

# 大数据视角下的网络新闻跟帖中地域攻击现象分析

——以网易新闻跟帖数据为例

## 引言

2003年12月3日网易的第一条新闻跟帖出现,到目前为止,国内各大门户网站的新闻版块都开设了新闻跟帖功能。新闻跟帖作为网民参与社会公共事务讨论的一种交流形式,能够真实反映网民的社会心理。在热点新闻中,网民跟帖少则几十条、几百条,多的可达到上万条或者更多。然而我们发现,人气最高、最受关注的跟帖中往往夹杂着地域歧视、相互谩骂的跟帖。虽然很多门户网站已在醒目位置提醒“文明社会,从理性发帖开始,谢绝地域攻击”,但仍然无济于事。

地域攻击一直是媒体与社会公众关注的热点问题。而在诸多新闻网站中,网友对网易的认知度较高,网易在新闻跟帖方面较其他同类网站特色也较明显。因此我们选取网易新闻跟帖中的地域攻击现象作为研究对象,希望能够呈现出具体的各省之间的印象标签和关系,以及针对这些刻板印象背后的规律给出我们的分析与解释。从而对我国的网络空间治理、舆情检测与管理提供有效的分析方法和有力的建议。

## 相关研究

项目涉及到了(1)网络、传媒地域歧视现象研究,和(2)新闻跟帖研究。到目前国内对此的研究主要还是从文科角度出发,用社会学和传播学等学科的知识加以探讨。

### (一) 网络、传媒地域歧视现象研究

南京师范大学詹雄斌的《天涯论坛中的苏北地域歧视现象研究》以具体地区为案例,使用内容分析法、个案研究法等方法详细分析网络社区产生地域歧视现象的原因,并从群体认同的视角出发,探讨网络空间对于群体的自我认同与互斥现象的影响。而徐硕强的《网络新闻跟帖中地域歧视现象的现实解读与理性反思》则认为,造成网络地域歧视现象的原因既有刻板印象的惯性影响,也有地域经济差异与网民自卑心理的共同作用。

除了分析网络地域歧视现象的论文以外,也有相当多的学者聚焦于传媒的地

域歧视问题。张汉娟的《浅析媒介中的地域歧视现象》分别从媒介和受众的角度分析造成传媒地域歧视的原因；而曾庆江的《事件新闻报道命名与地域歧视》、沈毅玲的《新闻报道中的地域歧视性语言现象探析》则从新闻的报道出发探究造成地域歧视的原因。

## （二）新闻跟帖研究

网民新闻跟帖不同于传统的新闻评论，有其独特的特性与价值。廖福生、江昀的《对门户网站新闻跟帖特点的分析——以网易新闻为例》从传播学角度出发，通过研究新闻跟帖的形式和网民心理，探究群体意见极化现象和跟帖失范现象。而梁慧婷的《网民新闻跟帖中的语言暴力研究》采用案例分析法和文献分析法，以心理学、语言学、传播学为主要理论支撑，对网民新闻跟帖中语言暴力的主要内容、成因、特点进行了全面和深入的解读。

上述论文从文科的不同层面对新闻跟帖的地域歧视现象进行了探讨，对本文研究有一定借鉴意义及理论支撑。

## 数据

数据来源于本课程与网易新闻网站合作所提供的新闻正文与跟帖文本数据，其详细介绍为：

### （一）时间范围

研究者按照建构周的随机抽样方式，将每年的周一组成一个集合、周二组成一个集合……以此类推构成了七个集合，然后在每个集合中随机抽取出1天，进行组合，构建了一个“组合周”。该周为2017年12月5日(周二)、10月4日(周三)、5月11日(周四)、2月24日(周五)、7月1日(周六)、2月12号(周日)，研究者获得该组合周中每一天的新闻和对应的跟贴数据，以尽可能地消除人口流动导致的IP地址定位误标记。

### （二）热度范围

考虑到长尾效应，为减小网易提取出的数据规模，我们舍弃掉所有的冷门新闻和跟帖，只获得评论数大于100的新闻及跟帖数据。

### （三）信息范围

新闻正文标题、日期等信息；评论正文、时间、楼层、赞/踩数等信息；用户

发帖地域、基本属性等。

#### （四）数据统计信息

根据最后与网易协商后获得的数据,我们获得的数据中,用户数共为 597802, 用户评论总数为 3176262, 新闻一共有 216889 条。

### 项目框架

项目框架主要可以概括为如下几个方面:

1. 对于新闻
  - a) 新闻标题: 通过地域词典探测标题中所含有的地区。
  - b) 新闻正文: 可提取每段的关键句用主题模型确定新闻主题。
2. 对于回帖
  - a) 回帖用户: 利用回帖用户的IP去判断回帖者所处地区。
  - b) 回帖文本: 通过地域词典探测回帖文本中提到的目标地域, 利用模型分析回帖用户对目标地域的情感倾向; 利用LDA主题模型对评论进行主题和关键词提取。
3. 对于新闻和回帖:
  - a) 探究新闻标题和正文是否会对网友的回帖产生倾向性的引导。

