

# Analysis of the COVID-19, Outbreak in Brazil Using Topological Weighted Centroid: An Intelligent Geographic Information System Approach

Masoud Asadi-Zeydabadi<sup>1</sup>, Marina Mizukoshi<sup>2</sup>, Massimo Buscema<sup>2,3</sup>, Giulia Massini<sup>3</sup>, Weldon Lodwick<sup>4</sup>

<sup>1</sup>Department of Physics, University of Colorado, Denver, Colorado, USA

<sup>2</sup>Institute of Mathematics and Statistics, Federal University of Goiás, Goiania, Goiás, Brazil

<sup>3</sup>Semeion Research Center, Rome, Italy

<sup>4</sup>Department of Mathematical and Statistical Sciences, University of Colorado, Denver, Colorado, USA

Email: Masoud.Asadi-Zeydabadi@UCDenver.edu, Tuyako@ufg.br, M.Buscema@Semeion.it, Weldon.Lodwick@UCDenver.edu

**How to cite this paper:** Asadi-Zeydabadi, M., Mizukoshi, M., Buscema, M., Massini, G. and Lodwick, W. (2024) Analysis of the COVID-19, Outbreak in Brazil Using Topological Weighted Centroid: An Intelligent Geographic Information System Approach. *Journal of Data Analysis and Information Processing*, 12, 248-266.  
<https://doi.org/10.4236/jdaip.2024.122014>

**Received:** April 4, 2024

**Accepted:** May 24, 2024

**Published:** May 27, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

This study used Topological Weighted Centroid (TWC) to analyze the Coronavirus outbreak in Brazil. This analysis only uses latitude and longitude in formation of the capitals with the confirmed cases on May 24, 2020 to illustrate the usefulness of TWC though any date could have been used. There are three types of TWC analyses, each type having five associated algorithms that produce fifteen maps, TWC-Original, TWC-Frequency and TWC-Windowing. We focus on TWC-Original to illustrate our approach. The TWC method without using the transportation information predicts the network for COVID-19 outbreak that matches very well with the main radial transportation routes network in Brazil.

## Keywords

COVID-19, Topological Weighted Centroid (TWC) Algorithms, TWC-Original, TWC-Frequency and TWC-Windowing

## 1. Introduction

Modeling the dynamics of pandemics or epidemics, mathematically, is done via model driven analyses, the susceptible-infected-recovered (SIR) [1] differential equations models, but in general some simplifications in the model are needed to do a mathematical analysis because the free parameters needed are obtained

via statistical methods. This requires levels of precision that typically cannot be guaranteed by the data. A way to supplement the differential equations approach is to consider observed properties of the diffusion dynamics [2] [3] [4] instead of considering the free parameters. More recently a data-driven mathematical analysis has been developed, the topological geometric approach called Topological Weighted Centroid (TWC). According to [2] [3] [4], the Topological Weighted Centroid makes it possible to estimate the location of the both source of the outbreak and the dynamic unfolding of the epidemic by working only on the spatial distribution of the events, without any reference to the chronology and frequency of the observations, that is, on the basis of the purely spatial characteristics of the phenomenon.

The World Health Organization (WHO) in China, reported, in December of 2019, a pneumonia with unknown cause in Wuhan, China. The new virus, the corona virus or SARS-COV2, caused the COVID-19 disease, which spread and is spreading over the world.

The first case of COVID-19 in Brazil was reported on February 26, 2020 in São Paulo [5]. A person came back on a flight from Italy together with 300 other passengers who went to other destinations in Brazil besides São Paulo including Rio de Janeiro, Porto Alegre, Salvador, Curitiba, Belo Horizonte, Fortaleza, Recife, Vitória and Florianópolis, contributing to the nationwide spread of the disease [5]. The first reported death was in São Paulo city on March 16, 2020 and after March 20, 2020 the community spread of COVID-19 was happening in all the country, according Ministry of Health [6]. In Feira de Santana, a city in the state of Bahia on March 6, 2020 [7], in Manaus, a city in the state of Amazonas on March 13, 2020 [8], in Brasília, a city Capital of Brazil on March 6, 2020 [9], in Porto Alegre, a city in the state of Rio Grande do Sul on March 10, 2020 [10] were the first reported infections in Northeast, North, Center-West and South regions, respectively.

Nicolelis *et al.* highlighted that while international airports served as primary entry points for SARS-CoV-2, the drivers behind the uneven geographic spread of COVID-19 cases and deaths in Brazil are largely undisclosed. Factors such as super-spreader cities, highways, and the accessibility of intensive care facilities are pivotal in the outbreak dynamics of COVID-19 [11].

Castro *et al.* conducted daily data analysis to grasp, quantify, and compare the spatiotemporal dynamics of COVID-19 spread across municipalities. This entailed examining clustering, trajectories, speed, and intensity of virus movement towards interior regions, in addition to assessing policy measure indices. The analysis reveals a multifaceted scenario where diverse factors play roles in the spread [12].

The Topological Weighted Centroid algorithms are used here in the analysis of COVID-19 in Brazil. Topological Weighted Centroid works on the basis of an optimization principle, that is, it adaptively explores the fitness landscapes with respect of certain quantities (free energy and entropy), and determines the optimal solution accordingly. The Topological Weighted Centroid approach is a

data driven intelligent geographical dynamical system that models disease spread in space and time [2] [3] [4], whereas in a model driven analysis, such as the SIR models encodes the understanding of the behavior of a system, corona virus causing COVID-19 in the case at hand. Model driven models fit a predetermined cause/effect relationship to the data into equations and mathematical relationships. Their cause/effect is the explanation of what occurs and the data fits into the model. On the other hand, a data driven analysis is one in which the set of cause/effect relationships is elicited from the data. Topological Weighted Centroid is the latter type of analysis. The confirmed cases on May 24, 2020 for this study are used [6] to illustrate the potential of Topological Weighted Centroid. We use May 24, 2020 since the data seems more reliable. Clearly, any date could be used given reliable data. In fact, the analysis can be done “real time” updating the data as needed.

The problems that Topological Weighted Centroid is used to address in this paper are: 1) Given data of COVID cases on various dates, determine, from which places the impetus for the spread is most accelerating. That is, from what focal areas, what geographical location, “hot spots”, would the central government locate, warehouse, its resources (for example, vaccines, ventilators, mobile medical personnel) to deal most efficiently, with the given data distribution of cases. The opposite view is the following. If the pandemic source (COVID-19 source) were to have located itself to most efficiently infect the population in the configuration of the reported data where would this be? We hypothesize that it is at the indicated “hot spot(s)” of our output maps. 2) Where will the effective location be with respect to causing tomorrow’s (future) spread? That is, given past data on various dates, determine where the new center of accelerated spread will be and locate resources at these “hot spots”. 3) What is the geographic network (in the sense of an undirected graph) over which the disease is spreading. Or positively, what is the network over the medical resource can flow in a most effective way to treat the outbreak.

It’s important to note that the focus of this presentation is descriptive, while the underlying theory is detailed in various publications, such as [2] [3]. Additionally, Nicolelis *et al.* [11] explored transportation data to analyze the impact of primary routes on the geographic spread of COVID-19 cases, pinpointing Brasília as the hub of Brazil’s main radial routes. Utilizing the TWC- $\alpha$  algorithm, they identified the epicenter near Brasília, employing statistical thermodynamics concepts. Remarkably, the network routes derived from transportation data in [11] closely resemble those obtained by the TWC- $\theta$  algorithm, which solely utilizes latitude and longitude data for state capitals, without incorporating transportation information.

