Research Article

Directed correlation networks, determined by the dynamics of COVID -19 distribution in various countries

Dmytro Lande [a], *Leonard Strashnoy* [b] (✉)

[a] *Institute for Information Recording of NAS of Ukraine*
[b] *UCLA, Institute Infectious disease department, USA*

ABSTRACT:

The approach of constructing correlation networks can be applied to countries, each of which is characterized by its own process of spreading the pandemic. Previously, it was shown that non-directional correlation networks of parameters allow defining clusters of objects. Non-directionality, on the one hand, reduces the quality of clustering, and on the other hand, it does not allow us to get closer to the problem of finding causal relationships.

A model of correlation networks is therefore proposed by the authors, which takes into account the absolute values of the compared measurement series and the mutual offsets of these series. As a result of the implementation of the model, in part, directional correlation networks are formed, determined by the dynamics of COVID-19 distribution in various countries. The paper shows that node sizes, link weights, and the clustering of such networks leads to easily interpreted results. The proposed methodology can be used both to study the spread of the pandemic in various countries and to study other social, political and economic processes.

KEYWORDS:

COVID-19, pandemic, correlation networks, directed networks, datasets, new incidents of infection.

✉ Corresponding Author E-mail: lstrashnoy@gmail.com

Introduction

The specific behavior of processes associated with the COVID-19 pandemic in various countries can become the basis of policies to counter the pandemic around the world, allowing for the unification of groups of countries in clusters with similar dynamics of infection, recovery or disability. Correlation analysis allows for the comparison of such processes, and it also allows the formation of correlation networks using the methods described in [1, 2].

Numerical series related to the dynamics of the spread of pandemics in different countries can act as parameters for building a network of links between these countries. However, generally accepted correlation networks are not directed. Along with certain advantages (logic, simplicity), they also have disadvantages that reduce the quality of results. In this paper, an example shows a case when the logic of interpretation is violated when analyzing the simplest network of three nodes.

The authors, therefore, propose a model of correlation networks, which takes into account the absolute values of the compared measurement series and the mutual offsets of these series. As a result of the model implementation, directional correlation networks are formed, determined by the dynamics of COVID-19 distribution in various countries. The paper shows that node sizes and link weights, and the clustering of such networks leads to easily interpreted results.

Generation of correlation matrices

When studying different time series associated with several events, great importance is paid to mutual correlations. Mutual-correlation defines similarity but does not always reflect causal relationships. However, when such a relationship exists, a high correlation will indicate that.

In works [1,2], a method for constructing a directed network of correlations of entities (countries) is proposed, consisting of the following stages, based on the method [2]:

1) For each object, numeric series are selected from the dataset for a specific period.
2) The set of maximal cross-correlations between the obtained vectors is calculated, and the corresponding correlation matrix is formed.
3) An adjacency matrix is formed in accordance with the formula for determining the maximum correlation based on mutual offsets, and this matrix is saved as a CSV file.
4) This matrix is uploaded in CSV format to the network structure analysis system Gephi (https://gephi.org/) [3].

5) Object group modularity classes are defined, and the loaded network structure is then clusterized and visualized.
6) At the last stage, an expert interpretation of the results takes place.

Advanced algorithm

It is proposed to take into account two points when constructing a correlation network, namely:

1) which process started first;
2) absolute values of time series for mutual correlation, i.e. to determine the value of the directed connection between nodes *A* and *B* in proportion to the sum of the values of the numerical series corresponding to node *A*.

Let s_k vector of parameter values be matched to each element from a set $S = \{s_k\}_{k=1}^{|S|}$ of objects $\overline{w^k} = \left(w_1^k, w_2^k, ..., w_n^k\right)$, where n is the number of elements in this set.

To implement the first point, the formula determining the relationship between objects i and j [2], is used:

$$a_{ij} = \max_{0<m\le K} \frac{\sum_{k=1}^{n-m} w_{k+m}^i w_k^j}{\sqrt{\sum_{k=m+1}^{n} \left(w_k^i\right)^2} \sqrt{\sum_{k=1}^{n-m} \left(w_k^j\right)^2}}, \quad (1)$$

where *K* is the window of possible time offsets.

The max function is used for the following reasons: processes that are similar in nature may have similar dynamic behavior, but it is possible with a time shift. In contrast to the method described in [2], calculation of m is performed not according to the spectrum of values $[-K, K]$, but in the interval $[1, K]$.

