Journal of Water Process Engineering 66 (2024) 106107

Contents lists available at ScienceDirect

Journal of Water Process Engineering

journal homepage: www.elsevier.com/locate/jwpe

From time series to visibility algorithms: A novel approach to study the spread of SARS-CoV-2 in wastewater

A. Simone [a,*], A. Cesaro [b], G. Esposito [b]

[a] Department of Engineering and Geology, University "G. D'Annunzio" of Chieti-Pescara, Viale Pindaro,42, 65127 Pescara, Italy
[b] Department of Civil, Architectural and Environmental Engineering, University of Naples Federico II, Via Claudio, 21, 80125 Napoli, Italy

ARTICLE INFO

Editor: Guangming Jiang

Keywords:
Time series analysis
Visibility algorithms
Graph theory
SARS-CoV-2
Spatial networks
Wastewater-based epidemiology

ABSTRACT

The detection of SARS-CoV-2 in faeces encouraged various studies exploring wastewater as a disease surveillance tool from a wastewater-based epidemiology (WBE) perspective. Virus concentration data in wastewater are collected and arranged in time series and generally analysed by using statistical approaches. However, for studying complex and non-linear phenomena, this procedure may not be effective. In this regard, the present work introduces an alternative and innovative approach to analyse time series of SARS-CoV-2 concentration in wastewater based on visibility algorithms. The temporal evolution of the epidemic is transformed into a visibility graph that allows the study of time series from a nonlinear perspective. The connectivity structure of the visibility graph encapsulates significant information of the starting time series. By investigating the topological characteristics of the graph, it is possible to extract nontrivial evidence to give a physical interpretation of the phenomenon and to identify the factors that mainly influence the virus transmission. The proposed approach has been applied to the time series data collected at ten wastewater treatment plants to interpret the trend of the epidemic and attempt to forecast the phenomenon in the analysed basins. Overall, using visibility algorithms to study COVID-19 in sewage is a valuable tool for monitoring the community, with potential for predicting epidemics and community behaviours.

1. Introduction

The spread of diseases attracted particular attention in recent years due to Covid-19 [1], a severe acute respiratory syndrome discovered at the end of 2019 that spread rapidly around the world. Airborne transmission has been recognized as one of the primary routes of conveyance of the virus [2], whose propagation and permanence was found to be influenced by both environmental and social factors [3,4]. The decisive actions adopted over time to mitigate the spread of the epidemic [5,6] marked the end of the health emergency state, although, given the uncertainty and lack of information about possible new variants, effective strategies for epidemic monitoring are still required.

A first paradigm shift in epidemiological surveillance occurred with the discovery that traces of SARS-CoV-2 are detectable in wastewater [7]. This finding encouraged several studies on the suitability of wastewater surveillance to determine infection and transmission levels and to produce early warning of outbreaks in local communities [8,9,38,39]. Most of the studies concerning monitoring and detection of SARS-CoV-2 in wastewater rely on the Wastewater-Based Epidemiology (WBE) approach [10], which assesses the presence and quantity of specific chemical or biological markers (e.g., drugs, antibiotics, alcohol, tobacco, etc.) in wastewater samples [11]. The WBE approach as tool for monitoring and detecting traces of SARS-CoV-2 in wastewater has been extensively tested and validated in many industrialized cities worldwide, where numerous procedures were initiated. In Italy, this procedure was formally defined via the project SARI (*Environmental surveillance of SARS-CoV-2 through urban wastewater in Italy*), which started on the 1st of July 2020. The project proposed an environmental surveillance activity based on the WBE model and planned the monitoring of 167 Italian wastewater treatment plants (WWTPs), focusing on big urban centres (more than 50,000 inhabitants) as well as on centres of touristic importance, even with less of 50,000 inhabitants. The results of the monitoring were provided as time series of SARS-CoV-2 RNA concentration detected in the influent at different WWTPs.

