

面向网络公害社群的社交网络特性分析方法

王博, 郭捷, 邱卫东, 黄征

(上海交通大学网络空间安全学院, 上海 200240)

摘要: 随着网络的普及和网民群体的壮大, 网络水军、网络公害社群等互联网公害组织严重威胁我国网络环境的安全。针对这类组织远离国内监管、隐蔽性和伪装性强的特点, 提出了面向网络公害社群的社交网络特性分析方法。在用户特性层面, 通过中心性测量对组内成员的影响力分布进行计算, 找出影响力最大的核心用户, 并分析核心成员的匿名性特点。在复杂网络特性层面, 通过小世界效应和无标度特性的分析, 研究了该社群内部的集群状况和资源分布状况。在时序特性层面, 对核心成员的发文内容进行预处理后, 利用BERTopic进行动态主题建模, 得到社群主题随时间的变化关系。大量的实验性能分析验证了本文方法的有效性, 研究成果可用于分析公害社群相关特征, 追踪犯罪活动轨迹, 为公害组织的侦查和打击工作以及自动化检测系统的设计工作提供理论参考。

关键词: 互联网公害; 影响力测量; 网络特性分析; 动态主题建模

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2024041

Analysis method on the properties of social networks of internet nuisance communities

WANG Bo, GUO Jie, QIU Weidong, HUANG Zheng

School of Cyber Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract: With the widespread adoption of the Internet and an increase in the number of Internet users, the security of the network environment has been seriously threatened by Internet public hazards organizations, such as the Internet water army and the internet public hazards community. Given the characteristics of these organizations, which include being distant from domestic supervision and possessing strong concealment and camouflage abilities, a social network characteristic analysis method was proposed for Internet public hazards organizations. Initially, at the user characteristic level, the influence distribution of group members was calculated through centrality measurement to identify the core users with the greatest influence, and an analysis of the anonymity characteristics of these core members was conducted. Subsequently, at the complex network characteristic level, the cluster status

收稿日期: 2023-02-28; 修回日期: 2023-09-01

通信作者: 郭捷, guojie@sjtu.edu.cn

基金项目: 国家自然科学基金项目 (61972249)

Foundation Item: The National Natural Science Foundation of China (61972249)

引用格式: 王博, 郭捷, 邱卫东, 等. 面向网络公害社群的社交网络特性分析方法[J]. 网络与信息安全学报, 2024, 10(3): 81-90.

Citation Format: WANG B, GUO J, QIU W D, et al. Analysis method on the properties of social networks of internet nuisance communities[J]. Chinese Journal of Network and Information Security, 2024, 10(3): 81-90.

and resource distribution within the community were studied through the analysis of small-world effects and scale-free characteristics. Finally, at the temporal characteristic level, after preprocessing the post content of core members, BERTopic was utilized for dynamic topic modeling to ascertain the relationship between community themes and time. A large number of experimental performance analyses have verified the effectiveness of the proposed method. The research results can be utilized to analyze the characteristics of public hazard communities, track criminal activity trajectories, and provide theoretical references for the investigation and crackdown of public hazard organizations, as well as for the design of automated detection systems.

Keywords: internet hazard, influence measurement, network property analysis, dynamic topic modeling

0 引言

近年来,中国互联网络飞速发展,截至2022年6月,中国网民规模达到10.51亿人,互联网普及率达74.4%^[1]。然而,包括谣言泛滥、侵犯隐私和群体极化在内的网络公害层出不穷,严重威胁网民正常生活。由于网络环境具有高度的灵活性和交互性,虚假、有害的信息很容易泛滥,仅在2020年,微博(Weibo)官方共处理了76 107条虚假信息^[2],网络水军^[3]、电信诈骗^[4]也严重扰乱了社交网络的正常秩序,此外,由于社交网络的“回音室”“信息茧房”等效应,群体意见可能会向两极发展,容易被网络暴力等网络恐怖主义利用^[5]。而随着社交平台的普及,部分活跃于Twitter、Facebook等社交软件上的组织和团体,借助诸如国际公益组织、非政府组织的外衣,从事黑客活动^[6]等犯罪行为,在输出不实消息、制造焦虑和恐慌等方面具有更强的伪装性、欺骗性和危害性^[7]。

没有网络安全就没有国家安全,相比较诈骗、水军等公害团伙,部分网络恐怖主义组织具有极强的误导性的同时有一定的隐蔽性,亟待从社会计算角度进行严格定性和定量的研究。

本文以某一公害社群为例^[8],利用中心性计算成员的影响力,找出核心成员。在此基础上,通过网络特性分析来研究网络社区的宏观特性,通过动态主题建模对社区讨论内容进行时间序列分析,把握其活动轨迹,突破了传统方法无法追踪主题变化的局限性,对社群能够更加深入了解。本文对网络公害社会网络特性的分析,将为网络公害组织的侦查打击工作、互联网公害检测系统的设计工作提供理论参考。