### 数据分析方法

#### （零）数据预处理

预处理的阶段包括了IP定位和构建地域词典两个方面。

##### 1. IP定位

- a) 任务定义: 根据网易回帖用户的IP地址, 编写代码获取IP归属地, 返回归属地对应的省份和城市。
- b) 解决方法
  - i. 方法一: 利用Python调用淘宝IP地址库, 省准确度超过99.8%, 市准确度超过96.8%, 缺点是接口调用次数和速度有限。
  - ii. 方法二: 下载现有的QQWry IP地址库到本地, 编写Python代码进行调用和比对, 速度极快, 缺点是准确率可能没那么高。

## 2. 地域词典构建

- a) 任务定义：人工构造一个地域词典，目的是能够做到关键词和地域的对应，方便后续通过该地域词典探测回帖文本中提到的目标地域。
- b) 解决方法：主要从地理、人文等方面，针对容易出现地域攻击的领域来考虑构造该词典，具体包括：各省各城市的下一级地理名称、别称、球队名称等。在构建词典的时候尽量做到保持中性和客观，防止加入构建者的“地域刻板印象”导致后期模型结果不准确。对于各地美食特产、景点和方言、网络热词等，我们发现它们中的大部分并不能与地域产生一一对应的关系，因此不作考虑。

### （一）情感分析

情感分析主要是为了用于分析用户回帖文本对目标地域的情感倾向，在实现上可以通过传统统计学和神经网络两种方法。

#### 1. 统计学方法

- a) 方法一：利用现有情感词典，简单统计回帖文本中正面词汇与负面词汇的个数，由二者的相对大小决定该回帖情感的正负向。
- b) 方法二：利用带情感分数的情感词典并考虑程度词和否定词的存在，以累计每个情感词的情感分数的方法计算回帖的情感分数。
  - i. 情感词典来源：[大连理工大学信息检索研究室情感词汇本体库](#)、[知网情感分析用词语集](#)、[BosonNLP情感词典](#)。
  - ii. 情感分数的具体计算方法： $Score += W * (SentiDict[word])$ 。其中，SentiDict代表一个映射情感词word与情感分数的字典，W为权重，如果在回帖文本的两个情感词之间存在程度词和否定词，则W应该对应的改变成程度词的权重大小或否定词的负向。

#### 2. 神经网络方法

- a) 中文词向量训练：采用[搜狗新闻数据集](#)预训练出中文词向量表示。
- b) 神经网络：采用预训练好的中文词向量作为输入，利用RNN循环神

经网络对回帖文本进行向量建模。可用一标注好的的新浪微博情感数据集训练情感分类器，让分类器自动为回帖做情感分类。

## （二）网络模型

图论方法：将省市作为节点，极端情绪指数作为权重，构建有向带权图，并得出图中所有节点的中心值。由于极端情绪拥有自身的传播特性，因此在地域极端情绪网络图中，中心值表示的是该地域受到极端情绪影响的大小。

在这里我们选取特征向量中心值，即一个节点的重要性既取决于其邻居节点的数量（即该节点的度），也取决于其邻居节点的重要性。特征向量中心性公式为：

$$C_e(v_i) = \frac{1}{\lambda} \sum_{j=1}^n A_{j,i} C_e(v_j),$$

其中， $C_e$ 是结点中心性，结点  $j$  是结点  $i$  的入度邻居， $A$  是邻接矩阵， $\lambda$  是固定常量。将所有结点中心性写成向量  $C_e$ ，则有：

$$\lambda C_e = A^T C_e,$$

因此， $C_e$ 是 $A^T$ 的特征向量， $\lambda$  则是对应的特征值。

## （二）LDA主题模型

在 LDA 主题模型中，一篇文档生成的方式如下：

1. 从狄利克雷分布 $\alpha$ 中取样生成文档  $i$  的主题分布 $\theta_i$
2. 从主题的多项式分布 $\theta_i$ 中取样生成文档  $i$  第  $j$  个词的主题 $z_{i,j}$
3. 从狄利克雷分布 $\beta$ 中取样生成主题 $z_{i,j}$ 的词语分布 $\varphi_{z_{i,j}}$
4. 从词语的多项式分布 $\varphi_{z_{i,j}}$ 中采样最终生成词语 $w_{i,j}$

其中主题和词采用 Gibbs 抽样。

## 实验结果及分析

### （一）地区间网络极端情绪的总体趋势

首先，通过构造的地域词典，我们对网易新闻用户发言中的各省市自治区关键词进行识别，从而筛选出与地域有关的评论，并确定这些评论针对的目标地域  $B$ 。对于这些评论我们利用 LDA 模型做主题和关键词提取，构建了如图 0 的中国

