \sum w_j Q_{j\tau} \quad (8)$$

④计算估计值与真实值之间的皮尔森相关系数, 表示两个参数之间的预测能力。计算如公式(9)、公式(10)所示:

$$\rho_{CCM_{Q_{i\tau}, \widehat{Q}_{i\tau}}} = \rho[Q_{i\tau}, \widehat{Q}_{i\tau} | Q_{j\tau}] \quad (9)$$

$$\rho_{CCM_{Q_{j\tau}, \widehat{Q}_{j\tau}}} = \rho[Q_{j\tau}, \widehat{Q}_{j\tau} | Q_{i\tau}] \quad (10)$$

公式(9)表示 i 对 j 的影响程度, 公式(10)表示 j 对 i 的影响程度。

⑤重复步骤①至④, 最终得到参数之间的交叉映射相关性系数矩阵 C 。 $C = \begin{bmatrix} c_{11} & \cdots & c_{1q} \\ \vdots & \ddots & \vdots \\ c_{q1} & \cdots & c_{qq} \end{bmatrix}$, c_{ij} 是第 i 个特征与第 j 个特征的交叉映射相关性系数。

⑥延长时间序列, 重复步骤①至⑤, 得到随时间变化的交叉映射相关性系数矩阵集合, $\mathbb{C} = \{C_1, C_2, \dots, C_p\}$ 。根据时间序列 β 检验法^[26]计算参数 i 、 j 之间相关系数收敛参数值 β_{ij} 、 μ_{ij} , 并生成收敛参数矩阵 $\Phi =$

$$\begin{bmatrix} \beta_{11} & \cdots & \beta_{1q} \\ \vdots & \ddots & \vdots \\ \beta_{q1} & \cdots & \beta_{qq} \end{bmatrix}, \Psi = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1q} \\ \vdots & \ddots & \vdots \\ \mu_{q1} & \cdots & \mu_{qq} \end{bmatrix}。$$

输出: 检查参数 i 、 j 之间相关系数取值及收敛参数取值, 判断变量之间是否存在因果关系。若 $c_{ij} > 0$ 且 $\beta_{ij} * \mu_{ij} < 0$ 则 i 是 j 的因, 若 $c_{ij} > 0$ 且 $\beta_{ij} * \mu_{ij} > 0$ 则 i 是 j 的伪因, 二者仅存在相关性。

3 CCMMM 建模方法应用

3.1 数据集介绍

综合已有研究发现, 不同类型突发事件的网络舆情影响因素和作用机制有区别, 其舆情传播模型也会存在差异^[31-32]。对此, 本文收集新浪微博 4 个不同类型的突发事件数据, 通过 CCMMM 方法发现不同网络舆情系统特征之间的因果关系, 挖掘相应传播模型。

在新浪微博平台以“双黄连+新冠”“甘肃省+白银越野赛”“十堰+爆炸”和“青海+地震”为关键词, 收集了双黄连事件、白银越野赛事件、十堰爆炸事件和青海地震事件网络舆情数据, 包括: 微博 id、用户 id、微博正文、转发数、发布时间、转发微博 id。将采集的数据进行处理, 生成 4 个不同的时序网络, 基本情况如表 1 所示。4 个事件的持续时间最长为 31 天, 最短为 26 天, 具体情况详见表 1。

表 1 数据集

Tab.1 Datasets

事件	事件类型	网络节点	网络连边	起止日期	持续时间/天
双黄连事件	公共卫生事件	199564	81728	2020/01/31—2020/02/29	30
白银越野赛事件	事故灾难事件	7282	746	2021/05/23—2021/06/21	30
十堰爆炸事件	社会安全事件	13908	2972	2021/06/13—2021/07/13	31
青海地震事件	自然灾害事件	23438	2125	2021/05/22—2021/06/16	26

图 3 展示 4 个事件的传播规模变化情况。双黄连事件的单日传播规模最大取值为 11 万, 是其他事件

的 10 倍。双黄连事件、白银越野赛事件及十堰爆炸事件传播规模分别在第 1 天、第 2 天就达到最高峰，之后在第 15 天、第 14 天、第 20 天出现二次发酵。而青海地震事件的传播规模则是在第 5 天达到高峰，之后缓慢消退。总之，突发事件网络舆情传播规模的变化差异有两点，一是是否存在二次发酵，二是消退时刻有早晚。