## 2. Topological Weighted Centroid Method

Topological Weighted Centroid only considers one of type of data, latitude/longitude of the COVID-19 cases. The Topological Weighted Centroid method has been outlined in detail and can be found in [2] [3]. This data driven

approach was developed based on statistical thermodynamics, diffusion theory, optimizing free energy and entropy. In this method, the probability of spread of the disease to  $(x_i, y_i)$  is given in term of energy, which is a function of a meta distance to the infected points. In statistical thermodynamics the probability distribution for  $i$ -th state is given by

$$p(E_i) = \frac{e^{-\frac{E_i}{K_B T}}}{Z} \quad (1)$$

where  $E_i$  is the energy for  $i$ -th state,  $K_B$  is Boltzmann constant,  $T$  is the absolute temperature and  $Z = \sum_{i=1}^N e^{-\frac{E_i}{K_B T}}$  is a partition function for  $N$  states. Generally speaking, in the context of COVID-19 and epidemics, the TWC algorithms identify the center locations from which diseases are accelerating. These locations are calculated as past sites, present sites, near future sites (over the next weeks), and longer future sites (several weeks ahead). There are three types of TWC, TWC-Original [2] [3], TWC-Frequency and TWC-Windowing. The first type, TWC-Original, is based on the disease's occurrence not considering the frequency of cases. The second type, TWC-Frequency, is where frequency is explicitly incorporated. The third type, TWC-Windowing, is a version that balances weak outliers with strong influences, that is, it is an approach that balances weaker local dynamics with stronger global dynamics. There are five algorithms, which are applied to each of these three types where one of these five creates an undirected graph between outbreak hotspots. The way this is done is by breaking the entire region in subregions, running TWC on the sub-regions then joining the regions together. The five TWC methods are called TWC- $\alpha$ , TWC- $\beta$ , TWC- $\gamma$ , TWC- $\theta$  and TWC- $\iota$  [2] [3].

1) The  $TWC_\alpha$  algorithm estimates the location of the origin where, in the past, according to the data, a statistical thermodynamic, optimization of free energy, and entropy point of view, where the acceleration (of COVID-19 in this case) is coming. It is the location that is driving the dynamics of the disease as manifest by the data. To use an analogy, if a group of COVID virus "benefactors" (CVB) were to affect the most COVID benefit (e.g. providing sufficient protective gear) to the locations as represented by the distribution of the data as reported and used in the analysis, from where would the CVB impetus arise to benefit the cases reported by the data in the most efficient way. The ideal point area is not the region with the high concentration of events, but where the global entropy of the distances from all the other points attains its minimum level. According to [2] [3], the TWC of all the points of the dataset for any specific value is (see [2] [3]):

$$TWC_x(\alpha_n) = \frac{\sum_{i=1}^N w_i(\alpha_n) x_n}{\sum_{i=1}^N w_i(\alpha_n)}, \quad (2)$$

$$TWC_y(\alpha_n) = \frac{\sum_{i=1}^N w_i(\alpha_n) y_n}{\sum_{i=1}^N w_i(\alpha_n)}, \quad (3)$$

where  $w_i(\alpha_n) = \frac{1}{N-1} \sum_{j=1, i \neq j}^N e^{-\frac{\bar{d}_{i,j}}{D} \alpha_n}$  and  $\bar{d}_{i,j} = \frac{1}{N-2} \sum_{k \neq j, k \neq i}^N d_{i,k}$ .

In the above equations  $\bar{d}_{i,j}$  is called the indirect distance that controls and is the contraction/expansion of original space considering the levels of organization and is the measure of the energy needed for a given distribution of points in space to collapse their spatial organization into a single point.  $d_{i,k}$  is the Euclidian distance between 2 points,  $\alpha \in [0, +\infty)$  modulates the attraction strength of points and it is equivalent to the inverse of the temperature of a thermodynamic system and  $\alpha_n$  are increment of the temperature  $\alpha_0 = 0$ ,  $\alpha_n = \alpha_{n-1} + \varepsilon$ . To analyze the potential attraction strength among the points of distribution, the free energy

$$F(\alpha_n) = \frac{-\sum_{i=0}^N w_i(\alpha_n)}{\alpha_n}, \tag{4}$$

is calculated. Iteratively,  $TWC_x(\alpha_n)$ ,  $TWC_y(\alpha_n)$  are calculated according to (2, 3) and (3) and subsequently the free energy according to (2.4). The free energy will increase  $F(\alpha_k) < F(\alpha_{k+1})$  until after some point  $k$ ,  $\alpha^* = \alpha_k$ , when  $F(\alpha_{k+1}) < F(\alpha_k)$ . That is  $F(\alpha^*)$  corresponds the optimal value [2]. Then, (2, 3) and (3) are written as

$$TWC_x(\alpha^*) = \frac{\sum_{i=1}^N w_i(\alpha^*) x_i}{\sum_{i=1}^N w_i(\alpha^*)}, \tag{5}$$

$$TWC_y(\alpha^*) = \frac{\sum_{i=1}^N w_i(\alpha^*) y_i}{\sum_{i=1}^N w_i(\alpha^*)}, \tag{6}$$

The more a given point is closer to the trajectory defined by the vector of  $\alpha$  points, then more that point has a strong activation value. It is important to stress that such an inference may only be correct if the observed distribution of points represents a statistically significant sample of the underlying phenomenon and in the absence of spurious observations. The process of estimation is  $m_{k,j} = \sqrt{(x_k - TWC_x(\alpha_j))^2 + (y_k - TWC_y(\alpha_j))^2}$ , where  $j$  is the index for each  $TWC-\alpha$  points,  $k$  index for each point of the plane, and the activation strength of each point of the 2D plane is

$$\alpha_k = \frac{1}{N_t} \sum_{j=1}^{N_t} e^{-\frac{m_{k,j} \beta^*}{d_M}}, \tag{7}$$

Here,  $N_t$  is the total numbers of TWC points,  $d_M = \max_{i,j} \{d_{i,j}\}$  distance between points and  $\alpha_k \in [0,1]$ .

2) The TWC- $\beta$  algorithm calculates the condition of the disease distribution of current time as it will affect tomorrow's outbreak. That is, to affect future disease spread tomorrow, TWC- $\beta$  positions itself where the accelerating impetus would locate itself to actuate the data that is reported tomorrow. Here the indirect distance, in which each point interacts also with itself. The difference in this algorithm is that TWC- $\beta$  moves back toward the center of mass, because the interaction of each point with itself takes over and the attraction strengths tend to

collapse into self-attraction. However, we obtain the maximum from

$$\beta^* = \max_{\beta_n} \{d(TWC(\beta_n) - CM)\}, \quad (8)$$

where  $d$  is the Euclidian distance and  $CM$  center of mass

$$TWC_x(\beta_n) = \frac{\sum_{i=1}^N w_i(\beta_n) x_i}{\sum_{i=1}^N w_i(\beta_n)}, \quad (9)$$

$$TWC_y(\beta_n) = \frac{\sum_{i=1}^N w_i(\beta_n) y_i}{\sum_{i=1}^N w_i(\beta_n)}, \quad (10)$$

where  $w_i(\beta_n) = \frac{1}{N-1} \sum_{j=1, j \neq i}^N e^{-\frac{\bar{d}_{i,j}}{D} \beta_n}$  and  $\beta_n = \beta_{n-1} + \varepsilon$ . Each point carries a specific value and meaning of its own. Given a certain distribution of points, each already existing point is going to play a role in the future unfolding of the phenomenon once the local/global tradeoff is further influenced by the emergence of new points. The  $\beta$  Map, instead, defines the probability to observe new points (events) that belong to the same distribution (that is, to the same data generating process). The  $\beta$  Map can therefore be regarded as a snapshot of a probability density function which is implicit into the structure of the data set. It is generated on the basis of the weighted distance between each generic point of the space and all of the points belonging to the observed distribution.

