Applied Soft Computing 138 (2023) 110177

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

An analytical tool to support public policies and isolation barriers against SARS-CoV-2 based on mobility patterns and socio-economic aspects

Julio Cezar Soares Silva [a], Diogo Ferreira de Lima Silva [b], Nivan Roberto Ferreira Júnior [a], Adiel Teixeira de Almeida Filho [a,*]

[a] Centro de informática, Universidade Federal de Pernambuco, Recife, Brazil
[b] Production Engineering Department, Universidade Federal Fluminense, Niterói, Brazil

ARTICLE INFO

Article history:
Received 21 October 2021
Received in revised form 23 January 2023
Accepted 3 March 2023
Available online 8 March 2023

Keywords:
Mobility patterns
COVID-19
SARS-CoV-2
Network model
Visualization
Decision support

ABSTRACT

It is crucial to develop spatiotemporal analysis tools to mitigate risks during a pandemic. Many dashboards encountered in the literature do not consider how the geolocation characteristics and travel patterns may influence the spread of the virus. This work brings an interactive tool that is capable of crossing information about mobility patterns, geolocation characteristics and epidemiologic variables. To do so, our system uses a mobility network, generated through anonymized mobile location data, which enables the division of a region into representative clusters. The clusters' aggregated socioeconomic, and epidemiologic indicators can be analyzed through multiple coordinated views. The proposal is to enable users to understand how different locations commute citizens, monitor risk over time, and understand what locations need more assistance, considering different layers of visualization, such as clusters and individual locations. The main novelty is the interactive way to construct the mobility network that defines the social distancing level and the way that risks are managed, since many different geolocation characteristics can be considered and visualized, such as socioeconomic indicators of a location, the economic importance of a set of locations, and the connection of important neighborhoods of a city with other cities. The proposed tool was built and verified by experts assembled to give scientific recommendations to the city administration of Recife, the capital city of Pernambuco. Our analysis shows how a policymaker could use the tool to evaluate different isolation scenarios considering the trade-off between economic activity and contamination risk, where the practical insights can also be used to tighten and relax mitigation measures in other phases of a pandemic.

1. Introduction

The detection of a new coronavirus in 2019, called COVID-19, forced governments to develop strategies to control the spread, such as the creation of social isolation rules, cancellation of flights, border control, and large-scale testing [1–3]. The high spread rate of COVID-19 has prompted governments to seek innovative approaches to fight the outbreak, considering computational epidemiology and big data.

Several articles have studied the dynamics of the disease through mathematical and simulation modeling [4–9], including estimations over the evolution of the curve of infected people, have been used for some governments in public reports. Of course, the uncertainty behind the parameters used in such models should be observed.

When dealing with any mathematical model, the results should be interpreted and validated as a mathematical model, which is a simplified representation of reality. Nevertheless, it is important to have models and results on hand to support decisions, and this can also be supported integrated with iconic models to enable structuring decision problems to create decision opportunities [10]. Furthermore, social infrastructure, cultural aspects, and individual behavior are key elements that may influence the dynamics of disease spread.

To better understand what has worked well and what has not in managing the response and recovering in the pandemic, the OECD elaborated a document on the lessons learned from governments evaluations of COVID-19 responses. They considered evaluations of policies according to phases of the risk management cycle: pandemic preparedness, crisis management, and response and recovery. [11].

* Corresponding author.
E-mail addresses: juliocss@greefo.org.br (J.C.S. Silva), diogofls@id.uff.br (D.F. de Lima Silva), nivan@cin.ufpe.br (N.R. Ferreira Júnior), adielfilho@cin.ufpe.br (A.T. de Almeida Filho).

https://doi.org/10.1016/j.asoc.2023.110177
1568-4946/

Most of the evaluations on response and recovery of 93% of the sample of countries have focused on four main types of policy or measure: economic and financial support, social policies, health policies, and lockdown and restriction measures. Lessons learned about lockdown and restriction measures conclude that they came at a high cost for society, but it is suggested that more efforts should be made to evaluate the costs and benefits associated with these measures, especially lockdowns, which are not always coherent or well implemented [11].

Decisions and studies regarding the recovery of economic activities have been conducted, which requires that decisions for public policies consider the reality of the threat of future waves of COVID-19 spreading or a future pandemic with an analytical perspective, including data-driven approaches [12,13].

We propose an approach to help policymakers in developing more strategic ways to perform lockdowns and restriction measures, such that the cost imposed to the society is minimized. Thus, instead of performing the same lockdown and restriction measures for all the locations in a given geographical region, policymakers could establish different actions to different clusters of locations. Each cluster is defined by locations more probable to suffer from isolation, because they are strongly connected due to frequent travels relative to family, work, and expenditure of services and goods.

In this paper, our goal was to design a visualization tool that integrates the risk-based approach in [13] with an interactive identification, exploration, and simulated restriction of mobility patterns by crossing information on socioeconomic, and epidemiologic variables, to develop isolation strategies for different clusters of locations.

To do this, we first considered anonymized mobile phone location data, provided by In Loco company [14], to obtain risk-based networks that explore the behavior of citizens in a particular location. The data is represented by origin–destination matrices, where an arc between an origin and a destination exists if and only if a certain volume of people, in terms of the percentage of the population of the origin, is commuted. Thus, if a certain *mobility threshold* is achieved, the connection between pairs of locations exists.

By exploring different values of this threshold, experts can investigate the trade-off between the level of isolation imposed and the macro-regions size that should be kept isolated. The limits of the clusters are represented geographically as isolation barriers. These isolation barriers support policymakers in differentiating strongly connected sets of locations. Each isolated region receives resources depending on the magnitude of its contamination risks.

Also, these isolation barriers are used to, for example, monitor the situation of the pandemic and identify the impact of the pandemic on different economic arrangements. Depending on the situation of the pandemic and the region under study, policymakers may impose sanitary barriers, travel restrictions, costs or other restriction measures to reduce the number of people traveling between locations that belong to different clusters.

The system uses multiple coordinated views [15] so that policymakers can understand the characteristics of these regions concerning indicators related to the population's socioeconomic aspects, the risk of infection, mobility patterns, and economic arrangements.

The case study included in this paper was made for the state of Pernambuco (PE), in Brazil. PE is one of the country's most popular states and has been primarily affected by the ongoing pandemic. The socioeconomic diversity found among the cities of Pernambuco also brings an essential demand for the present work, which can be replicated in other countries.

To summarize, the contributions of this work are:

- The proposed approach is able to cross mobility information with epidemiologic data and characteristics about the studied locations.
- The tool can be used in the first phase of a pandemic, to verify the impact of lockdowns on risk metrics, and in the subsequent phases to understand how the tightening and relaxing restriction measures can impact a neighborhood or municipality according to their characteristics.
- Case studies were performed to demonstrate the importance of the tool in monitoring risk metrics, evolution of the risk metrics, investigation of the impact in important economic arrangements, identify neighborhoods that most commute citizens from different cities, and how to prioritize mitigation actions in neighborhoods according to their socioeconomic variables.
- The visual decision support tool could provide policymakers with relevant information, helping them to better understand the pandemic situation, to learn how to partition the studied region using mobility data, and to support decision-making.
- This exploratory tool enables insights to enhance compartment epidemic models by including socioeconomic variables whose effects may be visually observed.
- It is possible to adapt and use the proposed visual analytics tool for any region given the availability of mobility, socioeconomic, and epidemiologic data. The source code can be accessed in https://github.com/nivan/covidClusters. The README file contains explanations on how to adapt the tool for the user's region. To interact with the tool, go to https://nivan.github.io/covidClusters/bairrosRecife/ and https://nivan.github.io/covidClusters/municipiosPE/.

The rest of the paper is organized as follows. Section 2 present related visual analytics studies for mobility data and networks in urban systems, in epidemic analysis and in the context of COVID-19 pandemic. Section 3 presents materials and methods used in this work. In Section 4, we detail the proposed visual analytics tool. The experimented use cases and validation with domain experts are presented in Section 5. An overall discussion about the findings concerning the proposed tool is presented in Section 6. Finally, Section 7 concludes the paper and presents perspectives for future work.

2. Related work

Our work comprises two fields of research. The first is related to visual analytics tools based on networks and mobility data, and the second is visual analytics tools for epidemic analysis and COVID-19.

2.1. Visual analytics based on mobility data and networks for urban dynamics

Analysis of massive data produced by different technological sources, such as mobility data, can be performed through visual analysis tools. This approach has been adopted by some researchers to transform data into robust information for the final user [16,17]. Visual analytics tools may incorporate mobility patterns to achieve a better analysis of urban dynamics.

These visual analytics tools can extend the comprehension about different aspects of mobility [18], by providing visual insights to support decisions relative to urban systems, such as public transportation [19–21], traffic analysis [22], tax evasion [23] and smart cities applications [17,24]. The study of mobility data brings challenges due to complex spatiotemporal patterns, which can be analyzed by the visual tools that present adequate mobility representations [25].

Network science has been used in different contexts where visual analytics tools are applied. Networks are widely applied to provide insights in the context of social networks [26–28]. Real-world situations can be represented more precisely by weighted directed networks, which contain more information about the relationship between pairs of nodes [29]. Thus, it enables us to characterize terrestrial transportation route networks [30], model and visualize scientific datasets [31], and to analyze patient flow in hospitals [32].

Urban mobility data can be combined with networks to study the relationship among neighborhoods. This analysis can include the detection of communities, which are common mobility patterns in an urban environment. Krueger et al. [33] proposed a visual interactive tool to provide insights for urban planning research, using mobility graphs and unsupervised learning to detect mobility patterns and behavior anomalies.

An interactive visual analytics tool was developed in [34] to enable the detection and changes in spatial–temporal patterns, and also provide clutter reduction when comparing flows of people in certain moments, by combining spatial aggregation of flows and temporal clustering. Wang et al. [35] used geo-textual data, which embodies geographical and human activity information, to generate sequential latent graphs and applied the Louvain algorithm for community detection method to identify and analyze different types of region transformations.

The works found in the literature use automated clustering approaches to detect and analyze mobility patterns. Also, these works did not consider interactive approaches to simulate mobility restriction, which makes it impossible to study the impact of relaxation or tightening of social distancing. In this work, an interactive clustering approach is used in a visual analytics tool to obtain constrained and relaxed mobility patterns using networks generated through anonymized mobile phone data.

2.2. Visual analytics for epidemic analysis and the COVID-19 outbreak

Data visualization encompasses essential tools to aid policymakers during an epidemic, offering statistical analysis, surveillance, and indicator monitoring. The tools are developed such that human cognitive effort is reduced and the capacity to analyze big datasets is augmented. Unified global visualization technologies can provide insights to support decisions to manage an epidemic, and consequently minimize its propagation [36,37].

The devastating power of COVID-19 became more evident as the pandemic advanced [38]. Therefore, several governments and institutions have used visualizations in their reports on the COVID-19 spread, including economic and labor factors [39–41]. Many challenges for deploying solutions for COVID-19 management and control arose, such as the development of problem-oriented big data acquisition systems and inference of population behavior [42].

