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.
Analysis of COVID-19 pandemic in USA, using Topological Weighted Centroid

Masoud Asadi-Zeydabadi a,∗, Massimo Buscema b,c, Weldon Lodwick b, Giulia Massini c, Francesca Della Torre c, Francis Newman b

a Department of Physics, University of Colorado Denver, Denver, CO, USA
b Department of Mathematical and Statistical Science, University of Colorado Denver, Denver, CO, USA
c Semeion Research Center of Sciences of Communication, Rome, Italy

ARTICLE INFO

Keywords:
COVID-19 pandemic
Topological weighted centroid (TWC)
TWC-Windowing

ABSTRACT

The first case of COVID-19 in USA was reported on January 20, 2020. The number of COVID-19 confirmed cases and death has increased since the first reported case and the outbreak has appeared in all states. This paper analyzes disease outbreak using Topological Weighted Centroid (TWC), which is a data driven intelligent geographical dynamical system that models disease spread in space and time. In this analysis the COVID-19 cases in USA on March 26, 2020 as provided by Johns Hopkins University is used. The COVID-19 outbreak is mapped by the TWC method. We were able to predict and capture some features of the pandemic spread using the early data. Although we have used the geographical distance from the latitude and longitude coordinates, our results indicate that one of the main paths of diseases spread are arguably airline routes. In this analysis, we used a large set of data. A modified version of TWC, is named TWC-Windowing to elaborate the effect of data from all places.

1. Introduction

The office of the World Health Organization (WHO) in China, reported, on December 2019 a case of pneumonia with unknown cause in Wuhan, China [1,2]. The new virus is known now as the novel corona virus or SARS COV-2 and it has spread throughout the world. The first case of COVID-19 in USA was confirmed on January 20, 2020 [3]. The Center for Disease Control and Prevention (CDC), on February 26, 2020 confirmed the first case of community spread of COVID-19 in California [4]. The first reported death from COVID-19 in USA was in the state of Washington, on February 20, 2020 [5,6]. On March 26, 2020, according to Johns Hopkins University, Coronavirus Resource Center, the number of confirmed and recovered cases were 83,836 and 681 and deaths were 1,209 [5,6]. The numbers of confirmed cases and death continue.

The Topological Weighted Centroid (TWC) algorithms are used in the analysis of COVID-19 in this paper. These algorithms have successfully predicted the pattern and the origin of variety of disease outbreaks around the world [7–12]. In this paper, a new set of TWC algorithms (TWC-Windowing) to analyze the COVID-19 data is introduced. The results show the TWC-Windowing is more suitable than other type of TWC to model a vast disease outbreak such as COVID-19 in USA.

TWC was compared with other epidemiological methods in many papers quoted in the references. In 2008 and 2009, we compared TWC with the geographic profile algorithms [13]. In 2013, TWC was compared with traditional data driven methods. In the 2013 paper, a comparison was made with 4 different well-known epidemics (Chikungunya fever, Italy, 2007 – The Foot and Mouth Disease Epidemic, UK, 1967 – The Golden Square Cholera Epidemic of London 1854 – the Russian Influenza in Sweden in 1889–1890). The algorithms used for the comparison were: a. The Rossmo Algorithm (RIGEL), b. The Negative Exponential Summation or Canter Algorithm (NES), c. The Likelihood Variance Maximization (LVM), d. The Maximum Probability Algorithm (MaxProb). In all the applications, TWC performed better than the other classic algorithms. The accuracy was measured using three different metrics: Distance, Sensibility, and Search Area [12]. Lastly, in the 2018 paper, TWC was applied to the problem of terrorism and TWC was shown to perform with excellent accuracy in terrorism event prediction (from high accuracy to average accuracy) [14].

