

2020-12-25~2021-1-12

数据获取

在此次新型冠状病毒（COVID-19）传播这一重大公共卫生事件情境下，为了深描中国大众的网络社会心态变化，分析从疫情最初不被重视，到逐渐扩散、爆发，再到全国上下联防联控、万众一心，新冠肺炎疫情得到有效控制，最后全国逐渐复工复产、经济逐渐恢复这一过程中中国大众网络社会心态的体现和分布情况，我们决定选择获取从 2019 年 12 月中旬至 2020 年 6 月的官方新闻文本，以及从 2020 年 1 月到 6 月初的各类疫情新闻的“热评”。除此之外，我们注意到：在 2020 年 12 月中旬至今（2021 年 1 月 19 日左右），北方疫情出现了较为严重的反弹，包括河北藁城、北京大兴、黑龙江黑河、绥化望奎、惠七等地均出现了大量的确诊病例和无症状感染者，在全国逐渐基本回到了正常生活轨道的时刻，疫情突然再度反弹，无疑会对大众心态产生较大的冲击，因此我们额外选取了 2020 年 12 月中旬至 2021 年 1 月中旬的微博及“热评”，分析在天气转冷、疫情反弹的情势下，中国大众网络社会心态又会发生什么样的变化。

为了获取最真实、最能反映中国大众网络社会心态的新闻评论，我们最终决定选择获取新浪微博相关正文内容及热评，原因及相关做法如下：

一、主要原因

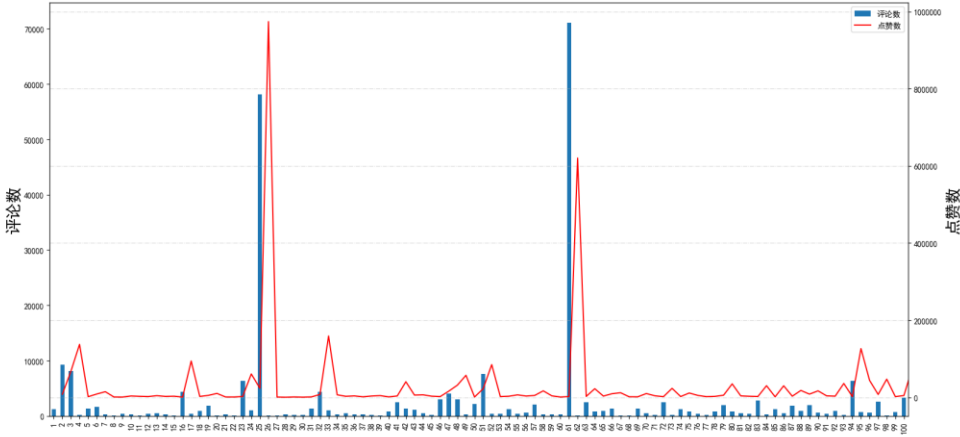
如今，新浪微博是中国网民的重要活动平台，微博上的许多“大 V 用户”拥有数以千万计甚至上亿计的粉丝量，其发布的微博具有权威性和引导性，其微博下的评论也可以反映出中国大众的主流网络社会心态。另外，微博平台不似其他众多平台，在微博评论中，网友可以不拘于条框，将自己内心真实想法用“评论”的方式表达出来，而微博上给评论“点赞”的功能则能够反映其他网友对于这条评论的认可程度，点赞数较多的前几条（约 5~8 条）评论将成为“热评”，而使用“热评”作数据的好处有：1、评论内容真实可信，可有效避免“机器粉”刷屏；2、评论内容为多数网友所赞同；3、减少了后期筛选数据的工作量，免去了数据“拟合”的过滤筛查，我们可以直接使用大量“热评”数据进行分词和提取。在本次作业中，我们爬取了热点新闻下相对应的“热评”，这些评论不仅点赞数最多，且反映出来的网络社会心态也最真实；

二、URL 选择

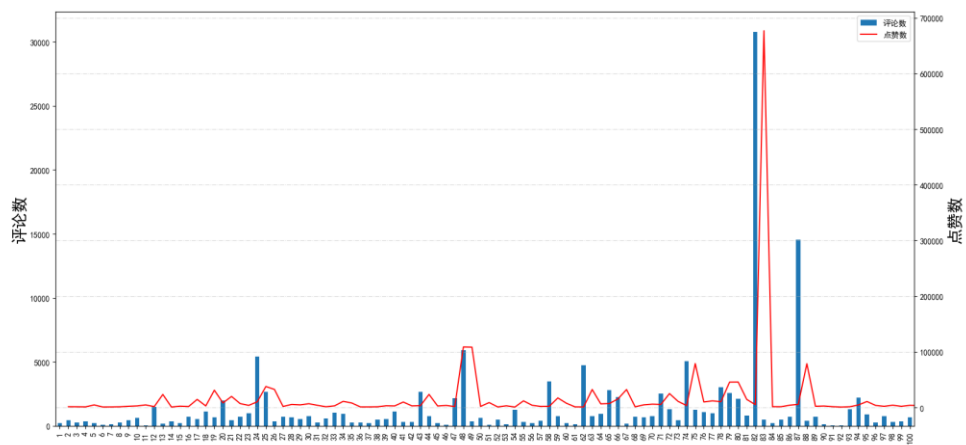
网页版微博（weibo.com）可以方便地获取对应时间日期段的微博内容（可以直接指定月份爬取），可以减小后期从大量数据中再划分数据的任务量，相应地最难获取；手机端微博（m.weibo.com）数据爬取难度相比前者更小，但无法指定对应日期的微博内容，后期数据处理时需要另外进行事件阶段性划分；移动端微博（weibo.cn）最易爬取，但数据最为粗糙，后期处理时更加需要精细的筛选。综上，我们最终选择爬取网页版微博的数据；

三、数据源选择

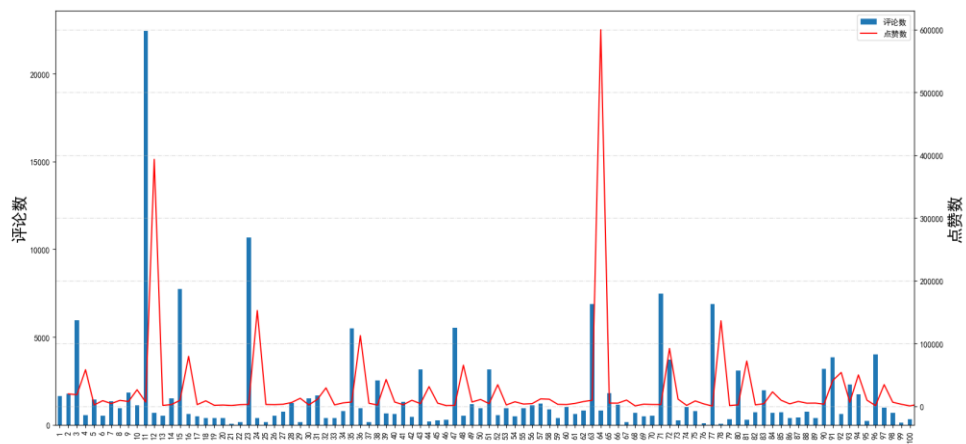
权威性的重点新闻及其下的评论数据是我们爬取的目标，何谓“重点新闻”呢？通俗理解就是指评论量、点赞量及转发量等远超其他新闻的新闻。我们横向对比了新浪微博上的“大 V”新闻官煤，包括：央视新闻、人民日报、人民网、荔枝网、新华视点、中国青年报等，通过对每家媒体的微博内容进行抽样调查，发现只有央视新闻和人民日报满足这一要求，我们分别对人民日报和央视新闻的微博点赞数和评论数进行了四次抽样，每次随机抽取 100 条微博内容，并获取其下的评论数和点赞数，得到如下图的结果：