Open information about the geographical distribution of COVID-19 is important to support scientific analysis and to provide a better understanding of the spread of COVID-19 [43]. A good example is a dashboard provided by John Hopkins University [44], which includes heat maps concerning the active number of cases, cumulative cases, hospitalizations, and incidence rates. The World Health Organization [45] also provides an interesting dashboard that includes specific visualizations regarding statistics and region comparisons.

Furthermore, articles exploring data visualization have been published. Dey et al. [39] present an exploratory data analysis of the pandemic with visualizations of the spreading of COVID-19 in China. Gao et al. [40] propose the use of cartograms to visualize the expansion and spread of COVID-19. Barone et al. [46] created a surveillance dashboard for COVID-19 containing graphs, figures, and criticality tables with data collected from the European Centre For Disease Prevention and Control. Zhang et al. [47] used a spread index and an extinction index to investigate when the pandemic grows and declines, and also used bubble charts and animations to visualize how the pandemic is changing over time.

Khan et al. [48] designed an ontology-based infrastructure that captures relationships between data streams, visualization functions, and web pages to allow the rapid deployment of web-based interactive dashboards across multiple data streams, in a context with limited resources and low programming expertise. Ipenza et al. [49] developed a public visual analytics tool, named QDSSUS, to interactively explore millions of Brazilian healthcare records through interactive queries. The authors discovered how patterns evolved over one year of COVID-19 in different regions of Brazil, such as the relationship between patient age groups and their corresponding dominant symptoms.

A survey on the use of visual analytics for public health was presented by Preim and Lawonn [50], which included research about computational epidemiology. The survey mentions that visualizations containing spatiotemporal data are often used to describe the disease's behavior. This is because epidemic spread simulation based on the analysis of connections between different locations or individuals over time can output more realistic results in what–if scenario analysis.

A view of epidemic modeling through networks is given by [51]. The use of social network analysis and visualization [51,52] is very important as an epidemic spread often depends on mobility patterns among locations in the studied region. Tao et al. [53] present a visual analytics framework, called HoNVis, which can be used to explore epidemic outbreak scenarios through domestic and international travels, identifying airports of interest and tracing the propagation of an epidemic.

Luo [54] proposed a visual analytics tool to evaluate the effect of control measures using face-to-face interaction patterns of a primary school. Geo-social mixing patterns are explored from human interaction network data, which enables the identification of critical individuals, locations, and clusters of locations. Then, a control measure can be designed, considering the interaction patterns, and evaluated through a compartment model that simulates the propagation of an epidemic.

Dynamic Network Visualization (DyNetVis) [55–57] was proposed to be an open-source visual analytics tool for dynamic network exploratory visualization. DyNetVis also implements dynamic processes, including epidemic compartment models. It is possible to analyze and explore the infection path of simulated epidemics with different parameters, visualizing the transmission tree and different groups of individuals (i.e. susceptible, infected, and recovered).

Some scientific visualization tools that adopted network analysis were proposed in the COVID-19 context, Abel and Gietel-Basten [58] proposed a tool to analyze the global economy interconnectedness, and thus investigate the economic and social impact of COVID-19, based on a chord diagram.

Saraswathi et al. [59] used software tools, Cytoscape and Gephi, to create social network visualizations and explore common transmission patterns. The visualization approach enabled the identification of evolving hotspots, such as those associated with international travel and principal cities. Also and key actors were identified in the transmission, by considering age-sex attributes in the nodes of the transmission network. The exploratory analysis occurred during different phases of the lockdown in India.

Luo et al. [60] used a visualization approach to construct disease transmission network graphs for the COVID-19 epidemic, according to relationships among cases. The proposed visualization tool for transmission networks can identify the transmission source and contacts, assess the current situation of transmission and prevention, and provide an adequate response to control the propagation of COVID-19.

It is possible to see that past studies for COVID-19 dashboards considered exploratory analysis and surveillance of epidemiologic variables and automated development generation of visualization components. Also, works that proposed graph-based visualization tools for epidemics and COVID-19 were mainly focused on exploratory analysis of the transmission network, the visualization of simulation output from compartment models, and analysis of economic impact based on the degree of connection between different locations.

To the best of our knowledge, this is the first tool that uses mobility patterns, generated from millions of anonymized mobile phones, to manage the prioritization of risk mitigation actions to a set of locations. The insights provided to mitigate risk across a set of locations is obtained by crossing mobility, socioeconomic and epidemiologic data. Thus, it is possible to interactively define the restriction or relaxation of common mobility patterns, defined by clusters, given the values of risk contamination and impact on socioeconomic variables, observe the impact of the pandemic in economic arrangements associated with the clusters, and monitor the clusters' (and its components) contamination risk over time.

3. Materials and methods

The authors of this study were part of a larger institutional project of the Recife's City Administration, whose goal was to provide data-driven scientific recommendations to the administration on how to fight the pandemic and also plan the recovery process.

Works that propose novel visual analytics tools commonly invite domain experts' to validate the proposed tool and also use this feedback to develop increments in future work [49,61,62]. In order to guide the formulation of use cases in the experiments, it is important to define system requirements and analytical tasks. The proposed tool and its supported analysis were developed following an adaptive and evolutionary approach [63] with feedback from experts with an interdisciplinary background.

The group of domain experts included (2) Epidemiologists, (2) Economists, (1) Architect and (3) Engineers. All of the broad experience in their areas (all of them PhD. and/or Masters degrees) and participated in regular meetings to provide quick responses to the administration's needs. Thus, it was possible to validate the tool and obtain important feedback about its usefulness.

As a result of this effort, it was identified the need to study strategies for setting isolation barriers and try to minimize the impact of these barriers on the population and local economy. In a closer collaboration with the three Engineers in our team (co-authors of this paper), we came up with the methodology of the analysis illustrated in Fig. 1 and described in the following.

Phase 1 represents the data collection process. The data used in this paper include socioeconomic indicators and COVID-19 statistics regarding cities and neighborhoods made available by Recife City Hall and the Secretary for Planning and Management of Pernambuco (SEPLAG) from the Pernambuco State Government.

Also, mobility data collected from mobile devices were supplied by In Loco, a company that works with location analytics. Mobility data concerns anonymized location data from millions of devices that were transiting in Pernambuco in the first half of 2020.

Phase 2 points out the need for database cleaning, processing and integration. There are three main databases to be processed: boundaries, mobility, and socioeconomic. The boundaries database contains the polygons that are used as basic types of the

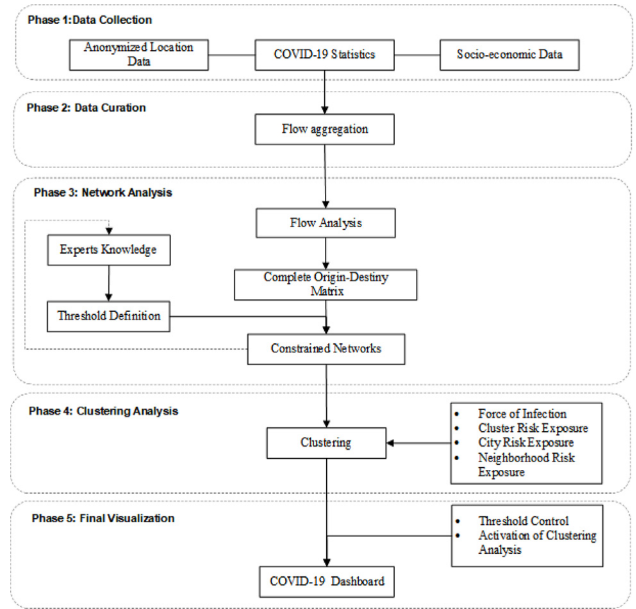


Fig. 1. Methodology used in the data analysis process used in this work. This process is divided into five phases, initializing with data collection and as a result, we get the dashboard.

leaflet.js map. The integration of these three databases is guided by the locations' names. The mobility data is constituted by the aggregation of daily origin–destination matrices. More details on data collection and data curation are given in Section 3.1.

Phase 3 concerns the flow analysis and the construction of a constrained graph. First, it is possible to perform different analyses on flow data, such as to study mobility behavior over time in each location. Secondly, the mobility data is transformed into a single representative origin–destination matrix (see Section 3.1). Then, thresholds defined by using experts' knowledge generate filtered graphs in a continuous refinement process, as illustrated by the dashed arrow. This definition concerns the division of macro-regions or clusters to be managed.

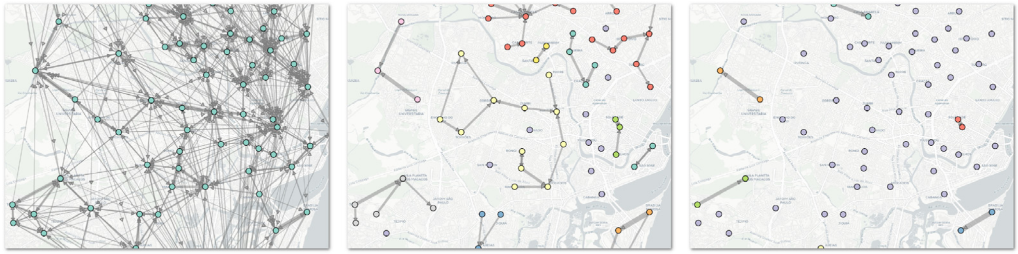
Managing many clusters may reduce interpretability due to high granularity. A small number of clusters will not divide the region to optimize resource allocation in planning mobility restrictions or economic recovery. Thus, the policymaker can define which is the adequate number of regions to be managed according to its preferences and its understanding of how connections in each cluster reflect the reality of the state or city under study. More details on Phase 3 is given in 3.2.

In **Phase 4**, risk metrics were defined to prioritize actions for two types of clusters: healthy and infected. Healthy clusters need to be ranked according to their risk of getting infected. Thus, risk exposure metrics were defined to model this risk. Infected clusters can be ranked according to their infection risk or infection rate. Thus, we adopted a risk metric to model the infection risk in this type of cluster. More details on Phase 4 is given in 3.2.

Finally, in **Phase 5**, a COVID-19 dashboard is constructed to visualize the analysis defined in the other phases. It was implemented using javascript, with the Leaflet.js and Turf.js libraries for the geographical visualizations and D3.js for data-driven multiple coordinated views. We discuss the requirements for this dashboard in Section 3.3.

3.1. Data used in the experiments

The mobility data was provided by the In Loco company, which collects anonymized location data from about 60 million



(a) $l = 0.01$ (b) $l = 0.05$ (c) $l = 0.10$

Fig. 2. Clusters of neighborhoods in the city of Recife corresponding to different values of the mobility threshold, *l*. The colors on the vertices of the networks correspond to the different clusters. Purple nodes represent singleton clusters. It can be observed that as the mobility threshold grows, the connectedness among vertices decreases forming "isolated islands" or clusters.

devices around the world, from which about 2.2 million devices' data has been collected in Pernambuco, which is considered an entirely reasonable number in light of the entire population of Pernambuco, found about 8.8 million in the last census in 2010 and recently estimated in 9.5 million.