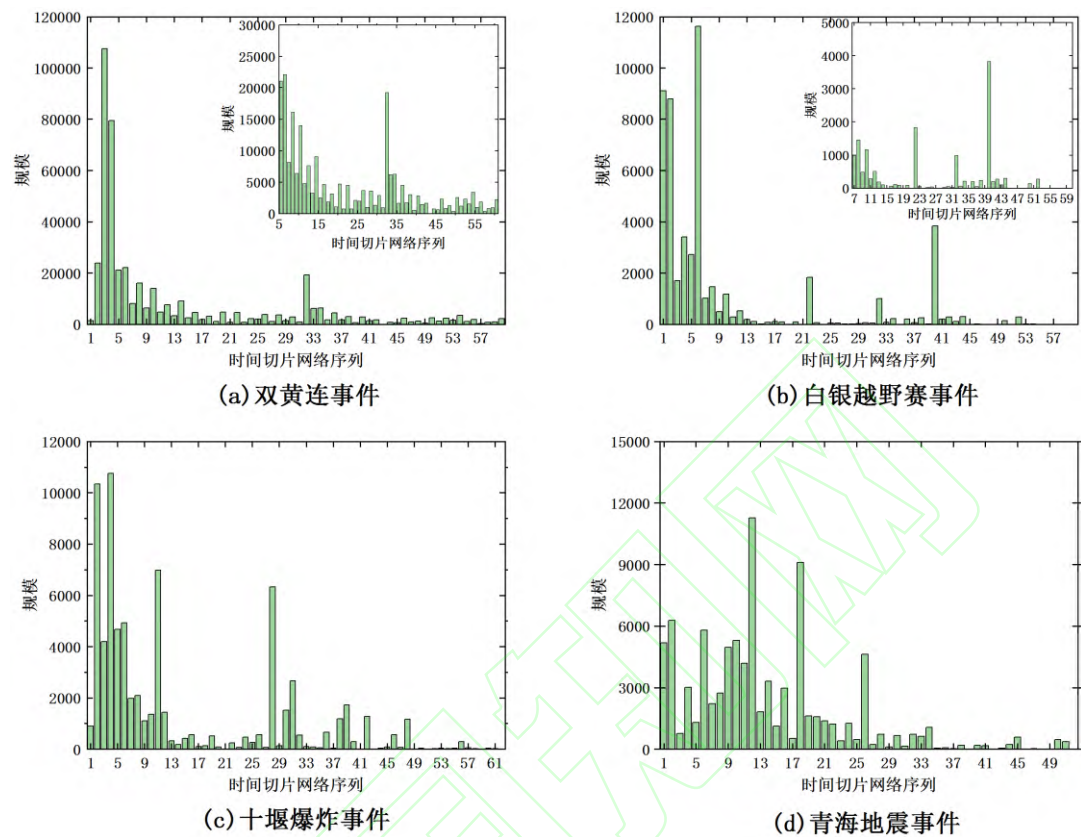


图 3 传播规模变化

Fig.3 The changes of propagation scale

3.2 CCMMM 参数设置

本研究以 12 个小时为单位，将时序网络分割成若干片段网络，分别统计网络舆情系统特征取值，并将这些时间序列数据作为 CCMMM 算法的输入。其中 G_i 网络的片段网络 $G_{i,j}$ 是它在时间段 $[t_j, t_{j+1}]$ 上的子网络， $T = \{[t_0, t_1], [t_1, t_2], \dots\}$ 。实验中设置初始时间序列长度 $L = 10$ 、滞后期 $\tau = 1$ 、时间序列增加数量 $\Delta = 4$ 。此外，通过 R 语言中 multispatialCCM 包中的 *SSR_pred_boot* 函数，计算各变量间的最佳嵌入维度 ξ ，如表 2 所示。

表 2 最佳嵌入维度

Tab.2 Optimum embedding dimension

CCM 对应关系	双黄连事件	白银越野赛事件	十堰爆炸事件	青海地震事件
$S \Rightarrow C \mid C \Rightarrow S$	2 2	3 3	3 2	4 3
$DA \Rightarrow S \mid S \Rightarrow DA$	2 2	2 3	3 3	4 4
$DA \Rightarrow C \mid C \Rightarrow DA$	2 2	2 3	3 2	4 3
$DA \Rightarrow K \mid K \Rightarrow DA$	2 3	2 4	3 3	4 3
$DA \Rightarrow DM \mid DM \Rightarrow DA$	2 2	2 4	3 5	4 2
$K \Rightarrow S \mid S \Rightarrow K$	3 2	4 3	3 3	3 4
$K \Rightarrow C \mid C \Rightarrow K$	3 2	4 3	3 2	3 3
$DM \Rightarrow S \mid S \Rightarrow DM$	2 2	4 3	5 3	2 4

$DM \Rightarrow C \mid C \Rightarrow DM$	2 2	4 3	5 2	2 3
$K \Rightarrow DM \mid DM \Rightarrow K$	3 2	4 4	3 5	3 2
$O \Rightarrow S \mid S \Rightarrow O$	4 3	2 2	2 2	2 4
$O \Rightarrow C \mid C \Rightarrow O$	4 4	2 4	2 2	2 4
$O \Rightarrow DA \mid DA \Rightarrow O$	4 2	2 2	2 3	2 4
$O \Rightarrow K \mid K \Rightarrow O$	4 3	2 4	2 3	2 4
$O \Rightarrow DM \mid DM \Rightarrow O$	4 3	2 5	2 4	2 2

