

大数据视角下的网络新闻跟帖中地域负面评论分析

——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板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，能在部分程度上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攻击、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负面评论作为研究对象，希望能够探究地域之间的负面对立情绪现状，并对背后的成因给出我们的分析和解释，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

互联网作为一种传播和社交的新型平台，越来越受到社会科学研究的关注（DiMaggio et al., 2001; Robison and Grenshaw, 2002; Selwyn, 2004; Juliet and Patton, 2013）。自1994年中国正式成为“互联网国家”起，学界就持续关注互联网对政治、经济与文化等领域的影响，特别是互联网用户的“在线表达”——公共领域与事件中的话语斗争和话语策略（桂勇等，2015）。由于互联网构建了一个悬浮于现实社会的互动平台，互联网中几乎所有个体被割裂为陌生人，陌生体验成为共同体验（张杰，2012）。于是，在互联网与现实社会“二元割裂”的情境下，互联网用户如何对社会议题进行评论与情绪表达，成为网络传播的关键。

对于网络传播的研究一般从三个角度出发，一是从社会学的角度，关注网络上的社会交往行为；二是从政治、法律的角度，关注网络传播对现有社会规范体

系的影响和改造；三是从与传统媒体想参照的角度，关注网络传播对信息传播活动带来的变化及其特点（陈红梅，2004），但这三个方面并非完全割裂的。最早对社交网络进行系统化分析的霍华德·莱茵戈德（Howard Rheingold）将这三个维度整合起来，并且提出了网络交流（computer-mediated communication）的概念（陈红梅，2004）。网络交流指的是个体或群体在以数字网络为基础的平台，发表言论并且进行互动的行为。

随着互联网的快速发展，网络交流的载体和内容也逐渐多样化，并产生了极化的现象（Wu，2014）。情绪化的表达和鲜明的对立态度滋养了针对性发言在网络空间内的生长，成为了网络交流的一个独特现象。桂勇等（2015）学者认为，“对社会化媒体用户价值和态度的研究需要超越理性和认知的视角，并具体考察‘态度—情绪’这一复合体”。换言之，对网络传播的研究，一个重要视角就是关注网络评论中的情绪表达。

研究表明，情绪有助于观点在社交媒体上的传播，充满情绪的话语能在更短的时间内被转发，被转发次数也更多（Stieglitz and Dang-Xuan，2013）。从易于文本分析操作的角度而言，网络情绪主要分为正向和负向两种，尽管中国的互联网是一个非政治化的空间，用户呈现政治冷漠的态势（Damm，2007），但在社会议题上，中国互联网用户依然积极爆发出自己的情感立场和现实诉求，其中最明显的就是地域之间的对立态度。

从态度上而言，中国各省市之间历来存在对立、防备和依赖的地缘政治（游家兴等，2018）；从情绪上而言，网络表达的易致性和“情感性”（Papacharissi and Oliveria，2012）又加剧了情绪对抗的张力，所以网络在线表达中，时常会出现地区间对抗性的负面表达，具体表现为地域歧视、偏见，甚至是相互攻击。当然，对网络地域间负面情绪的研究，需要从两个方面出发，一是负面情绪的表达特征；二是负面情绪的形成因素，但目前学界很少有将二者结合的研究，本研究试图填补这一空白。

在网络负面情绪的表达特征方面，首先要解决的是技术性的探测问题，即如何通过科学有效的指标反映负面情绪表达及遭受负面情绪的强度。莫勒罗等人（Molero et al.，2013）通过问卷和因子分析等方式构建了遭受负面情绪攻击的感知指数，试图将情绪感知进行量化。尽管测量指标依然是主观的，但是其量

化思维对本研究具有启发，我们试图将文本进行量化，探测出负面评论并构建负面情绪的客观指标。其次，就是要对网络情绪的表达做出界定，将其与现实社会中的情绪对应，因为网络表达通常是关乎情境和大众认知的（李贞芳等，2017）南京师范大学詹雄斌（2016）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间负面情绪的表达特征，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。而徐顽强与谭伟（2013）则认为，网络跟帖中的负面情绪表达实际上是一种表达的“失范”，换言之，是心理上的无序状态，是个体固有刻板印象和对不平等感知的总体性表达，尽管用于分析的评论文本是客观的，但文本背后的情绪依然是主观的。许多时候，地域间的刻板印象正在经历“社会货币化”的过程（喻国明，2018），负面情绪成为人们日常交流的谈资，对应到互联网社会中，负面情绪被不断放大。所以说，现实社会中的偏见和负面情绪会从现实社会转移到互联网的虚拟表达中去（Kuran and McCaffery, 2004），于是，对互联网言论中的负面评价和负面情绪进行研究，就具有了显示意义。

在网络负面情绪的成因方面，研究者也作出了诸多努力，如从个体心理特征、到社会地缘政治、再到地域间的经济差异等（徐顽强、谭伟，2013；Damon, 2013；Small, 2017），都成为了负面情绪表达的影响因素。但是，这些因素都是静态的社会和人口特征，而没有考虑动态的传播因素，例如人口流动带来的人际接触，以及互联网的信息流动等。中国是一个人口流动大国，省级人口流动为我国的劳动力市场带来了新鲜血液，也加速着不同区域之间的文化传播。但频繁的人际交往也会带来负面影响，例如加深刻板印象，并固化成为社会规范，产生强大的社会压力等（游家兴等，2018）。所以，按照接触理论，人口流动会促进个体的地域间负面情绪的产生，于是人口流入和流出的大省更有可能在互联网中表达负面情绪以对其他省市表示不满，同时也遭受来自其他省市的针对性话语。对于互联网中的信息流动而言，当信息源被拉长时，更有可能会出现泛化的攻击和对立态度（Blommaert et al., 2014），也就是说，当议题被几经传播和经过酝酿之后，竞争和对立的状态就会被加剧。而媒体在剪裁新闻的时候往往会对社会真实产生框架依赖（游家兴，2018），进而忽略了信息全貌，其结果就是信息源被拉长和狭窄化。于是，我们可以得出这样的假设，从动态视角而言，人际接触和网络传

播对网络负面情绪的表达，特别是地域之间的对立情绪，有加剧的作用。

以往的相关研究，至少存在几方面的不足：(1) 几乎没有对负面情绪的文本进行量化的分析，因此无法对网络整体舆论进行全面描绘；(2) 量化的尝试通常是主观的，客观的数据挖掘和统计分析较缺乏；(3) 对网络负面情绪的成因研究大多停留在静态的变量上，忽略了社会与网络的双重传播力量；(4) 缺少整合性的视角。基于这些研究不足，本研究以网易新闻跟帖文本为例，试图回答如下几个问题：

