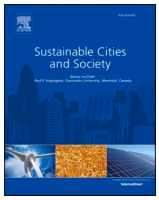
Sustainable Cities and Society 87 (2022) 104187

Contents lists available at ScienceDirect

Sustainable Cities and Society

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



An event-based model and a map visualization approach for spatiotemporal association relations discovery of diseases diffusion

Roya Habibi [a], Ali Asghar Alesheikh [a,*], Sayeh Bayat [b,c]

[a] Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, 19967 15433, Tehran, Iran
[b] Department of Biomedical Engineering, University of Calgary, Canada
[c] Department of Geomatics Engineering, University of Calgary, Canada

ARTICLE INFO

Keywords:
Spatiotemporal association rules
COVID-19
Infectious disease
New York city
Spatial mapping
Spatiotemporal dynamics

ABSTRACT

Infectious disease diffusion is inherently a complex spatiotemporal phenomenon. Simplifying this complexity to discover the associated structure of the city is of great importance. However, existing approaches mainly focus on distance property in geographic space to examine randomness, dispersion, or clustered structure of the disease distribution. While, the outbreak continuously changes its properties, shapes, or locations. Regardless of this adjacency-based structure, there may be associated spatial units that exhibit similar behaviors towards the outbreak fluctuations in a city. To reveal these characteristics, this research proposes a novel event-based spatiotemporal model, mining associated areas in space and time simultaneously. This model was applied to the cases rate of COVID-19 at the ZIP Code level in New York City. The results showed that the proposed approach could sufficiently address the spatiotemporal association relationships. To better understand the discovered associations, a map visualization approach is introduced, allowing recognition of these association relations at a glance. This approach develops a deep understanding of the spatiotemporal structure of the outbreak and better manifests the association and cause-and-effect relations between ZIP Code areas. The results provide good assets for the construction of healthy resilient cities with the function of preventing epidemic crises in the future.

1. Introduction

The world has already faced serious pandemics and epidemics such as plague, cholera, flu, severe acute respiratory syndrome coronavirus (SARS-CoV), and Middle East respiratory syndrome coronavirus (MERS-CoV) (Piret & Boivin, 2021). In the last three years, the world has witnessed coronavirus pandemic, also known as COVID-19 pandemic, that is caused by the SARS-CoV-2 virus. The first official case of human coronavirus disease 2019 (COVID-19) was reported on December 31, 2019 in Wuhan, China, and subsequently it was reported around the world. After a while, the World Health Organization (WHO) declared the outbreak a pandemic on March 11, 2020. In New York City (NYC), the first official case of COVID-19 was reported on March 3, 2020 (Lara-Reyna et al., 2020). Ever since, NYC has witnessed about two million confirmed cases of COVID-19 and more than thirty thousand deaths due to this virus (NYC-Health, 2022). The most cities have also experienced a similar trend, and the disease quickly spread throughout them. To limit the person-to-person transmission of the virus, governments have proposed guidelines for social distancing policies and preventative practices. In this regard, identifying areas that have had similar fluctuations in disease prevalence is very effective in developing policies to prevent and control infections.

The spread of the SARS-CoV-2, like other infectious diseases, can be considered as a spatiotemporal concept; that is, its distribution and diffusion happen in space and time and can be influenced by the surrounding environment. The geographical environments are complex in nature, and this complexity can be further intensified when temporal dimensions are considered. In fact, this complexity mainly lies in the relationships, autocorrelation, and heterogeneities of spatiotemporal data types, making it more difficult to identify patterns in spatiotemporal datasets compared to traditional numerical and categorical datasets (Wan & Zhou, 2008; Akbari et al., 2015). Therefore, what occur in this environment, such as the transmission of infectious diseases, will naturally inherit such complexities, which reflect the nature structure and dynamics of underlying processes. Furthermore, not only the prevalence of contagious disease is very high in some urban settings, but

* Corresponding author.
E-mail address: alesheikh@kntu.ac.ir (A.A. Alesheikh).

https://doi.org/10.1016/j.scs.2022.104187
Received 13 May 2022; Received in revised form 27 August 2022; Accepted 18 September 2022
Available online 19 September 2022

also its spatial distribution is heterogeneous within a city. This significant geographical variation in the disease intensity and extent of transmission occurs due to the various ways it affects the vulnerable groups and the environmental risk factors that facilitate the spatial spread of the pathogen (Cordes & Castro, 2020; Cuadros et al., 2020). In this context, understanding the nature of these spatiotemporal patterns is vital for identifying people who are most at risk, allocating resources, and determining effective intervention measures. To achieve this comprehension, getting insight into the spatial structure of the diffusion, its temporal development, and the interaction of space and time towards the evolution of disease are fundamental tasks (Jaya & Folmer, 2020). Significant efforts have been made to understand the relationships between COVID-19 and demographic characteristics and environmental factors such as air pollution (Srivastava, 2021; Travaglio et al., 2021), humidity (Haque & Rahman, 2020; Ma et al., 2020; Wang et al., 2021a), temperature (Menebo, 2020; Prata et al., 2020; Notari, 2021), and wind (Feng et al., 2020; Rendana, 2020). Moreover, several studies have explored the spatial or spatiotemporal structure of COVID-19 at different resolutions and in various regions. These studies have applied methods such as Spatial and Space-Time Scan Statistics (SaTScan) (Alkhamis et al., 2020; Wang et al., 2021b) and Moran's I indices (Liu et al., 2021; Zheng et al., 2021) to COVID-19 datasets. For instance, a study in NYC investigated spatial inequality in COVID-19 positivity rates across ZIP Codes using Bayesian spatial negative binomial models (Yang et al., 2021), and identified multiple sociodemographic factors contributing to disparities in positivity rates. Cordes & Castro (2020) analyzed clusters of testing rates, positivity rates, and proportion positive across NYC ZIP Codes. They also specified associations between clusters, and contextual factors. Furthermore, another study identified emerging hotspots of COVID-19 test percent positivity using space-time scan statistic across NYC (Greene et al., 2021). However, these studies have two main shortcomings.

First, their approaches mainly focus on the clustered structure of the region, which is based on adjacency relations and distance between features. Whereas, there are many underlying factors that contribute to the outbreak and result in a random or dispersed spatial structure. Human mobility is one of these factors which plays a partial role in the diffusion of COVID-19 or any other infectious disease. Due to the development of transportation networks and globalization, the infectious diseases spread rapidly and make it difficult to control the diffusion (Zhang et al., 2022). It may cause the pathogen to be diffused to areas that are not in the adjacency of the origin place. Socioeconomic status, worker characteristics (Baker et al., 2020), demographic characteristics, environmental characteristics, and urban structures are other examples of these factors. These factors can lead to the emergence of association relations and similar behavior towards the disease in specific areas that are not necessarily adjacent to each other. Namely, there might be some undetected homogeneity in the heterogeneous space caused by the dynamics of the disease. Second, these spatiotemporal analyses often divide the time interval into several time frames and investigate the clusters and spatial structure across the study area in each time frame separately. Integrating time in the model, however, provides a better understanding of the dynamics and complexities of the disease diffusion. Therefore, although several studies have investigated disparities in COVID-19, the spatiotemporal dynamics and complexities of this disease are still not fully considered. In light of the limitations of the previous studies, this study contributes to the literature with a novel event-based spatiotemporal model to spatiotemporal association relations discovery of disease diffusions. This model provides a way to identify the spatial units that exhibit similar behavior. The model was applied on COVID-19 disease infection at ZIP Code level in New York City. The proposed model is the adoption of traditional association mining. In general, the output of these methods is shown in the text form, as A → B. The interpretation or understanding of results it this text form may have difficulties in practice, in particular, when the results are too numerous to interpret. To overcome this issue, a map visualization method was introduced to better picture the association patterns and rules between spatial units. This visualization approach provides an asset to clarify cause and effect relations between areas in the diffusion process. Although predicting the spatiotemporal evolution of the infectious disease and recognizing the factors that lead to this evolution are of great importance for the prevention and control of the diffusion, these issues are not investigated in this paper and will be left for future studies.

2. Materials and methods

2.1. Study area and data sources

This study was conducted in NYC, located in New York State. With a population of approximately 8.4 million residents and an area of over 302.6 square miles (784 km^2), this city is the most populated city not only in its state, but also in the United States (Zangari et al., 2020). NYC consists of five boroughs: The Bronx, Brooklyn, Manhattan, Queens, and Staten Island. Manhattan is the most and Staten Island is the least populated borough (Fig. 1.a). The city's public hospitals have suffered from various diseases for centuries; among these are yellow fever, cholera, HIV/AIDS in the 18th, 19th, and 20th centuries, respectively, and Ebola just more recently (Chokshi & Katz, 2020). Shortly after the first coronavirus laboratory-confirmed case was reported (February 29, 2020), NYC experienced exponential growth in the COVID-19 rates and emerged as an early epicenter of the pandemic (Lara-Reyna et al., 2020). Each epidemic has brought some challenges for the governors. On March 20, the governor of New York announced the implementation of "New York State on PAUSE" executive order. It was a 10-point policy regarding lockdown measures, social distancing measures, closing all non-essential businesses statewide, and banning all non-essential gatherings of individuals temporarily (State, 2020). However, ever since, the corona outbreak in the city has fluctuated widely.

In this study, the data on rate of COVID-19 cases per 100,000 people COVID-19 were obtained from the NYC Health Department repository on GitHub (NYC-Health, 2022). This dataset was stratified by week and modified ZIP Code Tabulation Area (ZCTA) for 75 weeks from August 2, 2020, to January 8, 2022.The rate of cases is calculated through the following equation (Eq. 1):

$$Rate\ of\ cases = \frac{The\ count\ of\ COVID-19\ cases\ in\ a\ ZCTA}{The\ population\ of\ the\ ZCTA} \times\ 100,000 \tag{1}$$

The data excluded records with missing geographic information and was published using diagnosis date. This reporting approach better represents the time of onset of symptoms or infection. Moreover, this dataset contains information on residence of people's ZCTA at the time of reporting. Together, this information provides a more comprehensive understanding of how COVID-19 has spread in space and time in NYC.

In Fig. 1.b, the weekly trends of cases rate per 100,000 people in the city and each borough are displayed. As it shows, Staten Island has the highest cases rate compared to the other boroughs, while Manhattan has the lowest cases rate. Nonetheless, the critical points for each borough have occurred at approximately the same time. Fig. 2 shows the distribution of cumulative cases rate by ZCTA. It indicates that the outbreak was most severe in Staten Island and the central regions of the city, including the northern part of Brooklyn and the western part of Queens. The histogram of cases rate, stratified by ZCTA, is also shown in Fig. 2. For ease of discussion in the following sections, the map is labeled by ZIP Codes. Table 1 shows the summary statistics of data.

