Analysis of Covid-19 Transmission Using Complex Networks

Saschiko Shirai Reyna[1], Oroselfia Sánchez[2,3], Carmen A. García-Cerrud[3] and Idalia Flores-De la Mota[3]

[1] Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico
[2] Department of Chemical, Industrial and Food Engineering, Universidad Iberoamericana, Mexico City 01219, Mexico
[3] Facultad de Ingeniería, Universidad Nacional Autónoma de México, Mexico City 04510, México
oroselfia.sanchez@ibero.mx

Abstract. The global public health crisis caused by the SARS-CoV-2-19 (COVID-19) pandemic has highlighted the need for research into contagion complexity. This challenge necessitates the development and testing of various approaches to manage rapidly changing information with high impact. In this paper, we employ time series analysis and complex networks analysis to compare the evolution, spread, and containment of COVID-19 pandemics in eleven countries and globally. Our analysis enables us to observe the dynamics of spread and the impact of different strategies employed by each country in increasing and decreasing cases through complex network techniques. Additionally, we explore the transformation of data behavior over time as our understanding of the virus improves. Our findings provide important insights into the limitations of using statistical models and suggest that simulation of new cases of COVID-19 data can be modeled using complex networks. The complex network model provides a general description of contagion dynamics in the 11 countries and worldwide situation. This paper contributes by highlighting the limitations of using statistical models to infer and study early time series data and proposing the use of a complex network approach to study contagion dynamics.

Keywords: COVID-19, Visibility algorithm, Time series, Complex networks.

1 Introduction

In December 2019, a worldwide crisis was unleashed by a new coronavirus (SARS-CoV-2-19), first discovered in China, that quickly spread throughout the world, with millions of confirmed cases [1]. A behavior analysis on the spread of new cases using time series analysis enables us to observe patterns of behavior as regards trends, seasonality, and randomness. The different methods used for modeling aspects of the COVID-19 pandemic include mathematical models for studying its spread and complex networks for analyzing outbreaks like Wuhan [4]. Information about the dynamics of virus spread enables us to study the effects of the actions taken by the public health authorities and be able to estimate the numbers of the population that are infected [6]. Thus, we can make reliable predictions about its future evolution in a specific timeframe and consequently be able to prepare the health system to deal with it [9].

Using control approaches gives us quick treatments for a small part of the population but their effects help to fight the ravages of infectious diseases on large populations by administering the smallest possible amount of medicine, thus enabling its greater availability. Statistical methods for estimating representative distributions for certain variables of COVID-19 such as hospitalization, medical appointments when the symptoms appear, and other specific clinical cases have also been used to describe the dynamics of disease transmission [12]. These analyses support the decision-makers reports and the implementation of policies, such as closing borders and international flights, enacting national lockdowns, closing parts of a city or the economy, relaxing said restrictions, and deploying equipment and resources for field epidemiology [13]. The policies applied throughout the world for the mitigation of COVID-19 include [14] restrictions such as mobility, socio-economic, physical distancing, hygiene measures and changes in public communication about the situation. In this paper, we highlight the use of algorithms that, in combination with network models, contribute to modeling the efficiency of drugs that were not initially de-signed for viral treatments but that have been used to treat COVID-19 [18]. This comparison is based on time series characterization, a visualization algorithm to convert all data into networks, which will be analyzed using complex network metrics.

The paper is organized as follows: Section 2 presents the analysis of COVID-19 new cases data in many countries at the same time. Aiming to conserve and describe the infection behavior with population response to sanitary measures. All countries decided to implement different strategies to avoid the increase in COVID-19 new cases, the reason why every country has a particular infection behavior. Finally, the results are presented in section 3, preceding the conclusion and discussion.

2 Case Study Analysis

These countries were selected due to the data with the greatest spread of the epidemic and considering that all continents are considered.

The countries selected in this paper are: United States (US), Mexico (MX), Japan (JP), India (IN), Russia (RU), South Africa (ZA), Italy (IL), Germany (DE), Morocco (MA), Netherlands (NTH), Brazil (BZ) and we included the global information as a frame for the worldwide behavior of the COVID-19 pandemic.

The data on the new Covid-19 cases were obtained from the World Health Organization, defining the confirmed cases listed on its website, in daily reports as an analysis unit [19]. Fig. 1, shows the comparison between the calculated percentage ranking with the population density of each country versus the new cases published by the WHO ranking. Graph A and graph B contain the same new cases data but the first one presents the ranking of countries with more population contagion. Meanwhile, graph B shows the COVID-19 new cases reported ranking. For example, the US keeps the largest number of new cases reported and is the country with the most infected population. Besides, public portals report India as the second country with more cases in the world, but with a large demography population can be placed in the tenth place of the eleven countries analyzed in this work. Different from Italy and Netherlands ranked in lower order but both are countries with the largest population infected. Owing to the number of graphics generated for said analysis, we have used the ones corresponding to Mexico as an example. Graphics for the other countries are given in the Appendix A (Appendixes can be seen on: https://github.com/SashikoSR/COVID-19).