地域攻击词云。



图 0. 中国地域攻击词云

之后，我们通过 IP 定位函数确认了发表地域评论的用户的地域来源 A，并通过情感分析模型对评论进行情感分类；最后，将用户地域来源 A 对评论目标地域 B 的负面评论数除以地域 A 的总发帖人数，作为地域 A 对地域 B 的极端情绪指数，并构建了如图 1 的矩阵图。

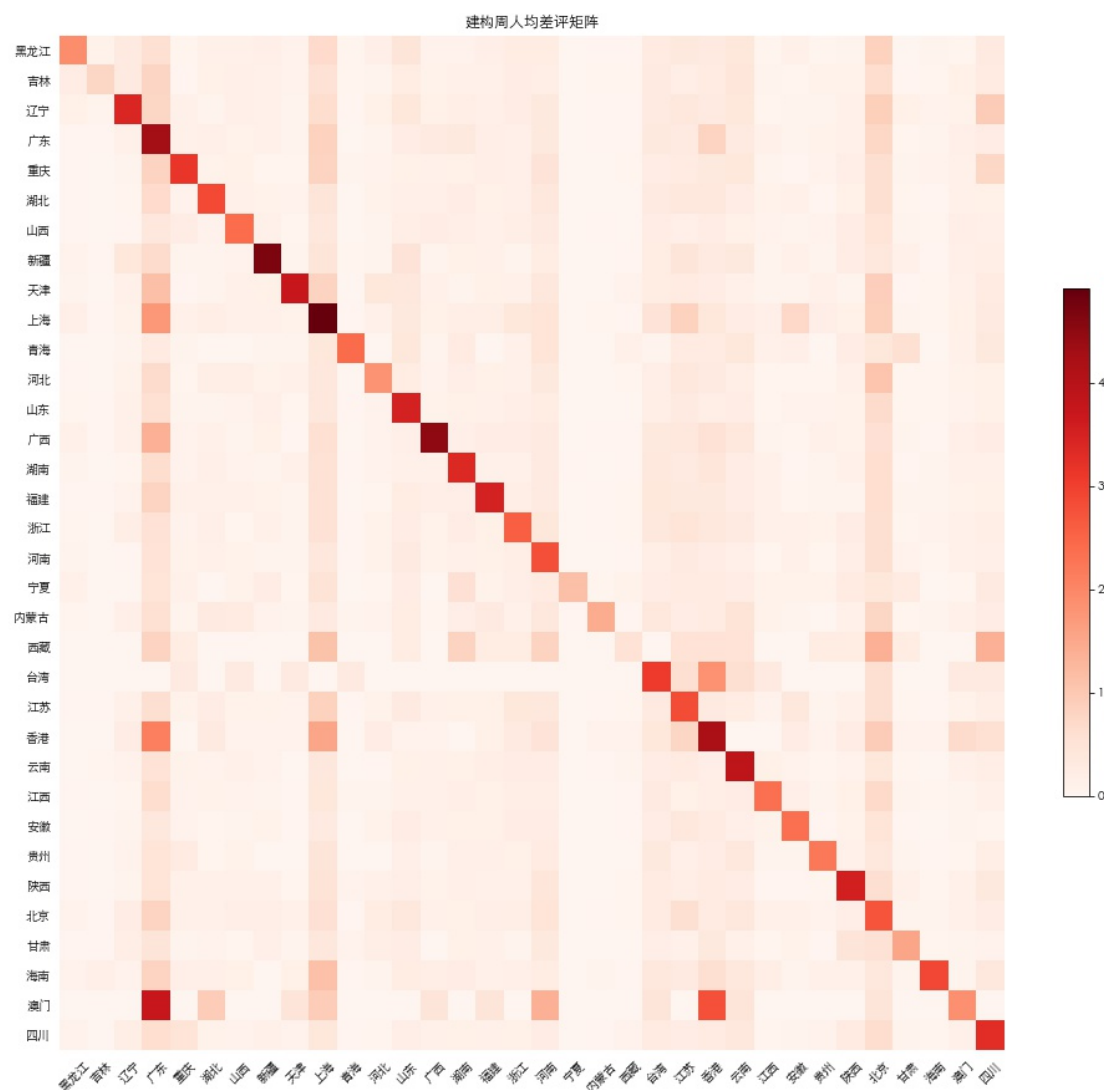


图 1. 构建周人均差评矩阵

在图 1 中，34 个省市自治区（包括港澳台）在行与列上一一对应，其中行代表 IP 所在地，列代表了现实物理空间，每一个矩阵单元格的色代表了地域极端情绪的激烈程度，方向为行地域对列地域的语言攻击。例如，矩阵第一行第一列单元格代表了在建构周中网易新闻的黑龙江网友对黑龙江的不满情绪和语言攻击程度；矩阵第一行第二列单元格代表了黑龙江用户对吉林的语言攻击强度，可见前者明显高于后者。

理解了矩阵图的对应关系之后，就不难理解，矩阵图的左上一右下对角线代表了网易新闻网友对所在地域的“自黑”，矩阵的行代表了该地区网友的总体极端情绪，列代表了该地区总体遭受其他地区的歧视程度。基于这些理解，我们至少能从矩阵图中得到四个发现。