1 研究现状

互联网公害吸引了国内外大量学者进行研究。石俊峥^[9]提出了面向加密移动流量的网络行为分析技术,在保护隐私的同时打击了谣言、暴恐等公害行为。Campana等^[10]对在线有组织的犯罪团伙进行了研究,讨论了声誉方法、社区检测等成员划分标准,指出该类问题分析的关键是团伙内部的等级制度和影响力。

在社交网络中,高影响力的关键用户会对整个社群的舆论环境产生重要影响,因此对社群影响力分布的分析具有重要意义。Jones等^[8]利用度中心性、中介中心性等中心性(centrality)度量标准对Anonymous黑客社群影响力进行了分析,证实了该组织影响力仅掌握在少数人手中。

复杂网络特性分析作为宏观层面的网络分析法,被应用到许多场景的网络分析中,如吕芳等^[11]通过对银行账户交易网络的定性分析,证实了该网络具有无标度、小世界、部分自组织和奇异吸引子等特性。

主题建模技术在诸多领域被应用于从非结构化文本中推断潜在主题。Kigerl等^[12]使用潜在狄利克雷分布(latent Dirichlet allocation, LDA)对非结构化文本进行聚类后,与结构化的数据结合到另一个聚类模型中以实现数据的分层聚类。动态主题建模可以在对给定的文档推断出潜在主题的同时,计算出主题随时间的变化。Grootendorst等^[13]提出了BERTopic动态主题建模算法,在主题连贯性、主题一致性多项指标方面均具有竞争力。

2 互联网公害社交网络分析方法

2.1 互联网公害社群的抽象数学模型

在Twitter等社交平台上,用户之间通过“关注”这一关系来订阅对方发布的内容。据此,可

以根据该关系,使用社交网络 G 来抽象化表示该社交网络,账户被视为节点,“关注”被视为从追随者指向被关注者的有向边,该社交网络 G 的定义如下。

$$G = (V, E, W)$$

其中, V 和 E 分别表示节点和边的集合, W 表示 E 根据连接的强度或容量对应的权重集合。在以“关注”关系构建的网络中, W 的值是1,可以从 G 中去除。

社交网络包含众多节点,而联系紧密的节点又会聚集成小的团体,要全面、客观地了解互联网公害社团网络,就需要兼顾微观和宏观,即用户的个体行为和网络的整体特性。此外,对社群讨论内容进行时序主题分析,有助于把握社群动态。互联网公害社群社交网络分析总览如图1所示,本文将从用户特性、复杂网络特性和时间序列特性3个方面对互联网公害社群进行分析。

2.2 用户特性分析

2.2.1 基于中心性测量的用户影响力分析

对于互联网公害社群,找出其“头目”或者“意见领袖”等拥有高影响力的用户至关重要。在社交网络中,通常使用中心性来衡量某个角色在整个社交网络中的位置,较高的中心性意味着可以更容易、更迅速地访问到其他成员,因此这部分成员也被称为“焦点”或“关键用户”。为了确定互联网公害社群中的“意见领袖”,本文对常见的中心性标准进行了考察。

(1) 度中心性

度中心性(degree centrality)通过节点的度数来衡量其中心性,也常常被视作“声望”。当

前节点的度越大,影响力越高,在社群网络中越占据重要的位置。对于有向图,节点的度中心性包括入度中心性和出度中心性,数学表达式为:

$$k_i^+ = \sum_{j=1}^n a_{ji} (i \neq j) \quad (1)$$

$$k_i^- = \sum_{j=1}^n a_{ij} (i \neq j) \quad (2)$$

其中, k_i^+ 和 k_i^- 分别表示有向图 G 节点 i 的入度中心性和出度中心性, n 为图 G 中的节点数, a_{ji} 表示从节点 j 指向节点 i 的边数, a_{ij} 表示从节点 i 指向节点 j 的边数。

(2) 中介中心性

中介中心性(betweenness centrality)度量了一个节点在网络中充当中介的程度,高中介中心性的节点在网络中占据了关键位置,起到了沟通几个小团体的“桥梁”作用,数学表达式为:

$$b_v = \sum_{s,t \in V(G) \setminus v} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (3)$$

其中, σ_{st} 表示节点 s 和节点 t 之间的最短路径数目, $\sigma_{st}(v)$ 表示其中经过节点 v 的最短路径数目。

(3) 特征向量中心性

特征向量中心性(eigenvector centrality)衡量了当前节点与其他良好连接节点的连接情况,它由邻接矩阵的第一特征向量计算得到,其基本思想为:当前节点的影响力与和它密切相关的节点的影响力有关,即当前节点 i 的中心性正比于所有邻居中心性之和,数学表达式如下。

