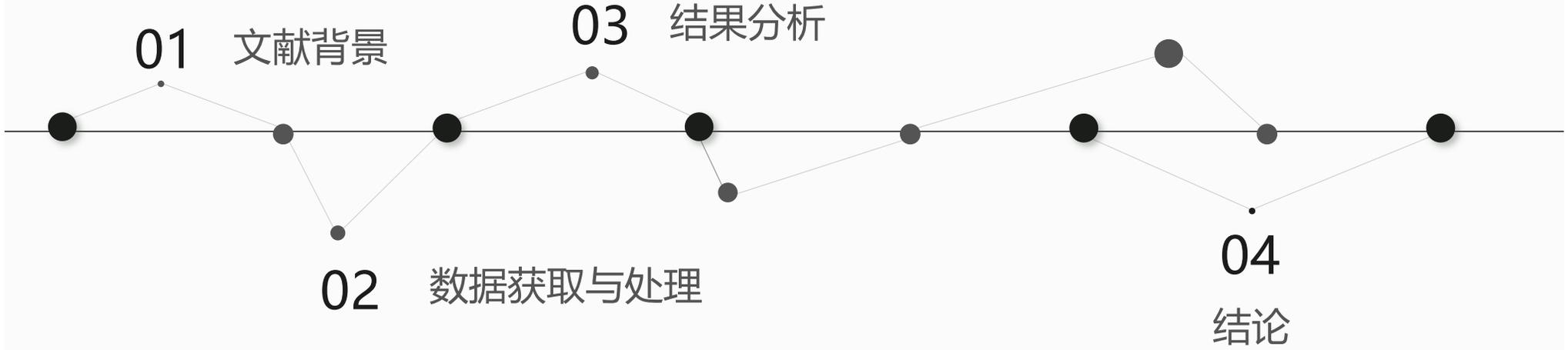




微博外交： 美驻华使馆新媒体传播 策略及效用分析

小组成员：陈震东、铁嘉欣、徐静 *拼音排序不分先后*

CONTENT



An abstract geometric design featuring several interconnected lines and dots of varying sizes, creating a network-like structure. The design is composed of thin black lines and solid black dots of different diameters, arranged in a way that suggests movement and connectivity. The overall aesthetic is clean and modern, with a focus on geometric forms and spatial relationships.

1

Part 01 背景

詹姆斯·格拉斯曼提出，基于社交网络技术的“公共外交2.0”是新媒介语境下美国公共外交应对挑战的最佳方法。

01

微博外交:

- 基于“推特”、“脸书”等微博平台的外交活动或行为
- 即时性、交互性

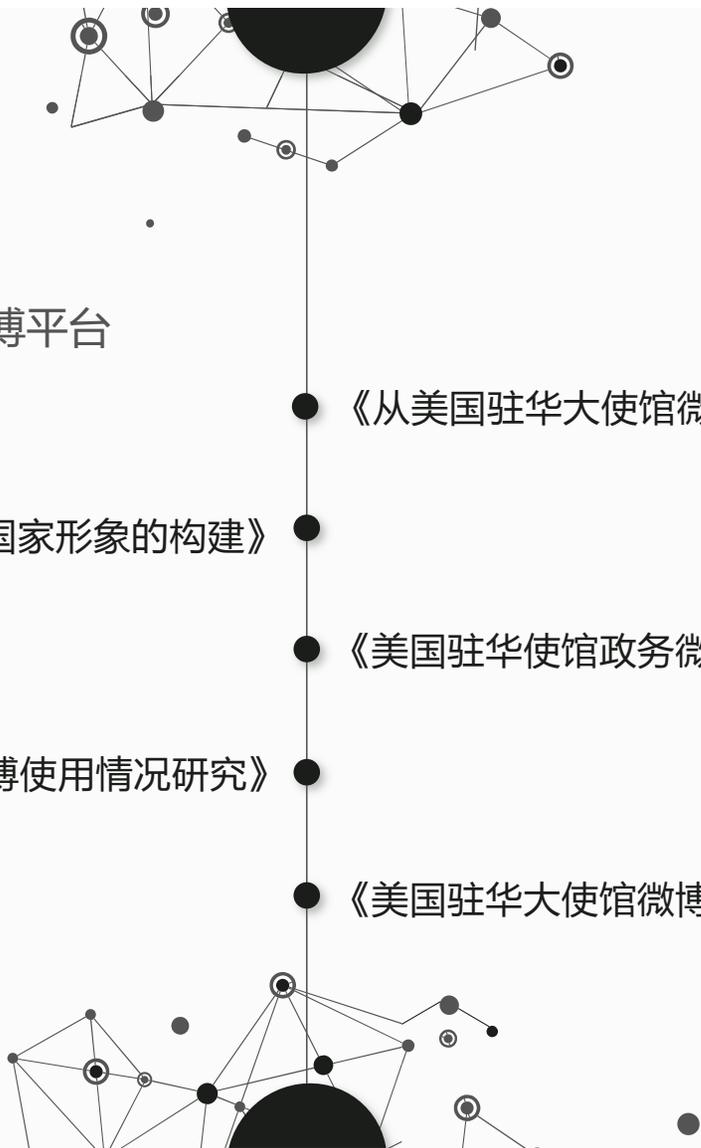
《社交媒体中国国家形象的构建》

《美国驻华使馆微博使用情况研究》

《从美国驻华大使馆微博运营看其对华公共外交》

《美国驻华使馆政务微博话语探析》

《美国驻华大使馆微博议题设置与策略》





2

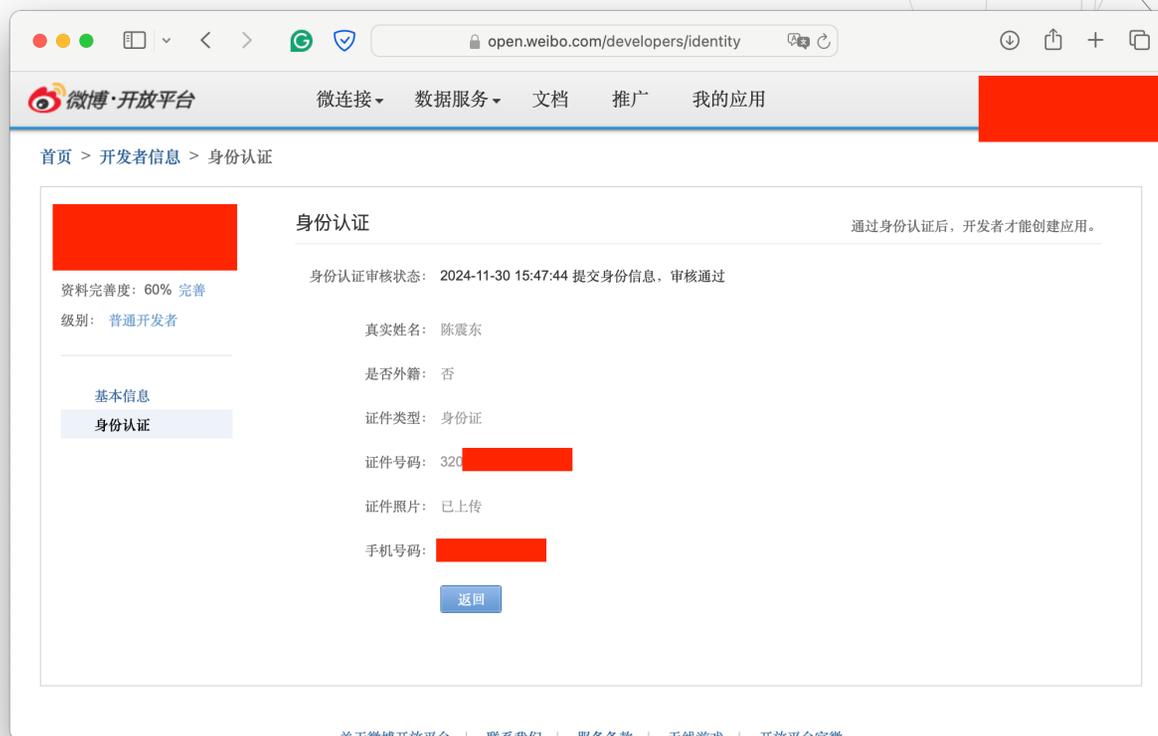
Part 02 数据

选取“美国驻华大使馆”新浪微博官方账号作为研究对象，因为该微博处于美国官方在华微博群中心地位，建立时间较早，更具有全国性、官方性和代表性。