∗ Corresponding author.
E-mail address: Masoud.Asadi-Zeydabadi@ucdenver.edu (M. Asadi-Zeydabadi).
https://doi.org/10.1016/j.compbiomed.2021.104670
Received 19 March 2021; Received in revised form 14 July 2021; Accepted 18 July 2021
Available online 21 July 2021
0010-4825/© 2021 Elsevier Ltd. All rights reserved.
2. Data, methods and results

2.1. Data

The COVID-19 cases in USA on March 26, 2020 as provided by Johns Hopkins University, Coronavirus Resource Center were used in this analysis [5, 6]. The COVID-19 data was collected from 3,146 locations scattered over all places in USA. There are about 31 (31.46) confirmed cases, on average, for each location. For each verified COVID-19 case, the data structure is the coordinates (latitude and longitude). The dataset therefore consists of as many rows as there are cases and two columns.

2.1.1. Methods

The TWC method is based on statistical thermodynamics and optimizing free energy and entropy. In this method the probability of dispersing the disease to \((x, y)\) is given in term of energy which is a function of a distance to the infected points. In statistical thermodynamics the probability distribution for \(i\)th state is given by

\[ p(E_i) = \frac{e^{-E_i/k_B T}}{Z}, \]

where \(E_i\) is energy for \(i\)th state, \(k_B\) is Boltzmann constant, \(T\) is absolute temperature and \(Z\) is the partition function that for a system of \(N\) states is defined as follows,

\[ Z = \sum_{i=1}^{N} e^{-E_i/k_B T}. \]

The free energy in statistical thermodynamics is defined as

\[ F = -k_B T \ln Z. \]

The entropy is

\[ S = -\frac{\partial F}{\partial T}. \]

In the TWC method the idea of statistical thermodynamics is adapted with some adjustments. Energy in TWC method is defined as functions of the Euclidean distance between infected points, and \(\frac{1}{k_B T}\) is replaced with an optimization parameter, \(C = \frac{1}{k_B T}\). The optimization parameter \(C\) in different TWC models are named as \(\alpha, \beta, \gamma\) or others for more details see Ref. [8].

| Algorithm | Description |
|-----------|-------------|
| TWC – \(\alpha\) | The possible past outbreak distribution according to the given data |
| TWC – \(\beta\) | The possible present outbreak distribution according to the given data |
| TWC – \(\gamma\) | The possible near future outbreak distribution and longer future according to the given data |
| TWC – \(\theta\) | The possible longer future outbreak distribution and constructs an associated undirected graph according to the given data |
| TWC – \(i\) | The possible vanishing outbreak locations according to the given data |

Fig. 1. The result of TWC – \(\alpha\) based on COVID-19 in USA on March 26, 2020.
There are associated scalar fields for these TWC methods and there are other types of TWC methods. For more details about the TWC method see Ref. [8]. There are three types of TWC analyses, each of these types having five associated algorithms. The three types are called TWC- Original algorithm \([1]\), TWC-Frequency and TWC-Windowing. The TWC method includes five main algorithms: \(TWC - \alpha\), \(TWC - \beta\), \(TWC - \gamma\), \(TWC - \theta\), and \(TWC - i\), see Table 1. The TWC algorithms identify the locations from which the disease dynamics can be thought to diffuse to other places. These locations are calculated as past, present, near future, and longer future sites by \(TWC - \alpha\), \(TWC - \beta\), \(TWC - \gamma\), and \(TWC - \theta\) respectively. In addition, \(TWC - \theta\) constructs an associated undirected graph that models the network through which the disease is thought to spread.

To use an analogy, suppose the activities of a group of Corona Virus “Terrorists” (CVT) with data of the distribution sites (locations of the COVID-19 cases), TWC will located the “hotspots”, centers, from which such a CVT group would most effectively attain the given distribution of cases. That is, the TWC algorithms use the reported data and predict from where the impetus arises to produce the data as exhibited by the reported values. In this analysis, \(TWC - \alpha\), \(TWC - \beta\), \(TWC - \gamma\), \(TWC - \theta\) and \(TWC - i\) algorithms have been used by implementing TWC-Windowing approach. That is, of the three types of TWC, we report results associated with one of these, TWC-Windowing. This method is introduced for the first time in this paper.

