Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Multi-scale causality analysis between COVID-19 cases and mobility level using ensemble empirical mode decomposition and causal decomposition

Jung-Hoon Cho, Dong-Kyu Kim, Eui-Jin Kim

Abstract

The global spread of the coronavirus disease 2019 (COVID-19) pandemic has affected the world in many ways. Due to the communicable nature of the disease, it is difficult to investigate the causal reason for the epidemic's spread sufficiently. This study comprehensively investigates the causal relationship between the spread of COVID-19 and mobility level on a multi time-scale and its influencing factors, by using ensemble empirical mode decomposition (EEMD) and the causal decomposition approach. Linear regression analysis investigates the significance and importance of the influential factors on the intrastate and interstate causal strength. The results of an EEMD analysis indicate that the mid-term and long-term domain portrays the macroscopic component of the states' mobility level and COVID-19 cases, which represents overall intrinsic characteristics. In particular, the mobility level is highly associated with the long-term variations of COVID-19 cases rather than short-term variations. Intrastate causality analysis identifies the significant effects of median age and political orientation on the causal strength at a specific time-scale, and some of them cannot be identified from the existing method. Interstate causality results show a negative association with the interstate distance and the positive one with the airline traffic in the long-term domain. Clustering analysis confirms that the states with the higher the gross domestic product and the more politically democratic tend to more adhere to social distancing. The findings of this study can provide practical implications to the policymakers that whether the social distancing policies are effectively working or not should be monitored by long-term trends of COVID-19 cases rather than short-term.

Article Info

Article history:
Received 23 September 2021
Received in revised form 11 January 2022
Available online 30 April 2022

Keywords:
COVID-19
Mobility
Ensemble empirical mode decomposition
Causal decomposition
Multi-scale causality analysis

1. Introduction

The recent coronavirus disease 2019 (COVID-19) pandemic has affected people’s daily lives in many ways. The United States has suffered from the pandemic by being one of the countries with many reported cases. Due to the nature of
Influential factors on the different time-scale. These multi-scale data can be analyzed by various signal decomposition methods to investigate how the relationship between social distancing policy and outbreaks of COVID-19 can be investigated; (ii) how the interstate characteristics and intrastate relationship can be considered in the analysis of the COVID-19 cases and social distancing.

The spread of infectious diseases and the travel behavior of people have in common a certain cyclicity. However, few studies have yet pointed out the mutual causal relationships between the spread of these diseases and periodic travel behavior. That aspect is crucial for clearly understanding the pandemic spread caused by the traveling people [5]. Signal decomposition analysis can help separate the time series for the number of confirmed COVID-19 cases and the number of trips made within states into intrinsic mode functions (IMFs) that represent intrinsic patterns on different time-scale. With the ensemble empirical mode decomposition (EEMD), a causal decomposition has merits when investigating the causality of the mutual causation system in different frequency domains [8]. The absolute and relative causal strengths derived from causal decomposition are also compared to find the causal relationships between the paired IMFs.

The purpose of this paper is to comprehensively investigate the spread of COVID-19 in different frequency domains (i.e., different time-scale) that have varied from the short-term to long-term influences by signal decomposition methods. Also, linear regression analysis is applied to identify the state-specific influential factors for intrastate causal strength between the number of COVID-19 cases and the number of daily trips (i.e., social distancing measure) and then compare their significance and importance. Further, interstate causality analysis is also performed to identify the mutual causation between different states for different time-scale and find the strength and direction of the causal relationship among the states. Lastly, clustering analysis is conducted to group the states by energy distribution of COVID-19 cases in each time-scale, and influential factors for those distributions is investigated. The findings of this study can provide practical implications to the policymakers: (i) how the relationship between social distancing policy and outbreaks of COVID-19 can be investigated; (ii) how the interstate characteristics and intrastate relationship can be considered in the analysis of the COVID-19 cases and social distancing.

The organization of this paper thus is as follows. The following section introduces previous work related to this study. The third section briefly introduces the methods applied to this particular study. The following section offers the detailed results of the analysis including data description and results of the interstate time-series analysis. The paper ends with a brief conclusion and offers outlooks for future research in the final section.

2. Literature review

Multiple statistical models have been developed to understand the spread of communicable diseases [9,10]. The susceptible–infectious–recovery (SIR) model and susceptible–exposed–infectious–recovery (SEIR) model are such models. However, these models have limitations when modeling the complex relationship with various factors [5,11]. To handle this issue, data-driven approaches are followed, for example, by applying the reinforcement learning (RL) algorithm to prioritize which people should be tested to increase the number of healthy people and using graph neural network to consider people as graph nodes for multi-graphs [12]. The mathematical SEIR model was also developed to analyze the ameliorating effects of social distancing mandates and the wearing of mask [5].

Due to the characteristics of superspreading of COVID-19, several studies were conducted to investigate the effects of travel restriction, interstate traffic control, and physical distancing. Lau et al. [13] reported that a strong linear correlation was found to exist between domestic air traffic volume and the spread of COVID-19 cases. The effects of COVID-19 on air transportation were analyzed based on the travel restriction measures and using the airline and customer satisfaction data [14]. In order to understand the timely intrinsic features of communicable disease, research has been conducted to decompose and interpret the time series using different time-scale [7,15–17]. Also, the societal response to the pandemic varied in terms of geographic, socioeconomic network, and governance [18].

A global disease transmission model was used to project the impact of travel restrictions on the national and international spread of the virus, and that model was found to have a modest delaying effect within the country [19]. In that sense, nonpharmaceutical measures reduced transmissibility by a maximum of 33% and without resorting to a strict lockdown strategy [20]. However, the increase in road traffic levels, the indication of contact between people, could lead to a risk of the COVID-19 spread [21]. Mobility and an epidemiology model using actual reported COVID-19 cases revealed that unconstrained mobility would significantly accelerate the global diffusion of the disease [22]. Similarly, the spread of COVID-19 decelerated the people’s trips of automobiles, public transits and air transportations [23–27], and modified their way of deciding their trips [28–30]. The correlation between the prevalence of the disease and local socio-economic conditions and turned out to indicate asymmetries in the level of urbanization, the population pyramid, racial distribution, and political orientation [31,32].