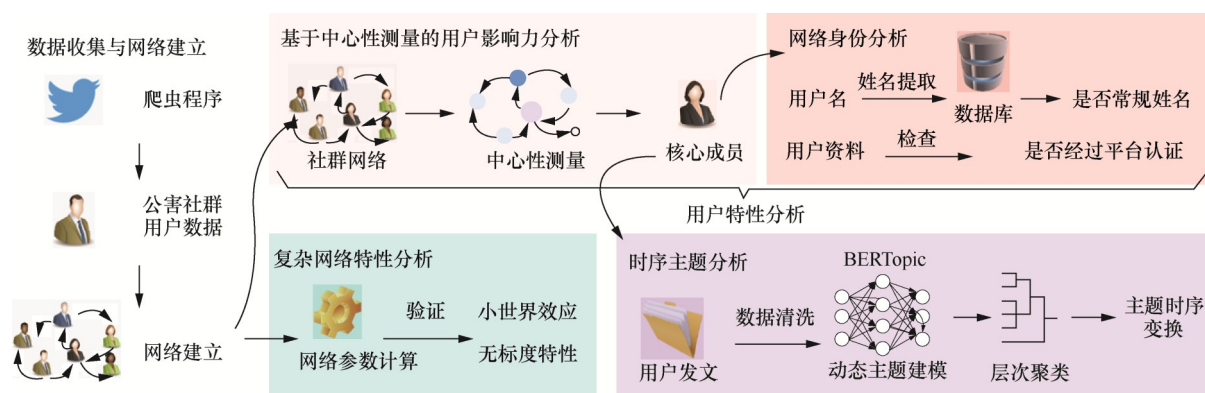


图1 互联网公害社群社交网络分析总览

Figure 1 Overview of social network analysis of Internet hazard communities

$$x_i = \frac{1}{\lambda} \sum_{j=1}^n a_{ij} x_j \quad (4)$$

其中, x_i 和 x_j 表示节点 i 和节点 j 的特征向量中心性, a_{ij} 是邻接矩阵 A 中的数据项, 代表节点间的连接情况, λ 是邻接矩阵 A 中最大的数据项。

(4) 佩奇排名中心性

佩奇排名中心性 (page rank centrality) 是衡量网络中节点重要性的指标, 它的核心思想是, 来自受欢迎节点的跳转应该比来自不受欢迎节点的跳转更为重要。社交网络中的有向线段可以看作节点之间的跳转, 每个节点都可以由它的追随者跳转来, 也可以继续跳转到它的关注者。该过程中, 可以用转移矩阵 M 来表示整个网络的跳转情况, 第 i 行第 j 列表示从节点 j 跳转到节点 i 的概率, 网络中所有节点的佩奇排名可以用向量 v 表示, 并且可以通过与转移矩阵相乘计算新的佩奇排名, 佩奇排名通过迭代计算直至收敛, 数学表达式为:

$$p' = dMp + e \frac{1-d}{n} \quad (5)$$

其中, p' 是本次得到的佩奇排名, p 是上次的佩奇排名, n 是节点个数, M 是转移矩阵, e 是单位矩阵, d 是阻尼系数, 代表随机跳转到不在关注列表中的新用户的概率。

2.2.2 网络身份分析

对于互联网公害群体而言, 为了方便实施非法行动, 成员大多不愿透露自己的身份。本节从姓名数据库认证和官方平台认证两个方面考察核心成员的网络身份。

(1) 姓名数据库认证

为考察用户是否愿意在用户名中使用常规姓名, 本文通过姓名数据库对用户名进行检查, 该数据库包含 105 个国家的 98.3 万个姓氏、73 万个名字, 以判断用户是否使用常规的姓名作为自己的账号名称。

(2) 官方平台认证

部分用户的身份得到了社交平台官方的认证, 在一定程度上具有真实性、可信性。

本节分别计算公开姓名用户和经过社交平台认证的用户所占比例 U_{pub} 和 P_{pub} , 并在整个社交平台随机抽样作为平均水平 U_{avg} 和 P_{avg} , 若 $U_{pub} < U_{avg}$ 且 $P_{pub} < P_{avg}$, 说明核心成员倾向于隐藏自己身份。

2.3 复杂网络特性分析

对于互联网公害组织, 研究成员的内部集群情况以及资源分布状况具有重要意义。真实世界中的复杂网络具备一些不同于随机网络的独特性质, 复杂网络具有自组织、自相似、吸引子、小世界、无标度中部分或全部性质。

本文通过研究 GROUP 群体的小世界效应, 分析该团体内部集群程度和消息传递的效率。通过验证群体内部的无标度特性, 分析社群的人脉资源、信息资源的掌握分布情况。

2.3.1 小世界效应

Milgram^[14]第一个指出现实社会网络中存在小世界效应, 即一对素不相识的个体可以通过几个熟人联系在一起, 也就是人们所熟知的六度空间理论。小世界特性可以通过聚集系数 (clustering coefficient) 和平均最短路径两个特征来体现。

(1) 聚集系数

聚集系数描述了节点之间联系的紧密程度, 网络的全局聚集系数计算思想^[15]为, 网络中组成闭三点组 (三点两两相连) 的点占组成闭三点组和开三点组 (缺一条边的三角形) 的点总和的比例, 数学表达式为:

$$C_{global} = \frac{3G_1}{3G_1 + 2G_2} \quad (6)$$

其中, G_1 为图中三角形的个数, G_2 为图中缺一条边的三角形的个数。

除全局聚集系数之外, 通常还使用局部聚集系数 C_a 来衡量某些节点的局部聚类程度。一个节点的局部聚集系数定义为该节点邻域内节点之间真实存在的连边与所有可能连边数之比, 所有节点局部聚集系数的均值即网络平均聚集系数。