2.2. Global spatial association

To assess the overall spatial association of COVID-19 in NYC, global Moran's I statistic was computed for each week. In a given period t, this statistic (I_t) measures the spatial autocorrelation through the following equation (Eq. 2) (O'sullivan & Unwin, 2003; Habibi et al., 2017):

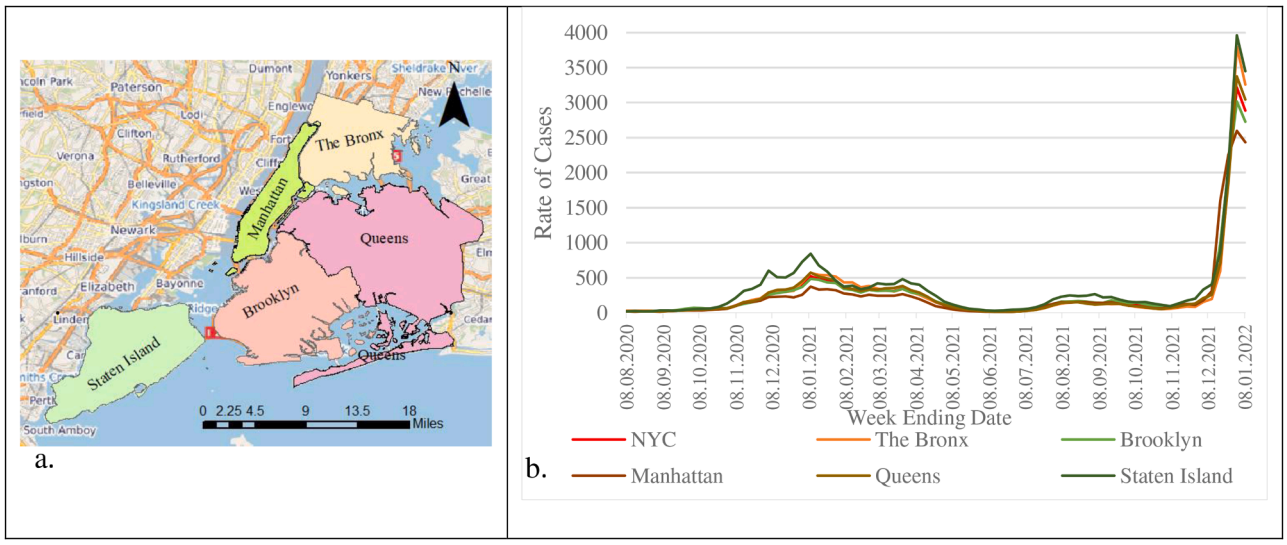


Fig. 1. a. New York City boroughs and modified ZIP Code Tabulation Areas (ZCTA) in each borough. b. Weekly trends of the rate of cases per 100,000 people throughout the New York City and its boroughs.

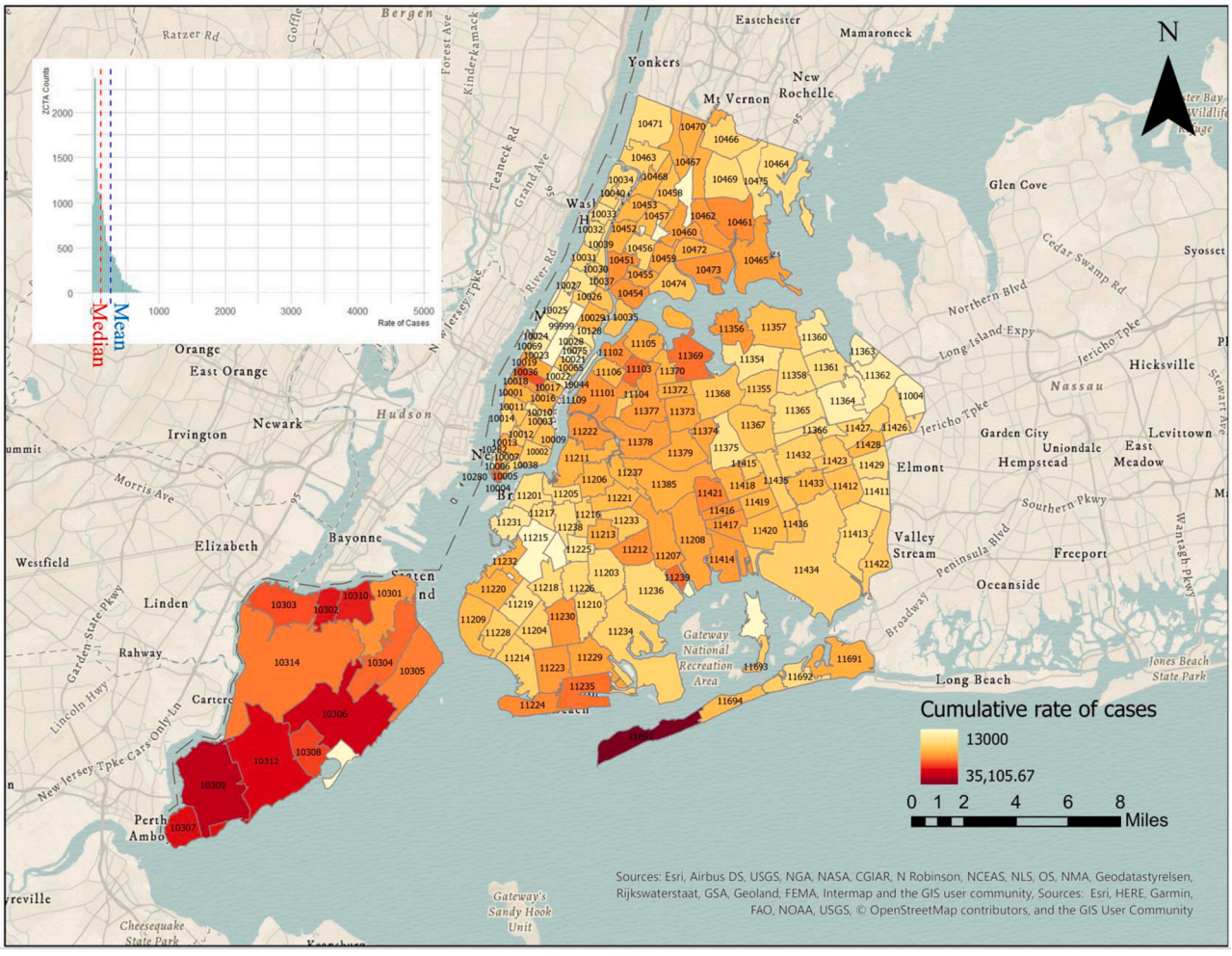


Fig. 2. Map of cumulative cases rate by ZCTA and the histogram of data.

$$I_t = \frac{n\sum_{i=1}\sum_{j=1} W_{ij}\left(x_i - \widehat{X}\right)\left(x_j - \widehat{X}\right)}{\sum_{i=1}\sum_{j=1} W_{ij}\sum_i \left(x_i - \widehat{X}\right)} \quad (2)$$

Where x_i and x_j, respectively, are cases rate in the i^{th} and j^{th} ZCTA, $\widehat{X}$ is the mean of cases rate in the entire city, *Wij* is the spatial weight between the ZIP Code area *i* and *j*, and *n* is the total number of ZCTAs.

Table 1
Summary of COVID-19 cases rate data in NYC.

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
0.00	44.4	125.6	273.4	268.9	4841.8

Index values fall between -1.0 and +1.0. The values greater than 0 show confirmed cases tend to cluster spatially, whereas negative values indicate a negative correlation and 0 values express insignificance. The closer the I statistic to one in absolute value, the stronger the spatial autocorrelation.

2.3. Reprasentaion event-based spatiotemporal model

2.3.1. Problem Definition

It is reported that the distribution of COVID-19 cases does not follow a random distribution and implies a spatial dependence structure globally (Shariati et al., 2020; Huang, 2021; Islam et al., 2021) or locally (Xiong et al., 2020; Das et al., 2021). There are also studies that concluded the clustered structure of COVID-19 diffusion in New York City (Cordes & Castro, 2020; Maroko et al., 2020). It means some locations behave similarly in response to the increase or decrease of the cases rate. This fact authenticates similar underlying processes that happen at these places. There is a hypothesis-testing approach to assessing overall spatial characteristics and association relations mining. The null hypothesis, in this regard, is an independent random process (IRP) or complete spatial randomness (CSR). The outbreak, like any other spatial event, is expected not to match IRP/CSR and show spatial dependence (O'sullivan & Unwin, 2003). It is noteworthy that even if the distribution of spatial events conforms to IRP/CSR, there may still be spatial interdependence relations in the events pattern. The desired associated structure cannot be discovered in a snapshot of the spatiotemporal event. It lies in examining the simultaneous spatiotemporal structure of the data, and refers to the existence of specific spatial events at particular places in several timestamps of an area. Although these events might not occur continuously, they can take place frequently in their specific places. This issue deals with the concept of frequent patterns of events occurring in many timestamps of data.

Dealing with frequent patterns would involve the realm of spatial co-location and co-occurrence patterns mining. Co-location patterns are subsets of spatial features whose instances frequently occur close to each other (Bao & Wang, 2019). Spatial co-location patterns can be interpreted as the result of interactions with spatial processes in which the outcomes at various locations are influenced by one another (Haining & Haining, 2003; Cai et al., 2019). "Induced spatial auto-correlations" lead to these patterns where an underlying variable that is itself spatially autocorrelated induces spatial autocorrelation of each feature (Dale & Fortin, 2014). On the other hand, co-occurrence patterns represent subsets of spatial features that are often located together in space and time. Though these methods enhance our understanding of the spatial characteristics of an area and its association patterns, there are two drawbacks to the current applications of these methods, in the context of the current study. First, these applications consider two or more feature types, investigating their prevalent co-location/co-occurrence patterns. Whereas, studying association patterns of one event type reveals probable spatial associations that are the result of interactions with spatial processes and determinants. Second, these applications consider spatial neighborhood constraints to generate candidate patterns. While,

