



开放科学
(资源服务)
标识码
(OSID)

突发事件下基于改进 K-Shell 分解的意见领袖识别研究

芦子涵 郑中团

上海工程技术大学数理与统计学院 上海 201600

摘要: [目的/意义] 为实现突发事件网络舆情的精准管控,对突发事件中的意见领袖识别进行研究。[方法/过程] 针对 K-Shell 分解使得同一核层的节点具有相同 K-Shell 值的粗粒化分解问题,结合用户自身属性与用户交互行为对核心用户进行用户重要度因子量化,并通过重构各节点 K-Shell 值的计算方法来加以改进;在此基础上定义以转发比例为权重的相邻用户重要度贡献值,从而构建一套意见领袖的识别方法。[结果/结论] 以“郑州地铁 7·20 事件”为例,进行实证分析。结果表明突发事件中意见领袖主要由主流媒体与自媒体两类用户组成,且意见领袖的特征与类型随舆情生命周期的变化而变化。本文提出的意见领袖识别方法能够精确地给出意见领袖的排名,较 K-Shell 分解法识别效率更高,较社交平台传统的排序方法更具可解释性。

关键词: 社会网络;突发事件;网络舆情;K-Shell 分解;意见领袖

中图分类号: G35; TP391

Research on the Identification of Opinion Leaders Based on Improved K-Shell Decomposition in Emergency

LU Zihan ZHENG Zhongtuan

School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai 201600, China

Abstract: [Objective/Significance] In order to realize the precise management and control of network public opinion in emergencies, the identification of opinion leaders in emergencies is studied. [Methods/Process] Aiming at the coarse-grained decomposition problem that K-Shell decomposition makes nodes in the same core layer have the same K-Shell value, the user importance factor of core users is quantified to reconstruction the calculation of K-Shell by combining the user's own attributes and user interaction behaviors; On this basis, define the contribution value of the importance of adjacent users with the forwarding ratio as the weight, so as to construct a set of opinion leaders identification methods. [Result/Conclusion] Take “Zhengzhou

作者简介 芦子涵(1998-), 硕士研究生, 研究方向为社会网络分析、文本挖掘; 郑中团(1979-), 博士, 副教授, 研究方向为机器学习与数据挖掘、应用统计与综合评价、区域可持续发展的统计测度、随机过程与复杂网络等, E-mail: zhongtuanzheng@163.com。

引用格式 芦子涵, 郑中团. 突发事件下基于改进 K-Shell 分解的意见领袖识别研究 [J]. 情报工程, 2023, 9(1): 30-42.

Metro 7.20 Incident” as an example to conduct empirical analysis. The results show that opinion leaders in emergencies are mainly composed of mainstream media and self-media users, and the characteristics and types of opinion leaders change with the life cycle of public opinion. The opinion leader identification method proposed in this paper can accurately give the ranking of opinion leaders, which is more efficient than the K-Shell decomposition method, and more interpretable than the traditional ranking method of social platforms.

Key words: Social Network; Emergency; Public Opinion; K-Shell Decomposition; Opinion Leader

引言

意见领袖 (Opinion Leader, OL) 这一概念最早由 Lazarsfeld 在其著作《人民的选择》中提出, 他认为意见领袖是指在信息传播网中进行信息传播、提供观点、发表看法, 引导舆论走向, 并间接或无形地影响其他用户的态度, 对整个事件的发展具有极大社会影响力的重要人物^[1]。

《中华人民共和国突发事件应对法》^[2]中指出, 突发事件是指突然发生, 造成或者可能造成严重社会危害, 需要采取应急处置措施予以应对的自然灾害、事故灾难、公共卫生事件和社会安全事件。近年来, 突发事件时有发生, 2020 年新冠肺炎疫情等公共卫生事件、2021 年 7·20 河南特大暴雨等自然灾害事件、2022 年 3·21 东航航班 MU5735 坠机等事故灾难事件等。与此同时, 随着网民规模的扩大与社交平台的普及, 像新浪微博这样传播范围广、普及率高的社交网络平台逐渐成为突发事件的曝光口与网络舆情的发生地。在突发事件网络舆情传播过程中, 有一类群体处于社交网络的中心, 能够在舆论传播的过程中引导舆论走向, 他们被称为意见领袖。由于突发事件具有不确定性、危害性等特点^[3], 通常会给社会大众带来负向的心理冲击。如若不能针对性地引导积极的舆论走向并

建立舆情治理机制, 则会放大社会大众的负面情绪, 引起不必要的激进言论, 甚至会对政府机构造成不良影响。因此, 有效的识别意见领袖方法能够帮助舆情管控部门实现精准化管控, 从而快速平复大众情绪, 营造积极的网络氛围。

针对意见领袖识别与特征分析问题, 为改进 K-Shell 分解^[4]使得同一核层的节点具有相同 K-Shell 值的粗粒化分解问题, 本文将结合用户自身属性与用户交互行为, 引入用户重要度因子并以此重构 K-Shell 值的计算方法。其次, 在此基础上定义以转发比例为权重的相邻用户重要度贡献值, 从而构建一套意见领袖的识别方法。

1 相关研究

目前, 对于突发事件中意见领袖的识别研究主要有以下方面。一是基于社交网络图的识别方法, 该方法多以构建社交网络图为基础, 以分析网络图中的相关指标为方法或以改进网络图中的排序算法为方法来识别意见领袖。D.Miorandi 等^[4]认为 K-shell 值越大, 节点影响力越大。康伟等^[5]以重大突发事件“7·23”动车事故为实证研究对象构建网络拓扑图, 进行了网络密度、可达性、聚类系数和中心性的测度,