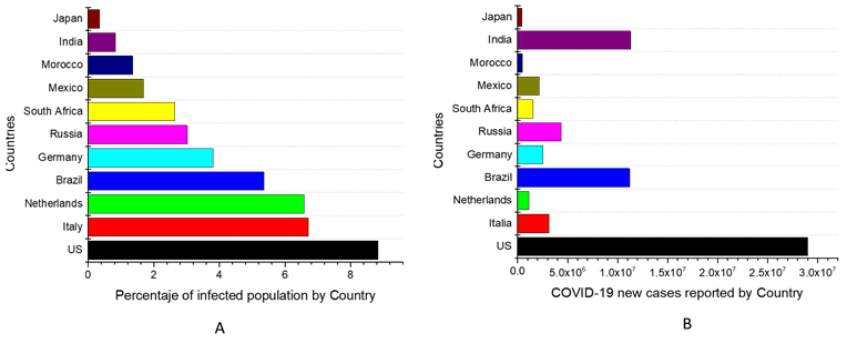

Fig. 1. Comparison between population density ranking and the World ranking of daily new cases published by the WHO.

2.1 Data Path: From the Repository to the Complex Network Model

The complex network model is built using the networks of new COVID-19 infection cases from each of the 11 countries. To determine the countries' analysis feasibility, the new cases data published in public and open repositories were compared. For example, in the portal of each country, the total new cases recorded were reviewed and compared on three additional databases (WHO, Hopkins, and the country's database). For some countries, there was a huge difference between the published records, both in the total number of new cases and in each day's records. These inconsistencies are a natural result of the methodological adjustments made by each country, but this does not conserve the propagation dynamics, it only accumulates the data to reach a total of new cases that is closer to reality. The difference between the databases and the data corresponding to each country is wide and there is a lot of inconsistency between the data, which has an impact on the feasibility of using this database in the complex network model owing to the impossibility of propagating data inconsistencies in a way that represents the dynamics of propagation.

For the model to describe the individual characteristics of the countries, a careful analysis of the published public data was made.

Fig. 2 available in GitHub, shows the data comparison, transformation, and preparation states for the complex network model.

The records of new cases are reported per day in the data sources. The data reports depend on factors such as the working day, the distance traveled to collect the data, to name but a few, and this generates noise in the databases. Grouping the data every week preserves the behavior while at the same time eliminating the noise that the daily period presents.

It is important to mention that the aim is to analyze whether the data present a seasonality or a specific trend. However, this prediction is not yet possible due to the state of the pandemic worldwide. In the future, with more collected and recorded data, trend and seasonality analysis can be applied to determine the stability of the time series. Once the data has been selected, pooled, and analyzed, the visibility algorithm is applied for its analysis. This algorithm connects the weeks that are visible to each other according to the newly reported cases. If there is a data segment that can be plotted in a polygonal shape, then all the nodes will be connected. Finally, the network is built from the newly reported cases for each week of the pandemic.

2.2 Hot Issues of Data in Complex Networks

New cases of COVID-19 reports reflect the different updates of methodologies for collecting the transmission. Over time, all the countries changed the criteria for the new cases records. In the data treatment framework, the natural evolution of the COVID-19 pandemic content the accumulation of the new cases total, but the time series do not show the propagation dynamics. It is common in complex networks to deal with different types of data, and always an opportunity to demonstrate the characteristics of such data with classical tests.

Fig. 45 available in GitHub, shows the classical treatments that demonstrate how the behavior of complex network data differs from others. It would be suitable to model the data under the same conditions for all the periods. Table 1 contains the results of the different tests applied to new case data. Table 2 shows the normality test results of the time series for every country analyzed in the complex networks model.

Table 1. Statistical tests for new cases of COVID-19.

Data Analysis by week						
Countries	KS (Two samples)	Conclusion	Chi square	Conclusion	Box-Ljung	Conclusion
Mexico-Italy	8.43E-03	Without similarities	2.12E-02	Without similarities	2.30E-14	Dependency
Mexico-South Africa	1.54E-04		1.10E-06		2.30E-14	
South Africa-Russia	4.95E-08		3.82E-04		9.66E-15	
Russia - Morocco	2.20E-16		3.82E-04		9.66E-15	
Morocco - India	5.77E-15		3.61E-03		7.77E-16	
India-Brazil	4.14E-02		3.37E-03		7.77E-16	
Brazil-Netherlands	1.33E-15		9.01E-07		6.66E-16	
Netherlands-Germany	4.14E-02		2.12E-02		2.00E-15	
Germany-Italy	7.17E-01	Similarity	1.80E-02		2.78E-15	
Italy-Japan	1.39E-05	Without similarities	8.47E-02	Similarity	2.11E-15	
Japan-US	< 2.2e-16		8.13E-02		1.27E-14	

Kolmogorov-Smirnov demonstrates if the samples are different enough between them to infer future behaviors. The model variation represented by Lilliefors is for review of the normal distribution adjustment. On the other hand, chi-squares relate two variables and show if the samples are associated or independent. And the run test allows testing if there is randomness in the subsequent order of the data. The statistical models were applied for the eleven selected countries according to the flow chart (see Fig. 44 available in GitHub). For each time series, data were prepared to be tested in the metamodel using RStudio packages.

