

# 流行病数据可视分析综述

孙国道 杨雨璠 潘翔 梁荣华

(浙江工业大学计算机科学与技术学院 杭州 310000)

**摘要** 流行病数据的多层面可视分析,可以加快流行病数据分析任务的交互式探索效率和加深对潜在模式的深刻理解.本文对流行病数据可视分析的相关工作展开综述,并主要通过四个方面进行总结和归纳:(1)流行病数据的时空可视分析,帮助用户发现和理解流行病数据在时间、空间以及时空维度中潜在的流行病特征和传播规律等;(2)流行病数据中涉及非结构化/半结构化文本的语义可视分析,辅助用户快速了解长、短文本内容中的流行病内容态势、情感走向等;(3)流行病传播模型的可视分析,增强用户对流行病传播过程中的预测、仿真、监测等任务的交互式理解;(4)流行病数据的静态信息图,直观反映流行病信息的上下文内容和提高公众的理解效率.在上述基础上,本文阐述了流行病数据可视分析在数据存储和获取,地图可视化形式多元扩展,流行病模拟传播分析和多学科交融的综合可视分析存在的问题和挑战.

**关键词** 流行病数据可视化;时空可视化;语义文本;流行病模型;信息图;可视分析

中图法分类号 TP18 DOI号 10.11897/SP.J.1016.2022.00601

## A Survey on Visual Analysis of Epidemic Data

SUN Guo-Dao YANG Yu-Fan PAN Xiang LIANG Rong-Hua

(College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310000)

**Abstract** The multi-facet visual analysis of epidemic data can speed up the deep understanding of epidemic related tasks and interactive exploration mode. In this paper, the related work of visual analysis on epidemic data is summarized and summarized in four aspects: (1) Spatio-temporal visual analysis of epidemic data can help users find and understand the potential epidemic characteristics and propagation rules of epidemic data in time, space and space-time dimensions. At the same time, it classifies the epidemic data based on three visual forms and classifies each type of model the epidemic visual analysis of the block is described. Among them, the visual analysis based on time series is mainly based on statistical charts, combined with related fields to carry out visual analysis on the development of epidemic and public opinion. The visual analysis based on spatial dimension is mainly based on the visual form of map, which shows the spread of epidemic in two-dimensional map or three-dimensional space through geographic location information. Visual analysis based on spatio-temporal data mainly combines temporal and spatial attributes, and its visual forms are diversified. (2) Semantic visual analysis of unstructured/semi-structured text in epidemiological data, assisting users to quickly understand the epidemic content situation and sentiment in long and short text content Trends. Based on the visual analysis of semantic structure, the text content of epidemic text is structured by combining structure diagram and tree diagram. Based on temporal

收稿日期:2020-07-28;在线发布日期:2021-07-13.本课题得到国家重点研发计划(2020YFB1707700)、国家自然科学基金项目(61972356, 62036009)和浙江省属高校基本科研业务费专项资金资助(RF-A2020001).孙国道,博士,副教授,中国计算机学会(CCF)会员,主要研究领域为大数据挖掘和分析、信息可视化与可视分析. E-mail: guodao@zjut.edu.cn. 杨雨璠,硕士研究生,主要研究领域为大数据挖掘和可视分析. 潘翔(通信作者),博士,教授,中国计算机学会(CCF)会员,主要研究领域为模式识别、多媒体搜索、图形图像处理. E-mail: panx@zjut.edu.cn. 梁荣华,博士,教授,中国计算机学会(CCF)会员,主要研究领域为医学图像可视化、信息可视化与可视分析、计算机视觉、网络与服务计算.

text semantic visual analysis, the visual techniques of topic stack diagram, story line and river map are the main visual methods, which mainly analyze the content and emotion of the text with time series content. In the visual analysis based on multivariable text, it is mainly represented by matrix graph and heterotaxis graph to visually analyze the multivariable text related to epidemic. (3) Visual analysis of epidemic spread models to enhance users' interactive understanding of tasks such as prediction, simulation, and monitoring during the spread of epidemics. Based on the transmission dynamics model mainly includes the epidemic dynamics model, aiming at different epidemic transmission mechanisms, there are different transmission methods, and their visual forms are also diverse. In the visualization based on individual model, we can simulate and predict the spread of epidemic by tracking the relationship between individuals and others in the region. The visualization based on model comparison can show different epidemic development based on different decision-making measures in the epidemic. (4) Static information graphs of epidemic data, which intuitively reflect epidemic information. The contextual content and improve the efficiency of public understanding. This paper describes the dynamic visualization of epidemic information, epidemic news information map, epidemic virus information map. Based on the above classification, this paper summarizes some problems and challenges faced by the visual analysis of epidemic data. For example, it is difficult to obtain and store epidemic data, the multiple expansion of map visual analysis method, the difficulty of simulation and reproduction of epidemic model, and the lack of multi-disciplinary epidemic visual analysis methods. Finally, this paper summarizes the whole work.

**Keywords** epidemic data visualization; spatio-temporal visualization; semantic text; information graph; visual analysis

## 1 引 言

流行病是指可以感染众多人口的疾病,它能在人与人、人与动物之间相互传播,且传播途径多样化。大多数流行病可以通过直接接触、感染者的飞沫、食物、水源进行传播,例如流感病毒, H1N1 病毒, COVID-19 病毒等。目前国内各类流行病发病率逐年上升<sup>[1]</sup>,其中 COVID-19 疫情是 21 世纪以来最为严重的公共卫生事件,不同领域、不同学科的研究者都在紧密关注这一事件。根据调查,从该疫情暴发至今,相关论文文献已经超过 13000 篇<sup>[2]</sup>。因此对流行病相关的防控工作以及分析总结工作迫在眉睫。

本文将对流行病数据的可视分析方法进行总结,首先对不同领域中涉及不同类型的流行病相关数据进行分析与调研,并将其统称为流行病相关数据。例如在医学领域中,从病原体角度出发,数据包含毒力、数量、变异性、基本传染数( $R_0$ )等属性;从受感染群体角度出发,数据包含发病率、死亡率、治愈率、病情拐点、感染症状、传播时间、人为干预措施等属性。在新闻媒体领域中,数据包括半结构化的疫情播报的文本数据,非结构化的疫

情新闻数据等。在传播学领域中,患病者的个人轨迹数据同样也是流行病传播可视分析的重要数据。目前传统的流行病数据库<sup>[3]</sup>已经较为成熟,国内外流行病数据库内容较为全面,涵盖流行病临床与基础方面的新动态、新进展、新技术和新方法,包括流行病的诊断、治疗、预防、护理以及卫生防疫、流行病学相关的知识和各种感染病诊治经验等<sup>[4]</sup>。流行病相关数据涉及医学领域,传播学领域,新闻媒体领域,计算机领域等,数据在不同领域中所呈现的形式不尽相同。因此本文根据流行病数据所涉及的不同领域,将流行病相关数据分为 4 个方面,从时空维度(T1)、语义层面(T2)、传播模型(T3)、相关上下文内容(T4)进行介绍和举例,并阐述了上述类别数据的表现形式。表 1 所示为对流行病数据类型和相关数据集的主要分析任务总结。

### (1) 时空维度

时空维度中包含了时间属性、空间属性。例如移动运营商类、授权非公开数据资源中包含了大量的位置信息,Monasterio 等人<sup>[9]</sup>获取了阿根廷和墨西哥两家移动运营商的匿名交通信息来分析两个拉丁美洲国家的出行方式和查加斯病的可能危险区。Afzal 等人<sup>[19]</sup>基于时间维度提出了一种决策空间并

表 1 流行病相关数据

数据类型	数据集	主要分析任务
流行病学相关数据 (T1, T3)	CDC <sup>[5]</sup> , Epidemiology and Ecological Data <sup>[6]</sup> , COVID-19 Cases in Italy <sup>[7]</sup> , Socio attens <sup>[8]</sup>	病毒学研究, 基因序列分析, 流行病病例研究分析.
移动运营商数据、交通数据 (T1, T3)	Mobile Phone Data <sup>[9]</sup> , Voice Calls and SMS <sup>[10]</sup> , Air-traffic Network <sup>[11]</sup> , transportation network data <sup>[12]</sup>	传播学研究, 时空数据模式分析
新闻媒体数据、医学文本数据 (T2)	Twitter API <sup>[13]</sup> , Microblogging Data <sup>[14]</sup> , New York Medical Journal <sup>[15]</sup>	话题演变, 实时疫情数据分析, 源头跟踪
科普信息数据 (T4)	丁香园 <sup>[16]</sup> , 凤凰新闻 <sup>[17]</sup> , Flowingdata <sup>[18]</sup>	科普流行病相关知识

用来模拟传染病传播, 以减轻疾病影响或防止流行病进一步传播. 时空数据是目前流行病传播可视分析方法中较为常见的数据.

### (2) 语义信息

语义信息中以文本数据为主, 例如新闻媒体、杂志、Twitter 中有关于流行病数据的短文本数据集<sup>[20]</sup>. Lee 和 Kanhabua 等人<sup>[21-22]</sup>对流行病相关的 Twitter 文本内容进行了可视分析. 中国国家卫生健康委员会对 2020 年暴发的新型冠状病毒肺炎疫情以文本形式进行实时通报, 同时微博、澎湃新闻、丁香园<sup>[16]</sup>等对疫情新闻的实时发布也使用了文本的形式. 社交媒体在现代计算机技术的帮助下, 不仅能为疫情文本信息进行及时可视监测和报道, 还能成为有益数据的获取平台, 为不同群体的信息获取提供重要帮助<sup>[23]</sup>.

### (3) 流行病传播模型数据

流行病模型数据包括模拟仿真数据, 模型相关参数等. 流行病模型包括传播动力学模型, 个体模型等. 常见的流行病模型按照流行病类型分为 SI、SIR、SIRS、SEIR 模型等<sup>[24]</sup>, 按照传播机理又分为基于常微分方程、偏微分方程、网络动力学等不同类型. 这些流行病模型数据可以对不同类型的流行病进行模拟仿真、预测发展趋势等.

### (4) 流行病相关上下文内容

流行病相关上下文内容中包含了病毒基因、科普信息<sup>[25-26]</sup>等数据. 基于流行病信息图的呈现方式, 例如病毒海报图、动态信息内容展示和基因图谱等, 加快公众对流行病上下文的理解.

本文基于流行病数据将流行病可视分析任务分为以下四类: (1) 流行病数据时空可视分析, 该分析任务侧重于流行病的分布、传播情况<sup>[27]</sup>; (2) 流行病数据语义可视分析, 该分析任务主要针对流行病文本数据的内容进行挖掘, 分析情感变化等; (3) 流行病模型的可视分析, 该分析基于流行病模型数据, 通过数据建模, 分析传播走向或预测疫情发展;

(4) 信息图可视分析, 该分析任务所要分析的内容较少, 基于直观的内容可视展现来解释数据. 因此, 根据不同的分析任务, 应用不同的可视化形式, 能对流行病数据有更好地可视展现. 上述四类可视分析任务中, 虽然存在了一定的任务重叠以及可视形式重叠, 但其侧重点不同. 因此本文根据其研究的主要内容以及所使用的流行病数据来源进行分类. 表 2 所示为本文对流行病数据可视化相关任务和可视化方法所作总结.

### (1) 时空演变可视分析

时空演变可视分析主要是对时间、空间的变化进行分析探索, 例如流行病数据集<sup>[19]</sup>中包含了大量的时间信息、地理位置信息等. 由于这类流行病数据有数据总量大、时序信息冗杂、空间位置信息不直观等特征, 因此可视分析技术能帮助这类数据以一种较为直观的方式进行呈现. 其中基于统计图表的可视技术、地图可视分析技术为这类数据集的主要可视分析手段. 在流行病数据可视分析任务中, 由于流行病的暴发往往具有时空性<sup>[85]</sup>, 统计图表可视分析技术能较好地展现流行病总体的发展趋势, 同时地图可视分析技术能更加直观地展示流行病的传播情况, 对比评估防控措施.

### (2) 语义内容可视分析

语义内容可视分析所要处理的数据主要是基于流行病相关的文本数据, 其中包括非结构化、半结构化等类型的流行病文本. 例如非结构化的流行病相关的推特数据集<sup>[20]</sup>中, 由于该数据集总量庞大、前期数据清理难度大、有效数据信息提取成本高、数据源变化更新快等特点, 运用可视分析技术能较好地解决这些问题. 目前 SAS 大数据挖掘软件<sup>[86]</sup>可以对大量数据进行评估可视分析, 运用文本自动可视技术<sup>[87]</sup>进行实时舆情分析, 了解大众对当下流行病的看法和情感状况, 对流行病相关的媒体数据实时获取, 在面对突发的舆情热潮, 政府机关能及时进行宏观调控等.

表 2 流行病数据可视分析的分类表

可视形式		相关论文/工作
时间序列可视分析	统计分析	Kévorkian <sup>[28]</sup> 等人、晴雨表 <sup>[29]</sup> 、Stoesser <sup>[30]</sup> 等人
	措施比较	EpiSimS <sup>[31]</sup> 、Siraj <sup>[32]</sup> 等人、Baumgartl <sup>[33]</sup> 等人
	二维地图统计分析	Kajornkasirat <sup>[34]</sup> 等人、Lee 等人 <sup>[21]</sup> 、Hockett <sup>[35]</sup> 等人
	模式探索分析	Monasterio 等人 <sup>[9]</sup> 、Boone <sup>[36]</sup> 等人、TVECTOR <sup>[37]</sup> 、Sun <sup>[38]</sup> 等人、Hoque <sup>[39]</sup> 等人
空间维度可视分析	高维地图统计分析	Curtis <sup>[40]</sup> 等人、Olislagers <sup>[41]</sup> 等人
	二维地图时空演变	Bertini <sup>[42]</sup> 等人、epSpread <sup>[14]</sup> 、Malik <sup>[43]</sup> 等人、Bosch <sup>[44]</sup> 等人
时空数据可视分析	三维地图时空演变	Proulx <sup>[45]</sup> 等人、Krrak <sup>[46]</sup> 等人
	抽象视图时空演变	Luo <sup>[47]</sup> 等人、Stehle <sup>[8]</sup> 等人
语义文本本可视分析	基于症状的文本分析	Astefanoaic <sup>[48]</sup> 等人、Amiri <sup>[15]</sup> 等人
	基于基因的文本分析	Ésteban <sup>[49]</sup> 等人、Collins <sup>[50]</sup> 等人、Huang <sup>[51]</sup> 等人、Schaufler <sup>[25]</sup> 等人
	基于空间的时序文本探索	Bosch <sup>[44]</sup> 等人、TSI <sup>[52]</sup> 、Bánfi <sup>[53]</sup> 等人、Mircea <sup>[54]</sup>
	文本时序探索	Dou <sup>[55]</sup> 等人、Tanahashi <sup>[56]</sup> 等人、Dou <sup>[57]</sup> 等人、Leite <sup>[58]</sup> 等人
多变量文本可视分析	多变量文本探索	Wang <sup>[59]</sup> 等人、Sun <sup>[60]</sup> 等人、Szabó <sup>[61]</sup> 等人、Alshaabi <sup>[62]</sup> 等人
基于传播动力学模型可视分析	模拟传播分析	Stefano <sup>[63]</sup> 等人、Poletto <sup>[64]</sup> 等人、Fox <sup>[65]</sup> 等人
	基于个体模型可视化	Rao <sup>[66]</sup> 等人、Grignard <sup>[67]</sup> 等人、Corley <sup>[68]</sup> 等人
	基于地图的分析	Perez <sup>[69]</sup> 等人、Rao <sup>[70]</sup> 等人、Lequime <sup>[71]</sup> 等人
	基于时序的分析	The Washington Post <sup>[72]</sup> 、Flowingdata <sup>[73]</sup> 、Anylogic Cloud <sup>[74]</sup>
模型对比的可视化	基于地图的分析	Shehzad 等人 <sup>[19]</sup> 、Ajelli <sup>[75]</sup> 等人、FRED <sup>[76]</sup> 、Rizzi <sup>[77]</sup> 等人
	基于时序的分析	Paparian <sup>[78]</sup> 等人、Times <sup>[79]</sup> 、Flowingdata <sup>[80]</sup>
流行病信息图	病毒上下文内容	Schaufler <sup>[25]</sup> 等人、Stoesser <sup>[30]</sup> 等人、DiPiazza <sup>[81]</sup> 等人
	动态信息图	Flowingdata <sup>[82][83]</sup> 、Ernst 等人 <sup>[84]</sup> 、Kurzahls <sup>[26]</sup> 等人
	新闻信息图	丁香园 <sup>[16]</sup> 、凤凰新闻 <sup>[17]</sup>