The spread of COVID-19 and the degree of social distancing are likely to be affected by the complex interaction of influential factors on the different time-scale. These multi-scale data can be analyzed by various signal decomposition methods.
methods, and each decomposition method has its own pros and cons according to the data conditions. For example, when the time-scale of data can be well-defined (i.e., stationary signals), Fourier analysis is suitable, and when we have the information on the data distribution, wavelet transform analysis tends to be effective. However, the above methods are not ideal for analyzing the non-stationary signal without prior information. Therefore, it is convincing to use data-adaptive methods such as empirical mode decomposition (EMD) and variational mode decomposition (VMD) useful for non-stationary signals [33,34]. The EMD is a fully data-driven method that extracts signal components through an adaptive filtering procedure. Among the variants of EMDs, EEMD is the most data-adaptive by adding white noise series to targeted data and can naturally separate scale without a priori subjective criterion selection [35].

EEMD is also useful when investigating the intrinsic multi-scale characteristics of various sources [36–38]. The IMFs obtained from the EEMD method were grouped into high-, medium-, and low-frequency components, which can represent the short-, medium-, and long-term variations of the original signal, respectively [34]. The IMFs extracted by EMD process well represent the local characteristics of the passenger flows [39].

Causal decomposition is the causality analysis method based on EEMD. It is known that the combination of various factors affects the transmission of infectious diseases in different frequency domains [7]. Therefore, if the effects of these factors could be analyzed by decomposing them into signals having different periods, the hidden characteristics caused by the interference of different periods could be disclosed more clearly. We thus applied the casual decomposition to reveal the unseen causal relationship by separating signals with different frequency domains. Other methods can be considered aside from the causal decomposition. For instance, Granger causality is superior in a linear stochastic situation where separability is guaranteed, but it is not proper to apply in a non-linear deterministic system. On the other hand, causal decomposition has the advantage of reflecting real-world data and phenomena based on instantaneous phase dependency between cause and effect, that is, oscillatory stochastic and deterministic mechanisms [8]. Several studies applied this technique to the change rate for the GDP timeseries between major countries on a different time-scale [40] and the investigation of Malaria epidemics [41]. The effect of social distancing and travel restrictions on the COVID-19 outbreaks has been investigated in the reviewed studies. Their efforts mainly focused on the influence of road or air traffic on the spread of COVID-19 or microscopic analysis of specific policies such as a lockdown. However, due to the difficulty of quantifying and decomposing the macroscopic data, most of the reviewed studies did not consider the macroscopic effect of COVID-19 cases and social distancing at the state-level. Therefore, EEMD and causal decomposition, which have been applied in analyzing other macroscopic time-series, can fill this research gap.

3. Methodology

Fig. 1 describes the overall research process and the data pipeline that was used.

3.1. Ensemble empirical mode decomposition (EEMD)

Empirical mode decomposition (EMD) was developed to decompose the non-linear and non-stationary signals into orthogonal sub-signals called intrinsic mode functions (IMFs) and trends, which represent the distinct time-scale [42]. The decomposed IMFs represent the short-term periodical patterns or part of long-term trend in each time-scale. Hilbert transform analysis was used to derive the instantaneous phase and frequency from the IMFs that were decomposed from EMD and each IMF represents the amplitude and frequency-modulated signal \( S(t) \) as noted below Eq. (1).

\[
S(t) = A(t) \cos(\phi(t))
\]

(1)

where \( S_H(t) = \frac{1}{\pi} \int \frac{S(t)}{t} dt' \) and instantaneous amplitude \( A(t) = \sqrt{S^2(t) + S_H^2(t)} \), phase \( \phi(t) = \arctan\left(\frac{S_H(t)}{3S(t)}\right) \), and instantaneous frequency \( \omega(t) = d\phi(t)/dt \). The original signal \( X(t) \) can be expressed as the summation of all IMFs and residual \( r \), as shown below Eq. (2).

\[
X(t) = \sum_{j=1}^{M} A_j(t) \exp\left(i \int \omega_j(t) dt\right) + r
\]

(2)

where \( M \) is the total number of IMFs, and \( A_j(t) \) and \( \omega_j(t) \) are the instantaneous amplitude and frequency of each IMF, respectively.

Ensemble EMD (EEMD) was developed to deal with the drawbacks of the original EMD [35]. EEMD is a noise-assisted data analysis technique used to extract the true signal from the data. These true IMF components \( S_j(t) \) are regarded as the mean of an ensemble of trials added by the white noise \( w^k(t) \) of the finite amplitude (Eq. (3)).

\[
S_j(t) = \sum_{k=1}^{N} \{S_j^k(t) + w^k(t)\}
\]

(3)

where \( S_j^k(t) \) is the \( k \)th trial of the \( j \)th IMF in the noise-added signal. The EEMD require two parameters: the amplitude of the white noise \( \varepsilon \) and the number of ensemble \( N \). These two parameters determine the standard deviation of error
Fig. 1. Research process for investigating intrastate and interstate causal strength of COVID-19 cases and the number of trips.

\( \varepsilon_n \) caused by the added noise, which is the difference between the input signal and the corresponding sum of IMFs as following Eq. (4):

\[
\varepsilon_n = \frac{\varepsilon}{\sqrt{N}} \tag{4}
\]

EEMD mainly uses the standard deviation of added white noise that ranges 0.1 to 0.4 [35]. This current paper applies 0.35 to it. The number of ensemble members is set to 1000, which should be enough to average-out the added noise.

3.2. Convergent cross mapping (CCM)

Convergent cross mapping (CCM) is developed to measure the causality that determines how well a historical record of \( Y \) can accurately predict the states of \( X \) [43]. It bases on the fact that same dynamical system shares a common attractor manifold \( M \), thus it could be used for estimating each other [44]. CCM detects the correspondence between the local neighborhood of the time delayed embedding shadow attractor manifold \( M_X \) and \( M_Y \), built from time-series variables \( X \) and \( Y \), respectively. Variable \( X \) could be said to causally influence variable \( Y \), if the state of variable \( Y \) can be used to predict the state of \( X \).

The estimated state of \( X \), \( \hat{X} \), is calculated as \( \sum_{j=1}^{m+1} w_j X_j \), where \( w_j \) denotes the weights, and \( m \) denotes given embedding dimension. The CCM detects the causation by following steps. The correlation coefficient of \( \rho_{X|Y} \) can be calculated as Eq. (5), where \( N \) denotes the sample size.

\[
\rho_{X|Y} = \frac{\sum \hat{X}_j X_j - \bar{X} \bar{X}_j}{\sqrt{\sum X^2 - N \bar{X}^2} \sqrt{\sum \hat{X}_j^2 - N \bar{X}^2}} \tag{5}
\]

The CCM coefficient from \( X \) to \( Y \) can be calculated vice versa, and the strength of causality between two directions can be determined by the difference of the coefficients.