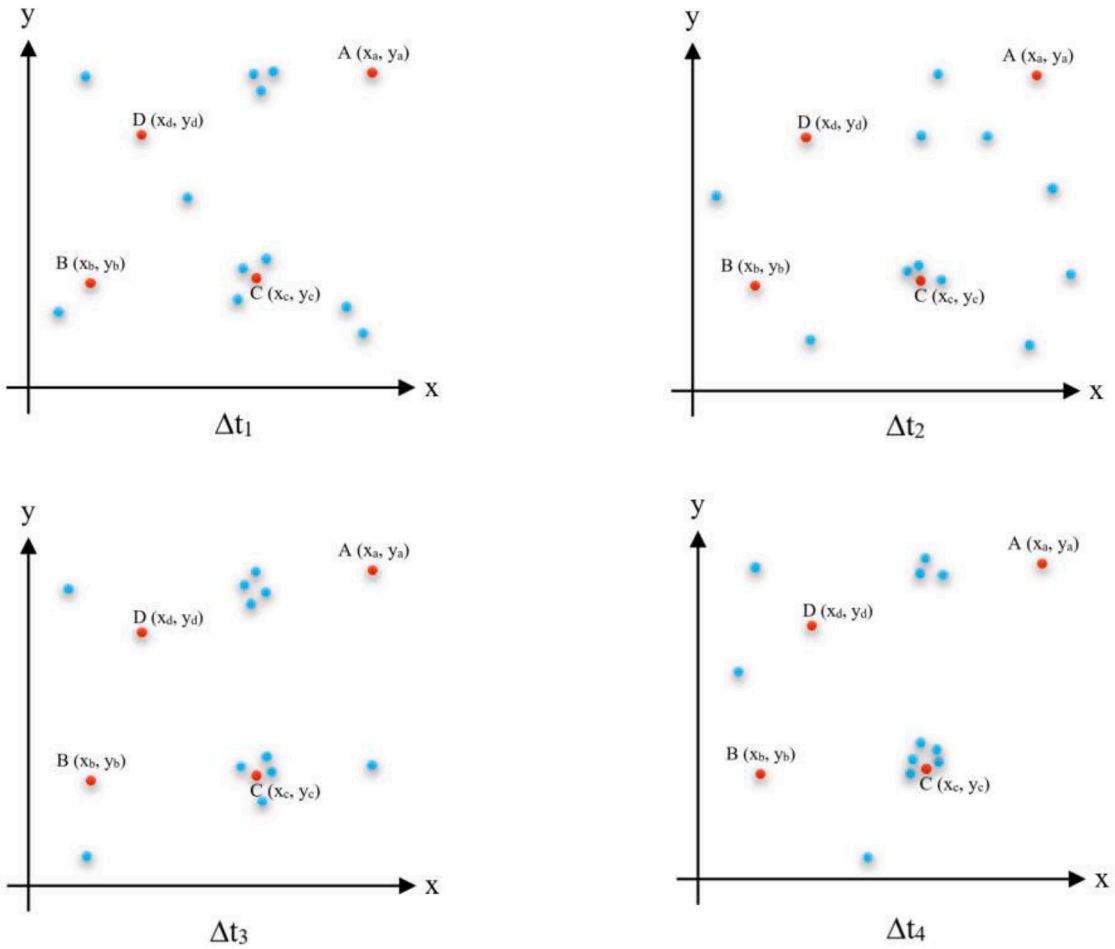


Fig. 3. An example of association relationships (between red points).

particular events of one feature might frequently happen in specific regions which are not close to each other. This fact implies underlying association relations between the regions towards the considered event, such as the outbreak of COVID-19. These relations are valuable resources for the spatial structure of the whole study area. Therefore, an extension to such pattern mining with no neighborhood relationships constraint should be considered to recognize patterns that indicate the same spatial characteristics towards one feature. Fig. 3 demonstrates an example of this extension where all points refer to specific events of one feature that occurred in four timestamps. Although the distribution of events is different, there are events that happened at the particular places repeatedly, which are shown in red. Also, the point C is located in a cluster, while the other red points are not. It is evidence of interdependence and association relations between these places which lead to this pattern, regardless of their distances. In current study, these relationships are investigated to understand the dynamics of COVID-19 outbreak better.

2.3.2. Proposed model

In the spatial or spatiotemporal analysis context, the concept of 'event' focuses more on the occurrence of the geographic phenomena and usually involved with spatiotemporal dynamics (He et al., 2020). It is considered as a significant change in a spatial unit in the observation sequences. This concept varies from spatial features. In Geographic Information Science (GIS) spatial features imply an object-oriented view in which objects are spatially homogeneous entities with distinct boundaries and locations. They usually exist in space and time constantly so that they can hold permanent identities.

In this paper, the spatiotemporal structure of the coronavirus outbreak is studied as a complex geographic event. This complexity is simplified by considering an event in both space and time. Therefore, a conceptual framework is proposed to explore the complexity of COVID-19 outbreak and to discover the associated spatial units across the study area. In this context, a frequent spatiotemporal pattern is considered a repetitive occurrence of events at a specific timestamp throughout the study area. Typically, the frequent pattern mining techniques search for recurring relationships in a dataset (Han et al., 2011). Market basket analysis is a typical example of this process. The traditional co-location pattern mining is an extension of this analysis in geographical space (Shekhar & Huang, 2001). In this way, spatial patterns are transformed into transaction-type data. Then, association analysis methods such as Apriori and FP-Growth will be used to discover co-location patterns. The proposed methodology has two major differences from previous methods for identifying co-occurrence patterns. First, Boolean spatial features are considered to be spatial units as areal events. That is, they are unique entities, their locations are fixed over time, and there are no instances of them in the geographical space. Second, the neighborhood condition was ignored. In other words, every two instances of spatial units satisfy the spatial neighbor relationship. In this way, any spatial units that emerge as an event can participate in the pattern. Based on the proposed framework, basic concepts are explained as follows.

Definition 1. Let the whole study area S is composed of m distinct spatial units, denoted as $S = \{s_1, s_2, \ldots, s_m\}$, and the time T is divided into n timespans, denoted as $T = \{\Delta t_1, \Delta t_2, \ldots, \Delta t_n\}$.

Definition 2. Considering each timespan, if a spatial unit experiences the conditions where an event has occurred, it is called an instance of that spatial unit.

Definition 3. An itemset (pattern) C ($C \subseteq S$) is a subset of spatial units whose instances emerge as events throughout the study area at a particular timespan Δt_i. The size of a pattern is the number of spatial units in it.

Definition 4. To assess the significance of the obtained rules, the following metrics are calculated: support (Eq. 3), confidence (Eq. 4), and lift (Eq. 5). Suppose itemset C, where $A \subset C$, $B \subset C$, and $A \cap B = \emptyset$. The support is a criterion indicating the probability of the simultaneous presence of A and B in rule A→B, and the confidence shows the accuracy of the generated rule of the Apriori Algorithm. The rules that satisfy minimum support and minimum confidence thresholds were considered as strong associated relations. The lift value is used to assess the goodness of the rule. It measures the correlation between two sets of spatial units so that if the lift value is smaller than one, the two sets are negatively correlated; otherwise, both sets are positively correlated. In addition, the values equal to one express independency between two sets with no correlation (Mirhashemi & Mirzaei, 2021).

$$\text{support}(A \Rightarrow B) = P(A \cup B) \tag{3}$$

$$\text{confidence}(A \Rightarrow B) = P(B|A) \tag{4}$$

$$\text{lift}(A \Rightarrow B) = \frac{P(B|A)}{P(A)} \tag{5}$$

Definition 5. To investigate spatial association relationships and rules, the Apriori algorithm (Agrawal & Srikant, 1994) is implemented on the generated itemsets according to definition 3. This algorithm is based on the Apriori property stating "*all nonempty subsets of a frequent itemset must also be frequent*" (Han et al., 2011). Apriori is a frequent pattern mining approach. It is based on a level-wise search, where k-itemsets are used to explore (k +1)-itemsets. First, the set of frequent 1-itemsets is determined by scanning all of the itemsets to count the overall incidence of each spatial unit in which events have occurred, as C_1, and collecting those that satisfy minimum support, as L_1. Next, $L1$ is used to find the set of frequent 2-itemsets, as L_2. L_2 is used to find L_3, and so on, until no more frequent k-itemsets can be found. To discover L_k for $k \geq 2$, a two-step process is followed, including *join* and *prune*. In the *join* step, a set of candidate k-itemsets is generated by joining L_{k-1} with itself, denoted as C_k. Then, during the *prune* step, any member of C_k that does not meet the minimum support will be removed from C_k. Thus, L_k is determined.

2.3.3. Events recognition

To better understand the behavior of the COVID-19 outbreak, it is necessary to determine the status and activation time of each unit, in order to know where and when the events occurred. Beacause the rate of cases in each ZCTA is typically non-zero, and almost all ZCTAs (units) have had COVID-19 confirmed cases in each week during the period, a significant change in the rates can be considered as an event occurrence in each unit. In this regard, two scenarios were considered. In the first scenario (SC1), the changes between the rate of cases in two consecutive weeks were investigated (Δt = one week). In the second scenario (SC2), the changes between the rate of cases for three consecutive weeks were investigated (Δt = two weeks). There are ZCTAs in the city that are likely to be in high-risk clusters of the disease and have the highest rates of cases for multiple weeks. However, what plays a decisive role in identifying an event is the significant change in the area, regardless of whether these changes are located in outbreak clusters or not. These two senarios are considered to account for incubation and contagious periods, as well as the median time between the exposure and onset of symptoms which were estimated to be between 1 to 14 days (Wassie et al. (2020); (Elias et al., 2021); ABC-News (2022)). The combination of the results of two scenarios provides a broader picture of the association relations between different city areas. In each scenario, considering the upward or downward trend of the outbreak, the changes between the two rates in Δt may be positive or negative. Accordingly, both positive and negative differences were considered to investigate interactions and association relations more precisely. These changes are simplified as positive and negative events, where the absolute values are significant and more than predefined thresholds. The advantage of this approach is

that both increasing and decreasing trends are studied in the areas; therefore, areas that behave similarly to the critical fluctuations in an epidemic can be accurately identified.

In this regard, the changes between the rate of cases in both scenarios (considering two different Δt) were calculated. The summary of resulting data and their distribution in each scenario are shown in Table 2 and Fig. 4.

As Fig. 4 depicts, the distribution of data is approximately bell-shaped, with most of the data clustered around the mean. Meanwhile, as shown in Fig. 2, the trend line grew very sharply in the last four weeks of the period, which skewed data slightly. However, despite the skewed distribution of the rate of cases, the growth of rate in both scenarios, weekly and biweekly basis of Δt, approximately follows a normal distribution. Therefore, the lower (first) and upper (third) quartiles of data in each scenario were selected as the thresholds. Regarding the data distribution, the lower quartile properly shows areas with higher negative growth, whereas the upper quartile indicates areas experiencing higher positive change in the rate of cases. Therefore, using this approach, areas that are statistically significant can be identified. Further, in each scenario, two situations were considered: positive events and negative events, making a total of four test cases (Table 3).

