

程序设计与数据可视化

课程大作业：
疫情下的谣言传播可视分析系统

作业背景和目标

- 背景：自2019年起，全球范围内爆发了一场新冠肺炎疫情。疫情的传播除了带来身体上的伤害外，还引发了大量的谣言传播，给疫情防控工作带来了巨大的挑战。本项目旨在通过分析2021年11月7日至2022年2月18日，今日头条收集的366条疫情谣言数据，探究疫情下的谣言传播现象。
- 目标：本项目旨在设计一个**谣言可视分析系统**，以实现疫情期间社交网络中谣言的可视分析。具体分析内容包括但不限于以下几个方面
 - 谣言主要分布内容和类别 **文本主题分析**
 - 谣言类别随事件发生的动态变化
 - 谣言的区域分布
 - 热点谣言以及媒体在其中的作用
 - ...

数据

- `rumor_data.csv`

- 本项目涉及到的谣言数据来自今日头条，收集了2021年11月7日至2022年2月18日，共计366条谣言，包含了谣言发布时间、谣言来源、谣言主要内容、谣言涉及省份、用户浏览谣言记录的相关信息。

- 数据文件列信息如下：

- ✓ `date`: 谣言发布的日期，用str记录
- ✓ `source`: 谣言的来源，用str记录
- ✓ `content`: 谣言的文本内容，用str记录
- ✓ `province`: 谣言涉及的境内省份，用str记录，值为空表示该谣言不涉及地理信息
- ✓ `user_0至user_17`: 用户浏览谣言的记录，用int记录，1表示对应用户浏览了相应的谣言，否则为0 一个用户浏览了两个谣言，这两个谣言之间可以连一条边；一个谣言被两个用户浏览，这两个用户也可以连一条边
- ✓ `like`: 点赞数

潜在任务

- 根据**文本**数据信息进行谣言主题挖掘，并将结果可视化呈现；
- 根据**地理**数据信息进行谣言区域分布挖掘，并将结果以地图形式可视化呈现；
- 结合**时间**维度信息，分析从文本和地理数据中挖掘到的内容随时间发生的动态变化，并将结果以图表形式可视化呈现；
- 根据用户浏览记录构建**网络**，并基于网络数据进行谣言主题分析与挖掘、用户偏好分析；
- 综合考量**以上各方面因素和外部因素**（如媒体、^{点赞数}谣言热度等），得出总结性结论，并将结果以图表形式可视化呈现；
- ...

提交方式和时间

- 提交文件包括两部分：
 1. **相关代码**：ipynb或py文件，请写明注释
 2. **中文报告**：pdf文件（20页以内），包含详细思路、实现过程、可视化结果、分析过程和最终结论
- 生成压缩文件后以“**name_sid_project.zip**”命名，提交至E-learning系统
- **截止日期：2023年6月23日**

评价标准

- 我们将从以下三个方面进行评分：
 - 相关代码（30%）
 - 可视化结果（30%）
 - 中文报告（40%）
- 禁止抄袭！



**对课程大作业有任何疑问，请邮件或微信联系
任课老师或助教！**

Contact: jiarongxu@fudan.edu.cn

Homepage: <https://galina0217.github.io>