2.1.2. TWC-windowing

In TWC-windowing, for each point a group of neighboring points (window) is selected. The size of the window can be small or large. Five TWC algorithms are applied in each window. Then all TWC maps are superimposed as a single map. The window maps, which depend on their size, might have large or small number of overlaps. This will make the superimposed TWC a network. Basically, each window around a point can be interpreted as a local TWC analysis between that point with its associated points and the superimposed TWC is the network of TWC over all data points.

In TWC-Windowing, an integer number \(K < N\) for segmenting (windowing) \(N\) data points is selected. The number of points in each window is \(P = \lfloor N / K \rfloor\), where \(\lfloor \cdot \rfloor\) is the floor function. A larger value of \(K\) gives a smaller \(P\) number, in practice \(K \ll N\). For each of the points, the \(P\) closest points will be found. For each data point there is a window. For
the $N$ data points there are $P$ windows with each window having $P$ points. Then, the five TWC algorithms are applied on each window. Next, all maps are superimposed and normalized as the final map. If $K = 1$, $P = \frac{N}{K} = N$, and the TWC-windowing becomes TWC-original. In this paper, since we have a large set of data over entire USA, TWC-windowing provides better results, because it elaborates the effect of all data where local behavior is important. The TWC analysis in this paper used $N = 3146$, $K = 50$ and $\frac{N}{K} \geq P = 62$. TWC-windowing captures both local and broad effect and connection between all places that cause disease outbreak.

3. Results

The results of TWC are compared with observation. The total number of confirmed cases for all 50 states are calculated. If the location of the confirmed case was not given in the data, then that case was not included in the calculation of the observation map. In the observation maps each point represents an infected place. However, in general, the number of confirmed cases varies from point to point. Since there are more than 3000 points, plotting the points by itself creates the map of USA. These maps depict both intensity and the distribution of the confirmed cases. The maximum, minimum and average, ($C_{\text{max}}$, $C_{\text{min}}$ and $C_{\text{ave}}$) values between 50 states of the USA for a given day were calculated. The confirmed cases were classified to four groups by blue, green, yellow, and red colors. The blue ($C_b$), green ($C_g$), yellow ($C_y$), and red ($C_r$) groups represent $C_{\text{min}} \leq C_b \leq \frac{C_{\text{ave}} + C_{\text{min}}}{2}, \frac{C_{\text{ave}} + C_{\text{min}}}{2} < C_g \leq C_{\text{ave}}, C_{\text{ave}} < C_y \leq C_{\text{max}}$ and $C_{\text{max}} < C_r$, respectively. Note that the color scale in TWC and observation maps do not represent exactly the same color level. In the observation maps, only four colors were used. However, in the TWC maps a wide range of colors (with almost a continuous intensity color spectrum) were used.

3.1. TWC – $\alpha$ Map

Fig. 1 shows the results of the TWC – $\alpha$ algorithm. The hotspots (the origin of the outbreak) are shown in this figure. Two possible outbreak origins are located in west, Washington and California states. The northwest hotspot that originated in Washington state is expanded toward the east and covers the neighboring states. There are several isolated hotspots, for example there is one hotspot in northeast.

Fig. 2 shows the map of COVID-19 in USA on March 22, 2020 [15]. By comparing TWC – $\alpha$ map, Fig. 1, which represents the past, with observation map on March 22, 2020 (four days before March 26), Fig. 2, the red hotspots TWC – $\alpha$ map match overall with the red to green groups in observation map of four days before. This gives some scale of comparison between TWC and observation maps. The hotspots in TWC maps are given by reddish spots of a yellow halo with different intensity level. Overall, those spots can be compared with red to green color in the observation maps. If in the observation maps, the states have been divided into subspaces with more color levels, we can see more detail. For example, in observation map, Fig. 2, Colorado and Texas are seen in green color, in TWC – $\alpha$ also both have hotspots. However, Texas has...
Fig. 6. The result of TWC – $\beta$ based on COVID-19 in USA on March 26, 2020.

