

# 在线热点事件的时空演变规律\*

龚凯 唐明<sup>†</sup> 尚明生 周涛

(电子科技大学互联网科学中心, 成都 610054)

(2011年7月17日收到; 2011年9月6日收到修改稿)

为了解舆论的时空演变规律并揭示其形成机理, 本文运用统计物理学方法, 通过对在线评论数据进行统计分析, 定量地研究了在线热点事件关注度(在线评论数)的时空演变规律. 实证表明, 虽然事件关注度在不同地区的分布存在极大的异质性, 即遵循双段幂律分布; 但是不同地区内事件的关注程度在时间演变过程中却表现出明显的一致性, 其不同时间内的 Zipf 分布变化很小. 关联性分析显示地区关注度受到该地区经济的显著影响, 而不同地区关注度演变行为的一致性来源于地区之间的强关联性. 另一方面, 舆论引导将显著影响事件的关注度, 导致传播速度(单位时间内评论数的增量)急剧增加. 通过计算不同地区传播速度的信息熵, 我们发现评论的地区分布在大部分时间内都具有有一致性, 而舆论引导有助于保持这种一致性. 地区传播速度之间的关联性分析表明在整个事件中经济较发达地区的舆论变化更趋于一致, 暗示这些地区对于舆论引导的响应更快, 因此加强发达地区的舆论引导有利于控制舆论的整体传播.

**关键词:** 在线舆论, 关注度, Zipf 分布, 信息熵

**PACS:** 89.20.-a, 89.75.Da, 89.70.Cf

## 1 引言

近年来, 越来越多的研究开始关注于认识与理解人类社会群集行为的涌现与演变规律, 其中社会网络<sup>[1,2]</sup>, 以及谣言<sup>[3]</sup>、流行病<sup>[4-6]</sup>传播动力学是当前统计物理的研究热点. 舆论传播研究主要关注交互系统中社会、人口、经济等各因素将如何影响舆论演变的过程, 如舆论产生、观点分裂和两极分化等, 从而揭示其中关键因素, 为舆论引导提供决策依据.

目前, 国外一些研究学者主要基于舆论传播机理提出了一些经典的舆论传播动力学模型, 如 Sznajd 模型<sup>[7]</sup>和 Deffuant 模型<sup>[8]</sup>. 前者主要描写两个具有相同观点的个体说服邻居个体认同这一观点; 而后者主要基于 Sznajd 模型提出了几种舆论传播规则. 最近, Galam 等人<sup>[9]</sup>给出一种有关舆论演化的重要模型, 简称 Galam 模型, 该模型以局部群体中个体总是按照少数服从多数的方式达成一致为基本原则. Kesten 等人<sup>[10]</sup>利用概率理论, 系统研究了谣言在移动人群中的传播行为及过程. Liu 等人<sup>[11]</sup>证明了随机网络是最容易传播谣言的; 汪小

帆等人<sup>[12]</sup>研究了具有可变聚类的无标度网络上的谣言传播行为. 何敏华等人<sup>[13]</sup>构建了自适应舆论演化模型, 该模型随时间演化表现出很强的趋同效应; 张力等人<sup>[14]</sup>利用网络论坛数据, 分析用户节点度的基本性质, 由此构建了度衰减的舆论传播模型; 叶作亮等人<sup>[15]</sup>分析了高校论坛的用户统计特征, 结果表明网络论坛度分布并不遵循严格的幂律分布.

舆论事件的产生和放大具有何种规律? 在舆论演变过程中, 哪些因素可能会推动舆论的放大? 针对我国基本国情, 研究舆论的形成、传播机理及其内在规律, 对于突发事件的舆论引导具有十分重要的理论和现实意义<sup>[2]</sup>. 对于具体舆论事件进行实证分析, 不仅有助于我们更好地认识与理解舆论的传播过程, 而且能够为其内在规律的揭示提供重要判据. 众所周知, 舆论的演变过程是由若干个体交互形成的群集动力行为, 因此借鉴群集动力学的研究, 能够启发式地认识并理解舆论的形成及演变机理. 一般而言, 舆论的产生通常是由非常规事件导致的, 并且往往具有突发性强、传播速度快、覆盖面广、持续时间长等特点. 目前, 在线社交网络中

\* 国家自然科学基金重大研究计划(批准号: 90924011, 91024026)和国家自然科学基金青年科学基金(批准号: 11105025)和资助的课题.