1. 矩阵图对角线的整体颜色偏深,代表存在明显的“自黑”倾向,即网络用户倾向于抨击自己生活的城市,可见网民总体的生活满意度和精神满意度并不高。
2. 矩阵图的行与行之间颜色分布差距不大,即基本不存在某些地域更喜欢进行地域攻击。
3. 矩阵图的某几列颜色明显偏深,说明存在几个省市成为“众矢之的”,遭受网络极端情绪的可能性较高,分别为北京、上海、广州、河南和四川。
4. 矩阵图中有少数点对于该列而言颜色偏深,说明存在一些点对点(地域之间)的明显攻击关系,如澳门对广东、香港对广东、西藏对四川等。

下文将从这四个方面加以阐明。

## (二) 地区自我抨击的社会因素

从差评矩阵中可以看出,几乎每个地区的网络用户都喜欢抨击自己所生活的省市,体现出他们美好生活期待与现实社会资源获得之间的差异,可见网民总体精神满意度并不高,这背后的因素可能分为以下三点:

### 1. 社会发展速度无法满足上层建筑的社会现实

首先,值得注意的是,互联网用户的网络表达与真实社会诉求之间存在差异,互联网容易夸大个体的情感,这种特征被称为“情感新闻”(Papacharissi and Oliveria, 2012)。换言之,网络用户表达出的不满和歧视,通过网络筛选和语言选择等机制,被夸大了。另外,互联网是个体表达的平台,时常爆发出强烈的网络民族主义情绪(Liu, 2011),是一种负向的网络情绪(Steiglitz and Dang-Xuan, 2013)。网络空间中的讨论经常会体现网民的不公平感和“弱势感”(袁光峰, 2015),所以围绕争议议题的在线争论中难免存在一定比例的极端观点(Yardi and Boyd, 2010)。因此,如果仅从网络极端情绪中去解读整体社会矛盾便显得言过其实,但依然可以反映人们可能存在的“不幸福感”。那么,这种“自我抨击”背后的原因就与社会矛盾有关,是人们精神世界的快速膨胀与现实社会建设速度缓慢之间的矛盾,才导致了人们对自己生活的“家乡”提出批评。



## 2. 社会排斥的现象依然存在，尤其是人口迁入强势区域

另一个值得注意的问题是，网络用户 IP 所在地并不代表他就是这个地区的当地人，也有可能是别的地区的迁入人口。这样一来，他们对这个地区的不满就应该理解为一种“社会排斥”，即无法融入地方社会的表现。例如，广东和上海两个城市的“自黑”程度非常高，但这种“自黑”并不一定意味着“当地人对广东和上海的不满”，更可能是“外来迁入人口对广东和上海的不满”。广东和上海是人口流入强势区域，根据 2010 年人口普查数据，广东和上海已经成为全国外来人口迁入率的前两名。在这两个人口杂居的城市中，难免发生社会排斥(Jackson, 2016)，进而导致了网络极端情绪的产生。

## 3. 省市内部可能存在各类社会矛盾

此外，另一些“自黑”程度较高的省市，存在省市内部的人民矛盾，例如新疆、香港、江苏、安徽等。新疆历来存在内部的民族问题；香港在“亲大陆”政策上存在内部的政治对立；而江苏和安徽的南部和北部存在很多历史遗留的社会问题，导致省内纷争频发。这些问题都导致了“自黑”程度的增加，其背后的逻辑是多样的。

这些原因共同作用导致了“地域自黑”心理倾向的产生。

### （三）为什么抨击？为什么被抨击？

那么，是什么因素决定了哪个地域的网络用户喜欢表达极端情绪，哪个地域成为网络用户极端情绪的“众矢之的”呢？在回答这个问题的时候，我们需要借助统计手段，考察省市自治区层面的经济水平、人口流动和网络普及是否对其有预测能力。中国大陆各省市的数据从《国家统计年鉴》和《第六次人口普查》数据获得。由于港澳台没有对应年份的数据（如分省市的人口流动数据等），故仅对大陆 31 个省市进行统计检验。由于观测值（31 个）较少，不宜设置太多变量，以免引起共线性偏误，所以仅需要简单线性回归模型对其进行探索性的研究，结果如表 1 所示。值得注意的是，在构建（遭受）网络极端情绪指数的时候，我们剔除了自己对自己的不满情绪表达。

表 1. 网络极端情绪的影响因素（省市自治区层面）

网络极端情绪指数	遭受网络极端情绪指数
----------	------------

变量	系数	标准误	系数	标准误
人均 GDP/1000	.408*	(.234)	.255	(.897)
人均 GDP 增长率	-1.637	(1.345)	-.192	(5.153)
城镇率	-.636	(.472)	.245	(1.809)
网络普及率(%)	.152	(.594)	1.430	(2.276)
迁入人口/1000	.001	(.111)	1.203**	(.425)
迁出人口/1000	-.226	(.163)	1.207*	(.623)
截距	84.902**	(30.946)	-83.168*	(118.294)
R 方	.384		.632	
观测数		31		

