



疫情期间舆情演化分析与热点事件模式挖掘

(ChinaVis 2020 数据可视分析挑战赛一等奖)

李泽宇 陈诺 王腾 李思思 贾世超 张加万

天津大学智能与计算学部可视计算实验室

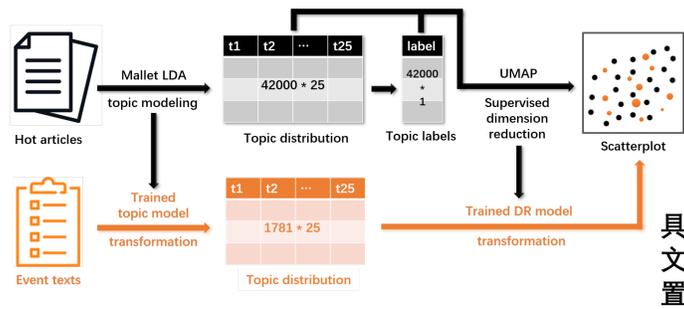
分析任务及数据

分析任务：明确疫情发展阶段；探究舆论关注点的转变；找出在情感演化上特殊的热点事件；挖掘热点事件的热度演化模式。

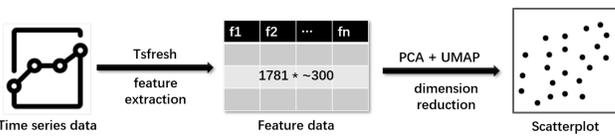
数据：时间跨度为2019年12月23日~2020年5月10日，具体描述如下表。

	热文数据	热点事件数据	疫情大事件列表
数量	300 * 140天 = 42000	1781	690
来源	榜单：清博大数据 文章：今日头条	清博大数据	知微数据
核心字段	标题、内容 发布日期 阅读数	标题、概述、热词 发生日期 14天的情感演化序列 14天的热度演化序列	标题 发生日期
类型	时序文本	时序文本，时间序列	时序文本