Through this additional data analysis, we conclude that it is not an option to transform data into other distributions despite inconsistencies in all the time series. The data used in the model represents a particular moment of COVID-19 transmission. Therefore, as mentioned before this type of data should not be transformed in other distributions.

Table 1 demonstrates that the data we collected cannot be used directly for statistical purposes or as input without prior characterization. Table 2, on the other hand, illustrates that it is not appropriate to transform the time series into other distributions due to two main reasons. First, complex network models commonly utilize the raw time series data to avoid losing important information during transformation. Additionally, time series analysis allows us to identify important features and patterns in the data that may be lost during the transformation process. Second, we propose characterizing the time series through time series analysis, which involves modeling the underlying trends and patterns of the time series. If this method is unable to infer information, a complex network model may be applied. This is particularly relevant in the case of epidemics, such as COVID-19, where the simulation of new cases can offer valuable insights about the available data.

Table 2. Normality Test.

Normality test KS (Lilliefors)			
Country	Value	Country	Value
Mexico	7.47E-02	Netherlands	6.56E-10
South Africa	1.23E-10	Germany	1.69E-09
Russia	2.13E-02	Italy	8.24E-09
Morocco	6.59E-09	Japan	1.42E-08
India	8.45E-06	US	1.71E-06
Brazil	3.45E-03	Global	1.64E-02

2.3 Time Series Analysis

Time series decomposition is a mathematical procedure that splits the series into its three components: Seasonality Patterns that repeat with a fixed periodicity. Trend underlying trend of the metrics, ascendant or descendant. And Random or noise: the remainder of the original time series after the seasonal and trend series are removed. ACF (Auto-Correlation Function) was used to determine the autocorrelation values for any series with its lagged values. It describes how well the present value of the series relates to its past values. ACF considers all the series components while finding correlations. The use of PACF (Partial Auto-Correlation Function) enables the determination of residual correlation by removing the effects of the anterior lag(s) and the posterior lag(s). An accurate correlation is obtained if there is any hidden data displayed by the following lag. The subsequent lag is taken as an element while modeling, but the elimination is preferable of correlated highlights that create multicollinearity issues [20]. The ACF and PACF plots are more commonly used to obtain the values of p and q to feed into ARIMA model frequently used for forecasts.

In this work we used time series as a based for the visibility algorithm to transform time series to complex network and then analyze the complex network topology. It transforms the time series into a complex network through a visibility graph that provides a complete analysis while showing the properties of the time series in the network. Section 2.5 explains the visibility algorithm and a small instance is also set to define how the algorithm sets the data as a network.

2.4 Time Series Analysis for the Eleven Countries Comparison

Time series metrics were used to analyze reported data for new cases of COVID-19. The data used correspond to daily new cases of COVID-19 in 66 weeks since January 2020 for the eleven countries and the global data reported. The data was analyzed using MicroStrategy, Excel, and R.

Fig. 3 available in GitHub, shows the graphs obtained in this analysis. The stage of the pandemic in each country (first wave and second wave), was used for its division. Graphs A and D correspond to the countries in the first wave, while graphs B and E represent countries in the second wave of infections. The behaviors analyzed were the seasonal variation index and non-seasonal data for each country vs. global behavior. Moreover, graphs C and F show Mexico's behavior in respect of the global behavior for both metrics.

These findings are only comparisons for the eleven studied countries, and the total and partial autocorrelation functions obtained for each country make it impossible to get accurate predictions for an overall view of the pandemic period of newly registered cases. However, when there is a higher number of new daily cases records, the correlational estimates between the values and the delays (Xt and Xt-2) shall be used to compare autoregressive models (Fig. 3, available in GitHub). It is important to observe that the non-seasonal data for Mexico (graph F) shows an acceptable control of the pandemic, as the initial peaks were much lower in respect of the behavior during the first wave (graphs A and D) in the other countries. Nevertheless, the data recorded by the countries have changed over time, which can influence the observed dynamics. As we can note in fig. 3, available in GitHub, no countries present seasonally or a clear tendency in COVID-19 new cases data. In other words, the lack of the two aspects mentioned before in the time series cannot allow (during the period under review) to know the future behavior dynamics. As a tool, time series analysis is useful for understanding behaviors. This analysis can be used with another data period to study if the behavior presents predictive characteristics. Meanwhile, complex networks can complete the methodology for describing the governmental strategies and the effects of society's response to the COVID-19 crisis dynamics.

The analysis of the time series conducted in this study did not reveal any clear seasonality or trend in accordance with the models mentioned before. However, upon examining the data, we found that organizing it on a weekly basis was necessary to address noise, despite the absence of any discernible pattern. Our analysis also revealed that peaks in the data that deviated from the norm corresponded with holidays, weekends, or non-working days, though it was difficult to quantify the overall impact of these events on the time series. Furthermore, the practice of governments adjusting the data by adding the total number of deaths or non-deaths on a specific day made it impossible to maintain or project any seasonality or trend over the entire time series. Instead, this practice only reflects a more accurate representation of the total number of new cases, which, in turn, affected the overall behavior of the time series. These findings can contribute to the preparation of data for visibility algorithms to be used in complex networks. The data analysis was performed using RStudio and Microsoft Excel.