From the provided Origin–Destination (OD) Matrices that estimate the flow amongst the cities of Pernambuco, Metropolitan Region of Recife, and Recife's neighborhoods from 2020-05-06 to 2020-05-26. The entries of the OD Matrix represent outflows from each city *i* to a city *j*. Thus it was calculated the percentage of people that remained in the city *i* and the percentage that left the city *i* to each of the other $N - 1$ cities at each date of the considered period. Given an OD matrix A, if a_{ij} = 0.2, then a 20% flow was estimated of population moving from city *i* to city *j*.

In this study, we considered the worst-case scenarios of the COVID-19 spread by estimating the worst-case mobility matrix for the studied period. Thus, given that a matrix A^t is the OD matrix for period *t*, the non-normalized representative matrix of the worst-case mobility for a period $t = 1, 2, \ldots, T$ scenario is given by $W'_{i,j} = \max_t(A^t_{i,j})$, where *T* is the final day available in the dataset. The worst-case mobility matrix is given by $W_{i,j} = W'_{i,j}/\sum_{j=1}^{N} W'_{i,j}$.

Furthermore, data concerning the COVID-19 spread in Pernambuco was collected from the Brazilian Ministry of Heathy reports [64], and the SEPLAG of Pernambuco [65]. Also, the socioeconomic variables of the cities were collected from IBGE [66]. A summary of the spatial variables for Recife neighborhoods and the group from which they belong is presented in Table 1

Table 1
Socioeconomic and epidemiologic variables that were collected and calculated for Recife neighborhoods.

Name	group
Probability of survival 60 years	Socioeconomic
Per capita income	Socioeconomic
Percentage of employed in the agricultural sector	Socioeconomic
Percentage of employed in the commerce sector	Socioeconomic
Percentage of employed in the construction sector	Socioeconomic
Percentage of employed in the extractive sector	Socioeconomic
Percentage of employed in the service sector	Socioeconomic
Percentage of employed in the industrial sector	Socioeconomic
Percentage of employed in the transformation ind. sector	Socioeconomic
Percentage of vulnerable households dependent on elderly	Socioeconomic
Population in vulnerable households dependent on elderly	Socioeconomic
Percentage of population in bedrooms with 2 or more people	Socioeconomic
HDI (Human Development Index)	Socioeconomic
HDI education	Socioeconomic
HDI longevity	Socioeconomic
HDI income	Socioeconomic
Active cases	Epidemiologic
Active cases per capita	Epidemiologic
Force of infection per capita	Epidemiologic
Risk exposure per capita	Epidemiologic

3.2. Networks and clustering analysis

This section presents the necessary concepts from network analysis used to obtain the results in this paper. Also, we provided some sources that can be used by the interested reader to explore more measures and definitions concerning the topics presented.

Let $D = (V, A)$, be a directed graph, or digraph, where V represents a set of vertices and A (disjoint from V) consists of a set of directed arcs together with an incidence function ϕ_D that associates each arc of D with an ordered pair of vertices [67].

In our analysis, we use directed graphs to represent mobility patterns where each node, v_i, represents a geographical location (*e.g.*, a district, a city, etc.) and each arc, (i, j), represents aggregated movement of one or more people from location v_i to location v_j.

Furthermore, for each arc $a = (i, j)$, we define the weight of an arc, $w(i, j)$, as the proportion of the population of each location *i* that travels region *j*. When considering the mobile location data, we collected the origin–destination matrix for each day from January to the date after social distance policies. Thus, we considered two possibilities for building the network scenarios: (i) the maximum flow observed before social distance policies and (ii) the maximum flow after social distance policies.

The maximum value was used to ensure a conservative risk approach, although other values such as the average could be used. These weights shall be considered according to the current social distance state. Then, the digraph, together with its arcs' weights, is called a weighted directed graph (D, w).

Our analysis is based on a threshold, *l*, used to filter the arcs in a weighted directed graph to build derived undirected graphs as described in the following. An arc between two vertices $v_i, v_j \in V$ exists (i.e., passes the filter) if and only if, $w_{i,j} \geq l$ or $w_{j,i} \geq l$.

In other words, this threshold represents the fractional volume of people, which implies a well-established communication amongst different locations, therefore indicating that there is a proportion of $w_{i,j} \geq l$ of people that travel from city *i* to *j*.

Fig. 2 presents the directed graphs obtained by choosing three different threshold values in a graph representing the mobility in the city of Recife, where the arc weights were obtained from the anonymized mobile phone location data described in Section 3.1.

The algorithm to generate clusters with a threshold value l is shown in Algorithm 1.

Algorithm 1 Pseudocode of the clustering method

```
    Input: W, Nodes, l
    Output: Partition U
 1: Edges ← {(i, j)|W_ij ≥ l, ∀i, j ∈ Nodes}
 2: g ← initGraph()
 3: g.vertices ← Nodes
 4: g.edges ← Edges
 5: notCovered ← [Nodes]
 6: U ← []
 7: while length(notCovered) > 0 do
 8:     v ← notCovered[0]
 9:     clusterExists ← False
10:     s ← []
11:     for u = 1 : length(U) do
12:         if |g.neighbors(v) ∩ U[u]| > 0 then
13:             clusterExists ← True
14:             S.push(u)
15:         end if
16:     end for
17:     if clusterExists then
18:         S.reverse()
19:         C ← {}
20:         for z = 1 : length(S) do
21:             C ← C ∪ U[S[z]]
22:             U.pop(S[z])
23:         end for
24:         U.push({C ∪ g.neighbors(v) ∪ {v}})
25:     else
26:         U.push({g.neighbors(v) ∪ {v}})
27:     end if
28:     notCovered.pop(0)
29: end while
```

The inputs are an origin–destination matrix W, a set of Nodes, and the mobility threshold l. The first step is to create a graph filtered by the mobility threshold l, as shown in lines 1–4. In line 5 we create a list of nodes to be covered by at least one cluster. The partition is initialized in line 6. The loop from lines 7–29 evaluates the cluster from which a node v and its neighbors belong.

In line 8 a node that was not evaluated is selected. In lines 9–16 we evaluate if the neighbors of v belong to at least one existing cluster. If it is true that the neighbors of v belong to at least one existing cluster, we identify the indexes of the clusters with a list S. Lines 17–24 merge the clusters that contains some neighbors of v. Note that the list S, which stores the index of the clusters to be merged, is reversed in line 18 to guarantee that the highest indexes from the list U will be removed first.

The clusters saved in S are merged into a new cluster C and are also deleted from U in Lines 18–22. In line 24, a v and its neighbors will be added to the merged cluster C to constitute a new cluster in U. As can be seen in Lines 25–27, if the neighbors of v do not belong to any cluster, a new cluster is created in U, containing v and its neighbors. In line 28, v is deleted from the set of nodes that are being evaluated. It can be observed that the sensitivity of the clustering method depends on the filtered graph, which depends only on the mobility threshold l.

We can see that the connection between pairs of locations in the considered region is sensitive to the threshold l. As the threshold is reduced, fewer locations will be connected, and clusters will begin to appear. These clusters are unconnected graphs, and each cluster is associated with a set of cities that commute at least a proportion l of people between them.

Common approaches that automatically cluster graphs mainly optimize a quality function, such as modularity [28,29,68]. This was not adequate for our context, since it is interesting for the policymaker to understand how isolating and relaxing isolation measures may affect the community and the contamination risk within a community.

The proposed clustering approach is the most adequate for our study since the only criteria needed to define connections between locations is the proportion of people commuting from one location to another, respecting the most probable connections, when l is increased or decreased, according to real mobility data. It is possible to obtain the most probable mobility patterns of a certain region and the impact of constraining and relaxing mobility in the risk metrics through this interactive clustering approach.

The adopted clustering approach represents mobility patterns that are more probable to occur according to the value of l. These mobility patterns are well-established connections between locations associated with frequent travels relative to family, work, and expenditure of services and goods.

Although singletons will exist as a result of the clustering, they will not contain representative mobility patterns relative to the region under study, since it is possible to define l to cover a significant part of the population of the region (state or city) under study. Thus, one can consider generating a set of clusters that cover at least 90% of the population of the region under study.

Different mobility patterns imply different cluster configurations which will impact the contact between people. Thus, contamination risk metrics will be affected depending on the clusters' configuration. The division of the region into clusters is important to allocate resources according to the risk of each cluster.

Therefore, this threshold implies a risk exposure considering the connectedness among cities. It is possible to see that this clustering approach makes it possible to configure clusters such that they represent real mobility patterns between locations in a region and to constrain or relax these connections to reduce contamination risk.

To make the importance of this approach clear, we can compare it with other common clustering methods. Clustering methods such as K-means generate a partition from a set of training examples evaluated in a set of observable features. The clustering method presented in Algorithm 1 generates a partition from a graph. It tries to agglomerate locations in different groups that share a well-established connection according to l.

K-means will cluster training examples according to their similarity in a feature set. Thus, if anonymized data from cell phones are not available, K-means can be used to generate clusters of locations when it is possible to collect other datasets and extract features that can be processed with this clustering method. Of course, K-means will not provide the same level of interactiveness as our approach, since the user is requested to choose the number of clusters, which will be automatically obtained through the considered feature set and the quality function.

We can now define risk measures for each cluster that will summarize the infection's possible spread based on the current state of the pandemic and the mobility data. These measures can help policymakers to identify and prioritize those regions that are critical. There are two types of regions to be prioritized: infected and healthy. First, we will present metrics that support plans for economic recovery. We defined the Risk Exposure Measure of a region, R_{v_j} by the following equation:

$$R_{v_j} = \sum_{i=1}^{n} w_{ij} \times p_i \tag{1}$$

where p_i is the likelihood of one individual from v_i having the infection. This equation quantifies the risk of infection in a healthy region. In other words, it represents the risk of a healthy region becoming an infected region. In Eq. (1), it was considered that the contamination within a region and its people moving to distinct regions are independent events.

The Cluster Risk Exposure Measure, $R_{C_k^0}$, is given by:

$$R_{C_k^0} = \sum_{v_j \in C_k^0} R_{v_j} \tag{2}$$

A conservative approach has been assumed for both risk exposure metrics R_{v_j} and $R_{C_k^0}$, thus including the redundant intersection of events, i.e., being contaminated by more than one city simultaneously. These risk metrics support the prioritization of regions where progressive economic recovery is possible. Thus, one may prioritize locations or clusters with small risk exposure to develop economic recovery plans.