依据测度结果进行关键节点即意见领袖的识别。周飞等^[6]采用粉丝数、获得赞同数、回答数来衡量用户自身影响力作为初值,以被浏览量的增长率作为权重来改进 PageRank 算法,提出名为 ZhihuRank 的算法来识别意见领袖。二是多指标的综合评价方法,该方法多以用户属性信息、用户交互行为等指标建立指标体系,并通过综合评价方法来计算综合得分,以此来识别意见领袖。彭丽徽等^[7]从用户的影响力、活跃度、认同度三个维度出发构建指标体系,并基于改进的模糊-层次分析法与灰色关联分析得到意见领袖的排序,从而识别出意见领袖。三是多角度融合的方法,该方法综合考虑网络结构、用户属性与交互行为指标、文本信息等要素,以此来识别意见领袖。李贺等^[8]从内容与结构两个维度出发,以社会网络与内容要素为一级指标,以中心度、关联度、活跃度、影响度等 8 个指标构成“8 度”二级指标,并采用模糊 TOPSIS 法综合评价并排序得出意见领袖。意见领袖的识别研究主要从基于社交网络图的识别方法、多指标的综合评价方法、多角度融合的方法展开。上述这些识别研究需要对参与事件讨论的每个用户或节点进行计算,然而现实中往往只有少数人具有成为意见领袖的潜质。故如何省去对不必要节点或用户的计算即如何筛选核心用户是本文关注的一个研究问题。

在社交网络中,用户的重要度不仅与其在网络中的位置有关,还与用户的自身属性与交互行为有关。例如,粉丝数 100 万的用户往往比粉丝数 10 万的用户具有更强的传播信息能力与影响力。为识别意见领袖,陈芬等^[9]在其研究中从影响力、活跃度、专业性三个角度出发,

构建指标体系。王佳敏等^[10]在其意见领袖影响力研究中从活跃度、影响力两个角度出发,以粉丝数、被转发数、被评论数、是否认证、关注人数与微博数等指标构建指标体系。彭丽徽等^[7]从影响力、活跃度与认同度构建指标体系,以此来计算用户影响力,从而得到意见领袖。Cha 等^[11]从用户交互行为的角度将粉丝数、转发数和被提及数 3 个指标进行对比分析。得出在舆情传播过程中转发数对传播舆情的影响力更大的结论。孙国梓^[12]在其微博影响力量化模型研究中提出微博博文影响力可由其影响的广度与深度体现,而影响广度可由转发数等指标体现,而影响深度可由评论数与信息活跃周期体现。在这些研究成果的基础上,同时遵循数据的可获取性,本文针对 K-Shell 分解算法存在无法给出节点的细粒度区分这一问题,将结合用户自身属性与用户交互行为从用户活跃度、用户影响力与用户权威性三个角度来构建指标体系,从而对部分核心用户进行用户重要度因子的度量以期给出用户的细粒度区分。

综上,大多学者所提出的识别方法通常需要对各个节点或用户进行计算,而对于突发事件所形成的大规模网络,其往往具有绝大多数节点的度值小,而只有少数节点具有度值大的性质,即只有少数节点具有较高的影响力,应当避免对不重要的节点的计算,降低算法复杂度;同时考虑到在 K-Shell 算法中, K-Shell 分解使得同一网络核层的节点具有相同的 K-Shell 值^[4,13],因而存在无法进一步区分节点的重要度这一缺陷。针对这两方面问题,本文提出基于 K-Shell 分解的 KSIN 算法来识别意见领袖,其优势与创新点具体体现在算法流程的 3 个步骤

上,具体如下:

(1) 首先,以 K-Shell 方法为基础将网络分解为 K 个核层以筛选出部分核心用户,避免了对网络中边缘节点不必要的计算;

(2) 其次,结合用户自身属性和用户交互行为引入用户重要度因子重构 K-Shell 值的计算方法,解决了 K-Shell 算法无法给出节点细粒度区分的缺陷。

(3) 最后,在(1)和(2)的基础上,定义了以转发比例为权重的相邻用户重要度贡献值,从而得到用户的意见领袖得分。

2 KSIN 算法与相关定义

2.1 K-Shell分解

K-Shell 分解^[13]是复杂网络中一种识别网络中核心节点的粗粒化方法,由 Kitsak 等人提出。其基本思想是以节点的度值为度量,迭代地删除网络中度数小于或等于 K 的节点及其连边,从而实现从外到内的核心节点发现。

具体步骤如下:假设网络中不存在度数为 0 的孤立节点。首先,将度数为 1 的节点及其连边删除(如图 1 中的节点 12、13、15、16、17、18),此时网络中会出现因删除上述节点而度数转变为 1 的节点,再将这些节点及其连边删除(如图 1 中的节点 7、8、11、14),循环该步,直至网络中不再出现度数为 1 的节点为止。由这些被删除的节点及其连边所构成的子图称之为 1-核,其中的节点 K-Shell 值(KS 值)为 1。此时,网络中所有节点的度数大于等于 2。接着去掉网络中度数为 2 的节点,以此类推循

环执行,直至网络中所有节点都得到 KS 值为止。由 K-Shell 分解得到的最大 K-核子图中的节点具有相同的 KS 值,而未能进一步区分这些节点的重要度差异,本文将引入用户重要度因子对此进行改进。