2.1 / 数据来源1 - 官方 API

- 数据来源正规可靠
- 需要实名认证
- 审核缓慢

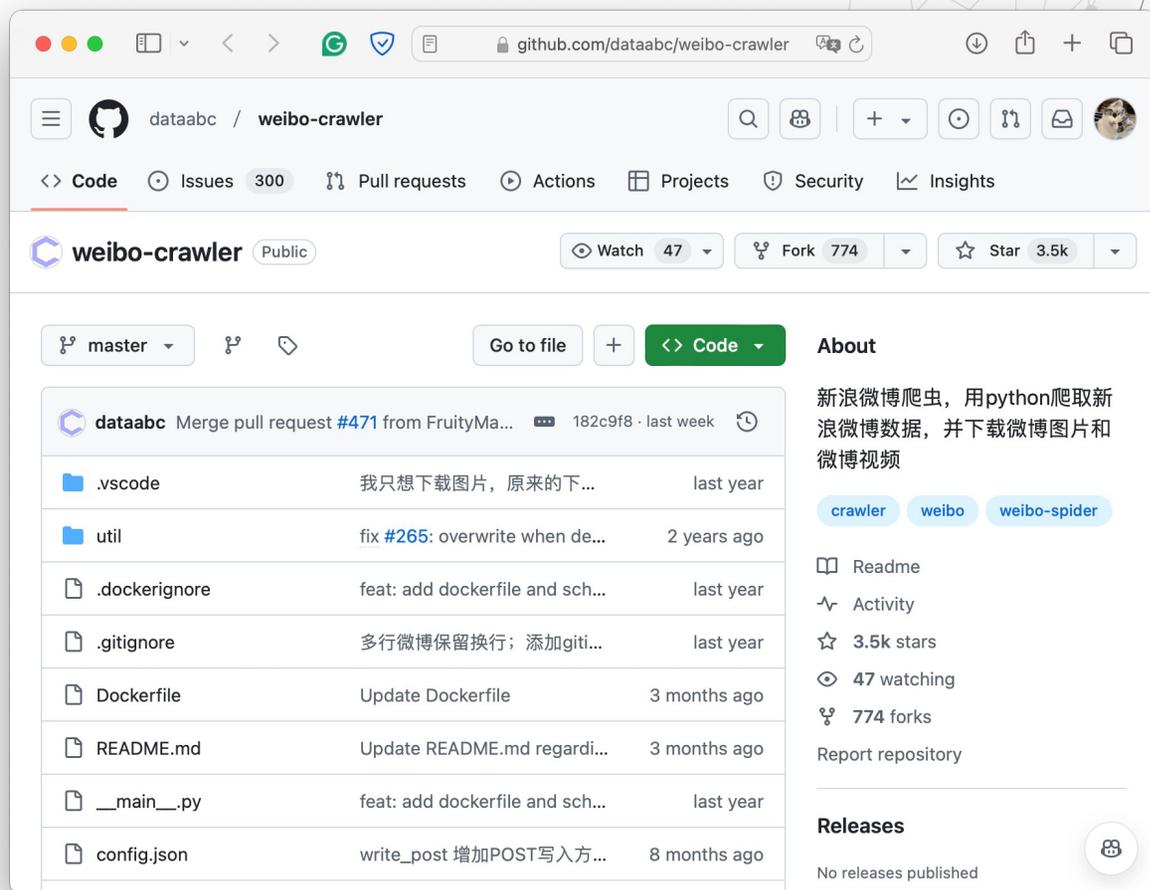
未赶上使用



2.1 数据来源2 – GitHub 爬虫项目

- 项目均无接口文档
- 难以使用 Python 调用
- 可能存在隐形问题
- Debug 困难

使用不便



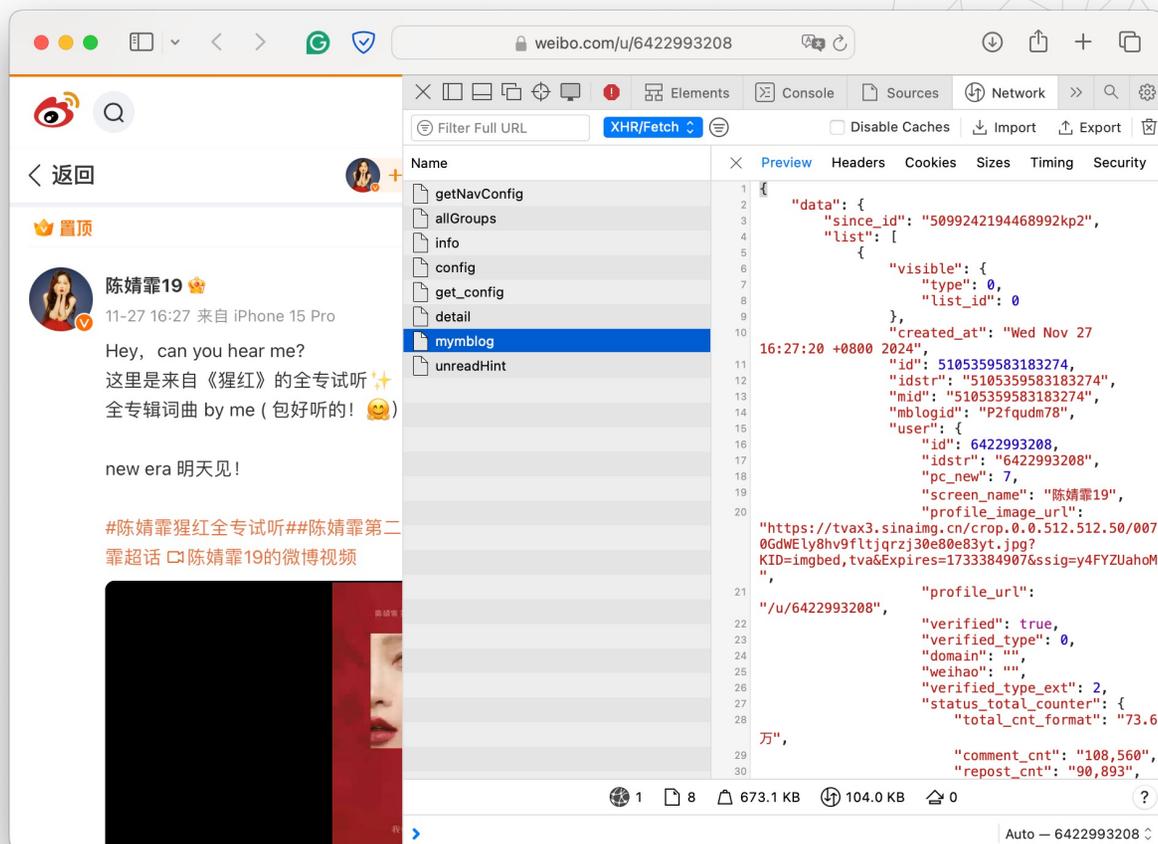
2.1 数据来源3 – 分析接口自行请求

- JSON 接口未加密