2.4. Proposed map visualization

As a rule of thumb, the extracted rules from the association rule mining techniques are shown in the text form, as A → B. Moreover, the huge volume of data probably leads to numerous generated rules in this form. Therefore, understanding or interpreting the results would be difficult and challenging. Since the proposed model is an adoption of the traditional association rule mining, it inherits these drawbacks. The generated association rules and relations from the proposed model may be too numerous to interpret. In addition, the interpretation of rules in the form of A → B may have difficulties in practice. In general, representing data in charts and maps makes the data understanding feasible and allows us to do it in much less time, especially when there is a large volume of data. To address this issue, a map visualization approach is introduced. This approach is adopted from the *Apriori Property*. Hence, all the significant rules of size 2, containing 2 spatial units, take part in the *join* step of the algorithm to generate candidate rules of size >2. In this regard, the proposed method represents association rules of size 2. This visualization approach is composed of a directional graph, in which each spatial unit is considered a node and If-Then relations to be mapped as directional edges. The length of each edge is the Euclidean distance between two spatial units. It means the rule "If *spatial_unit_A*, then *spatial_unit_B*" leads to an edge from *spatial_unit_A* to *spatial_unit_B*. In addition, the edges are colored by their Euclidean distance. The longer the distance, the darker the color. The size and the hue of each node are graphically determined by the number of edges pointing out of the node or the number of edges pointing to the node. In general, this approach provides a proper way to reflect causal relationships between spatial units that are considered as ZCTAs in this study. If the number of edges pointing out of the node is considered, the darker and bigger node shows the more significant impact of that node/ZCTA on the other nodes/ZCTAs. On the other hand, considering the number of edges pointing to the node, the darker and bigger node indicates node/ZCTA that have been more affected by the other nodes/ZCTAs.

3. Results and discussions

3.1. Global spatial association

Fig. 5 displays the Z-score and Moran's Index values of the global Moran's I. Overall, the p-values are 0 or so small that they can be considered as 0, except the fifth week ending on September 5, 2020, for which the p-value is 0.1. Statistical tests conducted on the rate of cases are indicative of relatively strong autocorrelation and significant clustering in most weeks, most significantly during the week 13th to the week 41st, as well as from the week 63rd, beginning of the sharp growth of rates in the city, to the end. The maximum and minimum values of Moran's I and Z-score take place in the 72nd week (0.78 and 20.77) and 5th week (0.09 and 2.57), respectively. Having Moran's Index values positive and close to +1, high-rate areas are close to each other, while areas with lower cases rate cluster near other low values, generally. The clustered spatial structure of the city implies the existence of associated pairs of adjacent locations. This structure is also probable for the spatial distribution of infectious diseases such as COVID-19 due to the contagious nature of the disease. Nevertheless, there are several underlying factors that have spatially varying effects on their propagation. However, there might be ZCTAs that are not neighbors and autocorrelated but are associated with each other. The following section is conducted to find such areas.

3.2. Modeling results

In this study, ZIP Code areas are considered spatial units of the proposed model (Section 3.2.2). Considering each scenario, the positive and negative events were detected in each ZIP Code area. Taking each Δt into account, the set of ZCTAs where the events occurred were identified as itemset of that period. After that, the proposed model was applied on all four test cases (Section 3.2.3). The minimum support and confidence values were set as 21% and 75% for SC1 and as 22% and 75% for SC2, respectively. In association rules mining, these thresholds are typically set subjectively, determining whether or not a pattern is prevalent, and consequently, a generated rule is interesting. Having a lower threshold may cause redundant rules, whereas setting a higher threshold may lead to the loss of rare relationships (Bao & Wang, 2019). In our case, the examination and implementation of the model suggest that the selected thresholds were the best for the purpose of the study. Fig. 6 reveals the basic steps of the proposed approach in which the diffusion of COVID-19 in NYC was simplified. In step 1, the changes between cases rate in each Δt across ZCTAs of NYC were calculated. Taking each scenario into account, the ZIP Code areas satisfying the conditions of event occurrence, the brownish colors in Fig. 6, were recognized in step 2. After that, the itemset of ZCTA instances in each timespan is collected. In the next step, frequent patterns of events in NYC were found using Apriori algorithm. Finally, the spatiotemporal association structure of the outbreak was discovered using generated rules.

In SC1, the model generates 36 rules for negative events (Test1) and 151 rules for positive events (Test2), whereas, in SC2, the model detected 276 rules for negative events (Test3) and 4389 rules for positive events (Test4). These results are evidence of significant association relations between ZCTAs that are not necessarily in their adjacency. In both scenarios, negative events generated fewer rules in comparison to positive events, confirming the heterogeneity in the spatial structure of the city with respect to the disease reduction, intervention, and control. Unsurprisingly, this finding also demonstrates that COVID-19 is more

Table 2
Summary of data considering each scenario.

SCENARIO	MIN.	1ST QUARTILE	MEDIAN	MEAN	3RD QUARTILE	MAX.
SC1	-1406.04	-21.11	3.18	38.32	32.29	2883.42
SC2	-4841.77	-30.02	5.25	38.04	51.31	4088.89

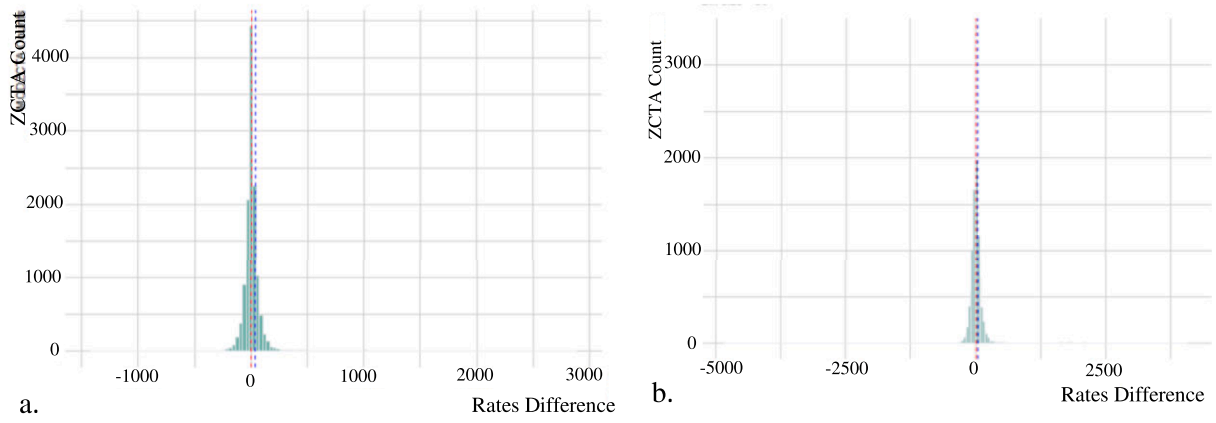


Fig. 4. . Histograms of changes between rate of cases considering two scenarios of Δt: a. SC1 (Δt = one week), and b. SC2, (Δt = two weeks). The red lines show the median values and the blue ones indicate the mean values of data .

Table 3
Details of four test cases in the study.

TEST NO.	Test-1	Test-2	Test-3	Test-4
EVENT_TYPE	Negative	Positive	Negative	Positive
SCENARIO	SC1	SC1	SC2	SC2

difficult to control than to spread. COVID-19, as an infectious disease, tends to diffuse into the area as the virus is seeded. Because of their proximity, autocorrelation, and social and cultural linkages, places that are adjacent to one another may be more likely to have similar infection rates and, consequently, shape spatial clusters (Arthur et al., 2017; Cordes & Castro, 2020). These reasons lead to more obtained association rules in positive events.

Another finding that emerged from this study is that NYC reveals more associations in the second scenario (SC2). This means that when the differences in the cases rates were considered between three consecutive weeks, more ZCTAs with similar behavior towards the increase or decrease of rates were identified. The likely reasons may be the average time of the transmission process, infectivity, the onset of symptoms, and the period that the patient can carry the disease. Table 4 presents the top 20 strongest rules ordered by their support, confidence and lift, respectively. Considering support values, the most discovered associated situations are ranked as *Test-4, Test-3, Test-2, Test-1*. As it shows, most of the determined rules involve adjacent ZCTAs. According to the results of the global Moran's I (Section 4.1), the spatial distribution of COVID-19 cases was clustered generally. Therefore, it is predictable that neighboring areas have association relationships. However, the discovered rules regarding for spatial structure of the city towards the trend of the outbreak are very interesting. Because they clarify implicit interdependence relationships between ZCTAs, which are not close together, and thus are not affected by adjacency relations such as autocorrelation.

Taking the cumulative weekly rates of cases into account (Fig. 2), the discovered rules are classified into four categories: low to low, low to high, high to low, and high to high. Since the nature of the diffusion of infectious disease, high to high and high to low, in adjacent areas, are excepted. High-rate areas usually emerge as the epicenters in local scales and function as determinants of rising the infections in neighboring areas. The other generated strong rules are critical and desirable.

ZIP Code numbers from 10,301 to 10,314 are located in Staten Island. As it is mentioned in Section 2.1, compared to other boroughs, this borough experienced the most severe outbreaks. Expectedly, most of the discovered rules are related to the ZCTAs of this borough. In addition to the rules involving ZIP Codes of this borough on both sides, other rules are as follows in each situation. For ease in writing and discussing the results, any ZCTA is written as "Z"+ code.

Test-1 (SC1- Negative Events): [11004] → [10006] is a very interesting result. Z11004 is located in the easternmost part of the city, while Z10006 is in the western part. Another important point is a major

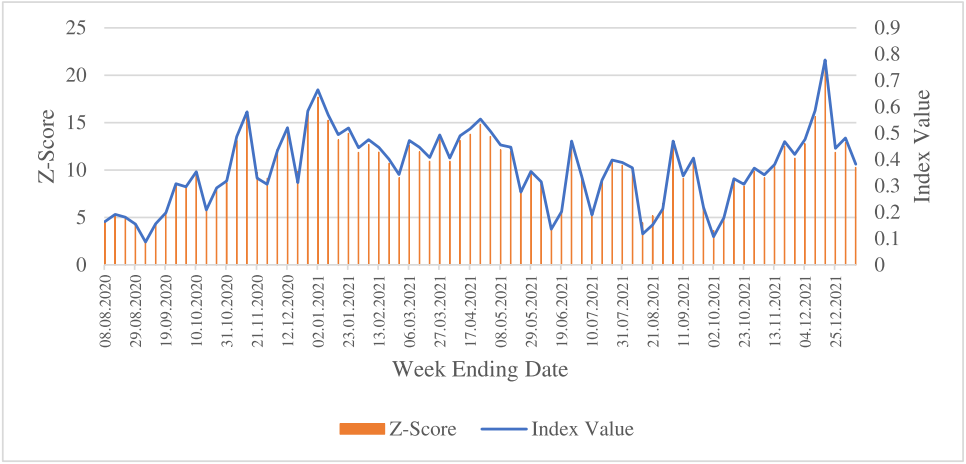


Fig. 5. Weekly results of global Moran's I statistic and Z-score on the rate of cases of COVID-19 at the ZCTA level.