3.3. Causal decomposition

The causal decomposition method was first proposed by Yang et al. [8]. Between \( j \)th IMF components of the two time series, 1 and 2, \( S_1(t) = A_1(t) \cos \phi_1(t) \), \( S_2(t) = A_2(t) \cos \phi_2(t) \), and the phase coherence can be calculated as \( \text{Coh}(S_1(t), S_2(t)) = \frac{1}{T} \int_0^T e^{i \Delta \phi_{12}(t)} dt \), where the instantaneous phase difference is simply expressed as \( \Delta \phi_{12}(t) = \)
\( \phi_j(t) \) \( \phi_2(t) \). Phase coherence allows the instantaneous phase dependency to be calculated regardless of the effect of any time lag between them. Causal decomposition analysis was originally based on Galilei’s principle: Re-decompose the time series after reducing a specific IMF. If the source time series affects the target IMF, subtracting the IMF from the original time series could lead to redistributing the phase dynamics into the emptied space of the corresponding IMF. The phase coherence between two IMFs can be regarded as the coordinates in multidimensional space, and the variance-weighted Euclidean distance is then calculated to quantify the causal strength [8]. Absolute causal strengths are defined as shown below (Eqs. (6) and (7)), and their range are between 0 and 1.

\[
D(S_{ij} \rightarrow S_{2j}) = \left( \sum_{j=1}^{M} W_j \left[ \text{Coh}(S_{ij}, S_{2j}) - \text{Coh}(S_{ij}^*, S_{2j}^*) \right]^2 \right)^{1/2} (6)
\]

\[
D(S_{2j} \rightarrow S_{ij}) = \left( \sum_{j=1}^{M} W_j \left[ \text{Coh}(S_{ij}, S_{2j}) - \text{Coh}(S_{ij}^*, S_{2j}^*) \right]^2 \right)^{1/2} (7)
\]

where \( S_{ij} \) and \( S_{2j} \) stand for re-decomposed time series, removing the corresponding IMF in the target time series and the weight \( W_j = (\text{Var}_{ij} \times \text{Var}_{2j}) / \sum_{j=1}^{M} (\text{Var}_{ij} \times \text{Var}_{2j}) \).

The relative causal strength between the IMFs is defined as the ratio of absolute cause strengths (Eqs. (8) and (9)).

\[
C(S_{ij} \rightarrow S_{2j}) = D(S_{ij} \rightarrow S_{2j}) / \left[ D(S_{ij} \rightarrow S_{2j}) + D(S_{2j} \rightarrow S_{ij}) \right] (8)
\]

\[
C(S_{2j} \rightarrow S_{ij}) = D(S_{2j} \rightarrow S_{ij}) / \left[ D(S_{ij} \rightarrow S_{2j}) + D(S_{2j} \rightarrow S_{ij}) \right] (9)
\]

To avoid any singularity problem, when both \( D(S_{ij} \rightarrow S_{2j}) \) and \( D(S_{2j} \rightarrow S_{ij}) \) are less than 0.05, the original absolute strength \( D \) is replaced by \( D + 1 \). The causal decomposition calculates causal strength for both directions, and we can identify the causal direction by comparing the strength of each direction.

4. Empirical analysis

4.1. Data description

Table 1 describes the data used in this study. The data for COVID-19 confirmed cases in the US were acquired from the USA Facts, a non-profit organization that provides data and reports on the US population [45]. To represent how many people are traveling in each state, we use the number of daily trips by states estimated by the Bureau of Transportation Statistics (BTS) by the Maryland Transportation Institute and the Center for Advanced Transportation Technology Laboratory at the University of Maryland [46]. The travel statistics are estimated using the anonymous national panel of mobile device data from multiple sources. The data for state policies on physical distancing, such as closures, shelter-in-place orders, and other related policies, were obtained from the database at the Inter-university Consortium for Political and Social Research (ICPSR) [47]. The number of state policies implemented at each time point were used as a proxy measure for the strength of those policies [5]. The selected data contained 222 days of records from 22 January to 30 August, 2020.

Apart from the time series data, various cross-sectional data were used to represent intrastate or interstate characteristics. Euclidean distance was calculated to represent the geographical distance between states. Socio-demographic data, such as Gross Domestic Product (GDP) [48], median age [49] or nonwhite population rate [50], were collected for each state from the U.S. Census Bureau (USCB). To review the compliance with government policy, political orientation was represented by the ratio of supporting Democrats in the 2018 U.S. Senate elections, which are collected by the MIT Election Data and Science Lab (MEDSL) [51]. To represent interstate traffic, we used the number of air travelers between states collected from the origin and destination survey of domestic airline passengers [52]. All the variables are standardized to compare the parameter estimates of variables with a different unit without the concern of multicollinearity.

We confirmed the correlation coefficients for intrastate variables and interstate variables, and they are used in the linear regression analysis in the later section. Regarding intrastate variables, the variance inflation factor values representing multicollinearity appeared to be lower than 1.50. The Pearson correlation coefficient between the intrastate variables was −0.036. Based on the results of the small correlation for the intrastate and interstate variables, we investigated the effect of those variables using linear regression analysis in a later section here.

4.2. EEMD analysis

Variability of amplitude and phase in the different frequency domains was reported from the previous influenza spread, which had a relationship with the geographical distance or temperature influences [53]. The travel pattern also was associated with the cyclicity and its variability based on the gender, race, income, and geographical attributes [54], and it has a nonlinear relationship with the COVID-19 cases [55]. Therefore, as with the impact of policy, EEMD is required to analyze daily trips and COVID-19 cases, as explained by different periods with multiple characteristics.
Table 1

| Name                  | Description                                      | Type        | Source              | Period                        | Area       |
|-----------------------|--------------------------------------------------|-------------|---------------------|--------------------------------|------------|
| **Time-series variables**                  |                                                  |             |                     |                                |            |
| COVID-19 cases        | Daily new COVID-19 confirmed cases               | Time-series | USA Facts           | Jan 22nd~Aug 30th, 2020        | Intrastate |
| Number of daily trips | Number of daily trips within the state           | Time-series | BTS                 | Jan 22nd~Aug 30th, 2020        | Intrastate |
| State policy          | Number of state policies on physical distancing  | Time-series | ICPSR               | Jan 22nd~Aug 30th, 2020        | Intrastate |
| **Intrastate variables**                |                                                  |             |                     |                                |            |
| GDP                   | Gross domestic product within the state          | Cross-sectional | USCB              | 2020 Q1                        | Intrastate |
| Median age            | Median age                                       | Cross-sectional | USCB              | 2019                           | Intrastate |
| Nonwhite population rate | Proportion of population of a race other than 'white alone' | Cross-sectional | USCB              | 2019                           | Intrastate |
| Political orientation | Ratio of supporting Democrats in the 2018 United States Senate elections | Cross-sectional | MEDSL              | 2018                           | Intrastate |
| **Interstate variables**               |                                                  |             |                     |                                |            |
| Distance              | Geographical distance between states             | Cross-sectional | –                  | 2020                           | Interstate |
| Air traffic           | Origin and destination survey by tickets reserved | Cross-sectional | BTS               | 2020                           | Interstate |