4 结果分析

4.1 突发事件网络舆情系统特征分析

按照 2.3 节中定义，首先统计 4 个事件的信息源规模、简单传播规模和多源传播规模的比例，展示突发事件网络舆情系统的群体结构变化，如图 4(a)~图 4(d)所示。

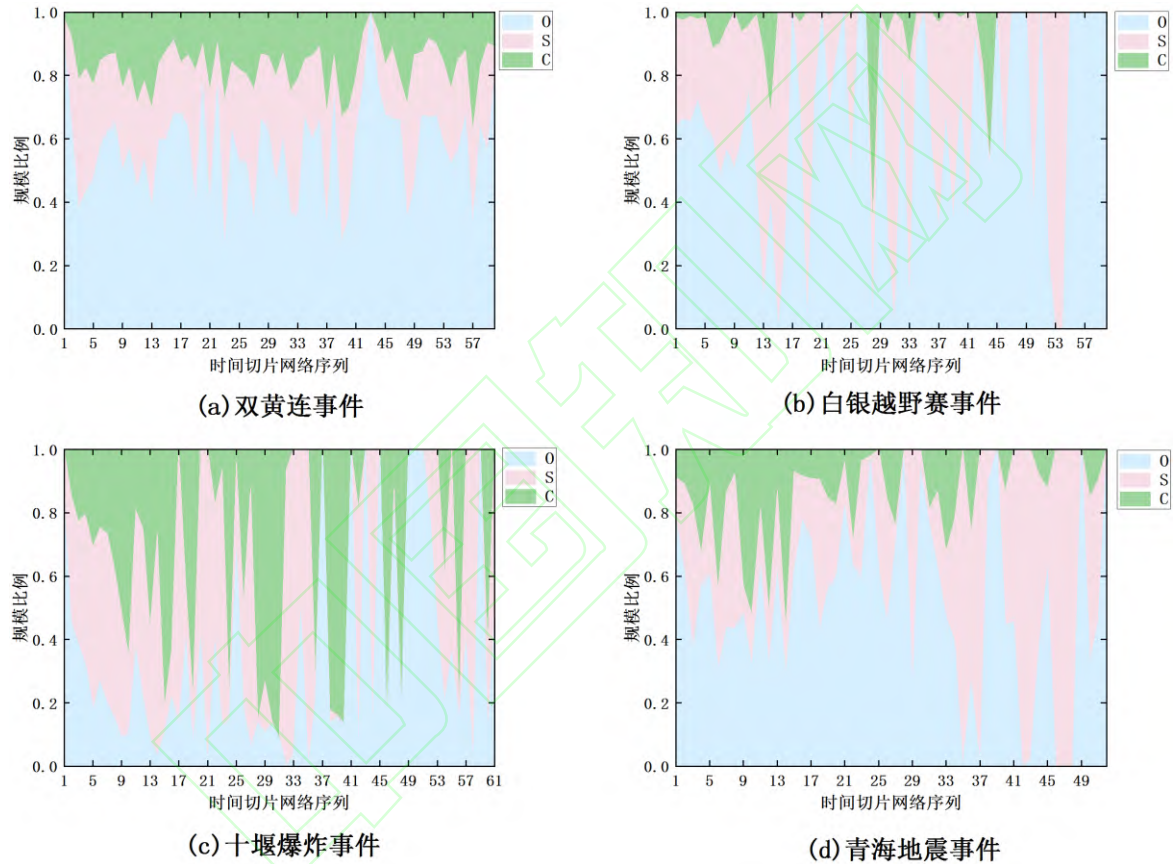


图 4 群体结构计算结果

Fig.4 The results of group structures

图 4 展示了信息源规模(浅蓝色)(计算机显示颜色,下同)与简单传播规模(粉红色)、多源传播规模(浅绿色)的变化趋势,说明 4 个事件所对应的群体结构动态变化模式存在差异。首先,各类事件中信息源规模占比都维持在 40%~80%之间,且信息源规模与简单传播规模变化趋势一致,这说明突发事件网络舆情过程中信息源带来的现场报道或者观点是引起舆情的重要动力。其次,在白银越野赛事件、青海地震事件后期多源传播规模基本为 0。原因是在此类事件中信息源发布的消息多为救灾或灾后内容,消息的争议性比较小,公众无需经过社会强化就能参与传播。第三,十堰爆炸事件中后期多源传播规模占优,是由于社会安全事件后续问责再次引起网民关注,在诸如燃气安全、旧城管理等议题讨论中出现了多源传播。

图 5 中展示了 4 个事件群体规范和群体过程计算结果。从图 5(a)中可以看出 4 个事件群体规范的作用方式不一样。双黄连事件的社会强化度取值维持了较高的水平,是因为公众对双黄连能否抑制新冠肺炎病毒一直持有争议。其他 3 个事件的早期社会强化度取值较高,但是会随着事件进展而衰减,即这是因为线下突发事件的合理处置可以消除公众对事件的不确定性,参与传播不需要多个邻居的社会强化。从图 5(b)