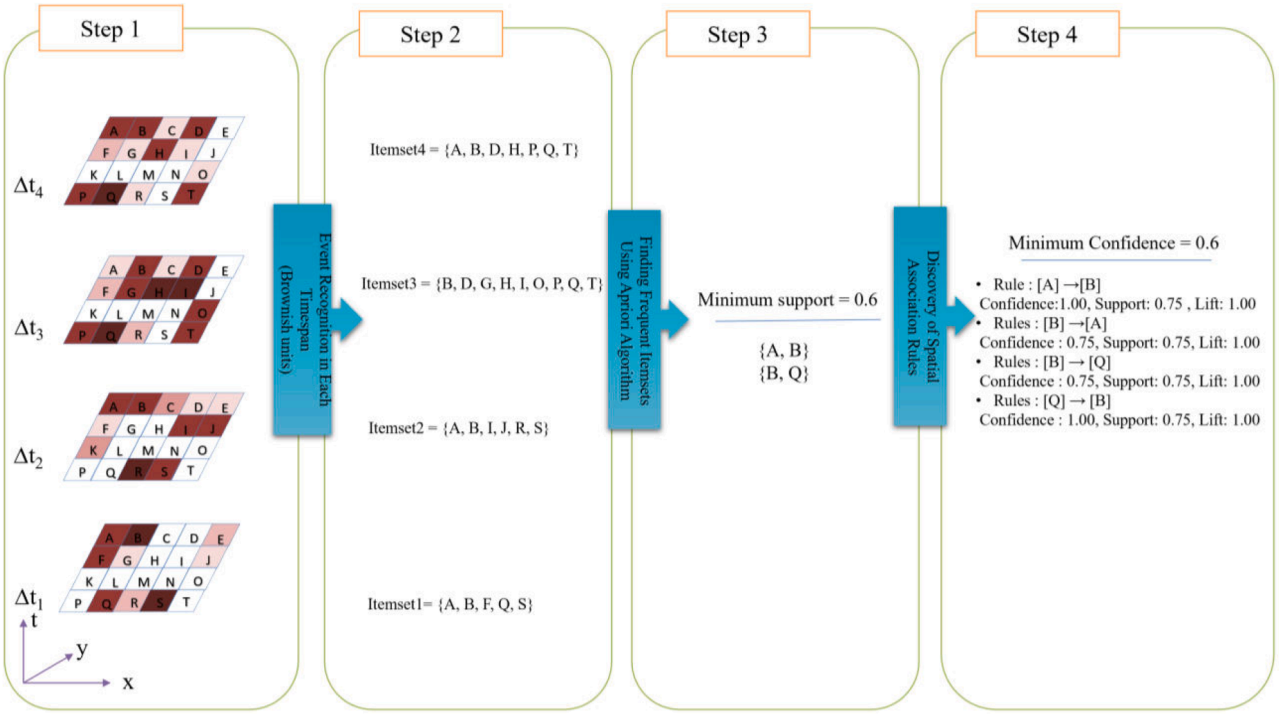


Fig. 6. Basic steps of the proposed approach.

difference in the total rate of cases between the two ZCTAs (Fig. 2). These values are 14,243 and 24,070, for Z11004 and Z10006, respectively (a low to high rule). [10029] → [11239] and [11370] → [10464] are other strong rules, involving four boroughs except Staten Island. They are low to high, and high to low rule types, respectively. In addition to these rules, the other rules involving Staten Island ZIP Codes show interactions between these areas.

Test-2 (SC1- Positive Events): Considering [11414] → [10282], Z11414 is in southern part of Queens with 23,146 cumulative value of cases rate. Z10282 is in Manhattan and its cumulative value was 13263. These two areas are distant and the higher cumulative cases rate area affects the lower one (High to low). In [11222] → [10005], Z11222 is in the north of Brooklyn and Z10005 is in the south of Manhattan; their cumulative values were 23370 and 21548, respectively. They have a common water border and the rule type is high to low. [11103] → [11414] is a rule from the center part of Queens to its southern part. The cumulative value of Z11103 was 25,517. So, it is also a high to low rule type.

Test-3 (SC2- Negative Events): [10012] → [10004] shows the association between close areas in the southern part of Manhattan. [10460] → [11422] is a distant high to low relation from The Bronx to Queens. The inverse of this rule is also a strong one. So, the relation is bilateral, as [10460] ↔ [11422]. The strong bilateral rules are remarkable, because they imply strong associations and interactions between areas. [11223] ↔ [11230] is an inter-borough rule which relates to areas in Brooklyn with almost the same cumulative values (23461 and 23373, respectively). [11419] ↔ [11421], [11370] → [11422], and [11434] → [11422] are indications of interdependence in near areas in Queens. [11214] → [11421], is a low to high rule from the southern part of Brooklyn to the western part of Queens. [11378] → [10312] is a high-value area of Queens (22930) to a higher value (29417) in Staten Island (high to high).

Test-4 (SC2- Positive Events): [11103] → [10307] and [10308] ↔ [11103] show associations between the northern part of Queens and Staten Island. The cumulative value of Z11103, Z10307, and Z10308 are 25517, 29123, and 27460, respectively. Therefore, they are high-to-high rules. Other top generated rules involve ZIP Code areas of Staten Island, and thus they are all high-to-high types. Almost all association rules are bilateral.

The above-mentioned rules show that in most cases, rules relate the regions with a higher cumulative rate of cases to the regions with lower cumulative rates. However, there were instances where the rules connected regions with lower rates to higher rates and even instances with associations between regions with almost the same number of cumulative rates. In addition, most of the association relations in all four tests belonged to ZIP Code areas that are distant and not adjacent to each other. Among all tests, there are areas associated to Staten Island. It is a significant result that this borough became the main driver of the outbreak within NYC.

To explore the finding rules better, the common rules in different tests are presented in Table 5. As it depicts, the common association rules are [10301] → [10312], [10312] → [10301], and [10304] → [10309]. These areas are located in Staten Island, but they are not in close proximity to each other. As mentioned previously, COVID-19 has been more prevalent in Staten Island compared to the other boroughs in NYC. This result addresses significant interdependence and association relations in this island, indicating that human mobilities and that human mobilities and other determinants affecting the behavior of the COVID-19 act similarly there. Although the effect of population is adjusted in calculating the rate of cases, another likely reason for this is the smaller population in this borough compared to the other boroughs. The higher population may lead to an increase in human mobility. This mobility interferes with spatial processes affecting the spread of the disease, and leads to the role of the other determinants being impaired. So, in this borough, the influence of the other underlying processes is more robust.

Beyond these findings, the obtained rules were visualized for all four tests using the proposed map visualization approach in Figs. 7 and 8. This approach is based on spatiotemporal association relations results, clarifying the causal relationships between ZCTAs. For the ease of measuring the influence of a ZIP Code area on the prevalence of the COVID-19 and determining how much it is affected by other infected ZIP Code areas, two maps were created for each test: 1) showing the number of edges pointing out of any ZCTA (*causality map*), and 2) representing the number of edges pointing to it (*affected map*). Therefore, novel maps

Table 4
Top 20 generated associated rules from each scenario.

Rule NO.	Test-1 (SC1- Negative Events)				Test-2 (SC1- Positive Events)			
	Rule	Support	Confidence	Lift	Rule	Support	Confidence	Lift
1	[10312] → [10308]	0.26	0.90	2.39	[11414] → [10282]	0.26	0.79	1.95
2	[10306] → [10310]	0.24	0.82	2.16	[10301] → [10309]	0.26	0.86	2.28
3	[10309] → [10307]	0.24	0.86	2.19	[10304] → [10309]	0.26	0.83	2.18
4	[11004] → [10006]	0.23	0.77	1.68	[10304] → [10312]	0.26	0.83	2.66
5	[10029] → [11239]	0.23	0.81	2.30	[10312] → [10304]	0.26	0.83	2.66
6	[10301] → [10308]	0.23	0.89	2.36	[10306] → [10309]	0.26	0.86	2.28
7	[10304] → [10308]	0.23	0.81	2.14	[11414] → [10307]	0.26	0.79	2.09
8	[10452] → [10307]	0.23	0.85	2.17	[10312] → [10309]	0.26	0.83	2.18
9	[10312] → [10310]	0.23	0.81	2.14	[11235] → [10309]	0.26	0.83	2.18
10	[10462] → [10310]	0.23	0.81	2.14	[11222] → [10005]	0.24	0.78	2.00
11	[11369] → [10310]	0.23	0.81	2.14	[10301] → [10304]	0.24	0.82	2.63
12	[10301, 10312] → [10308]	0.22	1.00	2.64	[10304] → [10301]	0.24	0.78	2.63
13	[10301, 10308] → [10312]	0.22	0.94	3.32	[10301] → [10312]	0.24	0.82	2.63
14	[10310, 10312] → [10308]	0.22	0.94	2.49	[10312] → [10301]	0.24	0.78	2.63
15	[10301] → [10312]	0.22	0.84	2.97	[10304] → [10307]	0.24	0.78	2.07
16	[11436] → [10310]	0.22	0.84	2.23	[10306] → [10307]	0.24	0.82	2.16
17	[10465] → [10464]	0.22	0.84	2.08	[10312] → [10307]	0.24	0.78	2.07
18	[11370] → [10464]	0.22	0.84	2.08	[10310] → [10309]	0.24	0.75	1.98
19	[10301] → [10308, 10312]	0.22	0.84	3.28	[11229] → [10309]	0.24	0.95	2.50
20	[10308, 10312] → [10301]	0.22	0.84	3.28	[11103] → [11414]	0.24	0.82	2.52
	Test-3 (SC2- Negative Events)				**Test-4 (SC2- Positive Events)**			
	Rule	Support	Confidence	Lift	Rule	Support	Confidence	Lift
1	[10308] → [10312]	0.27	0.91	2.80	[10307] → [10309]	0.31	0.79	2.10
2	[10312] → [10308]	0.27	0.83	2.80	[10309] → [10307]	0.31	0.82	2.10
3	[10312] → [10314]	0.27	0.83	2.94	[10306] → [10307]	0.30	0.92	2.34
4	[10314] → [10312]	0.27	0.95	2.94	[10307] → [10306]	0.30	0.76	2.34
5	[10012] → [10004]	0.26	0.79	1.83	[10306] → [10309]	0.30	0.92	2.42
6	[10460] → [11422]	0.26	0.90	2.91	[10309] → [10306]	0.30	0.79	2.42
7	[11422] → [10460]	0.26	0.83	2.91	[10308] → [10309]	0.30	0.81	2.15
8	[11223] → [11230]	0.26	0.79	2.66	[10309] → [10308]	0.30	0.79	2.15
9	[11230] → [11223]	0.26	0.86	2.66	[10301] → [10304]	0.28	0.88	2.59
10	[10301, 10308] → [10312]	0.24	1.00	3.08	[10304] → [10301]	0.28	0.84	2.59
11	[10301, 10312] → [10308]	0.24	1.00	3.36	[10306] → [10308]	0.28	0.88	2.40
12	[11419] → [11421]	0.24	0.95	3.34	[10308] → [10306]	0.28	0.78	2.40
13	[11378] → [10312]	0.24	0.90	2.78	[11103] → [10307]	0.28	0.88	2.23
14	[11370] → [10473]	0.24	0.90	3.03	[10308] → [11103]	0.28	0.78	2.40
15	[11214] → [11421]	0.24	0.90	3.17	[11103] → [10308]	0.28	0.88	2.40
16	[11370] → [11422]	0.24	0.90	2.90	[10306] → [10307, 10309]	0.28	0.88	2.82
17	[11434] → [11422]	0.24	0.90	2.90	[10309] → [10306, 10307]	0.28	0.75	2.52
18	[10308, 10312] → [10301]	0.24	0.90	3.17	[10306, 10307] → [10309]	0.28	0.95	2.52
19	[10308] → [10301, 10312]	0.24	0.82	3.36	[10306, 10309] → [10307]	0.28	0.95	2.44
20	[11421] → [11419]	0.24	0.86	3.34	[10307, 10309] → [10306]	0.28	0.91	2.82