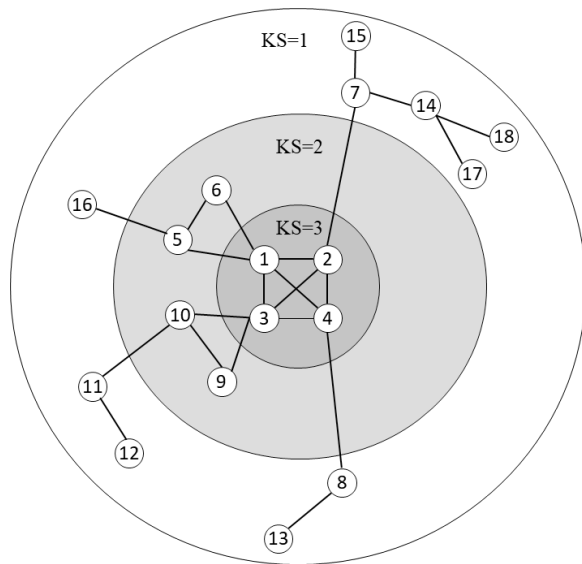


图 1 K-Shell 分解

2.2 用户重要度因子构建

新浪微博作为一个传播范围广、普及率高的社交平台,用户可以通过转发、点赞、评论参与到某一话题的讨论中来,进行信息的接收与传播。在社交网络中,用户的重要度不仅与其在网络中的位置有关,还与用户的自身属性与交互行为有关。通过借鉴已有的研究成果^[8-11],同时遵循数据的可获取性,将从用户活跃度、用户影响力与用户权威性三个角度来构建指标体系,从而构建用户重要度因子。

(1) 用户活跃度

Lazarsfeld 认为意见领袖是指在信息传播网中进行信息传播、提供观点、发表看法,

并间接或直接地影响其他用户态度的人物。因此,意见领袖应当是在突发事件舆情中持有看法与态度并将之发表出来的活跃用户。采用微博数 F_1 、关注人数 F_2 来度量用户的活跃度 $ACT(u_i)$, 定义如下:

$$ACT(u_i) = \omega_1 F_1 + \omega_2 F_2$$

(2) 用户影响力

意见领袖应当是某一话题中极具影响力的用户。若用户只具备活跃度,而不具备一定的影响力,其所发表的内容没有足够的曝光,则很难产生信息的交互与传播。若用户的追随者多,则其所发的博文更加容易产生转发、评论或点赞的交互行为,增加突发事件的曝光与传播,从而引导舆论方向。采用粉丝数 F_3 、历史被转发数 F_4 、历史被评论数 F_5 、历史被点赞数 F_6 来度量用户影响力 $INF(u_i)$, 定义如下:

$$INF(u_i) = \omega_3 F_3 + \omega_4 F_4 + \omega_5 F_5 + \omega_6 F_6$$

(3) 用户权威性

意见领袖应当具有被其他用户认同信服的能力,应当具有一定的权威性。用户权威性不同于用户影响力,用户权威性经常体现在专业性,具体表现为官方认证等级,是否为VIP用户等。新浪微博平台给出的官方认证需用户提交申请、官方多方面审查等多层程序,是对用户在其所在的领域权威性的一种官方认可。在彭丽徽等^[7]的研究中,构建了以影响力、活跃度与认同度为一级指标的指标体系,其中,认为“是否认证”是用户认同度的体现。在陈芬等^[9]的研究中,认为“是否认证”则是专业性的体现。本文采用是否认证 F_7 、是否VIP用户 F_8 来度量用户权威性 $AUT(u_i)$ 。其中,认证用户中蓝V表示机构认证,如政府、媒体、企业用户等

可申请机构认证;金V、橙V表示个人认证。将是否认证这一指标按照等级进行了划分,定义如下:

$$F_7 = \begin{cases} 3 & \text{蓝V用户} \\ 2 & \text{金V用户} \\ 1 & \text{橙V用户} \\ 0 & \text{无认证用户} \end{cases}$$

$$F_8 = \begin{cases} 1 & \text{VIP用户} \\ 0 & \text{非VIP用户} \end{cases}$$

$$ACT(u_i) = \omega_7 F_7 + \omega_8 F_8$$

在考虑用户自身属性与交互行为的情况下,构建了度量用户重要度因子的三元组 $(ACT(u_i), INF(u_i), AUT(u_i))$, 其中 $ACT(u_i)$ 代表用户活跃度, $INF(u_i)$ 代表用户影响力, $AUT(u_i)$ 代表用户权威性。参考张仰森^[14]在其用户权威度定量评价方法中的做法,将用户重要度因子 $IMP(u_i)$ 的计算转换为三元组 $(ACT(u_i), INF(u_i), AUT(u_i))$ 中各个特征的线性加权和,定义如下:

$$IMP(u_i) = \alpha \cdot ACT(u_i) + \beta \cdot INF(u_i) + \gamma \cdot AUT(u_i)$$

其中, u_i 代表用户, α 、 β 、 γ 代表各个特征的权重, $\omega_1 \sim \omega_8$ 代表其对应指标的权重。熵权法是一种客观赋权的方法,其原理是根据各评价指标数值的变异程度所反映的信息量大小来确定指标权重。在本节中,特征权重 α 、 β 、 γ 与 $\omega_1 \sim \omega_8$ 均使用熵权法确定。

2.3 基于改进K-Shell分解的KSIN算法

先前的大多研究多采用微博平台中用户间的关注关系来构建网络,然而因关注某一突发事件所形成的临时群体不一定都具有关注关系^[15]。在微博平台中,即使用户间不相互关注,仍然可以查看对方博文内容,并进行转发、点赞或评论等操作。考虑到微博这一特性,

且转发行为对舆情传播的影响力更大^[11], 本文将用户为节点, 以转发关系为边, 构建有向转发网络 G , 并得到用户集合 U 。

在突发事件中, 并非每一个参与该突发事件讨论并产生交互行为的用户都能够成为意见领袖。本文使用 K-Shell 分解得到 KS 值最大的核层, 引入用户重要度因子对 KS 值的计算进行重构, 得到 KSI 值。一方面可避免对边缘用户进行不必要的计算, 另一方面可区分同一核层具有相同 KS 值的用户。

其次, 考虑到用户的重要度不仅与其自身有关, 还与其相邻用户的重要度有关。为了兼顾考虑全局属性与局部属性, 全面识别网络中的意见领袖, 在上述改进 K-Shell 算法对用户进行区分所得到的 KSI 值的基础上, 以转发比例为权重考虑相邻用户所贡献的重要度, 得到最终的重要度得分排序, 得分高者, 认定为意见领袖。综上, 基于改进 K-Shell 分解的意见领袖识别算法 (简记为 $KSIN$ 算法) 的具体流程如图 2 所示:

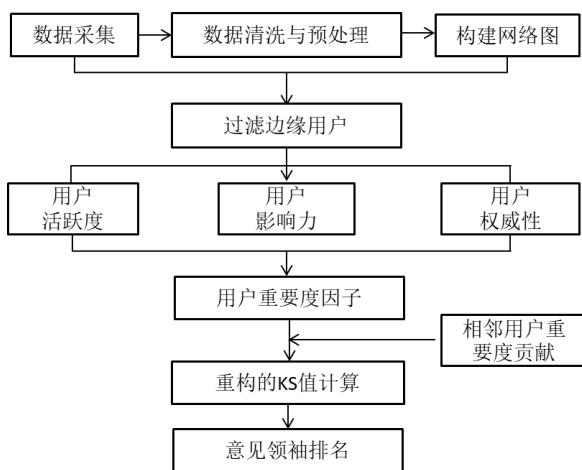


图 2 识别流程图

步骤 1: 采用 K-Shell 算法将网络分解

为 K 个核层, 得到网络中各节点的 KS 值, 并选择 KS 值最大的核层, 得到 KS 值向量

$$\mathbf{KS} = (ks_{u_1}, ks_{u_2}, \dots, ks_{u_n})$$

步骤 2: 构建用户重要度因子, 进一步区分节点重要度, 得到用户重要度向量

$$\mathbf{I} = (imp(u_1), imp(u_2), \dots, imp(u_n))$$

步骤 3: 引入用户重要度因子, 重构 KS 值, 得到 KSI 值。即:

$$ksi_{u_i} = imp(u_i) \times ks_{u_i} \quad (1)$$

其中 u_i 表示用户, $i=1,2,3,\dots,n$ 。将上式用矩阵表示为:

$$\mathbf{KS}' = \mathbf{I} \odot \mathbf{KS}$$

$$(ksi_{u_1} \quad ksi_{u_2} \quad \dots \quad ksi_{u_n}) = (imp_{u_1} \quad imp_{u_2} \quad \dots \quad imp_{u_n}) \odot (ks_{u_1} \quad ks_{u_2} \quad \dots \quad ks_{u_n})$$

其中矩阵 $\mathbf{KS}'$ 为各用户 KSI 值所构成的 n 维向量。

步骤 4: 计算相邻用户所贡献的重要度值, 最后按照各用户的 $KSIN$ 值排序。

$$ksin_{u_i} = ksi_{u_i} \times \sum_{u_j \in \tau_{u_i}} ksi_{u_j} \times r_{ji} \quad (2)$$

$$r_{ji} = \frac{R_{ji}}{SR_i} \quad (3)$$

其中, τ_{u_i} 表示用户 u_i 的相邻用户的集合, R_{ji} 为用户 u_j 转发用户 u_i 的次数, SR_i 为用户 u_i 被转发的总次数, r_{ji} 为转发比例。

用矩阵表示为:

$$\mathbf{KS}'' = \mathbf{R} \times \mathbf{KS}' + \mathbf{KS}'$$

$$\begin{pmatrix} ksin_{u_1} \\ ksin_{u_2} \\ \dots \\ ksin_{u_n} \end{pmatrix} = \begin{pmatrix} 0 & r_{21} & r_{31} & \dots & r_{n1} \\ r_{12} & 0 & r_{32} & \dots & r_{n2} \\ \dots & \dots & \dots & \dots & \dots \\ r_{1n} & r_{2n} & r_{3n} & \dots & 0 \end{pmatrix} \begin{pmatrix} ksi_{u_1} \\ ksi_{u_2} \\ \dots \\ ksi_{u_n} \end{pmatrix} + \begin{pmatrix} ksi_{u_1} \\ ksi_{u_2} \\ \dots \\ ksi_{u_n} \end{pmatrix}$$

其中, $\mathbf{KS}''$ 为各用户 $KSIN$ 值所构成的 n 维向量, $\mathbf{R}$ 为转发比例邻接矩阵。

3 实证与分析

3.1 突发事件概述与数据准备

2021年7月20日,河南郑州发生罕见特大暴雨。当日晚间19时左右,据郑州本地广播官方微博“@MyRadio”发布的微博称,郑州地铁5号线雨水倒灌,车厢内积水已到达乘客胸部,数名乘客被困。随后该条微博被澎湃新闻官方微博“@澎湃新闻”转发,转发人次5.2万,评论人次3.7万,事件爆发。截至当日晚间22时左右,消防救援人员陆续疏散被困人员500余人。7月21日上午,郑州地铁官方发布称此次事件导致12人遇难。随后,两名个人用户“@假装是个么得人发现的小号”“@大脸蛙8399”发布博文称有乘客邹某、沙某仍失联。26日,乘客邹某、沙某确认遇难,27日上午,

“@郑州”发布官方发布此次事件最终导致14人遇难,再次引起一波舆论高潮。2022年1月21日,国务院调查组调查认定郑州地铁5号线亡人系责任事件,是造成重大人员伤亡与财产损失的突发事件。

本文以“郑州地铁7.20事件”为实证话题,以2021年7月20日19时—2021年7月31日22时为时间区间,以2个小时为一个时间段,以“郑州地铁5号线”“多人被困”等为关键词,使用爬虫采集器Gooseeker^[16]采集数据,共采集到137990条转发博文数据与用户信息数据,微博发布趋势如图3所示。其中,转发博文数据包括用户id、博文内容、发布时间、被转发用户id、被转发博文内容、转发数、评论数、点赞数等;用户信息数据包括用户id、发布微博数、历史被转发数、历史被评论数、历史被点赞数、认证信息等。