<sup>†</sup> E-mail: tangminghuang521@hotmail.com

借助社会关系进行舆论传播的实证研究难度颇大,主要有三个方面的困难: 1) 无法获得全面而准确的社交关系, 难以构建社交网络; 2) 尚无完善且高效的语义分析方法, 无法从海量信息中鉴别出舆论相关内容; 3) 由于舆论传播自身的特点, 难以跟踪舆论的传播通路. 虽然网站发布新闻的评论无法体现社会关系, 但是具有两个不可忽视的优点: 1) 回帖内容与舆论事件完全相关; 2) 舆论时空演变过程的可跟踪性. 在本文中, 我们借助门户网站, 分析研究舆论热点事件的新闻公告的评论, 2010 年互联网舆论事件“奇虎与腾讯之争”<sup>[16]</sup> 牵涉了国内 90% 以上的网民, 该事件具有因事而变的舆论引导、影响范围广、用户反应强烈、用户参与度高、持续时间长等特点, 因此非常适合作为考量舆论热点演变过程的研究对象.

## 2 数据说明

本文所分析的数据全部来源于互联网<sup>1)</sup>, 集中在 2010 年热点事件“奇虎与腾讯之争”, 数据集记录了 110 天内全国 34 个省、4 个直辖市、港澳特区共计 661 个市对事件的评论情况, 统计数据中只包含行政级别为县级市以上的单位, 评论总量为 149335 条. 数据以一天为计时单位. 每条记录包含了评论用户 ID、用户地理位置、评论时间、新闻主题、评论内容等信息.

## 3 在线评论的实证统计与分析

正如流行病传播动力学研究<sup>[17-22]</sup>, 通常以感染人数和新增感染人数为衡量传染病危害大小及其传播速度的关键指标. 在社会舆论事件中, 当前新闻评论总数是衡量舆论事件影响大小的重要指标, 因此我们定义某一时刻  $t$  地区  $i$  的总评论数为该地区的舆论关注度, 简称地区关注度. 另外, 新增评论数是衡量舆论传播速度的一个重要指标, 这里定义每天新增的评论数量为舆论的传播速度. 这两个指标反映了舆论事件的演变过程, 如舆论影响的大小, 舆论演变的快慢等情况.

如图 1(a) 所示, 在舆论的演变过程中关注度明显呈现出多次激增的现象, 这些激增分别对应于图 1(b) 中传播速度的多峰现象. 有趣的是, 如表 1 中, 传播速度表现出依次增大的峰值, 如  $t = 2$  时第一个峰值仅有 1279 条新增评论数, 而  $t = 39$  时第

四个峰有 29621 条新增评论数, 传播速度增大了 22 倍. 另外, 峰与峰之间的时间间隔越来越小, 从第一个峰值到第二个峰值的时间间隔为  $\Delta t = 19 - 2$ , 而第三个峰值到第四个峰值仅用了  $\Delta t = 39 - 32$ . 结合表 1 中重要公告发布的时间, 很容易看出这些激增现象是外部公告激励的响应行为, 尤其是  $t = 32$  和  $t = 39$  时腾讯连续发布的公告激起了广大网民的强烈不满, 显著促进了舆论事件的进一步放大, 令舆论的关注度和传播速度急剧增加, 使该事件成为社会焦点.

表 1 重要公告发布的情况

时刻 $t$	公告标题	当日新增评论数 $\Delta N_p(t)$
2	腾讯称 360 借色情网站推广	1279
19	腾讯宣布已正式起诉 360	1550
32	腾讯、百度等发联合声明	6505
39	腾讯让用户二选一	29621
56	工信部通报批评 360 与腾讯	6670