### (3) 流行病模型的可视分析

流行病模型在传播学中研究广泛,其主要以传播机理可分为基于常微分方程、偏微分方程、网络动力学展开。基于上述方法的流行病模型会造成非该领域的研究人员较难入手,因此对流行病模型的可视分析同样重要。流行病模型可以有效地仿真预测不同类型流行病的传播过程。传统仅依靠流行病数据中的数字对传播过程进行解释的行为耗时耗力,可视化可以大大减少传统方法带来的弊端,从视觉角度观察模拟的传播过程,可以增强可信度,并且可以多角度同时展示流行病数据内容。例如石耀霖<sup>[88]</sup>基于传染病模型构造了 SARS 传播动力学模型。Jalayer<sup>[89]</sup>基于 ABM (Agent-based model),对 COVID-19 疫情进行传播模拟。

### (4) 信息图可视分析

由于流行病数据涉及领域较为广泛,涉及专业知识内容较多,对于非专业领域人员而言,面对复杂的专业知识较难快速理解掌握。信息图可视分析是将数据用可视手法进行直观展现的可视分析方

法,虽然缺少交互信息,但能友好地面向广大群众。因此信息图可视化成为面向非专业人员进行知识科普的主要呈现方式。例如使用信息图可视分析可以简单展示病原体的基因衍变和基因变异,也可以展示人体感染病原体的过程。

## 2 流行病数据的时空可视化

时空数据包括时间、空间、专题属性三维信息,具有多源性、数据量大、更新较快等综合特征。时空数据可视化在研究时空大数据多维关联描述的表达、关联关系动态建模与多尺度关联分析时提供了直观的展现手法<sup>[90]</sup>。因此,在包含大量时空数据的流行病数据中,进行时空可视化是最为直观的可视化方法之一,目前流行病数据的时空可视化任务是分析流行病在时空上传播规律。对流行病数据进行时空态势分析,可以监控流行病发展态势,及时发现异常情况<sup>[91]</sup>,处理缺失流行病数据<sup>[92]</sup>,评估流行病措施优劣等。本节将流行病数据的时空可视分析分为基于时间序列可视分析、基于空间维度可视分

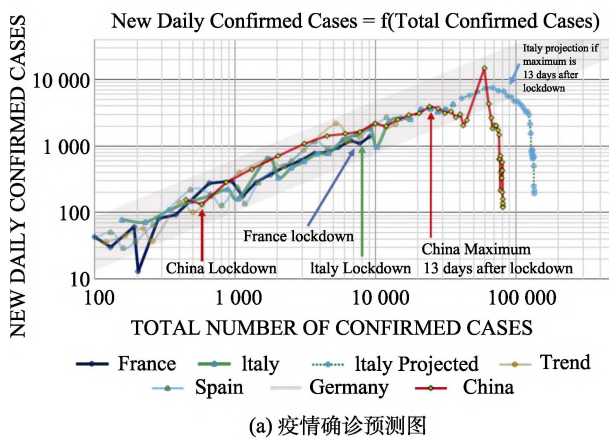
析和基于时空数据的可视分析。

### 2.1 时间序列可视分析

时间是描述事件发生过程或物体运动过程的一个参数，在流行病数据中，时序可视分析方法<sup>[93]</sup>能直观展示数据的发展趋势，例如折线图、柱状图、散点图等。本节对流行病数据的典型时间序列和非典型时间序列统计图表进行讲述。

时间序列可视分析能较好地展示不同时间段某流行病的发展演变。典型的时间序列可视分析中，将数据通过时序信息展示流行病发展趋势，其中 Kanhabua 等人<sup>[22]</sup>基于包含时间序列的推特数据，根据不同粒度来展示某疫情期间的时序可视化，但多重统计线条的组合易造成线条部分的重合，同时过多的线条展示也易造成视觉杂乱。Leite 等人<sup>[58]</sup>通过对流行病数据和新闻事件的时间分析来预测了疫情发展态势，基于时间序列可视分析方法，弥补了数据分析和新闻故事之间的距离，优化了折线图的之间的交互，降低了线条之间的杂乱程度。EpiSimS<sup>[31]</sup>针对不同的措施程度对流行病发展的影响进行了时序可视分析，通过线性分区可视化以避免线条部分重叠问题，但是也造成了比对困难。典型的时间序列可视分析大多基于时间坐标轴进行统计内容，缺少分析应用模块。因此非典型时间序列可视分析可以较好地弥补这些缺陷，如图 1(a)所示，Kévorkian 等人<sup>[28]</sup>将目标时序信息变成一个小点，以非常规的时间走向来绘制各个国家疫情下新增的人数，运用相空间图来对统计数据进行可视分析，虽然时序信息无法明确表示，但是可以直接比较其余非同步曲线，该方法无法较好地运用于海量时间点的数据集合中。

在时间序列可视分析中，如何更好地对比线与



线之间的区别至关重要。Siraj 等人<sup>[32]</sup>运用重叠面积图对不同类型的潜伏期下病毒密度的变化对比和可视展示。该种统计分析方法不仅能在有限的坐标区域中对发展趋势直观显示，同时还能获取重叠部分信息，对比多条线性的表示方法。该方法能让数据交集区间直观显示，但过多的重叠面积也会造成视觉错误。在数据量较大的基因演变数据中，如图 1(b)所示，Stoesser 等人<sup>[30]</sup>基于背景时间基线对大肠杆菌的基因序列进行了可视化，并结合时序属性进行额外标注。该可视分析方法能对不同时间下的流行病基因进行分类和对比，同时背景基线能减轻因过多的数据所造成的视觉杂乱。

在流行病可视化中，时间序列可视分析可以展现出数据整体的变化趋势，例如对疫情感染人数、治愈人数、病毒基因演变等和时间序列相关的数据的趋势的直观展现。基于简单的线性可视化，一些细节信息或者其他复杂的可视分析任务，无法较好地通过有限的坐标轴区域进行可视化表述。同时目前结合多个领域进行流行病综合评估的研究较少，挖掘和关联将会是未来流行病可视分析领域里较为重要的一部分。

### 2.2 空间维度可视分析

空间可视化中地理可视化可以直观反映出地理实体的空间特征和属性特征。1864 年如图 2(a)所示，John Snow 在手绘地图上标注了伦敦的霍乱发病案例，从而判断出水井污染是疫情暴发的根源，由此地图可视化成为了分析流行病患者人口分布的一种有效方法。目前地图可视化作为一种常见的时空可视化表现形式，能给人们带来地理位置分布的直观感受。在流行病数据地图可视化的研究中，位置属性可以衍生出较多有价值的信息，例如区域内的流

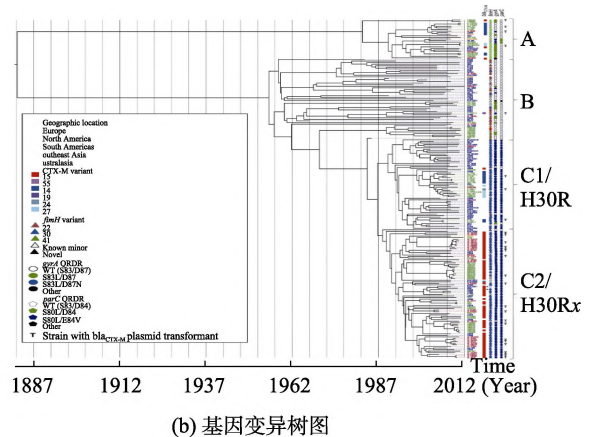


图 1 时间序列的可视分析

行病人口大致位置、人口感染密度、流行病的传播走向、流行病的发展趋势等。这些信息以可视化的方式可以展现出区域内的流行病发展态势，是较为常见的一种可视分析方法。

目前计算机科学技术发展迅猛，从早期人们通过手绘地图，人工收集病例数据，再进行手动标注的可视化方法，已经逐渐转变到运用计算机软件进行绘制地图，通过互联网进行流行病数据收集，自动生成更加美观且高效的地图，同时对病例数据进行自动标注。本节对地图可视化内部形式进行了概述，对于其所用到的流行病模型不予讲述，流行病模型内容将会在本文的第四章展开。

基于地理可视化是目前最为常见的流行病数据可视化手段。由于其可视化方式拥有不抽象，内容直观，通俗易懂等优点，因此地图可视化占据了目前流行病数据可视化的主要地位。本节对地理可视化分为：地图散点可视分析和热力图可视分析。

**地图散点可视分析** 该任务是发现地理数据集中目标关键词的频繁项集，并对项集进行聚合、分类和可视。如图 2(b)所示，City Sentinel<sup>[53]</sup>是一款“城市前哨流行病可视化工具”，从推文消息中检索隐藏信息来分析和定位流行病暴发。Twitter 文本具有实时性，该系统对庞大的文本内容进行挖掘，可以及时预测出流行病的暴发点。但由于基于地理位置的散点可视分析，大量散点极易造成杂乱、重叠等问题，由于目前数据更新较快，仅仅依靠地理位置进行散点标记，无法满足目前疫情发展趋势和疫情传播走向分析。因此 Boone 等人<sup>[36]</sup>使用 ArcGIS 软件，对疾病报告进行了挖掘分类，并将包含按症状分类的疾病报告的数据进行了颜色编码，然后将该数据覆盖在 Vastopolis 地图上，可以分别查看每次暴发情况，同时基于地理位置的“讲述故事”文本也可以运用于该可视化形式，逻辑进展会变得容易理解。该方法较好地解决了仅仅通过散点可视分析对区域内大流行病的布局限制，并设置时间戳功能对局部进行快照，对编码内容进行预测分析。

**热力图可视分析** 该任务是以特殊高亮的形式显示流量高的区域。热力图和散点图都可以表示数据分布情况，热力图的本质也基于散点图，通过高斯分布或者坐标点聚类计算来径向渐变的方式对散点图进行函数变换，生成目标热力图。在数据量较大的数据集中该方法优于散点图可视分析。如图 2(c)所示，Parikh 等人<sup>[37]</sup>使用基于像素格点的可视分析方法，对流行病相关的 Twitter 文本进行了聚类 and 可视分析，预测了暴发的可能来源。基于像素格点

的可视方法虽然缓解了因为数据量大而造成的散点杂乱，但同时也失去了较多真实细节，较为笼统地表示出某块区间内的大致暴发情况，无法对实际情况进行妥善标注。因此 Monasterio 等人<sup>[9]</sup>通过热力分布可视分析，使用色谱热力图来对移动电话记录进行分析，判断流行病的可能危险区，基于强度色谱的热力图是目前较为常见的可视分析方法。如图 2(d)所示，Mniszewski 等人<sup>[94]</sup>在模拟流行病在地理区域内的传播中使用热力图代替散点图，对比散点图能更好地地区分出不同区域之间的密度信息。同时该方法能较好解决因为数据量大而造成的视觉混乱，同时包含了不同时间段的传播信息，但目前已知流行病种类较多，该方法无法表达出两种或多种流行病的传播。

在空间维度可视分析中，散点图和热力图都能对离散数据进行表示，散点图可以是针对离散点的高斯分布或者聚类展现。但在密度过大的场景中，散点图容易造成杂乱、重叠，导致无法有效展示。而热力图则在数据量较少的空间中没有散点图反应的更加真实。目前在地图可视化中，如何展现更多的数据信息且不带来视觉杂乱是一个重要研究方向。

### 2.3 时空数据可视分析

时空数据来源于物体在时间和空间上发生的变化，本节对流行病数据的时空可视分析进行了总结。流行病时空数据运用可视分析方法，可以更全面地展现流行病数据的传播规律，直观展示数据内容等。本节对时空数据可视分析技术分为了以下几类：基于链接视图的可视分析、基于集成视图的可视分析和基于区域抽象视图的可视分析。

**基于链接视图的可视分析** 链接视图是将多视图单独摆放并相互关联的可视分析方法。例如以地理位置和时间轴组件分隔的方式，展示了流行病数据的时空可视化。如图 3(a)所示，Bosch 等人<sup>[44]</sup>提出 ScatterBlogs2 基于地图对区域进行实时危机监控，以交互式可视化方式进行展示，通过底部时间滑块对地理空间上的潜在的危机分布进行可视。该方法结合地图和其他功能组件对流行病数据实时监控，对比于空间可视分析中的地图可视分析，增强了实时性，交互性和任务性，通过交互变换能较好地解决基于单个地图但无法了解多种流行病的难题。因此如图 3(b)所示，Hoque 等人<sup>[39]</sup>展示了一款自然语言视觉应用界面的对话流程。用户输入问题例如：“Epidemic here”，结果显示基于可视分析的答案。该方法加入语义分析模块，可以根据用户自定义设置提问，同时基于地图实现相关统计可视分析，并