Notes. BTS = Bureau of Transportation Statistics; GDP = Gross Domestic Product; ICPSR = Inter-university Consortium for Political and Social Research; USCB = US Census Bureau; MEDSL = MIT Election Data and Science Lab.

We first obtained IMFs after applying the EEMD methods to the time series on COVID-19 cases and the number of daily trip data for all 50 states and a federal district (Washington, D.C.) in the United States. Fig. 2 depicts the decomposed IMFs after employing EEMD in California and New York state. The number of state policies represented as time-series and corresponding IMF are presented as gray and red lines, respectively. The left y-axis describes the normalized amplitude of IMF, while the right y-axis shows the number of state policies. Each IMF represents the orthogonal (i.e., mutual independent) signal that is associated with a specific frequency domain. To characterize those IMFs, the average period, energy strength, and its correlation coefficient to each state policy were calculated and displayed on the right side of the figure. The energy strength ranges from 0% to 100%, and it stands for how much of the energy density distribution corresponds to each IMF, as denoted in Eq. (10).

\[
\text{Energy strength, } E_j = \left( \frac{1}{N} \sum_{k=1}^{N} |S_j^k(t)|^2 \right) / \left( \frac{1}{M} \sum_{j=1}^{M} \sum_{k=1}^{N} |S_j^k(t)|^2 \right) \times 100
\]  

(10)

The IMF having the highest cross-correlation coefficient is highlighted in boldface and blue colored in Fig. 2.

The IMF with a high correlation value for each state turned out differently. In other words, the effect of state policy on COVID-19 cases or daily trips was reflected in the different level of period in each state. For COVID-19 cases in California, IMF 6 with 111 days of periods showed the highest positive correlation coefficient to the number of state policies. Meanwhile, IMF 5 decomposed from the number of daily trips was reported as having a negative correlation with the highest values. However, in New York state, both COVID-19 cases and the number of trips showed the highest correlation at the IMF 5 for the period of 74 days. These characteristics could be commonly found in that the correlation coefficient is the highest at the low frequency IMF; however, it turned out to be slightly different from state to state. These results show that the effect of state policy on the number of daily trips and further still, the number of COVID-19 cases appeared as a long-term effect. Therefore, the continuous intensity of policy measures may be more effective than the temporal ones.

Fig. 3 describes the distribution of energy strength and its corresponding dominant periods for the six decomposed IMFs, which were collected from 50 states and a federal district. In the case of the IMF for COVID-19 cases (See black circles), IMF 1 and IMF 2 show short-term patterns within approximately one week, while IMF 5 and IMF 6, representing long-term patterns, have a period of more than a couple of months. IMF 3 and IMF 4 denoted mid-term range patterns, having the largest variation among the states. In other words, the intrinsic and intricate pattern of the state could be recognized through gathering the mid-term pattern, while there was no significant difference between the average periods in the short-term and long-term (i.e., weekly pattern) by region. Therefore, this study mainly conducted state-level analysis focusing on the short-term and long-term because the time-scale of mid-term IMF varies from state to state. The decomposed function of the number of daily trips (See red triangles) also appeared to be similar to the COVID-19 cases in short, medium, and long-term patterns. However, the average periods were relatively steady and only small varied...
for the states in the short and long-term IMFs in daily trips, unlike the COVID-19 case. That may be because people’s trip making was dominated by the factors such as social distancing policies and quarantine rules that were not significantly different from state to state [56], while infectious diseases were affected by varied and complex human, socio-economic, and climate factors [2].

4.3. Causality analysis

4.3.1. Intrastate causal relationship

Since various factors entailed varying effects on COVID-19 cases in a different frequency, it is difficult to explain them all by discovering only the original time series. An EEMD that decomposes the time series into IMFs was introduced to identify those effects that are difficult to see through examining the original time series but could be seen through the IMFs representing different frequency domains [7,15–17]. Various influential factors affected the trip-making process, including weekly shared trips, and there are intrinsic social contexts for the different trip purposes. This section explores the causality between COVID-19 reported cases and the number of daily trips in each state in the United States. These impacts are quantitatively measured by an analysis of CCM and causal decomposition. Two methods calculate the extent of causation and the bidirectional causal strengths for each, and their value ranges between 0 and 1. To compare the causal strength between the COVID-19 case and the number of people’s trips within each state, we used the average of the causal strength of the number of trips on the COVID-19 case and that of COVID-19 cases on the number of trips. This analysis further aims to determine which factors can explain the mutual strength of causality between COVID-19 confirmed cases and the daily number of trips at a specific frequency. We covered 50 states and a federal district (Washington, D.C.) in the United States, and Table 2 shows the results of the linear regression analysis. The dependent variable was the CCM coefficients between each IMF of the COVID-19 cases and the corresponding IMF of the number of daily trips for each state. The independent variables were the intrastate variables, such as GDP, median age, nonwhite population ratio, and political orientation. All variables were standardized by subtracting the mean and dividing by the standard deviation.

The regression results for the causal strength between COVID-19 cases and the number of trips reveal the significant factors for the specific IMF not revealed by those results from CCM coefficient. For the IMF 2, GDP is positively associated in terms of CCM, which may support the fact that inequality in state-level income and spending on welfare and education

Fig. 2. Ensemble empirical mode decomposition (EEMD) results for COVID-19 cases and daily trips in California and New York states: Energy, period, and cross-correlation coefficients of each IMF are indicated.
Fig. 3. Energy strength and corresponding periods for each IMF gathered from the number of COVID-19 cases and the number of daily trips by states.