注：\*p<0.05 \*\*p<0.01 \*\*\*p<0.001

### 1. 表达网络极端情绪

从表 1 中可知，在网络极端情绪的表达方面，地域间的差异较小，这与前文的描述性发现是一致的。唯一的规律是，模型显示，人均 GDP 越大，该地区的网络用户越有可能表达出自己对其他地区的不满情绪。对这个现象的解释可以从“虚假优越感”的角度出发（Hu，2017），人们总是视自己为社会的贡献者，当贡献水平越高的时候（如经济越发达的地区），就越容易视别人为“搭便车的人”，于是对他们产生负面情绪。

另一方面，尽管网络普及率对极端情绪的表达的影响并不显著，但二者的拟合曲线基本呈现出线性变化的趋势（如图 2 所示），总体而言，网络普及度越高的地区，表达对其他省市负面情绪的可能性也越大。从图中也可知，二者之间的非相关性很有可能是西藏和上海两个异常值导致的，尤其是在观测值较少的时候，异常值的杠杆作用很大。去除异常值之后的检验发现，二者显著相关。从网络普及度与极端情绪之间的关系可以看出，互联网提供了一个情绪宣泄，尤其是负面情绪的平台，人们真的还是“沉默的大多数”吗？（桂勇等，2015）。

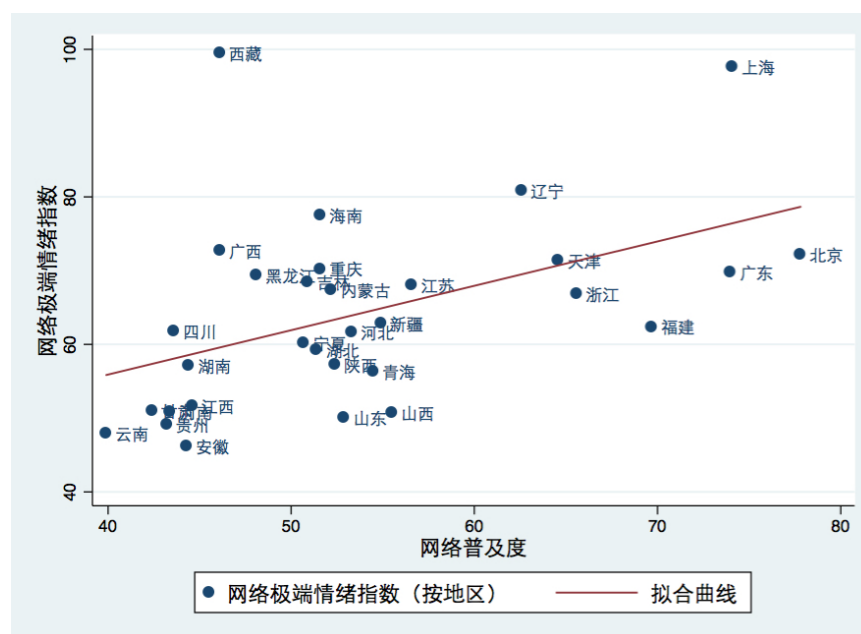


图 2. 网络普及度和网络极端情绪的拟合曲线

## 2. 遭受网络极端情绪

从表 1 的“遭受网络极端情绪”模型中可以看出，经济水平并不能解释是否被抨击，而社会流动却是关键的预测变量。无论是迁入还是迁出强势区域，都容易成为网络用户抨击的对象。从图 1 的矩阵图中也可以看出，在中国大陆，北京、上海、广东、河南和四川是最容易遭受抨击的五个省市。根据“六普”数据，北上广三个城市是人口流入强势区域，人口迁入分别已经超过 38 万、49 万和 138 万；河南和四川是人口流出强势区域，人口迁出已经超过 54 万和 49 万。可见人口流动会带来社会矛盾，进一步说明社会排斥的存在，迁入北上广的人对大城市不满，而河南、四川等地的迁入人口进入全国各地，不被当地人所接受，让当地人对河南和四川产生刻板印象。换言之，我国社会在社会融合方面依然存在各种阻力，这是本研究的现实意义。

为了进一步分析点对点的区域歧视关系，研究还绘制了地域极端情绪网络图（见图 3）。图中连线的箭头表示网络极端情绪的流向，例如香港指向广东，表明香港对广东的语言攻击；连线深浅及粗细表示极端情绪的强度。为了更清晰展现关系较强的几组对应关系，我们分别设置了强度阈值和选取强度最大的 10 组关系，构成了图 3。地域攻击最强烈的

十组关系分别为：香港对广东、澳门对广东、上海对广东、台湾对香港、香港对上海、广西对广东、澳门对香港、西藏对四川、西藏对北京、澳门对河南。这十组关系背后体现的正好是经济、政治和社会的三组对抗。