2.5 Visibility Algorithm

The main goal of the visibility algorithm is to map a time series into a complex network. To study and analyze it with all the techniques and properties of network theory, resulting in a more complete analysis and not just the time-series study. Another important thing is that this network inherits several properties of the time series. The algorithm criterion sets up two arbitrary data values (ta, ya) and (tb, yb) that will have visibility, and consequently will become two connected nodes of the associated graph if any other data (tc, yc) placed between them satisfies:

$$y_c < y_b + (y_a - y_b)\frac{tb-tc}{tb-ta} \qquad (1)$$

Fig. 4 available in GitHub, gives an example of a time series of 20 data values plotted as a periodic series by using vertical bars (the data values are displayed above the plot). Considering this as a landscape, we link every bar (every point on the time series could be days, weeks, months, years, etc.) with all those that can be seen from the top of the one being considered (gray lines), thus obtaining the associated network (shown in the lower part of the figure). In this network, every node corresponds in the same order to series data and two nodes are connected when there is visibility between the corresponding data and if there is a straight line connecting the series data, provided this "visibility line" does not intersect any intermediate data height. Using the algorithm, the associated network extracted from a time series is always Connected meaning that each node can "see" at least its nearest neighbors (left and right), the first and the last one at least sees one. And undirected, where there is no direct definition in the links. Also, invariant under affine transformations of the series data, the visibility criterion is invariant under rescaling of both horizontal and vertical axes and horizontal and vertical translations. Fig. 4b available in GitHub, shows the visibility algorithm generated for the Mexico case.

The visibility network of a time series remains invariant under several transformations of the time series as shown in Fig. 3 available in GitHub (a) Original time series with visibility links. (b) Translation of the data. (c) Vertical rescaling. (d) Horizontal rescaling. (e) Addition of a linear trend to the data. As can be seen in the bottom diagram, in all these cases the visibility network remains invariant. The key question is to know whether the associated network inherits some structure of the time series, and consequently whether the process that generated the time series may be characterized by using network theory. In the first step, we will consider periodic series. The example plotted in fig. 4a available in GitHub is nothing but a periodic series with period 4. The associated visibility network is regular, as long as it is constructed by periodic repetition of a pattern. The degree distribution of this network is formed by a finite number of peaks related to the series period, much in the vein of the Fourier power spectrum of a time series. Generically speaking, all periodic time series are mapped into regular networks, the discrete degree distribution being the fingerprint of the time series periods. In the case of periodic time series, its regularity seems, therefore, to be conserved or inherited structurally in the network through the visibility map.

2.6 Complex Network Analysis

The complex network model is made up of 11 networks that represent each of the chosen countries, in addition to the global behavior of new cases. Each network is made up of nodes that represent. every week of the pandemic, containing the accumulated cases for 7 days. In particular, the networks consist of 66 nodes each covering a reporting period for new cases from January 2020 to August 2021. The links are connected through the visibility algorithm, which determines all the weeks (nodes) that are seen from each point, meaning that this algorithm makes it possible to relate to each other those numbers of new cases for each week that have the same maximum number of closest cases, both: backward and forwards.

After the creation of the network, network metrics such as clustering, closeness, betweenness, assortativity, degree distribution, among others, are obtained to perform the complex network analysis. This provides an accurate network behavior approach. Fig. 5 shows the network created for Mexico's case.

When analyzing the metrics of the network associated with each country, the proportion of infected cases and the degree of control of the spread of the disease are compared.

2.7 The Analysis of Complex Networks Metrics for the Eleven Countries

The comparison of the metrics obtained by the complex network analysis allows the characterization of the spread and disease spread control degree to determine how several variables such as the implemented policies, the population density, the degree policies fulfillment by the population among others change the behavior dynamics.

Concepts such as degree (i.e., the number of links per node) or clustering (the number of neighbors that are also connected) are truly local quantities, depending on the state of a single node and its neighbors. Measurements of centrality, such as betweenness, depend on the state of the entire system and will be defined below as global. In between these two scales, we have the study of communities that range from a few nodes to the entire network [21].

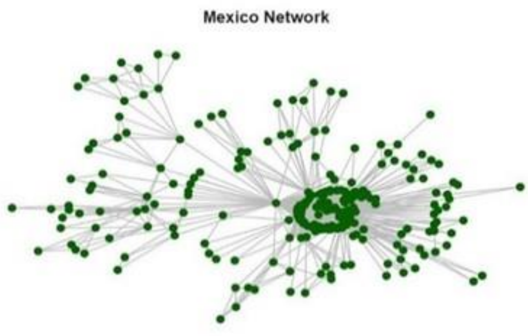

Fig. 2. Visibility algorithm for Mexico.

3 Results

The time series analysis was carried out in Excel and Origin. The visibility algorithm and the complex network were solved in RStudio. Table 3 presents the results obtained by complex networks metrics calculations.

These results mark the contribution of the use of complex networks. This tool helps to achieve a general description of the 11 countries without missing the specific behavior of each country. For instance, the form in which COVID-19 new cases are increasing because of the health measures implemented during the lockdown.