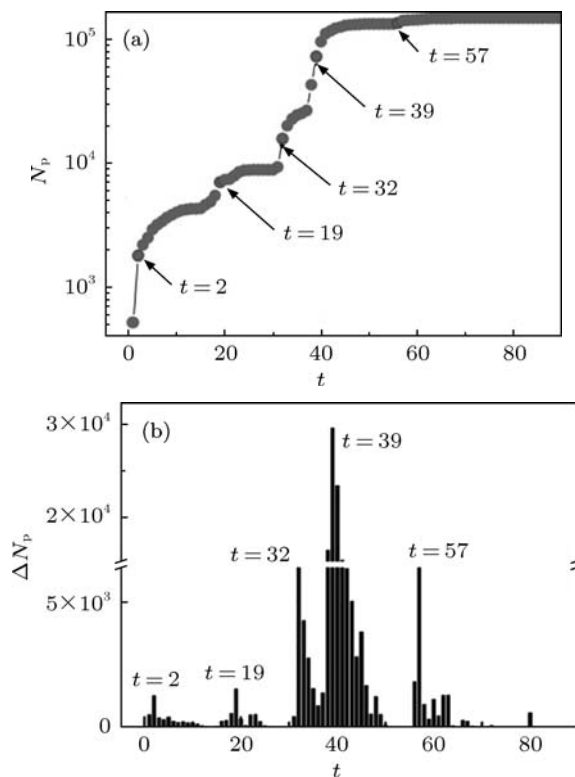


图 1 舆论关注度的演变过程 (a) 关注度随时间的变化; (b) 传播速度随时间的变化 (其中  $N_p(t)$  表示时刻  $t$  评论总量,  $\Delta N_p(t)$  表示评论增量)

### 3.1 关注度的时空演变

如前所述, 舆论传播是由若干个体交互形成的群集行为. 此处, 相对于图 1 的整体数据, 我们视

1) data from: www.163.com, www.sina.com.cn, www.sofu.com, www.ifeng.com

省市为个体,分析不同个体(省市)的演变过程.由图2明显看出:一方面,不同地区的舆论关注度存在明显的异质性,各地区的评论总数差别很大,如图2(a)中上海地区的评论数总是远高于宁夏;另一方面,不同地区在整个舆论事件中的演变趋势都呈现出一致性,如上海和宁夏两地评论数随时间变化的趋势基本一致.本部分研究重在揭示不同地区关注度的分布规律及其影响因素.

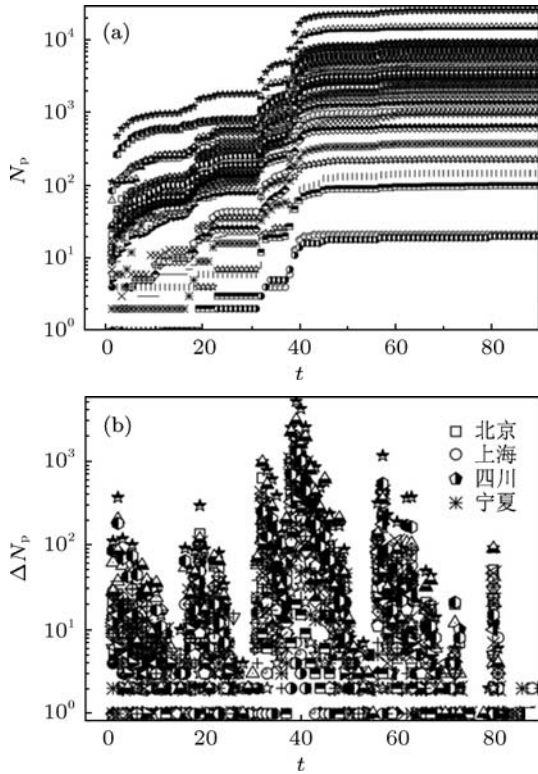


图2 省市舆论关注度的演变过程 (a) 关注度随时间的变化,每条曲线代表一个省(直辖市、特别行政区);(b) 传播速度随时间的变化(其中  $N_p(t)$  表示时刻  $t$  评论总量,  $\Delta N_p(t)$  表示评论增量)

### 3.1.1 关注度的地区分布

为了深入理解舆论的演变过程,揭示其内在规律,需要进一步考察舆论演变过程中不同地区的关注度分布情况.对此,Zipf分布是一个非常合适的重要指标,它最早是由Zipf于1949年提出的<sup>[23,24]</sup>.近年来,Zipf分析方法突破了语言学研究的界限,在复杂网络和复杂系统等领域得到了成功的应用<sup>[25-28]</sup>.

图3对不同地区内关注度从大到小进行排序,图3(a)给出了不同城市中关注度的Zipf分布,该分布服从双段幂律<sup>[29]</sup>;图3(b)给出了以省市为单位的Zipf分布,同样具有双段幂律特征,其中评论数利用归一化系数  $n_p^i(t) = N_p^i / \sum N_p^i$  进行了标准化处理,  $i$  表示不同省市单位.双段幂律分

布的物理意义在文献<sup>[30,31]</sup>中已经提到,即帕雷托法则(Pareto):在众多现象中,80%的结果取决于20%的原因.观察图3(a),前段幂律分布说明了舆论演化过程中部分城市的评论数量占总评论数的绝大多数,约27%的城市占据约89%的总评论数,遵循帕累托法则;另外,可以明显看出前段头部中出现少数靠前的地区(从高到低依次为广州、上海、北京、苏州、武汉、杭州、深圳),我们称这些少数地区为“引领者”.此外,图3(b)表示时间70天以内不同地区(省)关注度的Zipf分布,由图可见,随时间Zipf分布的斜率由-1.2增大到-0.8,在整个舆论事件中不同地区内关注度的相对比率变化较小,其演变形式具有某种程度上的一致性.造成这一较小变化的可能原因为:由于热点话题开始于部分地区,随着时间演化而舆论不断放大蔓延,体现在各地区关注度不断提高,其地区分布差异逐渐减小,从而分布的斜率随之增大.