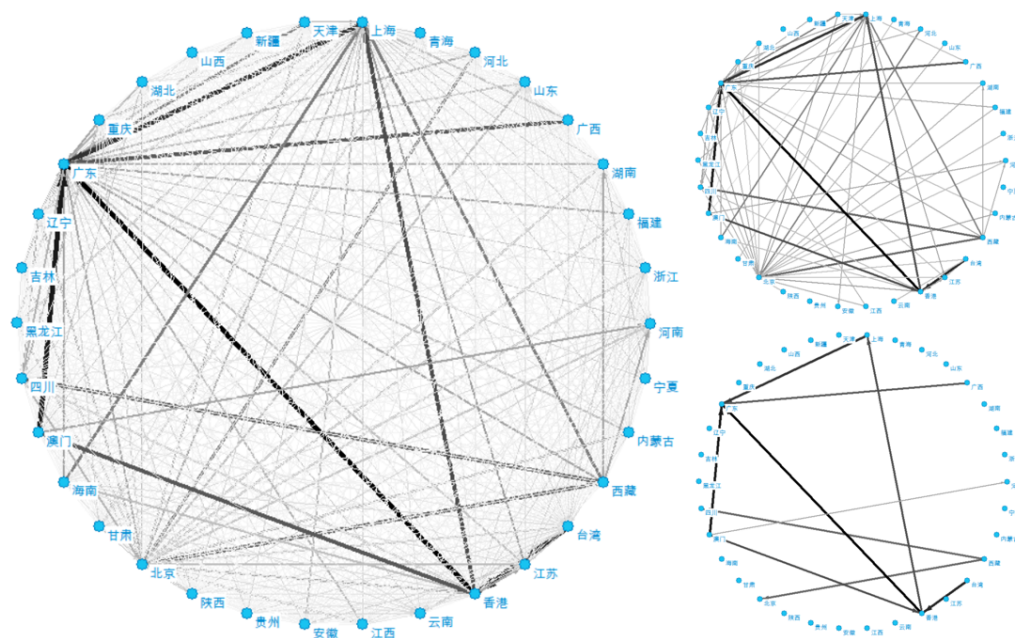


图 3. 地域极端情绪网络图

左:完整图 右上:强度阈值筛选图 右下:强度 Top10 图

首先，上海与广东、香港与上海、澳门与广东等关系中的双方是经济发展的竞争关系，二者之间容易出现较为激烈的负面情绪；此外，香港、台湾、西藏是民族分裂问题最为严重的三个地区，这三个地方的网络用户更容易选取大陆中的省市（通常是邻近的）进行言语攻击以表达自己的政治不满和政治诉求；最后，广西与广东、西藏与四川体现了相邻省市之间人员流动产生的冲突。仅以广西对广东为例，根据“五普”和“六普”数据，广东省迁入人口中以广西省最多（于潇等，2013），频繁的社会流动与接触并不能带来文化融合与社会融入，反而增加了冲突产生的可能性。

同时，借助网络图，我们也可以分析各结点（即省市自治区）的中心性，从而度量它们受到网络极端情绪的影响大小。我们将各地域自己对自己的不满情绪表达重新考虑进去，得出每个地域的特征向量中心性，并与遭受网络极端情绪指数进行对比（如表 2）。可以发现，遭受网络极端情绪指数基本能反映地域受到网络极端情绪的影响大小。因此，地域

受到极端情绪的影响的主要来源是其他地域对自身的攻击，而与自身对其他地域的攻击及对本地域的不满情绪关系较小。

表 2. 各地域特征向量中心性与其遭受网络极端情绪指数对比表

	特征向量中心性	排名	网络极端情绪指数	排名
黑龙江	0.028267655	28	0.019222332	28
吉林	0.00968454	33	0.008357286	32
辽宁	0.074776232	20	0.0413197	18
广东	0.541242906	1	0.270441349	1
重庆	0.068774272	21	0.040560544	19
湖北	0.091536993	15	0.049347208	15
山西	0.061536903	22	0.039235172	20
新疆	0.077636696	19	0.034749534	22
天津	0.057082876	23	0.029162421	26
上海	0.432770752	2	0.196487324	3
青海	0.018705897	30	0.011993836	31
河北	0.055020462	24	0.033364517	23
山东	0.143695742	10	0.084459353	10
广西	0.102932369	14	0.041525655	17
湖南	0.103757151	12	0.067209985	11
福建	0.103021869	13	0.057208708	13
浙江	0.118858011	11	0.065396613	12
河南	0.192689964	6	0.127918615	5
宁夏	0.002114068	34	0.001786933	34
内蒙古	0.012027034	32	0.006875827	33
西藏	0.016048338	31	0.01321367	30
台湾	0.18175063	7	0.099915973	9
江苏	0.227587807	5	0.116815164	6
香港	0.30437899	4	0.155289859	4
云南	0.178132753	8	0.105471765	7
江西	0.054021774	25	0.030295704	24
安徽	0.089800214	17	0.037377916	21
贵州	0.043149214	26	0.025659298	27
陕西	0.091536172	16	0.05464949	14
北京	0.338439951	3	0.218623767	2
甘肃	0.028622664	27	0.02933284	25
海南	0.025747214	29	0.014031255	29
澳门	0.087614442	18	0.045948444	16
四川	0.159078574	9	0.102124903	8

