18th International Conference on Environmental Science and Technology

Athens, Greece, 30 August to 2 September 2023

Analysis of the Sars-COVID 19 trend: from time series to visibility graphs

SIMONE A.[1*], CESARO A.[2], ESPOSITO G.[2]

[1]Department of Engineering and Geology, University "G. D'Annunzio" of Chieti Pescara, Pescara, Italy
[2]Department of Civil, Building and Environmental Engineering (DICEA), University Federico II, Naples, Italy

*corresponding author:
e-mail: antonietta.simone@unich.it

Abstract Sars COVID-19 epidemic continues to represent a relevant and current topic, which is of concern mainly with respect to possible variants. Predictive monitoring can address the need to reduce the risk of spreading the virus, but it needs to rely on non-invasive, as well as effective and inexpensive strategies. Wastewater-Based Epidemiology (WBE) fits into this context, representing an approach to surveillance of diseases and early warning for any outbreaks of pathogenic viruses, which provides results relating to the trend of the epidemic in the form of time series. An innovative approach that allows to infer information on the spread of Sars COVID-19 is to transform the data of these time series into visibility graphs using the so-called visibility algorithms. The connective structure of the visibility graph inherits many properties of the starting time series and allows to extract nontrivial information on the behavior of the system using topological metrics of the Complex Network Theory (CNT). In this work, the time series of Sars COVID-19 corresponding to a 12-month period for a treatment plant serving a large size basin is analyzed in order to provide useful data on the spread of the epidemic.

Keywords: Sars-COVID19, time series analysis, visibility algorithms, graph theory, Wastewater-Based Epidemiology

1. Introduction

The severe acute respiratory syndrome from SARS COVID-19 originated in late 2019 and spread rapidly around the world, causing cascading deaths and uncertainty for the future of global and national economies. Italy was among the first countries to interface with this state of emergency, supporting the development of various strategies for monitoring and controlling the virus. Initially, the monitoring of the virus was done through molecular and antigenic tests. Their use was, however, often questioned both for their costs and for the significant percentage of false negative results (Zhang et al., 2021). A significant revolution in monitoring techniques occurred with the discovery that traces of SARS COVID-19 were detectable in wastewater (La Rosa et al., 2020). This allowed to monitor entire communities simultaneously using non-invasive and inexpensive procedures, such as the Wastewater-Based Epidemiology (WBE) approach (Sims & Kasprzyk-Hordern, 2020), a technique that analyses wastewater to detect the presence of chemical or biomarkers (e.g., drugs, alcohol, etc.).

The approach was tested for monitoring SARS-CoV-2 and to detect traces of viral RNA in wastewater. In Italy, this happened within the project called SARI (*Environmental surveillance of Sars-cov-2 through urban wastewater in Italy*) (Rosa et al., 2021) started on the 1st of July 2020, which provided for an environmental surveillance activity for Sars Cov-2 based on the WBE model.

The WBE technique provides, downstream of the analysis, many data grouped in the form of time series, from which it is important to extract as much information as possible on the trend of the phenomenon in order to be able to interpret it and to develop adequate control measures.

The analysis of the time series is also possible through an innovative approach, which involves transforming the starting time series into a visibility graph through the so-called visibility algorithms (Lacasa et al., 2008), in order to identify specific characteristics of the phenomenon. The analysis of the visibility graph, extracted from the time series, is then analysed with metrics proposed by the Complex Networks Theory (CNT).

In this work, the time series of viral RNA concentration collected at a wastewater treatment plant (WWTP) located in Campania Region (Italy) is transformed into visibility graphs and then studied by using CNT metrics (Newman, 2010; Giustolisi et al., 2017), in order to extract useful information on the trend of the phenomenon that cannot be detected by the analysis of the time series alone, and validate this approach as a tool for the analysis of other substances for which concentration values over time, in specific sections, are available. The main objective is to produce results that allow to support virus control activities and also suggest ad-hoc monitoring measures.

2. Complex Network Theory (CNT)

2.1. Basic Concepts

The complex network theory (CNT) provides a novel perspective in the study of complex systems, associating the latter to graphs, i.e., a set of nodes, which represent the components of the system, connected to each other through links, which represent the relationships between the components. A graph is a mathematical object G= (N, L) where N = {1, . . ., n} is the set of nodes of G, and L = {l_1, . . ., l_m} is the set of links of G. The adjacency matrix A is the most used representation of a graph G and indicates whether couples of nodes are connected or not in the graph, and is defined by the conditions,

$$a_{ij} = \begin{cases} 1 \; if \; (i,j) \in L \\ 0 \; if \; (i,j) \notin L \end{cases} \qquad (1)$$

Graphs can be directed, if the links have a direction, and undirected otherwise. For indirect graphs, the relationship between nodes is symmetric, i.e., given two nodes (i, j) є G, $l_{ji} = l_{ij}$, while for direct graphs $l_{ji} \neq l_{ij}$. For undirected graphs, the adjacency matrix is symmetric with respect to the main diagonal, while for direct graph it is not necessarily. For the purposes of the analysis proposed, it is necessary to define some basic concepts and metrics of the CNT (Newman, 2010).