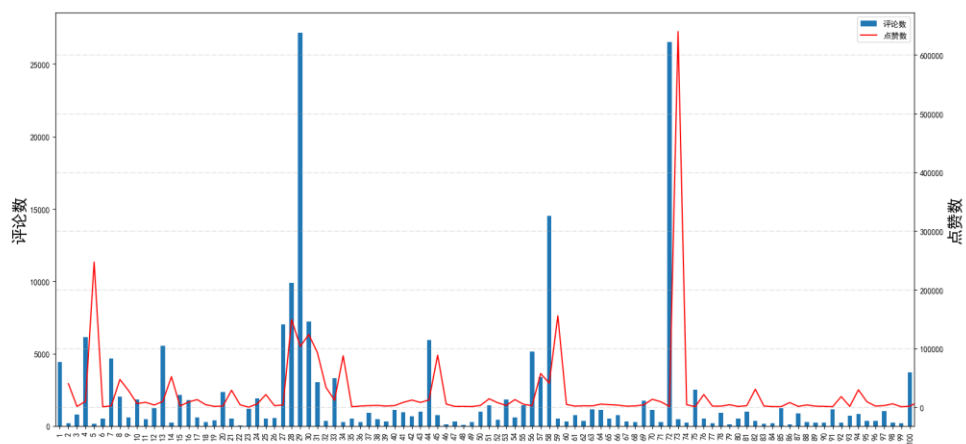
（央视新闻微博点赞、评论抽样分布图 1）



(人民日报微博点赞、评论抽样分布图 2)



(央视新闻微博点赞、评论抽样分布图 3)



(人民日报微博点赞、评论抽样分布图 4)

去除个别微博热度过高导致评论和点赞数异常高，我们发现：“央视新闻”和“人民日报”两家媒体平均每条微博的评论量可以达到 4000~8000，平均点赞量可以达到 8000~20000，且这一水平较为稳定。而对于热点微博事件评论量可以达到超过 10000，而点赞量更可以达到数十万，如图中“红线”的高凸处。而对比其他媒体的只有 1000~2000 的点赞量和几百评论量来说，央视新闻、人民日报发布的微博可以称为“重点新闻”，而对于如此庞大的数据量，我们有充分的理由认为，央视新闻和人民日报这两个新浪“大 V 用户”的新闻内容及相关热评可以反映中国大众网络社会心态；

四、数据爬取

虽然网页版微博(weibo.com)可以方便地获取对应时间日期段的微博内容，但是三者中最难获取的，由于微博反爬机制的存在，我们最终采用“模拟人浏览网页”的方案进行网页源码获取，即采用 Python 并引入 Selenium 库，通过半自动化操作来获取动态加载的网页，而后再对网页源代码进行解析、正则提取，最后获得了相关微博的正文、点赞数、转发数、评论数，并爬取了所有的“热评”，最终累计获取微博 19000 余条，微博“热评”共 1210000 余条。除微博外，我们在其他平台爬取的纯新闻数据共 30000 余条。我们接下来的任务就是对这些数据进行进一步的筛选和提取。

2021-1-13~2021-1-21

事件阶段划分

在获取了大量数据后，我们对这些数据进行了第一次筛选——先进行一次粗筛选，先根据如下事件阶段标准对数据进行划分：

第一阶段：2019 年 12 月末——2020 年 1 月下旬，疫情从刚显苗头、未得重视到全国人民逐渐意识到新冠肺炎疫情的严重性（由于 2019 年 12 月中旬以前关于疫情的报道相当少，因此并未选择获取那时的数据）；

第二阶段：2020 年 1 月末——2 月上旬——由于人民逐渐意识到疫情严重和逐渐主动居家、不走亲访友，但随之而来的便是物资匮乏、物价上涨，此时好多网友纷纷表示物资“太贵了”、“买不到”，食品、用品等价格上涨；

第三阶段：2020 年 2 月中旬至下旬，政府严格统一管控、物资配给，使得居民生活得到了保障，人民逐渐从恐慌到安心，全国上下有序联防联控；

第四阶段：2020 年 3 月以后，各省开始有序复工复产……进一步划分，我们发现，在 2020 年 3 月之中，我国逐渐聚焦于复工复产的稳健进行、各地防控工作的稳定开展，我们迎来了逆行天使和抗疫英雄们的凯旋，各省各地欢迎医护人员们“回家”……

第五阶段：2020 年 4 月~5 月，我国疫情形势已得到了稳定良好的控制，而国外疫情情势逐渐变得十分严峻，全世界范围内病魔肆虐，中国境外除澳门均为高风险地区，我国人民开始对外国疫情表示出了担忧等情绪……2020 年下半年，我国疫情防控工作一直顺利进行，各地偶有病例也迅速隔离、追根溯源，疫情一直得到了有效稳定的控制；

第六阶段（额外分析）：入冬以后，从 2020 年 12 月中旬开始至今，中国北方（如河北、吉林、黑龙江、北京等地）疫情由于聚集性活动和低温等原因严重反弹，最严重时日新增确诊病例可达近 100 例，而这正是大学生即将进入寒假、新春佳节将至的时刻，网友们对能否回家过年、大学生能否正常返乡表现出了担忧，全国上下为北方防疫工作加油鼓劲……

关键词提取

为了更加细致地分析中国网络大众在疫情防控至复工复产再到疫情反弹这一过程中的心态趋势，我们需要从他们的大量评论中提取关键词和情绪核心词，相关提取方法如下：

一、分词

由于获取到的“热评”均为长度不等的句子，因此首先需要粗略地将这些句子划分为词语，我们引入了一个强力的工具——jieba（“结巴”）分词库（<https://github.com/fxsjy/jieba/>），该库的主要功能是将长句切分成词语，同时用户可以创建自己的新词词典，以便分词更加精确。此外，jieba 库包含基于 TF-IDF 算法的关键词抽取功能，但该函数只能传入待提取文本，也即：此函数运作时是先分词，再基于 TF-IDF 算法进行词频分析和逆文本频率指数分析，

最终得出 Top-K 个关键词，因此如果使用该函数，我们只会得到最终分析出来的关键词，其分析过程中的分词成果对于我们是透明的，因此最终我们选择只使用 jieba 库的分词功能，不使用现成的关键词提取功能，以便更清晰地观察特定词语在评论中出现的情况，后续再通过精细的分词来获取更有效的词语频数表，进而自己实现词频统计和逆文本频率指数分析的相关代码并作图分析。

在初次分词的过程中，我们发现：句子中存在的大量中文标点（如，。、【】等）、大量无意义单字（如的、啊、以等许多单字）、无法反映网络心态的词汇（好像、看着、身边、网络等）对中国大众网络社会心态分析无益，因此我们在原有停用词模板（近 3000 条停用词）基础上通过一遍遍筛选过滤，添加了近 200 个新的停用词，这样即保证了文本数据的有效性，又不至于将文本数据过滤得过于干净而丧失有效数据；与此同时，我们发现 jieba 库虽然有新词识别算法，但是对于疫情期间出现的特定词汇如“抗疫”、“火神山”、“雷神山”、“境外输入”、“新型冠状病毒”、“输入病例”、“一级响应”、“逆行天使”、“核酸检测”等不能进行很好的识别和分词，为此，我们建立了新词词典并添加了 200 余条疫情期间网友们常用的新词和特定词汇，在分词时便能更好地屏蔽无意义符号和词语，分出和收录更多能切实反映大众心态的词汇了，同时我们还在新词表中添加了中国地名共 400 余个。

二、关键词提取

接下来我们对划分好的词汇进行关键词提取。

在剔除近 3000 个停用词及引入新词词典后，我们将 61000 余条“热评”语句划分成了对心态分析有利的词汇，我们引入 Python 中 collections 库中的 Counter 类，向 Counter 类中传入列表并将其传入 dict() 方法中，即可自动生成一个“关键词-频数”的字典，对于每个关键词，Counter 类都会自动计算出其在所有词汇中出现的总次数，而后再调用 sorted() 函数进行排序，我们就得到了逆序排列的“热评”高频词表了。

下一步我们选择自己实现词频 (TF: Term Frequency) 和逆文本频率指数 (IDF: Inverse Document Frequency) 的分析过程，以便更好地了解关键词在各月份的疫情微博“热评”中的分布情况。