综上所述，当我们讨论区域为什么被抨击、为什么被人抨击的时候，需要把社会流动、经济发展等因素考虑在内，才能从宏观和结构层面上解释这个问题。

#### （四）网络传播的案例说明

除了现实社会因素之外，一些网络因素（例如互联网传播效应等）也会煽动和引导网络用户的情绪。或者说，网络用户的极端情绪是潜在和不稳定的。

仅以云南为例。如图 4，在建构周的结果中，云南基本没有遭受其他省市对其的负面评价，而在每个日期单独的差评矩阵中，云南也基本远离网络争论的风暴中心。但在 2 月 24 日的评论数据中，云南却遭受了很多网络极端情绪的攻击，甚至超过了北上广等地区成为“众矢之的”。为什么呢？

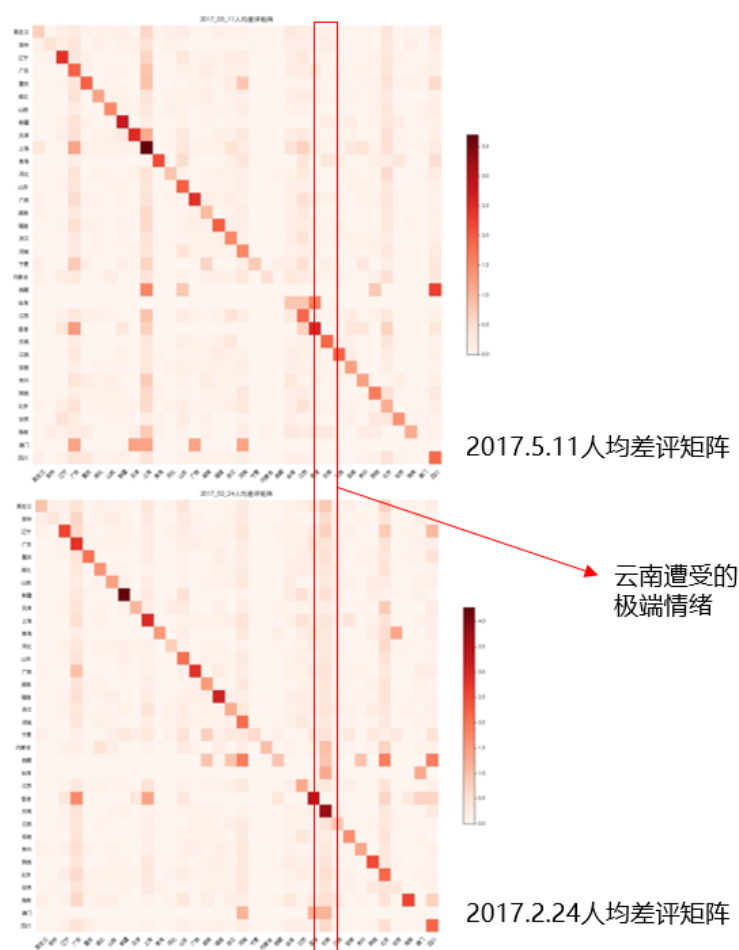


图 4. 云南遭受极端情绪对比图

这很有可能是网络传播导致的。2017 年 1 月 23 日，微博昵称为“琳哒是我”的用户发博称“丽江打人事件天理难容！”，其头条文章《天理难容》罗列了 10 张当事人被打后的照片。事件引起了社会各界的关注，包括旅游局和公安机关的回应。这起事件经过短时间的酝酿、发酵之后，爆发了长达两个月的全民声讨。我们追踪微博和网易的评论，发现在接下来两个月内，与云南相关的评论大多为用户表达自己在云南旅游的不愉快经历，以及对云南旅游业的痛诉，云南一时成





筛选出来的这部分评论中,可能不够准确。比如:由于“佛山无影脚”含有“佛山”,于是目标地域被探测成了“广东”。

- b) 地域探测存在的另一个重要的问题是召回率可能不够高,有些地域相关的评论由于并不显式含有地域词典中的词语而不能被探测出来,但其仍然隐式存在目标地域,比如:回帖楼层中的指代关系。举个例子:2楼的回帖人回帖“重庆是个好地方”,4楼的回帖人回复2楼“我觉得你说的对”,此时4楼的目标地域是重庆,但我们并不能探测出来。