The final risk measure is calculated for singletons or clusters that are infectious (contain at least one active case). This metric is important to rank clusters according to their infection rate. If a cluster's infection rate is high, more efforts are required to mitigate the risk of contamination by defining mobility restrictions. The name of the measure is the Force of Infection (FOI), and it represents the rate at which a single individual contracts a disease [69]. FOI is defined according to Eq. (3).

$$\lambda_{C_k^1} = \beta \frac{Y_{v_j}}{N_{C_k^1}} \tag{3}$$

where C_k^1 is an infected location, Y_{v_j} is the number of infected individuals from a city $v_j \in C_k^1$ and $N_{C_k^1}$ is the population size of C_k^1. $\beta = c \times t$, where t is the transmission probability and c models the contact rate. Similarly, a more conservative analysis was considered and, therefore, the transmission probability was settled to 100%.

The proxy adopted to represent the contact rate was the mean strength of the cluster [70]. We can see that the risk is reduced when the contact rate is reduced, where this reduction is directly associated with mobility restriction measures, considering the limits of isolation barriers.

The proposed approach enables a way to protect and isolate uncontaminated areas for reducing local economic losses since it allows policymakers to choose which weak mobility links shall be interrupted with risk-informed decision support. Using a simple approach for clustering the neighborhoods brings the advantage of easy understanding and usability by policymakers and epidemiologists. This feature is also an advantage for enhancing communication with society to better support isolation plans.

3.3. System requirements and analytical tasks

In our discussions with the domain experts in our team, we were able to identify four main requirements for the visualization tool developed:

R1 — Visualization of multiple data sources: The analysis of isolation barriers requires the investigation of multiple data sources: mobility networks, COVID-19, and socioeconomic indicators. One particularly important aspect of the analysis is how these sources interact to define the properties of the clusters (defined in Section 3.2).

R2 — Consider multiple scenarios: Mobility restriction policies, associated with the defined isolation barriers have big consequences on the daily activities of the society. For this reason, it is essential to consider the multiple levels of restriction. It will also enable decision-makers to adapt this policy for different locations and also to different moments of the pandemic development.

R3 — Simplicity: Designing policies for such an urgent pandemic situation requires the consideration and discussion with many stakeholders with a wide variety of backgrounds (outside the technical team). Thus, the visualizations designed should be simple and easy to discuss with all the people involved in the decision-making process.

R4 — Easy to share: To quickly promote discussion of policies associated with isolation barriers, it is important that the tool can be easily shared with the stakeholders and the city administration. Given that this is a general theme in the current pandemic situation, this also promotes a more extensive and quicker idea propagation and possible replication in other locations.

4. System design

We now discuss our tool's final design and how it satisfies the requirements presented in Section 3.3. The overall strategy concerning the visualizations used in the entire tool's design was to provide widely used and straightforward visual summaries (R3). The main interface, shown in Fig. 3, is comprised of four components (R1) described in the following:

The Controls widget (Fig. 3(a)) enables the user to control the parameters of the analysis. The mobility threshold (*Limiar de mobilidade*) slider sets the value of the threshold used to filter out arcs in the graph and thus to control the cluster generation following the definitions of Section 3.2 (R2). The other controls allow the users to show/hide different graphical layers presented on the map and to configure which data variable is presented in the visualization.

The set of checkboxes allows the user to show/hide the different layers on the map. The "Color by" (*Colorir por*) dropdown enables the selection of one of the socioeconomic and pandemic indicators (e.g., active cases, the population in vulnerable conditions, etc.) that are visualized as a choropleth map in the map widget (R1).

The Map widget (Fig. 3(b)) presents the geographical context of the analysis. The different layers on the map present the regions being considered (and how they cluster together) and the mobility network built using the mobility threshold. More specifically, the socioeconomic and pandemic data corresponding to the geographical regions being studied (e.g., neighborhoods, cities, etc.) are shown as choropleth maps (e.g., Figs. 5 and 9).

Also, as shown in Figs. 3(b) and 4, the colors of the regions can also be used to represent the clusters of regions (regions with the same color belong to the same cluster). In this mode, singleton clusters are represented in purple color. This allows us to represent isolated regions.

We also use the convex hull of the set of vertices to represent the clusters. These polygons are colored either red or blue if they are infected (have active COVID-19 cases in at least one of its member regions) or not-infected (no active cases in the cluster), respectively. This alternative representation of the clusters allows us to overlay the cluster representation with other data layers.

The mobility network is presented as a node-link diagram, where nodes are placed at the center of the corresponding regions. Nodes are colored according to the cluster they belong to, similarly to the regions described above (see also Fig. 2).

By clicking on the elements of one of the layers, the system presents a tool-tip with detailed information such as population, active cases, etc. (as shown in Fig. 3(b)). All the layers on the map can be set to be shown or hidden using the checkboxes in the controls widget.

The **Bar chart widget** (Fig. 3(c)) and **Scatterplot widget** (Fig. 3(d)) provide quantitative information on the administrative

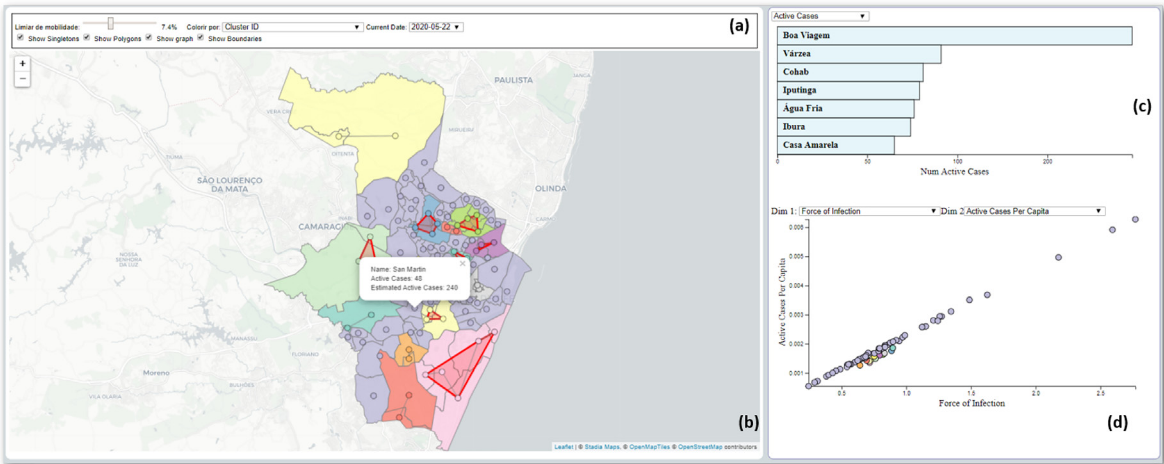


Fig. 3. Graphical interface of our tool showing the exploration of neighborhoods in the city of Recife. The Controls widget (a) is used to configure the visualization and to set the mobility threshold. The Map widget (b) shows the geographical context of the different layers of data. Finally, the Bar chart widget (c) and the Scatterplot widget (d) show quantitative information about the regions and clusters, respectively.

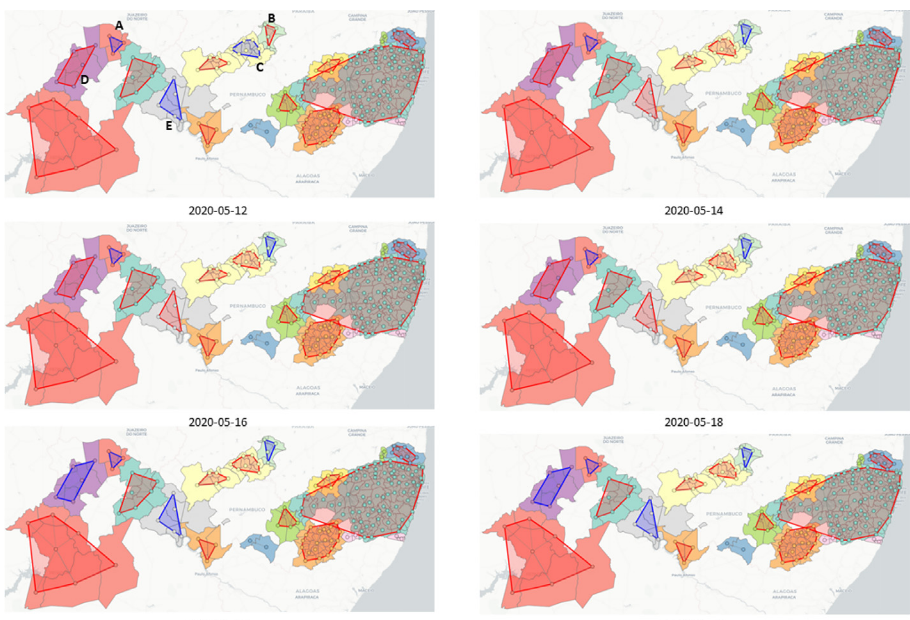


Fig. 4. The evolution of the clusters' pandemic state over time, where the clusters were generated with a mobility threshold of $l = 0.025$. The color of the polygons corresponds to the clusters' classification. The overlaid polygons show the 'infected' (red) and the 'not infected' (blue) clusters. Notice that some clusters had state transitions between infected and not infected.

regions and clusters, respectively. Notice that the dots on the scatterplot are colored with the color of the corresponding cluster. Similar to the controls widget, the dropdown menus control the indicators shown in the corresponding plots.

All the graphical components described in our system are linked, i.e., selecting an element in one of them highlights the corresponding element in the other widgets. For example, when the user clicks on one of the polygons representing region boundaries on the map, the dot corresponding to the cluster containing this region is highlighted. On the other hand, clicking a dot on the scatterplot caused the tool-tip corresponding to the selected cluster to open on the map widget.

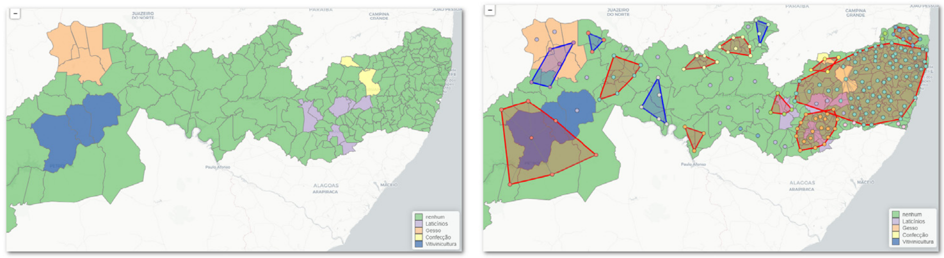
The tool includes data from Pernambuco's municipalities and Recife's neighborhood. Recife is the city of Pernambuco with the largest population of the state (1.6 million) and the most significant demographic density (7,039.64 $inhabit/km^2$). Most of the state's public and private health infrastructure is located in the Recife area. Therefore, it shall be difficult to control the epidemic once infected persons from other Pernambuco cities shall be medical treatment in Recife.

In light of these local features, the system considers Recife and its neighborhoods in detail and a general view of the entire state as presented with its components presented in Fig. 3.

Implementation. Our tool is implemented as a web-based dashboard (R4). It was implemented using the Leaflet.js and Turf.js libraries for the geographical visualizations and D3.js for the other components.