- 方便程序化调用
- Debug 简单
- 仅用于学术研究

最终方案

相关代码存储于

- post_crawler.py
- comment_crawler.py



2.2 数据爬取

- 提供 Cookie 并向接口请求
- 循环爬取最近600条博文
- 「长文本」单独请求
- 爬取每一条推文对应的评论
- 存储为 csv 格式

相关文件存储于

- data/posts.csv
- data/comments.csv

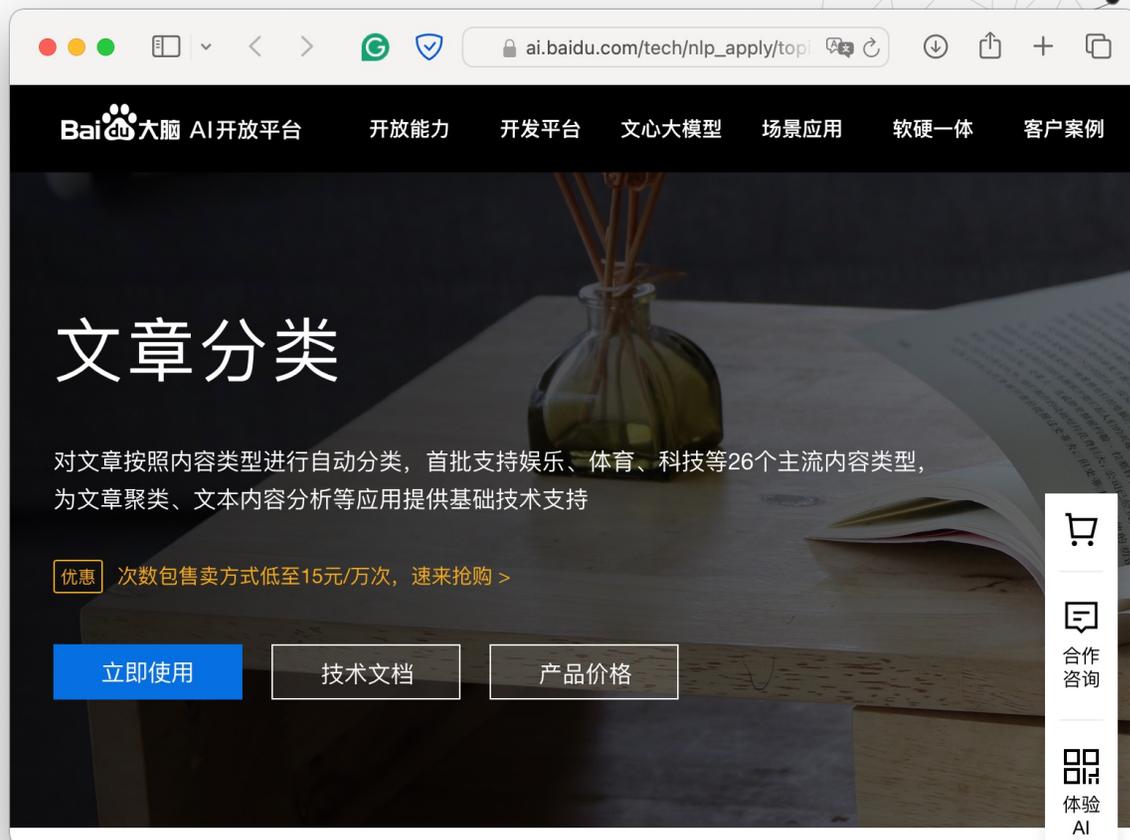


2.3 推文清理与主题分类