## 2. IP探测

- a) 在线探测IP的程序准确度高但对于接口访问的频率和速度有限制,而本地查询IP的程序精确度可能没那么高。

## 3. 情感分析

- a) 情感分析在自然语言处理中本就是一个比较棘手的问题,在该项目中网络用语、对比以及反讽现象的存在更是加大了对于评论情感的分析难度。比如:“去过山东,感觉山东的经济就像黄晓明的身高一样!”这条评论实际上是在讽刺山东的经济不景气,但是像这种回帖中的比较关系我们目前不能进行准确的识别。
- b) 在处理有多目标地域的评论时,我们的模型不能准确的区分同一条评论对不同目标地域的情感倾向。比如:“恭喜广东,在缺乏北京的系列赛里,谁输谁赢都显得无关紧要,希望双方能够贡献精彩的比赛,广东的总体实力确实太强,四川的差距还太大,系列赛估计也就这么回事了,加油,广东,加油,四川!”这句评论中我们探测到了广东、北京、四川三个目标地域,但是却不能确定对于这三个地域分别的情感倾向。

## (三) 主观影响

### 1. 来自分析者

- a) 研究者在筛选结果和进行分析的时候会受到研究者自身主观因素的影响。

### 2. 来自网易



- a) 网易官方会对回帖进行筛选, 对于不符合规定的帖子会采取删帖和用户禁言措施。

### 3. 来自回帖者

- a) 回帖者人为刷楼的现象会导致同一个地区的同一种情感倾向被多次计算, 造成结果的不准确。

## 参考文献

- [1] 徐硕强, 谭伟. 网络新闻跟帖中地域歧视现象的现实解读与理性反思[J]. 电子政务, 2013(9):9-13.
- [2] 曾庆江, ZENGQing-jiang. 事件新闻报道命名与地域歧视[J]. 海南师范大学学报(社会科学版), 2014, 27(9):104-107.
- [3] 廖福生, 江昀. 对门户网站新闻跟帖特点的分析——以网易新闻为例[J]. 东南传播, 2010(6):88-90.
- [4] 詹雄斌. 天涯论坛中的苏北地域歧视现象研究[D]. 南京师范大学, 2016.
- [5] 李维杰, 包红云. BBS 中信息传播模式的特征分析[J]. 计算机工程与应用, 2010, 46(29):18-22.
- [6] 岳伍东. 报纸新闻报道地域歧视与妖魔化河南研究[D]. 华中科技大学, 2010.
- [7] 张汉娟. 浅析媒介中的地域歧视现象[J]. 新闻研究导刊, 2016, 7(2):50-51.
- [8] 梁慧婷. 网民新闻跟帖中的语言暴力研究[D]. 湘潭大学, 2016.
- [9] 沈毅玲. 新闻报道中的地域歧视性语言现象探析[J]. 中国记者, 2017(5).
- [10] 冯健, 赵楠. 后现代地理语境下同性恋社会空间与社交网络——以北京为例[J]. 地理学报, 2016, 71(10):1815-1832.
- [11] 饶元, 吴连伟, 张君毅. 跨媒介舆情网络环境下信息传播机制研究与进展[J]. 中国科学:信息科学, 2017(12).
- [12] 江昀, 罗静雅. 微博传播中的地域歧视现象研究[J]. 新闻爱好者, 2017(9):41-43.
- [13] 汤博为. 微博新媒体时代的公共利益与公民权利——基于宪法学和创新社会管理机制的思考[J]. 四川理工学院学报(社会科学版), 2012(3):28-32.
- [14] 桂勇, 李秀玫, 郑雯, 等. 网络极端情绪人群的类型及其政治与社会意涵 基于中国网络社会心态调查数据(2014)的实证研究[J]. 社会, 2015, 35(5):78-100.
- [15] 于潇, 李袁园, 雷峻一. 我国省际人口迁移及其对区域经济发展的影响分析——“五普”和“六普”的比较[J]. 中国人口科学, 2013(3).

- [16]Hu, Anning. 2017. "On the Relation between Subjective Inequality and Generalized Trust." *Research in Social Stratification and Mobility* 49: 11-19
- [17]Jackson, Steven F. 2016. "Does China Have a Monroe Doctrine? Evidence for Regional Exclusion." *Strategic Studies Quarterly* 10(4): 64-89
- [18]Liu, Fengshu. 2011. "From Political Indifference to Vehement Nationalism: Chinese Young People Negotiating the Political Self in the Internet Age." In *Urban Youth in China: Modernity, the Internet and the Self*. New York: Routledge: 161-179
- [19]Papacharissi, Zizzi, and Mariade Fatima Oliveria. 2012. "Affective News and Networked Publics: The Rythms of News Storytelling on Egypt." *Journal of Communication* 62(2): 266-282
- [20]Steiglitz, Stefan, and Linh Dang-Xuan. 2013. "Emotions and Information Fiffusion in Social Media-Sentiment of Microblogs and Sharing Behavior." *Journal of Management Information Systems* 29(4): 217-248
- [21]Yardi, Sarita, and Danah Boyd. 2010. "Dynamic Debates: An Analysis of Group Populariztion over Time on Twitter." *Bulletin of Science, Technology & Science* 30(5): 316-327