Fig. 7. COVID-19 map in USA on March 26, 2020.

Fig. 8. The number of confirmed cases in Michigan and Wisconsin between March 22 and May 1, 2020.
Fig. 9. The result of TWC – γ based on COVID-19 in USA on March 26, 2020.

Fig. 10. COVID-19 map in USA on April 7, 2020.

Fig. 11. The number of confirmed cases in Idaho, Washington and Wyoming between March 22 and May 1, 2020.
Fig. 12. The result of $TWC - \theta$ based on COVID-19 in USA on March 26, 2020.

Fig. 13. COVID-19 map in USA on April 15, 2020.
two hotspots in $\text{TWC} - \alpha$. The number of infected points in Texas are more than in Colorado and Texas is larger than Colorado, it is possible to see more detail by dividing Texas to subspaces. TWC method is a data driven technique and this analysis used only confirmed cases on March 26, 2020, and any information before or after this date has not been used. As a result of this feature, the comparison with details in some cases such as an outbreak in a vast country such as USA is a challenge.

Fig. 3 shows the observation map on March 10, 2020 (sixteen days before March 26) [16]. The hotspots in this figure are in Washington State and California. This is in agreement with $\text{TWC} - \alpha$ map, that also shows the main hotspots are in Washington State and California. Comparing the $\text{TWC} - \alpha$ map with the real data from March 10, sixteen days before March 26, are more reasonable. However, since at the beginning of outbreak there are less number of confirmed cases we also showed the observation map for the confirmed cases on March 22. All TWC and observation maps were given based on the relative intensity. In the observation maps, four colors were used, however, in TWC maps with a wide range of colors.

Table 2
Attractors location in cluster A.

| Location Number | City        | State       | Lat      | Long     | Confirmed | Deaths | Recovered |
|-----------------|-------------|-------------|----------|----------|-----------|--------|-----------|
| 17              | Colbert     | Alabama     | 34.69847 | -87.8017 | 1         | 0      | 0         |
| 128             | Craighead   | Arkansas    | 35.83018 | -90.6324 | 4         | 0      | 0         |
| 159             | Mississippi | Arkansas    | 35.76271 | -90.0519 | 0         | 0      | 0         |
| 248             | Arapahoe    | Colorado    | 39.64977 | -104.335 | 119       | 0      | 0         |
| 287             | Moffat      | Colorado    | 40.61811 | -108.207 | 0         | 0      | 0         |
| 933             | Jefferson   | Kansas      | 39.23479 | -95.3829 | 1         | 0      | 0         |
| 943             | Linn        | Kansas      | 38.21268 | -94.8425 | 4         | 0      | 0         |
| 1709            | Keya Paha   | Nebraska    | 42.87991 | -99.7134 | 0         | 0      | 0         |
| 1781            | Camden      | New Jersey  | 39.80344 | -74.9639 | 73        | 1      | 0         |
| 2521            | Warren      | Tennessee   | 35.67283 | -85.7797 | 0         | 0      | 0         |
| 2539            | Baylor      | Texas       | 33.61641 | -99.2135 | 0         | 0      | 0         |
| 2671            | Lee         | Texas       | 30.31121 | -96.9705 | 0         | 0      | 0         |
| 2879            | Halifax     | Virginia    | 36.76655 | -78.9351 | 1         | 0      | 0         |
Figs. 4 and 5 are from CDC around the time of data for Figs. 2 and 3. They are comparable with Figs. 2 and 3. They show at the beginning of the pandemic how fast the disease dynamics changes. In March 8, the intensity at west is higher than at east. However, in March 24 the intensity at east is higher than the west. The intensity scale in Figs. 4 and 5 are different from Figs. 2 and 3. The map in Fig. 4 was released on March 9, 2020, shows confirmed cases as of 4 p.m. March 8, 2020. Fig. 5 is the COVID-19 map on March 24, 2020 that was released by CDC.