Two nodes are adjacent if they are connected by a link and the number of connections of a node i, denoted by k_i, represents its nodal degree. The average number of links per node in the graph is the Average degree (<k>).

A path is a sequence of nodes {n_1, n_2, …, n_n} connected by links and a shortest path represents the path with the minimum number of links between two nodes (i, j). The greatest shortest path refers to the diameter (D) of the graph G. The Average path length (APL) is the average of the shortest paths between all the nodes in the network.

The clustering coefficient of a node i (C_i) (Watts & Strogatz, 2011) measures to what extent its neighboring nodes tend to form a complete graph, i.e., a graph where each pair of nodes is connected by a link. The global clustering coefficient $C(G)$ of the graph is the average of the local values C_i.

Several models have been developed in order to classify complex systems according to its connectivity structures.

Erdos and Reny (1959) proposed random networks, i.e., networks characterized by a high homogeneity, with a degree distribution randomly distributed around an average value. Watts and Strogatz (2011) proposed small-world networks as systems similar to random ones (i.e., very homogeneous) but dominated by the so-called small world effect (Milgram, 1967) about six degrees of separation. In these networks, the most part of the nodes are not close to each other, but however each component in the network can reach each other through a short sequence of links. Barabasi and Albert (1999) proposed the scale-free networks, characterized by non-homogeneous nodal degree distributions, i.e., networks where many nodes have a low degree and few nodes (called hubs) have a high degree.

Small world and random networks, characterized by a homogeneous degree distribution (Poisson), present a significant structural resistance to both random failures and intentional threats, while scale free networks, characterized by a heterogeneous degree distribution (Pareto) show a very high structural resistance to random failures but a weak resistance to intentional threats.

2.2. Visibility Algorithms

The visibility algorithm is based on the idea of visibility between points/nodes, which correspond to the measurements of specific quantities within the series. Considering a time series composed of data of different heights (Figure 1-a) forming a histogram, it will be possible to draw links (red lines in Figure 2-b) between the various measured values only if there is visibility between them. The visibility lines thus traced represent the links of the visibility graph (Figure 1-c) associated with the time series.

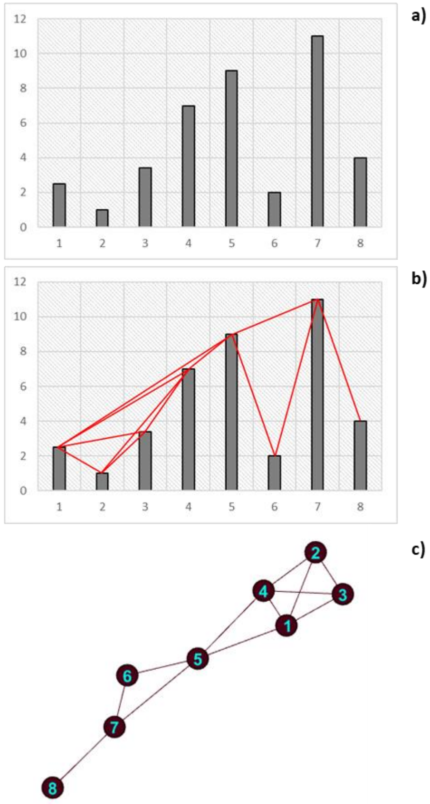

Figure 1. Example time series (a); visibility lines between points of the time series (red) (b); visibility graph of the starting time series (c).

In mathematical terms, if $\{x(t_i)\}_{i=1\ldots N}$ is a time series of N data, the natural visibility algorithm (NVg) (Lacasa et al., 2008) assigns each data of the series to a node of the visibility graph. Two nodes, i and j, in the graph are connected if a straight line can be drawn in the time series joining $x(t_i)$ and $x(t_j)$, which does not intersect any intermediate datum at height $x(t_k)$, i.e., the following geometric criterion is satisfied within the time series,

$$x(t_k) < x(t_i) + \left(x(t_j) - x(t_i)\right)\frac{t_k - t_i}{t_j - t_k} \qquad (2)$$

The graph associated with the time series is always:

- connected: each node sees at least its closest neighbors (left and right);
- undirected: there is no defined direction in the links;
- invariant under affine transformations of the series: it is invariant under horizontal and vertical axis scaling, as well as horizontal and vertical translations;
- lossy: some time series information is inevitably lost in the mapping due to the fact that the

network structure is completely determined in the adjacency matrix, and for two periodic series with the same period but different intensities, for example, there would be the same graph.