中可以看出，双黄连事件的持续能力最大值可达 165，平均取值是其他 3 个事件 2.6 倍。其他 3 个事件的持续能力则是前高后低，平均取值为 14.9 左右。这反映了当诸如新冠肺炎这一类的新问题时，个人和组织会在社交媒体平台上发布和搜寻新知识、新技能，群体过程的持续性会一直保持在较高的水平。当相似的突发事件重复发生时，其网络舆情传播过程中的群体交互会保持一定水平后消退。从图 5(c)中可得 4 个事件的社会亲密度水平最大取值为 0.24，平均值在 0.1 左右。按照其计算方法，若 DM 取值为 0.25，则此时段中参与者构成的交互网络可能连通，若 DM 取值为 0.1，则此时段的一半参与者可以发生交互，说明 4 个事件中突发舆情是典型的群体事件。

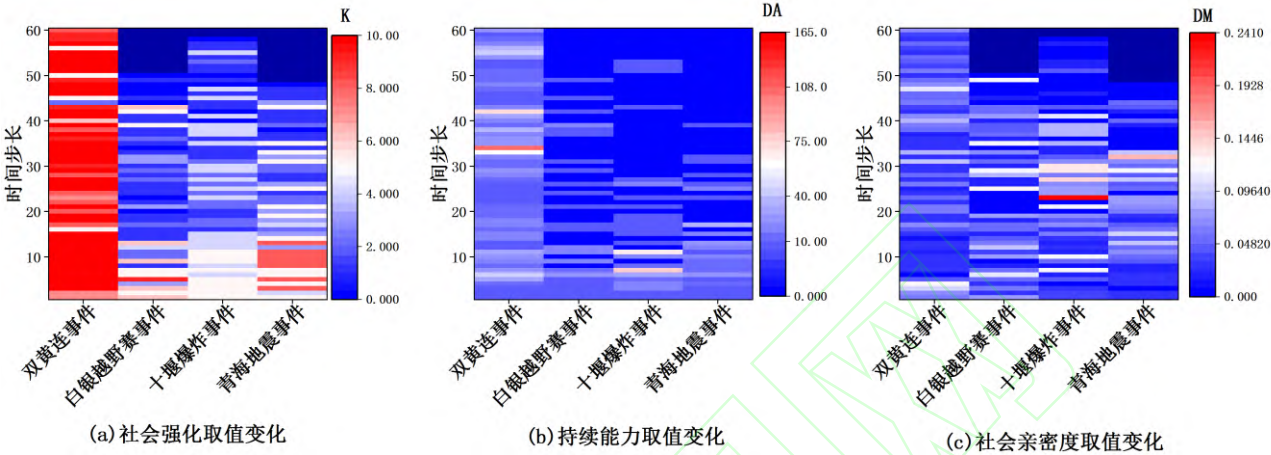


图 5 群体规范和群体过程计算结果

Fig.5 The results of group norms and group processes

4.2 因果关系识别结果

表 3 展示 CCMMM 算法所识别的因果关系。其中互为因果的有 4 组，分别是 S 和 C 、 O 和 C 、 K 和 S 、 K 和 C 。前两组结果证明了信息源、简单传播、多源传播之间互为因果，即群体结构中的各个要素都是不可或缺的。同时也验证了文献[36]的仿真实验结果：由于微博等在线社会网络的模块度高，给每个在线用户的交互提供了足够宽的“桥”，使得简单传播、多源传播相互为因果、相互促进。后两组结果证明了社会强化度、简单传播和多源传播之间互为因果，即群体规范与群体结构是相互作用的。因为在线社区上的社会强化等机制要发生效用，与个体如何“接触消息”的微观行为时间特征有关。尤其是突发事件网络舆情演化中，个体需要观察多个邻居状态，也即简单传播规模累积到一定数量后，才会社会强化作用下发生多源传播。此外， O 是 S 、 DA 、 K 、 DM 的因，这说明信息源对事件网络舆情的演化起到了直接推动作用。 C 与 DA 、 DA 与 K 、 DA 与 DM 、 DM 与 S 及 K 与 DM 之间也是单向因果关系。多源传播可以促进群体过程的持续，即对事件相关消息的在线讨论会保持群体过程的持续能力，进一步增加社会强化效应，并能促进群体的亲密程度增加。