5. Use cases and results

This section presents the results on the usefulness and features of the proposed system. To this end, we use data from different sources (described in Section 3.1). We highlight that the use



(a) (b)

Fig. 5. (a) Local productive arrangements colored in the map of the state of Pernambuco. Notice that there are islands of production (non-green regions). (b) The clusters of cities generated from a mobility threshold equal to $l = 0.025$. We can evaluate the pandemic situation in the context of local productive arrangements on 2020-05-26 observing the clusters that surround them. As can be seen, some of the local productive regions could have been isolated by imposing the isolation barriers through travel restrictions or other mobility restriction measures.

cases described were carried out by three domain experts (co-authors of the paper). The tool and its use cases were presented to policymakers and other domain experts of Project D.A.D.O, which validated the tool and supplied important feedback.

5.1. Pernambuco networks: State scale

After setting up the desired threshold using the mobility threshold slider, the user can monitor the clusters' states (infected or not infected) over time. Fig. 4 illustrates how clusters generated for $l = 0.025$ behaved in the period from 2020-05-12 until 2020-05-26 and included cities from other states bordering with Pernambuco that have relevant people flow.

We highlighted some clusters on the first day of analysis, manually associating a letter for each of them. Cluster A did not become infected during the considered period. Therefore, the policymaker may judge if this cluster is a candidate for a specific isolation barrier policy associated with an economic recovery plan or how to relax the isolation policy after observing a tolerance time.

On the other hand, the states of the clusters B, C, D, and E changed in the considered period. Cluster C became infected while Cluster B became not infected on 2020-05-14. Cluster D became not infected on 2020-05-22. Cluster E became infected on 2020-05-14 and then became not infected on 2020-05-22. It is important to highlight that during this period since the tool presented in this paper was not available, no condition-based policy was planned for these clusters of cities, so we expect that this proposed graphic tool may be useful for preserving the population of areas that are not infected.

Another analysis involving the selected cluster set is to investigate the situation of some local productive arrangements (LPAs), which constitute the state's strategic economic regions. These types of economic clusters are supported by government programs that promote regional development by stimulating the enterprises' technological development, competitiveness, and sustainability. The analysis is shown in Fig. 5.

Areas in green are not contained in any cluster. Purple areas represent dairy's LPA, pale orange areas represent plaster's LPA, yellow areas represent clothing production's LPA, and blue areas represent viticulture's LPA. On 2020-05-26, pale's LPA was the only one not infected. This shows that the majority of the considered strategic economic regions are infected, and policymakers should give special attention and develop a plan of action and monitoring for these areas.

5.2. Recife metropolitan area

An analysis including only the Recife metropolitan area gives more details about this region. In this scenario, instead of considering Recife as a single vertex, each of its neighborhoods was considered in the graph as a vertex to reveal these neighborhoods' relations with the cities of the Recife metropolitan area. Fig. 6 presents the clusters constructed with a threshold of $l = 0.05$, including singletons.

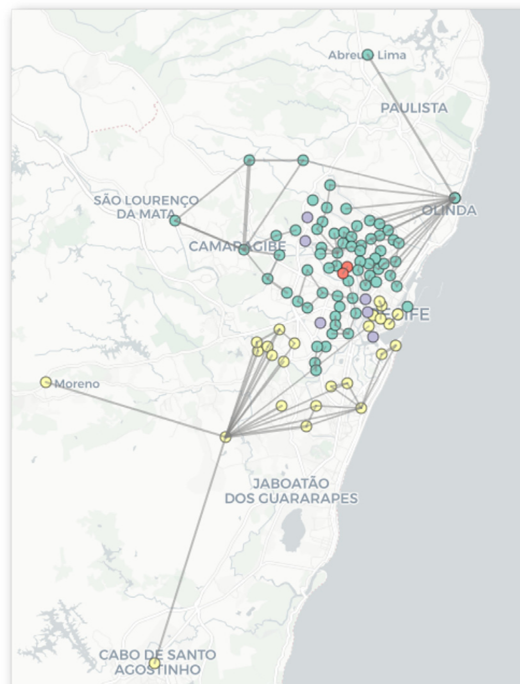


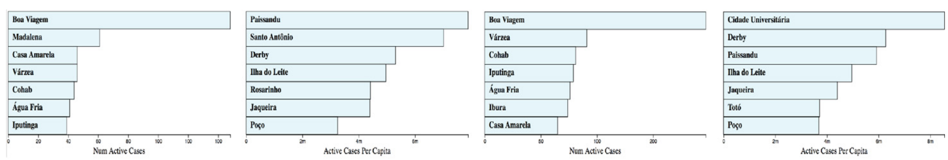
Fig. 6. Cities most related to Recife's neighborhood with $l = 0.05$. This analysis provides information on which cities are more probable to share infected people with Recife.

A more detailed analysis of Recife's neighborhoods is included in Section 5.3. Nevertheless, having a view of the relations between the neighborhoods and cities within the region is interesting. Since the flow of people in the metropolitan area is intense during working hours, it is essential to understand the networks before relaxing social distancing and working policies related to the pandemic. As shown in Fig. 6, Recife's neighborhoods are within clusters that contain cities such as Olinda, Paulista, Jaboatão dos Guararapes, and Itapissuma.

5.3. Neighborhoods of Recife

This section presents the analysis that considers only the neighborhoods of Recife. Fig. 7 presents the clusters construction

Fig. 7. Clusters (top) and associated scatter plots for two particular dates (bottom). The scatterplots show the relationship of the Force of Infection and Active Cases Per Capita for clusters generated from different mobility thresholds. We can see how this relationship changes over time by comparing the scatterplot in different periods. The values used to generate the three plots were, from the left to the right, $l = 0.01$, $l = 0.05$ and $l = 0.1$, respectively.



(a) 2020/05/06

(b) 2020/05/22

Fig. 8. Bar charts showing the active cases and active cases per capita bar plots in the neighborhoods of Recife corresponding to two dates: 2020/05/06 (top) and 2020/05/22 (bottom). This widget ranks individual locations according to the chosen indicator.

considering three different thresholds. The lowest threshold ($l = 0.01$ on the left) results in the identification of a single cluster, considering all Recife's neighborhoods. Increasing the threshold ($l = 0.05$ on the center and $l = 0.1$ on the right), more clusters are identified, this being interesting for policymakers when deciding over social isolation policies.

Of course, if the threshold is large enough, only singletons will be shown. The dispersion plots in Fig. 7 show the relationship between the Force of Infection of the clusters and their active cases per capita at two dates, considering the same thresholds of $l = 0.01$, $l = 0.05$, and $l = 0.1$.

In Fig. 8, bar plots illustrate the most infected Recife's neighborhoods in terms of the total number of cases and the number of cases per capita. While Boa Viagem has a higher number of cases on both dates, other neighborhoods (Paissandu and Cidade Universitária) have a higher number of cases per capita. Several hospitals and clinics of Recife are located in Paissandu, which can indicate that this neighborhood receives people from other locations that represent potential contamination cases.

Cidade Universitária is a neighborhood that is mostly centered on activities related to Universidade Federal de Pernambuco, which has a large Clinic Hospital that serves for teaching and provides general public health services. Although the university has stopped its teaching activities since March 15, 2020, and deployed the home office for all non-essential activities, it was not expected that this neighborhood would have such a high incidence of cases.

It is possible that this could be revealing a similar pattern that was found by [71] around two Wuhan hospitals related to a high concentration of SARS-CoV-2 RNA in aerosols, although the infectivity of the aerosolized virus still requires further studies.

Thus, while further studies are not carried out, authorities should reinforce rigorous sanitation procedures and protection for the surrounding population of these areas, which could include extra care for the elderly that lives in such a neighborhood.

Further analysis with the clusters generated with $l = 0.05$ and socioeconomic variables can be performed to highlight specific aspects that might be important considering individual behavior or risk factors that may influence public policy success or suggest additional preventive actions.

There are several socioeconomic composite indexes [72–74] that may reveal specific conditions of the population. Thus, besides considering HDI, other aspects that can provide insights for policymakers have been included due to its correlation with faster spread of disease, severe symptoms of SARS-CoV-2 and lethality.

Fig. 9 shows clusters and colored regions for different socioeconomic indicators and variables related to SARS-CoV-2. The five visualizations enable the investigation of different characteristics that may drive specific public policies for a cluster or even part of a cluster.

Evaluating the clusters and neighborhoods according to HDI, it is possible to verify in Fig. 9(a) that there is one specific cluster, C_1, which shows the highest social differences located in the north area of Recife. This reflects the mobility pattern of this low-income population who work, in general, to provide services for this wealthy area of the city.

As shown in Figs. 9(a) and 9(b), longevity is related with quality of life conditions. However, the elderly population can be an important aspect when analyzing FOI to minimize fatalities. Similarly, the number of persons per bedroom can bring insights and be an important aspect of policymakers. As shown in Fig. 9(c), there are specific neighborhoods associated with low HDI where several persons live in the same habitation, and therefore more than two persons share the same bedroom.

This is an important piece of information for policymakers as it can be impossible for such people to keep social distancing or even isolate the family member who develops COVID-19 symptoms. This means that if one family member is infected, all persons living in this house might be infected at the same time, increasing the disease contagion rate.

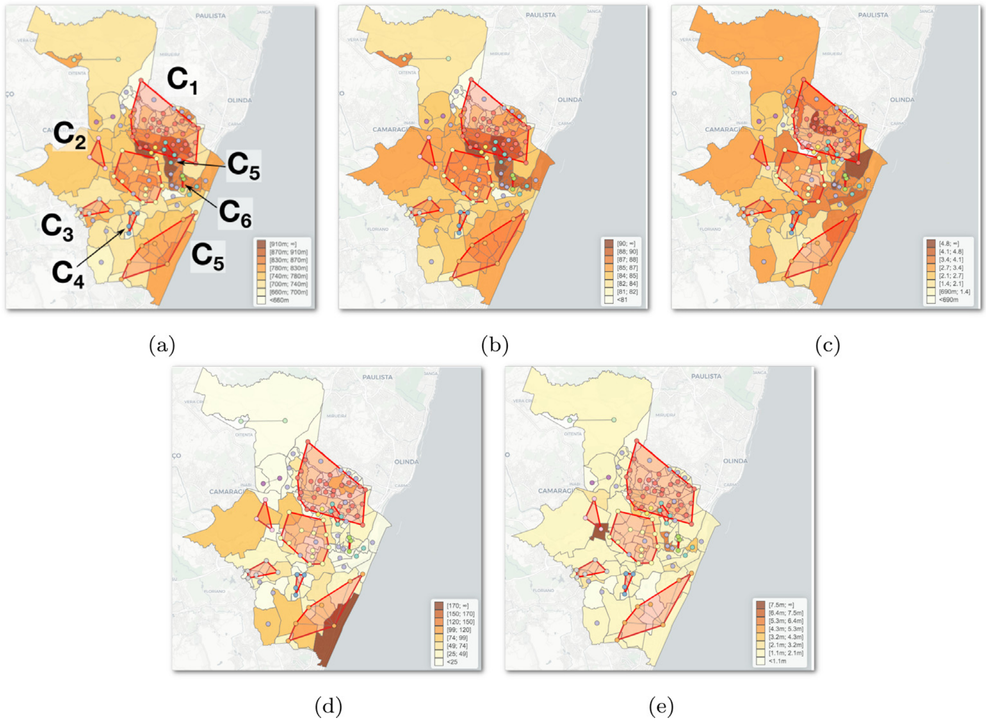


Fig. 9. Clusters with mobility threshold equal to $l = 0.05$ and regions colored by socioeconomic indicators and epidemiologic variables: (a) HDI (Human Development Index) - socioeconomic, (b) proportion of elderly people — socioeconomic, (c) percentage of houses with two or more people per room — socioeconomic, (d) number active COVID-19 cases — epidemiologic, and (e) number of active COVID-19 cases per capita — epidemiologic.