(2) 平均最短路径

对于社交网络图 G , 节点 i 和节点 j 之间的最短路径 d_{ij} 定义为从节点 i 到节点 j 需要经过的最少边数, 平均最短路径 L 即所有节点对之间最短路径的均值。

随机网络的平均路径长度为 L_{rand} , 全局聚集系数为 $C_{rand, global}$, 平均聚集系数为 C_{rand} , 根据 Sporns^[16]的小世界判断标准, 验证社交网络的小世界特性, 需要满足 $\frac{C_a}{C_{rand}} > \frac{L}{L_{rand}}$ 。

2.3.2 无标度特性

在随机网络中,边以恒定的概率出现,因此大多数的顶点具有相似或者相等的度数,度分布呈现二项分布或者泊松分布。

然而,不同于随机网络,在真实世界的社交网络图中,节点的度的分布非常不均匀,绝大多数的节点度数较低,少部分节点度数较高^[17]。这种度数的分布被称为幂律分布,网络的这种特性也被称为无标度特性,用数学公式表达为:

$$P(k) = ck^{-r} \quad (7)$$

其中, $P(k)$ 是度数为 k 的节点在社交网络中所占的比例, c 为幂律分布系数, r 为幂律分布指数。

网络的无标度特性由节点符合幂律分布判定,验证网络的无标度特性,即验证 $P(k) \sim k^{-r}$ 。

2.4 时间序列特性分析

为了更好地掌握互联网公害群体的动向,其在线讨论内容是一个重要的切入点,以时间为主线对讨论内容主题进行分析,既可以掌握其犯罪证据,又对预测社群的未来动向具有指导意义。

2.4.1 数据清洗

为了得到更好的主题聚类结果,在动态主题建模之前,需要对用户的推文进行清洗,如过滤掉特殊符号、人称代词等。数据清洗的步骤如下。① 利用谷歌翻译将所有非英语文本翻译为英语文本。② 将@用户内容和网址链接统一去除。③ 对#后的标签进行分词处理。④ 将所有的字符改为小写。⑤ 过滤掉所有的表情符号。⑥ 删除所有的停用词,包括英语常见的停用词(如人称代词)和社交媒体中常见的“RT”转发标记等。⑦ 将所有文本按照月份分组。

2.4.2 动态主题建模

对处理好的文本,本文采用BERTopic^[13]进行动态主题建模。为减少主题数量以方便查看,采用Ward层次聚类^[18]对相似的主题进行合并,最终得到主题随时间的动态变化关系。

BERTopic主题建模需要3个步骤来计算得到文本的主题集群。① 使用预训练的孪生双向编码器表征^[19](siamese bidirectional encoder representation from transformers, SBERT)模型得到句子的嵌入向量。② 使用均匀流形近似和投影^[20](uniform manifold approximation and projection,

UMAP)来降低向量的维数。③ 通过基于类的词频-逆文档频率^[21](term frequency-inverse document frequency)变体算法提取全局主题。

在全局主题的基础上,将每一个时间段内主题词频率与全局逆文档频率值相乘得到每个时间段的主题分布,经过层次聚类合并相似主题后,得到精简后的每个主题随时间变化的关系。

3 互联网公害社会网络分析实验

3.1 数据收集和网络构建

本文研究对象是近年来活跃在境外社交平台上的一个以网络暴力标签聚拢的社群,为表述方便,下文均以GROUP代指。

GROUP起源于一次网络暴力事件,随后事件的参与者联系越来越密切,社群也达到一定的规模,有了名字和专门的官方账号。该社群前期散布虚假消息,误导民众,后期输出暴力手段,扰乱治安,威胁居民的人身安全,已然成为网络空间的一大公害。

本文利用社交平台的开发者API对该社群2020年7月~2023年1月的有关推文进行收集,并爬取作者的关注者和追随者,得到共计604 837名用户,选取其中带有社群名称的发文用户,经过人工筛查用户资料,最终建立包含1 762个节点、35 695条有向边的GROUP核心社区。

3.2 用户特性分析实验

3.2.1 中心性

本文选取度中心性、中介中心性、特征向量中心性和佩奇排名中心性这4项指标,通过中心性测量来计算每个节点在整体社交网络中的影响力,主要回答两个问题:一是该网络社区的影响力分布是否均匀,即是否存在影响力较大的“意见领袖”;二是各项指标的计算结果是否一致,即计算出的“意见领袖”是否在多个标度上都有很高的数值。

GROUP的网络社区中心性分布如图2所示。GROUP的网络社区各用户的中心性统计如图3所示。GROUP的网络中心性分布极为不均衡,极少数账户占有较高的影响力,而大多数的账户影响力较低。GROUP用户中心性得分排名如表1所示,本文选取每个指标得分最高的6名用户进行对比,共得到13名用户。所有用户使用