表 3 舆情系统中的因果关系

Tab.3 The Causal relationships in public opinion system

待定关系	双黄连事件		白银越野赛事件		十堰爆炸事件		青海地震事件	
	收敛系数 ρ	判定结果	收敛系数 ρ	判定结果	收敛系数 ρ	判定结果	收敛系数 ρ	判定结果
$S \Rightarrow C$	0.784	互为因果	0.940	互为因果	0.575	互为因果	0.371	互为因果
$C \Rightarrow S$	0.886		0.956		0.722		0.597	
$DA \Rightarrow S$	不收敛	伪因果	不收敛	伪因果	不收敛	伪因果	不收敛	伪因果
$S \Rightarrow DA$	不收敛		不收敛		不收敛		不收敛	
$DA \Rightarrow C$	不收敛	单向因果	不收敛	单向因果	不收敛	单向因果	不收敛	单向因果
$C \Rightarrow DA$	0.296		0.291		0.239		0.336	
$DA \Rightarrow K$	0.254	单向	0.441	单向	0.278	单向	0.517	单向

$K \Rightarrow DA$	不收敛	因果	不收敛	因果	不收敛	因果	不收敛	因果
$DA \Rightarrow DM$	0.280	单向	0.286	单向	0.519	单向	0.313	单向
$DM \Rightarrow DA$	不收敛	因果	不收敛	因果	不收敛	因果	不收敛	因果
$K \Rightarrow S$	0.450	互为	0.619	互为	0.578	互为	0.287	互为
$S \Rightarrow K$	0.511	因果	0.620	因果	0.715	因果	0.387	因果
$K \Rightarrow C$	0.433	互为	0.906	互为	0.846	互为	0.595	互为
$C \Rightarrow K$	0.447	因果	0.729	因果	0.953	因果	0.788	因果
$DM \Rightarrow S$	0.193	单向	0.620	单向	0.267	单向	0.156	单向
$S \Rightarrow DM$	不收敛	因果	不收敛	因果	不收敛	因果	不收敛	因果
$DM \Rightarrow C$	不收敛	无因果	不收敛	无因果	不收敛	无因果	不收敛	无因果
$C \Rightarrow DM$	不收敛	因果	不收敛	因果	不收敛	因果	不收敛	因果
$K \Rightarrow DM$	0.352	单向	0.407	单向	0.256	单向	不收敛	无因果
$DM \Rightarrow K$	不收敛	因果	不收敛	因果	不收敛	因果	不收敛	因果
$O \Rightarrow S$	0.450	单向	0.936	单向	0.144	单向	0.445	单向
$S \Rightarrow O$	不收敛	因果	不收敛	因果	不收敛	因果	不收敛	因果
$O \Rightarrow C$	0.624	双向	0.982	双向	0.306	双向	0.530	双向
$C \Rightarrow O$	0.732	因果	0.990	因果	0.206	因果	0.401	因果
$O \Rightarrow DA$	0.485	单向	0.196	单向	0.263	单向	0.204	单向
$DA \Rightarrow O$	不收敛	因果	不收敛	因果	不收敛	因果	不收敛	因果
$O \Rightarrow K$	0.323	单向	0.667	单向	0.270	单向	0.626	单向
$K \Rightarrow O$	不收敛	因果	不收敛	因果	不收敛	因果	不收敛	因果
$O \Rightarrow DM$	0.136	单向	0.282	单向	0.132	单向	0.267	单向
$DM \Rightarrow O$	不收敛	因果	不收敛	因果	不收敛	因果	不收敛	因果

注： ρ 收敛时 A 与 B 存在因果关系，表示为 $A \Rightarrow B$ ， ρ 的大小表示因果关系强度。

4.3 突发事件网络舆情传播建模结果

综合表 3 中因果关系，最终得出的 4 个突发事件网络舆情的传播模型，如图 6 (a)~图 6(d)所示。图中左上区域节点 O 、 S 、 C 对应群体结构，右下区域 K 、 DA 、 DM 对应群体规范和群体过程。对比图 6 (a)~图 6(d)中的因果关系取值，可以看出双黄连事件中的因果关系有 12 组（占比 70.6%）取值低于 0.5，整体因果效应偏弱。此事件发生在疫情早期，主要议题是关于“双黄连体外可以抑制新冠肺炎病毒”。但是公众对新冠肺炎的医疗健康知识了解不够深入，一些错误解读在网上也同时传播，舆情复杂度和不确定性都很大，反应在模型中就表现为各个要素之间的因果关系取值较小。图 6(b)中 12 组因果关系取值大于 0.5，说明线上群体遵循了社会强化规范参与事件消息的讨论。此事件舆情主题以白银越野赛的发生、处置方式等内容为主，且由于线下相关企业和政府部门处理事件的举措不够透明和及时，令网民产生了争议，在模型中表现为各个要素之间的因果关系效应较强。图 6(d)与其他图有明显差异， K 和 DM 之间无因果关系，即社会强化与社会亲密度之间无关联。这是因为自然灾害事件中在线的消息多以实时报道为主，且在线群体关心的是自然灾害带来的后果，群体聚集是由新消息引起，而不是由于社会强化产生，因此模型中没有 K 和 DM 的关系。此外，图 6(a)与图 6(b)中均存在 O 和 C 的高强互因果关系，同时结合 3.1 节中的基本情况介绍，二者都存在二次发酵，这说明信息源及多源传播的相互促进是引起二次发酵的原因。