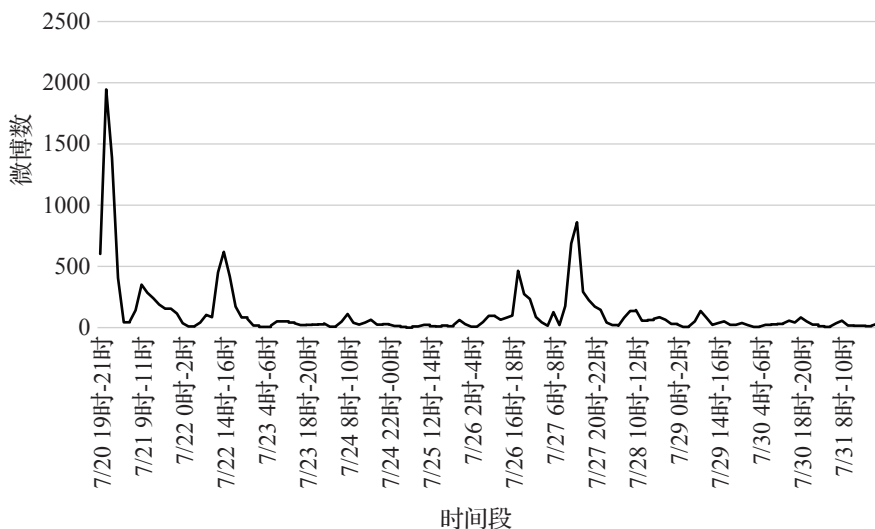


图3 郑州地铁事件微博发布趋势

舆情事件的形成和演化具有一定的生命周期特性^[17]。图3可以看出该事件发展规律总体上遵循从潜伏期到爆发期,从爆发期到冷淡期

这一规律,且潜伏期到爆发期的过程较短。同时,可以发现该事件在时间区间内引起两次爆发,且第二次爆发程度小于第一次,与事件发

展实际情况相符。考虑到同一用户可能在不同的生命周期阶段均发布或转发微博这一问题,且为探寻突发事件舆情不同生命周期的意见领袖及其特征,按照实际状况将该突发事件生命周期分为潜伏—爆发期、冷淡期、二次爆发期、二次冷淡期四个时期。

由于存在部分用户利用该关键词热搜词条发布与事件不相关的博文内容以获取热度,且存在数据爬取过程中所产生重复的转发博文数据,因此剔除重复转发博文与不相关博文。另外,根据微博转发博文的特点,其往往以“//@用户 A:***//@用户 B:***”的形式存在,需对转发博文内容利用 Python 编程中的正则表达式分离成两两转发关系。以用户为节点,以转发关系为边,以被转发为入度,以转发为出度,分别以四个周期构建有向网络,具体信息如表 1 所示。

表 1 网络图基本信息

时期	节点数	连边数	最大K核值
潜伏—爆发期	32405	33820	4
冷淡期	42051	44231	8
二次爆发期	37162	41133	3
二次冷淡期	26877	28240	4

3.2 实证分析

3.2.1 转发网络构建与分析

根据 K-Shell 分解对四个时期的转发网络进行分解过滤并得到各个用户的 KS 值,得到如图 4 所示的转发关系有向网络图,从转发网络图中可以看出各周期均存在少数占据网络中心的关键节点。各周期网络图的相关描述性统计指标如表 2 所示,以潜伏—爆发期为例,分解后

的节点数仅占总节点数的 0.5%,这说明了大多数用户在突发事件讨论所形成的临时群体中占据边缘位置。在意见领袖的评估中,这些大多数用户没有成为意见领袖的特质。

“被转发”即入度,可使用户的信息得到扩散,放大突发事件的扩散范围,从而达到更广的受众范围^[15]。以潜伏爆发期为例,根据用户之间的转发关系,可以得到入度的度分布序列,从而得到如图 5 所示的转发网络中的长尾现象。可以看出,基于这种转发关系所构建的社交网络属于“无标度网络”,其表现为绝大多数用户的入度很小,而只有少数用户的入度大,即被转发的次数多。这说明只有少数用户所发布的博文能够达到一定范围的受众面,具有一定的影响力。同时该“长尾现象”也再次印证了先前的论述,即只有参与突发事件讨论的少数用户具有成为意见领袖的特征。

3.2.2 意见领袖识别结果

根据基于 K-Shell 分解的 KSIN 算法对上节 K-Shell 分解过滤后的用户计算用户重要度因子与邻居用户所贡献的重要度值,得到最终的 KSIN 得分并依据此排名,截取排名前 10 位用户如表 3 所示。

结合图 6 认证用户分布情况可以看出,首先在该事件中,从整体来看表 3 所示的用户均为官方认证用户,即具有一定的权威性。爆发期的意见领袖多以蓝 V 用户为主,而冷淡期的意见领袖多以金 V 用户为主。其次,对比图 4,可以看出除“@假装是个么得人发现的小号”外,各时期排名前三的用户均处于各网络的中心。