数据模型

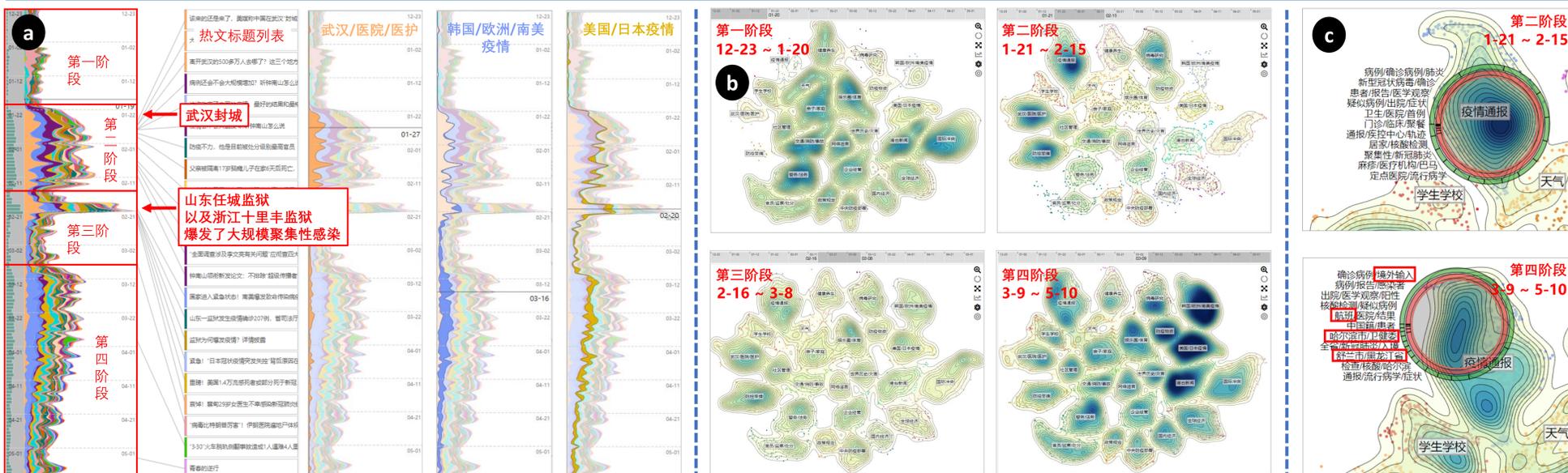


左图描述了热文和热点事件统一投影空间（主题地图）的构造过程。在主题地图中具有相似主题的热文和热点事件被放置在相近位置。



左图描述了事件演化范式投影空间的构造过程。具有相似热度演化范式的事件被放置在相近位置。

疫情发展阶段与舆论关注点在主题级别的演化



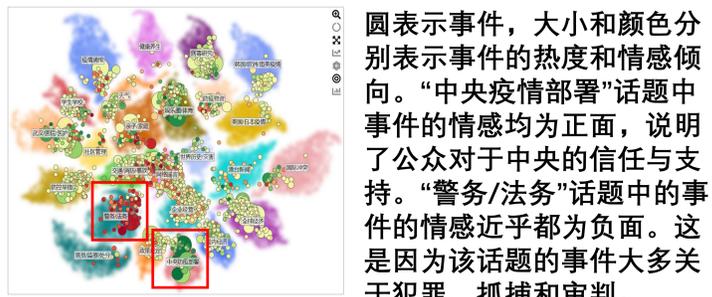
②中的堆叠面积图描述了各个主题的热文在阅读总数上的分布及其演化，我们发现在12月23日之后的140天里，疫情可分为四个发展阶段。在③中的主题地图中，具有相似主题的热文被放置在一起（参见数据模型）。配合等高线和时间刷，分析者能够直观地感知话题级别的舆论关注点的转变。④中的聚光灯组件可任意拖动，帮助分析者查看任意区域的语义。结合②中的热文列表列出的代表文章，我们解释这四个阶段如下：第一阶段（12月23日至1月20日），国内疫情舆论“密切关注期”。此时疫情尚未引起公众广泛的关注。第二阶段（1月21日至2月15日），国内疫情舆论“爆发期”。1月23日，“武汉封城”的新闻引起了全国范围对疫情的高度重视。第三阶段（2月16日至3月8日），国内疫情舆论“平缓期”。但在2月19、20日出现了一个爆发式增长，这是因为山东任城监狱以及浙江十里丰监狱发生了规模较大的聚集性感染。第四阶段（3月9日至5月10日），国际疫情舆论“爆发期”。此阶段，疫情在全球爆发。

舆论关注点在词汇级别的演化



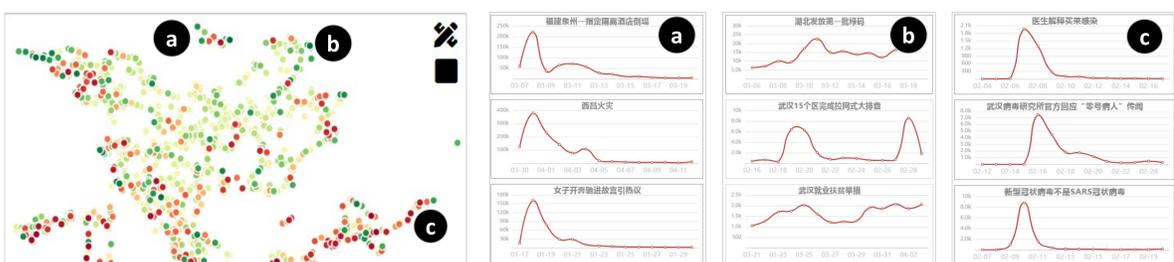
条带宽度表示主题的热度，词的大小正比于其在所处时段的显著性。“韩国/欧洲/南美疫情”话题和“美国/日本疫情”话题中热词的转变与疫情在国际的传播过程高度吻合：始于日韩，后转至欧洲，又蔓延到美国和南美。

热点事件的话题分布和情感倾向



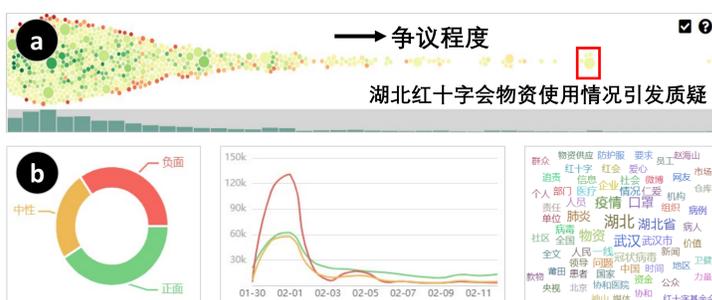
圆表示事件，大小和颜色分别表示事件的热度和情感倾向。“中央疫情部署”话题中事件的情感均为正面，说明了公众对于中央的信任与支持。“警务/法务”话题中的事件的情感近乎都为负面。这是因为该话题的事件大多关于犯罪、抓捕和审判。

热点事件的热度演化模式



左图中具有相似热度演化模式的事件被放置在相近位置（参见数据模型下图）。比如，a区域中的事件其热度会在中段有一个二次小波峰，b区域中的事件其热度具有较强的波动。c区域中的事件其热度波峰位置在中段。

在情感演化上特殊的热点事件



“湖北红十字会物资使用情况引发质疑”这一事件存在较大争议(ⓐ)。该事件发生14天内的正负面情感占比类似，情感从负面转向正面(ⓑ)。这是因为2月4日湖北省纪委监委对红会领导予以了严肃问责，并在2月8日起，红会每天都会公布捐赠物资使用情况，这两件事收获了公众的赞扬。

联系人：李泽宇
lzytianda@tju.edu.cn