3.2. TWC – β Map

The result of TWC – β, Fig. 6, shows the outbreak has expanded over the country. The cold (blue color) places get smaller and turn toward a green and yellow color. From this figure, the pattern of the disease spread over the country at the present time can be seen.

The map of COVID-19 on March 26, 2020 (present time) is given in Fig. 7 [17]. Wisconsin that was in the green group in Fig. 2 has been changed to the blue group in Fig. 7. There is the same trend in TWC approach for Wisconsin. The color of Wisconsin in TWC – α (Fig. 1), has been changed from reddish to yellowish in TWC – β (Fig. 6). TWC algorithm captures one of the main observational differences between past and present (here March 22 and March 26). It is noted that if, for example Wisconsin get cooler in the observation and TWC maps, this does not always mean that the number of cases decreases in this state. TWC and the observation maps are given in terms of relative number of cases. Fig. 8 shows the number of confirmed cases in Michigan and Wisconsin between on March 22 and March 26, 2020. In both states the number of cases increases, however, the rate of increase in Wisconsin is

| Location Number | City       | State       | Lat     | Long     | Confirmed | Deaths | Recovered |
|-----------------|------------|-------------|---------|----------|-----------|--------|-----------|
| 3               | Barbour    | Alabama     | 31.86826| -85.3871 | 0         | 0      | 0         |
| 177             | Sebastian  | Arkansas    | 35.19606| -94.2716 | 2         | 0      | 0         |
| 388             | Washington | Florida     | 30.61359| -85.66   | 0         | 0      | 0         |
| 400             | Bleckley   | Georgia     | 32.43829| -83.3304 | 0         | 0      | 0         |
| 677             | Rock Island| Illinois    | 41.46614| -90.5704 | 3         | 0      | 0         |
| 706             | Carroll    | Indiana     | 40.58078| -86.562  | 0         | 0      | 0         |
| 784             | Warren     | Indiana     | 40.34728| -87.356  | 1         | 0      | 0         |
| 813             | Clinton    | Iowa        | 41.89982| -90.5326 | 0         | 0      | 0         |
| 816             | Davis      | Iowa        | 40.74774| -92.4101 | 0         | 0      | 0         |
| 1211            | Prince George’s | Maryland | 38.8307 | -76.8496| 101       | 2     | 0         |
| 1433            | Jackson    | Mississippi | 30.54021| -88.6418 | 16        | 0      | 0         |
| 1501            | Cape Girardeau | Missouri | 37.38489| -89.6844 | 3         | 0      | 0         |
| 1614            | Fallon     | Montana     | 46.33478| -104.418 | 0         | 0      | 0         |
| 1674            | Cheyenne   | Nebraska    | 41.21999| -102.994 | 0         | 0      | 0         |
| 1680            | Dawes      | Nebraska    | 42.71985| -103.131 | 0         | 0      | 0         |
| 1752            | Churchill  | Nevada      | 39.58106| -118.339 | 0         | 0      | 0         |
| 1903            | Bladen     | North Carolina | 34.61296| -78.5618 | 0         | 0      | 0         |
| 2064            | Crawford   | Ohio        | 40.85065| -82.9199 | 1         | 0      | 0         |
| 2091            | Lawrence   | Ohio        | 38.59743| -82.5347 | 1         | 0      | 0         |
| 2431            | Yankton    | South Dakota | 43.00924| -97.3947 | 0         | 0      | 0         |
| 2450            | Cumberland | Tennessee   | 35.95272| -84.9984 | 6         | 0      | 0         |
| 2683            | Martin     | Texas       | 32.306  | -101.951 | 1         | 0      | 0         |
| 2855            | Culpeper   | Virginia    | 38.482  | -77.9563 | 2         | 0      | 0         |
| 2981            | Okanogan   | Washington  | 48.54855| -119.739 | 1         | 0      | 0         |
Table 4
Attractors location in cluster C.

| Location Number | City      | State     | Lat     | Long    | Confirmed | Deaths | Recovered |
|-----------------|-----------|-----------|---------|---------|-----------|--------|-----------|
| 150             | Lawrence  | Arkansas  | 36.04188| −91.1087| 1         | 0      | 0         |
| 737             | Jefferson | Indiana   | 38.78576| −85.4363| 0         | 0      | 0         |
| 1152            | Plaquemines| Louisiana| 29.42245| −89.6632| 13        | 0      | 0         |
| 2505            | Roane     | Tennessee | 35.84923| −84.5246| 1         | 0      | 0         |

Fig. 16. Attractors Loop in cluster B.