of the spatial inequality in COVID-19 considering the participation of any ZCTA in association relations were drawn. These maps provide noticeable information that could not be understood from the rules in text form alone. They can reflect the influential extent of each ZCTA in association relations. In this context, any ZCTA is considered a node. Taking the *causality map* into account, the greater number of edges pointing out of the nodes shows its ZCTA has more influence over the prevalence of COVID-19 in NYC. Moreover, considering the *affected map,* the more edges pointing to the nodes, the more affected it is. In both map types, as the number of edges increases, the circles of each relating node become larger and darker, demonstrating the stronger interdependence relationships of their ZCTAs with the other areas.

As is depicted in Fig. 7, in "*Test1_SC1, Negative Events*", there were some dispersed nodes in both "*affected map*" and "*causality map*" mainly concentrated on Staten Island. In addition, considering the *causality map,* Z10464 in The Bronx and Z10029 in Manhattan and regarding the *affected map,* Z10006 in Manhattan and Z10464 in The Bronx were relatively significant.

Considering "*Test2_SC1, Positive Events*", the distribution of causative nodes was more homogeneous than affected nodes. Further, Z10309 and Z10307 in Staten Island, and Z11414 in Queens emerged as affected centers in the southern part of the city.

As positive events investigate ZCTAs' behavior towards the increase of COVID-19 cases rate, it clarifies increased incidence of the COVID-19 in these ZIP Code areas is under the influence of the other regions of the city that have edges pointing to them. Additionally, the darker edges are those indicating association relations in far distances. As Fig. 7 shows, most of the edges are dark and reflect the interdependence relation between the areas located at long distances from each other. Another interesting rule is [11414] → [10307], indicating a distant association from Queen to Staten Island, and also a relation between the two affected centers.

In Fig. 8, the discovered association rules concerning $\Delta t=2$ weeks (SC2) are represented. As it shows, the rules are more than found rules in SC1, especially when positive events are considered (*Test4*). Similar to SC1, the discovered rules in negative events are fewer than the positive events. In *Test4,* the concentration of rules was mainly on Staten Island, whereas the distribution was more balanced considering *Test3*. While, both "*affected map*" and "*causality map*" have the same significant nodes, approximately. Z10476, Z11421, Z11233, Z10305, and Z10312 in *Test3* and also Z11103, Z11235, and Z10306 in *Test4* are such examples. Considering *Test4- affected map,* the most of significant nodes are located in Staten Island. While, the prominent nodes in the other three maps are relatively dispersed.

In addition, the dominant implicit spatial structure of the associated areas is different in *Test3* and *Test4*, which are north-south and southwest-northeast, respectively. However, there is no a very specific geospatial direction in the implicit connecting links between associated

Table 5
Common recognized spatiotemporal association rules between two scenarios and two event type.

Test NO.	Test-1	Test-2		Test-3	
Test-2	[10301] → [10312]				
	[10304] → [10309]				
	[10312] → [10301]				
Test-3	[10301] → [10308]	[10301] → [10304]	[10305] → [10312]		
	[10301] → [10312]	[10304] → [10301]	[10306] → [10309]		
	[10312] → [10301]	[10301] → [10312]	[10310] → [10309]		
	[10304] → [10309]	[10312] → [10301]	[10312] → [10309]		
	[10309] → [10306]	[10305] → [10304]	[10314] → [10312]		
	[10312] → [10308]	[10304] → [10309]	[10460] → [10312]		
		[10304] → [10312]	[11223] → [11230]		
Test-4	[10301] → [10308]	[10301] → [10304]	[11223] → [11230]	[10301] → [10304]	[11235] → [10473]
	[10301] → [10312]	[10304] → [10301]	[11229] → [11235]	[10304] → [10301]	[10473] → [11421]
	[10312] → [10301]	[10301] → [10306]	[10301, 10304] → [10309]	[10301] → [10308]	[11379] → [11103]
	[10304] → [10308]	[10306] → [10301]	[10301, 10309] → [10304]	[10301] → [10312]	[11103] → [11421]
	[10304] → [10309]	[10301] → [10307]	[10304, 10309] → [10301]	[10312] → [10301]	[11421] → [11103]
	[10309] → [10306]	[10301] → [10309]	[10301, 10304] → [10312]	[10302] → [10309]	[11223] → [11230]
	[10306] → [10310]	[10301] → [10312]	[10301, 10312] → [10304]	[10460] → [10302]	[11230] → [11223]
	[10312] → [10306]	[10312] → [10301]	[10304, 10312] → [10301]	[10304] → [10305]	[11230] → [11224]
	[10309] → [10307]	[11209] → [10301]	[10301, 10306] → [10309]	[10305] → [10304]	[11224] → [11421]
	[10312] → [10308]	[11694] → [10302]	[10301, 10309] → [10306]	[10304] → [10309]	[11421] → [11224]
	[10312] → [10310]	[10303] → [10306]	[10306, 10309] → [10301]	[10304] → [10312]	[11229] → [11230]
	[10301, 10308] → [10312]	[10303] → [10309]	[10301, 10306] → [10312]	[10314] → [10304]	[11230] → [11229]
	[10301, 10312] → [10308]	[10305] → [10304]	[10301, 10312] → [10306]	[10305] → [10308]	[11235] → [11421]
	[10308, 10312] → [10301]	[10304] → [10307]	[10306, 10312] → [10301]	[10305] → [10309]	[11421] → [11235]
	[10312] → [10308, 10310]	[10304] → [10309]	[10301, 10307] → [10309]	[10305] → [10312]	[10301, 10308] → [10312]
	[10308, 10310] → [10312]	[10304] → [10312]	[10301, 10309] → [10307]	[10312] → [10305]	[10301, 10312] → [10308]
	[10308, 10312] → [10310]	[10312] → [10304]	[10302, 10307] → [10309]	[10305] → [10314]	[10308, 10312] → [10301]
	[10310, 10312] → [10308]	[11694] → [10304]	[10302, 10309] → [10307]	[10314] → [10305]	[10314] → [10308, 10312]
		[10305] → [10312]	[10304, 10307] → [10309]	[10306] → [10309]	[10308, 10312] → [10314]
		[10306] → [10307]	[10304, 10309] → [10307]	[10309] → [10306]	[10308, 10314] → [10312]
		[10306] → [10309]	[10307, 10309] → [10304]	[10306] → [10312]	[10312, 10314] → [10308]
		[10454] → [10306]	[10304, 10307] → [10312]	[10308] → [10309]	
Test NO.	**Test-1**	**Test-2**		**Test-3**	
Test-4		[11209] → [10306]	[10304, 10312] → [10307]	[10312] → [10308]	
		[10312] → [10307]	[10307, 10312] → [10304]	[10314] → [10308]	
		[10314] → [10307]	[10304, 10309] → [10312]	[10310] → [10309]	
		[11369] → [10307]	[10304, 10312] → [10309]	[10312] → [10309]	
		[11379] → [10307]	[10309, 10312] → [10304]	[10314] → [10309]	
		[11414] → [10307]	[10306, 10307] → [10309]	[10465] → [10309]	
		[11433] → [10307]	[10306, 10309] → [10307]	[11378] → [10309]	
		[10310] → [10309]	[10307, 10309] → [10306]	[11421] → [10309]	
		[10312] → [10309]	[10306, 10309] → [10312]	[10312] → [10314]	
		[11209] → [10309]	[10306, 10312] → [10309]	[10314] → [10312]	
		[11229] → [10309]	[10309, 10312] → [10306]	[10460] → [10312]	
		[11235] → [10309]	[10306, 10309] → [11235]	[11103] → [10312]	
		[11379] → [10309]	[11235, 10306] → [10309]	[11378] → [10312]	
		[10314] → [10310]	[11235, 10309] → [10306]	[10454] → [11103]	
		[10314] → [10312]	[10307, 10309] → [10312]	[10465] → [11235]	
		[10460] → [10312]	[10307, 10312] → [10309]	[11224] → [10470]	
		[11420] → [10312]	[10309, 10312] → [10307]	[11235] → [10470]	
		[11414] → [11103]	[11235, 10307] → [10309]	[11691] → [10470]	
		[11433] → [11103]	[11235, 10309] → [10307]	[10473] → [11235]	

areas in SC1. This kind of spatial structure can be inferred as the impact of directional parameters in the severity and weakness of disease spread, such as wind direction. Another likely reason for this is human mobility, as the main driver of the outbreak, that does not happen in a particular direction.

Considering both scenarios, it results that SC2 (Δt=2 weeks) better reveals association relations between ZCTAs and clarifies the spatial structure of the outbreak in NYC.