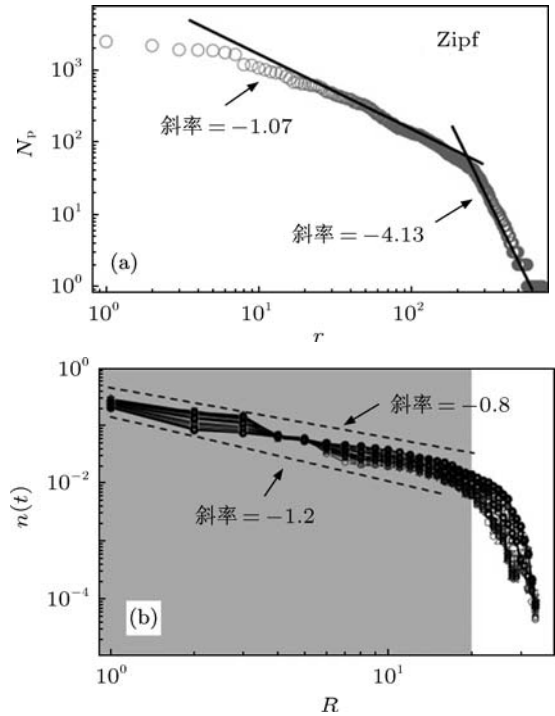


图3 不同地区内关注度的Zipf律 (a) 以地级市为基本单位;(b) 以省市为单位

### 3.1.2 影响地区关注度分布的关键因素分析

为了理解演变过程中不同地区内关注度分布的异质性特征,通过计算各省评论数与国民生产总值GDP<sup>[32]</sup>、人口数量<sup>[33]</sup>等的相关系数,从而衡量不同因素对地区分布异质性的影响程度<sup>[34]</sup>.Kendall's Tau系数<sup>[35]</sup>是用来描述两个序列之间相关性的指标,取值介于-1和1之间,Kendall's

Tau 系数越大说明两个序列一致性越高. 计算公式如下:

$$\tau = \frac{2}{m(m-1)} \sum_{i < j} \text{sgn}[(x_i - x_j)][(y_i - y_j)], \quad (1)$$

$\tau_{\text{GDP}}$  表示不同地区评论数与该地区 GDP 的相关系数;  $\tau_{\text{POPU}}$  表示与该地区人口的相关系数,  $\tau_{\text{primary}}$ ,  $\tau_{\text{middle}}$ ,  $\tau_{\text{high}}$ ,  $\tau_{\text{university}}$  表示与不同教育程度人数的相关系数. 上式中,  $m$  表示地区数;  $x_1, x_2, x_3, \dots, x_m$  表示不同地区的评论数,  $y_1, y_2, y_3, \dots, y_m$  表示不同地区的人口、GDP 及不同教育程度的人数;  $\text{sgn}(x)$  为分段函数, 即  $x > 0$  则值为 +1,  $x < 0$  则为 -1,  $x = 0$  则为 0.