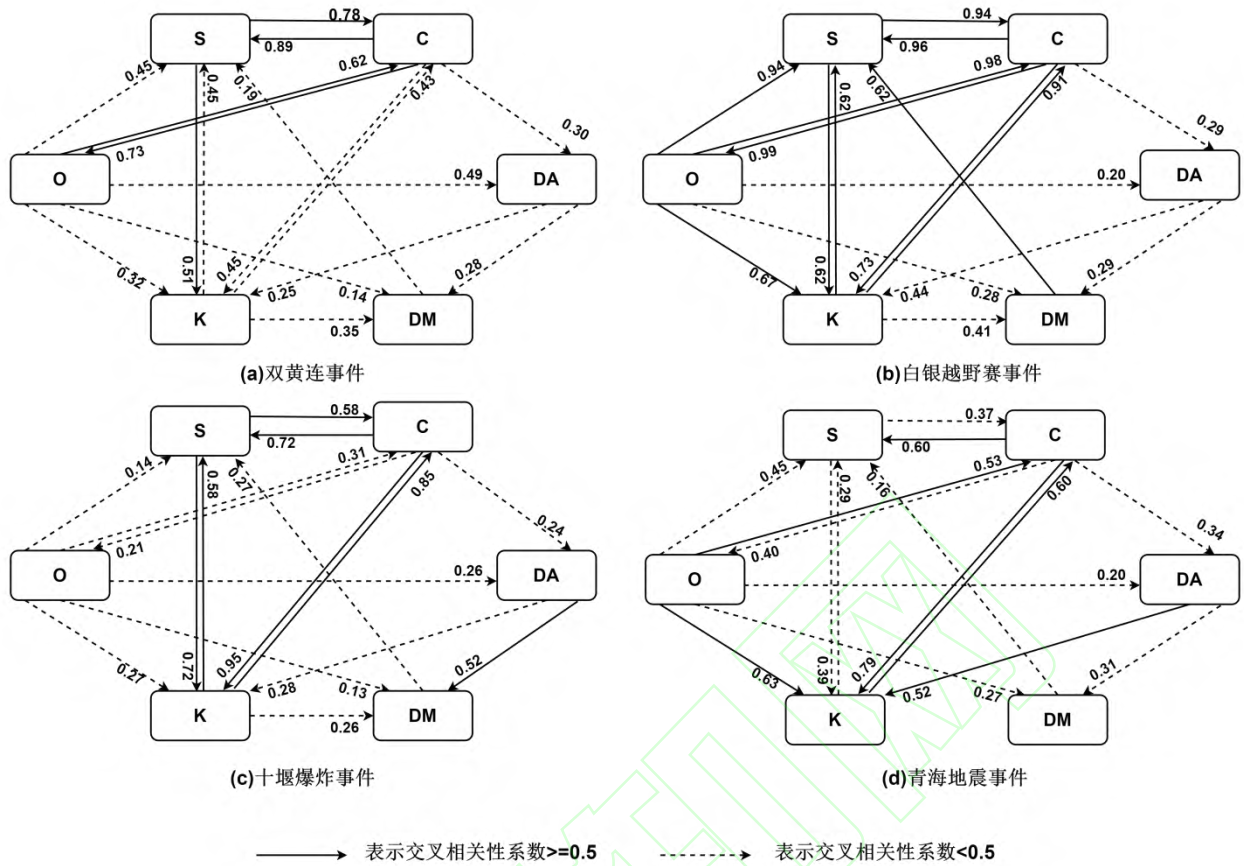


图 6 突发事件网络舆情传播模型

Fig.6 The models of Network public opinion propagation for different emergencies

4.4 讨论分析

从研究结果来看,本文所提出的传播建模方法既考虑了微观个体行为差异——舆情结果由3种不同行为汇集,又融合了群体行为机理——舆情演化由群体互动驱动,并用算法准确地揭示突发事件网络舆情演变中的因果关系,以应对和解决当前大数据时代背景下网络舆情的复杂性和动态性等问题。基于 CCM 的突发事件网络舆情传播建模方法要比传统的网络舆情建模方法具备可计算性、高效性以及更强的适应性等3个方面的优势。①可计算性是指能对舆情中的群体行为进行计算。在以疾病传染为基础而拓展的传播建模研究中,多关注个体如何在可疑、传染、免疫等状态中进行转换,难以刻画和捕捉舆情传播中的群体行为^[40]。而本研究从群体结构、群体规范和群体过程3个维度提出了测度方法,提高了传播过程的可解释性,为传播学与社会学相关的其他研究提供了更广阔的探索思路。②高效性是指建模方法时间效率高。只需对大规模舆情数据进行一次特征计算及执行一次算法流程,即可识别多个系统要素相互因果关系,输出精确建模结果。而通过 Agent 仿真探寻传播模型,需要各要素之间关系相互独立,重复多次实验后才能找到舆情传播中的关系强度^[41]。③适应性是指可以适用于不同类型突发事件网络舆情建模需求。尽管不同突发事件网络舆情传播结果有差异,但 CCMMM 方法不会因为参与者和互动关系的调整变化就失效,并能识别出自然灾害事件舆情传播模型的不同^[42]。而基于系统动力学的建模方法由建模者设置系统要素及互为因果的反馈函数,与建模者的知识理论和经验有关^[15]。