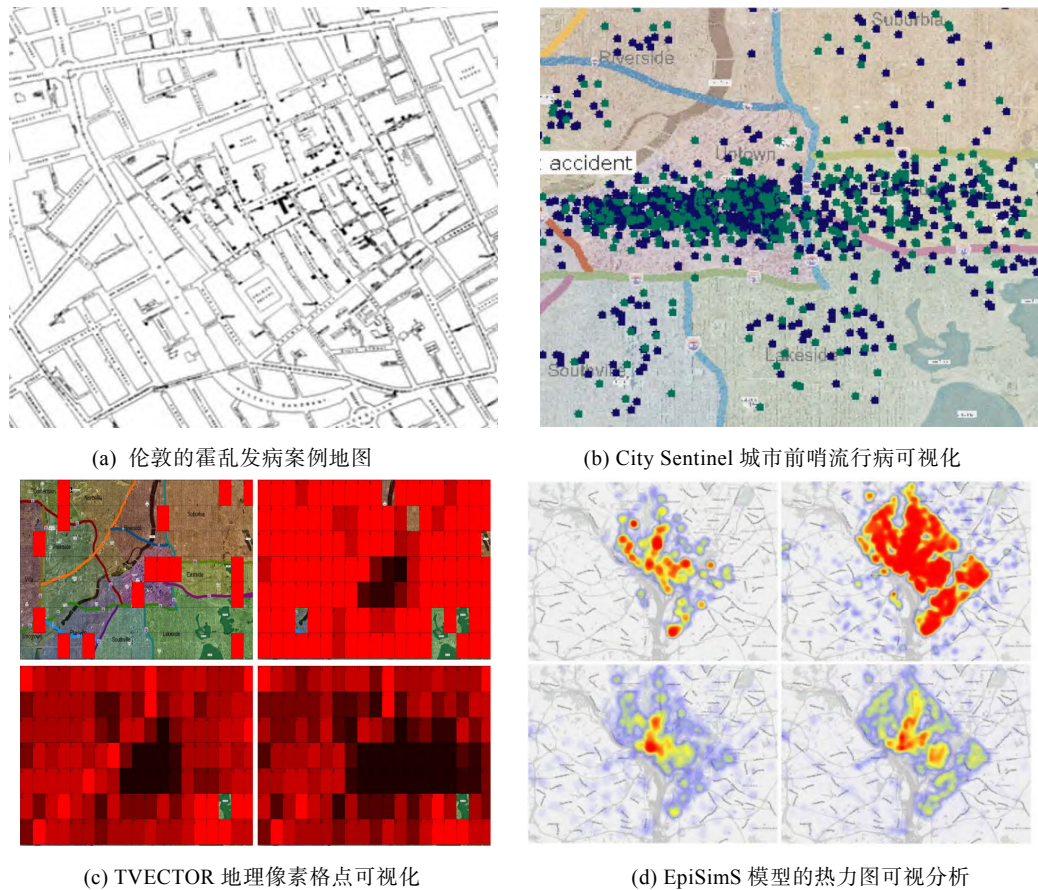


图 2 空间维度的可视分析

非局限于某一种流行病，拓宽了知识广度，但其研究深度降低，只能生成一些浅层可视分析图表。如图 3(c)所示，Bryan 等人<sup>[52]</sup>通过交互式注释生成和放置实现叙事可视化的方法，基于时间序列的河流堆栈图，显示流行病在地理位置上分布的变化，更加细节化。

**基于集成视图的可视分析** 集成视图是将不同视图同时集成于同一视图中，以减少用户关联负担的可视分析方法。例如在三维空间中，以高度作为时间轴，基于地面，抽象的表示出三维立体的可视形式来对流行病的传播进行分析。如图 4(a)所示，Proulx 等人<sup>[45]</sup>提出 nSpace 和 GeoTime 工具在原有地图的基础上对地图上散点进行时间表示，形成一个三维立体空间，在地图高度维度进行了传播过程的可视化，每一个散点都代表了一个三维坐标，相邻散点由于相互挤压无法立刻感知时间维度信息。因此如图 4(b)所示，Malik 等人<sup>[43]</sup>基于地图和时间轴，对美国暴发的 H1N1 流感进行了可视化探索 and 空间分析。较好地改善了在三维空间中时间信息不明确的问题。但是通过散点缺少了演变趋势，因此如图 3(d)所示，Kraak 等人<sup>[46]</sup>基于 STC<sup>[95]</sup>绘制出

历史感染云，提出了一个基于三维地图流行病传播可视化系统。将时间维度融入二维地图，每一个云层代表某地区该流行病的开始和结束，基于全局视角可以了解到该流行病在该地区上蔓延情况。是目前较好的一个集成视图的可视分析方法。

**基于区域抽象视图的可视分析** 区域抽象视图是将地图划分成抽象的区域模块，结合时间属性展现传播或人员之间的交流的可视分析方法。例如矩阵图是从多维问题的事件中，找出成对的因素，排列成矩阵图，然后根据矩阵图来分析问题，确定关键点的方法，是一种通过多因素综合思考，探索问题的好方法，其最大优点在于寻找对应元素的交点方便，且不会遗漏，同时显示对应元素的关系清楚。如图 4(c)所示，Lou 等人<sup>[47]</sup>提出了一个基于矩阵图的疾病有效控制可视框架 GS-EpiViz，Lou 将每个班级分成矩阵中的一个区域块，横纵轴为每个班级的排列，从而观察到班级与其他班级之间的交流程度。间接的模拟了国家地区之间的人口流动。Lou 基于矩阵二维对比的特性，进行了人口流动模拟，对每个人进行实时监测更加精准的同时，也带来了耗时耗力的缺点。矩阵散点交流图可对人员的流动进行

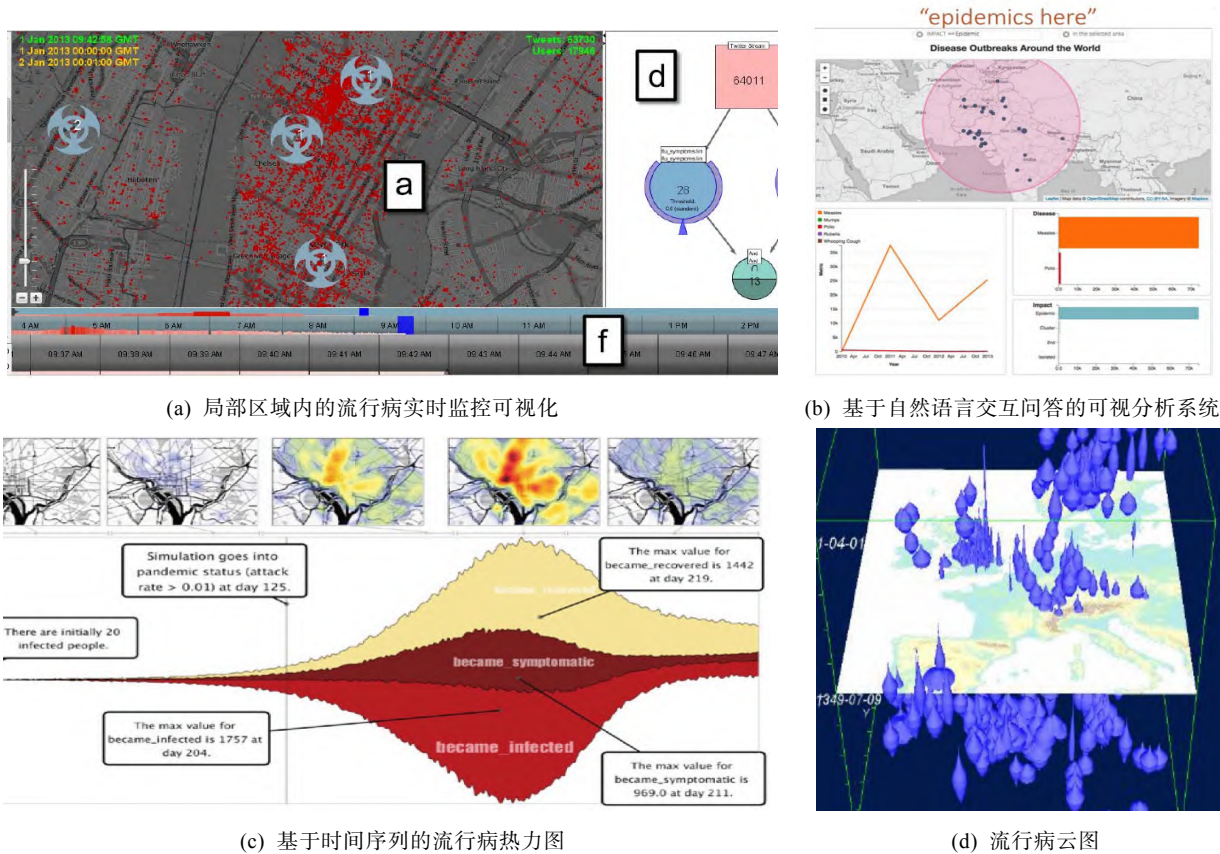


图3 基于流行病发展态势可视分析

抽象展现,但无法直观观察整体的行动轨迹.因此,如图4(d)所示,Stehle等人<sup>[8]</sup>基于整体来绘制时空轨迹线,抽象展现出不同集合在时间线上的交流情况.轨迹线中每一行对应一个特定位置,每一条彩色线对应一个班级的时空轨迹,可以展示不同时间,不同班级在不同区域的交流情况.该可视化方法能清晰展现单个集合在时间轴上的空间位置变化,但存在多区域重叠问题,如何优化轨迹线的重叠,自动生成最优的轨迹线<sup>[96]</sup>可视分析方法也是一个有价值的研究领域.

### 2.4 小结

流行病数据的时空可视分析主要方式基于统计图表、地理位置信息、时空交互可视分析,主要任务侧重于疫情数据统计、流行病传播情况、流行病地理分布、政策管控等.目前对于时空可视分析中对地理位置分布的疫情情况、传播分析、疫情数据趋势预测做到了较好地呈现,基于时空可视分析能让研究人员有对流行病数据的直观感受.目前流行病数据量庞大,病例数据复杂,时空中的可视分析大多只能对某一个人或某一个区域的信息粗略展示,无法较好地展现细节信息.因此对于该问题,

本文认为面对含有时空数据的数据集中,如果需要对其进行细节展现则需要高层次维度的加入,或者展开文字摘要标注同时结合地图分析将会成为较好的可视分析方法之一.

## 3 流行病数据语义可视化

本节介绍流行病数据语义可视化方法,根据文本层面情况,将语义可视化分为语义结构可视化、时序文本可视化、多变量文本可视化.

### 3.1 语义结构可视分析

流行病文本结构可视化基于处理过的结构化文本,本节介绍网络关系可视化和层次关系可视化.

**网络关系的可视分析** 目前较为常见的结构化文本可视分析方法是网络关系结构可视化,基于网络节点和连接的拓扑关系,直观地展示网络中潜在的模式关系<sup>[97]</sup>,例如节点或边聚集性,是网络关系可视化的主要内容之一.在流行病数据中,大量文本内容数据会导致难以快速了解其整体的拓扑结构,网络关系图能较好地解决这一难题,同时也能从中发现潜在的关联信息.

对于大量的报纸和医学杂志中和流行病相关的



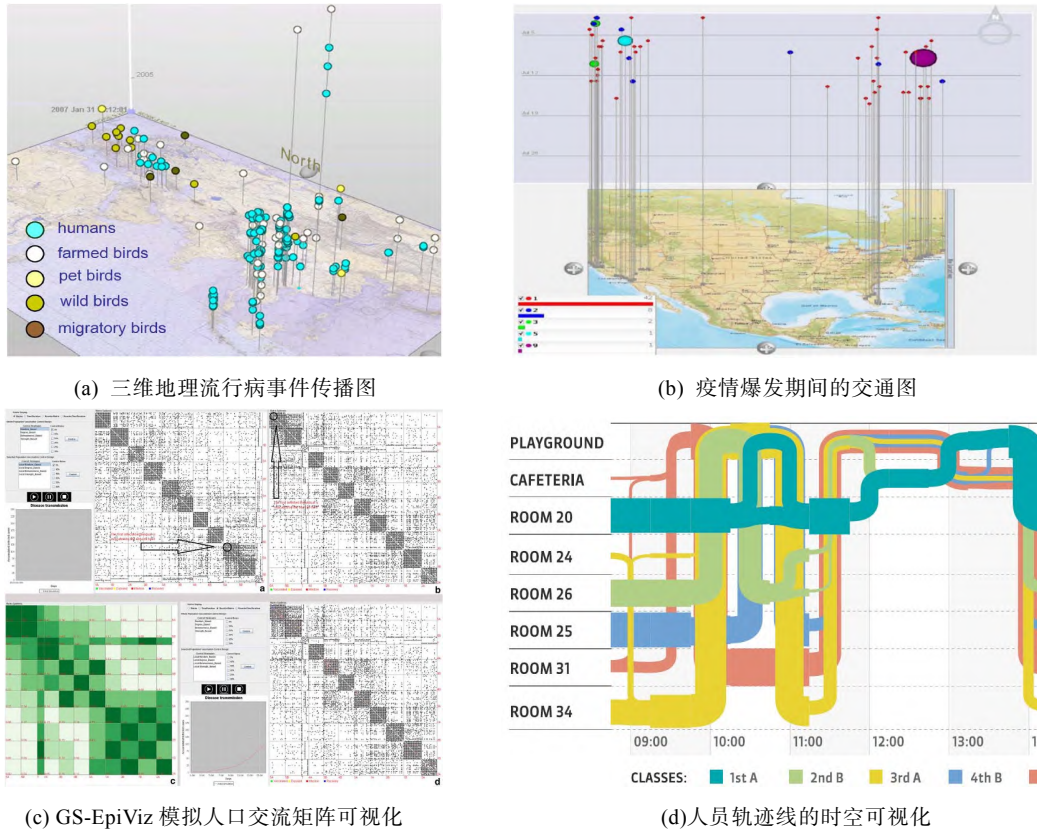


图 4 基于时空数据轨迹的可视分析

文章内容，快速了解该类数据集中最为核心的关键词是至关重要的，通过共现图可以快速认知该文章集合所代表学科中各主题之间的关系。如图 5(a)所示，Amiri 等人<sup>[15]</sup>基于共现图分析了有关于俄罗斯流行病的报纸和医学期刊。图中的节点数量为关键词数量，连线粗细代表共现频率，虽然用户能直观感受到关键词组合的共现，但是随着关键词数量的增加线条过多而会产生视觉杂乱。因此，如图 5(b)所示，Astefanoaie 等人<sup>[48]</sup>基于共现图的思想提出了一款弧形图，每个节点的大小表示症状出现的次数，同时依次排列，弧的宽度表示同时出现两种症状的记录的数量，可以快速了解某些流行病症状的相关性。弧形图的排列减少了节点的凌乱，通过分类、优化排列组合同样也能减少视觉凌乱。

上述两种可视化形式，是针对流行病的文本数据进行了关系展示的优化，但对具有复杂关联属性的流行病文本数据，需要更加详细的、有组织性的、可关联的、可操作的方法展示。知识图谱是一种较合适的、同时可以显示实体内容与结构关系的分析方法。目前 COVID-19 的知识图谱数据<sup>[98]</sup>日渐完善，Wise 等人<sup>[99]</sup>提出了可以加速文档信息检索和发现的新冠肺炎知识图谱。基于流行病的知识图谱分析

方法可帮助研究人员从科学文章语料库中提取实体和关系信息，同时可以回答关于流行病相关的科学问题。知识图谱是图分析可视化中图数据的重要组成部分，在知识图谱与可视化直接进行强关联的可视分析工作中，主要针对语义信息布局、关联信息呈现、大规模知识图谱构建等工作进行研究。

**层次化数据的可视分析** 层次关系作为常见的数据结构之一，它能表示出不同维度，节点信息、填充信息等属性，优点在于能概览整体结构和辨认层次结构的特征信息，易观察细节信息，拥有较好的人机交互机制，可视化结果具有一致性等。在流行病领域中，结构可视化在病毒的耐药性监测、流行病学分析和新临床谱系控制中起到了很好的可视效果。该方法利用包含关系表达层次化数据的可视化方法，其呈现数据的高效性和良好的伸缩性得到了广泛的应用，例如流行病病原体的变异发展、病原体基因族谱等。如图 5(c)所示，Collins 等人<sup>[50]</sup>使用聚类的方式，直观地展现出了流行病基因图谱的发育情况。如图 5(d)所示，Huang 等人<sup>[51]</sup>介绍了一种微生物的三层发育环状树图。发育环状树可以基于内部核心数据来对应添加环外数据，相较于基因发育结构树，更加体现出数据前后变化的内容。