- 使用百度 NLP 文章分类接口
- 筛除过短博文
- 应接口需要，提取标题
- 根据接口解析博文主题
- 整理并存储为 csv 格式

相关文件存储于

- processed/posts.csv



2.3

评论清理与情感分类

- 使用百度 NLP 情感接口
- 使用正则表达式将评论的 HTML 代码解析为纯文本
- 调用接口解析评论情感
- 整理并存储为 csv 格式

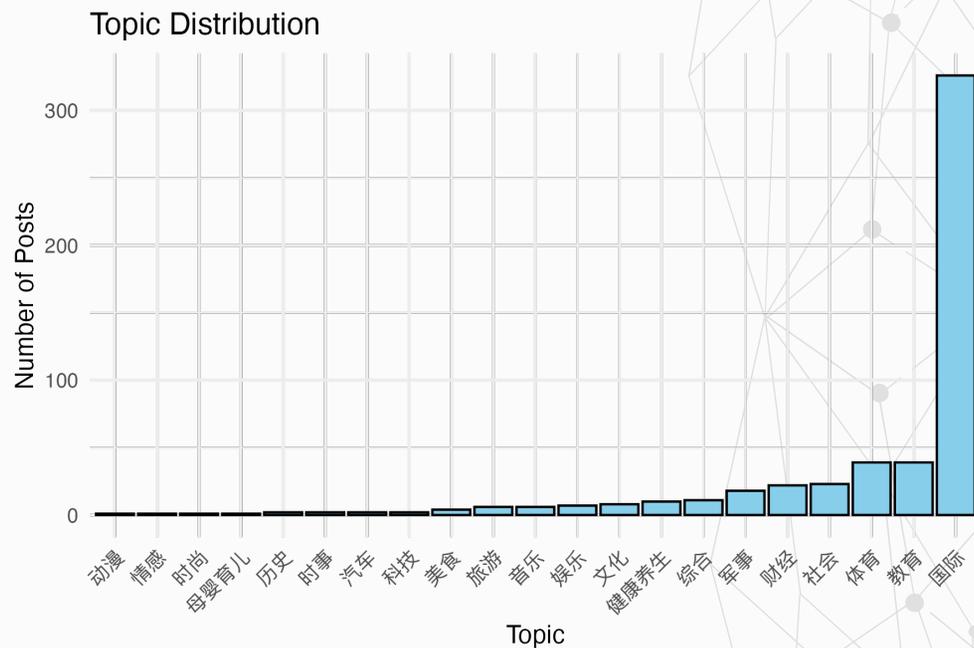
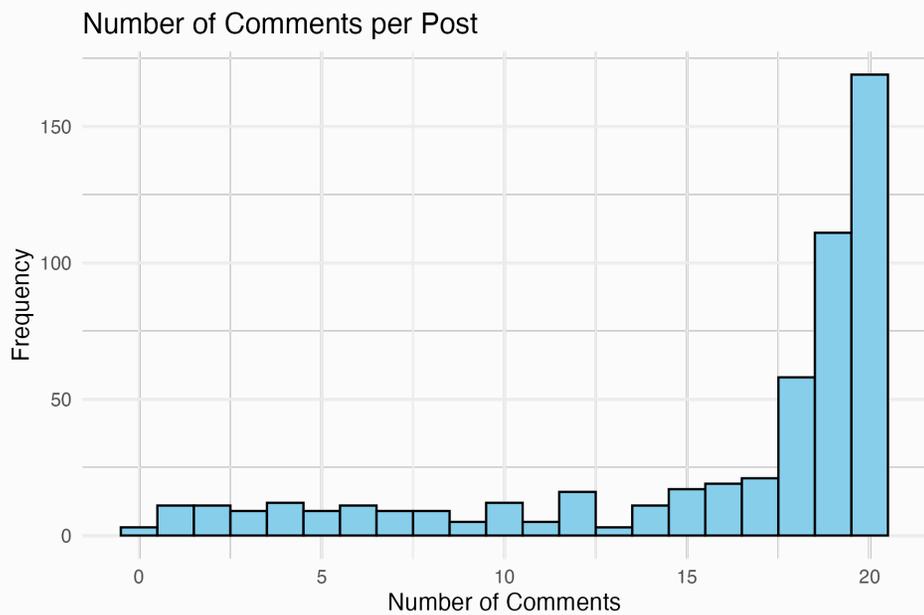
相关文件存储于