3) The TWC- $\gamma$  algorithm will make outbreak calculations over a short time span about one week into the future. The  $\gamma$  Map evaluates how the variations in the positions of certain points may influence the whole spatial structure from the perspective of any single point belonging to it, in its relations with the other points. It acts as a modifier of the Euclidean distance between points, so as to suitably “tune” the optimal deformation needed to capture the actual spatial organization of the points according to a logic that is similar to the one already followed for the construction of the  $\alpha$  and  $\beta$  Maps. It is the movement of some individual carrying the epidemic agent from one location to another, or in the case of a quasi-epidemic process such as a territorial network of towns, the interaction may consist in trade and social exchanges, migrations of families and workers, and so on. At the early stages of the process, when the phenomenon is still emerging, one may assume that the interactions are mostly stochastic, that is, the fact that the infectious agent travels in a certain direction or that certain trade relations begin to be established among certain locations rather than others may be due to random factors. If this is the case, we can conjecture that, statistically, most of such random early interactions are spatially mediated by the center of mass, which represents the average of the coordinates of the all the points belonging to the spatial distribution. For this reason, the matrix of all the TWC vectors describing the trajectories of convergence to the various points of the distribution moving away from the center of mass provide a good approximation of the statistics of the interaction among the points at early phase. Moreover, the matrix of trajectories gives useful information about the different attraction

strengths of the various points. The more a trajectory is curved, the more it is affected by the attraction of other points located in the corresponding direction (as a sort of “gravitational” effect). Therefore, the trajectories whose shape is close to a straight line identify the mainstream of the diffusion process, as they are relatively less affected by the attraction strength of the other points while at the same time possibly affecting the trajectories of such points. The  $\gamma$  scalar field may therefore be thought of as a further qualification of the  $\beta$  field, which transforms attraction strengths into intensities of network interaction and therefore highlights longer-term activation patterns within the spatial distribution. As the  $\beta$  field is a “prediction” of the evolution of the  $\alpha$  field, the  $\gamma$  field may in turn be regarded as a “prediction” of the evolution of the  $\beta$  field. The associated trajectory is calculated as follow

$$TWC_x(\gamma_i(t)) = \frac{1}{\sum_{j=1}^N p_{i,j}(\gamma_i(t))} \sum_{i=1}^N p_{i,j}(\gamma_i(t)) x_j, \quad (11)$$

$$TWC_y(\gamma_i(t)) = \frac{1}{\sum_{j=1}^N p_{i,j}(\gamma_i(t))} \sum_{i=1}^N p_{i,j}(\gamma_i(t)) y_j, \quad (12)$$

where  $p_{i,j}(\gamma_i(t)) = e^{-\frac{d_{i,j}}{D} \alpha_n}$ ,  $\gamma_i(t+1) = \gamma_i(t) + \varepsilon$ .  $\lim_{\gamma_i(t) \rightarrow \infty} TWC_x(\gamma_i(t)) = x_i$  and  $\lim_{\gamma_i(t) \rightarrow \infty} TWC_y(\gamma_i(t)) = y_i$ .

4) TWC- $\theta$  is used to predict the longer-range time evolution of the disease outbreak. In addition, associated with TWC- $\theta$  is an undirected graph between hot spots, which may be interpreted as the pathways of the disease. The Non-Linear Minimum Spanning Tree (NLMST) is computed as follows

$$\theta_{i,j} = \sum_{i,j}^{z_{i,j}-1} d_s(TWC_{\theta_{i,j}}(t), TWC_{\theta_{i,j}}(t+1)), \quad (13)$$

where  $d_s$  is the distance between two points of the same path.

$$TWC_{\theta_{i,j}x}(t) = x_j + (TWC_x(\gamma_i(t)) - TWC_x(\gamma_j(t))),$$

$$TWC_{\theta_{i,j}y}(t) = x_j + (TWC_y(\gamma_i(t)) - TWC_y(\gamma_j(t))),$$

$$z_{i,j} = \min\{Q_i, Q_j\},$$

and  $Q_i$  is the number of the new points belonging to the vector  $TWC(\gamma_i(t))$ ,  $i \in N$ . The  $\theta$  Map allows us to build the MST that links together the points of the spatial distribution in terms of a minimal network of influence. This corresponds to the deepest structural layer explored so far. The NLMST serves as the basis to build the  $\theta$  scalar field. The distance matrix  $\theta_{i,j}$  may be easily transformed into a probability matrix,  $p_{i,j}$ , which estimates the probability any given point in the spatial distribution communicates with another one as the effect of some random external shock (for instance, a local shortage of some key resource for quasi-epidemic processes, or the occurrence of a mutation of the infective agent in an epidemic process). The closer two points in the spatial distribution, the more likely that the random shock will force an interaction between

them. This Markovian process allows us to attain yet another level of depth in our analysis, in that we not only understand the minimal connectivity structure of the urban network, but also an estimate of the intensity of the exchange flows along the structure. We can therefore study an explicit dynamical model that provides still another layer of interpretation of our spatial distribution, and in particular allows us to compute the attractor of the dynamics.

5) The TWC- $l$  algorithm, TWC- $l$  estimates the points at which the outbreak is thought to vanish [2] [3]. In the context of the COVID-19 pandemic, TWC- $l$  predicts where the disease will begin to stop, the last “outpost” or location of the disease.

The three types of TWC analyses, each type having five associated algorithms, produce fifteen maps.

TWC-original which is the foundation of this method is discussed in [2] [3]. TWC-Frequency and TWC-Windowing include the number of confirmed cases, the outbreak frequency.

The TWC-Frequency produces new  $N(i)$  points corresponding to the number of cases reported at the city located at random in the neighborhood of the original  $i$ -th source. Here, “source”, for the Brazil data, is the state capitol. The location of new  $N(i)$  points is calculated by  $N(i) = \text{Ceil}(1 + \log f(i))$ , where  $f(i)$  is the frequency at the  $i$ -th source point. The  $x$  coordinate of the new points are given by  $x(i) = x_s(i) + 2(\text{rand} \times 0.05) - 1$ , where  $x_s(i)$  is the  $x$  coordinate of the  $i$ th source, and  $0 \leq \text{rand} \leq 1$  is a random number. The same method is applied for the  $y$  coordinates. Then TWC-Frequency applies TWC-Original to all the generated points.

TWC windowing uses a selected number of points for the window. The size of the window (the data points in each window) is the same for all points. Each window works as a network around a point. The TWC algorithms are applied on the data in each window. For a data set of  $N$  data points, there are  $K$  windows. The windows may (often do) have some common points. The common points make connection between the windows. Finally, all TWC maps that are produced for all windows are overlapped for the final TWC maps. In TWC procedure an integer number  $K < N$  is selected. Then the number of the point in each window is determined by  $P = N/K$ . Of course, as  $K$  get bigger, the number of the data points in a window,  $P$ , get smaller.

The Brazil data that is available does not give the infected person’s address, only the city. Thus, there are many cases occurring in the same city. Moreover, what is typical of the Brazil data set is that cases are reported only by state. So, one only knows the state in which the cases occur. However, more recently, outbreak by municipality has been reported. This study did not use municipality data in its analysis, only state data. The municipality data was used to compare the TWC with the reported results. However, the use of municipality data is straight forward in principle.