3. Methodology

The present work proposes an application of the visibility algorithms (Iacobello et al., 2018; Lacasa et al., 2008) to the time series of the Sars COV-2 concentration data, detected in the influent to the WWTP of Villa Literno, in Campania region (Italy). The WWTP covers a population of 631.714 population equivalent (PE) and the time series refers to samples collected from December 2021 to September 2022, whose data were obtained within the SARI project (Rosa et al., 2021), aimed at environmental surveillance of Sars COVID-2 in Italian sewer.

The visibility graph corresponding to the starting series is analyzed using the CNT tools, in order to capture useful information on the spread of the epidemic that cannot be detected by the analysis of the time series alone. Once the time series is transformed into a visibility graph, it is possible to obtain the adjacency matrix and then evaluate the previously defined CNT metrics. Finally, the association of the degree distribution of the graph to one of the models proposed in the literature can be useful in defining the behavior of the epidemiological phenomenon.

4. Results and discussion

Figure 2-a plots the starting time series of the analyzed data as histogram and Figure 2-b reports the visibility graph corresponding to the time series. The graph is characterized by many poorly connected nodes and few most connected nodes (hubs), generally corresponding to the highest concentration values in the time series. Table 1 reports relevant data of the basin serving the WWTP and the values of the CNT metrics.

Table 1. Relevant Data and CNT metrics for the visibility graph of the WWTP of Villa Literno.

# nodes	# links	<k>	diameter	Average CC	APL
75	236	6.29	6	0.787	2.94

Although the presence of hubs shortens the distances between pairs of nodes in the network, these elements do not create a huge gap, in terms of degree, between nodes. The only exception is represented by nodes 26 and 56, with more than 20 connections.

Degree and clustering coefficient for each node are shown in Figure 3. The higher the number of connections of a node, the lower its clustering coefficient, probably due to the fact that high values of the ordinate prevent visibility between neighboring nodes, and therefore the formation of triangles. The high values of the clustering coefficient (CC=0.787) and the values of the average degree (<k>=6.293) denote a very interconnected system, and the presence of the two hubs facilitates the connection between distant nodes even over short distances, as confirmed by the Average Path Length value (APL=2.944). Despite this, the diameter value is quite high considering the size of the network (D=6), although perfectly within the range defined by Milgram (Milgram, 1967) for small-world networks, based on the six degrees of separation in social networks.

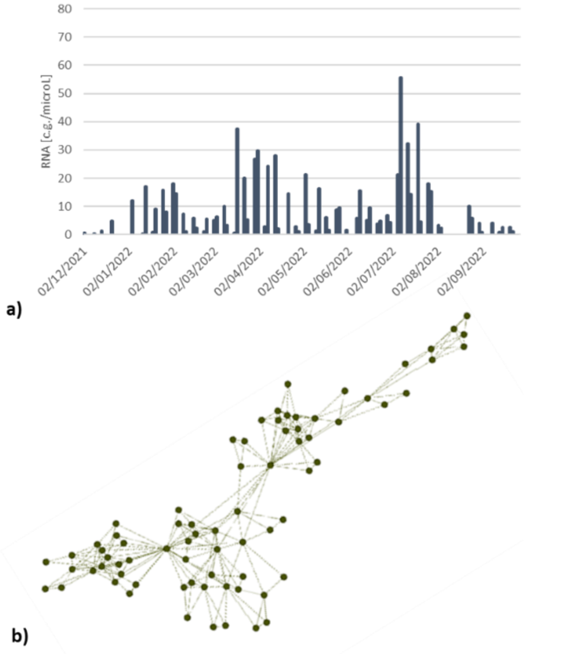

Figure 2. Starting time series of the WWTPs of Villa Literno as histogram (a) and visibility graph corresponding to the time series (b).

Figure 3. Degree (a) and clustering coefficient (b) for each node of the graph.

This combination of values of the metrics, which reflects characteristics of very homogeneous systems (random), indicates that the degree distribution of the graph follows a Poisson's law (Figure 4). Most of the nodes have similar degrees and are clustered around decreasing average values, meaning that people live in a space that influences their habits and act mainly over short distances. İt is possible to state that the spread of the epidemic identifies the graph analyzed as responding to spatial network, where the spatial features make it probable that the random contamination curve is repeated with decreasing average values since the phenomenon is decreasing over the analysed period, i.e., the concentration values detected in the time series are, however, dependent on the values of their neighbours. The presence of the two hubs on the final tail of the curve indicates a low probability of other contagion events, or in any case of minor magnitude.