“User” 和 ID 的后 4 位表示。 k 、 b 、 x 、 p 分别代表度中心性、中介中心性、特征向量中心性和佩奇排名中心性，4 项指标的计算结果基本一致，该社交网络的“意见领袖”在多个指标上取得较高分数。

例如，平均得分最高的用户“User9361”有两项指标排名第一，拥有 5.3 万粉丝，截至 2023 年 1 月，在账号建立 2 年多的时间内发布了 6 117 条推文，其中 596 条属于转发、引用和评论消息。从其影响力的得分上看，度中心性 k 和中介中心性 b 均排名第一，这得益于该用户有着巨大的受众群体，并且在一定程度上充当了沟通桥梁，这些从用户的关注者数量、发文数量以及转发数量中可以得到印证。

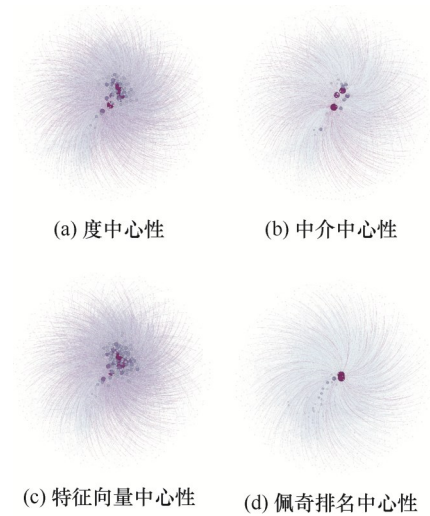


图2 GROUP 的网络社区中心性分布
Figure 2 Distribution of centrality in GROUP

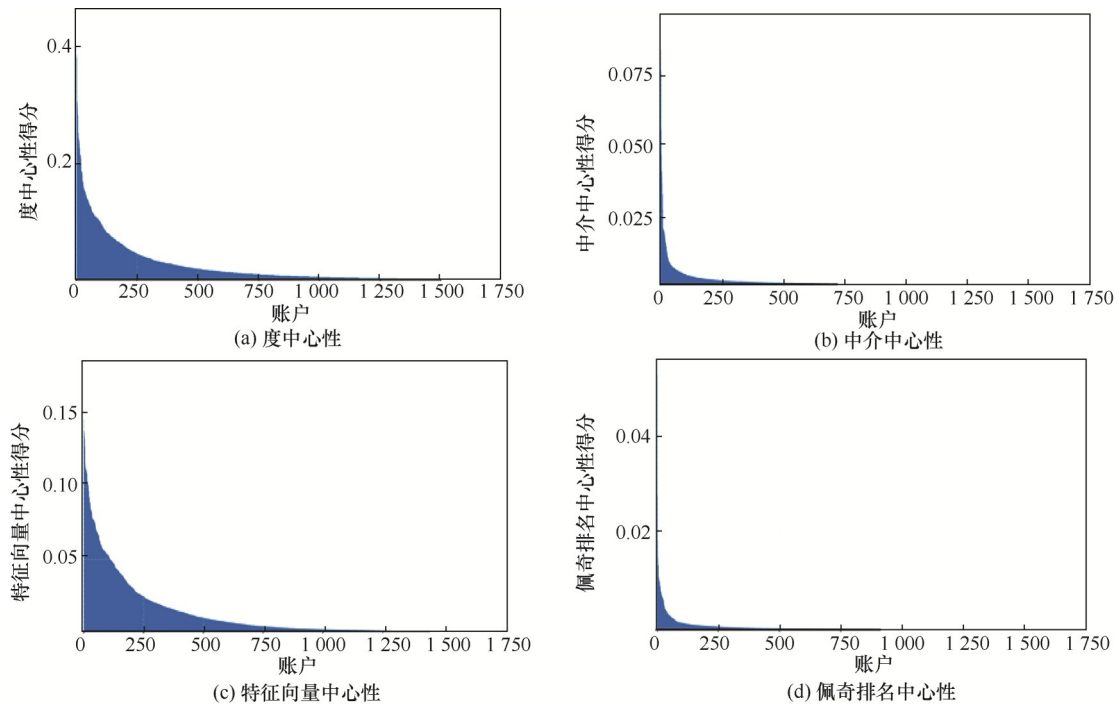


图3 GROUP 的网络社区各用户的中心性统计
Figure 3 Centrality statistics for each user in GROUP

表 1 GROUP 用户中心性得分排名
Table 1 Centrality score rankings of users in GROUP

指标	得分前 6 名用户(从高到低排列)
k	User9361 User1091 User7684 User6288 User7136 User7914
b	User9361 User7684 User7136 User6288 User1091 User6768
x	User1091 User6288 User9361 User4306 User3025 User7684
p	User4306 User9495 User9361 User3065 User0897 User8997
平均	User9361 User7684 User1091 User6288 User4306 User7136