To implement the second point, each of the matrix elements a_{ij} is multiplied by the value of the sum of the values of the corresponding vector $v_i = C\sum_{k=1}^{n} w_k^j$, where *C* is the normalizing constant.

When further using the Gephi visualization tools, the network was defined as directional, the node sizes corresponded to the node degrees of the weighted directed network, clusterization, if necessary, is calculated using the

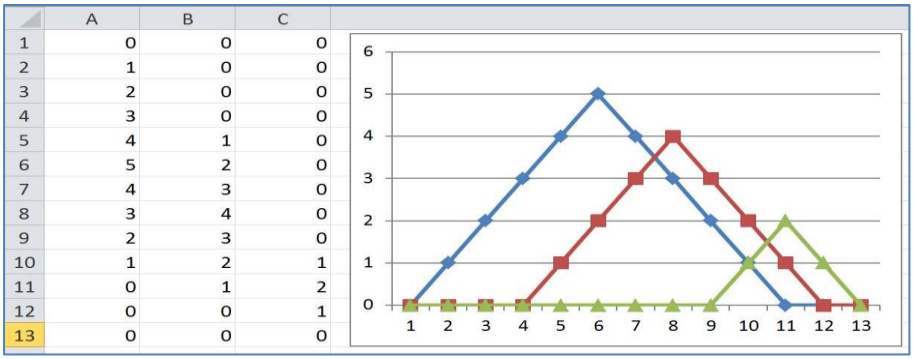
OpenOrd or Fruchterman Reingold algorithms, and node modularity is calculated with Resolution = .5 .

Examples

As a demonstration example, let's consider three entities (Node1, Node2, Node3), each of which corresponds to a time series:

Node1: (0, 1, 2, 3, 4, 5, 4, 3, 2, 1, 0, 0, 0)
Node2: (0, 0, 0, 0, 1, 2, 3, 4, 3, 2, 1, 0, 0)
Node3: (0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 1, 0)

The processes corresponding to these three nodes are shown in Figure 1.



	A	B	C
1	0	0	0
2	1	0	0
3	2	0	0
4	3	0	0
5	4	1	0
6	5	2	0
7	4	3	0
8	3	4	0
9	2	3	0
10	1	2	1
11	0	1	2
12	0	0	1
13	0	0	0

Figure 1: Processes that match the test example.

Visualization of a table corresponding to the correlation matrix calculated using the algorithm from [2]:

;Node1;Node2;Node3
Node1;0.000;0.818;0.623
Node2;0.818;0.000;0.766
Node3;0.623;0.766;0.000

displays the view shown in Figure 2.

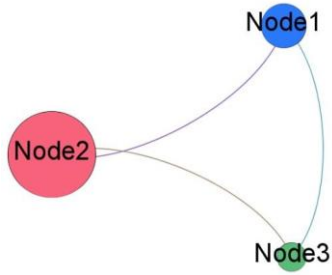


Figure 2: The correlation network of the example, calculated by the algorithm from [2].

O In this matrix, node 2 is represented by the largest circle, although it is obvious that the process corresponding to node 1 started earlier and has a larger amplitude.

To correct this discrepancy, the presented improved algorithm allows us to obtain the following matrix of node relationships, the visualization of which is shown in Figure 3:

;Node1;Node2;Node3
Node1;0.000;1.022;0.779
Node2;0.611;0.000;0.613
Node3;0.002;0.050;0.000

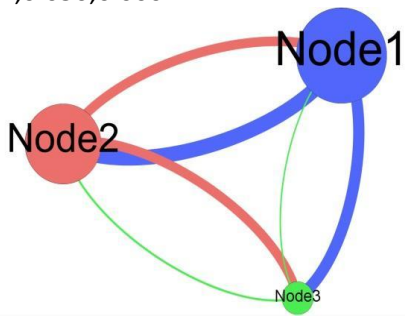


Figure 3: Directed weighted correlation network of the example, calculated using an improved algorithm. Here the direction of the arcs (links) is defined by the color of a node.