- processed/comments.csv



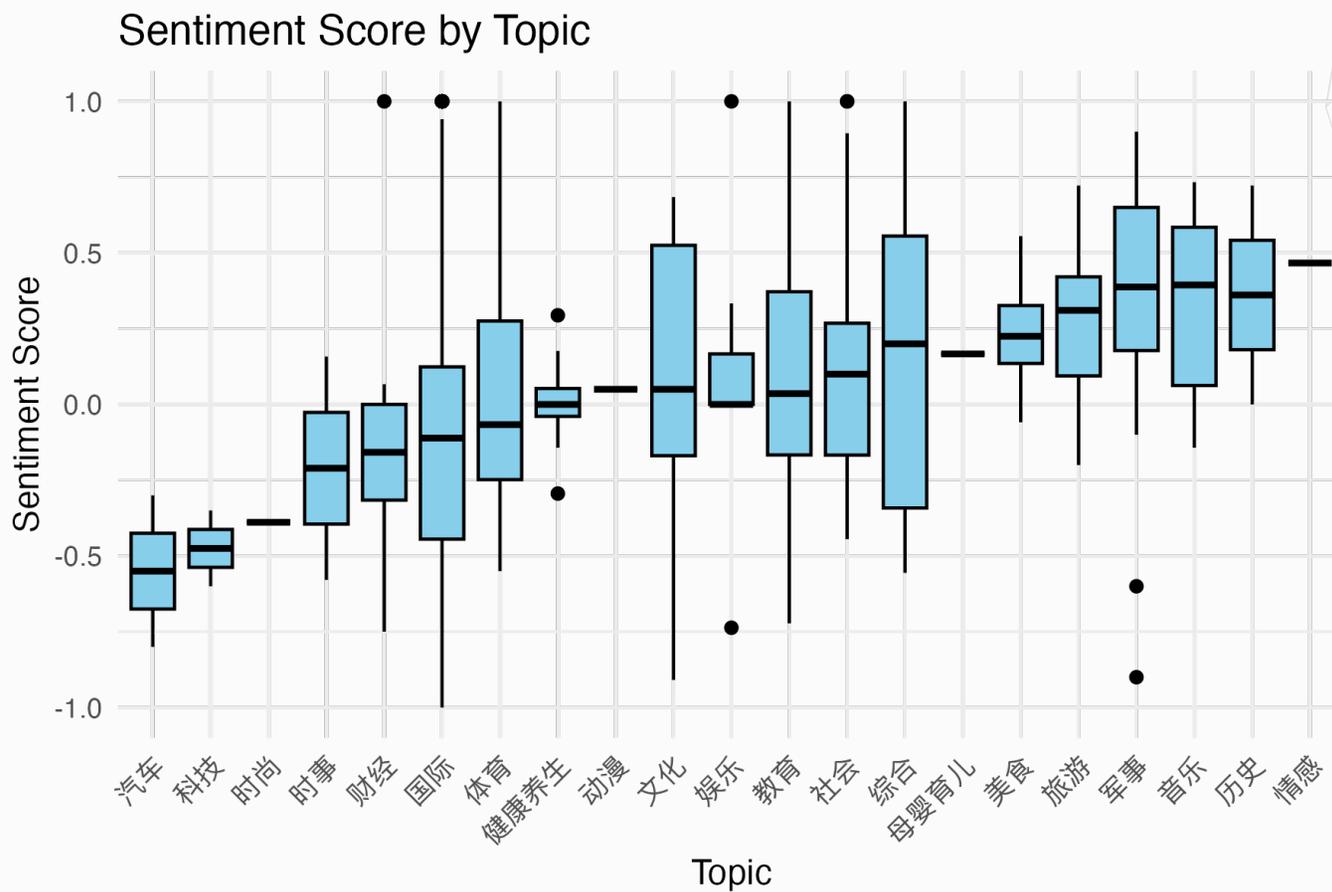
2.4

数据一览 - 描述性统计



2.4

数据一览 - 描述性统计





3

Part 3 分析

各国外交机构越来越多地运用微博平台发布外交信息、引发公众关注、收集公众信息并试图影响公众舆论。

03



- 数据分析1
聚类分析推文主题与主题下的关键词
- 数据分析2
各簇平均情感得分统计
- 数据分析3
计算推文主题强度与情感得分回归系数

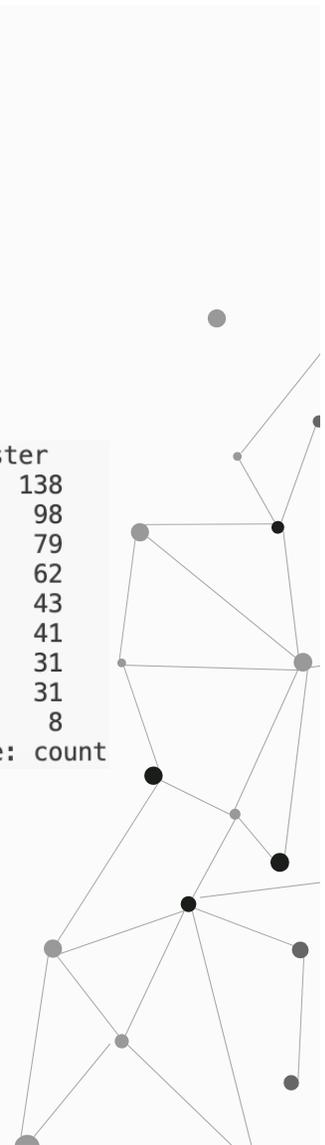
03

- 对推文内容进行K-means聚类分析