Fig. 17. Attractors Loop in cluster C.

Fig. 18. Attractors on USA maps including the data for the confirmed cases on March 26, 2020. Since a main transportation route in USA is via airport, we compare the locations these three attractor clusters with airports in USA. Figure 19 shows the map of the USA airports.
Fig. 19. The airport map: https://www.nationsonline.org/oneworld/map/google_map_major_US_and_Canadian_Airports.htm.

Fig. 20. Attractors on USA maps and Airports.
significantly slower than in Michigan. There are other differences between the number of cases on March 22 and March 26, 2020, such as outbreak in Pennsylvania, which is not captured clearly in TWC method.

3.3. **TWC – γ Map**

The result of TWC – γ, Fig. 9 predicts the outbreak in the future according to the TWC analysis. This algorithm predicts some cold regions get cooler. As mentioned before this might not mean these regions are moving toward a recovery stage. These maps represent the relative confirmed cases. From Figs. 6 and 9, it can be seen, for example, that Wyoming, in TWC – γ compared to TWC – β, gets cooler. However, this change is not seen in Fig. 10, the map of Covid-19 on April 7, 2020 (12 days after March 26) [18]. Fig. 11 shows the number of confirmed cases in Idaho, Washington and Wyoming states between March 22 and May 1, 2020. From this figure, it is clear that the rate of confirmed case in Wyoming compared to Idaho and Washington is increasing very slowly. The increase in rate of the confirmed case in Idaho compared to that of Washington is slow and compared to Wyoming is fast. The TWC – γ, Fig. 9, captures these changes. The color of Texas from Figs. 6–10, changes from green to yellow. As we mentioned both green and yellow represent the infected hotspots compare to the blue regions. In addition, there are two hotspots in Texas. For example, Texas compared to Colorado, which has only one hotspot, should be the hotter place.

3.4. **TWC – θ Map**

Fig. 12 shows the results of TWC – θ. The result of TWC – θ and TWC – β are very similar and the differences between them and TWC – γ and TWC – α are not significant. This means that the recovery is slow or that the relative differences between all places remain constant. The dynamics indicate that the disease is not moving toward the “cool places”. Fig. 13 shows the COVID-19 map on April 15, 2020 [19].

### Table 5

|       | A          | B          | C          |
|-------|------------|------------|------------|
| X     | 0.97895    | 0.987922   | 0.883241   |
| Y     | 0.946345   | 0.908965   | 0.984906   |

Fig. 21. Attractors on USA maps including the data for the confirmed cases on March 26, 2020 and airports.

Fig. 22. Attractors with nearest airports.
Fig. 23. TWC – $\alpha$ paths and TWC – $\beta$ maps for attractor points $A$ and the nearest airport to the attractor $A$. 
Fig. 24. $TWC - \alpha$ paths and $TWC - \beta$ maps for attractor points $B$ and the nearest airport to the attractor $B$. 
Fig. 25. TWC $\alpha$ paths and TWC $\beta$ maps for attractor points C and the nearest airport to the attractor C.
the observation maps of COVID-19 on March 22, March 26, April 7 and April 15 (Figs. 2, 6, 9 and 12), show the outbreak pattern remain relatively similar. The number of infected places increases in east side of country, particularly in northeast. Although, the pattern of the outbreak overall is not changing, the outbreak until April 15 is spreading out over the country. It means the intensity differences between them remain almost constant. We know that one of the main hotspots is in New York. In all observations and TWC maps, New York is seen as a hotspot.