Q1：如何从大数据视角对文本内容进行分析，并探究互联网空间中地域之间的负面对立情绪现状？

Q2：互联网中各地域之间负面评价的表达现状是什么样的？

Q3：现实社会中的人口流动和社会经济发展水平如何影响网络地域负面评价？

Q4：新闻传播和发酵如何影响地域间负面评价？

Q5：是否还有影响地域间负面评价的其他因素？

数据来源

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) 原始数据统计信息：

按照建构周的随机抽样方式，我们将某一年所有的星期 N 构建为集合 N ，然后在每个集合中($1 \leq N \leq 7$)随机抽取出1天，构建了一个“组合周”。我们从网易新闻方最终获取了随机选取五天的数据，很遗憾没有获得完整的组合周。这五天数据的完整数据字段分别为：

- 新闻信息：新闻编号、新闻发布时间、新闻标题
- 评论信息：评论编号、评论文本内容
- 用户信息：新闻编号、用户通行证、评论编号、发帖时间、发帖终端、用户信息完善度、发帖IP、发帖时间、所在楼层、第一层楼编号、上一层楼编号

加粗部分为我们所用数据字段，其中根据去重后的新闻编号计算出新闻总数、评论编号计算出评论总数、用户通行证计算出用户总数。则最原始数据统计信息如表1所示，值得注意的是，由于每天出现的用户不同，且总用户数会进行去重处理，所以总体用户数小于每天的用户数之和。

表1：原始数据统计信息表

日期	星期	新闻数(占比)	评论数(占比)	用户数(占比)	用户平均评论数
2017-02-24	五	29112 (13.44%)	624329 (19.66%)	205837 (34.43%)	3.03
2017-05-11	四	37163 (17.15%)	678006 (21.35%)	224272 (37.52%)	3.02
2017-07-01	六	36318 (16.76%)	647486 (20.39%)	213387 (38.70%)	3.03
2017-10-04	三	40110 (18.51%)	605156 (19.05%)	191274 (32.00%)	3.16
2017-12-05	二	74186 (34.24%)	621285 (19.56)	191338 (32.01%)	3.25
总体		216641 (100%)	3176262 (100%)	597802 (100%)	5.31

(二) 预处理后数据统计信息：

我们观察到跟帖中会出现大量的重复评论，一般是关于上访、揭发等主题的负面情绪信息，如“XX省XX县派出所所长XXX贪污X千万”。此类信息值得管理者关注，但对于我们的研究而言，信息如果重复次数过多，则会影响到探究负面情绪的比例，因此我们只保留重复信息中的第一条。

其次，当研究负面评论的比例信息时，过少的统计样本会严重影响到比例的计算，导致极端值的出现。同时，因为负面样本过少会导致指标出现异常值，比如：

在我们的数据中，台湾网民仅占全体网民的0.55%，探测出的发布负面评论的台湾用户仅24个，偶然因素对比例影响过大，在我们后续的分析中，台湾、澳门、青海等用户数过小的地域分析指数呈现出异常值。

我们不认为少量用户的评论习惯可以代表该地域的负面评论状况。因此，我们设定负面评论的最小样本量阈值500，筛选出样本量足够大的地域相关评论，希望能够得出相对可靠的结论。

但是，若计算人均负向评论数量，则小数量不会对此产生影响。所以我们给出删除小样本和未删除小样本的两种情况统计信息，如表2所示。其中筛选地域和情感评论的方法会在后文中具体解释。

表2：预处理后数据统计信息表

样本类型	用户数	评论数	用户平均评论数	地域评论数(占比)	用户平均地域评论数
未删除小样本	597802	3176262	5.31	160022 (5.04%)	91102
删除小样本	485769	2647318	5.50	110297 (4.17%)	65712

数据分析方法

在展示分析结果之前，我们先简单地介绍我们所用的技术框架、模型效度和指标设计，为后续的研究发现展示提供帮助。

(零) 地域探测

我们将A地对B地的评论动作定义为A地“发出”和B地“遭受”，则地域探测的阶段，我们首先使用IP定位获取评论的发出地，再用地域词典获取评论的遭受地。

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。
- c) 选择方法：在IP定位问题上我们选择了方法二，因为同样大小的数据集，在相同的Intel Xeon处理器上，开16个进程，方法一需

要约1小时才能跑完，而方法二只需要4分钟，且错误率也在我们可以接受的误差范围内。

2. 地域词典构建

- a) 任务：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) 构建思路：主要考虑地理、人文领域，针对容易出现地域负面评论的关键词来构建词典，具体包括：各省市自治区的别称、下一级地理名称、代表球队名称等。
- c) 注意事项：
 - i. 在构建词典的时候尽量做到保持中性和客观，防止我们的“地域刻板印象”导致的后期模型结果不准确。
 - ii. 对于各地美食、特产、景点、方言、网络热词等关键词，它们中的绝大部分并不能与地域产生一一对应的关系，因此不作考虑。
 - iii. 对于其他无法产生一一对应关系的关键词也进行删除，如西藏简称“藏”会探测出“躲藏”等评论而不一定是“西藏”相关评论。
- d) 基本特征：平均每个地域拥有15.3个关键词。
- e) 效果说明：我们采取人工抽样评测的方式，随机抽取了100个样本，三个人进行标注，最终测试我们的地域词典探测准确率为98%。

(一) 情感分析

情感分析主要是针对已被地域词典匹配出目标地域的用户回帖文本，进行用户对目标地域的情感倾向分析，在具体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。

3. 选择方法

在情感分析时, 我们选择了统计学方法中的方法一, 因为用户回帖的平均长度在15字左右甚至更少, 而神经网络对于长文本的分类效果会更优。经过简单的试验对比, 我们发现针对回帖这种短文本, 直接匹配情感词的方法是准确率高且速度快的。在我们的人工标注测试中, 我们抽样了100个样本, 由三人进行标注, 获得的情感识别准确率为70%。

(二) 网络模型

图论方法: 将地域作为节点, 地域间负面评价指数(地域间负面评论数÷评论发出地网络用户数)作为权重, 构建有向带权图, 并得出图中所有节点的中心值。由于负面情绪拥有自身的传播特性, 因此在地域负面评价指数网络图中, 中心值表示的是各地域在网络负面情绪传播中的重要性。