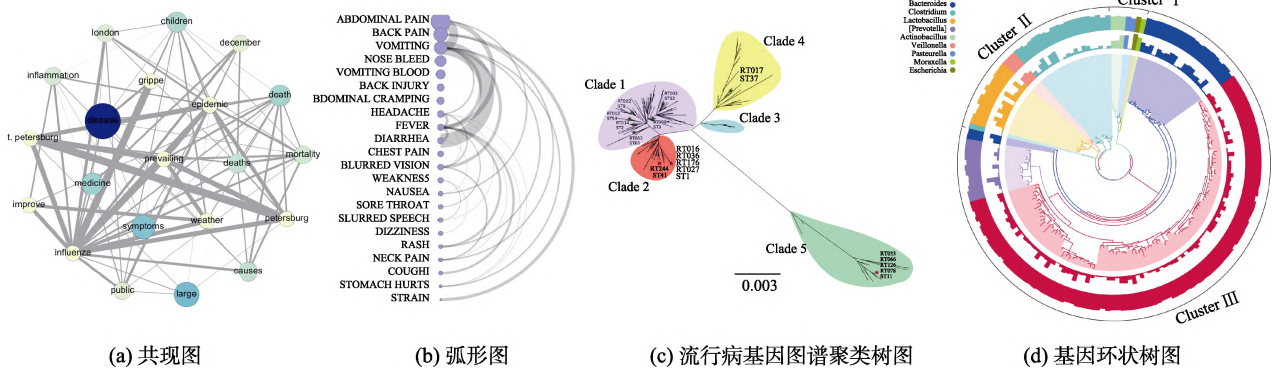


图 5 语义结构的可视分析

### 3.2 时序文本语义可视分析

时序文本是半结构化文本数据,其可视化形式基于时间轴或含有时序内容的展现形式,目前该可视化领域主要有主题河流、故事线等可视化方法。

**基于主题河流的可视分析** 主题河流可视分析方法用以表示事件或主题在一段时间内的权重变化,主题河流中不同颜色代表不同主题并将划分好的主题进行堆叠,其中条带的宽度大小代表该主题的影响程度或者当前条带属性的数值。基于这个特性,Tanahashi 等人<sup>[56]</sup>将流行病文本中的症状特征引入主题堆叠图中,如图 6(a)所示,Tanahashi 对医院和保健中心的结构化文本数据进行了主题河流可视化,可视化图中显示了多个症状的堆叠,能够直观地同时比较不同区域或不同症状之间的数据以及不同流行病的潜伏期,但是无法对非结构化文本进行有效处理。因此如图 6(b)所示,Dou 等人<sup>[57]</sup>使用 Twitter 文本,借鉴了主题河流的方式,提出了一个基于层次主题(Hierarchical Topics)可视分析系统,并将其运用到了流行病数据文本的可视化中。它可以有意义地将主题列表组织成一个层次结构,同时展现出流行病何时开始传播,以及症状如何随时间演变。

**基于故事线形式的可视分析** 故事线形式的可视化方法通常用来展示故事(如电影、小说等)中角色及角色间的互动(如是否发生交流等),同时结合颜色、标签,对某些事件的发展或群众情感变化等进行可视化。它可运用于疫情事件的展现和回顾。基于流行病数据的故事流可视化侧重于整个疫情事件的热点、情感走向和重大事件,及时了解大众对流行病的相关想法和其情感变化,有助于理解在社交媒体上的信息、观点和想法的传递。在疫情背景下,Baumgartl 等人<sup>[33]</sup>对故事布局进行了优化并设计了可视分析模块,通过调整布局参数来建立患者接触图,来追踪病毒初始来源,该方法适用于范围较

小,且在疫情大背景下其传播数据较难完全统计。在疫情背景的舆情中,如图 6(c)所示,Wang 等人<sup>[59]</sup>提出一个视觉分析系统 IdeaFlow,用于分析预定义的社会群体内部和群体之间的变化,并结合埃博拉病毒相关数据进行了可视化展示。如图 6(d)所示,Bryan 等人<sup>[52]</sup>提出了一个基于故事线的时间摘要的可视化,通过多时态视野,以及对重要的元素突出显示,其中对于流行病关键信息显示进行了重要程度的排名,用户可以选择注释数量进行自定义显示。鉴于流行病数据的数据本身具有更新快,数据量大等特征,并且对实时性要求较高,目前自动文本可视化技术<sup>[87]</sup>也逐渐成为了流行病数据可视分析研究的一部分。

**基于时间戳的词云可视分析** 该可视化方法通过对文本中出现频率较高的“关键词”予以视觉上的突出,结合时序信息形成含有时间戳的词云可视分析。基于流行病的词云可视化中内容是基于流行病文本数据,表现形式结合其他属性进行展现。如图 7(a)所示,Bosch 等人<sup>[44]</sup>提出了新颖的可视分析系统 ScatterBlogs2。在系统主窗口中,通过透镜滑动不同区域,可显示该区域内 Twitter 数据集中涉及到的流行病词云信息。这种方法能有效地监控流行病、重大灾难的发展趋势。Bosch 等人举例了一组用于检测来自报告流感相关症状的用户的消息和一组大雨检测天气的情况,并描述了交互式的、可视化分类器的创建。Twitter 信息的实时性,能快速了解社会生活中的舆情信息,由于 Twitter 信息具有无组织性、广度有限、缺乏可靠性等缺点,还需要结合传统媒体,两者相辅相成,舆情监测能及时对网络舆论危机的积极化解,对维护社会稳定、促进国家发展具有重要的现实意义。

### 3.3 多变量文本可视分析

多变量文本是一种含有结构化信息的文本类型

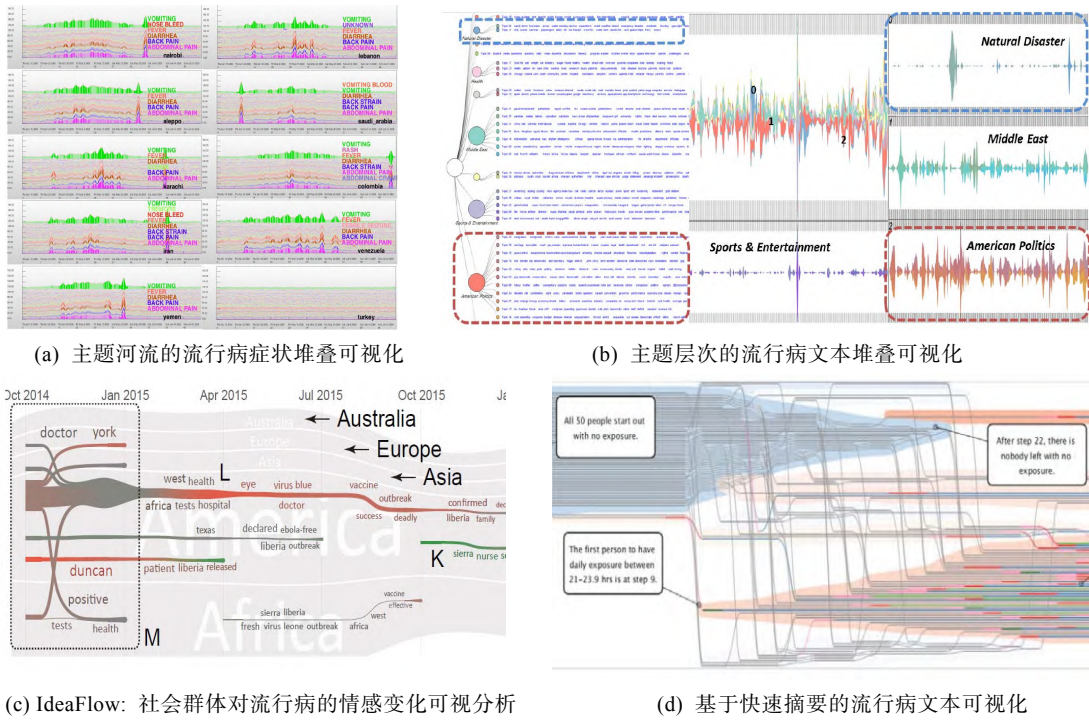


图 6 基于时间序列的文本可视分析

数据集,其可视化形式多样,拥有层次语义结构.本节将多变量文本可视化分类为语义传播结构图、矩阵语义图、异轴图,并展开综述.

在对大数据进行可视分析的过程中,界面中仅支持数据的可视表现无法达到解决问题、决策分析等各个环节的任务需求.界面还需要提供可视分析结果,或相应的交互式组件供分析者对任务需求进行管理交互.根据流行病数据背后影藏的模式和规律,在人机交互界面中包含了界面隐喻和交互组件和功能模块.例如对流行病数据的传播预测、舆情分析、病毒病理研究、网络调查、内容筛选等.

流行病语义传播结构可视分析 微博是目前大众分享简短实时信息的常用社交媒体软件,微博中包含了大量价值内容,国内外日均发布微博的数量已达到阻碍相关信息有效检索的水平.分析人员需

要新的可视分析方法来监控话题热点和信息传播走向,同时处理数据量及其动态性质.在时间策略任务中,为决策提供态势感知尤为重要.目前监控微博的工具通常根据用户定义的关键字查询和元数据限制过滤消息.这些方法单独使用时,在过滤精度和对趋势和信息传播变化的适应性方面可能存在缺点.

如图 7(b)所示,Sun 等人<sup>[60]</sup>提出 SocialWave,较好地解决了分析人员对媒体信息时空扩散的可视探索需求.Socialwave以 Lucene 文本数据库的形式,通过分时间分区域的数据分段存储分区存储,可以针对具体的业务场景展开检索,提高检索效率.同时该系统针对 2014 年埃博拉疫情暴发期间的社交媒体信息,对其时空传播模式展开了可视分析,允许用户快速识别整个时空扩散模式.运用动态模型

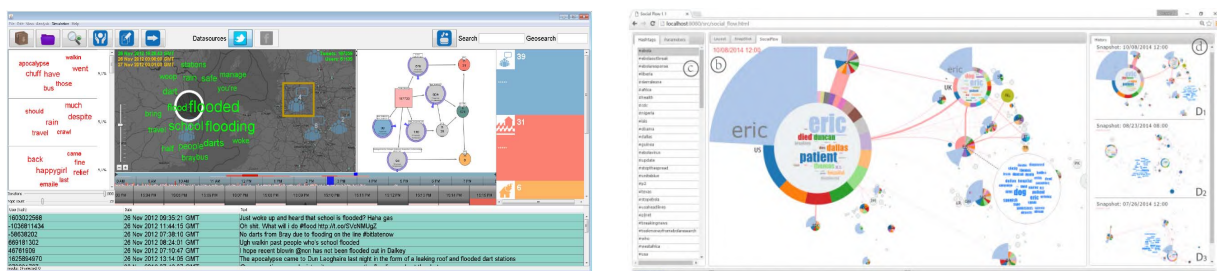


图 7 基于时序文本的疫情可视分析系统

(SGM)来量化信息交互行为, 优化整体主题结构, 较好地促进和加速各地区用户的信息传播, 但无法较好地比较各个主题之间内容的用户需求性. Mircea 等人<sup>[54]</sup>提出了为社交媒体共享的疫情信息实时分类可视分析方法, 帮助用户获得其中最有用的信息和最新消息. 研究人员基于语义传播结构的可视分析, 可以更好地了解某个大事件背景下的信息传播变化和趋势.

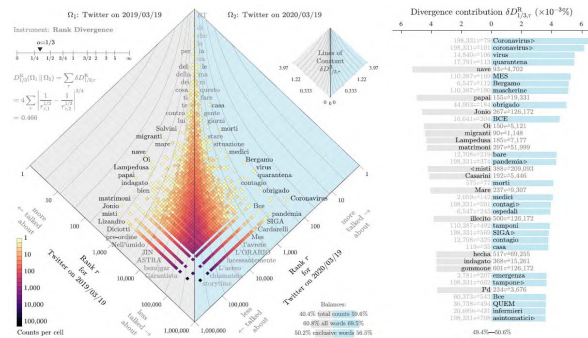
**基于矩阵图的流行病症状可视分析** 矩阵图的流行病症状可视化方法结合流行病症状属性进行可视分析. 如图 8(a)所示, Szabó 等人<sup>[61]</sup>提出了一款能够引起人们对任何时间序列数据剧烈变化的矩阵热图(heatmap), 并运用于流行病数据进行展示. 基于统计模型, 不同的症状显示在二维矩阵中, 每个矩阵对应一个位置, 它们表示某一天的患者数量. 每一列和每一行代表患者的症状或显著特征, 在行和列的交叉处, 表示同时具有相应症状的患者数量. 通过矩阵格点的颜色反映了患者的数量: 如果一个

矩阵格点有几天是黄色的, 那么人们应该保持警惕; 如果更多的矩阵格点是红色的, 则考虑这是暴发的流行病. 该工具可以帮助尽早查明流行病暴发, 同时矩阵可视分析方便研究人员将其中所包含的信息内容进行向量化表示, 易于储存和理解, 但是缺少了对其他外界因素的详细分析.

**基于异轴图的流行病话题热点可视分析** 异轴图的流行病话题热点可视分析能表示出很多维度的数据, 例如以矩阵格点的形式进行文字标注、排名对比、热力图展现、时间分析. 表达出了四个维度上的信息, 同时以中心轴为主, 横向对比了两个时间线中, 话题或关键词讨论的热度. 如图 8(b)所示, Alshaabi 等人<sup>[62]</sup>策划了 1000 个时间序列点从 2019 年 3 月到 2020 年 3 月, 对一些时间序列做了一些基本的观察, 包括与 COVID-19 引起的死亡人数的比较, 每个月话题或关键词的提及排名. 使用异轴图进行可视化, 可以利用这些时间序列来加强在流行病传播期间可能使用的任何分析以及回顾性调查.



(a) 基于矩阵的流行病症状可视分析



(b) 基于异轴图的流行病相关话题可视化

图 8 基于矩阵的流行病语义可视分析

### 3.4 小结

流行病数据的语义可视分析主要基于语义结构、时序文本和多变量文本可视分析, 主要任务侧重于流行病的文本内容可视分析、话题情感可视分析、舆情可视分析等. 基于语义可视分析的任务将文本内容以可视分析的方法展现, 大大减少了对海量新闻媒体数据的探索的时间, 帮助研究人员更快识别模式, 让用户深入了解细节层次, 帮助研究人员对长篇幅的流行病文本的层次分析和直观认识. 但是对于流行病数据的语义可视分析同样也存在一些问题和挑战, 面对日新月异的数据量更迭, 同时数据集合的复杂度的日渐提高, 对于新闻媒体领域中的媒体数据, 需要更好的数据规范化处理. 在实现文本可视化的同时, 如何同时更好地表达数据中

不同重要程度的信息内容是一个挑战, 在未显示的信息中, 可能也会存在用户或研究人员所需要认识的相关信息, 因此自定义信息粗度的文本可视分析系统一个较好的解决方案.