Table 2
Linear regression results for Convergent Cross Mapping (CCM) coefficient and the causal strength between COVID-19 cases and the number of daily trips obtained from a causal decomposition for each state.

| Dependent variable: Convergent Cross Mapping value between COVID-19 cases and the number of daily trips for each state | Estimate | IMF1 | IMF2 | IMF3 | IMF4 | IMF5 | IMF6 |
|---------------------------------------------------------------|---------|------|------|------|------|------|------|
| Average CCM coefficient | 0.08277 | 0.2782 | 0.2367 | 0.5834 | 0.6667 | 0.8056 |
| (Intercept) | 0.08277*** | 0.2782*** | 0.2367*** | 0.5834*** | 0.6667*** | 0.8056*** |
| GDP | 0.02148 | 0.04258* | 0.004062 | 0.01854 | 0.01787 | 0.01196 |
| Median age | −0.008538 | −0.01339 | 0.001278 | −0.002171 | 0.002108 | −0.01478 |
| Political orientation | −0.01235 | −0.000493 | 0.01154 | 0.01146 | 0.007396 | −0.04477* |
| Nonwhite rate | −0.003596 | −0.04582* | 0.02402 | −0.02784 | 0.01900 | 0.01421 |
| Adjusted R² | 0.03849 | 0.1359 | 0.05889 | −0.01222 | −0.02631 | 0.06024 |

Dependent variable: causal strength between COVID-19 cases and the number of daily trips for each state

| Estimate | IMF1 | IMF2 | IMF3 | IMF4 | IMF5 | IMF6 |
|---------|------|------|------|------|------|------|
| Average causal strength | 0.0702 | 0.0429 | 0.0361 | 0.0612 | 0.1498 | 0.0603 |
| (Intercept) | 0.0702*** | 0.0429*** | 0.0361*** | 0.0612*** | 0.1498*** | 0.0603*** |
| GDP | 0.0016 | −0.0009 | 0.0008 | 0.0010 | 0.0248 | −0.0031 |
| Median age | 0.0029 | −0.0012 | 0.0120*** | 0.0206*** | −0.0031 | 0.010 |
| Political orientation | −0.0084 | −0.0024 | −0.0039 | 0.0047 | 0.0348* | 0.0067 |
| Nonwhite rate | −0.0040 | −0.0005 | 0.0050 | −0.0000 | −0.0219 | 0.0094 |
| Adjusted R² | 0.0471 | −0.0538 | 0.2299 | 0.2266 | 0.1302 | 0.1078 |

Notes: * = p < 0.05; ** = p < 0.01; *** = p < 0.001.

are related to the risks of mortality from COVID-19 [18]. Also, the reduction of their trip is also associated to the land price and the association is higher in the wealthier areas [24], while less wealthy people tend to keep their trips during the pandemic [57].

Several factors not revealed by the CCM coefficient appear differently and more significant in terms of causal strength. It may be due to different causation patterns and their significance according to the different frequency domains. Over the CCM analysis, the causal decomposition method has its merits in two ways: (i) it helps accommodate the complex causal interaction with oscillatory stochastic mechanisms in real-world data at the comparable time-scale and (ii) causal decomposition also reflects other frequency domains, considering it calculates the phase coherence between the original IMFs and the re-decomposed IMFs removing certain IMFs, whilst the CCM coefficient represents the correlation between the same frequency domains.
To find the mutual causal relationship between the COVID-19 cases and the number of daily trips, the causal decomposition was conducted followed by regression analysis. The bottom of Table 2 presents each IMF’s average causal strength between the COVID-19 cases and the number of daily trips and the results of that regression using intrastate variables and causal strength. For causality analysis between the case of COVID-19 and the number of trips, it is necessary to compute the average of the causal strength of each direction, since it is not known whether the number of confirmed cases or the social distancing mandate is the cause, and these patterns are expected to change continuously over time and space. The average causal strength of IMF 5, representing long-term causality, is the largest among the IMFs, which indicates that the long-term effects dominate the influences between COVID-19 and the number of daily trips (i.e., degree of social distancing). The average CCM coefficient also tends to be larger in the long-term effects. While the short-term effect of social distancing policies is reported to shift the peak of epidemics keeping the value of the peak retained at the same level, those policies significantly flatten the long-term trend of the spread [58]. Therefore, the energy strength of the long-term domain of COVID-19 cases may partially represent the degree of social distancing.

The causal strength of causal decomposition from the COVID-19 cases to the daily trips becomes stronger when the state’s median age is high. The higher the median age, the more successful was the practice of reducing trips or social distancing according to the spread of diseases. The elderly people considerably reduced their trips during the COVID-19 pandemic, as they are thought to be one of the most susceptible populations [31]. This is also comparable to the results from Hu et al. [3] that higher proportions of the 40 to 65 age groups are more likely to reduce more trips than the 0 to 24 age group. In the causal strength at the low-frequency IMF, the stronger the state’s democratic tendency, the greater was the effect of reducing trips according to outbreaks of the COVID-19. Political orientation has been known to be a significant factor in affecting trips by case. Democrats are more likely to adapt to the government’s policy and social distancing mandates, and this tendency significantly affects the spread of COVID-19 [18,59]. However, the regression results from CCM and the causal decomposition appear differently. The CCM coefficient for political orientation in low-frequency domain presents a negative estimate in IMF 6, while causal decomposition shows a positive estimate in IMF 5. It was also found that neither GDP nor proportion of the nonwhite population affected intrastate causal strength in any frequency domain.

The significant impacts of intrastate variables are mainly found in IMF 3 through 6. In other words, the high-frequency domain is inclined to be difficult at providing significant explanations since various microscopic and noisy components could be derived from exogenous features. In contrast, the low- and mid-frequency domain may represent the macroscopic component which characterizes overall general or unique patterns of the different states.

4.3.2. **Interstate causal relationship**

The spread of epidemics does not only occur within states. It can also occur through the interaction between states. Reports on the number of confirmed cases disclosed and the spread of an epidemic may influence people’s decisions to travel. We applied causal decomposition methods to discover mutual causation and place it into two folds: between the COVID-19 cases of two states, and the COVID-19 case and daily trips of two states. The former is to see the transmission of COVID-19 between states, and the latter is to investigate the impact of other states’ COVID-19 cases on people’s mobility levels. In this study, a total of 2550 (51 × 50) pairs of causal strengths are calculated to investigate the mutual causal relationship among 50 states and a federal district. Similar to the case of intrastate causality, we only measure the average causal strength of each direction.