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。特征向量中心性算法是当前最常用的重要节点的评判方法之一, 有着广泛的应用。地域负面评价指数网络图的节点数较少, 且为完全图, 可以避免特征向量中心性算法缺陷(可能会导致网络中存在被大量关注但中心值为0的点)的发生。

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, λ 是

固定常量。将所有结点中心性写成向量 \mathbf{C}_e ，则有：

$$\lambda \mathbf{C}_e = \mathbf{A}^T \mathbf{C}_e,$$

因此， \mathbf{C}_e 是 \mathbf{A}^T 的特征向量， λ 则是对应的特征值。

（三）主题分析

对于评论文本中出现的每一段评论，我们将它视作一个文档，因此可以利用LDA模型探究所有评论的主题词，以此期望挖掘出用户评论的共性和个性。

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 α 中取样生成文档 i 的主题分布 θ_i
2. 从主题的多项式分布 θ_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 β 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。

（四）指标设计

首先，我们从传播者，即评论发出者角度出发，希望探究各省之间发帖用户的负面情绪强烈程度的异同。因此设计了基于用户的人均负面评论指标：

1. **（A地对B地）人均负面评价指数：** A地对B地负面评论数/A地用户数
2. **（A地）人均发出负面评价指数：** A地对其他地区负面评价指数之和
3. **（B地）人均遭受负面评价指数：** 其他地区对B地的负面评价指数之和

其次，我们以地域角度出发，希望获取该地域获得的负面评价和发出负面评价的特征，因此设计了基于地域的负面评论比例指标：

1. **（A地的）本省负面评价比例：** A地对A地的负面评价数量/A地对A地地区的评价总数
2. **（A地的）发出负面评价比例（除A地）：** A地对其他地区的负面评价数量/A地对其他地区的评价总数
3. **（B地的）遭受负面评价比例（除B地）：** 其他地区对B地的负面评价数量/其他地区对B地的评价总数

注：我们研究的粒度为省级，以下直辖市、特别行政区、自治区均称为“省份”。

称，是因为城市名在一定程度上可以反映负面评论所针对的城市（比如“福建莆田”）。

图 1 中，除了一些通用的骂人词语（“傻”、“逼”等），部分地域的词云中出现了较有“特色”的关键词，比如广西的“吃”、“狗”、福建的“医院”、河南的“井盖”等（见图 2）。通过图 1，我们可以比较直观地获知网民在针对地域展开负面评论时容易使用的词语。



图 2. 广西、福建、河南印象词云（基于网易新闻地域负面评论）

2. 基于用户的人均负面评论指数矩阵

我们从传播者，即评论发出者角度出发，希望探究各省之间发帖用户的负面情绪强烈程度的异同。因此设计了基于用户的人均负面评论指标。

实现方法是通过 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 的人均负面评价指数，并构建了如图 3 的矩阵图。

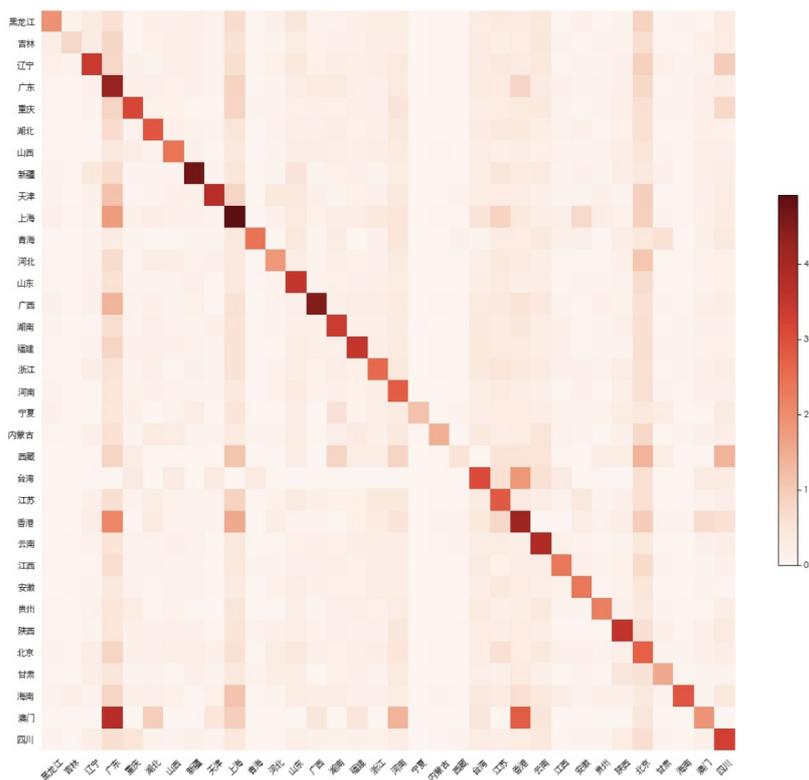


图 3. 构建周负面评价指数矩阵

在图 3 中，34 个省市自治区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顏色代表了每个发帖用户发出地域负面评价的激烈程度，方向与常用的邻接矩阵相同，为行地域对列地域。

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负面评价指数；矩阵第一行第二列单元格代表了黑龙江用户对吉林的负面评价指数，前者明显高于后者。

从人均地域负面评价指数矩阵图中，我们至少能得到两个发现。

1. 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”，遭受负面评论的可能性较高。遭受负面评价指数总和（去除自己对自身的负面评价指数）排名前五的地域分别为广东（0.270，见表 4，下同）、北京（0.219）、上海（0.196）、香港（0.155）、河南（0.128）。
2. 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）的负面评价倾向。

需要注意的是，由于这里的负面评价指数是人均负面评论数，而网民一般更喜欢将话题聚焦在自己所在的地域，谈论自己所在地域的评论也会比其他省份高，负面评论也随之增多，因此图 3 对角线的整体颜色偏深，但这并不一定代表网民

更喜欢“自黑”。不过，对于其他省份而言，其他省对我省的人均负面评价指数则具有比较意义，因此接下来的分析中将会剔除自己对自己的负面评价指数，而针对相异省份之间的负面评价指数展开分析。