## 4 流行病模型和可视化

流行病数学模型, 是研究流行病的传播速度、空间范围、传播途径、动力学机理等问题, 以指导对流行病的有效地预防和控制. 基于传播动力学模型和基于个体的模型占据了流行病模拟、预测模型的主导地位, 随着机器学习的发展, 基于深度学习的流行病传播预测模型也逐渐成熟<sup>[100]</sup>. 本节对基于传播动力学和基于个体的传播模型及可视化进行介绍.

### 4.1 基于传播动力学模型的可视化

**传播动力学** 传播动力学是研究事物运动变化的一种因果关系，一般通过求解微分方程找出事物运动变化的原因，传播动力学模型中主要包括了流行病动力学模型、谣言传播模型等。流行病数学模型起源于 1927 年的 Kermack-McKendrick 模型<sup>[101]</sup>，K-M 模型等同于 SIR 模型。

如图 9(a)所示，Blue1Brown 等人<sup>[73]</sup>深入研究了 SIR 模型，改变输入变量的值，模拟不同情况下的传播情况。但该可视分析缺少了对其他因素介入的影响，例如执行隔离、疫苗介入、健康指导等，仅仅针对 SIR 模型进行了可视分析。因此如图 9(b)所示，Liu 等人<sup>[74]</sup>基于 SEIRS 模型，并且考虑流行病的介入行为对流行病控制的影响，通过人机交互的形式展现传播动力学模型内部的表现形式，该可视化方式更加贴近实际。

在流行病传播过程中，种群的流行病传播分析大多基于结构点，同时结合流行病模型进行演变。如图 9(c)所示，Poletto 等人<sup>[64]</sup>基于 SIR 流行病模型对种群流行病传播过程进行了可视分析，在该模型中，流动者拥有初始状态，并且他们停驻的时间为

非均匀分布的时间间隔。通过分析计算和数值模拟表明，图中较大宽度的灰色链接对应于感染路径，较浅的灰色链接对应于可见节点之间的现有连接。该模型对节点的开始时间、大小和程度对节点进行颜色编码，根据节点的程度对第一层节点进行排序，以突出不同程度节点在两种情况下的分层入侵模式中的作用。如图 9(d)所示，Fox 等人<sup>[65]</sup>基于随机模型对易感人群在流感中的演变进行了可视分析，橙色代表易受大流行病毒感染的个体及其之间的联系，灰色表示当前对大流行性感染免疫的个体。将结构点改变位置或颜色应用于流行病流动走向，以及对易感人群随时间的进化结构，展现了流行病传播模型流动模式。

### 4.2 基于个体模型的可视化

在流行病传播学中，通常使用个体模型进行可视分析。该方法通过跟踪每个个体在相关社会网络和地理区域（如同事、同学）中与他人的联系，或是个体的流动（例如往返工作场所和/或学校）以及与他人接触，来模拟流行病的传播。常见适用流行病模拟的基于个体的模型有元胞自动机模型、蒙特卡洛模拟、ABM 等。这些模型同时也能直观地展现

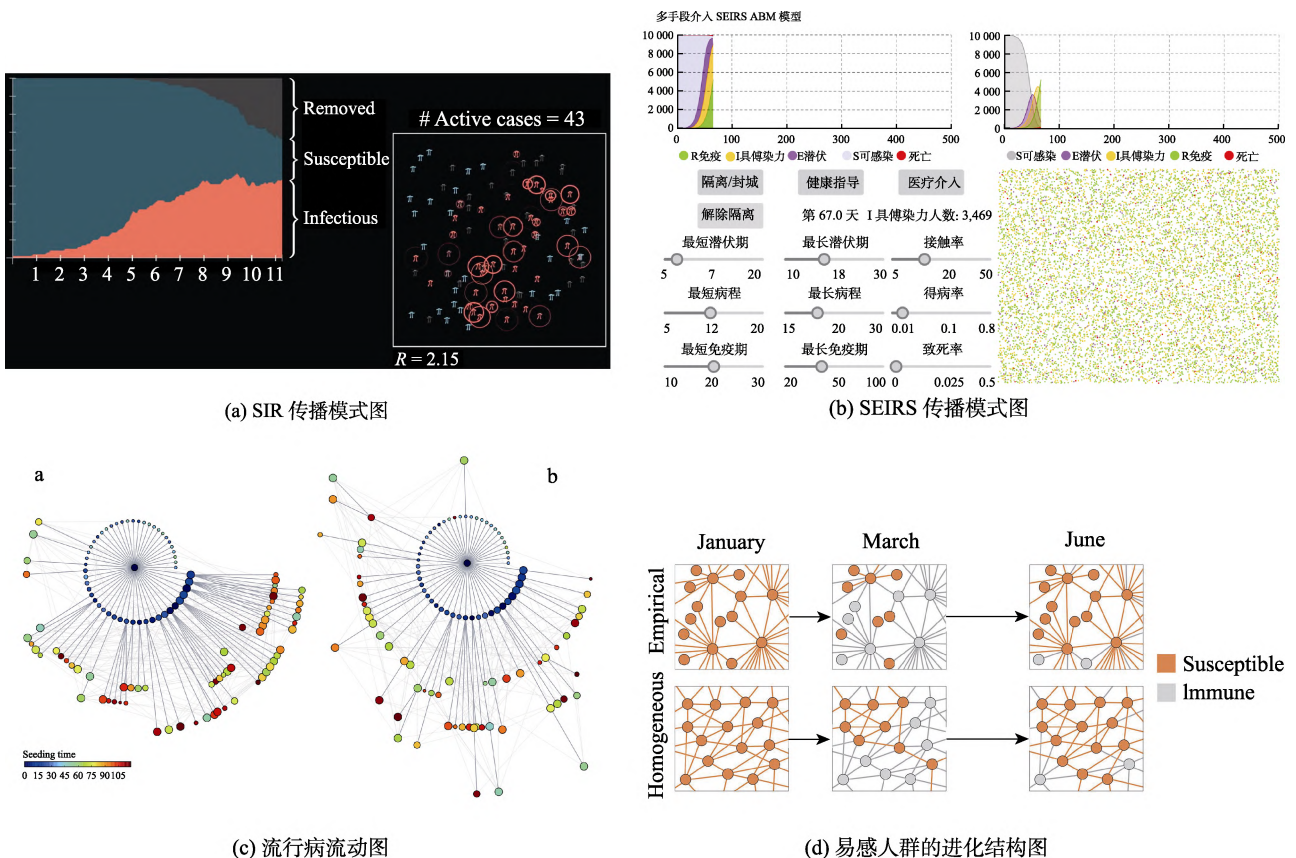


图 9 流行病模型的可视分析

流行病的传播模拟、预测分析. 本节将结合可视分析对基于个体的模型展开综述.

**元胞自动机的流行病传播模拟可视分析** 为了更好地预测和策略来对抗流行病的传播, 仅依靠建立微分方程组往往不够, 需要开发新的数学方法和计算方法. 在物理和流体动力学问题的背景下, “元胞自动机 (Cellular Automata, CA)” 算法成功运用, 元胞自动机构建基于个人之间的联系过程, 元胞自动机可以轻松地研究流行病传播的空间方面和疫苗接种策略, 并解决传染过程的自然随机性. 能够减少或消除“经典”微分方程建模方法局限性地使用 CA 的流行病模型引起了广泛关注.

如图 10(a)所示, Stefano 等人<sup>[63]</sup>提出了一种基于 CA/LGCA 模型, 并介绍了一种用于模拟流行病传播的软件工具. 该模型的主要特点是基于元胞自动机, 可以帮助识别流行病的空间特征. 该模型可以很容易地被修改, 来考虑流行病影响的人群的行为变化. 与基于偏微分方程的模型不同, 该模型是基于“个体”的. 流行病的传播是由于个体的运动和相互作用而发生的, 该模型对每个“个体”单独设置参数, 通过个体到全局的方法对流行病的传播进行可视分析.

但元胞自动机的一些限制因素制约了其模拟真实世界的的能力, 元胞自动机每一个固定的像素格按照规律有序排列, 无法真实代表一个个体, 涉及到的传播方式较为简单. 在现实生活中很少 (教室内人口流动模拟) 有如此规则和一致性相似的空间系统.

**基于代理模型的传播模拟可视分析** 代理模型 (Agent-based model, ABM) 是一种基于自主意识体系的模拟模型, 可用于计算独立个体或共同群体之间相互作用, 并通过可视分析评估个体在系统整体中的作用. ABM 已经广泛地应用于流行病的传播、疫情发展评估等, 该模型可以帮助研究者深入了解

不可先验的流行病模拟过程.

如图 10(b)所示, Perez 等人<sup>[69]</sup>基于 GIS-agent 的流行病传播模型, 对城市易感染人群进行空间分布的可视分析, 其中黑色圆圈代表易感人群, 黑色星星代表暴露人群, 黑色三角形代表受感染人群, 黑色正方形代表恢复(免疫)人群, 通过不同的散点可以模拟出全局感染过程, 并成功地复现了在复杂而现实的地理城市环境中流行病暴发的各种情况. 流动性的增加使人们能够真实地模拟一个人口个体的日常行为, 这些个体之间相互作用并且在从一个地方移到另一个地方后在固定的空间定位区域执行固定活动, 该模型的优点在于可简单地调整建模的疾病时间表和/或感染模型, 以修改传播过程来模拟任何其他流行病传播.

**个体传播模型可视分析系统** 个体传播模型可视分析系统能直观展现出每个个体的感染、传播情况, 但是一旦涉及到较多个体时, 个体的概念性移动会导致大量进程间消息, 这会显著增加同步开销并降低总体性能. Jon 等人<sup>[102]</sup>介绍了一种全球规模代理模型 (GSAM), GSAM 是基于病原体流行病建模的高性能分布式平台, 并且可以有效执行全球范围的流行病模拟, 高性能平台较好地缓解了基于个体传播建模的开销. Rao 等人<sup>[70]</sup>提出的基于加速并行代理的流行病学模拟, 发了一个称为 SEARUMS 的建模、并行仿真和分析环境, 对比了两种个体建模方法, 提出并评估了一种多重主动代理 (MAP) 方法, 该方法提高了性能, 显著减少了模拟执行时间. 如图 10(c)所示, Grignard 等人<sup>[67]</sup>介绍了一款 GAMA 建模仿真平台, 可以应对基于代理的模型复杂度的增加, 其中结合地理位置的方法, 展示了人口在区域内的流动情况.

如图 11(a)所示, Lequime 等人<sup>[71]</sup>提出的基于随机代理的传输链仿真框架 Nosoi, 是可调整和扩展的基于个体的框架, 可以在多种流行病学情况下, 是

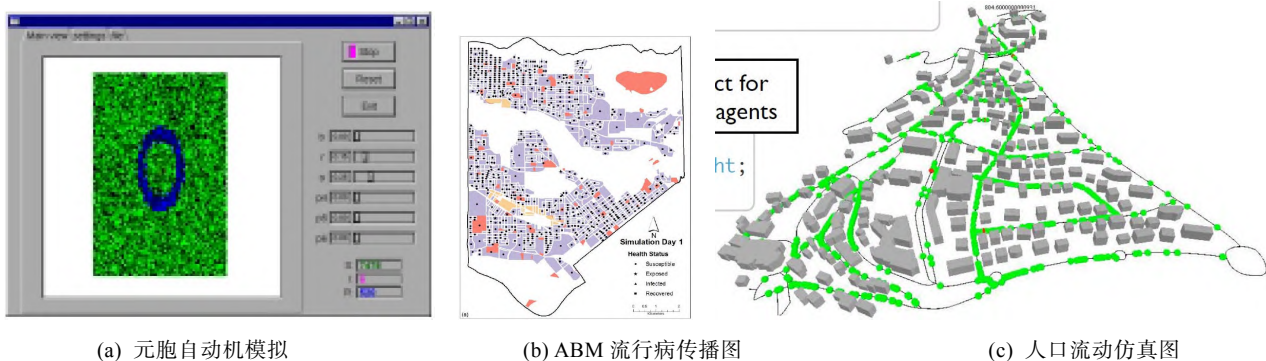


图 10 基于个体传播模型的可视分析

针对单宿主和双宿主传播情况的模拟传播链，其中网络图和树图中每个内部节点都是一个传播事件，每个尖端都代表一个代理的退出时间点，而传播链的根是最初感染的开始时间点，传播链的分支集合代表每个个体的生命周期。

### 4.3 模型对比的可视化

**流行病措施的线性比较可视分析** 流行病措施的线性比较可视分析是流行病模型的线性对比方法之一，如图 11(bc)所示，Papararian 等人<sup>[78]</sup>基于 FRED 进行了对流行病措施可视化的研究，FRED<sup>[103]</sup>是一个建模系统，支持流行病流行动力学的研究，以及缓解策略、病毒进化和个人健康行为的工具，Papararian 等人开发了一个名为 FRED Navigator 的工具，允许用户交互地探索先前计算的大规模模拟练习的结果，其中关于流行病传播期间学校关闭政策进行了可视化对比展示，通过 FRED Navigator 显示示例，图 11(b)中的横轴代表天数，纵轴代表影响人数，根据学校关闭持续的不同周数对每天受感染的人数进行了预测对比，图中不同颜色代表了关闭的

时长，图 11(c)中相应的“攻击率”曲线表明，在比较所有关闭学校政策的情况下，最终影响的人口百分比略有差距。如图 11(d)所示，纽约时报<sup>[79]</sup>展示了一个预测模型，对新冠肺炎疫情在不同时间点进行干预：结束公共集会、关闭工作场所和一些学校、大规模检测和加强医院使感染率降低并减少死亡，在可视界面加入人机互动的滑动条，对比不同时间点进行干预下疫情传播情况。

如图 12(a)所示，the Reich Lab<sup>[80]</sup>比较了六种对新冠肺炎疫情模型的预测结果，这六种模型使用不同的数学模型，不同的假设说明，得出了不同的轨迹，可以帮助决策者做出明智的决定。如图 12(b)所示，华盛顿邮报<sup>[72]</sup>使用 Simulitis 在 200 个人的小镇上进行传播模拟，每一个人在一个随机的位置开始，以一个随机的角度移动，动画的模式展现传播情况，同时对比了半隔离、社交隔离、硬隔离等隔离手段的传播情况。这些模拟大大简化了现实生活中流行病传播的复杂性，但同时降低了流行病传播的真实性。