Fig. 4 shows the relative causal strength of COVID-19 cases in Nevada and California adjacent to each other. In the association between Nevada and California, the causal strength become clearer for the long-term trend. These patterns show they are different among other states, since the causal relationship for each IMF is known to differ [40], which mirrors the fact that the epidemic transmission pattern is fluctuating and changeable over the varying periods. The IMFs having the higher average causal strength are highlighted in boldface.

For a comprehensive understanding of mutual causation, it is essential to discover the importance and significance of its heterogeneous features. The linear regression analysis consists of the absolute causal strength as a dependent variable and the distance for states and aviation traffic as the independent variables. It is necessary to compare the average of the four sets (i.e., case of A to trip of B, case of B to trip of A, trip of A to case of B, and trip of B to case of A) of causal strengths between their COVID-19 case and the number of trips in a couple of states.

Table 3 presents the average causal strength and the parameter estimates for each IMF. Similar to the causality of the intrastate, the average causal strength of IMF 5 is the largest among the IMFs, which indicates that the long-term effects also dominate the influences between the COVID-19 cases of a pair of states and those between COVID-19 and the number of daily trips. In the United States, the distance between two states greatly influences road traffic, so the amount of road traffic could be considered through examining the distance between states. The distance variable has a negative association in IMF 3, 4, and 5. This result implies that the smaller the distance, the more likely it is to have mutual travel, and a higher causal strength between COVID-19 cases. However, the IMF 6 representing the longest term shows a positive relationship between the geographic distance and the causal strength of IMF 6. The IMF 6 shows a period of 74 or 111 days, nearly half or a third of the total period of 222 days of the pandemic. Therefore, this trend appears to be determined dependently on the entire period such as seasonality, rather than on specific factors.

On the other hand, since there is a tendency to use airplanes for relatively long-distance travel, the amount of airplane traffic refers to the number of travels made between states over a longer distance. Airline traffic showed a positive
association in IMF 4 and 5. It is in line with the expectation that the more air traffic that occurs in the long-term period, the greater the impact on other states. It is acknowledged that there is a significant linear correlation between domestic air traffic volume and the spread of COVID-19 [13]. However, the results show that air traffic in a high-frequency IMF has a negative association, which is distinct from the hypothesis. It implies that this difference is due to the greater effect of not only traffic but also other unobserved factors including noises that were not considered in the regression model [60].
Due to the seriousness of COVID-19, the number of confirmed cases worldwide is disclosed on a daily basis. As this information is widely accessible to the public, the spread of COVID-19 in one state may greatly influence the trip decision-making processes of another state. These linear regression results describe the causality of the COVID-19 reported cases in one state on the number of daily trips in another state as an influence of that distance or the estimated amount of airline traffic between the states. The results show that the distance between states is significant in all time frequency domains. It also appears the negative association from the IMF 1 to 5 on the causal strengths is less likely to be affected by the new COVID-19 cases between distant states. This association is also consistent with the notion that near things are more related than distant things. On the other hand, the positive influence in IMF 6, which also appeared in the case to case, could be seen as the effect of an entire trend rather than any specific factor. This effect could be examined for longer data collection period lengths. The variable of domestic airline traffic shows a significant positive association in IMF 5 and a negative association with IMF 1 to 3. It was found that, in the long-term domain, the higher the air traffic volume, the higher the causality they have, whereas, in the short-term, the states with more air traffic decreased the causal strength of the two states. This can be seen as the same trend as revealed in the relationship between the COVID-19 cases shown above. These results might stem from the phenomenon that we have not considered or the limitation of causal decomposition itself. Some previous studies also exhibited less obvious and unclear direction of the causality or sometimes contradictory results in interpreting the causal decomposition results [40,41]. The energy strength of IMFs and their causal strength varies from different factors, and the causality results might differ in severe fluctuation among different frequency domains.

4.4. Clustering analysis

The preceding results show that each state has a certain frequency domain that dominantly affects or is affected. We conduct the clustering analysis to group states having similar time-scale energy strengths of COVID-19 cases. IMF 1 and 2 are classified as short-term, IMF 3 and 4 as mid-term, and IMF 5 and 6 as the long-term components. Clustering methods can be divided into simultaneous clustering and hierarchical clustering. We have considered the basic method in each approach, K-means clustering and agglomerative hierarchical clustering, which can cluster points in 3D space into several groups. The K-means clustering is sensitive to outliers and has disadvantages in that only circular clusters can be found. Fig. 5 depicts the scatter plots for each method according to the number of clusters predetermined. Since the three variables used for clustering are the relative ratio of the energy strength of short-term, mid-term, and long-term domains, their summation is 1.0, which can be drawn on one plane (i.e., \(x + y + z = 1\)). Each color represents the clustering results.

The silhouette method is applied to evaluate the clustering performances, which measures the similarity of each element within the cluster compared to the other clusters [61]. The better clustering result indicates a higher overall average silhouette coefficient. Table 4 describes the calculated average silhouette coefficient of the K-means and agglomerative hierarchical clustering. Both methods require the prior determination of the number of clusters. To derive the best combination of the clustering method and the number of clusters, we evaluated the clustering performance using the average silhouette coefficient, which measures the variations of inner distance. The evaluation results showed that the average silhouette coefficient of K-means clustering at \(k = 2\) shows the highest value, 0.52, so we adopted it in this study.

According to the energy strength in the short-, mid-, and long-term IMFs, the states are clustered into two groups. Fig. 6 presents the geographical distribution of the clustered states and corresponding energy strength distributions. Table 5 describes the energy strength of each frequency domain IMFs of COVID-19 cases and corresponding standardized values of the intrastate variables. Cluster 1, consisting of 29 states, has relatively higher energy strength in the short-term IMFs; Cluster 2 includes 21 states, having higher energy strength in long-term IMFs. The logistic regression to classify clusters reveals the statistically significant association with the intrastate variables, including GDP and political orientation, as shown in Table 6. Positive coefficients indicate that the attribute increase the probability of the state to be classified as a Cluster 2. Compared to Cluster 1, states belonging to Cluster 2 tend to produce more GDP and are politically closer to the Democrats. It reveals the relationship between clusters and socio-demographic characteristics that the higher the GDP and the more politically democratic, the higher the energy strength of the long-term IMFs. The differences of energy distribution could be interpreted through the socio-economic features of the states in each cluster. As shown in the results in Sections 4.2 and 4.3, the long-term IMFs mainly represent the causality between the COVID-19 cases and the degree of the social distancing. States belonging to Cluster 1 are more likely to be politically conservative and
less likely to adhere to the social distancing rules or nonpharmaceutical interventions [55,62]. Whereas, states in Cluster 2 tend to support Democrats and produce higher GDP, and these states are more likely to embrace social distancing policies. Therefore, in Cluster 2, COVID-19 cases would be mainly affected by long-term variations stemmed from the social distancing policies and global trends rather than short-term variations of COVID-19 outbreaks. These results revisit the findings that the causality of social distancing and the COVID-19 cases mainly comes from the long-term rather than the short-term scale. These findings suggest that state policy makers need to continuously monitor long-term trends of COVID-19 cases to determine whether the social distancing policies are effectively working or not.
Table 5