Other approaches aimed at tracking the epidemic involved the use of Complex Network Theory (CNT) tools to study the spread of diseases [40], also with reference to the analysis of social networks [12]. In this regard, the main interest in the use of CNT approaches lays in the possibility of analysing several social networks (e.g., friendships, work,

* Corresponding author.
E-mail addresses: antonietta.simone@unich.it (A. Simone), alessandra.cesaro@unina.it (A. Cesaro), giovanni.esposito1@unina.it (G. Esposito).

https://doi.org/10.1016/j.jwpe.2024.106107
Received 22 March 2024; Received in revised form 15 July 2024; Accepted 31 August 2024

Nomenclature

$A = (a_{ij})$	Adjacency matrix
APL	Average shortest path length
ARPAC	Regional Agency for Environmental Protection in Campania
CNT	Complex network theory
C_i	Clustering coefficient of a node *i*
C	Average Clustering Coefficient
COVID-19	SARS-CoV-2 acute respiratory disease
D	Diameter
G	Graph
k_i	degree of a node *i*
$<k>$	Average degree of the graph G
L	set of links in the graph G
N	set of nodes in the graph G
NVg	Natural Visibility Algorithm
P (k)	Degree Distribution
RNA	genome of the SARS-CoV-2 virus
SARS-CoV-2	severe acute respiratory syndrome
WBE	Wastewater-Based Epidemiology
WWTP	wastewater treatment plant

family, etc.) simplifying their representation using graphs. In fact, since the spread of diseases mainly takes place through contact between people, it is possible to represent this phenomenon through a graph where the nodes represent the people and the connections between the nodes represent their relationships [13], by assuming that two individuals are probabilistically connected based on the distance between them.

Interestingly, the CNT can support the study and analysis of epidemics with realistic and high-performance capabilities. This is achieved by recurring at the time series data of the evolution of the phenomenon over time, as produced within the WBE approach and employing visibility algorithms. The idea of transforming time series into graphs proved to be an effective method for representing and studying time series, especially for non-linear phenomena, so much so that it has attracted the interest of the scientific community in different fields.

One of the main applications concerns the medical sector [14–16], where it has been used also with respect to the propagation of epidemics [17]. Many environmental topics have been studied by visibility algorithms, including the analysis of CO_2 emission time series relating to the carbon price [18], the impacts on people's health and daily life through the analysis of air quality patterns [19], the daily and monthly sunspot series [20], the lag between two hydrogeological time series [21], and the velocity components of a fully developed turbulent channel flow [22].

In this regard, the present work aims at proposing an innovative approach for the analysis of SARS-CoV-2 RNA concentration time series. The approach involves transforming each viral RNA concentration time series into a visibility graph through the visibility algorithms [23] proposed by the CNT. The connective structure of the visibility graph inherits many characteristics of the starting time series and facilitates the extraction of valuable insights into the behaviour of the epidemic. This is achieved through CNT topological metrics [24], with the main aim of highlighting how the spread of the epidemic varies across different contexts. The findings aim to suggest more effective measures for containing and managing the epidemic.

The added value of this approach is its usefulness in gathering information on the non-linear trend of the phenomena that cannot be detected by the analysis of the time series via conventional methods. To validate the proposed approach, the time series of viral RNA concentration collected at ten WWTPs serving sewer network catchment areas (basins) with different characteristics (e.g., number of inhabitants, etc.) were analysed.

Experimental outcomes addressed the development of a promising tool to study time series from a new perspective as well as the discussion about the potential applications of the proposed approach for environmental surveillance purposes. Although this application focused on the analysis of the collected data, its aim is to represent a first step towards defining a multidisciplinary approach, which addresses not only engineering and environmental aspects, but also incorporates social and health considerations.

2. Complex Network theory (CNT)

The Complex Network Theory (CNT) integrates mathematical and technical tools to analyse the nature, complexity and dynamics of many real complex systems. These systems consist of many components that interact with each other, exhibiting behaviours that are difficult to model due to the properties such as nonlinearity, emergence and spontaneous order, as well as the interaction with the surrounding environment.

The basic concepts and models of CNT considered for the purposes of this study are reported in the Section 1 of the Supplementary Material.

2.1. Visibility algorithms

The study of specific phenomena, such as temperature of a basin or level of rainfall, characterized by temporal variability, generally relies on collecting data at discrete time interval, in a specific point or section. These data are often stored as time series, usually displayed as line graphs or histograms, which provide the primary basis for describing and analysing the trend of a phenomenon. The visibility algorithms [23] provide the transformation of such time series into graphs. These graphs are then used to identify specific characteristics of the phenomenon that cannot be detected via the sole analysis of the time series by using CNT metrics and models.