3. 基于地域的负面评论比例图

我们用最小样本阈值进行数据筛选，然后进行了和上一节相似的数据处理操作，仅更换指标的分母为A地到B地的评论总数。获得了整体上的一系列发现：

首先，与地域负面评论人均数量分布不同，基于地域的负面评论比例（如图4）表明，在我们选取的大样本量的23个省份中，每个省份对于自己省份的差评比例，均远低于对于其他省份的差评比例。

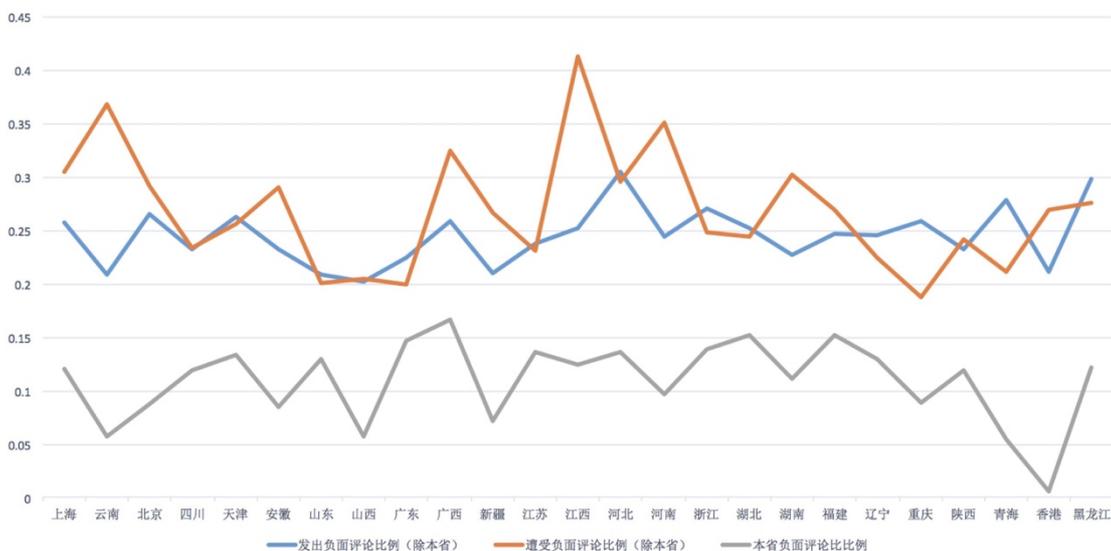


图4：负面评论比例统计图

可以想见，如果我们是网易发帖者，根据上一节人均负面评论数量指标规律，我们会更多地聚焦在谈论自己所在地域的主题上，因此关于本地的评论数量相对于其他地域则会变高。而根据本指标负面评论比例的规律，我们感受到用户虽然对于本省的负面评论数量较多，但负面评论比例则保持低水平，在5.73%~16.66%之间浮动，对于自己的平均负面评论数仅占对本省的总评论数的11.67%，即对于本省负面评论比例相对较小。

相反，用户对于其他省份的负面评论比例则会高出很多，在20.23%~30.49%浮动，对其他省份平均负面评论比例则为24.44%，比对本省的比例高出一倍有

余。因此可以看出网易新闻发帖用户呈现出对本省份“最为关切”（人均负面评价指数高），“也最为仁慈”（本省负面评价比例低）的特点。

因为各省对于其他省份的相对较高负面评论比例，我们可以得知所有省份的遭受负面评论比例也是相对本省负面评论比例较高。遭受负面评论比例在18.73%~41.24%之间浮动，我们将遭受负面评价进行排序，结果如图5所示。

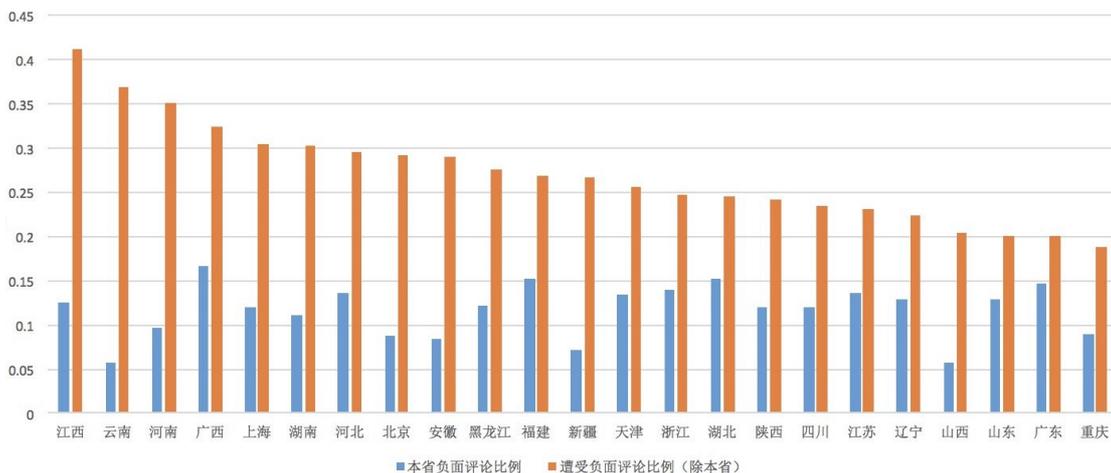


图5：遭受负面评价比例排序柱状图

4. 一个尝试：社会网络分析方法：

社会网络分析提供了网络中顶点和边分析方法，如果我们将各省份负面评论交互构建成“社会网络”，则我们可以用特征向量中心值指标表明各省份在这个网络中的中心程度，即这个节点在重要性。我们希望通过社会网络模型提供关于省份负面评价互动的另一种分析视角。

我们首先考虑网络中边和点如何定义，在众多的可能定义方式中，我们选择将某个地域的一个用户视为该地域的代表，即社会网络的顶点。而该用户的发出、遭受的负面评价数量，则是该顶点的出边和入边权重，即以上文中所说的人均负面评价指数为权重。我们进行了社交网络分析中的特征向量中心值计算，构成的社会网络示意图如下：