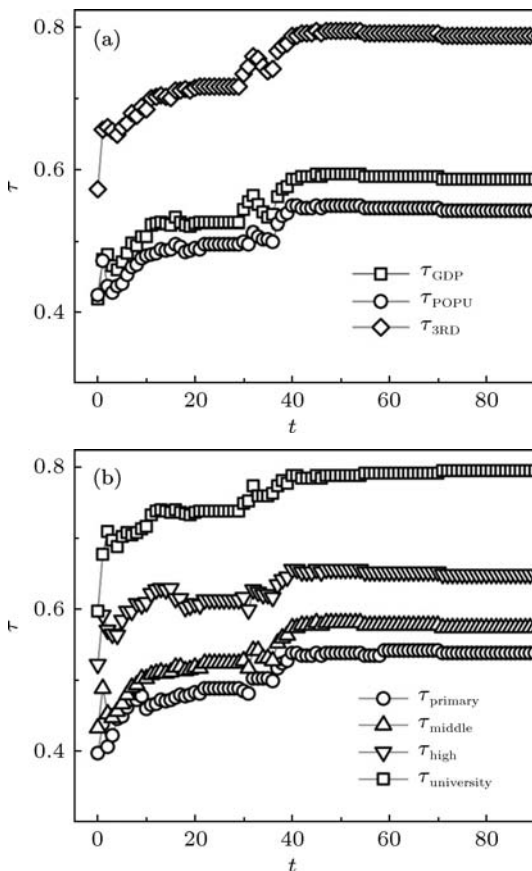


图4 地区关注度的相关性 (a) GDP、人口、第三产业的相关性; (b) 教育程度的相关性

图4描述了地区关注度和该地经济、人口、教育程度等的关联系数. 由图可见, 经济因素对地区关注度的影响程度高于人口因素. 这不但表明经济越发达的地区可能对舆论的关注度越高, 还暗示了这些经济较发达的地区可能就是“引领者”, 例如北京、上海、广东等省市. 其中在经济水平里, 第三产业具有更高的相关性, 如图4(a)所示, 省市的第三产业产值是影响不同地区舆论关注度的重要因素. 通过图4(b)明显看出, 学历越高的人群接

触互联网频率也越高, 与地区关注度的相关系数值也就越大. 第三产业值和教育学历等数据均来源于《中国统计年鉴2010》.

### 3.1.3 地区关注度之间的关联性分析

为了更深入理解地区关注度之间的规律性, 本文利用 Pearson 系数考察舆论演变过程中各省市地区间的相关性. Pearson 系数是一种衡量变量之间相关度的评价方法. 当对一个样本进行计算时, 系数大小由  $r$  决定,  $r$  取值介于 -1 和 1 之间. 系数等于 1 表示变量完全正相关, 0 则表示无关, -1 表示完全负相关. Pearson 系数计算公式如下:

$$r_{ij} = \frac{\sum_{t=1}^n (x_i(t) - \langle x_i \rangle)(x_j(t) - \langle x_j \rangle)}{n S_i S_j}, \quad (2)$$

其中  $n$  表示省市地区数,  $x_i(t)$  表示时刻  $t$  地区  $i$  的评论数,  $\langle x_i \rangle$  表示地区  $i$  评论数的均值,  $S_i$  表示地区  $i$  评论数的标准差.

图5展示了地区关注度之间的关联性, 很容易看出地区之间的关联性极高, 介于 0.97—1.00 之间. 这一分析证实了(1)式中的猜测, 整个舆论事件中各地区关注度的变化趋势基本一致, 具有某种意义上的同步特性.

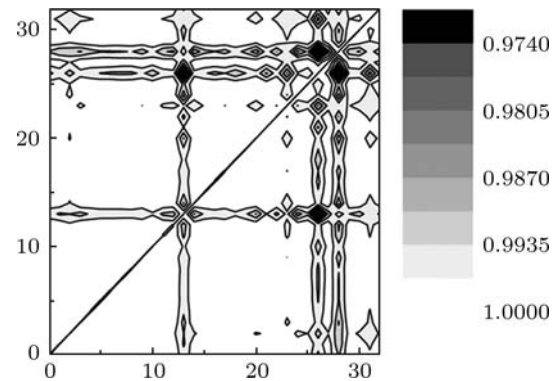


图5 各地区关注度之间的关联性

## 3.2 舆论引导对评论速度的影响

针对具体舆论事件的舆论引导的影响进行实证分析, 不仅有助于我们更好地认识与理解舆论的演变规律, 而且能够为合理、有效地进行舆论控制提供决策依据. 在这一部分, 我们分析了重要公告的发布对于各地区舆论传播速度的影响.

### 3.2.1 各地区传播速度的信息熵分析

统计物理学通常用熵表示分子不规则运动的程度, 信息论则把它作为随机变量无约束程度的一种刻画指标<sup>[36,37]</sup>. 本文借用了信息熵理论来表征

舆论事件中地区传播速度分布的异质性.  $H$  作为熵值用来描述舆论过程中的地区传播速度分布的异质性, 公式<sup>[38]</sup>如下:

$$H(t) = -\frac{1}{\log V} \sum_j p_j(t) \log p_j(t), \quad (3)$$

其中,  $p_i(t) = \Delta N_p^i(t) / \Delta N_p(t)$  表示时刻  $t$  地区  $i$  新增评论数占新增评论总数的比例,  $V$  表示地区数,  $\sum_i p_i(t) = 1$ . 当舆论只出现于一个地区时, 传播速度分布的异质性最高, 其指数  $H$  为 0; 随着参与地区的增加, 舆论分布的异质性有可能降低, 其指数  $H$  增大, 当所有地区都参与评论时, 满足熵最大化条件时,  $H = 1$ .