总之,借助数据驱动的机器学习方法解决舆情传播建模问题,是大数据及人工智能时代发展的研究范式和研究手段。

5 结论

针对目前舆情建模领域中在可计算化和社会过程研究缺乏的问题,本文提出基于 CCM 的突发事件网络舆情建模方法。为了验证所提方法的有效性,本文分别选择新浪微博上4个不同突发事件网络舆情数据进行分析。实验结果表明,该方法能够清晰地识别出每类突发事件网络舆情传播的差异,并对舆情动态过程和其中的因果关系做出定量解析,能为网络舆情治理的同时提供定性和定量决策参考。基于因果关系挖

掘探测舆情传播过程的思路是更适应当前大规模复杂网络信息时代的新兴研究范式。方法的局限是基于历史数据的事后建模,后续将进一步改进核心算法,结合时序长度、训练轮次等进行优化,围绕舆情传播过程的实时监控和预警开展探索和研究。

参考文献

- [1] 郑昱.突发公共事件中舆论信息传播倾向的影响因素——基于民众负性情绪的研究视角[J].情报理论与实践,2017,40(7):80-87.
- [2] 吴俊杰,郑凌方,杜文字,等.从风险预测到风险溯源:大数据赋能城市安全管理的行动设计研究[J].管理世界,2020,36(8):189-202.
- [3] 巢乃鹏,黄文森.范式转型与科学意识:计算传播学的新思考[J].新闻与写作,2020(5):6.
- [4] 赵雪芹,王青青,蔡铨.网络问答社区意见领袖的知识分享行为特征分析——以知乎“旅行”话题为例[J].情报科学,2021,39(6):68-74.
- [5] 王晰巍,邢云菲,韦雅楠,等.大数据驱动的社交网络舆情用户情感主题分类模型构建研究——以“移民”主题为例[J].信息资源管理学报,2020,10(1):11.
- [6] MIRBABAIE M, FROMM J, LPPENBERG S, et al. The potential of social media analytics for improving social media communication of emergency agencies[J]. arXiv preprint arXiv:2004.08570, 2020.
- [7] SHARAD G, ASHTON A, HOFMAN J, et al. The structural virality of online diffusion[J]. Management Science, 2016, 62(1): 180-196.
- [8] PIERRI F, PICCARDI C, CERI S. Topology comparison of twitter diffusion networks effectively reveals misleading information[J]. Scientific Reports, 2020, 10(1): 1-9.
- [9] Porter M A. Nonlinearity+ networks: A 2020 vision[M]//Emerging frontiers in nonlinear science. Springer, Cham, 2020: 131-159.
- [10] 杜运周,李佳馨,刘秋辰,等.复杂动态视角下的组态理论与QCA方法:研究进展与未来方向[J].管理世界,2021,37(3):18.
- [11] 瞿英,田红梅,种子鹏.基于改进SIR模型的突发公共事件网络舆情情感演化研究[J].河北科技大学学报(社会科学版),2021,21(2):80-85.
- [12] 赵剑华,万克文.基于信息传播模型——SIR传染病模型的社交网络舆情传播动力学模型研究[J].情报科学,2017,35(12):34-38.
- [13] 田世海,张家毓,孙美琪.基于改进SIR的网络舆情信息生态群落衍生研究[J].情报科学,2020,38(1):3-9,16.
- [14] 邓建高,张璇,傅柱,等.基于系统动力学的突发事件网络舆情传播研究:以“江苏响水爆炸事故”为例[J].数据分析与知识发现,2020,4(Z1):110-121.
- [15] 殷飞,张鹏,兰月新,等.基于系统动力学的突发事件网络谣言治理研究[J].情报科学,2018,36(4):57-63.
- [16] 叶琼元,兰月新,王强,等.面向突发事件的网民情绪演化系统动力学模型研究[J].情报杂志,2017,36(9):153-159,105.
- [17] 黄传超,胡斌,闫钰炜,等.网络暴力下突发事件中观点决策与舆情反转[J].管理工程学报,2019,33(1):252-258.
- [18] 吴鹏,王夏婷,金贝贝.基于BDI-Agent模型的网民集群行为建模研究[J].情报理论与实践,2019,42(4):136-144.
- [19] 彭程,祁凯,黎冰雪.基于SIR-EGM模型的复杂网络舆情传播与预警机制研究[J].情报科学,2020,38(3):145-153.
- [20] 王家坤,于灏,王新华,等.基于用户相对权重的在线社交网络舆情传播控制模型[J].系统工程理论与实践,2019,39(6):1565-1579.
- [21] 满媛媛,刘佳宁.国内突发事件网络舆情研究进展[J].情报科学,2020,38(12):170-177.
- [22] PERAL J, ROBINS J M, GREENLAND S. Confounding and collapsibility in causal inference. Statistical Science[J], 1999, 14(1):29-46.
- [23] NIZZOLI L, TARDELLI S, AVVENUTI M, et al. Coordinated behavior on social media in 2019 UK general election[J]. arXiv preprint arXiv:2008.08370, 2020.
- [24] 李振鹏,唐锡晋.集体行动的阈值模型[J].系统科学与数学,2014,34(5):550-564.
- [25] Tapia A G, Ramirez-Marquez J E. Understanding collective action through social media-based disaster data analytics[M]//Understanding Disaster Risk. Elsevier, 2021: 297-318.