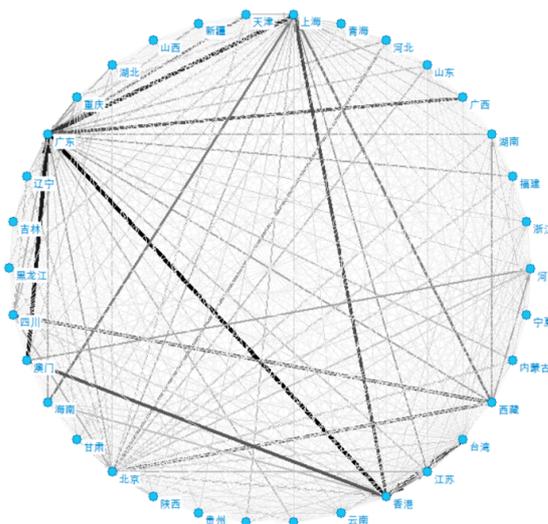


图 6. 地域负面评价指数网络图

用图论方法得出每个地域（节点）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 3）。

表 3. 各地域负面评价网络特征向量中心性与其遭受负面评价指数对比表

	负面评价网络 特征向量中心性	排名	人均遭受负面 评价指数	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

针对该图，我们有如下两点发现：

- 遭受负面评价指数基本能反映地域在负面情绪传播网络中的影响力，而后者与自身对其他地域的负面情绪及对本地域的负面情绪关系较小。这也和我们上文中发出负面评论比例与接收负面评论比例的相关性不显著有吻合之处。
- 负面评价网络图特征向量中心性和人均遭受负面评价指数的排序和大小并没有非常大的改变，两者所代表的含义在此问题中差异不大，我们最终认为在此可以不使用社交网络进行建模分析。

（二）形成地域之间负面评论的原因探究

我们针对上文所述的地域之间负面评论的原因进行了有趣而细致的探究。

1. 各个地区的用户负面评论与地域社会经济等宏观因素有怎样的关系？

是什么因素决定了哪个地域的网络用户更喜欢对其他地域展开负面评论，哪个地域成为网络用户负面评论的“众矢之的”呢？在回答这个问题的时候，我们希望借助统计手段，先从宏观层面进行相应分析。

我们使用了上述的人均负面评价指数指标，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（四）地域负面感受网络的特性探究。

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4所示。值得注意的是,在构建总体负面评价指数(对其他地区负面评价指数之和)和遭受负面评价指数(其他地区对自己的负面评价指数之和)的时候,我们都剔除了自己对自己的负面评价指数。

表4. 网络负面评论的影响因素(省市自治区层面)

变量	总体负面评价指数(“攻击”)		遭受负面评价指数(“被攻击”)	
	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. 总体负面评价指数(“攻击”)

模型显示,人均GDP能够部分解释网络用户对其他地域展开负面评论的程度。另一方面,尽管网络普及率对总体负面评价指数的影响并不显著,但二者的拟合曲线基本呈现出线性变化的趋势(如图7所示),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。总体而言,从模型中可以看出,人均GDP越高、网络普及度越高的地区,似乎表达对其他省市负面情绪的可能性也越大。

的强势区域，对预测结果并不会产生重大影响。基于此，我们使用“迁入/迁出人口”的绝对数量作为预测变量。

通过统计手段，我们发现，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。在我们的探究中，GDP和互联网普及度，均和人均负面评论指数有一定的正相关性。

2. 地域发出负面评价比例和遭受负面评价比例有怎样的关系？越爱发出负面评价的省份，就越容易遭受负面评价吗？

根据我们的直觉，喜欢攻击其他省份的网友会被其他网友“围攻嘲笑”，那么，是不是说明我们的发出负面评论比例与遭受负面评论比例会呈现这样的正相关性呢？我们以上文中基于地域的负面评价比例指标，对此问题进行了探究：

但结果出乎我们的意料。根据出发出的负面评论比例与遭受负面评论比例的散点图（图8），我们计算出皮尔逊相关系数为0.18，这表明两者相关性并不显著，我们进行线性回归也表明了这一点：

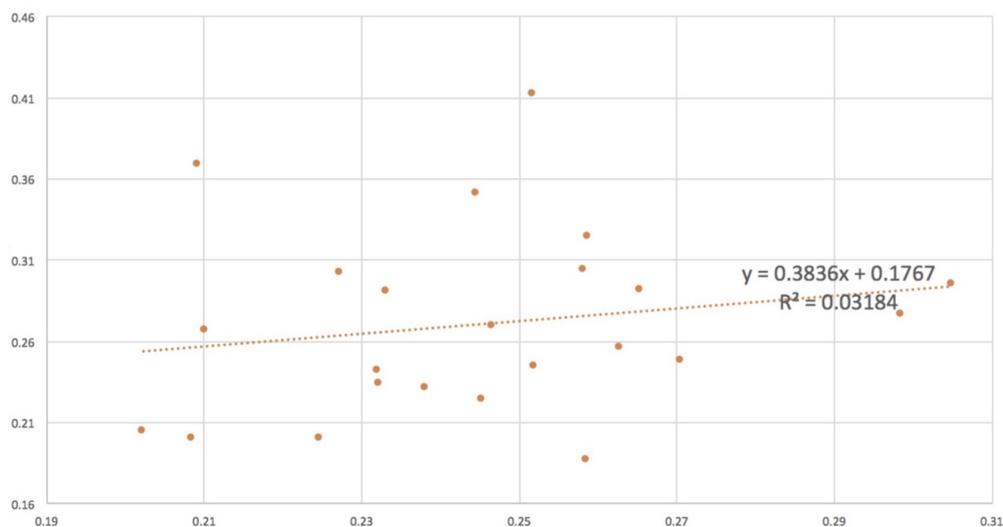


图8：发出负面评论比例（除本省）与遭受负面评论比例（除本省）两个变量的散点图及线性回归结果

由于我们只使用了23个省份的数据，由于观测数较少，而且情感识别的算法有一定的错误率，我们不能断言用户发出的负面评论对本省遭受的负面评论

没有任何相关性。但我们可以分析几个典型情况，通过人工的文本分析获取一些信息。

如果说，一个省份发出负面评论的比例可能并不是影响遭受负面评论的最主要因素，那么他们之间会有怎样的差距呢？会出现“某省份对其他省份很少有负面评价，但是依然被其他省份给以高比例的负面评价”的情况吗？我们首先对这两种指标进行排序，然后对排序结果作差：结果为负，则综合来看该省份偏向于遭受负面评价，结果为正，则综合来看该省份偏向于发出负面评价，以此识别出“差距”最大的省份。