由图 6 可以明显看出, 熵值在很长时期内维持在一个高位,  $H$  平均值约为 0.65, 说明了舆论演变行为是一个地区异质性较为稳定的过程, 这可能由舆论本身的内在特点所决定. 首先, 观察图 1 可以明显看出不同的公告发布前后的评论数增幅变化的差异, 体现了舆论因事而变的传播特性; 同时, 多个公告使得对该舆论关注度居高不下, 表现了舆论具有持续性的特点. 值得注意的是, 在  $t = 2, 19, 32, 56$  之前  $H$  都会出现急剧下降的现象, 这可能是由于焦点舆论随时间推移逐渐淡出网民视野, 参与舆论事件的地区数减少所致. 图 6 在  $t > 60$  时异质性逐渐增加, 甚至到大于 80 的时候出现  $H = 0$  的情况, 这可能是由于不同地区对事件关注度的衰减程度不同造成的.

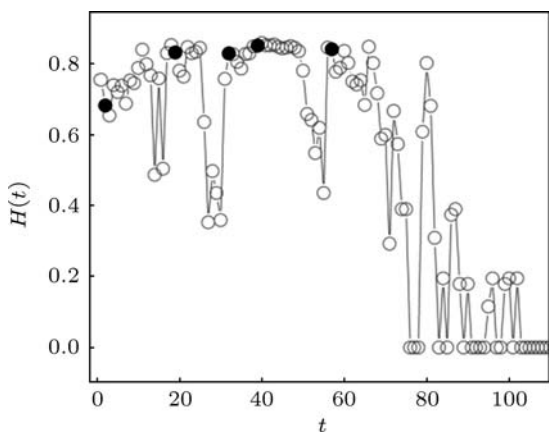


图 6 各地区传播速度的信息熵随时间变化的情况, 其中实心圆对应表 1 中的重要公告

### 3.2.2 地区传播速度的关联性分析

为了解各地区对于舆论引导的响应规律, 我们利用 Pearson 相关系数考察各地区传播速度之间的关联性. 从图 7 中可以看出, 某些地区在舆论引导下具有强关联性特征 ( $r$  值在 0.95—1.00 之间),

但一些地区与其他地区之间的关联性较弱 ( $r$  值在 0.65—0.85 之间). 通过统计数据, 我们发现, 这些地区具有一个显著的共同特征, 即地区经济相对落后, 如甘肃、内蒙古、青海、新疆、贵州、西藏等. 一般来说, 地区经济的发达程度直接决定了该地区的计算机普及率. 结合本文对异质性特征的分析及经济因素对舆论演变过程的显著影响, 我们推测, 地区经济越发达 (如北京、上海、广州等), 对于舆论引导的响应越迅速. 这一结论暗示, 对于经济发达区域, 有针对性的舆论监督能够有效地控制和引导舆论的传播过程. 但是, 经济因素能否表征地区在舆论演变过程中的引导能力, 仍需要大量的实证数据和更深入的研究.

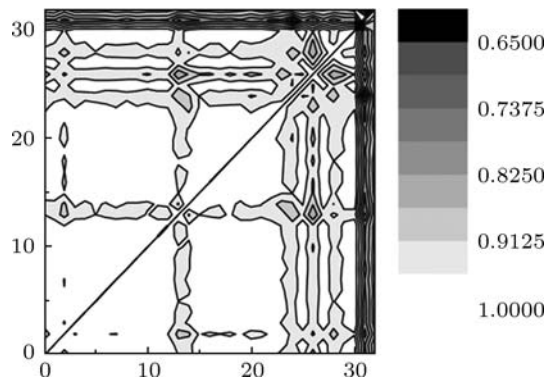


图 7 地区舆论传播速度的关联性

## 4 结论

本文运用统计物理学方法, 利用 2010 年“奇虎与腾讯之争”事件的在线评论数据, 统计分析了舆论演变中的内在规律. 首先, 我们研究了各地区关注度的时空演变规律. 分析表明, 虽然事件关注度在不同地区的分布存在极大的异质性, 即遵循双段幂律分布; 但是不同地区关注度在演变过程中却表现出明显的一致性, 其不同时间内的 Zipf 分布变化很小. 以上结论暗示了“引领者”的存在. 通过关联性分析, 我们发现地区分布的异质性受到当地经济与人口的显著影响, 而经济因素对于舆论的影响明显高于人口因素. 这不但表明了经济越发达的地区可能对舆论的关注度越高, 还暗示了这些经济较发达的地区可能就是“引领者”. 另外, 通过地区关注度之间的关联性分析, 我们验证了整个事件中各地区舆论的变化趋势基本一致, 具有某种意义上的同步特性.