TF-IDF 是一种统计方法，用来评估词语对于一个文件集或一个语料库中一

份文件的重要程度，其在文件中出现次数越高，可以初步认为该词语重要性越大，但同时会随着它在语料库中出现的频率成反比下降。

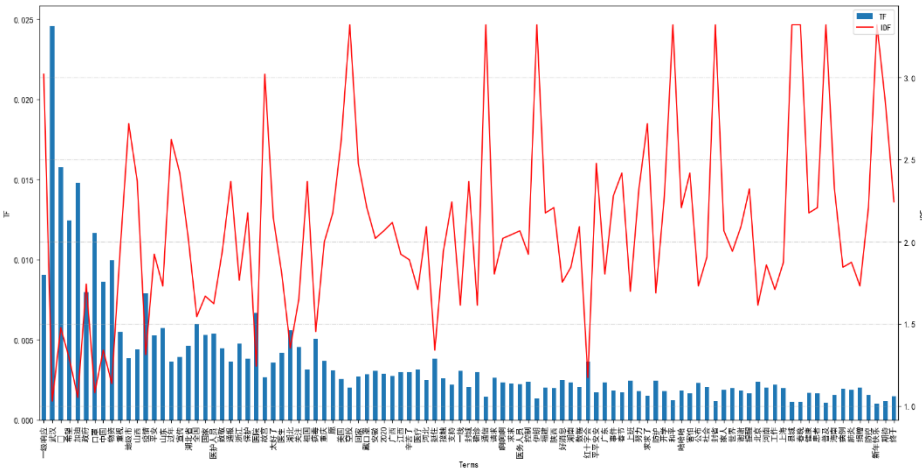
词频(TF)是指某一个给定的词语在文件中出现的频率，逆文本频率指数(IDF)是一个词语普遍重要性的度量，而由于特定词语可能在文件中一次都没出现过，因此实际计算时应将分母加 1。但不能单纯地认为词频越高，重要性越大，因为语料库中某一文件可能会反复出现该词语而使 TF 值偏高，因此我们引入 TF-IDF 值来对词语重要性进行考量：一个词语的 TF-IDF 值即为一个文件或语料库中该词语的 TF 值与 IDF 值的乘积，某一词语越重要，其在该文件内即具有高词语频率，在整个文件集合中即具有低文件频率，这样便会产生高权重的 TF-IDF 词语，进而筛选出我们需要的“关键词”。非常值得注意的是，我们之所以要不断完善停词表 (stop words)，就是因为类似“的”、“以”之类的无意义词可能会占据总词频的大部头，而它们对项目分析几乎无用，所以应将其忽略掉。

根据 TF 和 IDF 的计算公式 (IDF 计算时分母加 1)，我们取某一阶段的所有评论语句构成该阶段的完整语料库，认为每一条热评构成一个文件，在计算 IDF 时，需要判断某词语是否在某条“热评”(即语料库中的文件)中出现。

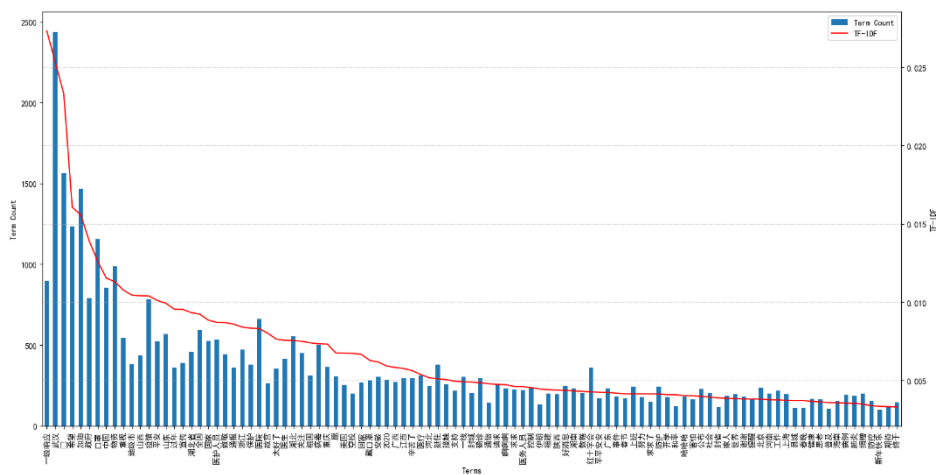
$$TF = \frac{\text{关键词的出现次数}}{\text{文档中所有词的数量}}$$

$$IDF_w = \log \frac{N}{\sum_{i=1}^N I(w, D_i)}$$

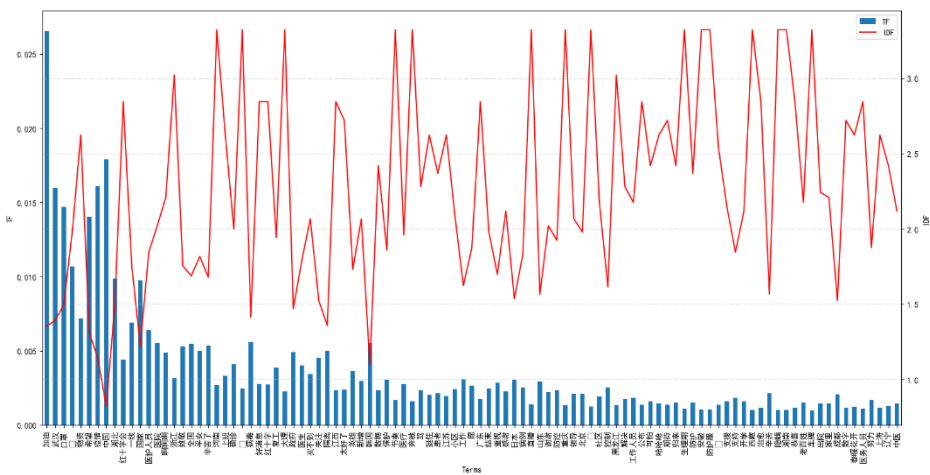
首先我们统计、计算、绘制并观察各个阶段“热评”词汇的 Term Count (词数)、Term Frequency (TF: 词频)、Inverse Document Frequency (IDF: 逆文本频率指数) 和 TF-IDF 综合值的条形图和折线图。



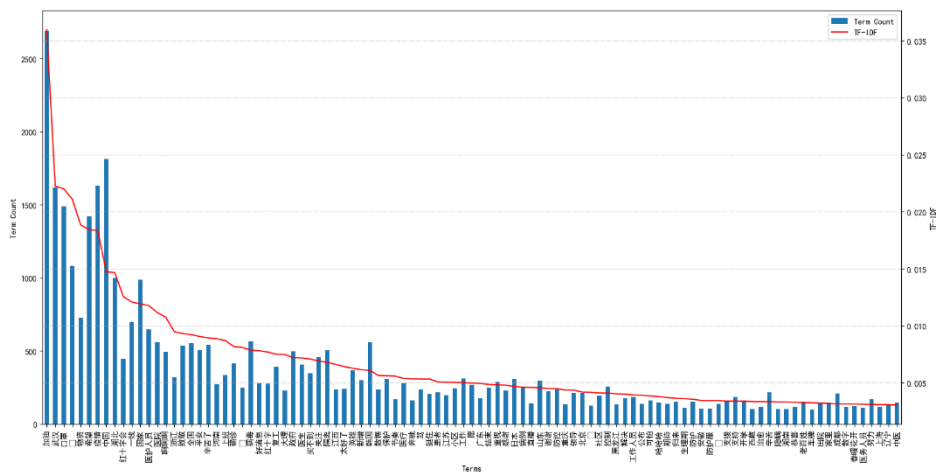
(一阶段词语 TF (条形) 及 IDF (折线) 统计图)