“@假装是个么得人发现的小号”为事件当事人家属,其所发布的博文虽得到 2.3 万次的转发,

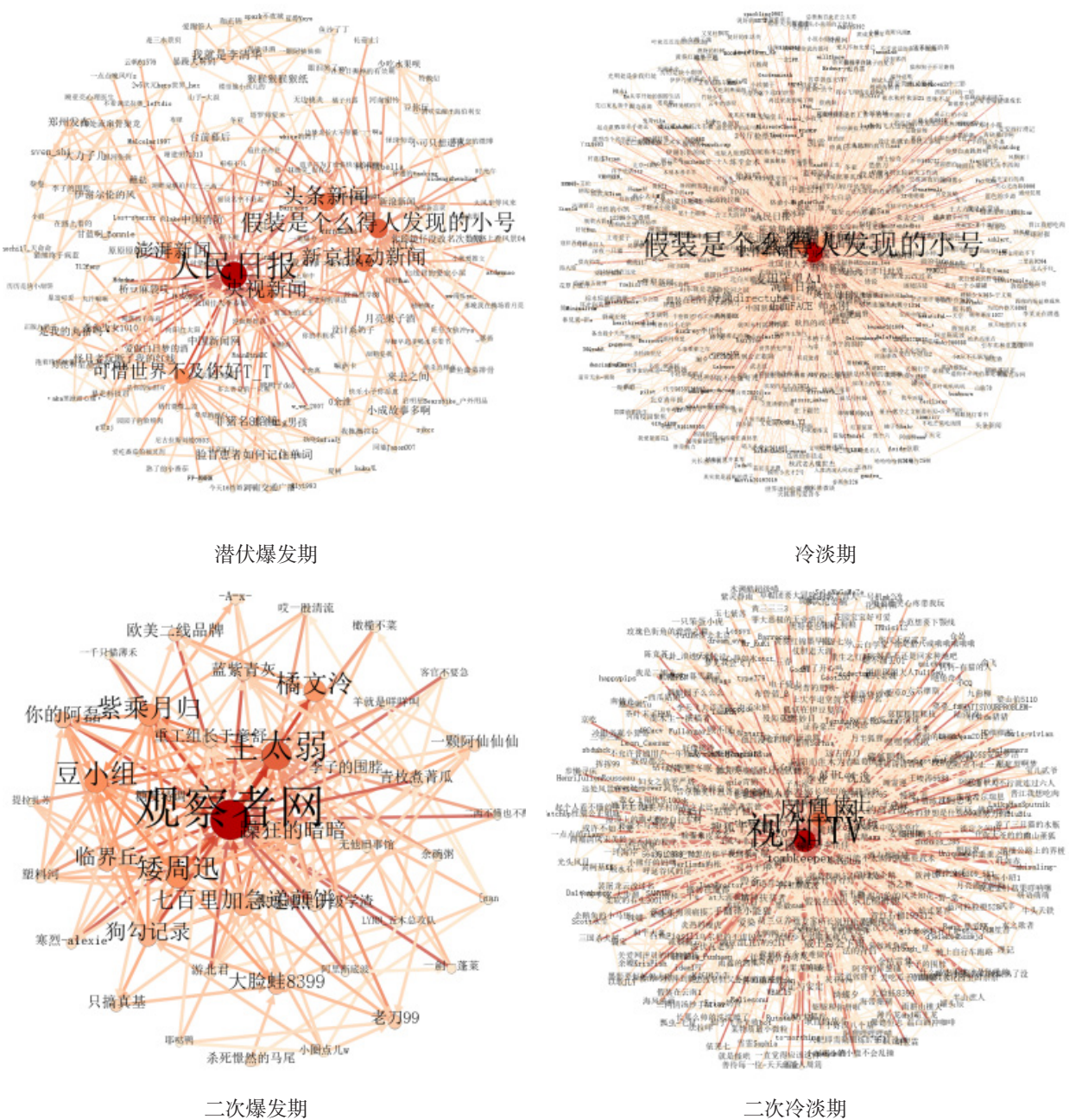


图4 各时期转发网络图

表2 描述性统计指标

时期	最大K核值	所含节点数	平均加权度	网络直径	图密度	平均路径长度
潜伏一爆发期	4	170	2.359	8	0.014	1.906
冷淡期	8	42	4.071	6	0.099	2.457
二次爆发期	3	522	2.55	13	0.005	4.038
二次冷淡期	4	366	2.508	8	0.007	2.358

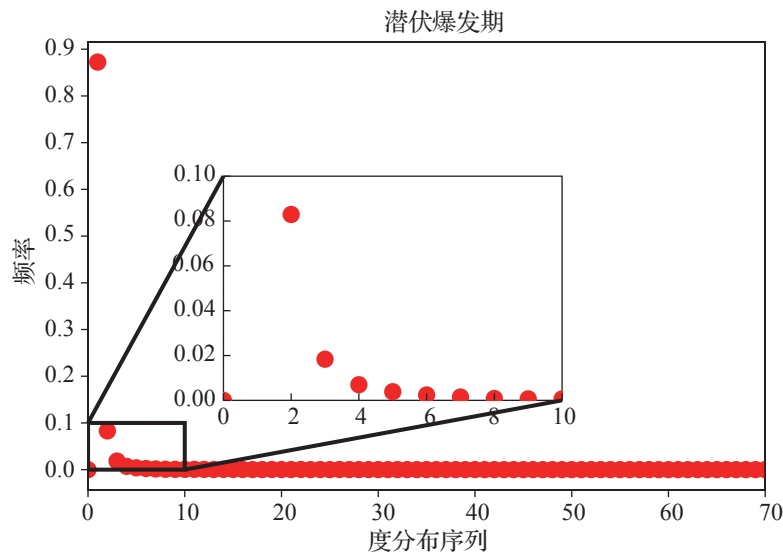


图 5 度分布

表 3 意见领袖排名

潜伏—爆发期			冷淡期		二次爆发期		二次冷淡期	
ID		得分	ID	得分	ID	得分	ID	得分
1	人民日报	4.134	观察者网	4.787	央视新闻	3.094	凤凰网	3.007
2	央视新闻	2.637	快乐十级学渣	4.323	人民日报	2.907	司马平邦	2.313
3	头条新闻	2.040	七百里加急	3.872	头条新闻	1.843	所长是我	2.001
4	一颗阿仙仙仙	1.681	狗勾记录	3.851	央证公开课	1.686	视知TV	1.787
5	中国新闻网	1.655	紫乘月归	2.854	3号厅检票员工	1.594	子午侠士	1.770
6	澎湃新闻	1.580	一颗阿仙仙仙	2.854	新浪新闻	1.568	罐头辰	1.684
7	珈正锦	1.552	塑料河	2.756	环球时报	1.331	大公文匯網	1.648
8	我就是李清华	1.507	王太弱	2.483	我是落生	1.328	假装在纽约	1.579
9	中国消防	1.507	-A-x-	2.478	中国新闻网	1.323	沈逸	1.558
10	新浪新闻	1.333	临界丘	2.413	封面新闻	1.320	沸点视频	1.496