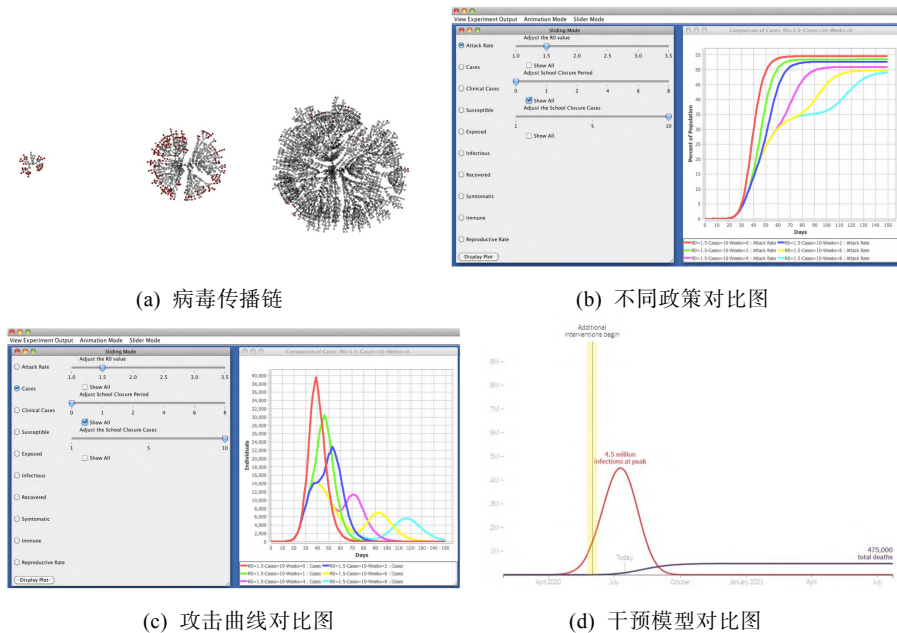


图 11 基于流行病模型的比较可视分析

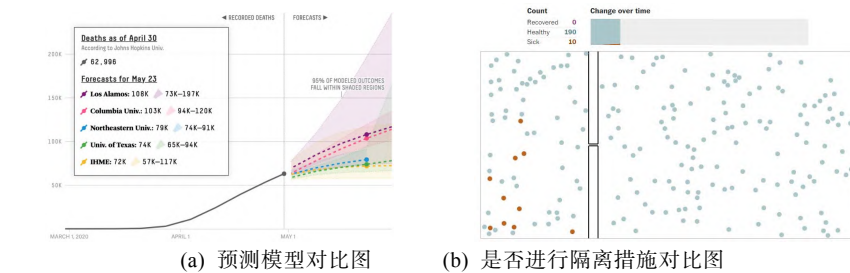


图 12 模型比较信息图

**基于代理模型的决策对比可视分析** 如图 13 所示, Afzal 等人<sup>[19]</sup>提出了一个交互式时空决策环境, 用户可以在不同时间点选择不同的决策措施进行模拟流行病传播, 可以对不同决策措施下的死亡率和感染率进行比较. 在该可视分析方法中包含了时空模型视图, 用户可以在时空上查看流行病的传播走向. 在历史决策视图中, 用户选择不同的决策条件进行交互. 决策路径随时间在  $x$  轴上绘制,  $y$  轴表示与基线模拟的累积偏差. 图 13 为大流行性流感病例研究, 并在不同的时间点引入了各种不同的决策度量, 探索地理空间中的模拟情况.

#### 4.4 小 结

在本节流行病模型的可视化中, 介绍了目前流行的流行病数学模型和基于个体的模型, 同时对其模型输出结果进行了可视分析, 其可视化任务主要为仿真模拟流行病在区域内的感染传播情况和对比使用不同政策管理调控的模型输出的防控情况. 目前对于流行病的预测、传播模拟大多都基于传播动力学模型进行分析, 借助可视化的方法对流行病模型的预测模拟结果进行展现. 流行病往往涉及到较多领域和学科, 因此不同领域的专家和研究人员都会参与其中, 构建针对不同领域任务的可视分析系统, 让研究专家或用户不必了解模型细节的前提下

结合自身领域知识来调整模型<sup>[104]</sup>也是很有价值的方向.

## 5 流行病信息图

很多综述<sup>[105-106]</sup>并没有针对信息图进行整理和总结, 但信息图是可视化领域中重要的一部分, 信息图虽然无法交互, 但是承担了一部分直观的数据理解任务. 要特殊强调的是, 信息图对分析来说并不是很适合, 但是本文认为这些信息图对人类快速理解流行病传播、衍变等内容承担不可或缺的作用.

**病毒信息图** 病毒信息图是对病毒的变异发展, 感染情况的可视化展现, 内容涉及到医学、生物学等领域的专业知识, 其优点是对整体内容的可视化表达, 能宏观地了解其中的内容信息. 如图 14(a)所示, Schaufler 等人<sup>[25]</sup>将变异的病原体进行了对比, 通过基因图谱的可视化方式, 对变异的病原体的基因序列对比. 如图 14(b)所示, DiPiazza 等人<sup>[81]</sup>对感染后的细胞进行可视化, 病毒侵入细胞后, 在细胞中的一系列变化, 病毒如何逃过免疫细胞的追踪、病毒如何繁衍等. 如图 14(c)所示, Corum 等人<sup>[82]</sup>描述 SARS-CoV-2 冠状病毒如何劫持健康细胞, 复制自己, 通过身体传播, 感染他人.

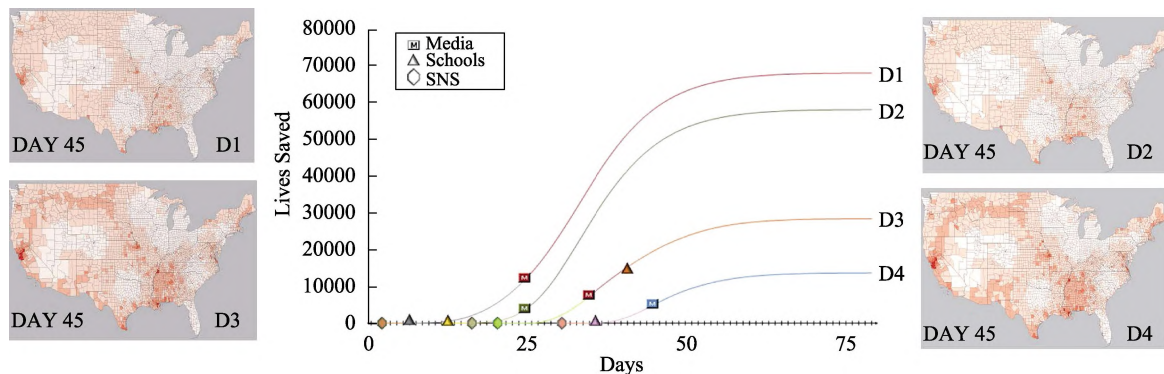


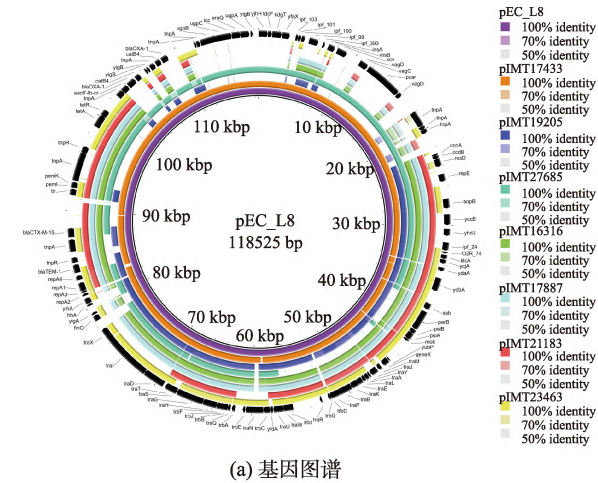
图 13 基于不同措施的流行病可视分析

**流行病信息动态可视化** 流行病信息动态可视化是一种对复杂的流行病知识以动态形式进行可视展现的可视分析方法. 如图 15(a)所示, Kurzgesagt<sup>[83]</sup>以一种动态信息可视方法, 解释新冠病毒中复杂而抽象的概念, 提出面对冠状病毒公众应该做出哪些措施手段的建议. 该方法展现病毒的变异发展过程, 病毒感染情况等动态信息内容, 其优点是抽象复现病毒在人体内的感染过程. 如图 15(b)所示, Ernst 等人<sup>[84]</sup>提出 BioNex, 一个挖掘、

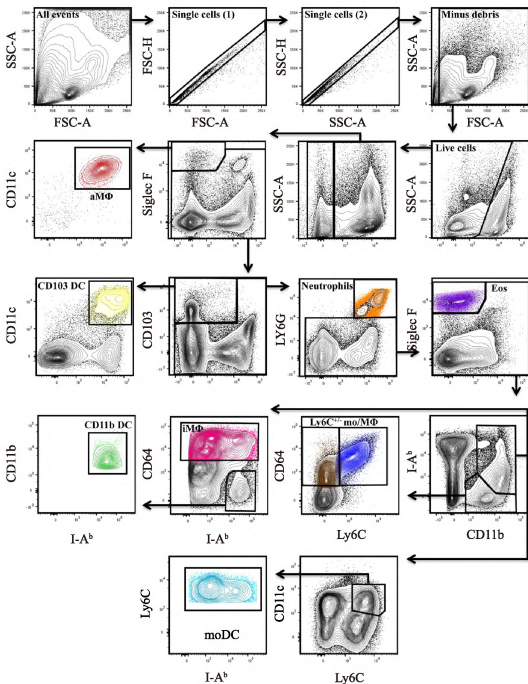
排序和可视化生物医学新闻事件的系统, 使用一系列历史相关的新闻事件集群和同一天的其他非生物医学事件为新闻事件, 同时提供上下文内容.

**流行病相关新闻信息图** 新闻数据由于具有实时性、可靠性等特点, 因此对于这些实时性的信息, 通过信息图进行表示能提高公众对流行病上下文信息和当前态势的理解效率. 很多自媒体通过信息统计, 制作简单的疫情信息图, 由于其具有信息实时更新性, 内容直接, 易于理解, 方便获取

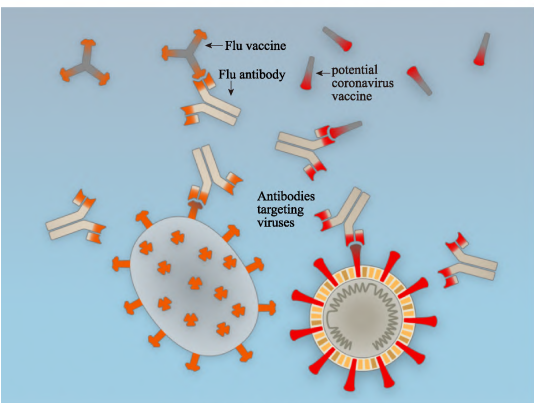




(a) 基因图谱



(b) 流行病学传播图



(c) 病毒信息图

图 14 基于病毒相关内容的信息图

疫情信息，受到了大众非常高的关注度。如图 16(a)<sup>[107]</sup>所示，展示了冠状病毒的测试精度，如图

16(b)<sup>[18]</sup>展示了在 2020 年的新冠肺炎疫情中因病毒而死亡的人的损失展示。如图 16(c)所示，DeBold 等人<sup>[108]</sup>在热图中显示每 100,000 人中患流行病的病例数，并高亮疫苗出现时间，对比疫苗对流行病传播的影响。

流行病学信息图包含病毒信息图、病毒感染信息图、疫情信息图、流行病学知识图。其可视化方式直观易懂，简化复杂概念，快速展示内容亮点。信息图是一种面相非专业人士的数据可视化方法，同时也是可视化领域中的重要组成部分。

### 6 问题和挑战

#### (1) 流行病学数据的获取和存储

本文将流行病学数据分为病例数据(其中包括患病情况、个人情况等)、流行病学相关文本数据(医学文本、新闻媒体等)、流行病学相关模型数据(模拟仿真数据、模型参数等)。在早期的流行病学可视化研究中，流行病学数据通常是对某个地区进行实地调查进行数据获取，这样的方法无法获取完整的流行病学病例数据集，造成数据存在偏差，重复，无效，单一，质量欠佳等问题，从而影响流行病学数据可视化的科学进展。近年来随着存储条件的高速发展以及获取数据手段多样化，大部分流行病学患者在就医后，个人病例数据保存于专门的存储管理系统中，并在整理和清洗后可供科研工作者们使用。由于并非全部的患病人员会选择就医，无法获取较为完整的数据集合，统计的数据容易出现漏洞或错误，因此预测技术应运而生。例如 2020 年初，出现的新型冠状病毒肺炎疫情，我国通过健康码来对国内每一个使用智能手机的居民进行实时轨迹监控，根据疫情的严重程度，不同地区显示不同颜色的健康码，从而控制人口流通率，减少传播患病概率。利用智能手机来追踪患者行迹。虽然这样的方法还是无法较为全面地了解每一个居民的健康状况，同时“健康码”技术<sup>[109]</sup>易产生隐私泄露问题，但是通过展示颜色来表示自己可能处于的健康状况，较好地降低了各个区域之间的人口流动，从而减少病毒传播。在我国法律的保护下，流行病学数据可视化的数据获取问题将会是可视化分析面临的第一个挑战。

#### (2) 地图可视化形式的多元拓展

在流行病学时空数据中，可视分析方法基本设定在基于地图等的形式。由于位置属性是流行病学数据较重要的数据属性之一，结合地图可视化呈现位置信息的方法较为普遍和通用。对于流行病学时空数据而言，如何转变可视化形式，将其他维度的信息融

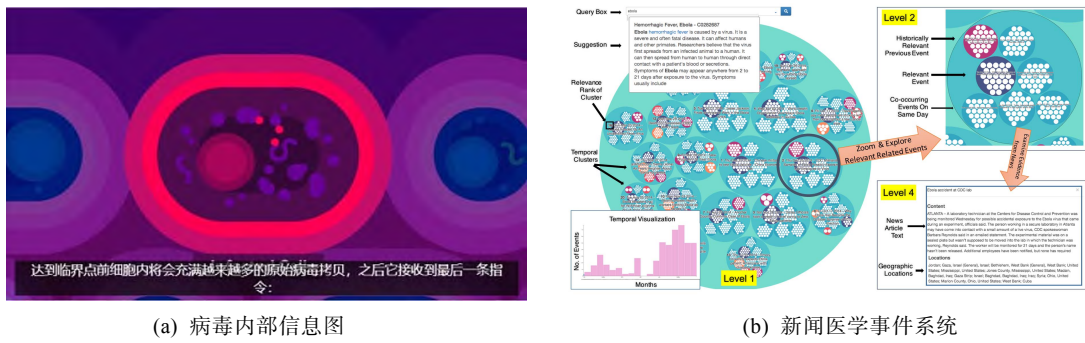


图 15 基于流行病上下文内容的信息图



图 16 基于新冠肺炎疫情的信息图

入到时空数据中是目前的一个挑战。例如目前地图可视化大多以某一时刻的快照形式展现，当数据集中包含轨迹数据时，地图可视化较难将其通过地图进行轨迹复现，除了因为地图比例尺过大，导致移动不明显外，在地图上散点的移动，极易造成视觉杂乱，只能大致感知某区域内交流频率。如何在地图上合理运用轨迹信息，并进行可视分析呈现也是一个挑战。其次，基于疫情传播速度而言，极易造成局部地区疫情极其严重，而其他地区疫情较轻的情况，导致在对疫情数据的地图可视分析时出现局部地区患病人口值极其巨大，而其他地区则极小，常见的地图散点分析、地图热力图分析已经无法将患病人口极大值处的区域更具有特征性的表示。目前对地图形状进行改变，来强调某区域的患病人口异常大，而这样的方法改变了地图形状，从而也丧失了对整体疫情的了解。因此对数据值差异极大的时空数据的可视分析呈现是一个挑战。