- 测试最佳聚类数量为九个簇



cluster	count
8	138
5	98
6	79
7	62
1	43
0	41
3	31
2	31
4	8

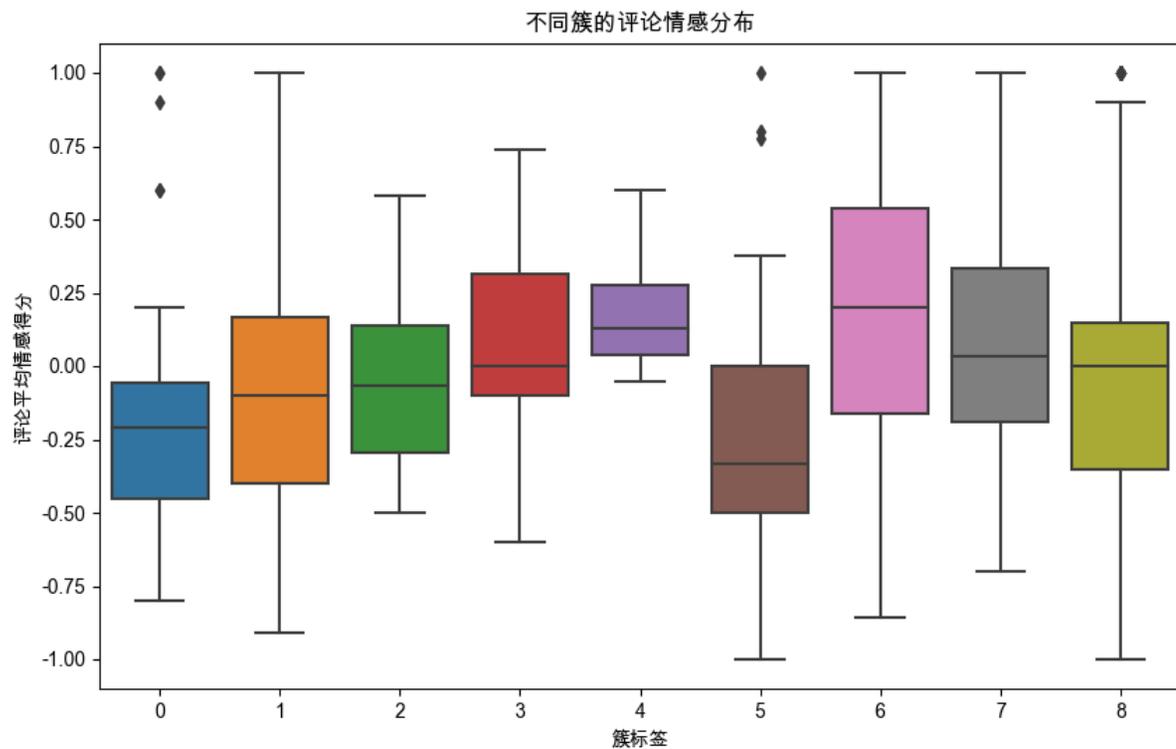
Name: count



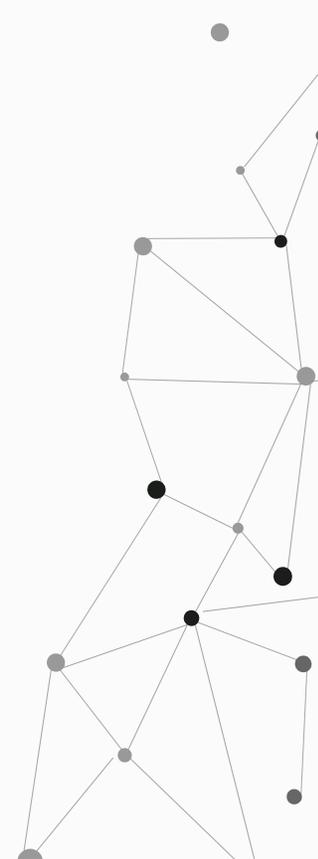
- 簇 0: 国际合作 (东盟、中国等)
- 簇 1: 拜登总统与拉美事务
- 簇 2: 奥运会与残奥会
- 簇 3: 乌克兰战争与国际援助。
- 簇 4: 大熊猫外交与动物园
- 簇 5: 以色列及中东安全问题
- 簇 6: 中美关系与文化交流
- 簇 7: 美国选举与英文教育
- 簇 8: 全球话题