Table 3 shows the metrics obtained by using complex network analysis, and their interpretation is presented below:

Diameter: The diameter D of a network is defined as the longest distance you can find between two nodes in the network. Some other definitions (e.g., average distance) are possible [21]. In this complex network model, the diameter indicates how much time there is between weeks with a higher or lower number of infections, in other words, the cycles of recorded new cases. The results of the complex network analysis (Table 3) show that Japan and Italy presented the smaller diameter, while Brazil, Morocco, Germany, Russia, and Mexico presented the bigger diameter. It is possible to determine the control of the spread of the virus using the strategies applied if population density is considered.

Mean distance: This metric gives the distance in a more general way since it tells us the average of the distances between every pair of nodes in the network. In our case, the mean distance represents the variation of new cases between the weeks before and after each node. If the diameter is small, then the weekly rise in new cases will be similar from one week to the next. For high diameter values, the new cases between one week and the weeks before and after will change without distinction, representing increasing uncertainty. In this case, for all countries, the mean distance is very similar with a very low variation, meaning that all countries have almost the same distance between nodes. According to the mean distance results (Table 3), the average number of weeks needed to observe a change in the behavior of the propagation dynamics that occurs between 2 or 3 weeks.

Cliques: One of the first papers on community structure, published in 1949, defined a community as a group of individuals whose members all know each other. In terms of graph theory, this means that a community is a complete subgraph or a clique. A clique is a connected subgraph with maximal link density. Two k-cliques are considered adjacent if they share k – 1 nodes [22]. In our problem, cliques indicate the time between the changes (adjacency), in other words, the behavior of the virus spread and the pandemic control period. The clique value translates into two behaviors: the number of weeks when the virus is spreading or the number of weeks when it is under control. A high clique value could mean more stability from one node to another (valley or peaks), while a smaller value represents a virus spread variability. The dynamics of propagation in the US are characterized by being very unstable and consistent with the value obtained from the complex network analysis (10 cliques). India, South Africa, and Italy have the highest number of cliques in the entire network (17 cliques). These values show minor variation in new cases from the previous week to the next one. The differences between the clique values for each country show that a higher clique value represents the countries that have contained the pandemic for extended periods, or when the increase in new cases takes more time. India and South Africa have the same clique value. However, for South Africa, it represents the number of weeks that the pandemic has been under control (valley), and for India, it indicates the number of weeks the new cases have increased (peak).

Density: The density of a network is defined as the ratio between <k>, the average value of the degree (the degree of a node), and N, which is the order (number of links) of the network: $\rho = \langle k \rangle / N\text{-}1$, with $\rho \leq 1$ [21]. For this complex network model, this metric establishes whether each country's dynamics are scattered or dense, in other words, the connectivity degree between one week and the following ones or the impact of one on the other. India, South Africa, and Italy (0.32, 0.30, 0.29 respectively) present the highest density values, while Brazil, Mexico, and the US have the lowest values (see Table 3). This metric indicates new cases recorded over the weeks (66). For example, South Africa shows that new cases are mostly together in the peaks of the propagation dynamic. In contrast to the situations in Brazil, Mexico, and the US, where all new cases are recorded over 60 or 50 weeks which causes a propagation dynamic more scatter.

Assortativity: We can identify the presence of a community from nodes with equivalent properties. The level of 'similarity' of two nodes is usually computed using a mathematical quantity called correlation. Usually, since the most immediate property of a node is the degree, we look for the presence of a correlation between nodes with a similar degree. There is no reason, in principle, to expect a particular correlation. Actually, in some situations, there is a tendency for high-degree nodes to be connected to other high-degree nodes. In this case, the network displays what is called assortative mixing or assortativity [21]. In our networks assortativity measures two connected nodes that have a similar increase (or decrease) in the number of new cases, being positive when these changes vary slightly from week to week, and negative when this behavior is not alike. Japan is the only country with negative assortativity (-0.03; see Table 3), which describes the occurrence of three gradual peaks and some valleys. As a metric, this behavior implies that there are several nodes of different degrees, meaning that it has many weeks in valleys or peaks. India has the highest assortativity value because these propagation dynamics (data for 66 weeks) show a noticeable increase and a long duration of a high number of cases, which implies that the degrees of the nodes are remarkably similar to each other throughout the entire time series. Another relevant behavior is the one presented by South Africa (0.149, see Table 3), where the new cases present two marked increments, which determines a valley that remains for several weeks. Despite that, it is an assortative network like India, the results of the model show the difference between the countries.

Clustering: The clustering coefficient Ci of node i is a measurement of the number of links 'around' vertex i. Ci is given by the average fraction of pairs of neighbors (of the same node) that are also neighbors of each other. In general, we can write the clustering coefficient as the fraction of actual links over the possible ones between vertices i, j, k. [21]. Global Clustering refers to the probability of forming triangles between the nodes. It measures the mean proportion of neighbors a node has and presents two variations: global clustering refers to the entire network and local clustering that is calculated for each of the nodes, providing the mean of all the nodes. In this model, global clustering shows the influence of the government measures adopted by each country to contain the COVID-19 contagion rate. Global clustering for the eleven countries is between 0.58 and 0.81 (Table 3), meaning that many groups have formed during the 66 weeks. On the other hand, the local clustering represents the increase of recorded new cases. For example, South Africa and Japan have high increments (Table 3). The eleven countries display similar values, but their behavior differs. If local clustering is higher, then there is also a higher probability of transitivity per node.