3.2.2 用户身份分析

对于公害组织掌握核心成员的基本信息有助于判断该社群的基本情况，本节对 GROUP 影响力最高的 13 名用户公开资料进行分析。

在 13 名用户中，仅有 4 名用户使用常规姓名作为用户名，其余用户则选用组织名称中的单词，或者其他有实际意义的英语单词作为用户名。在平台认证方面，仅有 1 名用户经过了社交网络平台的认证，其余核心成员均未认证。GROUP 核心成员的常规姓名率为 30.77%，认证率为 7.69%。

作为对比，本文在整个社交网络平台随机采样 20 133 名用户，结果显示，有 7 958 名用户公开自己的姓氏和名字，2 299 名用户经过平台认证。常规姓名率为 39.53%，认证率为 11.42%。

该组织的核心成员的常规姓名率和认证率均低于整个网络的平均水平，从该社群性质的角度，这很容易解释，即互联网公害群体需要尽量隐藏自己的身份或者伪造身份，以便于在网络空间非法活动。

3.3 复杂网络特性分析实验

3.3.1 小世界效应分析实验

本节通过测定 GROUP 的社交网络的平均最短路径和全局聚集系数来分析该社群中的小世界效应。作为对照，本节同时生成与该社群节点数、边数相当的随机网络，测量其平均最短路径和聚集系数，随机图生成 10 次，最终取 10 次的平均值。

根据定义，本节计算得到网络的平均最短路径长度 $L = 2.616 5$ ，这意味着在该社群中，任意互不认识的两个成员中间仅相隔 1~2 名成员，可见小世界效应在该社区中得到显著体现。

根据聚集系数的定义，该网络的平均聚集系

数 C_a 为 0.322 5，全局聚集系数 C_{global} 为 0.178 1，网络的局部聚集程度略大于网络的全局聚集程度，说明与该社群网络的平均水平相比，部分用户之间的关系更为紧密。

同等规模随机网络的小世界效应相关指标测量结果如表 2 所示。同等规模的随机网络中，测得平均最短路径长度 $L_{\text{rand}} = 2.648 9 \pm 0.004 7$ ，随机网络平均聚集系数 $C_{\text{rand}} = (1.488 8 \pm 0.039 6) \times 10^{-2}$ ，全局聚集系数 $\bar{C}_{\text{rand,global}} = (1.489 0 \pm 0.038 4) \times 10^{-2}$ 。与 GROUP 相比，随机网络具有略长的平均最短路径和更小的聚集系

数。经计算验证可得， $\frac{C_a}{C_{\text{rand}}} = 21.661 7$ ， $\frac{L}{L_{\text{rand}}} = 0.987 8$ 。GROUP 满足条件 $\frac{C_a}{C_{\text{rand}}} > \frac{L}{L_{\text{rand}}}$ ，具备小世界特性。

3.3.2 无标度特性分析实验

本节对 GROUP 社交网络中的无标度特性进行分析，研究用户拥有的粉丝数和用户发文量是否符合幂律分布。首先，对网络中节点的入度（即粉丝数）进行了统计，在实际计算中，因为粉丝数极差太大，采用 100 的刻度进行计数以方便统计，并分别绘制了散点图和双对数散点图。从图 4(a)、图 4(b)可以看出，粉丝数分布在对数刻度下呈现出衰减趋势，符合幂律分布。在互联网公害组织中，这种现象很容易解释，即新加入的成员总是会优先关注拥有大量受众的“意见领袖”，从而使“意见领袖”的听众越来越多，而普通成员的受众则始终保持在较低水平。同样地，用户的发文量在统计时以 1 000 为刻度计数，从图 4(c)、图 4(d)可以看出，用户的发文量也呈现出类似的对数关系，原因可能是该社群中存在拥有大量信息的“意见领袖”，其通过独立撰写或大量转发自己所认同的消息，达到联络成员、

表 2 同等规模随机网络的小世界效应相关指标测量结果

Table 2 Measurement results of relevant Indicators of the small-world effect of random networks of the same size

指标	次数									
	1	2	3	4	5	6	7	8	9	10
L_{rand}	2.66 091	2.65 054	2.64 799	2.65 051	2.64 750	2.64 299	2.64 668	2.64 358	2.64 967	2.64 819
$C_{\text{rand}} (\times 10^{-2})$	1.43 988	1.46 349	1.51 157	1.43 543	1.53 339	1.53 112	1.52 116	1.46 365	1.53 746	1.45 100
$C_{\text{rand}}^{\Delta} (\times 10^{-2})$	1.45 405	1.46 501	1.50 573	1.44 002	1.52 255	1.53 558	1.53 012	1.46 852	1.53 340	1.43 544