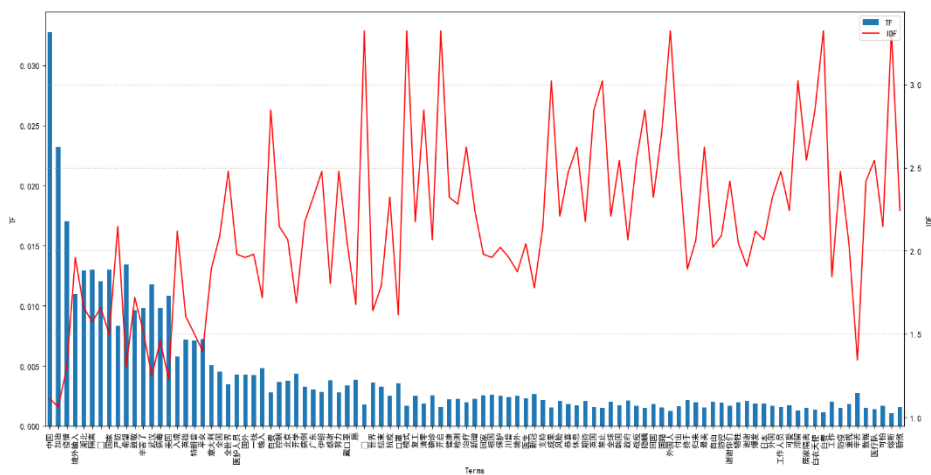
（一阶段词语词数（条形）及TF-IDF（折线）统计图）



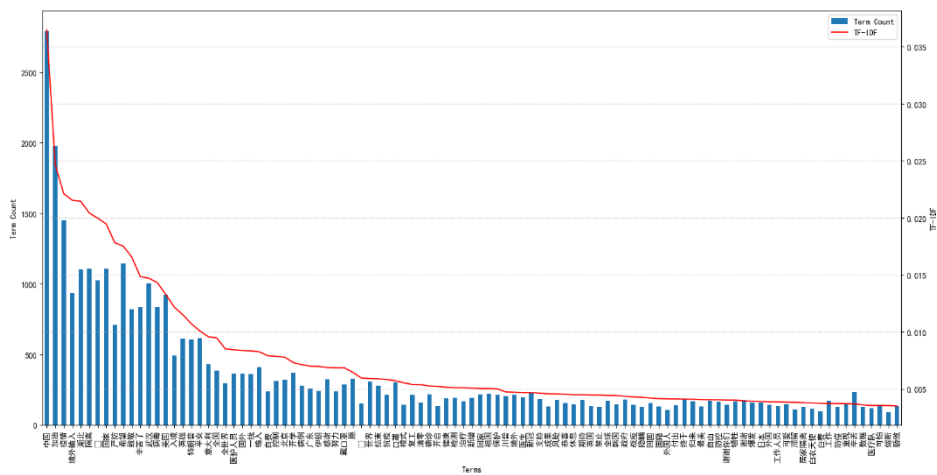
（二阶段词语 TF（条形）及 IDF（折线）统计图）



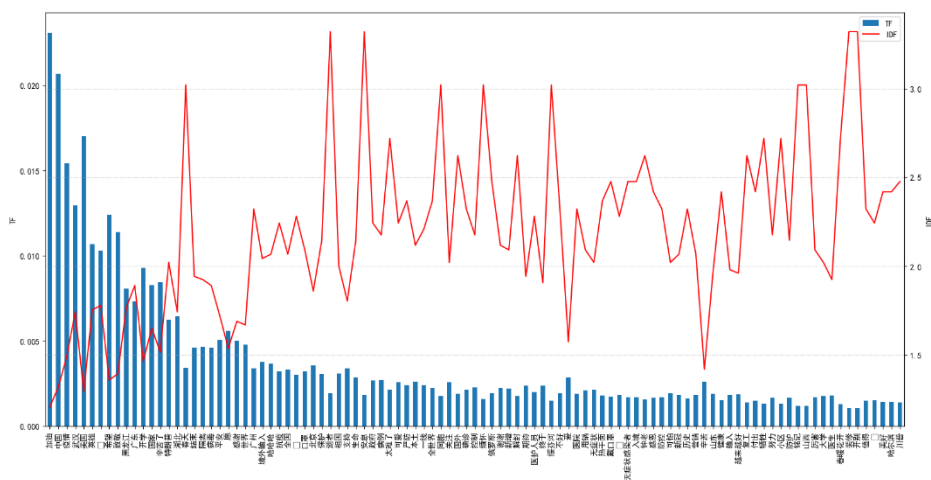
（二阶段词语词数（条形）及TF-IDF（折线）统计图）



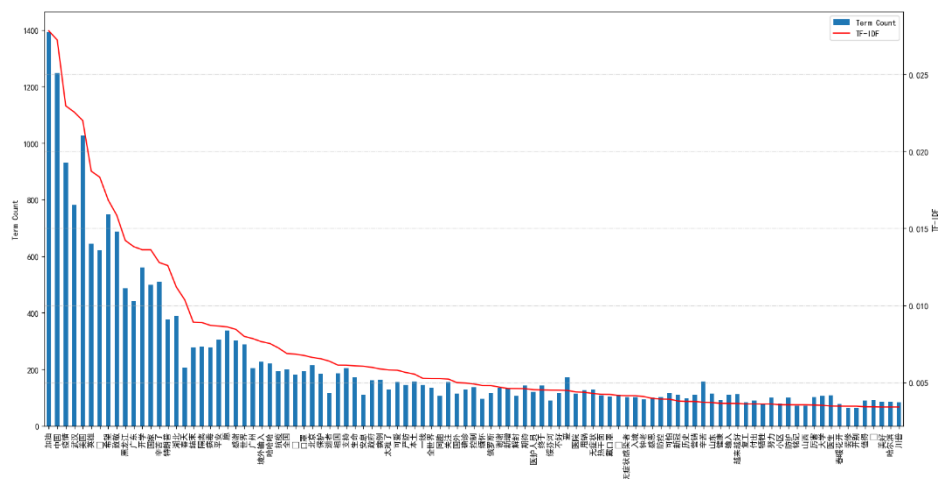
(三阶段词语 TF (条形) 及 IDF (折线) 统计图)



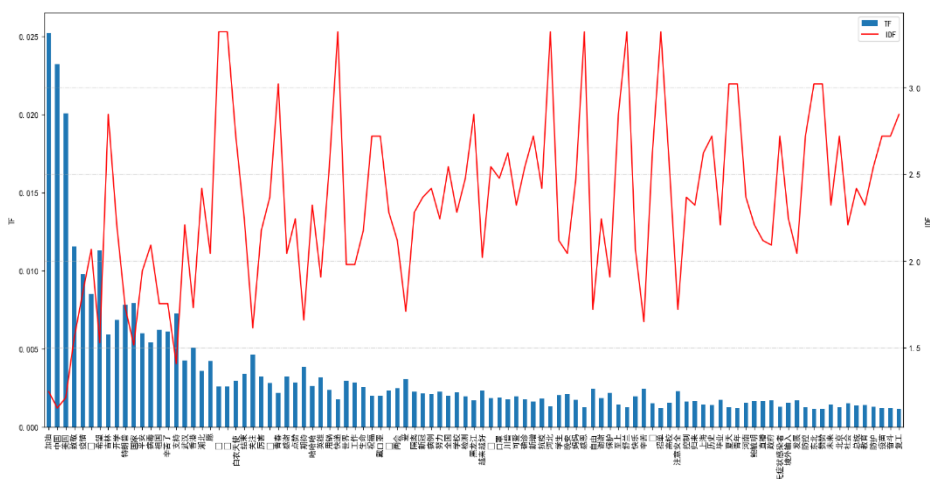
(三阶段词语词数 (条形) 及 TF-IDF (折线) 统计图)



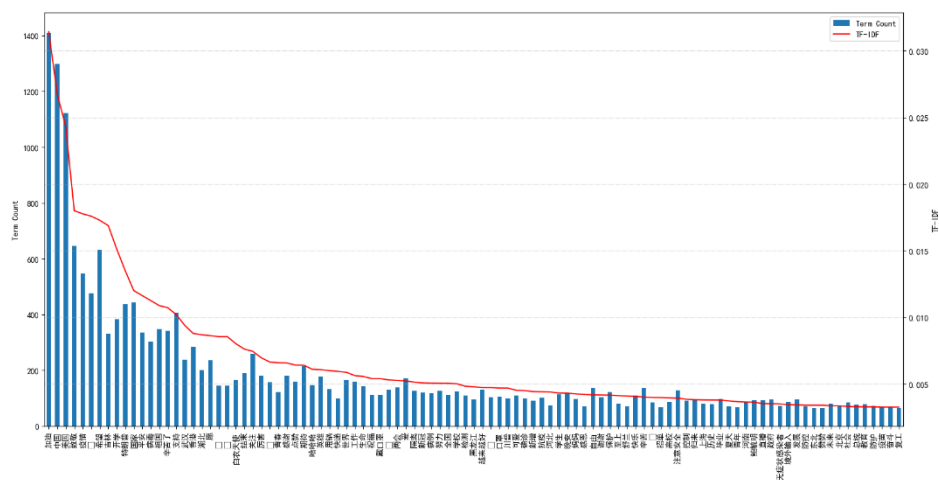
(四阶段词语 TF (条形) 及 IDF (折线) 统计图)