其次, 我们研究了舆论引导 (重要公告的发布) 对于各地区传播速度的影响. 研究表明, 舆论引导

明显影响事件的关注度, 导致传播速度的多峰现象. 这表明了重要公告的发布将显著促进舆论事件的进一步放大, 令舆论的关注度和传播速度急剧增加, 使该事件成为焦点话题. 通过计算不同地区传播速度的信息熵, 我们发现评论的地区分布情况在大部分时间内都相对稳定, 而舆论引导在保持该稳定性上起着重要作用. 地区传播速度之间的关联性分析表明在整个事件中经济较发达地区的舆论变化更趋于一致, 暗示了这些地区对于舆论引导的响应可能更快. 因此, 对于经济发达地区, 有针对性的舆论

监督能够更有利地控制和引导舆论的传播过程. 本文的工作有助于进一步理解网络舆情的传播行为, 并为人类群集行为的深入研究提供了借鉴<sup>[39,40]</sup>.

由于影响舆论演变过程的因素众多, 加上个体行为交互的复杂性, 本文工作只针对评论数进行了定量分析. 在接下来的研究中, 我们将针对演变过程中的用户个体行为特征进行实证分析; 另一方面, 我们将基于实证分析的结果进行舆论演变模型的构建工作, 进一步揭示其内在规律.