**Table 1** shows the latitude and longitude coordinates of the capitals with con-

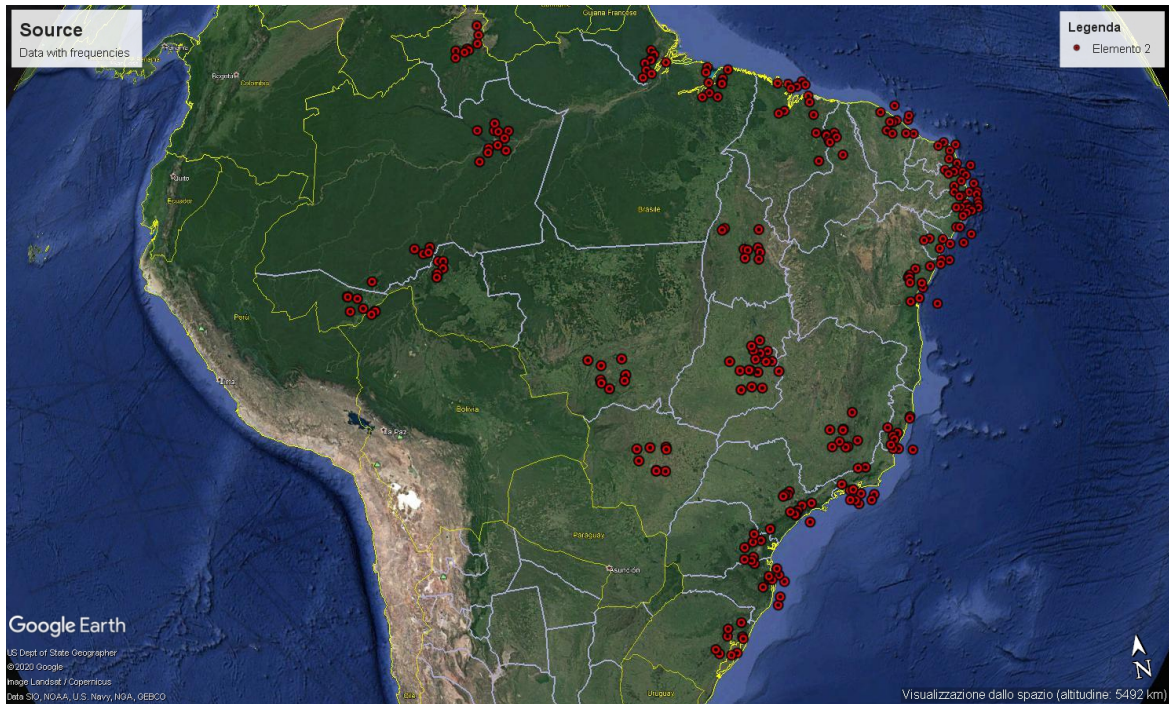
firmed COVID-19 cases on May 24, 2020 in Brazil. The exact locations of the cases of the original, as mentioned, are not given. As was mentioned, the capitals of each state were used as the location of the outbreak. In this paper the results of TWC-Original are discussed.

### 3. Results

This study used TWC method for analyses of COVID-19 in Brazil and compares the TWC results with some given maps and data by Brazilian authorities as empirical sources. The first map, **Figure 1**, that is given, is just that of the capital of

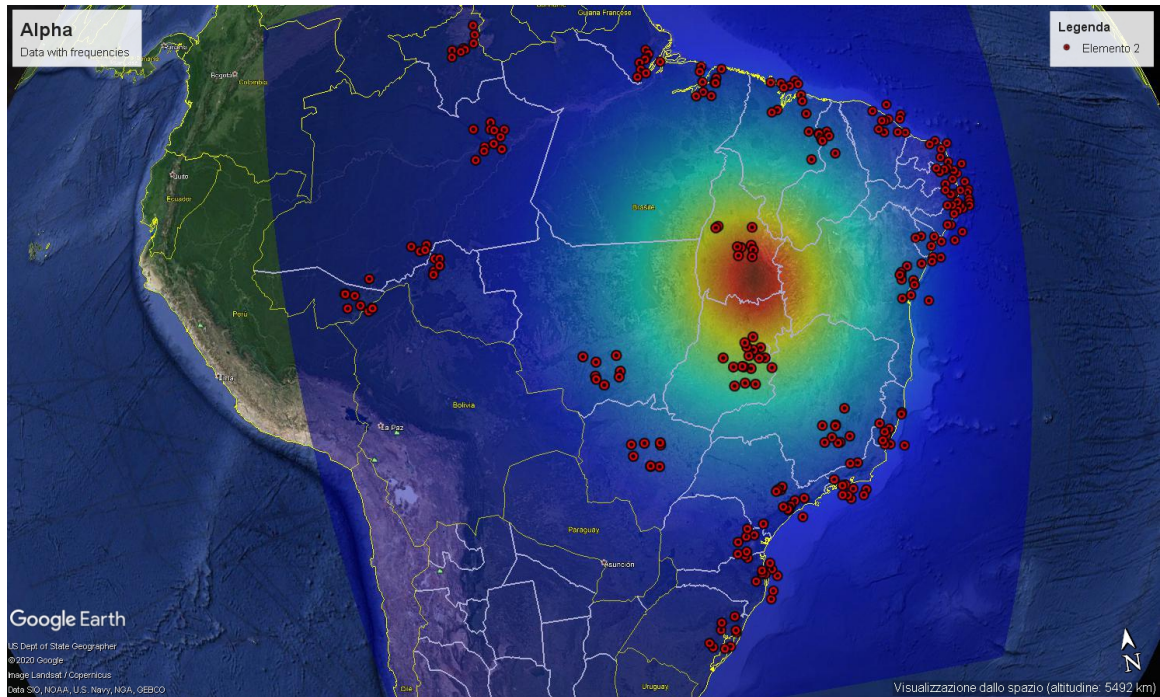
**Table 1.** Latitude and longitude coordinates of the capitals with confirmed COVID-19 cases on May 24, 2020 in Brazil.

State	Capital	Latitude	Longitude
Alagoas	Maceió	-9.66625	-35.7351
Bahia	Salvador	-12.9704	-38.5124
Amazonas	Manaus	-3.10719	-60.0261
Amapá	Macapá	0.0344566	-51.0666
Distrito Federal	Brasília	-15.7801	-47.9292
Ceará	Fortaleza	-3.71839	-38.5434
Espirito Santo	Vitória	-20.3222	-40.3381
Minas Gerais	Belo Horizonte	-19.8157	-43.9542
Goiás	Goiânia	-16.6799	-49.255
Maranhão	São Luís	-2.53073	-44.3068
Mato Grosso Sul	Campo Grande	-20.4435	-54.6478
Mato Grosso	Cuiabá	-15.5989	-56.0949
Pernambuco	Recife	-8.05428	-34.8813
Paraíba	João Pessoa	-7.11532	-34.861
Pará	Belém	-1.45502	-48.5024
Piauí	Teresina	-5.08921	-42.8016
Paraná	Curitiba	-25.4284	-49.2733
Rio de Janeiro	Rio de Janeiro	-22.9035	-43.2096
Rio Grande do Norte	Natal	-5.79448	-35.211
Rondônia	Porto Velho	-8.76183	-63.902
São Paulo	São Paulo	-23.5489	-46.6388
Rio Grande do Sul	Porto Alegre	-30.0277	-51.228
Santa Catarina	Florianópolis	-27.5969	-48.5495
Roraima	Boa Vista	2.81954	-60.6714
Acre	Rio Branco	-9.974	-67.8076
Tocantins	Palmas	-10.1689	-48.3317
Sergipe	Aracaju	-10.9095	-37.0748

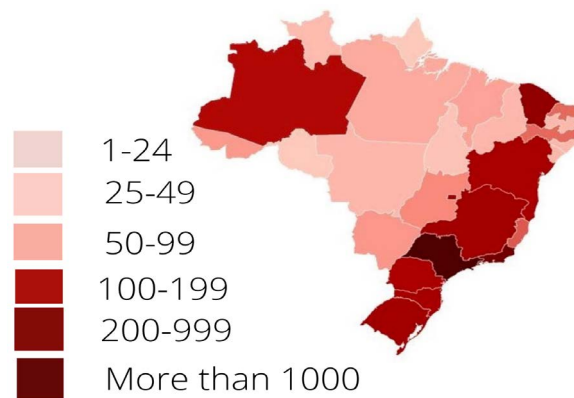


**Figure 1.** The source data, May 24, 2020 in Brazil.

the states with confirmed cases on May 24, 2020. We have used only the location information (latitude and longitude) of the infected places for the TWC analyses. We did not use the number of cases (frequency) for May 24, 2020. One of the advantages of TWC method is using minimum information for predicting disease outbreak. This is particularly useful, since at the beginning of an outbreak, the amount of information is insufficient and access to the reliable information is limited. **Figure 2** shows the results of the TWC- $\alpha$  algorithm and shows the acceleration area, where the impetus for COVID-19 spread centers itself. For the outbreak distribution in Brazil on May 24th the impetus is near Brasília, based on statistical thermodynamics concepts, on border between Goiás and Minas Gerais states. This is an area where the disease would locate itself in the past if one were to affect the infection distribution “today” (May 24th). Basically, it says that Brasília is in effect centrally located with respect to the initial distribution of COVID cases. A hot spot on the TWC- $\alpha$  map, Brasília, can be viewed as the activation location for the disease dynamic. It means if the dynamics start from a location in order to obtain the given outbreak data (the confirmed case on May 24) the geographical coordinates start from the hot spot of the TWC- $\alpha$  map. The TWC- $\alpha$  map (in **Figure 2**) shows a hot spot that can interpreted as on average or a global perspective of the outbreak. The COVID-19 map of Brazil on March 26 as developed by the Health Ministry is shown on **Figure 3**. This figure shows, in addition to the hot region in São Paulo, there is a high number of infections in a very small region (point) that is located in Brasília. This observation to some extent, is in agreement with TWC- $\alpha$  results. The peak of the hot region on the TWC- $\alpha$  map is located on the south of