Betweenness centrality: Sensitive measurement of centrality is given by the number of times we cross one node k in going from one node I to another j following the path of minimal length (distance d (I, j)). This number is called site-betweenness b(i) [21]. Another definition is a link or link betweenness, defining xij as the number of shorter paths that go through the link (I, j). Links connecting different communities are expected to have large xij while links within a community have small xij [22]. Betweenness centrality indicates the proportion of increased infections in respect of the previous and subsequent weeks. As a complex metric, this can be interpreted as the propagation period, including the rise in new cases, peak width, and decrease, representing how many weeks is COVID-19 active in the whole period. According to Table 3, there is not a wide difference in the countries' values. The betweenness centrality values for Germany and the US are the same, corresponding to 0.33 both. This metric analysis and interpretation conclude that the COVID-19 pandemic is still propagating.

Closeness: Centrality closeness is aimed at measuring how close a node is to other nodes in the network. This is done in terms of communication distance, as measured by the number of links between two nodes if connected by the shortest path, as the closeness metric is the mean shortest path to all other nodes [23]. Closeness represents the time when the new cases remain low or high for each country. In complex networks, a high closeness considers the existence of a few nodes with an extremely high degree. These nodes are called hubs. Table 3 shows that Russia (0.28) and Brazil (0.33) present the lowest closeness value, meaning these countries have the highest variation between subsequent weeks. Japan (0.61) and the Netherlands (0.54) present several similar nodes, which means that the number of new cases is similar for a longer period, or it is remaining in the valley.

Degree centrality: The degree centrality of a node refers to the number of links attached to the node. The standardized score is found by dividing each score by n-1 (n = the number of nodes). As for the centrality degree, the pattern of behavior regarding closeness is accentuated for the following countries: Japan (0.52), Germany (.42), Netherlands (0.41), and Italy (0.35). Brazil (0.18), Mexico (0.19), India (0.19), and Russia (0.19) have similar values and the lowest values of the eleven countries. However, the dynamics between them are different, for Brazil and Mexico, this value means that the centrality degree is low because of the variability in the number of infections that arise on consecutive weeks. Whereas, in the case of Russia and India, the similarity between the nodes represents the accumulation or drop in newly recorded cases. The highest centrality degree corresponds to Japan (0.52), which represents an accelerated rise in the number of COVID-19 infections in a few weeks, implying closed borders.

Degree distribution: Degree distributions can be calculated by the ratio between the number of nodes in network G of degree k and N denotes the size of G (number of nodes). The equation $p(k) = |dk(n)| / N$ is exactly the proportion of nodes in G having degree k. Degree ki of node I is the number of links connected with node i. From this, it follows that this equation also has the meaning that a randomly chosen node in the network has, with probability p(k), k links.

It is an interesting and important fact that many real-world networks like the World Wide Web (www), the Internet, social networks, citation networks, or food webs are not Poisson distributed but follow a power law: p(k) = k-γ, γ>1 [23]. According to the complex network analysis, the eleven countries are free-scale networks. Free-scale networks consider that few nodes have a lot of connections while several nodes have few connections. It means that there are many weeks with records of COVID-19 new cases and that nodes with high values are connected to the rest of the nodes, which in complex network metrics are known as hubs.

Table 3. Complex networks results.

Global Ranking		1	2	3	6	8	9	15	20	21	36	41
Country	Global	US	India	Brazil	Russia	Italy	Germany	Mexico	South Africa	Netherlands	Japan	Morocco
Max degree	33	28	34	24	28	42	43	25	36	43	48	37
Min degree	2	1	1	1	1	1	1	1	5	1	4	2
Mean degree	9.7	12.7	21.2	11.9	15.2	19.1	15.6	12.2	20.1	16.1	13.9	16.4
Diameter	5	5	5	6	6	4	6	6	5	5	4	6
Mean distance	2.57	2.30	2.34	2.93	3.09	2.29	2.37	2.68	2.20	2.22	2.08	2.23
Cliques	11	10	17	13	14	17	16	14	17	15	13	12
Density	0.147	0.195	0.326	0.183	0.233	0.294	0.240	0.188	0.309	0.248	0.214	0.252
Assortativity	0.084	0.087	0.712	0.675	0.610	0.579	0.217	0.609	0.149	0.176	-0.04	0.05
Global Clustering	0.589	0.563	0.816	0.755	0.707	0.777	0.713	0.762	0.807	0.671	0.659	0.607
Mean Local Clustering	0.792	0.688	0.799	0.769	0.748	0.795	0.821	0.790	0.845	0.763	0.822	0.738
Closeness Centrality	0.476	0.367	0.375	0.331	0.286	0.432	0.457	0.374	0.412	0.545	0.611	0.351
Degree Centrality	0.353	0.236	0.197	0.186	0.198	0.352	0.422	0.197	0.245	0.414	0.524	0.316
Betweenness Centrality	0.514	0.331	0.513	0.490	0.376	0.410	0.332	0.395	0.485	0.496	0.476	0.418
Eigencentrality	0.702	0.645	0.559	0.685	0.639	0.554	0.623	0.676	0.564	0.625	0.675	0.616

Fig. 6F available in GitHub, shows the distribution degree for Mexico, and the global distribution degree (Fig. 6A available in GitHub), which presents a p-law distribution, even the tail is longer. Graphs E and F show the evolution of the pandemic over time. Comparing plot E which contains new cases reported until September 2020, and plot F that corresponds to January 2020 until March 2021, the degree distribution shows the differences between nodes' degrees.