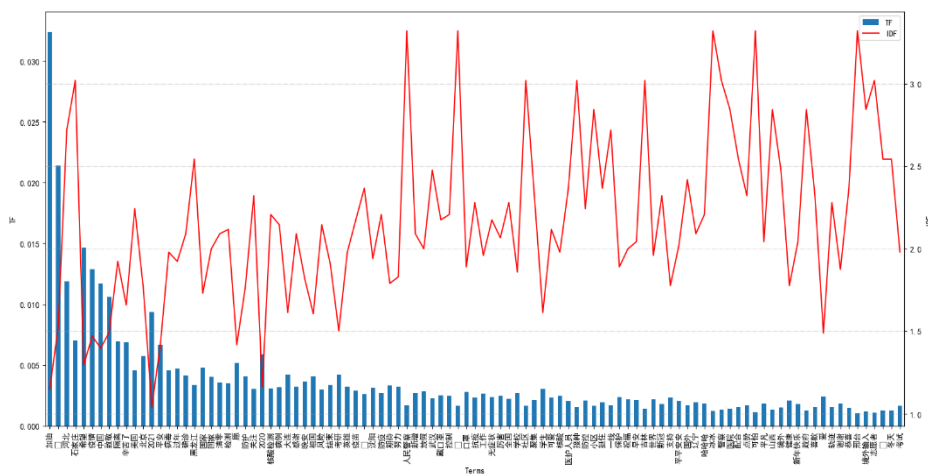
(四阶段词语词数 (条形) 及 TF-IDF (折线) 统计图)



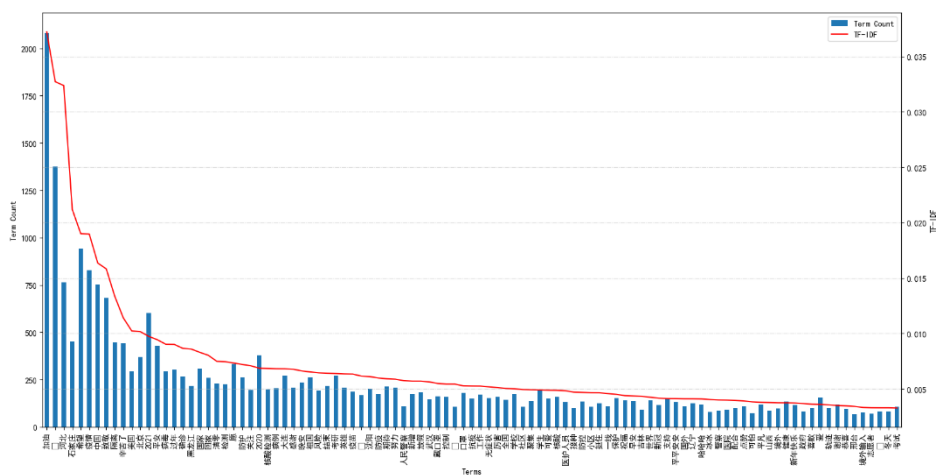
(五阶段词语 TF (条形) 及 IDF (折线) 统计图)



(五阶段词语词数 (条形) 及 TF-IDF (折线) 统计图)



（六阶段词语 TF（条形）及 IDF（折线）统计图）



（六阶段词语词数（条形）及 TF-IDF（折线）统计图）

通过统计图，我们非常直观地看到，并不是 TF 越大的词语，其 TF-IDF 值就越高，因为还会有 IDF 值的制约。通过对每个阶段 TF & IDF 统计图和 Term Count & TF-IDF 统计图的综合分析，选取每个阶段 TF-IDF 权重前 100 的词汇，我们就得到了每个阶段前 100 个关键词——

第一阶段	第二阶段	第三阶段	第四阶段	第五阶段	第六阶段
一级响应	加油	中国	加油	加油	加油
武汉	武汉	加油	中国	中国	🇨🇳
🇨🇳	口罩	疫情	疫情	美国	河北
希望	🇨🇳	境外输入	武汉	致敬	石家庄
加油	物资	湖北	美国	疫情	疫情
政府	希望	隔离	英雄	🇨🇳	希望
口罩	疫情	🇨🇳	🇨🇳	希望	中国
中国	中国	国家	希望	吉林	致敬
物资	湖北	严防	致敬	开学	隔离
重视	红十字会	希望	黑龙江	特朗普	辛苦了
地级市	一线	致敬	广东	国家	美国
山西	国家	辛苦了	开学	平安	北京
疫情	医护人员	武汉	国家	病毒	2021
平安	医院	病毒	辛苦了	祖国	平安
山东	啊啊啊	美国	特朗普	辛苦了	病毒
过年	浙江	入境	湖北	支持	过年
宣传	致敬	英雄	春天	武汉	确诊
湖北省	全国	特朗普	结束	香港	黑龙江
全国	平安	平安	隔离	湖北	国家
国家	辛苦了	意大利	病毒	愿	回家
医护人员	河南	全国	平安	🤔	清零
致敬	上班	全世界	愿	白衣天使	检测
通报	确诊	医护人员	感谢	结束	愿
浙江	病毒	国外	世界	关注	防护
保护	坚持	一线	广州	厉害	关注
医院	好消息	输入	境外输入	👍	2020
故宫	红十字	自费	哈哈哈	青春	核酸检测
太好了	复工	控制	抗疫	感谢	病例
医生	大理	北京	全国	点赞	大连
湖北	政府	开学	👍	期待	感谢
关注	医生	病例	口罩	哈哈哈	晚安

第一阶段	第二阶段	第三阶段	第四阶段	第五阶段	第六阶段
祖国	买不到	广东	北京	英雄	祖国
病毒	关注	伊朗	保护	甩锅	风险
重庆	隔离	感谢	逝者	快递	结束
愿	太好了	努力	祖国	世界	考研
美国	英雄	戴口罩	支持	工作	英雄
空投	新增	愿	生命	生命	疫苗
回家	韩国	👨👩👧👦	安息	祝福	👨👩👧👦
戴口罩	数据	世界	政府	戴口罩	沈阳
安徽	保护	结束	病例	👨👩👧👦	防疫
2020	节奏	抗疫	太难了	两会	期待
广西	医疗	口罩	可爱	爱	努力
江西	奔驰	模式	严防	隔离	人民警察
辛苦了	骂	复工	本土	新冠	新增
医疗	挺住	清零	一线	病例	放假
河北	患者	确诊	全世界	努力	武汉
挺住	江苏	开启	同胞	全国	戴口罩
接触	小区	健康	关注	学校	控制
支持	工作	检测	国外	检测	👨👩👧👦
一线	愿	治疗	确诊	黑龙江	口罩
封城	广东	新增	控制	越来越好	抗疫
确诊	结束	回家	缅怀	👨👩👧👦	工作
通信	重视	祖国	俄罗斯	口罩	无症状
请求	感谢	保护	谢谢	川普	厉害
啊啊啊	日本	川普	新增	可爱	全国
求求	病例	境外	解封	确诊	学校
医务人员	直播	医生	期待	新增	社区
控制	山东	新冠	医护人员	抗疫	聚集
伊朗	谢谢	支持	终于	河北	学生
福建	防控	成果	绥芬河	学生	可爱
陕西	重庆	风险	不好	晚安	核酸
好消息	领导	恭喜	爱	妈妈	医护人员