**Figure 2.** TWC- $\alpha$  map.



**Figure 3.** COVID-19 map in Brazil on March 26, 2020.

Tocantins near the border of Goiás which is not place with high infection in **Figure 3**. We may explain this difference because of several reasons: TWC-original is not using the frequency (the number of cases) and the effect of the hot region on other places like Amazônia, shifts the location of the center of the TWC- $\alpha$ . The TWC- $\alpha$  result using TWC-Original does not capture the local hot spots as they are distributed over the country. However, if we were to use the TWC-windowing, then the hot spot distribution can be captured as occurred with the analysis done for USA-COVID (see [4]). As we mentioned the focus of this study is using TWC-original. In practical perspective, we may also consider the location of TWC- $\alpha$  point or the hotspots of TWC- $\alpha$  map as a useful place for managing the COVID-19 outbreak. For example, we may locate the vaccination center or head-quarter for controlling the pandemic. **Figure 3** is an empirical evidence for TWC- $\alpha$ .

As mentioned, the TWC- $\alpha$  map shows a hotspot region around Brasília, this is in agreement with a small hotspot in Brasília on March 24, 2020. According Nicolelis *et al.* [11] Brasília should be considered as an important hotspot to the propagation of COVID-19, because it is the center of the main radial routes in Brazil.

**Table 2** reports COVID-19 cases for May 15, May 24, May 28 and June 7, 2020 [6]. The data in **Table 2** is an empirical source and they were not used as input for TWC method.

The Relative Changes ( $RC$ ) of COVID-19 cases from day  $n$  to day  $n'$  for data given in **Table 2** were calculated as follows,  $RC(n, n') = \frac{N' - N}{N}$ , where  $N$  and  $N'$  are the infected cases in relation the days  $n$  and  $n'$ , respectively.

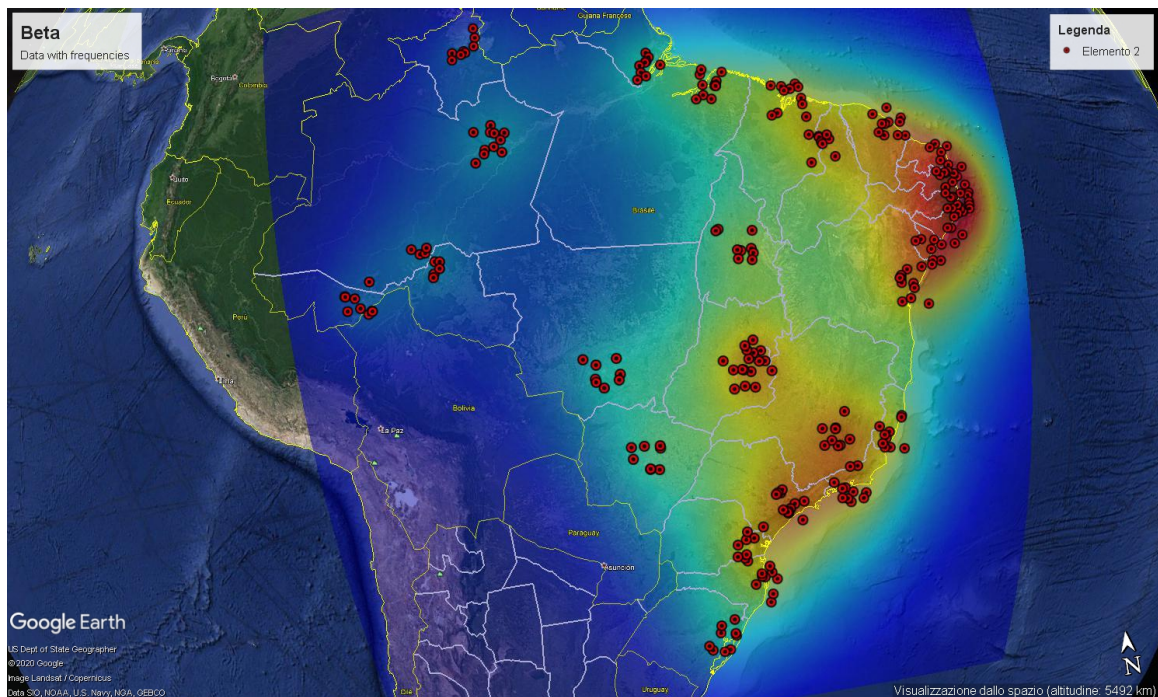
**Table 2.** Data: COVID-19 cases for May 15, May 23, May 28 and June 7, 2020.

State	May, 15th	May, 24th	May, 28th	June, 7th
Acre	1785	4160	5600	7983
Amazonas	18,392	29,867	36,146	49,269
Amapá	3630	6353	8152	13,045
Pará	12,109	24,125	33,699	54,271
Rondônia	1789	3201	4252	8110
Roraima	1789	2439	2959	5529
Tocantins	1179	2591	3277	5644
Alagoas	3212	6214	7580	15,706
Bahia	8128	13,899	15,963	28,250
Ceará	22,490	35,595	37,821	64,271
Maranhão	10,739	21,191	27,979	47,593
Paraíba	3739	7823	11,132	20,310
Pernambuco	16,209	27,759	30,713	40,242
Piauí	1905	3390	4243	7478
Rio Grande do Norte	2786	4697	5630	9452
Sergipe	2868	5314	6156	9290
Espírito Santo	6198	10,007	12,203	19,619
Minas Gerais	4196	6668	8686	15,703
São Paulo	58,378	82,161	95,865	143,073
Rio de Janeiro	19,987	37,912	44,886	67,756
Distrito Federal	3787	6638	8300	16,623
Goiás	1572	2443	3090	5813
Mato Grosso Sul	479	924	1262	2253
Mato Grosso	779	1462	1951	3878
Paraná	2139	3212	3984	6897
Rio Grande do Sul	3597	6470	8234	12,139
Santa Catarina	4562	6696	8000	11,565