| Energy strength of COVID-19 cases (%) | Intrastate variables (standardized) |
|--------------------------------------|-------------------------------------|
|                                       | Short-term | Mid-term | Long-term | GDP        | Median age | Political orientation | Median income |
| Cluster 1                            | 58.89      | 10.65    | 30.46     | -0.3291    | -0.0124    | -0.4431            | -0.3946       |
| Cluster 2                            | 29.17      | 11.72    | 59.11     | 0.4338     | 0.0164     | 0.5841             | 0.5201        |

Table 6

| Logistic regression results for the clustering results. |
|--------------------------------------------------------|
| Estimate | Standard error | p-value |
| Intercept | -0.2303 | 0.3779 | 0.5423 |
| GDP | 1.2977 | 0.6854 | 0.0583 |
| Median age | -0.2173 | 0.4494 | 0.6287 |
| Political orientation | 1.3105 | 0.7686 | 0.0882 |
| Median income | 0.2576 | 0.5700 | 0.6513 |

5. Conclusions

The prior works have focused on the influence of road or airline traffic on the spread of COVID-19 and figured out the significant correlations between them [6,13,21]. While some literature deeply dives into microscopic influences [3,20], it still needs to consider macroscopic influences for different time-scale signals by investigating state-level correlations. This study explored the intrastate and interstate causality of COVID-19 among the states in the United States by applying EEMD and the causal decomposition method that finds the intrinsic features and causal relationships using the decomposed signals (i.e., IMFs) in different frequency domains. Also, cluster analysis revealed the association of energy distribution of IMFs with socio-economic variables of the states.

The findings of this study provide useful and interesting insights into the causal dependencies of COVID-19 cases and social distancing behaviors. The energy distribution of IMFs from the EEMD analysis indicates that the mid-term and long-term domain portrays the macroscopic component of the states that characterizes overall intrinsic characteristics. For example, state policy for social distancing tends to affect COVID-19 cases in the long-term, further suggesting the need for continuous implementation of policy measures rather than the temporal one.

Intrastate causality analysis identified the significant effects of median age, political orientation, and nonwhite rate on the causal strength, and it revealed the hidden relationship between COVID-19 cases and the number of trips not identified from the CCM. This result implies the capability of causal decomposition to elucidate the causative factors hidden in a local time domain. Interstate causality results show a negative association with the interstate distance, which implies that the distance plays a significant role in the spread of COVID-19. It is also found that, in the long-term domain, the higher the air traffic volume, the higher the causality they have, whereas, in the short-term, the states with more air traffic decreased the causal strength of the two states due to noisy and unobserved factors.

Clustering analysis help reveal the relationship between clusters and socio-demographic characteristics. Cluster 1 has relatively higher energy strength in the short-term IMFs, and Cluster 2 shows higher energy strength in long-term IMFs. It turned out that the higher the GDP and the more politically democratic, the higher the energy strength of the long-term IMFs. Considering the tendency to adhere to the social distancing of states that are high GDP and democratic [55,62], this result reconfirms the long-term and global effects of social distancing. These findings suggest the need for monitoring of long-term trends of COVID-19 cases rather than short-term, to determine whether the social distancing policies are effectively working or not.

This study emphasizes the achievement in revealing the intrastate and interstate level causal relationship between several features of states and the spread of COVID-19 in different frequency domains. We confirmed that several influential factors on causality only could be revealed by the proposed multi-scale causality analysis. In other words, this study offers originality in its investigation of the association between the pandemic to socioeconomic features and travel behavior. The findings offer new implications for legislators to establish policies such as travel restrictions to mitigate the spread of such diseases, including both national and international aspects.

Although the present study deeply elucidated the causal relationship between the COVID-19 cases and the mobility levels by investigating various attributes, we need to consider several limitations in future research. First, the causal decomposition lacks the representation of the time lag between the original time series and IMFs due to the nature of the methods. Further enhancement can be achieved by considering autoregressive terms to learn the temporal relationship between the features. Second, since the causal decomposition only considers the overall direction of the causality, we only analyze the averaged causal strength between the states without considering the direction. The direction of causality between the social distancing and the outbreak of COVID-19 may be dynamically changed and nonlinearly related. Also, these patterns could vary in each state. Investigating this dynamic causality is the important research topic. Work in this regard is ongoing. Third, this study revealed the significant association of socio-economic attributes with the causality.
between the COVID-19 and mobility level. To verify these findings apart from US, it is necessary to apply this study to other regions with different cultural backgrounds. Lastly, the recent relaxation of the strict social distancing mandates has been progressed around the world after the vaccination. Thus, future research should explore the possible influence of the vaccination and the modified mandates on the spread of COVID-19.

CRediT authorship contribution statement

Jung-Hoon Cho: Data curation, Formal analysis, Investigation, Writing – original draft, Methodology. Dong-Kyu Kim: Writing – review & editing, Supervision, Funding acquisition, Project administration. Eui-Jin Kim: Conceptualization, Writing – review & editing, Methodology, Validation.

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

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF), South Korea funded by the Ministry of Science and ICT (2020R1F1A1074395).