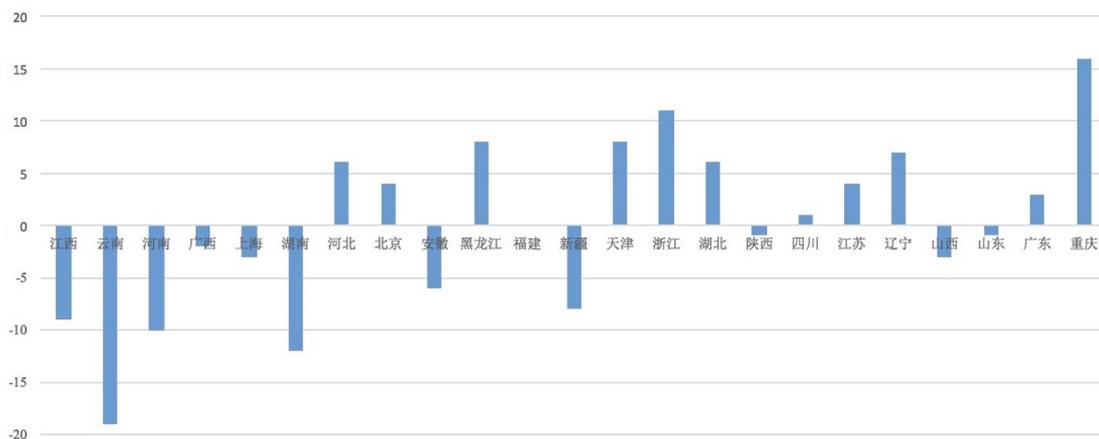


图9：各省市遭受负面评论和发出负面评论排序之差

我们可以发现负值中，云南的绝对值最高，云南在发出负面评论的各省排名中位列第21，但在遭受负面评论的各省排名中位列第2。而重庆在正值中绝对值最高，其发出负面评论比例在各省排名中位列第7，然而没有因此遭受极多的负面评论，而在各省该指标中排名最末（第23），可以说是“脾气暴烈，却不惹人厌”。这样的现象，从数据上和上文所说的相关性不显著现象吻合，我们对此非常好奇，于是深入重庆相关的评论文本，通过人工分析了解一些原因：

- 首先通过简单浏览，发现重庆的攻击范围很广，而且话语侵略性并不弱，并且不是集中在某一两个用户之中，经统计重庆发出的地域负面评价文本有2091条，涉及到的用户有1267个，其中评论示意如表5所示。

表5：重庆发出的地域负面评价示例表

攻击地域	评论内容	用户名
广东	广东二逼，糟腻马鼻	...@wx...
福建	你没有爸爸，你是个野种，福建的杂种坏子	...297_1@163...

湖南	垃圾，湖南是土匪发源地	... 142@qq...
河南	哈哈河南狗有资格喷？大家来骂河南狗，这么穷，骗子，偷井盖还喷！！	... 680@sina...
注：此处为真实示例，并无统计学意义，用户名作脱敏处理。		

➤ 那么，发出负面评价比例高达第7的省份重庆，为什么遭受负面评价比例竟然是最低水平呢？我们进行了以下的探究：

1. 重庆的负面评价地域非常广泛，每个省份都有涉及，但有侧重：

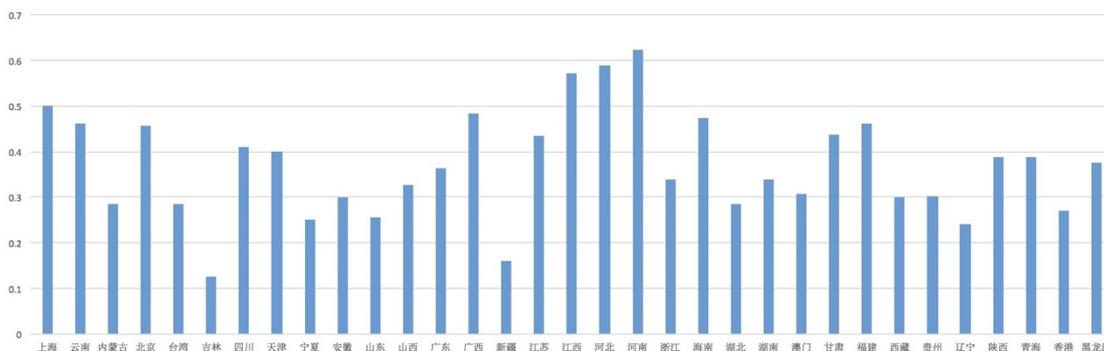


图10: 各省市遭受重庆负面评价比例

我们可以看出，重庆的负面评论比例最高省份前三名为河南、河北、江苏，这三个省份在发出负面评论比例的排名中，分别为第13，第1和第14。可能是这些省份的发出负面评论的倾向不同，导致了重庆是否经常遭受负面评论。我们以重庆-河南的关系为例：

表6: 重庆-河南地域评论模式比较示意

重庆 → 河南	河南 → 重庆
哈哈我只百度了河南人，怎么弹出来偷井盖的由来呢！！哈哈肚子都笑痛了…… 百度你太城市了我要打你...	重庆骗子还少？各大一大板
又是尼玛河南人，这地方的人咋那么坏呢？	哇塞！GPS 导航到了重庆不得哭啊？
守护井盖，防止有河人偷井盖	成都清凉，重庆热，热到死
注：以上仅为真实示例，不具有统计学意义。	

通过LDA主题模型抽取重庆对河南的评论可以发现出现最多的主题词是：井盖、骗、穷，而河南对于重庆的负面评论回复比较温和，通常是“又是地

域黑，大家不要理他”这样的类型。而河南对于重庆的其他负面评论基本是关于：热、绕路，并没有出现非常多的侮辱性词汇。

河南、江苏分别是发出负面评论排名的第13和第14名，我们发现他们对于重庆的负面评论回击并不多。经过人工文本分析之后，我们认为这可能是重庆的遭受负面评论比例相对比较低的原因之一——重庆负面评论比例较高的对象，拥有较低的发出负面评论的倾向。