第一阶段	第二阶段	第三阶段	第四阶段	第五阶段	第六阶段
湖南	北京	休息	医院	感恩	接种
数据	👉	期待	甩锅	自由	防控
红十字会	社区	英国	无症状	谢谢	小区
平平安安	控制	禁止	热干面	保护	挺住
广东	黑龙江	全球	戴口罩	至上	一线
事件	解决	韩国	👉	舒兰	保护
春节	工作人员	政府	无症状感染者	快乐	祝福
上班	公布	战疫	入境	辛苦	早安
求求了	可怕	隐瞒	钟老	❤️	吉林
防护	哈哈哈	回国	感恩	高校	世界
开学	期待	国籍	防控	注意安全	新冠
和平	归来	外国人	可怕	控制	支持
哈哈哈	生理期	付出	新冠	归来	平平安安
害怕	防护	终于	历史	上海	国外
公布	安徽	归来	营销	历史	辽宁
社会	防护服	最美	辛苦	毕业	哈哈哈
封省	👉	自由	山东	夏天	冰冰
家人	支援	防控	健康	青年	警察
世界	支持	谢谢你们	输入	河南	医院
谢谢	开学	牺牲	越来越好	直播	配合
提醒	西藏	谢谢	复工	政府	点赞
北京	治愈	爆发	付出	无症状感染者	可怕
河南	辛苦	日本	牺牲	境外输入	平凡
工作	隐瞒	外国	努力	发展	山西
上海	湖南	工作人员	小区	防控	境外
县城	恭喜	可爱	防护	东北	健康
春晚	老百姓	滞留	铭记	赞	新年快乐
健康	车牌	居家隔离	山西	未来	政府
患者	出院	白衣天使	厉害	北京	喜欢
普及	家里	白费	大学	社会	爱
海南	成都	工作	医生	总统	轨迹

第一阶段	第二阶段	第三阶段	第四阶段	第五阶段	第六阶段
病例	数字	防疫	春暖花开	教育	谢谢
肺炎	春暖花开	重视	装修	防护	恭喜
捐赠	医务人员	辛苦	值得	疫苗	邢台
防控	努力	数据	❤️	奋斗	境外输入
新年快乐	上海	医疗队	美好	鼓劲	志愿者
期待	辽宁	可怕	哈尔滨	天津	👉

三、心态分析

（一）心态词典建立与关键词碰撞计划方案

得到了每个阶段的关键词，我们就可以对这些关键词进行心态分析了。

我们首先将心态词所表达的心态归一化，即将心态具体量化，映射到(0.00, 1.00)区间上，对每个心态赋一个“心态值”，即建立一个“心态词(attitude)→心态值(value)”的映射。具体做法是：

以“平静”（中立）心态为中线，赋值为 0.50；

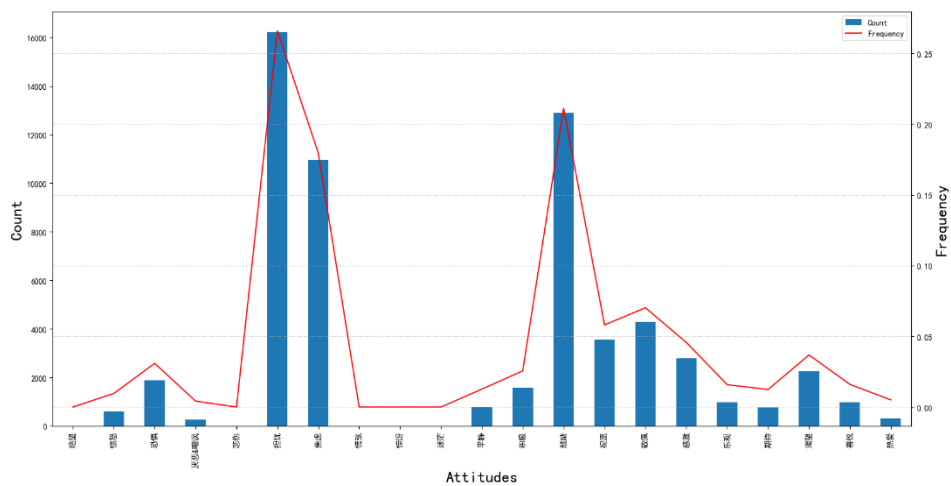
“心态值”小于 0.50 的心态词被认定为“负面心态”，包括：绝望 0.00，愤怒 0.05，恐惧 0.10，厌恶&嘲讽 0.15，悲伤 0.20，担忧 0.25，焦虑 0.30，慌张 0.35，惊讶 0.40，迷茫 0.45。其中越接近 0.00，表示负面心态越严重；

“心态值”大于 0.50 的心态词被认定为是“正面心态”，包括：积极 0.55，鼓励 0.60，祝愿 0.65，敬佩 0.70，感激 0.75，乐观 0.80，期待 0.85，渴望 0.90，喜悦 0.95，热爱 1.00。其中越接近 1.00，表示正面心态越积极。

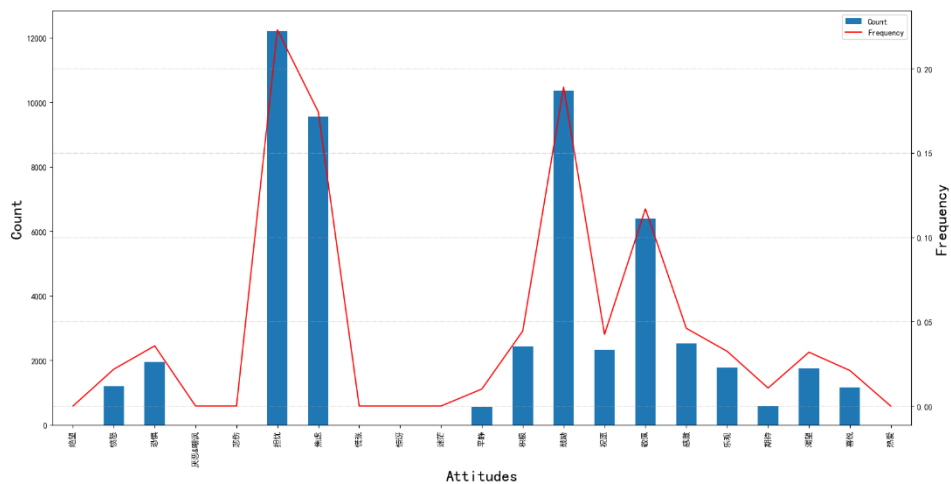
而后我们建立“心态词典”，即建立一个“核心词(key)→心态值(value)”的映射，通过对各阶段评论内容的浏览和分析，并借助提取后的逆序词频表，我们人工建立了一个 600 余条的“情绪/关键词→心态值”的心态词典。额外地，我们向心态词典中加入了各地地名，经过大量评论分析后我们认为，评论中出现各地地名可以经典地反映出网友对这些地疫情的“焦虑 0.30”、“担忧 0.25”和“鼓励 0.60”的心态。我们将之前得到的各个阶段的关键词与心态词典进行碰撞，若某关键词与心态词典的键重合，则可由“键”对应的“值”找到该关键词对应心态词，进而根据该关键词的词频数（Term Count）得到对应心态词的词频数，最后即可得到每个碰撞成功的关键词对应心态词的词频数，进而分析各阶段中国大众的网络社会心态。

（二）关键词与心态词典的碰撞

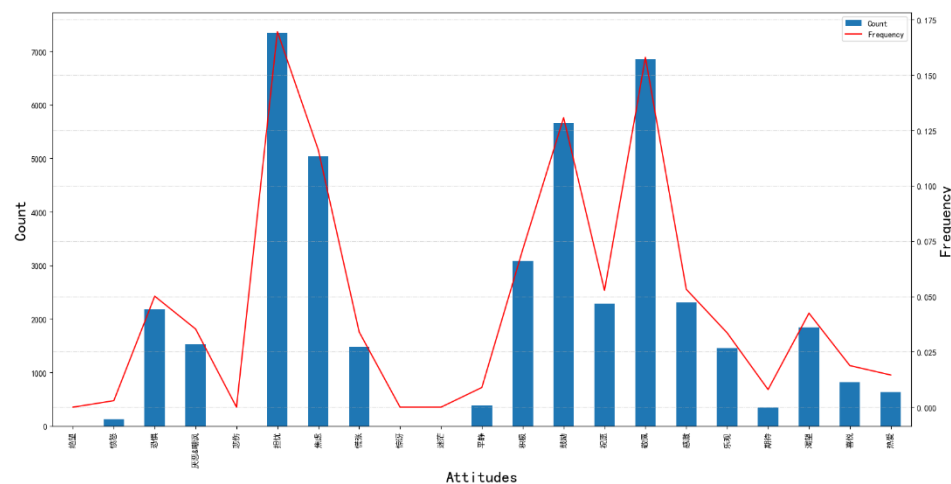
接下来我们按照上述方案进行每个阶段前 100 个关键词向心态词典的碰撞，并得到各个阶段每个心态的分布情况（横轴为心态词，纵轴为心态频数（条形）和心态词频率（折线））：