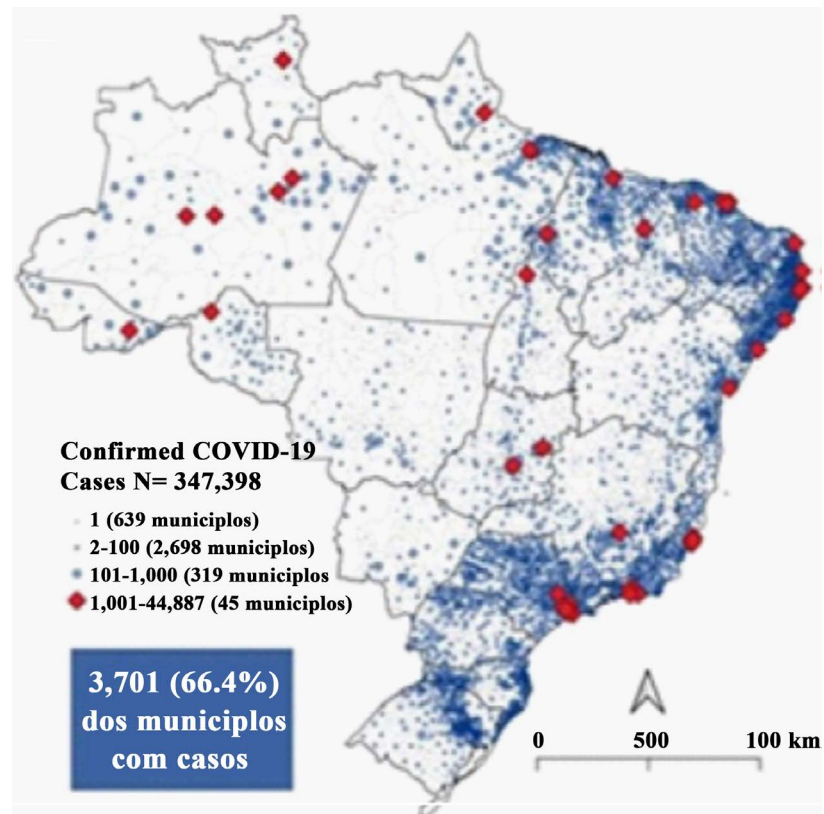
The relative changes from data on **Table 2** from May 24 to May 28, 2020 for Brasília (Distrito Federal) are 0.25 and for Sao Paulo is 0.17. This empirical result again is in agree with the result of TWC- $\alpha$  and it is an evidence that Brasília is an important hotspot. The results in Nicoletis *et al.* (see Fig. 2 [11]) show the contribution of Brasília in covid-19 outbreak around March 20, 2020.

Castro *et al.* [12] used daily data from epidemiological week 9 (February 23-29, 2020) to week 41 (October 4-10, 2020). They calculated the geographic center of the epidemic. They found the national trajectory of the centers started from São Paulo and moved north then turned south. This trajectory is located in the TWC- $\alpha$  hotspot and has a behavior similar to TWC- $\alpha$  path (see Fig. 2A in [12]).

**Figure 4** shows the TWC- $\beta$  map, which is a map of where the disease acceleration would locate itself today (May 24th) to affect tomorrow's future outbreak. Basically, TWC- $\beta$  shows the outbreak around the present time (around May 24). The outbreak is dominated in Northeast, Southeast and Central regions. The hotspot with high intensity is located more towards the Northeast region. The comparison between the results of TWC- $\alpha$  and TWC- $\beta$  indicates the outbreak whose initial "energy" started from the Central region (Brasília) expanded to the Northeast, Southeast and Central regions. Comparing **Figure 4**, TWC- $\beta$  map, with **Figure 5**, the map of the COVID-19 cases on May 23, indicates the outbreak spreads intensively over the eastern side of Brazil. The TWC- $\beta$  does not seem to capture very well, the reported outbreak in the northwest of the country. However, we need to note that the TWC maps represent the centers of intensity of the outbreak. Moreover, the COVID-19 maps that are given by authorities are given by the number of



**Figure 4.** TWC- $\beta$  map.



**Figure 5.** COVID-19 map of Brazil on May 23, 2020.

confirmed cases without consideration of the population density and the area of the regions. By comparing **Figure 3** and **Figure 5**, one can see, in spite of a large increase in the number of cases, the dynamics in the geographical perspective is not changing significantly. **Figure 5** shows the outbreak map on May 23 and this can be used as an empirical map for  $TWC-\beta$ . This map shows a result well comparable with  $TWC-\beta$  map.

We can also use the relative changes for COVID-19 cases from May 24 to May 28 (using data in **Table 2**). The largest relative changes are in Mato Grosso Sul, Mato Grosso and Paraíba (0.36, 0.33 and 0.42, respectively). Since the cases in Mato Grosso Sul and Mato Grosso are low, these changes indicate the starting of disease dynamics in these states. However, the large relative changes in Paraíba can be considered as an evidence or empirical data, which is in agreement with  $TWC-\beta$  method that shows there is an important hotspot region in northeast.

**Figure 6** shows the  $TWC-\gamma$  map. It can be seen from comparison of **Figure 4** and **Figure 5**, the outbreak would expand through northeast, southeast and central regions and the hot spots merge together and the high intensity moves from northeast back toward the central region. If we compare the maps in **Figure 3**, **Figure 5** and **Figure 8**, there are major outbreak in northeast and southeast/south of the country. These two hotspot places approached together and moved away from each other. The results of  $TWC-\beta$ ,  $TWC-\gamma$  and  $TWC-\theta$  maps (**Figure 4**, **Figure 6** and **Figure 7**) indicate this behavior. This kind of dynamics might

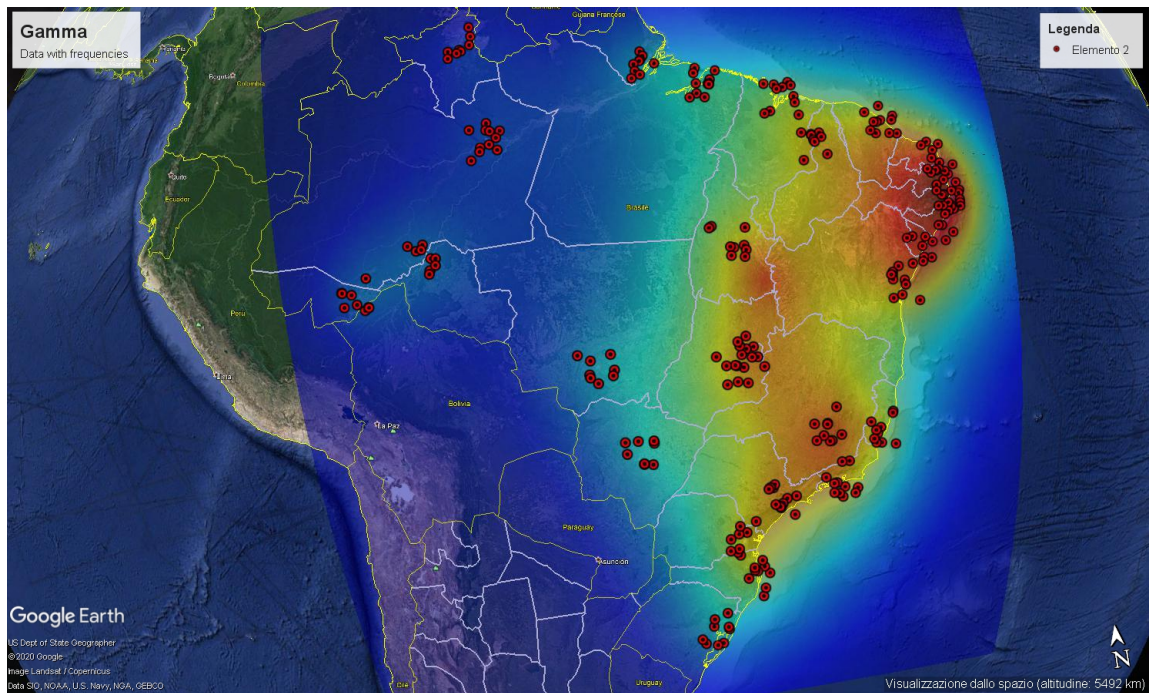


Figure 6. TWC- $\gamma$  map.