Concerning the infection-related variables, the absolute number of active cases can hide the severity of epidemic spread over neighborhoods with a lower population, as shown in Figs. 9(d) and 9(e). For instance, although the neighborhood of Boa Viagem has the highest number of active cases in Fig. 9(d), it is one of the less critical neighborhoods according to Fig. 9(e) because this neighborhood has about 120 thousand inhabitants.

Fig. 10 presents three scatter plots concerning FOI against socioeconomic indicators. From Fig. 10(a), we can see that until the date of this analysis, the majority of neighborhoods that contain the above-average value of the indicator "more than two persons per bedroom" are concentrated in lower FOI values. This reveals that there should be preventive or educational actions to keep these values under control. Otherwise, there will be fast growth in infected cases that originated in these areas.

Paissandu is a neighborhood with one of the highest FOI and a percentage of the population with more than two persons per bedroom. Therefore public services should plan on how to support this population by aiding social distancing and isolation of infected persons since the virus can infect more people per domicile.

In Fig. 10(b), it can be observed that most of the neighborhoods with a low-educated population still have a low FOI. Therefore behavioral aspects should be reinforced to avoid epidemic spread amongst this population. The Totó neighborhood has the highest FOI amongst these neighborhoods with a low-educated population.

Fig. 10(c) shows that neighborhoods with a higher elderly population and higher FOI, such as Derby, are in dangerous situations. Similarly, public services can seek to offer specialized support for these neighborhoods to reduce the number of fatalities.

5.4. Expert feedback

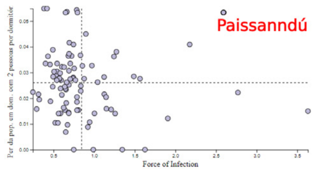
We presented the developed system and case studies above to our project team, agents from the city administration, and researchers working on similar initiatives in other states of Brazil. All of them gave positive feedback concerning the usefulness and easiness of usage of the tool. In general, they were particularly pleased with the functionality of testing multiple isolation scenarios interactively by changing the mobility threshold.

Important feedback was about using socioeconomic variables such as the service level of water supply per neighborhood, homes with more than two persons per bedroom, HDI, and education. Depending on how these factors are associated with a neighborhood and the epidemic advance, there are possible actions to be deployed in terms of enhancing water supply and educational campaigns to stress the population about hygiene and protective measures. Furthermore, the ability to inspect different urban data sources in the context of the pandemic spread data was also appreciated.

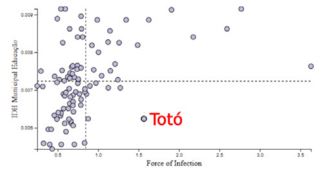
Finally, the domain experts made many suggestions for extensions to the tool. For example, they wanted to see different spatial data aggregation levels to investigate in more depth the isolation strategy that could be used. While meeting with experts from the city administration's mobility intelligence department, we had access to the mobility survey carried out by experts of the Secretariat for Urban Planning.

They were very interested in this tool and posed that it could be interesting to integrate our tool with their mobility survey dataset to perform a similar analysis. This survey was performed in 2018 with a sample of two hundred thousand questionnaires that include the modes of transportation used (bus, bike, private cars, etc.).

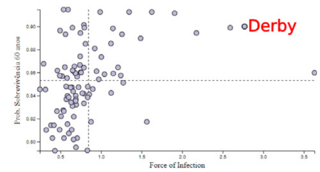
Thus, one direction for future work is to include this piece of information for statistical Bayesian models to enhance new case estimates according to different scenarios. Another constructive feedback was related to the integration of this tool with the mobility survey to optimize the schedule of opening hours in different economic activities, which shall be a new project to be developed after the COVID-19 outbreak.



(a) FOI vs percentage of the population with more than two persons per bedroom.



(b) FOI vs Education HDI.



(c) FOI vs elderly population.

Fig. 10. Comparison between FOI and some socioeconomic indicators using the scatter plot widget.

Finally, one important concern raised by the experts is that the sub-notification of COVID-19 cases could very much influence the results due to low testing rates in the region. While this is out of the scope of our tool, we acknowledge that this is an important issue, and strategies to compensate for this problem could be used. We intend to investigate this in the future.

6. Discussions

As in all affected countries, policymakers face complex tasks and decisions due to the spread of the coronavirus. Society needs resilience and fast responses not only during this pandemic but others that may appear in the future. We believe that the proposed graphical tool can support the policymakers' decisions to monitor and mitigate the effects of COVID-19 in the studied state and municipality.

One can understand how mobility restriction can mitigate or aggravate the spread of the virus. The generated clusters provide meaningful relationships among each location. Also, one can try to investigate how the socioeconomic characteristics of a specific singleton/cluster can influence the impacts of the epidemic. Decision-makers may include and remove indicators to increase the output quality of the analysis.

The active case analysis can be performed with period variation, thus opening opportunities for infectious individuals' time series analysis in some locations or groups of locations and consistency checking of new risk measures.

We decided to use scatter plots to provide visual identification of the pandemic situation for different types of neighborhoods characteristics. The objective was not to obtain correlations or a multivariate regression model but to orient decisions based on daily photography of epidemic conditions of each neighborhood type (with less per capita income, with less education, and others). The barplot panel was adopted to provide location ranking and intensities associated with chosen variables available in the dropdown.

With the obtained results, the developed visual analytics system is capable of providing the necessary information to support disease monitoring and progressive economic recovery in the studied state and city and also in other regions. The benefits of the proposed visual analytics approach are many. It provides a better way to allocate resources by dividing the state while observing important socioeconomic indicators in each division. Thus, the user can easily compare groups of locations, and investigate these locations individually if necessary.

This tool is very flexible concerning the addition of new variables in the analysis. Thus, other indicators or data about the hospital's capacity in each location can be easily integrated.

Also, in visual analytics for epidemics, it may be necessary to compare the evolution of the infection risk of each location over time. This analysis involves the visualization of multiple overlapping time series, which is associated with clutter problems [75,76].

The tool developed in this work does not require the use of time series data to evaluate how the risk of the pandemic evolves locally over time. It is possible to compare the most affected regions by monitoring the clusters' evolution over time.

The drawbacks of this paper are associated with assumptions about mobility and epidemiologic data. We considered that the population used the same mode of transport. But, there are many different modes of transport, such as bus, car, motorcycle, bicycle, and on foot.

If mobility data on different modes of transport is available, one can make the risk model more precise by considering that the contact rate is distinct in each mode of transport. Also, sub-notifications of the active cases due to low testing rates in the region may affect the value of the risk metrics and cluster classification.

Some assumptions on the contact rate and infection rate were made to simplify the risk model, which was sufficient to analyze the current state of the pandemic and to monitor this state over time. But, if the user wants to predict future scenarios with complex simulations in the generated clusters, it is necessary to estimate more precise parameters for the risk model, using, for instance, bayesian inference.

Since the tool developed in this work was integrated with the risk-based approach from [13], we compare the contributions of each work. Table 2 compares the work developed in [13] and the current work concerning the data used in the analysis, regions, expert interaction, software developed, and analysis performed.

The work developed in [13] studied how the lockdown would impact the mobility of the population and how to divide the state for a better allocation of resources. A clustering approach was developed to divide Pernambuco, where the partition was defined as the maximum number of clusters that cover a representative part of the population of Pernambuco (more than 90%).

The experiments evaluated the sensibility of the cluster approach when considering and not considering lockdown measures. Also, it was possible to rank the clusters to evaluate which were the most severe in the state and which were able to plan a safe economic recovery. These analyses were static (performed on a unique day).

In this work, a visual analytics tool was developed to support isolation and economic recovery policies. Experts responsible for defining scientific recommendations for Recife's city hall contributed, interactively, to the development of the tool. The tool enabled the interactive parameterization of the clustering method from [13] and monitored the clusters' risk over time.

This visual analytics tool could be used to cross information on mobility, epidemiologic and socioeconomic data, which enabled the analysis of clusters' risk over time, evaluate socioeconomic impacts due to the virus propagation in the studied period, the impact of the pandemic in productive arrangements, and individual analysis of cluster components. Some of these analyses were dynamic (performed over multiple days).

Silva et al. [13] presented initial modeling aspects in the beginning of lockdown process, considering only the State level

Table 2
Comparison between the contributions of [13] and this study.

Study	Data	Region	Interaction with experts	Software	Analysis performed
[13]	Epidemiologic, Mobility	Pernambuco	No	No	Static Risk analysis, Lockdown effect
This study	Epidemiologic, Mobility, Socioeconomic	Pernambuco, Recife	Yes	Yes	Dynamic Risk analysis, Socioeconomic impact, Impact on productive arrangements, Impact on cluster constituents

to provide intelligence for policymakers in a higher level. Firstly, some information could not be easily used to make decisions in other locations since the analyses were performed for a specific region. The proposed tool's shared repository contains instructions on how to adapt it to other locations. It can also be extended to incorporate any epidemiologic variable and other indicators for the region under study.

Secondly, the main objective of [13] was to evaluate the lockdown effect on the risk of contamination, which was more adequate for the first wave of the pandemic. The tool proposed in this work can be used in any phase of the pandemic, since one can evaluate the impact of mobility in the contamination risk and on different aspects of the region, such as socioeconomic variables and economic arrangements. This enables the policymakers to incorporate more information on the decision making process for the tightening or relaxation of isolation measures over time.

Thirdly, a visualization approach can be helpful in understanding the mobility dynamics and characteristics of a city with more precision. For instance, it was possible to analyze which neighborhoods of Recife, the capital of the estate of Pernambuco, received the most citizens from other cities of Pernambuco. This enabled the development of risk mitigation actions in neighborhoods of Recife that are more likely to commute people from other cities.

Also, the proposed tool enables the visualization of characteristics of neighborhoods within a city, which facilitates the visualization of the income among regions, or the work sectors of its residents, to develop strategies to prioritize resources for reducing the damage of the pandemic in specific neighborhoods according to their socioeconomic characteristics.