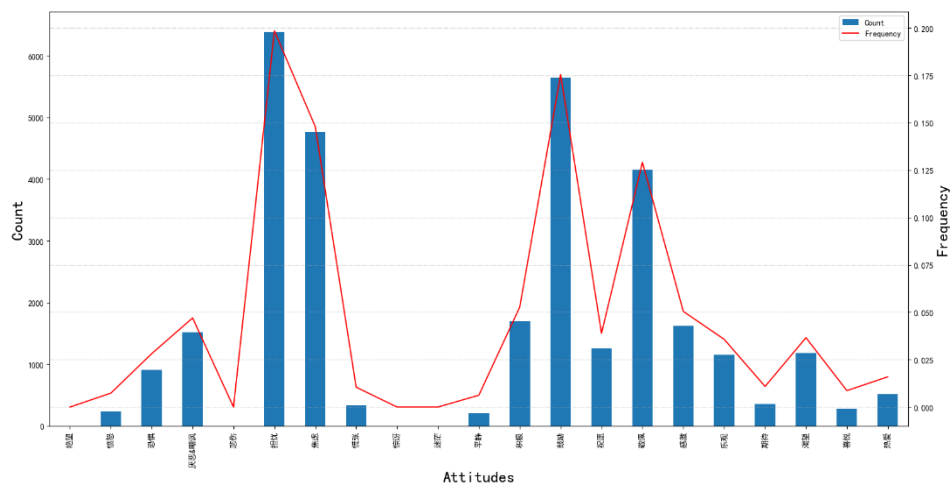
(一阶段)



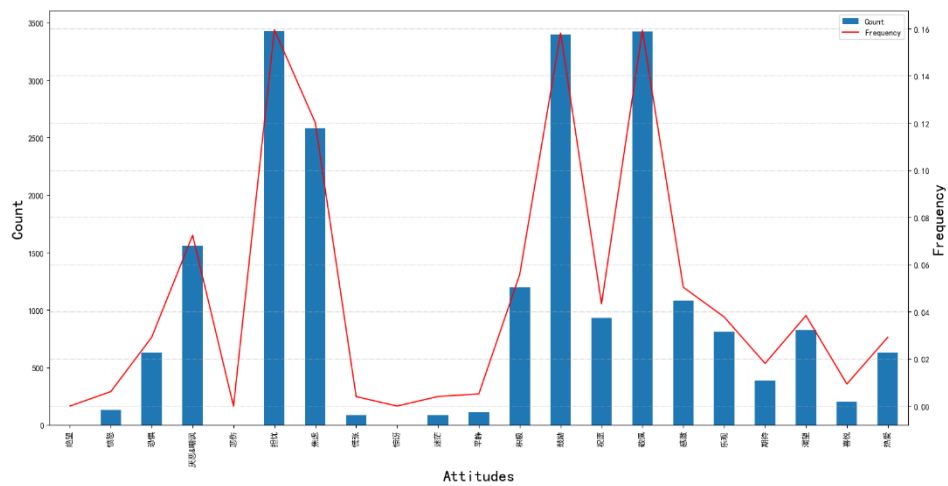
(二阶段)



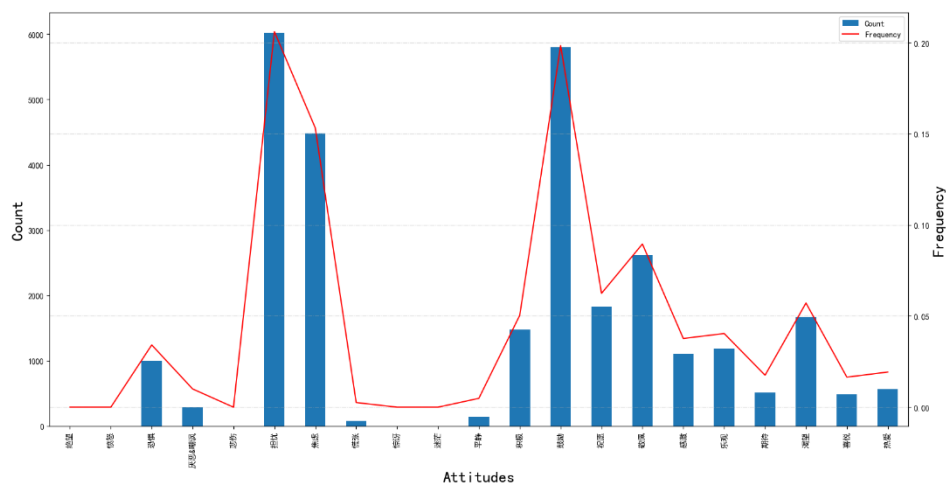
(三阶段)



(四阶段)



(五阶段)



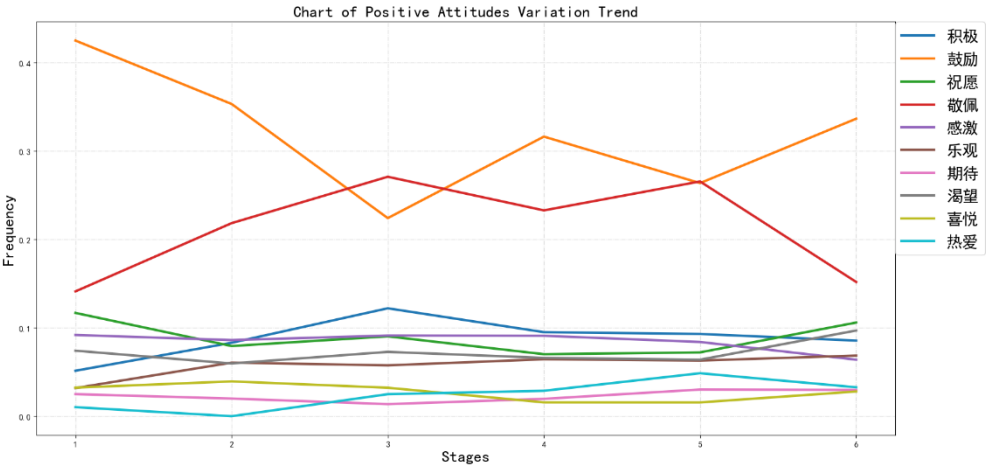
(六阶段)

通过对六个阶段心态频数及频率统计图的观察分析,我们发现:每个阶段中,中国大众网络社会心态分布情况大致相同,即每个阶段的“心态图形”都近似呈现一种“驼峰”形状,而积极心态(心态值 >0.5)和消极心态(心态值 <0.5)中都有某1~2种心态在图中为“驼峰”,即为特定阶段中中国大众表现出的主流网络社会心态。初步观察来看,在每个阶段的心态图中,积极心态的“鼓励”和“敬佩”占比较高,而消极心态的“担忧”和“焦虑”占比较高,总体心态分布中,过于乐观(心态值 ≥ 0.8)和过于悲观(心态值 ≤ 0.2)的心态占比很少,心态值接近0.50的心态占比较高。且由此我们初步假设:

(1) 疫情背景下,中国大众网络社会心态主要表现为“鼓励”、“敬佩”、“担忧”和“焦虑”;