Fig. 14a and b, c show the graph of the attractors according to the Markov Chain inferred from TWC – θ. The Non-Linear Minimum Spanning Tree (NL-MST) that represents the minimum-energy tree configuration that connects all the points of the distribution, pruning away all the edges that are not indispensable to maintain the connectedness among points by means of the most relevant interactions [20].

Figs. 14a, b, c show Fig. 22 three clusters A, B and C. The blue dots show the Attractor points. These points correspond to the attractors of the dynamics of the whole process. They are obtained by transforming the nonlinear distance matrix (from which the NLMST is derived) into a Discrete Time Markov Chain [21]. More details can be found in Ref. [8]. Table 2 and Fig. 15 show the Attractors loop for cluster A. Table 3 and Fig. 16 show the Attractors loop for cluster B. Table 4 and Fig. 17 show the Attractors loop for cluster C.

Fig. 18 shows the attractor points for clusters A, B and C on the map of the confirmed cases.

By comparing Figs. 18 and 19 we can see the airport distribution is more intense in the east side (Fig. 19) and the case distribution at March 26, 2020 is more concentrated on the east side (black points in Fig. 18). Figs. 20 and 21 show the attractor on USA maps including the data for the confirmed cases on March 26, 2020 and airports Fig. 22 shows the attractors with nearest airports.

The autocorrelation between the coordinates of the attractor A, B and C with corresponding nearest airports is given in Table 5. The correlation between the x coordinates of attractor C and the corresponding nearest airport is low compared to other attractors. Our hypothesis is that this is due to the rather small number of points of attractors of cluster (4 points).

In order to understand the correlation between the attractor clusters and corresponding nearest airports we use TWC method and find the TWC – α paths and TWC – β maps for these data. Figs. 23–25 show TWC – α paths and TWC – β maps for attractor A, B and C with corresponding nearest airports. Table 6 shows correlation coefficients between the TWC – α paths and TWC – β maps from the attractors to corresponding nearest airports. Since there are only 4 points in attractor C, the correlation coefficients is small.

The other factors such as population density that is an obvious reason can be considered as a cause of outbreak that is mostly a local effect. TWC method also shows the effect of population. However, the transportation is a major factor for outbreak spread over a big country such as USA. If there were no transportation, movement from one location to another, it would essentially a global quarantine.

Table 6
The correlation coefficients between the TWC – α paths and TWC – β maps for the attractors and corresponding nearest airports.

|         | A               | B               | C               |
|---------|-----------------|-----------------|-----------------|
| TWC-α paths | 0.999996732  | 0.99991004 | 0.99991764 |
| TWC-β maps | 0.974477       | 0.978051       | 0.796972       |

3.4. TWC – ι Map

Fig. 26 shows the TWC – ι results. The results of TWC – θ and TWC – ι are in agreement that the outbreak spreads out dominantly via airline routes. Fig. 12 14 shows how TWC – ι identifies the airline routes. Two TWC – ι clusters are located over the main airline routes that connect the USA to the world. The appearance and vanishing of this outbreak and spreading the pandemics seem to be happening mainly through airline routes. Notice that we have used the confirm data in USA on March 26, 2020. The TWC method hypothesizes that the cause of the outbreak is via airline routes coming from outside of USA and spreads mainly inside USA by airline connections.

4. Conclusion

The TWC approach was used to study the COVID-19 in USA. The TWC algorithms are a group of data driven algorithms. Since this is a data driven method, a prior mathematical model is not used in the analysis to predict the disease dynamics. In this paper a new TWC method called TWC-Windowing is used. In this study the confirmed cases on March 26, 2020 were used as a source of the data analysis and no information from other days was used. This method, that has been developed based on the concept of the statistical thermodynamics, the outbreak pattern is projected according to the relationship between the
source points and all other points. The TWC method predicts that the relative differences between all places remain constant and therefore the overall the hotspots (relative to other points) also remain constant. In this analysis, it is hypothesized that the main routes of COVID-19 spread are the airline connection. The TWC maps, including the hotspots locations, are also useful in vaccination and they might help for an effective vaccination strategy.