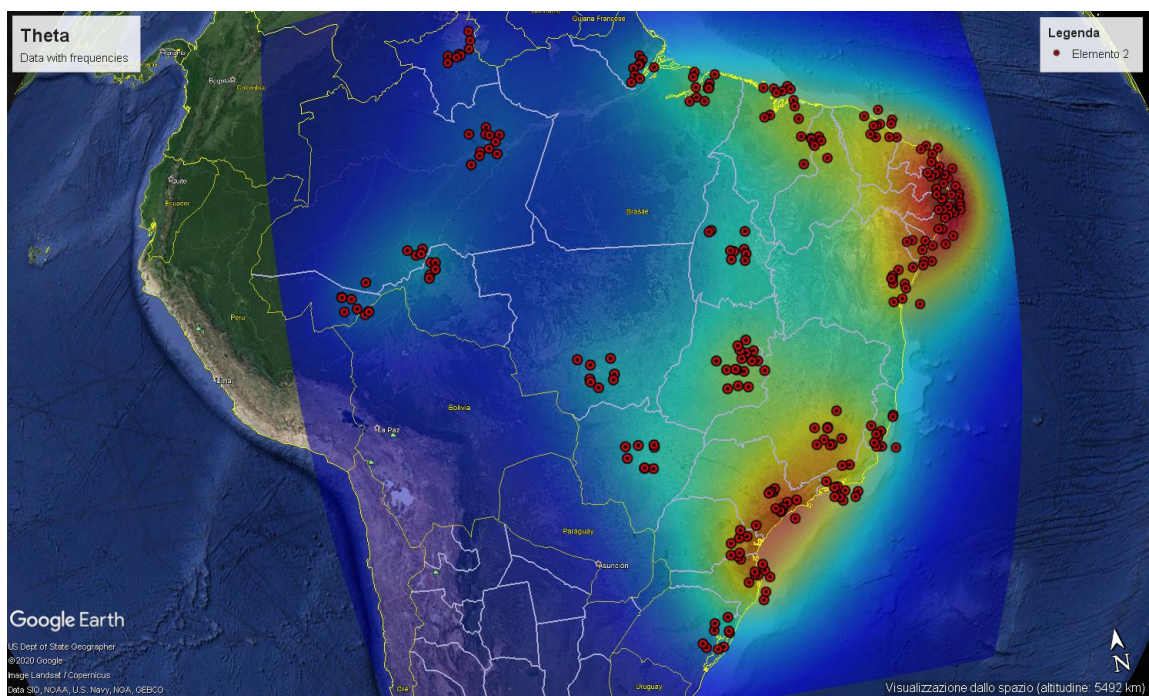


Figure 7. TWC- $\theta$  map.

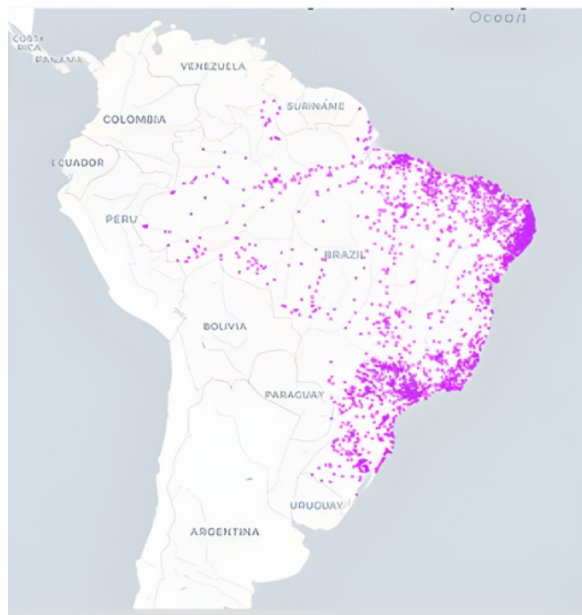
happen during different covid-19 peaks. In TWC- $\beta$  (Figure 4) and TWC- $\theta$  (Figure 7) maps, the hotspot regions are separated from each other. However, in TWC- $\theta$  these regions are expanded. It means the outbreak is expanding. The relative changes, using the data in Table 2 between May24 and May 28, in Minas Gerais are higher than Bahia and São Paulo (north and south of Minas Gerais).

This is in agree with TWC- $\gamma$  map that shows these two hotspots are approaching each other.

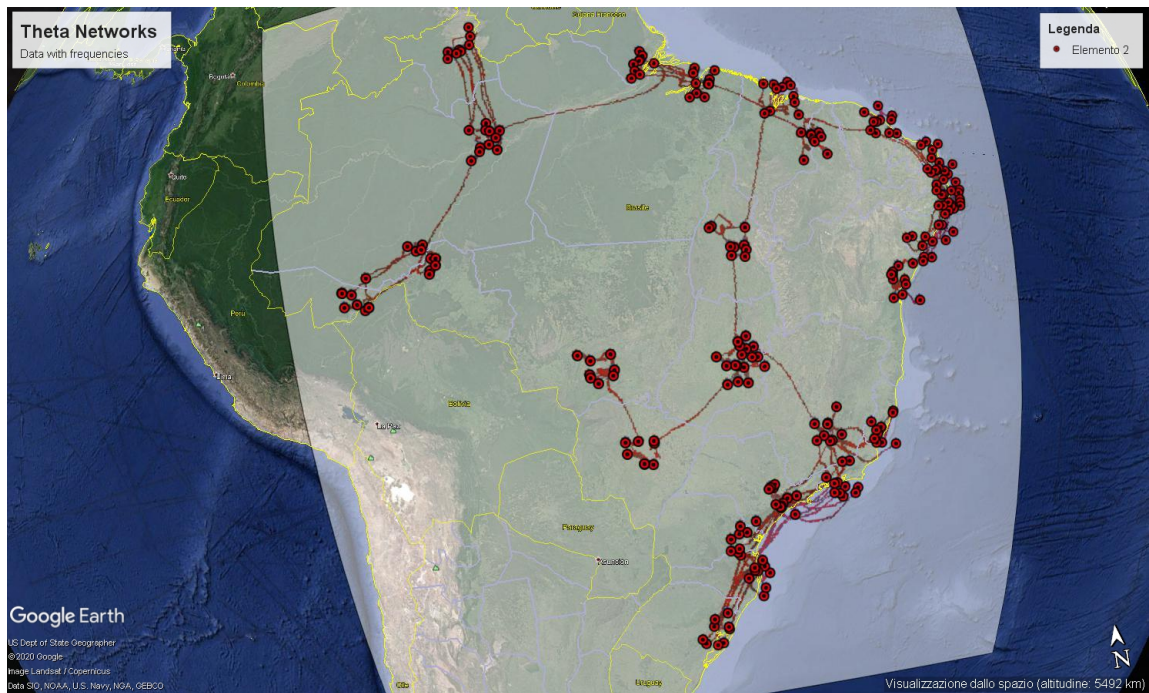
We can compare the TWC maps with other results in (Nicolelis *et al.* [11]) for more empirical evidence. The contribution of the northeast region in the outbreak from the beginning of April until almost May 20 increases, however, the contribution of southeast and south during this period decreases (see Figure 2, Nicolelis *et al.* [11]). This trend is significant particularly around May 20, 2020. This is an evidence for having a hotspot in northeast particularly in TWC- $\beta$ . Besides, the contribution of northeast in the outbreak decreases from the last week of May to the first week of July (see Figure 2, Nicolelis *et al.* [11]). This means the intensity of two hotspots are getting closer and the two regions are approaching to each other as shown in TWC- $\gamma$ .

**Figure 7** shows the TWC- $\theta$  map. The comparison of **Figure 6** and **Figure 7** shows the hot spot breaks into two parts, one hotspot mainly over the northeast region and the second over the central and south regions. **Figure 8** shows the map of COVID-19 on June 6, 2020 (the future outbreak relative to May 24, 2020). The results of TWC- $\gamma$  and TWC- $\theta$  compare well to **Figure 8** as the future map, that is, the result of **Figure 7** matches very well with **Figure 8**.