(3) 流行病模型模拟传播可视分析

流行病模型的可视化模拟中，首先对流行病模型的选择是至关重要的。然而目前我们熟知的流行病相关模型都无法对任意一个流行病的传播进行真实的模拟复现，除了因为流行病传播数据的缺失和难以获取以外，同时我们无法获知病毒是通过什么途径对被感染者进行侵犯。目前大多数的模型可视化，只是对某个病毒传播的模式抽象复现。通过可视分析手段，可以直观地感受不同流行病模型对

某种流行病的模拟真实程度，从而进行选择模型或微调模型参数。其次，流行病真实传播可视需要极其庞大的数据支持，同时硬件设备要求较高，目前的数据获取条件和平均硬件水平难以实现理想的流行病真实传播轨迹重现。在流行病模型可视化中，目前亟待一个可以较好比较相关模型或模拟较为真实传播的模型和可视分析系统。

(4) 面向多学科交融的流行病可视分析方法

目前不同领域如生物医学、网络动力学等对流行病或流行病数据都展开了不同层面的研究，如何将各领域研究得到的成果与计算机科学有效结合是目前的一个挑战。例如在流行病数据的获取中，自动综合不同领域或学科的数据并展开融合的交互式系统目前缺乏。在突发重大卫生事件时，疫情初期流行病数据往往难以获取和统计，流行病数据的深度和广度往往超过人们历来的先验理解。

与此同时，在数据处理过程中，由于流行病数据来源广泛，其数据类型结构不统一，在对数据进行规则标准统一过程中，需要涉及到大量医学、传播学等领域的知识。例如在面对半结构化的临床病例数据，非结构化的医学影像信息，数据流结构的传播轨迹数据等较为繁多的数据信息时，应该以多角度，多学科进行交融分析。在针对流行病数据的可视化方面，交互式新闻所呈现的内容如丁香网、凤凰新闻等往往更适合大众，并不适合专家用户展开深入探索。因此，如何在可视分析系统中，

体现领域需求和用户意图存在诸多挑战。例如, 针对新冠病毒不同毒株的差异<sup>[110]</sup>可视分析及其时空变迁方面的探索, 需要自定义的可视视图以综合生物信息学算法<sup>[111]</sup>、网络动力学知识等知识。

## 7 结束语

流行病数据可视化发展经历了十多年, 本文基于不同领域或学科对流行病数据的可视分析进行了研究, 总结了当前较为流行的流行病数据的可视分析方法, 并将其分类为: 时空可视化、语义可视化、模型和可视化和信息图可视化。

时空可视分析中基于地图对流行病的分析占据了目前主要的地位, 未来研究可以将相关联的多种流行病在地图上进行可视分析, 从而帮助相关领域的专家进行深入研究。语义信息对流行病的传播和发展也是至关重要的, 通过新闻媒体, 或者文献杂志, 对其中的流行病文本进行语义提取分析可视化, 可以监测或了解某些流行病的发展, 是一个较好的研究领域。流行病模型可以对流行病进行模拟、仿真、预测, 结合可视分析能更加全面地了解某种流行病的传播方法和特点, 目前流行病模型无法做到完全真实模拟某种流行病的发展情况, 结合可视化可以作为对某种流行病直观认识。流行病信息图是一种面向公众进行流行病科普的可视分析方法, 能帮助专业领域传播流行病知识, 快速直观认知流行病内容。最后, 本文对目前流行病数据可视分析存在的问题和挑战进行了总结。

## 参 考 文 献

- [1] Yuan Yunxiao, Wang lining, Wang Baohai, Li Dongyan. Spatial econometric analysis of the agglomeration and diffusion of infectious diseases in China--Based on spatial panel data. *Practice and Understanding of Mathematics*, 2020, 50 (21): 144-150 (in Chinese)  
(原云霄, 王丽宁, 王宝海, 李东燕. 中国传染病集聚与扩散的空间计量分析——基于空间面板数据. *数学的实践与认识*, 2020, 50 (21): 144-150)
- [2] Pei Tao, Wang Xi, Song Ci, Liu Yaxi, Huang Qiang, Shu Hua, Chen Xiao, Guo Sihui, Zhou Chenghu. Research progress on spatiotemporal analysis and modeling of Covid-19 epidemic. *Acta Geoinformatics*, 2021, 23 (2): 188-210 (in Chinese)  
(裴韬, 王席, 宋辞, 刘亚溪, 黄强, 舒华, 陈晓, 郭思慧, 周成虎. Covid-19 疫情时空分析与建模研究进展. *地理信息科学学报*, 2021, 23(2): 188-210)
- [3] Collier, Nigel, Son Doan. GENI-DB: a database of global events for epidemic intelligence. *Bioinform-Matics*, 2012, 28 (8): 1186-1188
- [4] Brudey K, Driscoll J R, Rigouts L, et al. Mycobacterium tuberculosis complex genetic diversity: Mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiology*, 2006, 6(1): 1-17
- [5] INFLUENZA S. Centers for Disease Control and Prevention. <http://www.cdc.gov/flu>, 2012
- [6] Lopez D, Gunasekaran M, Murugan B S, et al. Spatial big data analytics of influenza epidemic in Vellore, India//*Proceedings of the IEEE International Conference on Big Data*. Washington, USA, 2014: 19-24
- [7] Zhan C, Tse C K, Lai Z, et al. Prediction of COVID-19 spreading profiles in South Korea, Italy and Iran by data-driven coding. *PloS One*, 2020, 15(7): e0234763
- [8] Stehlé J, Voirin N, Barrat A, et al. High-resolution measurements of face-to-face contact patterns in a primary school. *PloS One*, 2011, 6(8): e23176
- [9] de Monasterio J, Salles A, Lang C, et al. Analyzing the spread of chagas disease with mobile phone data//*Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. Davis, USA, 2016: 607-612
- [10] Saravanan M, Karthikeyan P, Arathi A, et al. Mobile agent-based approach for modeling the epidemics of communicable diseases//*Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. Niagara Ontario, Canada, 2013: 16-20
- [11] Ribeiro L F R, Saverese P H P, Figueiredo D R. struc2vec: Learning node representations from structural identity//*Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, USA, 2017: 385-394
- [12] The Greater Vancouver transit authority. <http://www.translink.bc.ca>, 2020
- [13] Twitter streaming api. <https://dev.twitter.com/docs/streaming-apis>, 2020
- [14] Ap Cenydd L, Walker R, Pop S, et al. epSpread-Storyboarding for visual analytics //*Proceedings of the IEEE Conference on Visual Analytics Science and Technology*. Providence, USA, 2011: 311-312
- [15] Amiri S E, Tabassum A, Ewing E T, et al. Tracking and analyzing dynamics of news-cycles during global pandemics: a historical perspective. *ACM SIGKDD Explorations Newsletter*, 2019, 21(2): 91-100
- [16] Ding Xiangyuan, Real-time Epidemic Situation of New COVID-19 Pneumonia. <https://ncov.dxy.cn/ncovh5/view/pneumonia>, 2020 (in Chinese)  
(丁香园, 新型冠状病毒肺炎疫情实时动态. <https://ncov.dxy.cn/ncovh5/view/pneumonia>, 2020)
- [17] Phoenix News, Real-time Situation of the National COVID-19 Epidemic. <http://news.ifeng.com/c/special/7uLj4F83Cqm>, 2020 (in Chinese)  
(凤凰新闻, 新冠肺炎 COVID-19 全国疫情实时动态. <http://news.ifeng.com/c/special/7uLj4F83Cqm>, 2020)
- [18] FLOWINGDATA. An Incalculable Loss. <https://flowingdata.com/2020/05/24/an-incalculable-loss/>, 2020
- [19] Afzal S, Maciejewski R, Ebert D S. Visual analytics decision support environment for epidemic modeling and response evaluation//*Proceedings of the IEEE Conference on Visual Analytics Science and Technology*. Providence, USA, 2011: 191-200
- [20] Cao N, Lin Y R, Sun X, et al. Whisper: Tracing the spatiote-

- mporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2649-2658
- [21] Lee K, Agrawal A, Choudhary A. Real-time disease surveillance using twitter data: demonstration on flu and cancer//*Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Chicago, USA, 2013: 1474-1477
- [22] Kanhabua N, Romano S, Stewart A, et al. Supporting temporal analytics for health-related events in microblogs//*Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. Maui Hawaii, USA, 2012: 2686-2688
- [23] Lu jinlang, Wang Wanzhou, Zhang Yunjing, Zhou Jiangjie, Wang Shengfeng. Progress in novel coronavirus pneumonia (COVID-19) based on social media. *Chinese Journal of Disease Control*, 2020, 24 (10): 1209-1214 (in Chinese)  
(吕瑾莩, 王万州, 张云静, 周江杰, 王胜锋. 基于社交媒体的新冠肺炎(COVID-19)研究进展. *中华疾病控制杂志*, 2020, 24(10): 1209-1214)
- [24] Li Hao, Duan Guang, Tao Xueqiang, Chen en, Gao Shu Tian. Infectious disease dynamics model and its application in New Coronavirus pneumonia epidemic simulation prediction. *Medical and Health Equipment*, 2020, 41 (03): 7-12 (in Chinese)  
(李昊, 段德光, 陶学强, 陈恩高, 树田. 传染病动力学模型及其在新型冠状病毒肺炎疫情仿真预测中的应用综述. *医疗卫生装备*, 2020, 41(03): 7-12)
- [25] Schaufler K, Semmler T, Pickard D J, et al. Carriage of extended-spectrum beta-lactamase -plasmids does not reduce fitness but enhances virulence in some strains of pandemic E. coli Lineages. *Frontiers in Microbiology*, 2016, 7(336): 1-12
- [26] Kurzhals, K., Hlawatsch, M., Heimerl, F., Burch, M., Ertl, T., & Weiskopf, D. Gaze stripes: Image-based visualization of eye tracking data. *IEEE Transactions on Visualization and Computer Graphics*, 2016, 22(1): 1005-1014
- [27] Guo D. Visual analytics of spatial interaction patterns for pandemic decision support. *International Journal of Geographical Information Science*, 2007, 21(8): 859-877
- [28] Kevorkian A, Grenet T, Gallee H. Tracking the Covid-19 pandemic: Simple visualization of the epidemic states and trajectories of select European countries & assessing the effects of delays in official response. *medRxiv*, 2020: 1-6
- [29] Peking University, COVID-19 Pneumonia Epidemic Analysis System. <http://vis.pku.edu.cn/ncov>, 2020(in Chinese)  
(北京大学网新冠肺炎疫情网可视分析系统. <http://vis.pku.edu.cn/ncov>, 2020)
- [30] Stoesser N, Sheppard A E, Pankhurst L, et al. Evolutionary history of the global emergence of the Escherichia coli epidemic clone ST131. *MBio*, 2016, 7(2): 1-15
- [31] Mniszewski S M, Del Valle S Y, Stroud P D, et al. EpiSimS simulation of a multi-component strategy for pandemic influenza//*Proceedings of the Spring Simulation Multiconference*. Ottawa, Canada, 2008: 556-563
- [32] Siraj A S, Oidman R J, Huber J H, et al. Temperature modulates dengue virus epidemic growth rates through its effects on reproduction numbers and generation intervals. *PLoS Neglected Tropical Diseases*, 2017, 11(7): e0005797
- [33] Baumgartl T, Petzold M, Wunderlich M, et al. In Search of patient zero: Visual analytics of pathogen transmission pathways in hospitals. *IEEE Transactions on Visualization and Computer Graphics*, 2020, 27(2): 711-721
- [34] Kajornkasirat S, Muangprathub J, Rachpibool N, et al. Real-time analytics and visualization: Dengue hemorrhagic fever epidemic applying mobile augmented Reality//*Proceedings of the International Conference on Computational Science and Its Applications*. Melbourne, Australia, 2018: 735-742
- [35] Hockett J. Analyzing influenza incidence and policy data at different temporal and spatial scales//*Proceedings of the IEEE VIS 2016 Poster Program*. Baltimore, USA, 2016(03): 3-4
- [36] Boone K, Swing E. Mapping an epidemic outbreak: Effective analysis and presentation//*Proceedings of the IEEE Conference on Visual Analytics Science and Technology*. Providence, USA, 2011: 305-306
- [37] Parikh R. TVECTOR-Toolkit for visualization of epidemic occurrence, trending, origin and relief//*Proceedings of the IEEE Conference on Visual Analytics Science and Technology*. Providence, USA, 2011: 2-3
- [38] Sun C, Dong W, Bak P, et al. Geo-spatial temporal visual analytics//*Proceedings of the IEEE Conference on Visual Analytics Science and Technology*. Providence, USA, 2011: 1-2
- [39] Hoque E, Setlur V, Tory M, et al. Applying pragmatics principles for interaction with visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 2018, 24(1): 309-318
- [40] Curtis A J. Three-dimensional visualization of cultural clusters in the 1878 yellow fever epidemic of New Orleans. *International Journal of Health Geographics*, 2008, 7(1): 1-10
- [41] Olislagers F, Worring M. The spatiotemporal multivariate hypercube for discovery of patterns in event data//*Proceedings of the IEEE Conference on Visual Analytics Science and Technology*. Washington, USA, 2012: 135-136
- [42] Bertini E, Buchmüller J, Fischer F, et al. Visual analytics of terrorist activities related to epidemics//*Proceedings of the IEEE Conference on Visual Analytics Science and Technology*. Providence, USA, 2011: 329-330
- [43] Malik A, Abdalla R. Mapping the impact of air travelers on the pandemic spread of (H1N1) influenza. *Modeling Earth Systems and Environment*, 2016, 2(2): 1-15
- [44] Bosch H, Thom D, Heimerl F, et al. Scatterblogs2: Real-time monitoring of microblog messages through user-guided filtering. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(12): 2022-2031
- [45] Proulx P, Tandon S, Bodnar A, et al. Avian flu case study with nspace and geotime//*Proceedings of the IEEE Symposium On Visual Analytics Science and Technology*. Baltimore, USA, 2006: 27-34
- [46] Kraak M J, Madzudzo P F. Space time visualization for epidemiological research//*Proceedings of the 23rd international cartographic conference ICC: Cartography for everyone and for you*. International Cartographic Association. Moscow, Russia, 2007: 1-13
- [47] Luo W. Visual analytics of geo-social interaction patterns for epidemic control. *International Journal of Health Geographics*, 2016, 15(1): 1-16
- [48] Astefanoaie A, Bozianu R, Broghammer M, et al. Visual analytics of a pandemic spread//*Proceedings of the Visual*