若用社交网络模型类比，则可以说重庆负面评价所指向的许多省份节点，他们的入度较高、出度较低，一定程度上导致了重庆节点的负面评价入度较低。

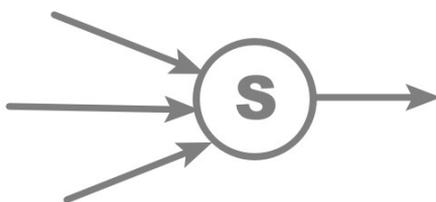


图11：重庆评论对象的网络结点示意图

当然，在这之中也有例外，比如重庆和河北两省的发出负面评价比例都位于前列，他们之间的负面评价互动“异常激烈”，太过激烈恕不在此处放出示例。

2. 我们通过，总结出主题比较多的是关于“火锅”、“高温”、“房价”、“地形”，这些主题大多数引发的负面评价并不具有侮辱性，而与之相反的河南等地，则在网友心目中形成了带有侮辱性的刻板印象，在这里我们将不需要根据评论内容就能获得的带有侮辱性刻板印象称为“侮辱性记忆点”，这些记忆点成为了其他省份网友进行负面评价的便利的“武器”，而没有这样的侮辱性记忆点的重庆，则虽然发出负面评论比例较高，但是其他省份更多地是就事论事的进行还击，这降低了负面评价的侮辱性，也提高了负面评论的门槛。

比如用LDA分析的两省的主题词，选取除地名外的主题词前五，可以发现重庆相比河南，主题词几乎不具有侮辱性。当然，对于此处的因果性需

要进一步分析，包括侮辱性记忆点的成因，以及主题词与记忆点的联系，我们期待进一步的研究。

表7：利用 LDA主题模型提取的重庆和河南负面评价文本的主题词比较表

重庆负面评价文本LDA主题词前五	河南负面评价文本LDA主题词前五
爹妈	井盖
儿女	傻
妈	狗
孙子	偷
孙女	骗子

综上所述，省份的发出负面评价比例与遭受负面评价比例是否正相关？我们的分析结果呈现并非如此。通过以重庆为典例，我们认为遭受负面评价比例还会与该省负面评价的对象有关、也与网友脑海中是否存在该省的侮辱性记忆点有关，我们期待进一步的研究去揭示其中的道理。

3. 新闻标题对地域评论会有怎样的影响？

我们对于新闻类型与地域负面评论的关系非常好奇，但是我们在实践中对于数据遇到了如下问题：

- ▶ 网易方提供的新闻正文数据不包括新闻类型字段。
- ▶ 如果使用pLSA或LDA模型进行主题分类，对于我们来说比较优秀的分析对象是新闻正文而非新闻标题，在没有既定的分类信息前提下，很难通过多样性较大而词数较少的标题归类出新闻类别。

因此，我们只对新闻进行了粗略的统计分析，希望能够有更多的时间和完善的数据进一步分析，尤其是分析新闻对于用户形成侮辱性记忆点的影响。

在此处我们进行一些统计，通过占比变化探究一些简单的影响地域评论的新闻标题变量。

表8：新闻数据统计表

标题总数	附有地域评论新闻标题数
215737	14926

带地域标题总数	附有地域评论且带地域标题数
32217	5097

我们使用概率论框架进行分析。设带地域标题为T，则带地域标题的新闻占比为 $P(T)=15.93\%$ ，设附有地域评论新闻为R，则其占比为 $P(R)=6.72\%$ 。而通过统计我们发现 $P(TR)=2.36\%$ ，高于 $P(T)P(R)=1.07\%$ ，这说明变量T和R之间具有相关关系。 $P(R)$ 为 6.72% 且 $P(R|\sim T) = P(R\sim T) / P(\sim T) = 5.36\%$ ，但此处 $P(R|T)$ 高达 15.82% 。该概率变化表明，若新闻在标题中给定地域，则评论中关于地域的言论的出现概率则会大大增高。

4. 网络热点事件对于地域负面评论会产生怎样的影响呢？

除了现实社会经济因素之外，我们猜想一些网络因素（例如互联网传播效应等）也会煽动和引导网络用户的负面情绪。标题的差异性正好验证了这一点，而我们在分析数据时也正好发现了一个有趣的网络热点导致的地域负面评论的现象，在此进行案例分析：

以云南为例。如图 12，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论，而在每个日期单独的负面评价矩阵中，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负面评论，甚至超过了北上广等地区成为“众矢之的”。为什么呢？

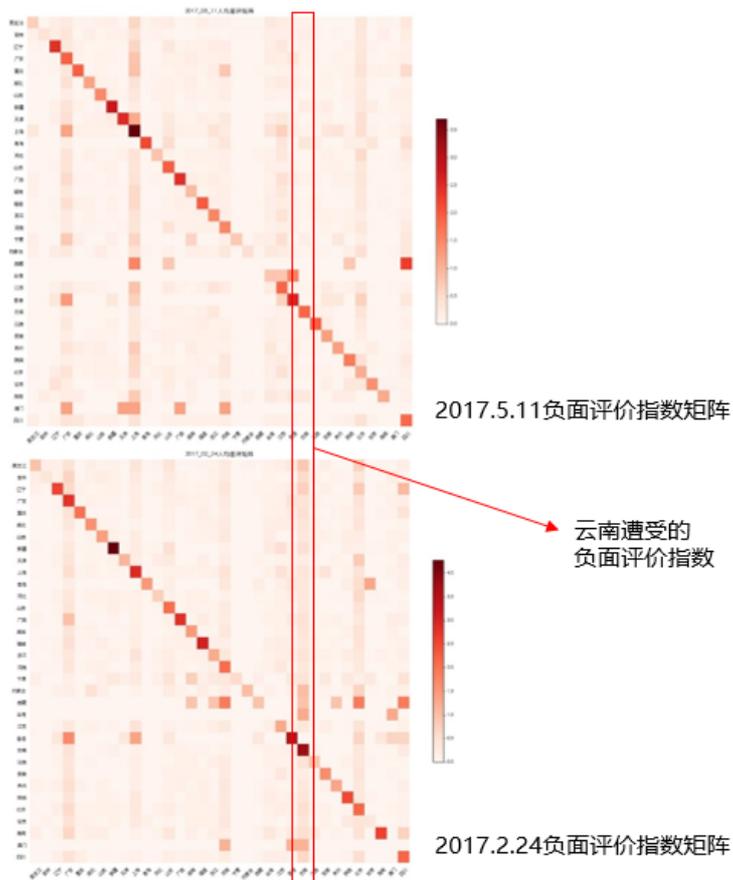


图 12. 云南遭受负面评价指数对比图

通过文本分析,我们发现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(图13),以及对云南旅游业的痛诉,云南一时成为网络话题的风口浪尖。