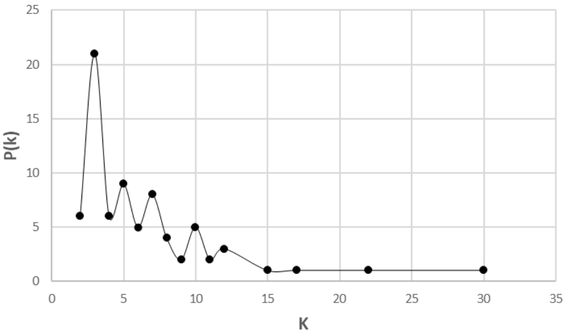

Figure 4. Degree distribution of the visibility graph.

This first approach to the study of Sars-Covid 2 time series through the use of visibility algorithms was useful for defining the characteristics of the phenomenon not previously defined, or at least not defined with standard techniques. The fact that this phenomenon responds to a spatial network system has already been previously announced in the literature (Sharma et al., 2021) using the analysis of social and contact networks. However, the implications for more in-depth studies are promising.

5. Conclusions

The present work proposes an alternative approach for studying the Sars-Covid 2 time series. The visibility algorithm approach is used to convert the time series of RNA concentration detected in the influent to a WWTP into visibility graph for analyzing it with CNT tools and capture information on the behavior of the epidemiological phenomenon that cannot be detected by the analysis of the time series alone.

The work evaluates the feasibility of the approach for the study of the phenomenon proposing a first preliminary analysis, which allowed to evaluate the variability of the phenomenon and to associate it to random/small world networks, i.e., homogeneous systems strongly influenced by spatial constraints. This initial information about the behavior of the phenomenon could be of support to the planning of control and monitoring strategies. In this regard, it can be concluded that the work opens towards future perspectives in the in-depth analysis of both the same phenomenon (i.e., Sars-COV-19 spread) and that involving different substances detectable in wastewater.

References

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*. https://doi.org/10.1126/science.286.5439.509

Erdös, P., & Rėnyi, A. (1959). . . ., P,, by selecting N edges from the ; i) possible edges Px (1 5. *Publicationes Mathematicae*, *6*, 290–297.

Giustolisi, O., Simone, A., & Ridolfi, L. (2017). Network structure classification and features of water distribution systems. *Water Resources Research*, *53*(4), 3407–3423. https://doi.org/10.1002/2016WR020071

Iacobello, G., Scarsoglio, S., & Ridolfi, L. (2018). Visibility graph analysis of wall turbulence time-series. *Physics Letters, Section A: General, Atomic and Solid State Physics*. https://doi.org/10.1016/j.physleta.2017.10.027

La Rosa, G., Bonadonna, L., Lucentini, L., Kenmoe, S., & Suffredini, E. (2020). Coronavirus in water environments: Occurrence, persistence and concentration methods - A scoping review. In *Water Research*. https://doi.org/10.1016/j.watres.2020.115899

Lacasa, L., Luque, B., Ballesteros, F., Luque, J., & Nuño, J. C. (2008). From time series to complex networks: The visibility graph. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0709247105

Milgram, S. (1967). small world problem. *Psychological Today*.

Newman, M. (2010). Networks: An Introduction. In *Networks: An Introduction*. https://doi.org/10.1093/acprof:oso/9780199206650.001.0001

Polo, D., Quintela-Baluja, M., Corbishley, A., Jones, D. L., Singer, A. C., Graham, D. W., & Romalde, J. L. (2020). Making waves: Wastewater-based epidemiology for COVID-19 – approaches and challenges for surveillance and prediction. *Water Research*. https://doi.org/10.1016/j.watres.2020.116404

Rosa, G. La, Ferraro, G. B., Mancini, P., Veneri, C., Lucentini, L., Bonadonna, L., Brandtner, D., Grigioni, M., Rossi, M., & Suffredini, E. (2021). *Flash survey on SARS-CoV-2 variants in urban wastewater in Italy 1 st Report (Investigation period : 04 – 12 July 2021). July*, 4–12. https://doi.org/10.1016/j.scitotenv.2020.139652.2

Sharma, N., Verma, A. K., & Gupta, A. K. (2021). Spatial network based model forecasting transmission and control of COVID-19. *Physica A: Statistical Mechanics and Its Applications*. https://doi.org/10.1016/j.physa.2021.126223

Sims, N., & Kasprzyk-Hordern, B. (2020). Future perspectives of wastewater-based epidemiology: Monitoring infectious disease spread and resistance to the community level. In *Environment International*. https://doi.org/10.1016/j.envint.2020.105689

Watts, D. J., & Strogatz, S. H. (2011). Collective dynamics of “small-world” networks. In *The Structure and Dynamics of Networks*. https://doi.org/10.1515/9781400841356.301

Zhang, Q., Zheng, F., Jia, Y., Savic, D., & Kapelan, Z. (2021). Real-time foul sewer hydraulic modelling driven by water consumption data from water distribution systems. *Water Research*, *188*, 116544. https://doi.org/10.1016/j.watres.2020.116544