Other information obtained from the degree distribution metric is that countries with more infected populations behave similarly to graphs B, D, or F, such as the US, that reported 8.84% of its population infected. Countries with a lower number of infected populations follow a pattern like graphs A, C, or E, such as Japan, which reported 0.35 % of its population infected, both these findings are based on data up to March 2021. All of this means that with the new cases recorded and the increased number of weeks elapsed, the distribution degree tends to function as a homogenous network.

The advantage of using a complex network technique is to describe the entire dynamic behavior of the selected countries. The methodology proposed in this research does not use only time series analysis, due to the dynamic behavior of the data that is not reflected if this method is used alone. On the other hand, statistical methods only describe proportions and parts of the total behavior, while modeling a system by complex networks shows the variations and dynamic nature of data.

When the complex networks metrics are understood in a social behavior context, the inference about the relation of data behavior with sanitary measures can be done.

4 Conclusions

Time series analysis was used to organize each country's behavior, which made it possible to determine that it is too early for there to be enough data (owing to the time the pandemic has lasted) to be able to establish a seasonality or trend of the dynamics of virus spread. The techniques used in this methodology can conserve particular and individual dynamics for each country, thus providing a specific contribution using complex network modeling in pandemic data while performing a whole analysis of the behavior of the eleven countries and overall behavior of covid across the world. We can discuss here the correlation between the metrics and the countries listed, as well as the exceptional cases where it can be assumed that this is because of how the policies for the contention of the pandemic were implemented in each country.

A ratio has already been mentioned between the network's diameter and population density and to illustrate this we are giving the following tables that present the countries with smaller and larger diameters and their respective population densities. In the case of these countries, it can be expected that for a high population density, there is a shorter time between weeks for peaks or valleys and the dynamics of the pandemic are faster. Cliques refer to a higher number of weeks of contention or new cases, while density refers to newly recorded cases, which is where the two metrics coincide. On the other hand, the density metric is low for Brazil, Mexico, and the US, which means that the spread is low over several weeks, which coincides with their low number of cliques. However, there are countries, such as Morocco, that have a low number of cliques, but higher density, so we would have to examine, among other factors, the country 's contention policies once again. For the results to be better understood, we should clarify that the density of the network is equivalent to the mean degree divided by the number of nodes minus 1. So, for example, in the case of Mexico has a density of 0.1878, this is equivalent to the mean degree, which is 12.212 between the 66 nodes minus 1, in other words, 65 nodes, resulting in 0.1878. Global Clustering and Betweenness centrality: In these two metrics, we can observe that South Africa and India have a high probability of the nodes (weeks) forming triangles (global clustering) while having at the same time a high (table 4) Betweenness metric as well as more cliques and higher density. This is important because, at the same time as triangles are formed in India and South Africa, they have a higher number of shorter paths that involve a node; in this case the weeks that have reported more cases.

Table 4. Comparison between clustering and betweenness.

Country	Clustering	Betweenness
India	0.815658	0.512773
South Africa	0.806672	0.484567

In a unique way from the papers found during the literary review, our paper seeks to describe, through complex network metrics, how government policies and the social response to them, affect the propagation dynamics in each country. But with the use of techniques, the components can be analyzed and described noticing that, from an appropriate association through modeling, we get results that contribute to an understanding of the evolution of the pandemic as a function of each country's singularities such as population density, economic situation, type of government, or geography. The contribution of this work is given by using complex networks with their metrics and how each country applied the restrictions and the influence on the global behavior is confirmed. Furthermore, we can analyze cases over the world, analyze the impact of the vaccination and other analysis like mortality or analyze the different stages or waves from the pandemic.

The main contributions of this work are that we analyzed the different time series for COVID-19, we applied the visibility algorithm to transform the time series into complex networks, analyze and compare the complex network metrics. As part of the future work, we can extend this work for the new cases, analyze the different stages of the pandemic, analyze the influence of the vaccination in different countries and apply the methodology to other countries and make a comparison among them. Simulation of complex networks offers several advantages over traditional statistical approaches, as demonstrated in our study. By using simulation, we were able to analyze the time series data from eleven different countries simultaneously, while preserving each country's unique characteristics. This approach would not have been possible using classical statistical methods that rely on analyzing new cases of COVID-19 individually. The networks we constructed for each country and their global behavior enabled us not only to examine the data as a network and observe its behavior but also to compare the behavior between countries and the global data. This comparison allowed us to gain insights into the similarities and differences in the spread of COVID-19 among the different countries and how it impacted the global community.