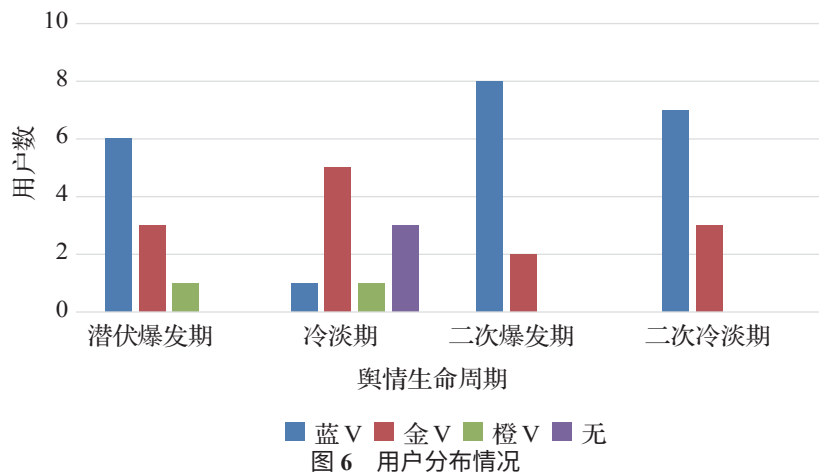


图 6 用户分布情况

但综合考量活跃度、影响力、权威性后并未跻身于前十位意见领袖中。综上,可认定为爆发期的意见领袖多以央视新闻、人民日报为代表的主流新闻媒体为主,而冷淡期的意见领袖多以个人或企业所运营的自媒体用户为主。即认为在突发事件舆情中,意见领袖主要由主流新闻媒体与自媒体两类组成,且爆发期以主流新闻媒体为主、冷淡期以自媒体为主。另外,主流新闻媒体、自媒体用户的热度、影响力远大于事件当事人这类个人非认证用户。

为说明 KSIN 算法的有效性与不同之处,以潜伏—爆发期的意见领袖排名为例,表 4 分别给出按 KSIN 算法、按粉丝数量与按 K-Shell 分解法进行排名的结果。其中,粉丝数量是微博等社交平台用户排行榜的主要依据。可以看出,K-Shell 分解法排序给出的是同一核层用户的集合。KSIN 算法在此基础上给出各个用户的 KSIN 得分,可进一步区分用户的重要度。其中,与粉丝数量排序相比,在 KSIN 算法排序中前三位与之相同。另外,用户“@一颗阿仙仙仙”提升了 4 个位次,而“@新浪微博”下降了 5 个位次。查看其在此次事件所发布的博文,有以下几点不同。首先,博文发布时间不同。用户“@一颗阿仙仙仙”的博文发布时间为 7 月 20 日 20 时,早于用户“@新浪微博”,更接近于事件爆发点,其微博被转发 1 470 余次,而“@新浪微博”的博文仅被转发 230 余次。其次,转发网络中相邻用户不同。用户“@一颗阿仙仙仙”的微博被同为认证用户的“@青崖寻酒”转发,二者均为金 V 认证的自媒体用户,且垂直领域相同。“@青崖寻酒”通过转发这一交互行为贡献其以转发比例为权重的自

身重要度。这印证了用户的重要度不仅与其自身有关,也与其邻居用户有关。同时也说明意见领袖具有同领域的特征,同领域用户之间的相互转发行为,可达到更广的受众范围。综上,KSIN 算法可以相对全面地考虑多方因素,有效地评估用户重要度,从而准确识别意见领袖。

表 4 排名对比

排名	按KSIN算法 排序	按粉丝数量 排序	按K-Shell分解法 排序
1	人民日报	人民日报	人民日报、澎湃新闻等5位用户
2	央视新闻	央视新闻	央视新闻、中国消防等5位用户
3	头条新闻	头条新闻	
4	一颗阿仙仙仙	中国新闻网	
5	中国新闻网	新浪新闻	
6	澎湃新闻	澎湃新闻	
7	珈正锦	中国消防	
8	我就是李清华	一颗阿仙仙仙	
9	中国消防	我就是李清华	
10	新浪新闻	珈正锦	

3.3 识别效果评价

为进一步地说明本文方法的有效性,采用节点核心率(Coverage Ratio)^[18]将本文方法与 K-Shell 分解法的识别结果进行对比,具体公式如下:

$$\text{Coverage Ratio} = \frac{\text{前}n\%\text{关键节点覆盖节点数}}{\text{总节点数}}$$

其中,关键节点覆盖节点数指与该节点直接相邻的节点数,节点核心率可以对关键节点影响力进行描述,通过比较 n 值较小时曲线的上升率对识别效果进行分析。具体来讲,曲线上升越快,说明该方法识别出的关键节点所覆盖的节点越多,其识别效率越高、识别效果越好^[18]。以潜伏爆发期为例,将本文识别方法

KSIN 与 K-Shell 分解法的识别效果进行对比,如图 7 所示。可以看出,两条曲线均在 0~10% 时较为陡峭,斜率较大;在 10% 之后开始趋于平缓,斜率减小。其中,在本文方法中,10% 的用户覆盖了全体用户的 88%;而在 K-Shell 分解法中,10% 的用户覆盖了全体用户的 75%。因此,本文方法的节点核心率曲线较 K-Shell 分解法更为陡峭、斜率更大、识别效率更高,可以有效地识别微博平台中突发事件的意见领袖。