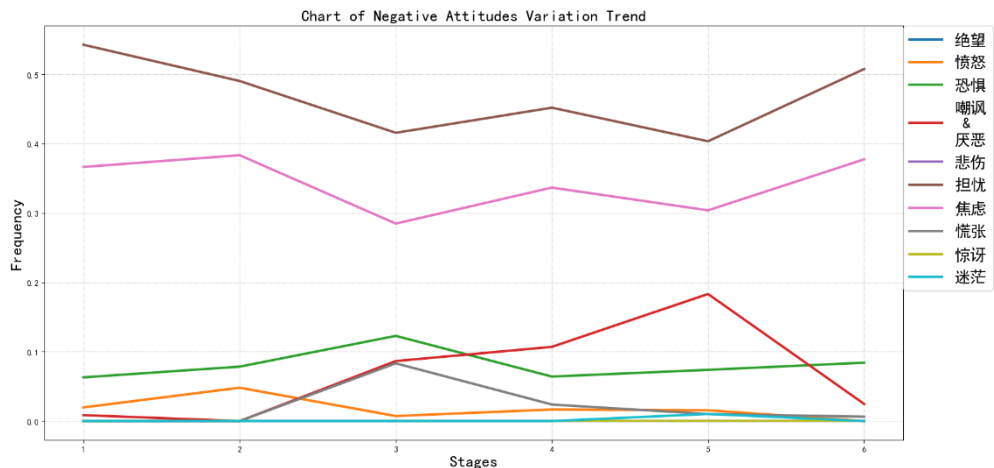
(2) 其余心态则随着各阶段疫情防控形势的不同而产生变化,但总不会超过上述四种心态的占比。

下面我们具体分析在整个六个阶段中每种心态占比的变化情况,我们绘制并观察各个阶段所有积极心态(心态值 >0.5)的占比及变化:



(六个阶段积极心态变化折线图, 纵轴为 Frequency)

我们绘制并观察各个阶段所有消极心态(心态值 <0.5)的占比及变化。



（六个阶段消极心态变化折线图，纵轴为 Frequency）

绘制好六个阶段中每个积极/消极心态的频率折线图，我们就能够非常轻松地看出各阶段中国大众主流网络社会心态及各心态由阶段的相应变化。

由“六个阶段积极心态变化折线图”和“六个阶段消极心态变化折线图”，我们验证了先前的假设（1），即在心态分布图的右侧“驼峰”中，“敬佩”和“鼓励”两种心态占比最高，在左侧“驼峰”中，“焦虑”和“担忧”两种心态占比最高。且其占比远远领先于其他心态。

下面我们单独分析具体的积极心态和消极心态的变化趋势，并简要分析其中原因。

分析具体心态时，我们着眼于：

（1）“鼓励”、“担忧”、“焦虑”、“恐惧”

“鼓励”、“担忧”和“焦虑”这三种心态总体变化趋势均为：1~5 阶段呈下降趋势，6 阶段回升。

“恐惧”的总体变化趋势是：1~3 阶段逐渐上升后下降并趋于稳定，在除“担忧”、“焦虑”的其他消极心态中占有较高的比重。

我们知道，在第 1 阶段初，人们尚未意识到新冠肺炎的严重性，对其有一种漠视的态度，以致后来疫情的爆发和逐渐严重，在这一阶段，大众从漠视到逐渐经历新冠肺炎爆发的恐怖后，“担忧”和“焦虑”自然成为负面心态中占比较高的部分，大众第一次面对新冠病毒，心中固然有许多困惑，乃至对未知病毒的恐惧：病毒会不会致命？病毒会不会持续传播？存不存在人传人的可能？……而在许多人担忧、焦虑甚至恐惧之时，“鼓励”便是必不可少的，它既是一种乐观积

极的心态，又是中国人在病魔肆虐下坚强渡过难关的一剂良药；而随着针对新冠病毒的医疗手段和相关政策的不断完善，人们开始逐渐习惯居家生活、积极配合国家防疫工作，因此在 2~5 阶段时，我们欣然地看到“担忧”、“焦虑”这两种心态占比逐渐下降，相应地“鼓励”的心态占比也随之下降了。

就在全中国疫情防控看似稳定之时，随着新年的到来和气温的下降，中国北部出现了多个聚集性和冷链传播等新冠事件，如河北藁城、黑龙江望奎、惠七、吉林通化超级传染源等，这些时间大都发生在 2020 年 12 月中旬到 2021 年 1 月至今，且目前新冠确诊病例仍在每天增长。面对新冠病毒的第二次肆虐，大众开始担心能否过年回家、病毒是否会变异、高校学生能否正常放假等关键问题，因此“担忧”和“焦虑”的占比明显回升，而网友们“鼓励”的声音也相应地增多了。

（2）“敬佩”、“感激”、“担忧”

“敬佩”的总体变化趋势是：1~5 阶段持续上升后保持平缓，在第 6 阶段有所下降。

“感激”的总体变化趋势是：1~6 阶段基本稳定，在除“鼓励”、“敬佩”的其他积极心态中占有较高的比重。

在（1）中分析的“担忧”心态，除了大众对新冠病毒本身的担忧，更有对“逆行天使”（医护人员）们生命安全的担忧。疫情期间，广大“勇士”纷纷请战疫情防控一线，他们穿着密不透气的防护服，克服着生理和心理上的种种不适，在抗疫一线夙夜匪懈地与新冠肺炎抗争，他们是疫情期间“最美的天使”。我们作为宅家的普通大众，从各种媒体了解到他们的抗疫故事和抗疫精神，有时亦会心痛地看到某位“逆行天使”牺牲的悲剧。我们对他们的逆行精神无比敬佩和对他们心生感激的同时，对他们的安全也是十分担忧。

而在第 2 阶段，正是疫情抗击的白热化阶段，这时大众对医护人员的敬佩心态占比持续上升；在第 3 阶段，疫情抗击逐渐见到曙光、夺取初步胜利，医护人员纷纷从抗疫一线凯旋，各地欢迎英雄回家，我们由衷歌颂逆行者的伟大精神，并由衷敬佩他们，因此“敬佩”心态占比持续增加。整个过程中“感激”心态占比基本稳定。

（3）“嘲讽” & “厌恶”、“愤怒”

“嘲讽” & “厌恶”是变化趋势较为显著的一种心态，其总体变化趋势为：

1、2 阶段占比很低，第 3~5 阶段其占比以较快的速度上升，且超过“恐惧”心态位于所有消极心态占比的第三位。

“愤怒”的总体变化趋势为：第 1 阶段到第 2 阶段增长幅度较大，到第 3 阶段又下降，而到了第 4、5 阶段又回升。

我们注意到，第二阶段的排名很靠前的一个关键词便是“红十字会”（指湖北省红十字会），当时因湖北省红十字会对社会各界捐助的抗议资源分配出现了严重问题，致使当地医院抗疫物资严重短缺，进而关联到许多人的生命安全问题。为此网友们义愤填膺，在微博上声讨“红十字会”，可以集中反映出大众的“愤怒”心态，因此第 2 阶段“愤怒”心态占比升高，第 3 阶段时红十字会时间已基本得到完善解决，故而占比下降。

到了第 4、5 阶段，“愤怒”和“嘲讽&厌恶”心态占比同时上升，究其原因，我们回顾相应时间点的大事件：美国新冠疫情集中爆发于 2020 年 3 月 15 日，新增确诊拐点日为 4 月 5 日，美国新冠疫情确诊病例自此持续以较快速度增加，与此同时，除中国外世界范围内疫情逐渐严重，不少国家的网民开始怨怼于中国，声称新冠病毒是“中国病毒”、“武汉病毒”，身为美国总统的特朗普也在社交平台上公然宣称新冠病毒是“中国病毒”，这让中国人十分愤怒和厌恶。同时，随着美国疫情的逐渐严重，特朗普的“群体免疫”策略失效，其本人出尔反尔、言行不一，遭到了中国网友们的嘲讽。

除以上分析的具体心态外，我们注意到，类似“祝愿”、“乐观”、“期待”、“渴望”等积极心态在总体阶段中均保持着较稳定的占比，且在积极心态中占比较高；而类似“悲伤”、“慌张”、“惊讶”、“迷茫”等消极心态虽然占比较为稳定，但是占比较低，个别消极心态如“绝望”等占比基本为 0。纵观心态阶段变化趋势图，我们有理由认为：

（1）疫情背景下，中国大众网络社会心态主要表现为“鼓励”、“敬佩”、“担忧”和“焦虑”；

（2）其余心态则随着各阶段疫情防控形势的不同而产生变化，但总不会超过上述四种心态的占比。

这两条假设成立。此外，我们也欣慰地看到：即使身处疫情防控的紧张情形下，我们中国大众仍然能全力配合国家防控工作，保持较高的积极心态，虽有消

极心态，但人们总会互相鼓励，会被疫情防控工作中的中国力量震撼，为逆行精神感动，为每一个参与疫情防控工作的逆行天使、志愿服务者等由衷敬佩和感激。