- [26] HOLME P, SARAMAKI J . Temporal networks[J]. Physics Reports, 2012, 519(3): 97-125.
- [27] 张海峰,王文旭.复杂系统重构[J].物理学报,2020,69(8):92-102.
- [28] GEORGE SUGIHARA et al. Detecting causality in complex ecosystems[J]. Science, 2012, 338(6106) : 496-500.
- [29] PEARL J, GLYMOUR M, JEWELL N P . Causal inference in statistics: a primer[J]. Biometrics, 2019, 75(2) : 708-709.
- [30] MARARSHALL M . How to prove cause leads to effect[J]. New Scientist, 2012, 215(2884):8-9.
- [31] LIU Huajun, LEI Mingyu . The causality between traffic congestion and smog pollution—an empirical study using convergent cross mapping[J]. Statistical Research, 2019,36(10):43-57.
- [32] STOLBOV M ,SHCHEPELEVA M, SEGOT T L . Systemic risk, economic policy uncertainty and firm bankruptcies: evidence from multivariate causal inference[J]. Research in International Business and Finance, 2019,52:101172. <https://doi.org/10.1016/j.ribaf.2019.101172>.
- [33] WU T ,GAO X , AN S , et al. Time-varying pattern causality inference in global stock markets[J]. International Review of Financial Analysis, 2021,77:101806. <https://doi.org/10.1016/j.irfa.2021.101806>.
- [34] 张军.网络舆情时变演化机制及应对策略研究[M].北京:中国社会科学出版社,2020:45-50.
- [35] 张宸瑀. 非常规突发事件中自组织群体结构的发展及其对群体行为有效性的影响[D].杭州: 浙江大学,2018.
- [36] BARASH V, CAMERON C , MACY M . Critical phenomena in complex contagions[J]. Social Networks, 2012, 34(4):451-461.
- [37] FELDMAN D C . The development and enforcement of group norms[J]. Academy of Management Review, 1984, 9(1):47-53.
- [38] LINYUAN L, DUAN-BING C , TAO Z.The small world yields the most effective information spreading[J]. New Journal of Physics, 2011,13:1-10.
- [39] CAMPBELL E, SALATHE M. Complex social contagion makes networks more vulnerable to disease outbreaks[J]. Scientific Reports, 2013, 3(1):1905.
- [40] FAN C, JIANG Y ,YANG Y, et al. Crowd or hubs: information diffusion patterns in online social networks in disasters[J]. International Journal of Disaster Risk Reduction, 2020, 46:101498. <https://doi.org/10.1016/j.ijdr.2020.101498>.
- [41] QIANG Z,PASILIAO E L, ZHENG Q P . Model-based learning of information diffusion in social media networks[J]. Applied Network Science, 2019,4(1):1-16.
- [42] 赵飞,廖永丰.突发自然灾害事件网络舆情传播特征及影响因素研究[J].地球信息科学学报,2021,23(6):992-1001.

作者简介: 张军 (ORCID:0000-0001-5745-1281), 女, 1978 年生, 博士, 教授, 硕士生导师。研究方向: 复杂系统重构。王学金 (ORCID:0000-0002-3480-1979), 男, 1996 年生, 硕士生。研究方向: 数据挖掘与数据分析。李鹏 (ORCID:0000-0002-9481-1642, 通信作者, Email:lp_sdut@163.com), 男, 1977 年生, 硕士, 副教授。研究方向: 情报学。庄云蓓 (ORCID:0000-0002-5562-5775), 1990 年生, 博士, 讲师。研究方向: 文本挖掘。

作者贡献声明: 张军, 提出研究命题、研究思路、撰写论文及修订论文。王学金, 数据分析与实验。李鹏, 理论模型构建及论文修改。庄云蓓, 数据分析与实验。

录用日期: 2021-12-20