4. Conclusions

More than two years from the first officially reported case of COVID-19, while new variants have emerged, the world continues to fight the COVID19 pandemic. Investigating the spatial structure of the pandemic makes it feasible to determine its underlying driving processes. The existing methods mainly investigate the association relations considering randomness, dispersion, or clustered structure of the space in each time frame, while there might be associated areas that are not adjacent and have the same behavior towards the severity or weakness of the outbreak. In this context, the main aim of this study was to recognize association relations in spatiotemporal dynamics of infectious disease, in which their severity and spatial structure continuously change in space and time. To this goal, an event-based spatiotemporal model was proposed and applied on the weekly rate of cases of COVID-19 in New York City at the ZIP Code level.

The first advantage of the model was the exploration of COVID-19 as a spatiotemporal event considering two different periods scenarios in NYC (one and two weeks). Moreover, the event is derived from the trend of cases rate over a period, not the number of cases themselves. Hence, the effect of changes in underlying processes and human interactions is better understood. Second, the proposed model does not consider adjacency relationships and searches the whole study area to find associated ZCTAs. Also, the discovered relationships were independent of the cumulative values of cases rate or its cluster structure in the city. The third advantage was the integration of time into the association relations discovery. Since weekly positive and negative events were used in the

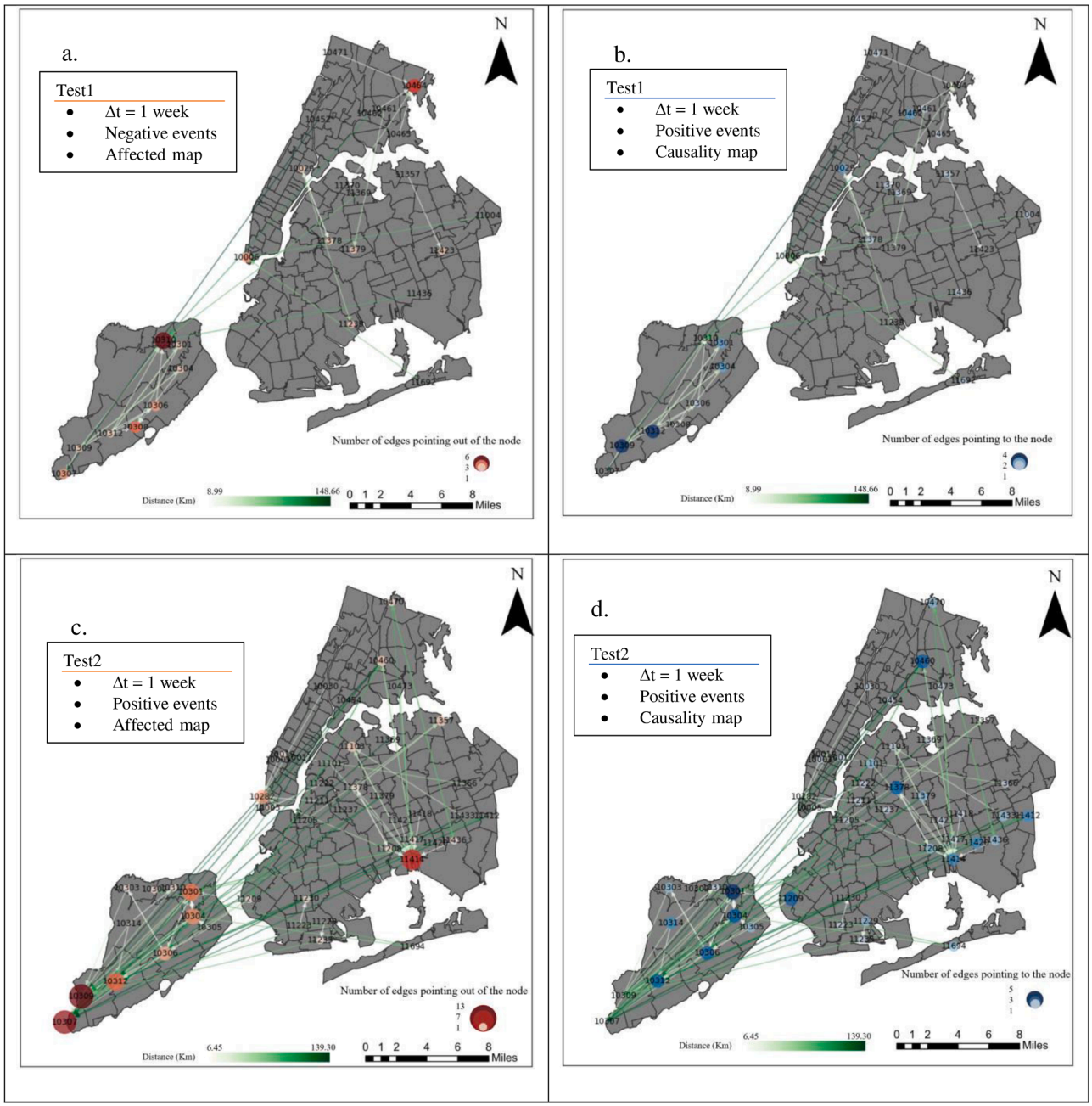


Fig. 7. (a) Test1, affected map. (b) Test1, causality map. (c) Test2, affected map. (d) Test2, causality map.

proposed model, it can detect associated areas that are seamlessly integrated into spatiotemporal complexities and dynamics of the diffusion. In other words, the proposed model sifts through the dynamics of the disease and clarifies associations in a heterogeneous space.

The results showed that in addition to interdependence relations between close areas, there exist remote areas having association relations. Moreover, there were three types of generated rules: (1) rules involving a high rate of cases areas, in particular, detected rules in Staten Island, (2) rules that were a combination of high and low rate of cases, mainly between Staten Island and other boroughs, (3) rules regarding low-rate areas. The latter was very interesting due to their long distances and the low value of cases rate. Regarding autocorrelation and human mobility, type 1 and 2 are very probable. The local epicenter affects other areas directly, whether high rate or low rate. At the same time, association relations among remote areas experiencing a lower rate of cases implies that the same underlying processes have been triggered consequently.

Different prevention and control policies and programs in several areas can vary the behavior of these regions about the spread of the disease and thus the event occurrence. Since this study was performed in a city, this factor had the same effect on all the results. Hence, the obtained results purely reflect interdependent relationships between various ZIP code areas, considering both the human mobility and the behavior towards the pandemic and also underlying factors. However, there are several factors involved in this association relations that should be investigated in detail in the future.

In addition to the proposed model, a map visualization approach was introduced that maps a massive number of rules into a single 2D map. This method provides a comprehensive picture of the discovered association patterns and rules. In this study, the method represented causative areas and those are affected by other areas concerning the diffusion of COVID-19 on the geographical surface of the map. In this way, distant

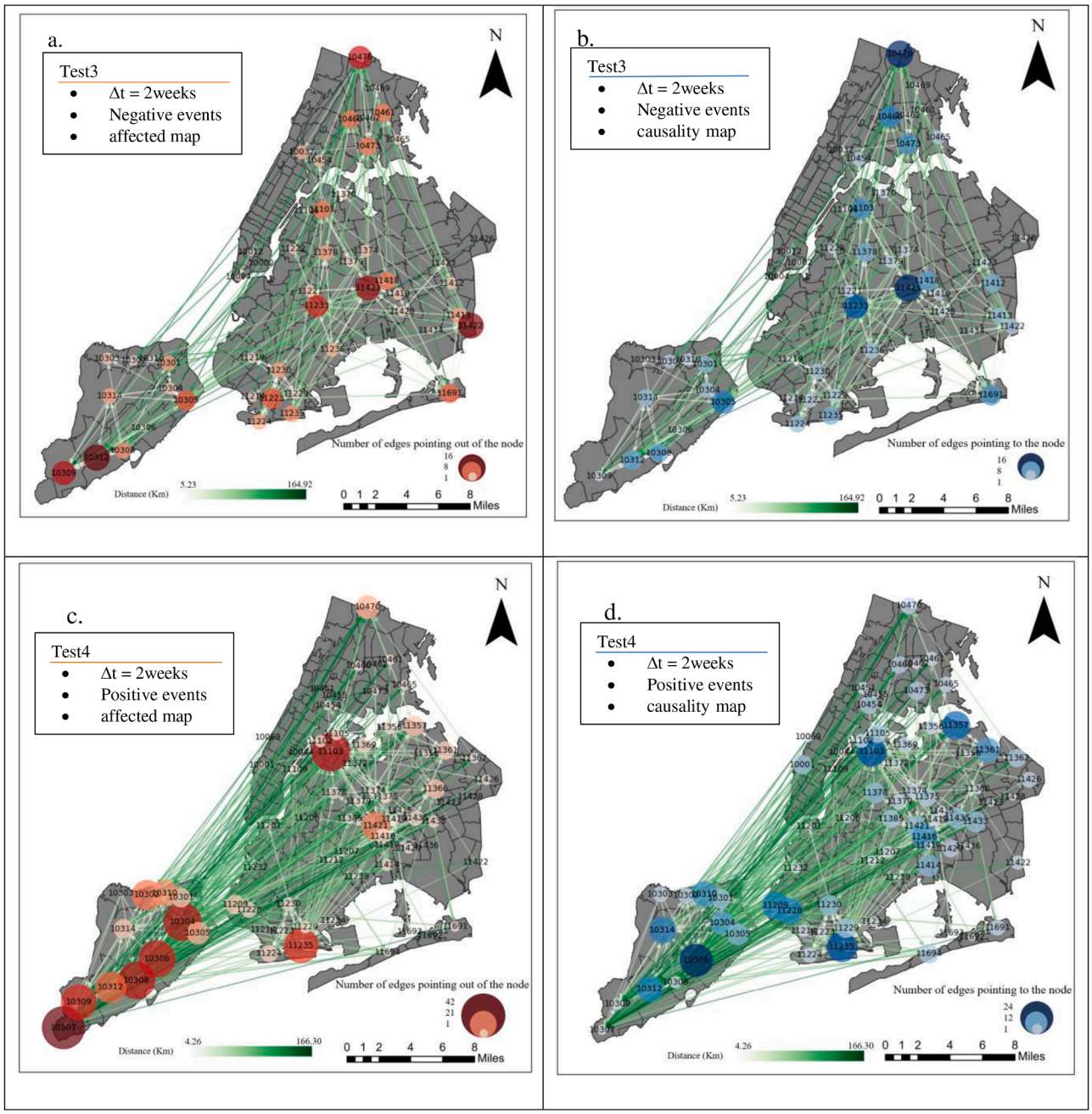


Fig. 8. (a) Test3, affected map. (b) Test3, causality map. (c) Test4, affected map. (d) Test4, causality map.

associated areas and critical ZCTAs are properly revealed. This achievement is not possible via traditional association rule mining results that have been shown in text format. The findings and proposed visualization approach of this study shed some light on how ZIP code areas contribute to the COVID-19 inequality. It will be helpful for governors and policy makers to adopt better local policies to overcome COVID-19.