References

1. WHO. World Health Organization. https://covid19.who.int/table, (2020)
2. Snacken, R.: Pandemic planning. Vaccine, 88–90 (2002).
3. Waterer, G. W., Hui, D. S. & Jenkin, C.: Public Health Management of Pandemic (H1N1) 2009 infection in Australia: ¡A Failure!. Official Journal of the Asian Pacific Society of Respirology, 51–56 (2010).
4. Liu, C., Wu, X., Niu, R. & Wu, X.: A new SAIR model on complex networks for analyzing the 2019 novel coronavirus (COVID-19). Nonlinear Dynamics, 1–11 (2020).
5. Emanuele, J. &. K. L.: Workflow Opportunities and Challenges in Healthcare. BPM & Workflow Handbook, 2–11 (2007).
6. Liu, D. Cliulemente. L., Poirier, C., Ding, X., Chinazzi, M., Davis, J., Vespignani, A., Santillana, M.: A machine learning methodology for real-time forecasting of the 2019-2020 COVID-19

outbreak using Internet searches, news alerts, and estimates from mechanistic models. e-prints posted on arXiv, (2020).

7. Zhang, J. Litvinova, M., Liang, Y., Wang Y., Wang, W., Zhao, S., Wu, Q., Merler, S., Viboud, C., Vespignani, A., Ajelli, M., Yu, H.: Changes in contact patterns shape of dynamics of the COVID-19 outbreak in China. AAAS Public Health Emergency Collection, Science. 368 (6498), 1481–1486 (2020). DOI: 10.1126/science.abb8001
8. Davis, J.T., Chinazzi, M., Perra, N., Mu, K., Pastore, A., Ajelli, M., Dean, N., Gioannini, C., Litvinova, M., Merler, S., Rossi, L., Sun, K., Xiong, X., Longini Jr. I., Halloran M.E., Viboud, C., Vespignani, A.: Cryptic Transmission of SARS-COV-2 and the first COVID-19 wave. Nature 600, 127–132 (2021). https://doi.org/10.1038/s41586-021-04130-w
9. Díaz-Pinzón, J. E.: Precisión del pronóstico de la propagación del COVID-19 en Colombia. Revista Repertorio De Medicina Y Cirugía. (2020).
10. Scabini, L. Ribas, L., Neiva. M., Junior, A., Farfán, A., Bruno, O.: Social Interaction Layers in Complex Networks for the Dynamical Epidemic Modeling of COVID-19 in Brazil. Physica A: Statistical Mechanics and its Applications. 564 (2020). DOI: 10.1016/j.physa.2020.125498
11. Alanis, A. Y., Hernández-Vargas, E. A., Ramírez, N. F. & Ríos-Rivera, D.: Neural Control for Epidemic Model of Covid-19 with a Complex Network Approach. IEEE Latin America Transactions, 19(6) 866-873 (2020).
12. Zhang, J. Litvinova, M., Wang, W., Wang, Y., Deng, X., Chen X.: Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modeling study. The Lancet Infection Diseases. 20(7), 793–802 (2020).
13. Villamizar, E. González-Casabianca, F., Herrera, S., Rodríguez-Barraquera, T., Angel, A., Corredor, V., Feged-Rivadeneira, A.: Políticas públicas, grandes datos, teoría de redes y COVID-19. Desafíos. 1–19 (2020).
14. Bruinen de Bruin, Y. Lequarre, A., McCourt, J.,Clevestig, P., Pigazzani, F., Jeddi, M., Colosio, C., Goulart, M.: Initial impacts of global risk mitigation measures taken during the combatting of the COVID-19 pandemic. Safety Science, 128 (2020).
15. Stella, L., Pinel, A., Bausso, D. & Colaneri, P. The role of Asymptomatic Individuals in the COVID-19 Pandemic via Complex Networks. SSRN Electronic Journal. (2020).
16. Montes-Orozco, E. E., Mora-Gutierrez R., De-Los Cobos-Silva, S., Rincón-García, E., Torres-Cockrell G., Juárez Gómez, J., Obregón-Quintana, B., Lara-Velázquez, P., Gutierrez-Andrade M. Identification of COVID-9 Spreaders using Multiplex Networks Approach. (2020).
17. Vespignani, A. Tian, H., Dye, C., Lloyd-Smith, J., Eggo, R., Shrestha, M., Scarpino, S., Gutierrez, B., Kraemer, M., Wu, J., Leung, K., Leung, G.: Modeling COVID-19. Nature Review Physics, 2, 279–281 (2020).
18. Morselli Gysi, D., Do Valle, I., Zitnik, M., Ameli, A., Gan, X., Varol, O., Ghiassian, S.D., Patten, J.J., Davey, R.A., Loscalzo, J., Barabási, A. Network medicine framework for identifying drug-repurposing opportunities for COVID-19. Proceedings of the National Academy of Sciences of the United States of America, 118(19) (2021).
19. WHO. World Health Organization. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports, (2020).
20. Salvi, J.: Towards Data Science - Significance of ACF and PACF Plots In Time Series Analysis. https://towardsdatascience.com/significance-of-acf-and-pacf-plots-in-time-series-analysis-2fa11a5d10a8, (2020) last accessed: 2020.
21. Caldarelli, G.: Scale-Free Networks. Complex webs in nature and technology. s.l.: Oxford University Press. (2013).
22. Barabási, A.: Network Science. s.l.: Cambridge University Press. (2016).
23. Dehmer, M.: Structural Analysis of Complex Networks. s.l.:Birkhäuser-Springer. (2011).