- [1] Castellano C, Fortunato S, Loreto V 2009 *Rev. Mod. Phys.* **81** 591
- [2] Zhou T, Wang B H, Han X P, Shang M S 2010 *Jour. Syst. Eng.* **25** 742 (in Chinese) [周涛, 汪秉宏, 韩筱璞, 尚明生 2010 系统工程学报 **25** 742]
- [3] Moreno Y, Nekovee M, Pacheco A F 2004 *Phys. Rev. E* **69** 066130
- [4] Eubank S, Guclu H, Kumar V S A, Marathe M V, Srinivasan A, Toroczkai Z, Wang N 2004 *Nature* **429** 180
- [5] Liu J Z, Wu J S, Yang Z R 2004 *Physica A* **341** 273
- [6] Zheng D F, Hui P M, Trimper S, Zheng B 2005 *Physica A* **352** 659
- [7] Sznajd W K, Sznajd J 2000 *Int. J. Mod. Phys. C* **11** 1157
- [8] Deffuant G, Neau D, Amblard F, Weisbuch G 2000 *Adv. Compl. Sys.* **3** 87
- [9] Galam S 2008 *Int. J. Mod. Phys. C* **19** 409
- [10] H Kesten, V Sidoravicius 2005 *Annals Prob* **33** 2402
- [11] Zhou J, Liu Z H, Li B 2007 *Phys. Lett. A* **368** 458
- [12] Pan Z F, Wang X F, Li X 2006 *Jornal of System Simulation* **18** 2346 (in Chinese) [潘灶烽, 汪小帆, 李翔 2006 系统仿真学报 **18** 2346]
- [13] He M H, Zhang D M, Wang H Y, Li X G 2010 *Acta Phys. Sin.* **59** 5175 (in Chinese) [何敏华, 张端明, 王海艳, 李小刚, 方频捷 2010 物理学报 **59** 5175]
- [14] Zhang L, Liu Y 2008 *Journal of Beijing Jiaotong University.* **2** 67 (in Chinese) [张力, 刘云 2008 北京交通大学学报 **2** 67]
- [15] Ye Z L, Wang X Q, Wang X L, Li J 2010 *Complex Systems and Complexity Science* 1 (in Chinese) [叶作亮, 王雪乔, 王仙玲, 李静 2010 复杂系统与复杂性科学 01 (in Chinese)]
- [16] <http://zh.wikipedia.org/wiki/>
- [17] Zhou T, Liu J G, Bai W J, Chen G R, Wang B H 2006 *Phys. Rev. E* **74** 056109
- [18] Barrett C, Eubank S, Marathe M V 2005 *Modeling and simulation of large biological, information and socio-technical systems: an interaction based approach* (Springer Verlag)
- [19] Crepey P, Barthélemy M 2007 *Am. J. Epidemiol.* **166** 1244
- [20] Ferguson N M, Cummings D A, Cauchemez S, Fraser C, Riley S, Meechai A, Iamsirithaworn S, Burke D S 2005 *Nature* **437** 209
- [21] Tang M, Liu L, Liu Z H, 2009 *Phys. Rev. E* **79** 016108
- [22] Tang M, Liu Z H, Li B W 2009 *Europhys. Lett.* **87** 18005
- [23] Zipf G K 1949 *Human Behavior and the Principle of Least Effort* (Addison-Wesley)
- [24] Zipf G K 1968 *The Psycho-Biology of Language: An Introduction to Dynamic Psychology* (Addison-Wesley)
- [25] Cattuto C, Loreto V, Pietronero L 2007 *PNAS* **104** 1461
- [26] Lü L Y, Zhang Z K, Zhou T 2010 *PLoS One* **5** e14139
- [27] Blasius B, Tonjes R 2009 *Phys. Rev. Lett.* **103** 218701
- [28] Shao J, Lvanov P C, Urosevic Urošević, Stanley H E, Podobnik B 2011 *Europhys. Lett.* **84** 48001
- [29] Liu H K, Zhou T 2007 *Acta Phys. Sin.* **56** 106 (in Chinese) [刘宏鲲, 周涛 2007 物理学报 **56** 106]
- [30] Reed W J 2003 *Physica A* **319** 469
- [31] Reed W J, Jorgensen M 2004 *Stats-Theory & Methods* **33** 1733
- [32] [http://www.ce.cn/macro/more/201102/15/t20110215\\_22211061.shtml](http://www.ce.cn/macro/more/201102/15/t20110215_22211061.shtml)
- [33] [http://www.stats.gov.cn/tjfx/jdfx/t20110429\\_402722512.htm](http://www.stats.gov.cn/tjfx/jdfx/t20110429_402722512.htm)
- [34] Han X P, Wang B H, Zhou C S, Zhou T, Zhu J F 2009 e-print arXiv. 0912.1390
- [35] Kendall M 1938 *Biometrika* **30** 81
- [36] Shannon C E 1951 *Bell Systems Technical Journal* **30** 50
- [37] Thomas M C 2006 *Elements of Information Theory* (Wiley)
- [38] Colizza V, Barrat A, Barthelemy M, Vespignani A 2006 *PNAS* **103** 2015
- [39] Li M J, Wu Y, Liu W Q, Xiao J H 2009 *Acta. Phys. Sin.* **58** 8 (in Chinese) [李明杰, 吴晔, 刘维清, 肖井华 2009 物理学报 **58** 8]
- [40] Xiong F, Liu Y, Si X M, Ding F 2010 *Acta. Phys. Sin.* **59** 10 (in Chinese) [熊菲, 刘云, 司夏萌, 丁飞 2010 物理学报 **59** 10]

# Empirical study on spatiotemporal evolution of online public opinion\*

Gong Kai Tang Ming<sup>†</sup> Shang Ming-Sheng Zhou Tao

1) (*Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 610054, China*)

(Received 17 July 2011; revised manuscript received 6 September 2011)

## Abstract

To understand the spatiotemporal evolution of online public opinion and reveal its formation mechanism, we investigate the data from several popular online comments by means of statistical physics. Although the empirical results show that the heterogeneity of concerns exists in different areas, which follows a double power law, an obvious consistency of such concerns occurs during the evolution of online public opinion. Through correlation analysis, we reveal that the regional population and economy may have a significant influence on the concern about the event, and find that the consistency of concerns in different areas derives from the strong correlations among regions. On the other hand, the public opinion guide can significantly affect the concern about the event, and lead to the rapid increase of propagation velocity. By calculating the information entropy of propagation velocity, we find that the geographical distribution of online comments is relatively stable in most time, and the public opinion guide may help to maintain this consistency. Furthermore, the correlation analysis shows that the more developed areas tend to be more synchronized, which suggests the responses of these areas to the public opinion guide may be faster. Therefore, enhancing the guide of public opinion in developed areas can help our government to control the spread of the online public opinion.

**Keywords:** online public opinion, attention, Zipf distribution, information entropy

**PACS:** 89.20.-a, 89.75.Da, 89.70.Cf

---

\* Project supported by the Major Research Plan of the National Natural Science Foundation of China(Grant Nos. 90924011, 91024026), and the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 11105025).

<sup>†</sup> E-mail: tangminghuang521@hotmail.com