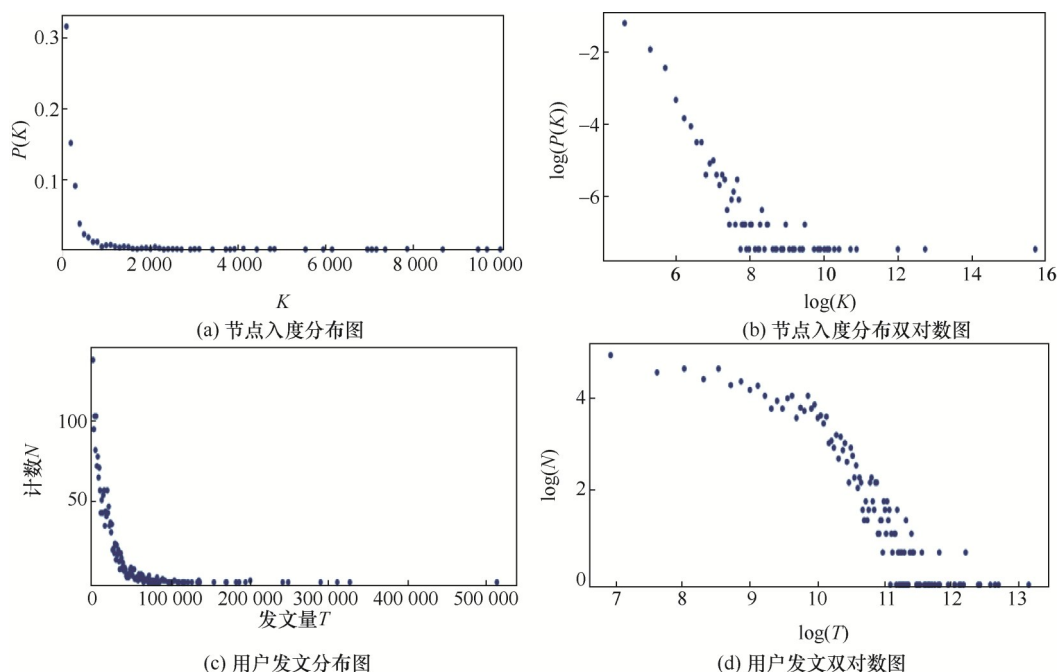


图4 GROUP中Twitter社区入度和发量分布

Figure 4 Distribution of in-degree and count of tweets for each user in the GROUP on Twitter

统一行动的目的。

综上所述, GROUP组织内部不仅粉丝数所代表的人脉资源服从幂律分布, 代表信息资源的发量也服从幂律分布, 从而说明了该社群拥有复杂网络的无标度特性。

3.4 时间序列特性分析实验

根据3.2.1节的节点中心性得分结果, 本文选取4项指标和平均得分各自最高的前6名用户作为需要关注的“关键用户”进行实验, 共有13名用户被选取。对每名用户, 本文爬取自2020年7月至2023年1月发布的全部内容, 共获得40587条推文。

本文对GROUP意见领袖每月的发量进行了统计, 结果如图5所示, 据图可见, 该社群分别在2020年秋季、2021年夏季、2021年冬季、2022年春季、2022年秋季和2022年冬季各有一次发文高峰。

对推文进行预处理后, 本文利用BERTopic进行动态主题建模, 以把握GROUP组织发布内容的主题以及变化情况, 经过层次聚类合并较小的主题后, 对每次发文高峰的主题进行分析, 并选取各高峰时段频率最高的主题词。

本节的重要发现是GROUP主题词在第一次

(2020秋季)和第二次发文高峰(2021夏季)之间发生了转变, 在查看第一次发文高峰时的推文内容后发现, 该组织发布大量的不实信息和误导性言论。

造成这样的原因可能与“意见领袖”的更换有关, 如图5所示。第一次的发文高峰与用户“User8997”的活动有关, 尽管之后其并未停止发文, 最后一次发文时间为2022年11月27日, 但随着社群的壮大, 其他地区的用户不断加入, 受众群体不断更新, 也产生了新的“意见领袖”, 从而以“User8997”为代表的核心成员在社群中的话语权和影响力减弱。

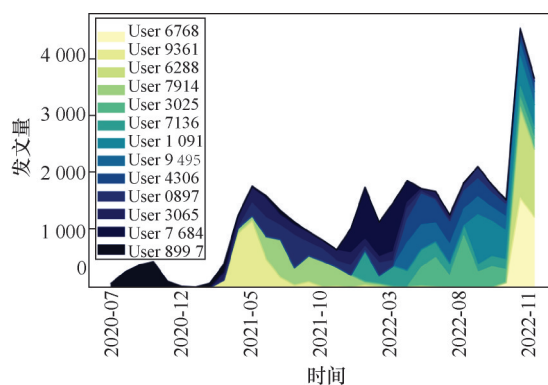


图5 GROUP社群的意见领袖每月的发量统计

Figure 5 Monthly volume of opinion leaders in GROUP

4 结束语

本文通过社会网络分析技术对互联网公害社群网络的特性进行了计算和分析。首先, 基于图论对互联网公害社群建立了数学模型。然后, 分别从用户特性、复杂网络特性和时间序列主题3个方面对公害社群展开了分析。在对真实案例GROUP的分析中, 本文发现该联盟影响力分布极不均匀, 存在影响力极高的“意见领袖”, 并且“意见领袖”为了方便在互联网的活动, 实名率和认证率低于网络平均水平。此外, 本文运用复杂网络分析验证了该网络的小世界效应和无标度特性。最后, 本文对几名“意见领袖”的发文进行了时序主题建模, 发现该社群的主题变化与“意见领袖”更换具有较强的关系。