- Analytics Science and Technology. Salt Lake City, USA, 2010: 277-278
- [49] Muñoz-Medina J E, Garcia-Knight M A, Sanchez-Flores A, et al. Evolutionary analysis of the Chikungunya virus epidemic in Mexico reveals intra-host mutational hotspots in the E1 protein. *PLoS One*, 2018, 13(12): e0209292
- [50] Collins J, Robinson C, Danhof H, et al. Dietary trehalose enhances virulence of epidemic *clostridium difficile*. *Nature*, 2018, 553(7688): 291-294
- [51] Huang M Z, Wang S Y, Wang H, et al. Differences in the intestinal microbiota between uninfected piglets and piglets infected with porcine epidemic diarrhea virus. *PLoS One*, 2018, 13(2): e0192992
- [52] Bryan C, Ma K L, Woodring J. Temporal summary images: An approach to narrative visualization via interactive annotation generation and placement. *IEEE Transactions on Visualization and Computer Graphics*, 2016, 23(1): 511-520
- [53] Bánfi N, Dudás L, Fekete Z, et al. City sentinel-VAST 2011 mini challenge 1 award: "Outstanding integration of computational and visual methods" // *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*. Providence, USA, 2011: 305-306
- [54] Mircea A. Real-time classification, geolocation and interactive visualization of COVID-19 information shared on social media to better understand global developments // *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP*. Edinburgh, UK, 2020: 1-5
- [55] Dou W, Wang X, Kraft T, et al. Identifying topical trends in social media with topic modeling // *Proceedings of IEEE Workshop on Interactive Visual Text Analytics for Decision Making*. Providence, USA, 2011: 1-4
- [56] Tanahashi Y, Wu Y. Analyzing hospitalization records of a pandemic // *Proceedings of the Visual Analytics Science and Technology*. Salt Lake City, USA, 2010: 1-2
- [57] Dou W, Yu L, Wang X, et al. Hierarchical topics: Visually exploring large text collections using topic hierarchies. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(12): 2002-2011
- [58] Leite R A, Schetinger V, Ceneda D, et al. COVIs: Supporting temporal visual analysis of Covid-19 events usable in data-driven journalism // *Proceedings of the IEEE Visualization Conference*. Salt Lake City, USA, 2020: 56-60
- [59] Wang X, Liu S, Chen Y, et al. How ideas flow across multiple social groups // *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*. Baltimore, USA, 2016: 51-60
- [60] Sun G, Tang T, Peng T Q, et al. Socialwave: visual analysis of spatio-temporal diffusion of information on social media. *ACM Transactions on Intelligent Systems and Technology*, 2017, 9(2): 1-23
- [61] Szabó A, Friedman E. Epidemic outbreak visualizer // *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*. Salt Lake City, USA, 2010: 1-2
- [62] Alshaabi T, Arnold M V, Minot J R, et al. How the world's collective attention is being paid to a pandemic: COVID-19 related n-gram time series for 24 languages on Twitter. *Plos One*, 2021, 16(1): e0244476
- [63] Di Stefano B, Fuks H, Lawniczak A T. Object-oriented implementation of CA/LGCA modelling applied to the spread of epidemics // *Proceedings of the Canadian Electrical and Computer Engineering*. Halifax, Canada, 2000: 26-31
- [64] Poletto C, Tizzoni M, Colizza V. Human mobility and time spent at destination: impact on spatial epidemic spreading. *Journal of Theoretical Biology*, 2013, 338(1): 41-58
- [65] Fox S J, Miller J C, Meyers L A. Seasonality in risk of pandemic influenza emergence. *PLoS Computational Biology*, 2017, 13(10): e1005749
- [66] Rao D M, Chernyakhovsky A, Rao V. Modeling and analysis of global epidemiology of avian influenza. *Environmental Modelling & Software*, 2009, 24(1): 124-134
- [67] Grignard A, Taillandier P, Gaudou B, et al. GAMA 1.6: Advancing the art of complex agent-based modeling and simulation // *Proceedings of the International Conference on Principles and Practice of Multi-agent Systems*. Shanghai, China, 2013: 117-131
- [68] Corley C D, Lancaster M J, Brigantic R T, et al. Outside the continental United States international travel and contagion impact quick look tool // *Proceedings of the First ACM SIGSPATIAL International Workshop on Use of GIS in Public Health*. California, USA, 2012: 81-86
- [69] Perez L, Dragicevic S. An agent-based approach for modeling dynamics of contagious disease spread. *International Journal of Health Geographics*, 2009, 8(1): 1-17
- [70] Rao D M. Accelerating parallel agent-based epidemiological simulations // *Proceedings of the 2nd ACM SIGSIM Conference on Principles of Advanced Discrete Simulation*. New York, USA, 2014: 127-138
- [71] Lequime S, Bastide P, Dellicour S, et al. nosoi: A stochastic agent-based transmission chain simulation framework in r. *Methods in Ecology and Evolution*, 2020, 11(8): 1002-1007
- [72] The Washington Post. Why outbreaks like coronavirus spread exponentially, and how to "flatten the curve". <https://www.washingtonpost.com/graphics/2020/world/corona-simulator/>. 2020
- [73] Flowingdata. Simulating an epidemic. <https://flowingdata.com/2020/03/30/simulating-an-epidemic>. 2020
- [74] Anylogic cloud. SEIRS ABM intervene. AnyLogic simulation software. <https://cloud.anylogic.com/mod-el/c23a19eb-5132-48db-9b61-57b444e8a8c0>. 2020
- [75] Ajelli M, Gonçalves B, Balcan D, et al. Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models. *BMC Infectious Diseases*, 2010, 10:190
- [76] Grefenstette J, Brown S T, Rosenfeld R, et al. FRED (A framework for reconstructing epidemic dynamics): An open-source software system for modeling infectious diseases and control strategies using census-based populations. *BMC Public Health*, 2013, 13: 940
- [77] Rizzi R L, Kaizer W L, Rizzi C B, et al. Modeling direct transmission diseases using parallel bitstring agent-based models. *IEEE Transactions on Computational Social Systems*, 2018, 5(4): 1109-1120
- [78] Papparian J, Brown S, Burke D, et al. FRED Navigator: An interactive system for visualizing results from large-scale epidemic simulations // *Proceedings of the 8th International Conference on E-Science*. Chicago, USA, 2012: 1-5

- [79] The New York Times. How much worse the coronavirus could get, in Charts <https://www.nytimes.com/interactive/2020/03/13/opinion/coronavirus-trump-response.html>, 2020
- [80] Flowingdata. Comparing Covid-19 models. <https://flowingdata.com/2020/05/05/comparing-covid-19-models> 2020
- [81] DiPiazza A, Nogales A, Poulton N, et al. Pandemic 2009 H1N1 influenza virus reporter virus reveals broad diversity of MHC class II-positive antigen-bearing cells following infection in vivo. *Scientific Reports*, 2017, 7(1): 1-17
- [82] Flowingdata. Coronavirus life cycle. <https://flowingdata.com/2020/03/11/coronavirus-life-cycle/>. 2020
- [83] Flowingdata. Coronavirus explained and what you should do. <https://flowingdata.com/2020/03/20/coronavirus-explained-and-what-you-should-do/>. 2020
- [84] Ernst P, Mishra A, Anand A, et al. Bionex: A system for biomedical news event exploration//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Tokyo, Japan, 2017: 1277-1280
- [85] Jin sichen, Tao Yubo, Yan Yuyu, Dai Haoran. Analysis of infectious disease model based on multi dimensional spatiotemporal data visualization. *Journal of Computer Aided Design and Graphics*, 2019, 31(2): 241-255 (in Chinese)  
(金思辰, 陶煜波, 严宇宇, 戴浩然. 基于多维时空数据可视化的传染病模式分析. *计算机辅助设计与图形学学报*, 2019, 31(2): 241-255)
- [86] Qiu jiongliang, Wu Wei, Zheng Jianning. Automatic drawing of global infectious disease risk early warning map based on text mining. *China Health Insecticides*, 2020, 26(3): 270-273 (in Chinese)  
(裘炯良, 吴薇, 郑剑宇. 基于文本挖掘的全球传染病风险预警地图计算机自动绘制. *中华卫生杀虫药械*, 2020, 26(3): 270-273)
- [87] Cui W, Zhang X, Wang Y, et al Text-to-viz: Automatic generation of infographics from proportion-related natural language statements. *IEEE Transactions on Visualization and Computer Graphics*, 2019, 26(1): 906-916
- [88] Shi Yaolin. A dynamic stochastic model for the spread of SARS. *Science Bulletin*, 2003, 48(13): 1373-1377 (in Chinese)  
(石耀霖. SARS 传染扩散的动力学随机模型. *科学通报*, 2003, 48(13): 1373-1377)
- [89] Jalayer M, Orsenigo C, Vercellis C. CoV-ABM: A stochastic discrete-event agent-based framework to simulate spatiotemporal dynamics of COVID-19. *arXiv preprint arXiv:2007.13231*, 2020
- [90] Sun G D, Wu Y C, Liang R H, et al. A survey of visual analytics techniques and applications: State-of-the-art research and future challenges. *Journal of Computer Science and Technology*, 2013, 28(5): 852-867
- [91] Xu K, Wang Y, Yang L, Yifang W, Bo Q, Si Q, Yong X, Haidong Z, Huamin Q. H. Clouddet: Interactive visual analysis of anomalous performances in cloud computing systems. *IEEE Transactions on Visualization and Computer Graphics*, 2019, 26(1): 1107-1117
- [92] Song H, Szafir D A. Where's my data? evaluating visualizations with missing data. *IEEE Transactions on Visualization and Computer Graphics*, 2018, 25(1): 914-924
- [93] Yang T, Shen K, He S, et al. Covidnet: To bring the data transparency in era of covid-19. *arXiv preprint arXiv:2005.10948*
- [94] Mniszewski, Susan M, et al. Towards a hybrid agent-based model for mosquito borne disease. *Simulation Series*, 2014, 46(10): 66-73
- [95] Kraak M. The space-time cube revisited from a geovisualization perspective//Proceedings of the International Cartographic Conference. Durban, South Africa, 2003: 10-16
- [96] Chen Z, Wang Y, Wang Q, et al. Towards automated infographic design: Deep learning-based auto-extraction of extensible timeline. *IEEE Transactions on Visualization and Computer Graphics*, 2019, 26(1): 917-926
- [97] Keeling, Matt J, Ken TD Eames. Networks and epidemic models. *Journal of the Royal Society Interface*, 2005, 2(4): 295-307
- [98] OpenKG.CN 新冠开放知识图谱. <http://openkg.cn/group/coronavirus>, 2021
- [99] Wise, Colby, Vassilis N. Ioannidis, Miguel Romero Calvo, et al. COVID-19 knowledge graph: accelerating information retrieval and discovery for scientific literature. *arXiv*, 2020: 1-7
- [100] Wang Xiaoxuan, Zhu Gaopei, Feng Jianing, Xiao Yufei, Sun Na, Xu Xiaoshan, Wang Suzhen. Visualization analysis of hot spot trend of domestic infectious disease prediction model based on CiteSpace. *Modern Preventive Medicine*, 2020, 47(16): 2886-2890 (in Chinese)  
(王晓璇, 朱高培, 冯佳宁, 肖宇飞, 孙娜, 许小珊, 王素珍. 基于 CiteSpace 国内传染病预测模型热点趋势可视化分析. *现代预防医学*, 2020, 47(16): 2886-2890)
- [101] Brauer F. The Kermack-McKendrick epidemic model revisited. *Mathematical Biosciences*, 2005, 198(2): 119-131
- [102] Parker J, Epstein J M. A distributed platform for global-scale agent-based models of disease transmission. *ACM Transactions on Modeling and Computer Simulation*, 2011, 22(1): 1-25
- [103] FRED. A Framework for Reconstructing Epidemiological Dynamics. <https://fred.publichealth.pitt.edu/>. 2020
- [104] El-Assady, Mennatallah, et al. Semantic concept spaces: guided topic model refinement using word-embedding projections. *IEEE Transactions on Visualization and Computer Graphics*, 2019, 26(1): 1001-1011
- [105] Ren Lei, Du Yi, Ma Shuai, Zhang Xiaolong, Dai Guozhong. Overview of big data visual analysis. *Journal of Software*, 2014, 25(9): 1909-1936 (in Chinese)  
(任磊, 杜一, 马帅, 张小龙, 戴国忠. 大数据可视分析综述. *软件学报*, 2014, 25(9): 1909-1936)
- [106] Chen Xiaohui, Xu Li, Ge Lei, Zhang Bing, Che Sen, Liu Haiyan. Review on visual analysis of infectious disease transmission data. *Journal of Computer Aided Design and Graphics*, 2020, 32(10): 1581-1593 (in Chinese)  
(陈晓慧, 徐立, 葛磊, 张兵, 车森, 刘海砚. 传染病传播数据可视分析综述. *计算机辅助设计与图形学学报*, 2020, 32(10): 1581-1593)
- [107] Flowingdata. Coronavirus testing accuracy. <https://flowingdata.com/2020/05/05/coronavirus-testing-accuracy/>, 2020
- [108] The Wall Street Journal. FRIEDMAN T D and D. Battling infectious diseases in the 20th century: the impact of vaccines. <http://graphics.wsj.com/infectious-diseases-and-vaccines> 2020
- [109] Yuan Ning. Regulation of personal information protection in the use of health code. *Legal Review*, 2020, 38(6): 111-121 (in Chinese)  
(园宁. 健康码运用中的个人信息保护规制. *法学评论*, 2020,

38(6): 111-121)

[110] Jin X, Xu K, Jiang P, et al. Virus strain from a mild COVID-19 patient in Hangzhou represents a new trend in SARS-CoV-2 evolution potentially related to Furin cleavage site. *Emerging Microbes & Infections*, 2020, 9(1): 1474-1488

[111] Chen Jiayuan, Shi Jinsong, Qiu Dongan, Liu Chang, Li Xin,

Zhao Qiang, Ruan Ji Shou, Gao Shan. 2019 bioinformatics analysis of New Coronavirus genome. *Bioinformatics*, 2020, 18(02): 96-102 (in Chinese)

(陈嘉源, 施劲松, 丘栋安, 刘畅, 李鑫, 赵强, 阮吉寿, 高山. 2019 新型冠状病毒基因组的生物信息学分析. *生物信息学*, 2020, 18(02): 96-102)



**SUN Guo-Dao**, Ph.D., associate professor. His research interests include big data mining and analysis, information visualization and visual analysis.

**YANG Yu-Fan**, M.S. candidate. Her research interests include big data mining and visual analysis.

**PAN Xiang**, Ph.D., professor. His research interests include pattern recognition, multimedia search, graphics, and image processing.

**LIANG Rong-Hua**, Ph.D., professor. His research interests include medical image visualization, information visualization and visual analysis, and computer vision, network and service computing.

## Background

Under the impact of COVID-19 epidemic, the content of epidemic data is increasing. Facing the outbreak of epidemic, complex epidemiological data, different types of needs and unintelligible contextual knowledge, traditional understanding methods are hard to adapt to the huge data and knowledge. In recent years, the visual analysis method of epidemic data has been popular among scholars due to its comprehensibility and adaptability to experts in many fields.

Based on full investigation and analysis, the categories of epidemic data and the research tasks are summarized. This paper collects common visual analysis methods of

epidemic data, and summarizes them into spatiotemporal visualization, semantic visualization, model visualization, and information graph. Finally, the future research directions and challenges in the field of visual analysis of epidemic data are discussed.

This work was partially support by National Key Research and Development Program of China (No. 2020YFB1707700), National Natural Science Foundation of China (No. 61972356 and 62036009), Fundamental Research Funds for the Provincial Universities of Zhejiang (No. RF-A2020001)