Author contributions

Masoud Asadi-Zeydabadi: Drafting the original manuscript, theory, adding to the data analysis, and programing.

Massimo Buscema: Reviewing the manuscript, theory, data analysis, and developing the algorithms.

Weldon Lodwick: Reviewing and editing the manuscript and adding to the data analysis.

Giulia Massini: Reviewing the manuscript Theory and adding to the data analysis.

Francesca Della Torre: Reviewing the manuscript and adding to the data analysis.

Francis Newman: Reviewing the manuscript and adding to the data analysis.

Declaration of competing interest

The authors of this manuscript have NO conflict of interest.

References

[1] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen.
[2] https://wwwnc.cdc.gov/eid/article/26-6/20-0233_article.
[3] M.L. Holshue, C. DeBolt, S. Lindquist, K.H. Lofy, , et al.March, First case of 2019 novel Coronavirus in the United States”, N. Engl. J. Med. 382 (10) (2020) 929–936, https://doi.org/10.1056/NEJMoa2001191.
[4] https://www.cdc.gov/media/releases/2020/s0226-Covid-19-spread.html.
[5] https://coronavirus.jhu.edu/map.html.
[6] https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_daily_reports/03-26-2020.csv.
[7] https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_daily_reports/02-20-2020.csv.
[8] M. Buscema, G. Massini, P. Sacco, The Topological Weighted Centroid (TWC): a topological approach to the time-space structure of epidemic and pseudo-epidemic processes, Physica A (2018) 492.
[9] M. Buscema, M. Breda, G. Catzola, in: M. Buscema, E. Grossi (Eds.), The Topological Weighted Centroid, and the Semantic of the Physical Space – Theory in Artificial Adaptive Systems in Medicine, Bentham, 2009, pp. 69–78.
[10] E. Grossi, M. Buscema, T. Jefferson, The topological weighted centroid, and the semantic of the physical space – application, in: M. Buscema, E. Grossi (Eds.), Artificial Adaptive Systems in Medicine, Bentham, 2009, pp. 79–89.
[11] M. Buscema, E. Grossi, A. Brunstein, W. Lodwick, M. Asadi-Zeydabadi, R. Benzi, F. Newman, A new algorithm for identifying possible epidemic sources with application to the German Escherichia Coli outbreak, ISPRS Int. J. Geo-Inf. 2 (2013) 155–200, https://doi.org/10.3390/ijgi2010155.
[12] Massimo Buscema, Massoud Asadi-Zeydabadi, Weldon Lodwick, , Alphonse Nde Nembot, Alvin Brunstein, Francis Newman, Ebola Paper, Analysis of the Ebola Outbreak in 2014 and 2018 in West Africa and Congo by Using Artificial Adaptive System, Applied Artificial Intelligence, 2020.
[13] T. Jefferson, E. Grossi, M. Buscema, Powerful computerized spatial epidemiology and semantics through the use of novel mathematical objects: can artificial intelligence systems identify outbreak sources?, in: NAFIPS 2008 - 2008 Annual Meeting of the North American Fuzzy Information Processing Society IEEE, New York City, NY, USA, 2008, pp. 19–22.
[14] M. Buscema, E. Grossi, M. Breda, T. Jefferson, Outbreaks source: a new mathematical approach to identify their possible location, Physica A 388 (2009) 4736–4762.
[15] https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_daily_reports/03-22-2020.csv.
[16] https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_daily_reports/03-10-2020.csv.
[17] https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_daily_reports/03-26-2020.csv.
[18] https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_daily_reports/04-07-2020.csv.
[19] https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_daily_reports/04-15-2020.csv.
[20] M. Buscema, P.L. Sacco, MST Fitness Index and implicit data narratives: a comparative test on alternative unsupervised algorithms, Physica A 461 (2016) 726–746.
[21] G. Bolch, S. Greiner, H. de Meer, K.S. Trivedi, Queueing Networks and Markov Chains, second ed., John Wiley, New York, 2006.