本文分别从微观和宏观对互联网公害社群进行了社交网络特性分析, 为侦查和打击网络公害组织、设计互联网公害检测系统提供了理论参考。

后续工作, 一方面将对主题分析的工作扩展到“意见领袖”之外, 以利于加深理解主题在网络中的传播过程; 另一方面将关注网络公害社群在其他社交平台的组织和活动情况, 以对这种社群的特点有更全面的掌握。

参考文献:

- [1] 中国互联网信息中心. 第50次中国互联网发展状况统计报告[R]. 2022.
CNNIC. The 50th statistical report on china's internet development [R]. 2022.
- [2] XU S, LIU X, MA K, et al. Rumor detection on social media using hierarchically aggregated feature via graph neural networks[J]. Applied Intelligence, 2023, 53(3): 3136-3149.
- [3] HE Y, YANG P, CHENG P. Semi-supervised internet water army detection based on graph embedding[J]. Multimedia Tools and Applications, 2023, 82(7): 9891-9912.
- [4] JIANG Y, LIU G, WU J, et al. Telecom fraud detection via hawkes-enhanced sequence model[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(5): 5311-5324.
- [5] DAI J, ZHU J, WANG G. Opinion influence maximization problem in online social networks based on group polarization effect[J]. Information Sciences, 2022, (609): 195-214.
- [6] JONES K, NURSE J R C, LI S. Out of the shadows: analyzing anonymous' twitter resurgence during the 2020 black lives matter protests[C]//Proceedings of the International AAAI Conference on Web and Social Media. 2022, (16): 417-428.
- [7] BADA M, NURSE J R C. Profiling the cybercriminal: a systematic review of research[C]//2021 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA). 2021: 1-8.
- [8] JONES K, NURSE J R C, LI S. Behind the mask: a computational study of Anonymous' presence on Twitter[C]//Proceedings of the International AAAI Conference on Web and Social Media. 2020: 327-338.
- [9] 石俊峰. 基于节点时空属性的移动应用网络行为分析技术研究[D]. 北京: 中国科学院大学, 2021.
SHI J Z. Research on mobile application network behavior analysis technology based on node space-time attribute[D]. Beijing: University of Chinese Academy of Sciences, 2021.
- [10] CAMPANA P, VARESE F. Studying organized crime networks: data sources, boundaries and the limits of structural measures[J]. Social networks, 2022, 69: 149-159.
- [11] 吕芳, 张新琳, 王巍, 等. 银行账户交易复杂网络特性分析[J]. 网络与信息安全学报, 2019, 5(4): 99-107.
LYU F, ZHANG X L, WANG W, et al. Analyzing the complex properties of bank account transaction network[J]. Chinese Journal of Network and Information Security, 2019, 5(4): 99-107.
- [12] KIGERL A. Behind the scenes of the underworld: hierarchical clustering of two leaked carding forum databases[J]. Social Science Computer Review, 2022, 40(3): 618-640.
- [13] GROOTENDORST M. BERTopic: neural topic modeling with a class-based TF-IDF procedure[J]. arXiv Preprint arXiv: 2203. 05794, 2022.
- [14] MILGRAM S. The small world problem[J]. Psychology Today, 1967, 2(1): 60-67.
- [15] NEWMAN M E J, MOORE C, WATTS D J. Mean-field solution of the small-world network model[J]. Physical Review Letters, 2000, 84(14): 3201.
- [16] SPORNS O, ZWI J D. The small world of the cerebral cortex[J]. Neuroinformatics, 2004, 2: 145-162.
- [17] BARABÁSI A L, ALBERT R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509-512.
- [18] SHARMA S, BATRA N. Comparative study of single linkage, complete linkage, and ward method of agglomerative clustering[C]//2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). 2019: 568-573.
- [19] REIMERS N, GUREVYCH I. Sentence-BERT: sentence embeddings using siamese BERT-networks[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019: 3982-3992.
- [20] MCINNES L, HEALY J, SAUL N, et al. UMAP: uniform manifold approximation and projection[J]. Journal of Open Source Software, 2018, 3(29): 861.
- [21] KIM D, SEO D, CHO S, et al. Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec[J]. Information Sciences, 2019, 477: 15-29.

[作者简介]



王博 (1997-), 男, 山东潍坊人, 上海交通大学硕士生, 主要研究方向为人工智能安全、信息内容安全。



邱卫东 (1973-), 男, 江西修水人, 上海交通大学教授、博士生导师, 主要研究方向为计算机取证、密码分析、人工智能安全、大数据隐私保护。



郭捷 (1976-), 女, 河南南阳人, 上海交通大学副研究员, 主要研究方向为多媒体安全、人工智能安全、大数据隐私保护。



黄征 (1975-), 男, 四川南充人, 上海交通大学副教授, 主要研究方向为人工智能、隐私保护计算。