In order to build a network of countries, related to the dissemination of COVID-19, data collected by the World Health Organization were used as sources of information https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ and in Our World In Data (the project of the Global Change Data Lab). In this source [4] data on the pandemic (https://ourworldindata.org/coronavirus-source-data) presented in a daily

updated "cleared" form, without abnormal jumps associated with technical failures, with a high degree of integration, in formats .xlsx, .csv, .json. In part, the following datasets from this source were used in this work:

New confirmed cases:
https://covid.ourworldindata.org/data/ecdc/new_cases.csv

New deaths:
https://covid.ourworldindata.org/data/ecdc/new_deaths.csv

This source was used to determine the vectors of dynamics of the pandemic process in various countries for a certain period (selected from 15.03.2020 to 31.07.2020), corresponding to infectivity.

As a result of the analysis of 50 countries, the corresponding correlation matrix was obtained, a network was formed and its clustering was carried out (Fig. 4).

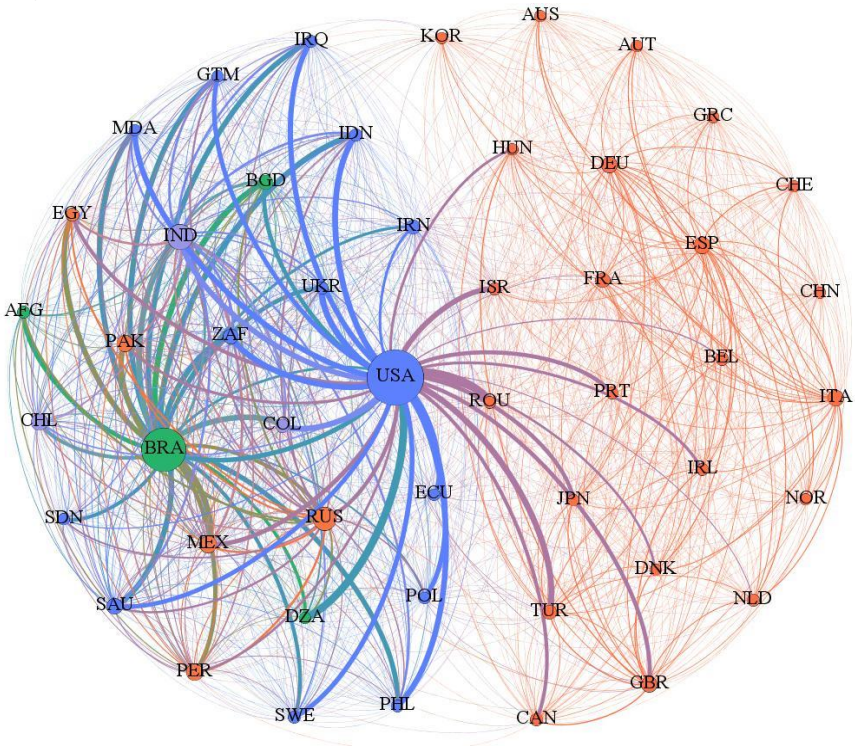


Figure 4: Directional weighted correlation network of countries (country codes according to ISO [5]). Here again, the direction of the arcs (links) is defined by the color of a node.

Conclusion and discussion

The paper describes the concept of a directed weighted correlation network, provides a methodology for its formation, clustering, and visualization of nodes that correspond to the vectors of the pandemic dynamics.

This approach, unlike currently existing ones, has the following advantages:

- intuitive, close-to-reality rules for determining the weight of nodes and links;
- a reliable mathematical basis for correlation analysis;
- objectivity – the dataset is responsible for the "purity" of data;
- ease of implementation, use of standard software tools (Gephi, Matlab, etc.)

Examples of subjects that you can apply the presented method:

1) political leaders, parties characterized by their attitude to various spheres of public life;
2) consumers of products – the parameters here are sellers and the sources of products [1];
3) entities and concepts reflected in social media, in this case, parameters can be time-series of published volumes for certain time periods.

References

1 Using Data Science to Transform Information into Insight Data Smart. John W. Foreman. Wiley, 2013.

2 Leonard Strashnoy & Dmitry Lande. Networks of Countries Defined by the Dynamics of the COVID-19 Pandemic (July 8, 2020). Available at SSRN: https://ssrn.com/abstract=3647570.

3 Ken Cherven. Mastering Gephi Network Visualization. – Packt Publishing, 2015.

4 Our World In Data Project. https://ourworldindata.org/coronavirus-source-data

5 ISO 3166-1 alpha-3. https://www.iso.org/iso-3166-country-codes.html.