03

➤ 对每个簇求相关推文的平均情感得分



簇 0: 国际合作 (东盟、中国等) 簇 1: 拜登总统与拉美事务 簇 2: 奥运会与残奥会
簇 3: 乌克兰战争与国际援助。 簇 4: 大熊猫外交与动物园 簇 5: 以色列及中东安全问题
簇 6: 中美关系与文化交流 簇 7: 美国选举与英文教育 簇 8: 全球话题



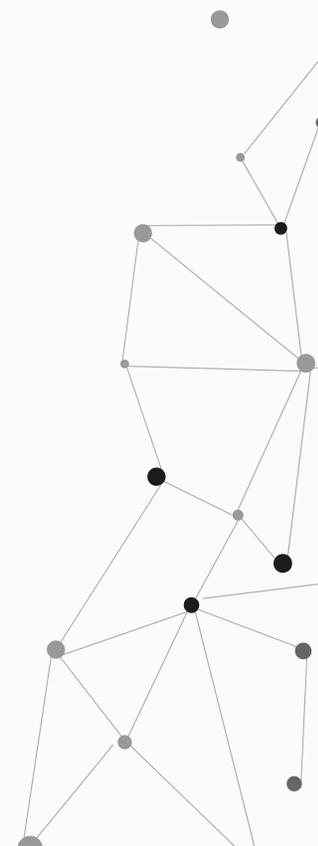
03

➤ 对每个簇求相关推文的平均情感得分

各簇情感统计：

	cluster	mean	std	count
0	0	-0.157076	0.445809	41
1	1	-0.069904	0.397011	43
2	2	-0.039811	0.309065	31
3	3	0.046216	0.336137	31
4	4	0.201335	0.251350	8
5	5	-0.236386	0.391704	98
6	6	0.186379	0.473636	79
7	7	0.105671	0.388424	62
8	8	-0.059642	0.449971	138

簇 0：国际合作（东盟、中国等） 簇 1：拜登总统与拉美事务 簇 2：奥运会与残奥会
簇 3：乌克兰战争与国际援助。 簇 4：大熊猫外交与动物园 簇 5：以色列及中东安全问题
簇 6：中美关系与文化交流 簇 7：美国选举与英文教育 簇 8：全球话题



03

- 为每条推文计算与每个簇关键词的匹配次数，作为主题强度
- 对主题强度与平均情感进行回归

R² 得分：0.0761207870960493

均方误差 (MSE)：0.2229506443499436

簇 0 的主题强度回归系数：-0.017199228223398148

簇 1 的主题强度回归系数：-0.013872443637726446

簇 2 的主题强度回归系数：-0.026458877537918903

簇 3 的主题强度回归系数：0.01761307049868006

簇 4 的主题强度回归系数：-0.004504054118157659

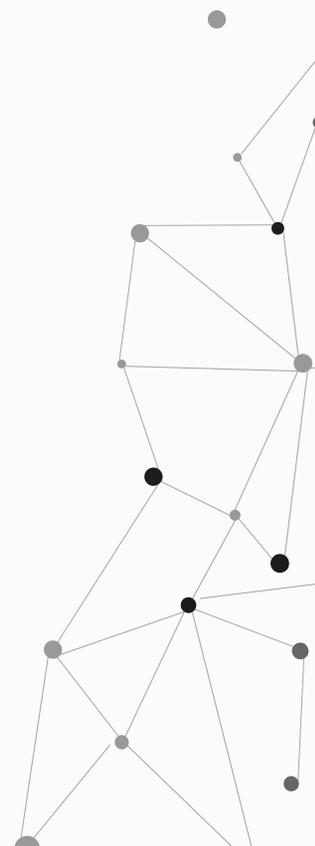
簇 5 的主题强度回归系数：-0.04668181421157623

簇 6 的主题强度回归系数：0.04081614791811359

簇 7 的主题强度回归系数：0.011087567116380167

簇 8 的主题强度回归系数：-0.003624020666113934

簇 0：国际合作（东盟、中国等） 簇 1：拜登总统与拉美事务 簇 2：奥运会与残奥会
簇 3：乌克兰战争与国际援助。 簇 4：大熊猫外交与动物园 簇 5：以色列及中东安全问题
簇 6：中美关系与文化交流 簇 7：美国选举与英文教育 簇 8：全球话题





4

Part 04 结论

通过美国驻华大使馆微博博文主题与评论情感的关系，
分析美驻华使馆新媒体传播策略及效应。

04



结论1

美国大使馆的推文发布主要分为九个主题，最近全球话题、以色列问题以及中美关系为最多的主题



结论3

推文主题强度与评论情感回归 R^2 不高可能是因为数据量偏少



结论2

与以色列问题、国际政治合作话题相关推文的评论情感最为负面，而与大熊猫外交、中美文化交流相关话题情感最积极



结论4

情感分析部分还存在一定问题，特别是难以识别情绪的强度，以及区分讽刺性内容





Thank You!