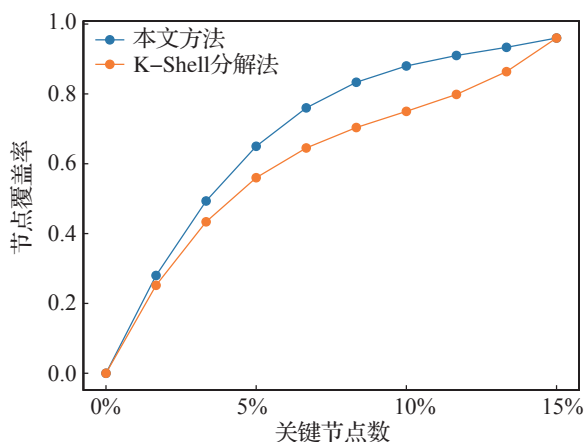


图 7 识别效果对比

4 结论与建议

本文在 K-Shell 分解的基础上,引入了用户活跃度、用户权威性与用户影响力三个维度所度量的用户重要度因子,并以转发比例为权重定义了相邻用户的重要度贡献力,从而构建了突发事件下意见领袖识别的 KSIN 算法。结论表明,该方法能够更全面地评估用户重要度,识别意见领袖。突发事件中意见领袖主要以主流新闻媒体与自媒体两类为主,突发事件的相关信息往往先流向主流新闻媒体,再流向自媒体用户与大众;主流新闻媒体、自媒体用户的热度、影响力远大于事件当事人这类个人非认

证用户;且突发事件中的自媒体类意见领袖具有同领域的特征。突发事件的网络舆情治理应当针对不同类型不同级别的微博用户实施相对应的介入措施,实现分级化管控。据此,将提出以下建议。对于主流新闻媒体微博用户,应当采取推动其引导舆论方向、聚焦于评论等介入措施。这类用户通常是机构认证用户,具有粉丝基础庞大、发布信息客观且影响力大、不倾向于输出主观看法等特征,且在突发事件传播中承担发布事件事实、追踪事件报道的角色。对于自媒体用户,应当采取常态化监控其产出内容为主、在不当言论出现时以干预为辅的介入措施。自媒体用户是指在微博等平台上以现代化、电子化的手段向其追随者传播自主化、私有化信息的一类用户。其具有在个人垂直领域中有一定影响力、倾向于对某一突发或热点事件进行转发使其获得更多的曝光并发布个人主观看法的特征。

本文也存在一定的局限性与不足,如只考虑了用户的位置属性、自身属性、行为属性等因素,并未涉及用户博文评论文本的分析。在未来的研究中,将以博文评论文本的主题建模展开研究,以检测特定突发事件中的子话题,从而挖掘出群众对特定突发事件的不同关注热点。

参考文献

- [1] Lazarsfeld P F, Berelson B, Gaudet H. The people's choice [M]. Oxford, England: Duell, Sloan & Pearce, 1944.
- [2] 中华人民共和国突发事件应对法. [EB/OL] <http://www.gov.cn>.
- [3] 李纲, 李阳. 关于突发事件情报失察的若干探讨 [J]. 情报理论与实践, 2015, 38(7):1-6.

- [4] Miorand I D, Pellegrini F D. K-shell Decomposition for dynamic complex networks [C] // Proceedings of the 8th International Symposium on Modeling and Optimization in Mobile, Ad-Hoc and Wireless Networks (WiOpt). Piscataway: IEEE, 2010: 488-496.
- [5] 康伟. 基于 SNA 的突发事件网络舆情关键节点识别——以“7·23 动车事故”为例 [J]. 公共管理学报, 2012, 9(3): 101-111+127-128.
- [6] 周飞, 高茂庭. 基于 PageRank 的网络社区意见领袖发现算法 [J]. 计算机工程, 2018, 44(02):203-209+219.
- [7] 彭丽徽, 李贺, 张艳丰. 基于灰色关联分析的网络舆情意见领袖识别及影响力排序研究——以新浪微博“8·12 滨海爆炸事件”为例 [J]. 情报理论与实践, 2017, 40(9): 90-94.
- [8] 彭丽徽, 李贺, 张艳丰. 基于 SNA 与模糊 TOPSIS 的网络舆情关键节点识别分类模型研究 [J]. 现代情报, 2017, 37(8): 17-25.
- [9] 陈芬, 高小欢, 彭玥, 等. 融合文本倾向性分析的微博意见领袖识别 [J]. 数据分析与知识发现, 2019, 3(11):120-128.
- [10] 王佳敏, 吴鹏, 陈芬, 等. 突发事件中意见领袖的识别和影响力实证研究 [J]. 情报学报, 2016, 35(2): 169-176.
- [11] Cha M, Haddadi H, Benevenuto F, et al. Measuring user influence in twitter:the million follower fallacy [J]. IC-WSM, 2010(10): 10-17.
- [12] 孙国梓, 仇呈燕, 李华康. 基于线性加权的微博影响力量化模型 [J]. 四川大学学报 (工程科学版), 2016, 48(01):78-84
- [13] Kitsak M, Gallos L K, Havlin S, et al. Identification of influential spreaders in complex networks [J]. Nature physics, 2010, 6(11): 888-893.
- [14] 张仰森, 郑佳, 唐安杰. 基于多特征融合的微博用户权威度定量评价方法 [J]. 电子学报, 2017, 45(11): 2800-2809.
- [15] 熊涛, 何跃. 微博转发网络中意见领袖的识别与分析 [J]. 现代图书情报技术, 2013(6): 55-62.
- [16] Gooseeker. Metaseeker. [EB/OL]. <https://www.gooseeker.com/product>.
- [17] 吴江, 赵颖慧, 高嘉慧. 医疗舆情事件的微博意见领袖识别与分析研究 [J]. 数据分析与知识发现, 2019, 3(4): 53-62.
- [18] 李纲, 李显鑫, 巴志超, 等. 微信群信息交流网络中的关键节点识别研究 [J]. 情报理论与实践, 2018, 41(7): 65-71.