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.

Data availability

The NYC COVID-19 Data was used in the article and the access link to the data has mentioned in the manuscript.

References

ABC-News, 2022. What's COVID's incubation period? Can Omicron spread if you're asymptomatic? Retrieved from https://www.abc.net.au/news/2022-01-07/covid-omicron-symptoms-incubation-period-infectious/100739260?utm_campaign=abc_news_web&utm_content=link&utm_medium=co%20ntent_shared&utm_source=abc_news_web. Accessed February, 4, 2022.

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In *Proceedings of the 20th international conference on very large data bases, VLDB. Citeseer* (pp. 487–499).

Akbari, M., Samadzadegan, F., & Weibel, R. (2015). A generic regional spatio-temporal co-occurrence pattern mining model: A case study for air pollution. *Journal of Geographical Systems, 17*, 249–274.

Alkhamis, M. A., Al Youha, S., Khajah, M. M., Haider, N. B., Alhardan, S., Nabeel, A., Al Mazeedi, S., & Al-Sabah, S. K. (2020). Spatiotemporal dynamics of the COVID-19 pandemic in the State of Kuwait. *International Journal of Infectious Diseases, 98*, 153–160.

Arthur, R. F., Gurley, E. S., Salje, H., Bloomfield, L. S., & Jones, J. H. (2017). Contact structure, mobility, environmental impact and behaviour: The importance of social forces to infectious disease dynamics and disease ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*, Article 20160454.

Baker, M. G., Peckham, T. K., & Seixas, N. S. (2020). Estimating the burden of United States workers exposed to infection or disease: A key factor in containing risk of COVID-19 infection. *PloS One, 15*, Article e0232452.

Bao, X., & Wang, L. (2019). A clique-based approach for co-location pattern mining. *Information Sciences, 490*, 244–264.

Cai, J., Deng, M., Liu, Q., He, Z., Tang, J., & Yang, X. (2019). Nonparametric significance test for discovery of network-constrained spatial co-location patterns. *Geographical Analysis, 51*, 3–22.

Chokshi, D. A., & Katz, M. H. (2020). Emerging lessons from COVID-19 response in New York City. *JAMA, 323*, 1996–1997.

Cordes, J., & Castro, M. C. (2020). Spatial analysis of COVID-19 clusters and contextual factors in New York City. *Spatial and Spatio-temporal Epidemiology, 34*, Article 100355.

Cuadros, D. F., Xiao, Y., Mukandavire, Z., Correa-Agudelo, E., Hernández, A., Kim, H., & MacKinnon, N. J. (2020). Spatiotemporal transmission dynamics of the COVID-19 pandemic and its impact on critical healthcare capacity. *Health & Place, 64*, Article 102404.

Dale, M. R., & Fortin, M.-J. (2014). *Spatial analysis: a guide for ecologists*. Cambridge University Press.

Das, A., Ghosh, S., Das, K., Basu, T., Dutta, I., & Das, M. (2021). Living environment matters: Unravelling the spatial clustering of COVID-19 hotspots in Kolkata megacity, India. *Sustainable Cities and Society, 65*, Article 102577.

Elias, C., Sekri, A., Leblanc, P., Cucherat, M., & Vanhems, P. (2021). The incubation period of COVID-19: A meta-analysis. *International Journal of Infectious Diseases, 104*, 708–710.

Feng, Y., Marchal, T., Sperry, T., & Yi, H. (2020). Influence of wind and relative humidity on the social distancing effectiveness to prevent COVID-19 airborne transmission: A numerical study. *Journal of Aerosol Science, 147*, Article 105585.

Greene, S. K., Peterson, E. R., Balan, D., Jones, L., Culp, G. M., Fine, A. D., & Kulldorff, M. (2021). Detecting COVID-19 clusters at high spatiotemporal resolution, New York City, New York, USA, June–July 2020. *Emerging Infectious Diseases, 27*, 1500.

Habibi, R., Alesheikh, A. A., Mohammadinia, A., & Sharif, M. (2017). An assessment of spatial pattern characterization of air pollution: A case study of CO and PM2. 5 in Tehran, Iran. *ISPRS International Journal of Geo-Information, 6*, 270.

Haining, R. P., & Haining, R. (2003). *Spatial data analysis: theory and practice*. Cambridge university press.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Haque, S. E., & Rahman, M. (2020). Association between temperature, humidity, and COVID-19 outbreaks in Bangladesh. *Environmental Science & Policy, 114*, 253–255.

He, Z., Deng, M., Cai, J., Xie, Z., Guan, Q., & Yang, C. (2020). Mining spatiotemporal association patterns from complex geographic phenomena. *International Journal of Geographical Information Science, 34*, 1162–1187.

Huang, Z. (2021). Spatiotemporal evolution patterns of the COVID-19 pandemic using space-time aggregation and spatial statistics: A global perspective. *ISPRS International Journal of Geo-Information, 10*, 519.

Islam, A., Sayeed, M. A., Rahman, M. K., Ferdous, J., Islam, S., & Hassan, M. M. (2021). Geospatial dynamics of COVID-19 clusters and hotspots in Bangladesh. *Transboundary and Emerging Diseases*.

Jaya, I., & Folmer, H. (2020). Bayesian spatiotemporal mapping of relative dengue disease risk in Bandung, Indonesia. *Journal of Geographical Systems, 22*, 105–142.

Lara-Reyna, J., Yaeger, K. A., Rossitto, C. P., Camara, D., Wedderburn, R., Ghatan, S., Bederson, J. B., & Margetis, K. (2020). Staying home"—early changes in patterns of neurotrauma in New York City during the COVID-19 pandemic. *World Neurosurgery, 143*, e344–e350.

Liu, L., Hu, T., Bao, S., Wu, H., Peng, Z., & Wang, R. (2021). The spatiotemporal interaction effect of COVID-19 transmission in the United States. *ISPRS International Journal of Geo-Information, 10*, 387.

Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., Yan, J., Niu, J., Zhou, J., & Luo, B. (2020). Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Science of the Total Environment, 724*, Article 138226.

Maroko, A. R., Nash, D., & Pavilonis, B. T. (2020). COVID-19 and inequity: A comparative spatial analysis of New York City and Chicago hot spots. *Journal of Urban Health, 97*, 461–470.

Menebo, M. M. (2020). Temperature and precipitation associate with COVID-19 new daily cases: A correlation study between weather and COVID-19 pandemic in Oslo, Norway. *Science of The Total Environment, 737*, Article 139659.

Mirhashemi, S. H., & Mirzaei, F. (2021). Extracting association rules from changes in aquifer drawdown in irrigation areas of Qazvin plain, Iran. *Groundwater for Sustainable Development, 12*, Article 100495.

Notari, A. (2021). Temperature dependence of COVID-19 transmission. *Science of The Total Environment, 763*, Article 144390.

NYC-Health, 2022. NYC COVID-19 Data. Retrieved from https://github.com/nychealth/coronavirus-data. Accessed February 21, 2022.

O'sullivan, D., & Unwin, D. (2003). *Geographic information analysis*. John Wiley & Sons.

Piret, J., & Boivin, G. (2021). Pandemics throughout history. *Frontiers in Microbiology*, 3594.

Prata, D. N., Rodrigues, W., & Bermejo, P. H. (2020). Temperature significantly changes COVID-19 transmission in (sub) tropical cities of Brazil. *Science of the Total Environment, 729*, Article 138862.

Rendana, M. (2020). Impact of the wind conditions on COVID-19 pandemic: A new insight for direction of the spread of the virus. *Urban Climate, 34*, Article 100680.

Shariati, M., Mesgari, T., Kasraee, M., & Jahangiri-Rad, M. (2020). Spatiotemporal analysis and hotspots detection of COVID-19 using geographic information system (March and April, 2020). *Journal of Environmental Health Science and Engineering, 18*, 1499–1507.

Shekhar, S., & Huang, Y. (2001). Discovering spatial co-location patterns: A summary of results. In *Proceedings of the international symposium on spatial and temporal databases* (pp. 236–256). Springer.

Srivastava, A. (2021). COVID-19 and air pollution and meteorology-an intricate relationship: A review. *Chemosphere, 263*, Article 128297.

State, N.Y., 2020. Governor cuomo signs the 'New York state on PAUSE' executive order.

Travaglio, M., Yu, Y., Popovic, R., Selley, L., Leal, N. S., & Martins, L. M. (2021). Links between air pollution and COVID-19 in England. *Environmental Pollution, 268*, Article 115859.

Wan, Y., & Zhou, J. (2008). KNFCOM-T: A k-nearest features-based co-location pattern mining algorithm for large spatial data sets by using T-trees. *International Journal of Business Intelligence and Data Mining, 3*, 375–389.

Wang, J., Tang, K., Feng, K., Lin, X., Lv, W., Chen, K., & Wang, F. (2021a). Impact of temperature and relative humidity on the transmission of COVID-19: A modelling study in China and the United States. *BMJ Open, 11*, Article e043863.

Wang, Y., Liu, Y., Struthers, J., & Lian, M. (2021b). Spatiotemporal characteristics of the COVID-19 epidemic in the United States. *Clinical Infectious Diseases, 72*, 643–651.

Wassie, G. T., Azene, A. G., Bantie, G. M., Dessie, G., & Aragaw, A. M. (2020). Incubation period of severe acute respiratory syndrome novel coronavirus 2 that causes coronavirus disease 2019: A systematic review and meta-analysis. *Current Therapeutic Research, 93*, Article 100607.

Xiong, Y., Wang, Y., Chen, F., & Zhu, M. (2020). Spatial statistics and influencing factors of the COVID-19 epidemic at both prefecture and county levels in Hubei Province, China. *International Journal of Environmental Research and Public Health, 17*, 3903.

Yang, T.-C., Kim, S., Zhao, Y., & Choi, S. W. E. (2021). Examining spatial inequality in COVID-19 positivity rates across New York City ZIP codes. *Health & Place, 69*, Article 102574.

Zangari, S., Hill, D. T., Charette, A. T., & Mirowsky, J. E. (2020). Air quality changes in New York City during the COVID-19 pandemic. *Science of the Total Environment, 742*, Article 140496.

Zhang, M., Wang, S., Hu, T., Fu, X., Wang, X., Hu, Y., Halloran, B., Li, Z., Cui, Y., & Liu, H. (2022). Human mobility and COVID-19 transmission: A systematic review and future directions. *Annals of GIS*, 1–14.

Zheng, A., Wang, T., & Li, X. (2021). Spatiotemporal characteristics and risk factors of the COVID-19 pandemic in New York State: Implication of future policies. *ISPRS International Journal of Geo-Information, 10*, 627.