Finally, since the proposed tool contains visual and interactive elements, it enabled the policymakers to define mobility thresholds by visualizing the whole geographical configuration and to better understand the problem considering the generated partitions and the characteristics of each location inside a partition.

The visualization of the evolution of the pandemic situation over time, interactiveness, and the possibility to cross multiple sources of information were not available in the past approach [13]. Thus, the proposed tool makes it possible for policymakers to understand the purpose of the clustering approach, learn new information about the current pandemic situation, and make better decisions considering the characteristics of each location.

The main challenge of a real-time application for this tool is to guarantee the quality of data associated with mobility and epidemiologic variables to maintain or even increment the precision of the estimated risk metrics. It is necessary to gather mobility data on a daily basis to update the worst-case mobility OD matrix, and thus consider precise connections between the studied locations. Since some of these locations did not have mobility data monitored, such work can be challenging. Therefore, this work considered the dataset available but considered the most conservative approach by taking the maximum observed mobility in the risk model.

Also, it is necessary to maintain an updated database of epidemiologic variables, monitoring the status of the contaminated citizens over time, to provide an accurate number of active cases over time. During the first weeks of contagion, the daily updated database had several inconsistencies, therefore requiring to be cleaned and verified on a daily basis. Thus, by handling these datasets properly it is possible to update the risk metrics correctly and make the correct decisions.

7. Conclusions

In this work, a dashboard that supports the definition of restriction policies, based on isolated macro-regions, was proposed. The developed tool can cross information about socioeconomic indicators, mobility, and epidemic variables through multiple coordinated views. Considering the amount of data available, this data-driven approach enabled an analytical process to support experts responsible to provide scientific recommendations for health officials from Recife's city hall in Brazil.

The tool was integrated with the risk-based approach of [13], where the user controls a threshold parameter that defines mobility restrictions. The mobility restrictions divide the location into representative clusters based on anonymized mobility data. The multiple coordinated views support the user in exploring the generated clusters' information. Each cluster contains information about the epidemic variables' state and their socioeconomic aggregated information. These features support the user in managing the risk of the pandemic by comparing which cluster is more healthy or riskier.

Plans for progressive economic recovery can be traced for more healthy clusters and risk mitigation plans can be traced for more infected clusters. Other features include the possibility to monitor the cluster state over time and the analysis of a cluster's constituents, including individual risk and socioeconomic indicators, which are summarized in dispersion and bar plots. Thus, the tool can generate, compare and monitor clusters, and explore the constituents of a cluster, based on mobility, socioeconomic, and epidemiologic variables.

The proposed dashboard was applied to Recife, one of the most affected cities in Brazil, and its state Pernambuco. It was possible to visualize the construction of clusters amongst cities and districts, which can be used to prioritize mitigation strategies and monitor the risk based on different regions.

It was possible to understand how Recife is affected by nearby cities. Furthermore, findings over the most affected districts of Recife were discussed. An ex-post analysis using the scatter plot could provide visual clues, considering the cumulative epidemic variables, to provide insights about how socioeconomic factors influenced the outbreak in the clusters' constituents.

Future work includes the development of other analytical components to support policymakers, which may be included as new features of this decision support tool, or developed in a new specific decision support system. Some examples of new features could be the inclusion of more socioeconomic and epidemiologic variables, or other features associated with a given municipality or neighborhood.

New decision support systems that use our approach for mobility data could offer time series visualization widgets to compare risk measures (risk of infection and exposure risk) over time for different locations, by embedding techniques such as the one developed in [75].

An important increment is to consider experts' perceived risk [50,77,78] to estimate an automated risk model to compare the locations under study. Thus, a decision-maker could classify some locations, given information about the cluster they belong and risk measures, and a learning algorithm would periodically estimate, for each location, risk perception for different levels of contamination, exposure, and other indicators. Thus, the learning algorithm would support policymakers' decisions concerning resource allocation for each location according to its estimated risk.

The insights obtained with this visualization tool enable new research opportunities with analytics, such as data mining, machine learning, multicriteria decision analysis, and multiobjective optimization for risk mitigation and economic recovery. Those methods can use the information generated by our clustering approach and generate better models for epidemic propagation simulation and prescriptive risk management, to be used in what–if scenario analysis. Thus, it is possible to consider the impact of mobility dynamics and characteristics of a given region in the model.

CRediT authorship contribution statement

Julio Cezar Soares Silva: Conceptualization, Writing, Methodology, Data curation, Formal analysis. **Diogo Ferreira de Lima Silva:** Conceptualization, Writing, Methodology, Data curation, Formal analysis. **Nivan Roberto Ferreira Júnior:** Conceptualization, Writing, Methodology, Data curation, Formal analysis. **Adiel Teixeira de Almeida Filho:** Conceptualization, Writing, Methodology, Data curation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank the In Loco company, D.A.D.O Team from Porto Digital and Prefeitura da Cidade do Recife for all the cooperation for this study. This study was partially supported by Facepe, Brazil (IBPG-0753-3.08/17, IBPG-0373-1.03/19), CNPq, Brazil (315245/2020-4) and CAPES, Brazil (001).

References

[1] X. Gao, J. Yu, Public governance mechanism in the prevention and control of the COVID-19: information, decision-making and execution, J. Chin. Gov. 5 (2) (2020) 178–197.
[2] S. Shrivastava, P. Shrivastava, Minimizing the risk of international spread of coronavirus disease 2019 (COVID-19) outbreak by targeting travelers, J. Acute Dis. 9 (2) (2020).
[3] M.U.G. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D.M. Pigott, L. du Plessis, N.R. Faria, R. Li, W.P. Hanage, J.S. Brownstein, M. Layan, A. Vespignani, H. Tian, C. Dye, O.G. Pybus, S.V. Scarpino, The effect of human mobility and control measures on the COVID-19 epidemic in China, Science 368 (6490) (2020) 493–497.
[4] A. Katal, L. (Leon) Wang, M. Albettar, A real-time web tool for monitoring and mitigating indoor airborne COVID-19 transmission risks at city scale, Sustainable Cities Soc. 80 (2022).
[5] B. Fatimah, P. Aggarwal, P. Singh, A. Gupta, A comparative study for predictive monitoring of COVID-19 pandemic, Appl. Soft Comput. 122 (2022).
[6] B. Zhou, G. Yang, Z. Shi, S. Ma, Interpretable temporal attention network for COVID-19 forecasting, Appl. Soft Comput. 120 (2022).
[7] M. Gupta, R. Jain, S. Taneja, G. Chaudhary, M. Khari, E. Verdú, Real-time measurement of the uncertain epidemiological appearances of COVID-19 infections, Appl. Soft Comput. 101 (2021).
[8] B. Ghosh, A. Biswas, Status evaluation of provinces affected by COVID-19: A qualitative assessment using fuzzy system, Appl. Soft Comput. 109 (2021).
[9] L. Aggarwal, P. Goswami, S. Sachdeva, Multi-criterion intelligent decision support system for COVID-19, Appl. Soft Comput. 101 (2021).
[10] R.L. Keeney, Value-focused thinking: Identifying decision opportunities and creating alternatives, European J. Oper. Res. 92 (3) (1996) 537–549.
[11] OECD, First lessons from government evaluations of COVID-19 responses: A synthesis, 2022, URL: https://www.oecd.org/coronavirus/policy-responses/first-lessons-from-government-evaluations-of-covid-19-responses-a-synthesis-483507d6/.
[12] S. Xu, Y. Li, Beware of the second wave of COVID-19, Lancet 395 (10233) (2020) 1321–1322.
[13] J.C.S. Silva, D.F. de Lima Silva, A. de Sá Delgado Neto, A. Ferraz, J.L. Melo, N.R. Ferreira Júnior, A.T. de Almeida Filho, A city cluster risk-based approach for Sars-CoV-2 and isolation barriers based on anonymized mobile phone users' location data, Sustainable Cities Soc. 65 (2021).
[14] InLoco Company, Location data to control COVID-19 while respecting individual privacy, 2020.
[15] N. Andrienko, G. Andrienko, Informed spatial decisions through coordinated views, Inf. Vis. 2 (4) (2003) 270–285.
[16] W. Chen, Z. Huang, F. Wu, M. Zhu, H. Guan, R. Maciejewski, VAUD: a visual analysis approach for exploring spatio-temporal urban data, IEEE Trans. Vis. Comput. Graph. 24 (9) (2018) 2636–2648.
[17] N. Andrienko, G. Andrienko, G. Fuchs, P. Jankowski, Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces, Inf. Vis. 15 (2) (2016) 117–153.
[18] N. Andrienko, G. Andrienko, Visual analytics of movement: an overview of methods, tools and procedures, Inf. Vis. 12 (2013) 3–24.
[19] W. Zeng, C.-W. Fu, S.M. Arisona, H. Qu, Visualizing interchange patterns in massive movement data, in: Proceedings of the 15th Eurographics Conference on Visualization, EuroVis '13, The Eurographs Association & John Wiley & Sons, Ltd., Chichester, GBR, 2013, pp. 271–280.
[20] G. Di Lorenzo, M. Sbodio, F. Calabrese, M. Berlingerio, F. Pinelli, R. Nair, AllAboard: visual exploration of cellphone mobility data to optimise public transport, IEEE Trans. Vis. Comput. Graphics 22 (2) (2016) 1036–1050.
[21] J. Alejandro Triana, D. Zeckzer, H. Hagen, J. Tiberio Hernandez, VafusQ: A methodology to build visual analysis applications with data quality features, Inf. Vis. 18 (4) (2019) 384–404.
[22] J. Poco, H. Doraiswamy, H.T. Vo, J.L.D. Comba, J. Freire, C.T. Silva, Exploring traffic dynamics in urban environments using vector-valued functions, Comput. Graph. Forum 34 (3) (2015) 161–170.
[23] W. Didimo, L. Giamminonni, G. Liotta, F. Montecchiani, D. Pagliuca, A visual analytics system to support tax evasion discovery, Decis. Support Syst. 110 (2018) 71–83.
[24] T. Sobral, T. Galvao, J. Borges, Visualization of urban mobility data from intelligent transportation systems, Sensors 19 (2) (2019).
[25] I. Boyandin, E. Bertini, P. Bak, D. Lalanne, Flowstrates: an approach for visual exploration of temporal origin-destination data, Comput. Graph. Forum 30 (2011) 971–980.
[26] D. Coelho, I. Chase, K. Mueller, PeckVis: a visual analytics tool to analyze dominance hierarchies in small groups, IEEE Trans. Vis. Comput. Graph. 26 (4) (2020) 1650–1660.
[27] J. Chou, C. Bryan, K. Ma, Privacy preserving visualization for social network data with ontology information, in: 2017 IEEE Pacific Visualization Symposium (PacificVis), 2017, pp. 11–20.
[28] R. Liu, S. Feng, R. Shi, W. Guo, Weighted graph clustering for community detection of large social networks, Procedia Comput. Sci. 31 (2014) 85–94.
[29] S. Fortunato, Community detection in graphs, Phys. Rep. 486 (3–5) (2010) 75–174.
[30] L. Prignano, I. Morer, F. Fulminante, S. Lozano, Modelling terrestrial route networks to understand inter-polity interactions (southern Etruria, 950–500 BC), J. Archaeol. Sci. 105 (2019) 46–58.
[31] J. Hua, M.L. Huang, W. Huang, C. Zhao, Applying graph centrality metrics in visual analytics of scientific standard datasets, Symmetry 11 (1) (2019).
[32] D.M. Bean, C. Stringer, N. Beeknoo, J. Teo, R.J.B. Dobson, Network analysis of patient flow in two UK acute care hospitals identifies key sub-networks for A&E performance, PLoS One 12 (10) (2017).
[33] R. Krueger, Q. Han, N. Ivanov, S. Mahtal, D. Thom, H. Pfister, T. Ertl, Birds-eye - large-scale visual analytics of city dynamics using social location data, Comput. Graph. Forum 38 (3) (2019) 595–607.