指责被打游客放荡 涉事民警被处分”，对评论起到了引导作用，使得关于云南的负面评论急剧上涨。

总结

在阅读了相关文献，对于数据进行了详尽的观测之后，我们使用一系列的具有创新性的探究方法，对网易新闻跟帖的地域负面评论进行了原因探究。该文的主要贡献有：

（一）地域负面评论的呈现手段：

1. 使用 **LDA 主题模型和词云地图** 构建了简明而信息量丰富的地域负面评价地图，并且 LDA 模型结果在后续案例分析中进行了多次使用。
2. 设计了基于用户的人均负面评论指数，探究用户视角下的地域负面信息关注程度，使用类似邻接矩阵的可视化方式呈现。
3. 运用传统柱状图、折线图和散点图分析负面评论比例与其他因素的相关性，以此探究基于地域的负面评论呈现方式。
4. 探究了社交网络对本任务的建模方法。

（二）地域负面评论的因素探究：

1. 探究宏观的社会经济因素对用户发出地域负面评价的影响，提出用户的地域负面评价与 GDP 以及网络普及度相关的看法。
2. 探究发出地域负面评价与遭受地域负面评价的关系，并探究重庆案例，提出遭受负面评价的影响多种可能的其他影响因素。
3. 在给定数据类型不足的前提下，对标题与新闻的关系进行了简单探究，给出了标题中出现地域会提高地域相关评论出现概率的结论。
4. 针对特殊的网络热点案例进行分析，分析了这种新的可能导致地域负面评论的模式。

当然，我们的工作还很不完善，对于需要改进的地方，我们写入了后续改进方向探讨之中。另外值得一提的是，对于我们每个人的工作量划分，可以视作三人等分，恕不再赘述。我相信缺少我们之中的任何一个人，都无法进行这样繁琐的数据交涉和处理、这样考验能力的探究工作。我们虚心接受任何改进意见，也始终为我们的三人团队感到自豪：)。

后续改进方向探讨

（一）数据

若能对完整数据进行分析，结果的可靠性将提高，并且可以对时序信息进行分析，挖掘地域攻击现象与时间的关系以及其演变模式。

（二）模型

1. 地域探测

- a) 研究者手动构造的地域词典可以筛选出绝大部分目标地域，但在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] 陈红梅. 网络表达及其对社会的影响——近十年来国外网络传播研究述略[J]. 新闻记者, 2004(9):33-35.
- [2] 徐顽强, 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.
- [3] 李贞芳, 方新子, 刘练. 影响社交媒体公众传播行为的情境因素——基于问题解决情境理论的视角[J]. 新闻大学, 2017(2):73-82.
- [4] 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.
- [5] 游家兴, 陈志锋, 肖曾昱, 等. 财经媒体地域偏见实证研究[J]. 经济研究, 2018(4).
- [6] 喻国明. 网络谣言的文本结构与表达特征——基于腾讯大数据筛选鉴定的6000+谣言文本的分析[J]. 新闻与写作, 2018(2).
- [7] 张杰. “陌生人”视角下社会化媒体与网络社会“不确定性”研究[J]. 国际新闻界, 2012(1):34-40.
- [8] 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.
- [9] 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.

- [10]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.
- [11]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.
- [12]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.
- [13]梁慧婷. 网民新闻跟帖中的语言暴力研究[D]. 湘潭大学, 2016.
- [14]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).
- [15]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.
- [16]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).
- [17]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.
- [18]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.
- [19]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.
- [20]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).
- [21]Blommaert L, Coenders M, Tubergen F V. Discrimination of Arabic-Named Applicants in the Netherlands: An Internet-Based Field Experiment Examining Different Phases in Online Recruitment Procedures[J]. Social Forces, 2014, 92(3):957-982.

- [22]Jens Damm. The Internet and the Fragmentation of Chinese Society[J]. *Critical Asian Studies*, 2007, 39(2):273-294.
- [23]Mayrl D, Saperstein A. When white people report racial discrimination: The role of region, religion, and politics[J]. *Soc Sci Res*, 2013, 42(3):742-754.
- [24]Dimaggio P, Hargittai E, Neuman W R, et al. Social Implications of the Internet[J]. *Annual Review of Sociology*, 2001, 27(1):307-336.
- [25]Carlisle J E, Patton R C. Is Social Media Changing How We Understand Political Engagement? An Analysis of Facebook and the 2008 Presidential Election[J]. *Political Research Quarterly*, 2013, 66(4):883-895.
- [26]Kuran T, Mccaffery E J. Expanding Discrimination Research: Beyond Ethnicity and to the Web[J]. *Social Science Quarterly*, 2004, 85(3):713 - 730.
- [27]Molero F, Recio P, García-Ael C, et al. Measuring Dimensions of Perceived Discrimination in Five Stigmatized Groups[J]. *Social Indicators Research*, 2013, 114(3):901-914.
- [28]Robison K K, Greshaw E M. Post-industrial transformations and cyber-space: a cross-national analysis of Internet development[J]. *Social Science Research*, 2002, 31(3):334-363.
- [29]Selwyn N. Reconsidering political and popular understandings of the 'digital divide'[J]. *New Media & Society*, 2004, 6(3):341-362.
- [30]Small T A. Online negativity in Canada: Do party leaders attack on Twitter?[J]. *Journal of Language & Politics*, 2017(1).

- [31]Wu A. Ideological Polarization Over a China-as-Superpower Mindset: An Exploratory Charting of Belief Systems Among Chinese Internet Users, 2008-2011[J]. Social Science Electronic Publishing, 2014, 8.
- [32]Hu, Anning. 2017. "On the Relation between Subjective Inequality and Generalized Trust." *Research in Social Stratification and Mobility* 49: 11-19
- [33]Jackson, Steven F. 2016. "Does China Have a Monroe Doctrine? Evidence for Regional Exclusion." *Strategic Studies Quarterly* 10(4): 64-89
- [34]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
- [35]Papacharissi, Zizzi, and Mariade Fatima Oliveria. 2012. "Affective News and Networked Publics: The Rhythms of News Storytelling on Egypt." *Journal of Communication* 62(2): 266-282
- [36]Steiglitz, Stefan, and Linh Dang-Xuan. 2013. "Emotions and Information Diffusion in Social Media-Sentiment of Microblogs and Sharing Behavior." *Journal of Management Information Systems* 29(4): 217-248
- [37]Yardi, Sarita, and Danah Boyd. 2010. "Dynamic Debates: An Analysis of Group Popularization over Time on Twitter." *Bulletin of Science, Technology & Science* 30(5): 316-327