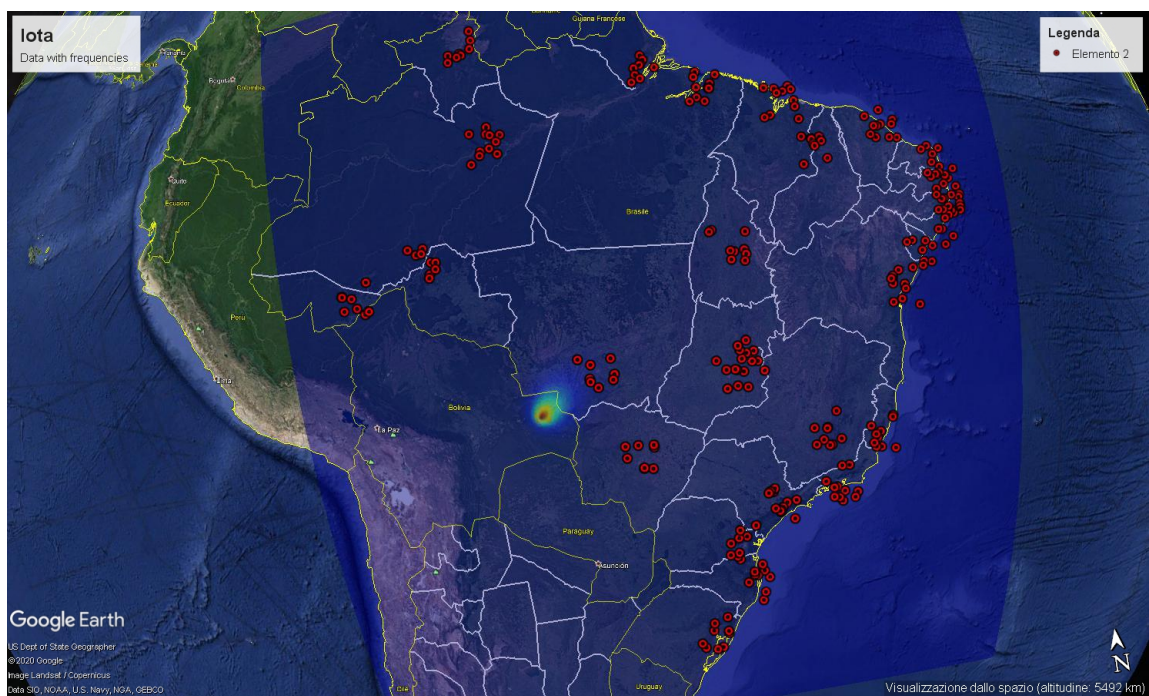
**Figure 9** shows the network of the disease outbreak. This network is given from the algorithm. This figure shows Brasilia is a main node of this network. This is in agreement with our observation (see **Figure 3** and **Figure 5**). As we mentioned we did not use the transportation information, however the TWC- $\theta$  network (**Figure 9**) is comparable with the results obtained by Nicolelis *et al.* (see Figure 1-M, N, O and P in [11]). Moreover, Brasilia is also the second busiest airport in Brazil, which is another piece of evidence for confirming the TWC- $\theta$  network.



**Figure 8.** COVID-19 map of Brazil on June 6, 2020.



**Figure 9.** TWC- $\theta$  network.



**Figure 10.** TWC- $\iota$  map.

**Figure 10** shows TWC- $\iota$  map. It shows how the outbreak point moves away from the east mainly toward the central region. This region can be the final destination of the outbreak, if there are no new sources that come into the play in the meantime. The results TWC- $\alpha$ , TWC- $\beta$ , TWC- $\gamma$ , TWC- $\theta$  and TWC- $\iota$  show the disease dynamics has some propagation toward west. This result is in agree

with relative changes 0.36 and 0.33 of COVID-19 cases between May 24 and May 28 (**Table 2**) in Mato Grosso Sul and Mato Grosso, respectively.

#### 4. Conclusions

Topological Weighted Centroid (TWC) was applied to May 24, 2020 COVID-19 cases in Brazil updated and more detailed data can be used. However, our focus is on illustrating the efficacy of our approach. The algorithms located hotspots, that is, the centers from which on succeeding days new cases of the disease are generated. In other words, if there were new case of the corona virus, the hotspots would be their centers of distribution from which they would most effectively cause of the cases as reported.

The paths, from which these cases center is located, are indicated by the undirected graph that is produced (**Figure 9**, the outbreak network). These analyses are given for the beginning of the outbreak, and we had limited information. Additionally, we focused on the outbreak over the entire country not the local level. Therefore, we used TWC-original. If we want to know more about the detail of the outbreak, we can use TWC-windowing. These analyses use only latitude and longitude information for the capitals of the states with confirmed cases on May 24, 2020. One of the advantages of the TWC method is that it can use a minimum of information for disease outbreak, which is especially useful and beginning of outbreak.

This study did not use the transportation information. However, the results of TWC- $\theta$  are comparable with the results obtained by Nicolelis *et al.* (see the pictures M, N, O and P in [11]). The TWC- $\theta$  network identified how COVID-19 spread through the country. It is reasonable that COVID-19 spread through the country by road and air transportation and the TWC- $\theta$  network **Figure 9** is very similar with the main Brazilian roads and air network.

#### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

#### References

- [1] Edelstein-Keshet, L. (2005) *Mathematical Models in Biology*. In: *Classics in Applied Mathematics*, Society for Industrial and Applied Mathematics, USA. <https://doi.org/10.1137/1.9780898719147>
- [2] Buscema, M., Massini, G. and Sacco, P.L. (2018) The Topological Weighted Centroid (TWC): A Topological Approach to the Time-Space Structure of Epidemic and Pseudo-Epidemic Processes. *Physica A: Statistical Mechanics and Its Applications*, **492**, 582-627. <https://doi.org/10.1016/j.physa.2017.09.050>
- [3] Buscema, M., Asadi-Zeydabadi, M., Massini, G., Lodwick, W., Breda, M., Petritoli, R., Newman, F. and Della Torre, F. (2023) The Topological Weighted Centroid: A New Vision of Geographic Profiling. In: *Theory and Applications*, Springer, Berlin. <https://doi.org/10.1007/978-3-031-28901-9>

- [4] FAPESP (2020) 54.8% Covid-19 Cases Imported to Brazil by March 5 Came from Italy. <http://agencia.fapesp.br/548-of-covid-19-cases-imported-to-brazil-by-march-5-came-from-italy/32854/>
- [5] Ministry of Health (2020) Coronavirus Covid-19. <https://coronavirus.saude.gov.br>
- [6] Secretary of Health in Bahia (2020) Central integrada de comando e controle da saúde. <https://bi.saude.ba.gov.br/transparencia/>
- [7] Jornal o Globo (2020) Amazonas tem primeiro caso confirmado de novo coronavírus. <https://g1.globo.com>
- [8] Agência Brasil (2020) Primeira paciente internada no DF está em estado gravíssimo. <https://agenciabrasil.ebc.com.br/saude/noticia/2020-03/primeira-paciente-internada-no-df-esta-em-estado-gravissimo>
- [9] Secretary of Health, Rio Grande do Sul (2020) Confirmado o primeiro caso de novo coronavírus no Rio Grande do Sul. <https://saude.rs.gov.br/confirmado-o-primeiro-caso-de-novo-coronavirus-no-rio-grande-do-sul>
- [10] Asadi-Zeydabadi, M., Buscema, M., Lodwick, W., Massini, G., Della Torre, F. and Newman, F. (2021) Analysis of COVID-19 Pandemic in USA, Using Topological Weighted Centroid. *Computers in Biology and Medicine*, **136**, 1-17.
- [11] Nicolelis, M.A.L., Raimundo, R.L.G., Peixoto, P.S. and Andreazzi, C.S. (2021) The Impact of Super-Spreader Cities, Highways, and Intensive Care Availability in the Early Stages of the Covid-19 Epidemic in Brazil. *Scientific Reports*, **11**, Article Number: 13001. <https://doi.org/10.1038/s41598-021-92263-3>
- [12] Castro, M., Kim, S., Barveris, L., Ribeiro, A.F., *et al.* (2021) Spatiotemporal Pattern of COVID-19 Spread in Brazil. *Science*, **372**, 821-826. <https://doi.org/10.1126/science.abh1558>