[34] T. von Landesberger, F. Brodkorb, P. Roskosch, N. Andrienko, G. Andrienko, A. Kerren, MobilityGraphs: visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering, IEEE Trans. Vis. Comput. Graphics 22 (1) (2016) 11–20.
[35] Y. Wang, G. Baciu, C. Li, Visualizing dynamics of urban regions through a geo-semantic graph-based method, Comput. Graph. Forum 39 (1) (2020) 405–419.
[36] X. Lu, Web-GIS-based SARS epidemic situation visualization, in: J. Sun, Z. Pan (Eds.), Fourth International Conference on Virtual Reality and its Applications in Industry, Vol. 5444, International Society for Optics and Photonics, SPIE, 2004, pp. 445–452.
[37] J.A. Polonsky, A. Baidjoe, Z.N. Kamvar, A. Cori, K. Durski, W.J. Edmunds, R.M. Eggo, S. Funk, L. Kaiser, P. Keating, O.l.P. de Waroux, M. Marks, P. Moraga, O. Morgan, P. Nouvellet, R. Ratnayake, C.H. Roberts, J. Whitworth, T. Jombart, Outbreak analytics: a developing data science for informing the response to emerging pathogens, Philos. Trans. R. Soc. B 374 (1776) (2019).
[38] World Health Organization, Coronavirus disease (COVID-19) situation report – 136, 2020, URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.
[39] S.K. Dey, M.M. Rahman, U.R. Siddiqi, A. Howlader, Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach, J. Med. Virol. 92 (6) (2020) 632–638.
[40] P. Gao, H. Zhang, Z. Wu, J. Wang, Visualising the expansion and spread of coronavirus disease 2019 by cartograms, Environ. Plan. A: Economy Space 52 (4) (2020) 698–701.
[41] P. Nikolaou, L. Dimitriou, Identification of critical airports for controlling global infectious disease outbreaks: Stress-tests focusing in Europe, J. Air Transp. Manag. 85 (March) (2020).
[42] C. Zhou, F. Su, T. Pei, A. Zhang, Y. Du, B. Luo, Z. Cao, J. Wang, W. Yuan, Y. Zhu, C. Song, J. Chen, J. Xu, F. Li, T. Ma, L. Jiang, F. Yan, J. Yi, Y. Hu, Y. Liao, H. Xiao, COVID-19: challenges to GIS with big data, Geogr. Sustain. 1 (1) (2020) 77–87.
[43] X. Xie, E. Naminse, S. Liu, Q. Yi, The spatial and temporal pattern of COVID-19 and its effect on humans' development in China, Global J. Environ. Sci. Manag. 6 (Special Issue (Covid-19)) (2020).
[44] E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time, Lancet Infect. Dis. 20 (5) (2020) 533–534.
[45] World Health Organization, WHO coronavirus disease (COVID-19) dashboard, 2020, URL: https://covid19.who.int/.
[46] S. Barone, A. Chakhunashvili, A. Comelli, Building a statistical surveillance dashboard for COVID-19 infection worldwide, Qual. Eng. (2020).
[47] S.-H. Zhang, Y. Cai, J. Li, Visualization of COVID-19 spread based on spread and extinction indexes, Sci. China Inf. Sci. 63 (2020).
[48] S. Khan, P.H. Nguyen, A. Abdul-Rahman, B. Bach, M. Chen, E. Freeman, C. Turkay, Propagating visual designs to numerous plots and dashboards, IEEE Trans. Vis. Comput. Graphics 28 (1) (2022) 86–95.
[49] J.C.C. Ipenza, N.M.L. Romero, M. Loreto, N.F. Júnior, J.L.D. Comba, QDS-COVID: A visual analytics system for interactive exploration of millions of COVID-19 healthcare records in Brazil, Appl. Soft Comput. 124 (2022).
[50] B. Preim, K. Lawonn, A survey of visual analytics for public health, Comput. Graph. Forum 39 (1) (2020) 543–580.
[51] A. Barabasi, Network Science, Cambridge University Press, 2016.
[52] K. Reda, C. Tantipathananandh, A. Johnson, J. Leigh, T. Berger-Wolf, Visualizing the evolution of community structures in dynamic social networks, Comput. Graph. Forum 30 (3) (2011) 1061–1070.
[53] J. Tao, J. Xu, C. Wang, N.V. Chawla, HoNVis: Visualizing and exploring higher-order networks, in: 2017 IEEE Pacific Visualization Symposium (PacificVis), IEEE, 2017, pp. 1–10.
[54] W. Luo, Visual analytics of geo-social interaction patterns for epidemic control, Int. J. Health Geogr. 15 (1) (2016).
[55] C.D.G. Linhares, B.A.N. Travençolo, J.G.S. Paiva, L.E.C. Rocha, DyNetVis: a system for visualization of dynamic networks, in: Proceedings of the Symposium on Applied Computing, SAC '17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 187–194, http://dx.doi.org/10.1145/3019612.3019686.
[56] C.D.G. Linhares, J.R. Ponciano, J.G.S. Paiva, L.E.C. Rocha, B.A.N. Travençolo, DyNetVis - An interactive software to visualize structure and epidemics on temporal networks, in: 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM, 2020, pp. 933–936, http://dx.doi.org/10.1109/ASONAM49781.2020.9381304.
[57] C.D.G. Linhares, J.R. Ponciano, J.G.S. Paiva, B.A.N. Travençolo, L.E.C. Rocha, Visualisation of structure and processes on temporal networks, in: Computational Social Sciences, Springer International Publishing, 2019, pp. 83–105, http://dx.doi.org/10.1007/978-3-030-23495-9_5.
[58] G.J. Abel, S. Gietel-Basten, International remittance flows and the economic and social consequences of COVID-19, Environ. Plan. A: Economy Space (2020).
[59] S. Saraswathi, A. Mukhopadhyay, H. Shah, T.S. Ranganath, Social network analysis of COVID-19 transmission in Karnataka, India, Epidemiol. Infect. 148 (2020).
[60] C. Luo, Y. Ma, P. Jiang, T. Zhang, F. Yin, The construction and visualization of the transmission networks for COVID-19: A potential solution for contact tracing and assessments of epidemics, Sci. Rep. 11 (1) (2021).
[61] H. Xu, A. Berres, G. Thakur, J. Sanyal, S. Chinthavali, EPIsembleVis: A geo-visual analysis and comparison of the prediction ensembles of multiple COVID-19 models, J. Biomed. Inform. 124 (2021) 103941.
[62] S. Jamonnak, D. Bhati, M. Amiruzzaman, Y. Zhao, X. Ye, A. Curtis, VisualCommunity: a platform for archiving and studying communities, J. Comput. Soc. Sci. 5 (2) (2022) 1257–1279.
[63] L.P. Willcocks, C. Sauer, M.C. Lacity, Formulating Research Methods for Information Systems, Vol. 2, Palgrave Macmillan UK, London, 2015.
[64] M. da Saúde, Coronavírus Brasil, 2020, URL: https://covid.saude.gov.br/.
[65] S. de Planejamento e Gestão, COVID-19 no Mundo, no Brasil e em Pernambuco, 2020, URL: https://dados.seplag.pe.gov.br/apps/corona.html.
[66] IBGE, Intituto Brasileiro de Geografia e Estatística, 2020, URL: https://www.ibge.gov.br/.
[67] J.A. Bondy, U.S.R. Murty, Graph Theory, in: Graduate Texts in Mathematics, vol. 244, Springer London, London, 2008.
[68] V.A. Traag, L. Waltman, N.J. van Eck, From Louvain to Leiden: guaranteeing well-connected communities, Sci. Rep. 9 (1) (2019).
[69] M.J. Keeling, P. Rohani, Modeling Infectious Diseases in Humans and Animals, Princeton University Press, 2008.
[70] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks, Proc. Natl. Acad. Sci. USA 101 (11) (2004) 3747–3752.
[71] Y. Liu, Z. Ning, Y. Chen, M. Guo, Y. Liu, N.K. Gali, L. Sun, Y. Duan, J. Cai, D. Westerdahl, X. Liu, K. Xu, K. fai Ho, H. Kan, Q. Fu, K. Lan, Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals, Nature (2020).
[72] R.L. do Carvalhal Monteiro, V. Pereira, H.G. Costa, Dependence analysis between childhood social indicators and human development index through canonical correlation analysis, Child Indic. Res. (2020).
[73] R.L. do Carvalhal Monteiro, V. Pereira, H.G. Costa, Analysis of the better life index trough a cluster algorithm, Soc. Indic. Res. 142 (2) (2019) 477–506.
[74] R.L. do Carvalhal Monteiro, V. Pereira, H.G. Costa, A multicriteria approach to the human development index classification, Soc. Indic. Res. 136 (2) (2018) 417–438.
[75] Y. Zhao, Y. Wang, J. Zhang, C.-W. Fu, M. Xu, D. Moritz, KD-box: line-segment-based KD-tree for interactive exploration of large-scale time-series data, IEEE Trans. Vis. Comput. Graphics 28 (1) (2022) 890–900.
[76] J.C.S. Silva, N.R. Ferreira Júnior, A.T. de Almeida Filho, A tool for visualization and analysis of neighbourhoods, clusters, and indicators during the COVID-19 pandemic, Math. Probl. Eng. (2023) 9524753, http://dx.doi.org/10.1155/2023/9524753.
[77] Y. Zhang, Y. Sun, L. Padilla, S. Barua, E. Bertini, A.G. Parker, Mapping the landscape of COVID-19 crisis visualizations, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21, Association for Computing Machinery, New York, NY, USA, 2021.
[78] K.H. Pine, M. Lee, S.A. Whitman, Y. Chen, K. Henne, Making sense of risk information amidst uncertainty: individuals' perceived risks associated with the COVID-19 pandemic, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21, Association for Computing Machinery, 2021.