References

[1] D.C. Adam, P. Wu, J.Y. Wong, E.H.Y. Lau, T.K. Tsang, S. Cauchemez, G.M. Leung, B.J. Cowling, Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong, Nat. Med. 26 (2020) 1714–1719, http://dx.doi.org/10.1038/s41591-020-1092-0.
[2] S. Hsiang, D. Allen, S. Annan-Phan, K. Bell, I. Bolliger, T. Chong, Druckenmiller, L.Y. Huang, A. Hultgren, E. Krasovich, P. Lau, J. Lee, E. Rolf, J. Tseng, T. Wu, The effect of large-scale anti-contagion policies on the COVID-19 pandemic, Nature 584 (2020) 262–267, http://dx.doi.org/10.1038/s41586-020-2404-8.
[3] S. Hu, C. Xiong, M. Yang, H. Younes, W. Luo, L. Zhang, A big-data driven approach to analyzing and modeling human mobility trend under non-pharmaceutical interventions during COVID-19 pandemic, Transp. Res. Part C Emerg. Technolog. 124 (2021) 102955, http://dx.doi.org/10.1016/j.trc.2020.102955.
[4] V.S. Huang, S. Sutermaster, Y. Caplan, H. Kemp, D. Schmutz, S.K. Sgier, Social distancing across vulnerability, race, politics, and employment: How different Americans changed behaviors before and after major COVID-19 policy announcements, 2020, http://dx.doi.org/10.1101/2020.06.04.20119131, medRxiv 2020.06.04.20119131.
[5] IHME COVID-19 Forecasting Team, Modeling COVID-19 scenarios for the United States, Nat. Med. 27 (2021) 94–105, http://dx.doi.org/10.1038/s41591-020-1132-9.
[6] K.-B. Lee, S. Han, Y. Jeong, Covid-19, flattening the curve, and benford’s law, Phys. A Stat. Mech. Appl. 559 (2020) 125090, http://dx.doi.org/10.1016/j.physa.2020.125090.
[7] X. Liu, J. Huang, C. Li, Y. Zhao, D. Wang, Z. Huang, K. Yang, The role of seasonality in the spread of COVID-19 pandemic, Environ. Res. 195 (2021) 110784, http://dx.doi.org/10.1016/j.envres.2021.110784.
[8] A.C. Yang, C.-K. Peng, N.E. Huang, Causal decomposition in the mutual causation system, Nature Commun. 9 (2018) 3378, http://dx.doi.org/10.1038/s41467-018-05845-7.
[9] H.W. Hethcote, Qualitative analyses of communicable disease models, Math. Biosci. 28 (1976) 335–356, http://dx.doi.org/10.1016/0025-5564(76)90132-2.
[10] W.O. Kermack, A.G. McKendrick, Contributions to the mathematical theory of epidemics. II.—The problem of endemicity, Proc. R. Soc. London. Ser. A. Contain. Pap. A Math. Phys. Character 138 (1932) 55–83, http://dx.doi.org/10.1098/rspa.1932.0171.
[11] M. Mandal, S. Jana, S.K. Nandi, A. Khattar, S. Adak, T.K. Kar, A model based study on the dynamics of COVID-19: Prediction and control, Chaos Solit. Fractals 136 (2020) 109889, http://dx.doi.org/10.1016/j.chaos.2020.109889.
[12] E.A. Meirom, H. Maron, S. Mannor, G. Chechik, How to stop epidemics: Controlling graph dynamics with reinforcement learning and graph neural networks, 2020, pp. 1–26, arXiv.
[13] H. Lau, V. Khorasripour, P. Kochach, A. Mikolajczyk, H. Ichii, M. Zacharski, J. Bania, T. Khorasripour, The association between international and domestic air traffic and the coronavirus (COVID-19) outbreak, J. Microbiol. Immunol. Infect. 53 (2020) 467–472, http://dx.doi.org/10.1016/j.jmii.2020.03.026.
[14] P. Monmousseau, A. Marzuoli, E. Feron, D. Delahaye, Impact of Covid-19 on passengers and airlines from passenger measurements: Managing customer satisfaction while putting the US air transportation system to sleep, Transp. Res. Interdiscip. Perspect. 7 (2020) 100179, http://dx.doi.org/10.1016/j.trip.2020.100179.
[15] K. Bhaskaran, A. Gasparrini, S. Hajat, L. Smeeth, B. Armstrong, Time series regression studies in environmental epidemiology, Int. J. Epidemiol. 42 (2013) 1187–1195, http://dx.doi.org/10.1093/ije/dyt092.
[16] B. Cazelles, K. Cazelles, M. Chavez, Wavelet analysis in ecology and epidemiology: impact of statistical tests, J. R. Soc. Interface 11 (2014) 20130585, http://dx.doi.org/10.1098/rsif.2013.0585.
[17] N. Nava, T. Di Matteo, T. Aste, Dynamic correlations at different time-scales with empirical mode decomposition, Phys. A Stat. Mech. Appl. 502 (2018) 534–544, http://dx.doi.org/10.1016/j.physa.2018.02.108.
[18] W.P. Hanage, C. Testa, J.T. Chen, L. Davis, E. Pechter, P. Seminario, M. Santillana, N. Krieger, COVID-19: US federal accountability for entry, spread, and inequities—lessons for the future, Eur. J. Epidemiol. 35 (2020) 995–1006, http://dx.doi.org/10.1007/s10654-020-00689-2.
[19] M. Chinazzi, J.T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A.Pastore-y.Piontti, P. Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M.E. Halloran, I.M. Longini, A. Vespiugiani, The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak, Science (80-.) 368 (2020) 395–400, http://dx.doi.org/10.1126/science.abb9757.
[20] S. Ryu, S.T. Ali, C. Jang, B. Kim, B.J. Cowling, Effect of nonpharmaceutical interventions on transmission of severe acute respiratory syndrome coronavirus 2, South Korea, 2020, Emerg. Infect. Diseases 26 (2020) 2406–2410, http://dx.doi.org/10.3202/eid2610.201886.
[59] H. Allcott, L. Boxell, J. Conway, M. Gentzkow, M. Thaler, D. Yang. Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic, J. Public Econ. 191 (2020) 104254, http://dx.doi.org/10.1016/j.jpubeco.2020.104254.

[60] Y. Xiao, X. Tian, M. Xiao. Tourism traffic demand prediction using google trends based on EEMD-DBN, Engineering 12 (2020) 194–215, http://dx.doi.org/10.4236/eng.2020.123016.

[61] P.J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65, http://dx.doi.org/10.1016/0377-0427(87)90125-7.

[62] A. Gollwitzer, C. Martel, W.J. Brady, P. Parnamets, I.G. Freedman, E.D. Knowles, J.J. Van Bavel. Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic, Nat. Hum. Behav. 4 (2020) 1186–1197, http://dx.doi.org/10.1038/s41562-020-00977-7.