The most used algorithm for time series analysis is the Natural Visibility Algorithm (NVg) proposed by Lacasa et al. (2008), used throughout the present work. A time series with n observations, represented as lines of a histogram (Fig. 1-a), corresponds to a visibility graph $G = (N,L)$, with a set of nodes $N = \{1, \ldots, n\}$ connected by a set of links, $L = \{1, \ldots, l\}$, which occur only between nodes whose visibility is not obstructed, i.e., between which it is possible to draw lines without intersecting others (Fig. 1-b). The number n of nodes, together with the drawn links, form the visibility graph of the starting time series (Fig. 1-c). (See Section 1.3 of the Supplementary Material for further information).

The visibility algorithm can be envisioned as a geometric transform, which decomposes a signal/series in a concatenation of graph's motifs, and the degree distribution simply makes a histogram of such 'geometric modes', thus showing clearly distinct, discrete structures that characterize the periodic time series.

3. Methodology: from time series to visibility graph

The present work proposes a novel approach for studying the time series of SARS-CoV-2 RNA concentration for several WWTPs by using the CNT tools, to evaluate the different diffusion trends of the epidemic and identify the mechanisms that mainly influenced its propagation. More specifically, the original data were provided by the Regional Agency for Environmental Protection in Campania (ARPAC) within the SARI project [25], and they referred to the viral RNA concentration detected over a period of approximately 12 months (from October 2021 to September 2022) in the influent of ten WWTPs serving small, medium and large sewer network catchment areas. The approach has been applied to the raw time series without carrying out any data pre-

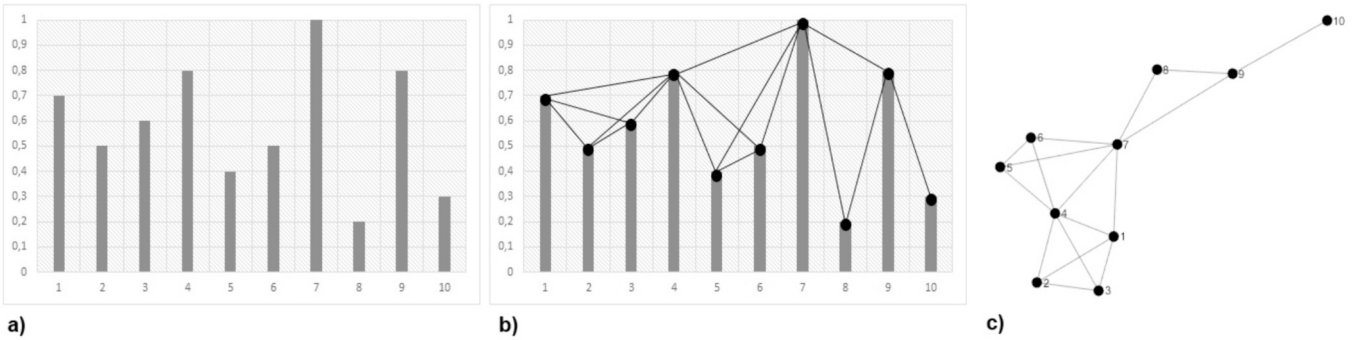

Fig. 1. Bar plot for an example time series (a); visibility lines between nodes (b); visibility graph of the starting time series (c).

processing or removal of outliers. The reason is mainly linked to the inaccessibility of the collected samples. In fact, the collection of samples was carried out by an external public agency (ARPAC) as part of a government project (SARI) that had set up maximum safety conditions, considering the probable risk linked to the contact with the samples themselves.

The same agency rearranged the data in the time series herein used. Nonetheless, this condition made it possible to validate the approach even in the presence of anomalous data. However, it is important to note that despite potential noisy data and variability in flow rates typical of wastewater analyses, these challenges are inherent to the analysed series rather than specific to the proposed approach. Therefore, this general limitation applies to both conventional and innovative methodologies alike. Table 1 reports relevant data for each WWTP.

The main steps of the proposed methodology, plotted in Fig. 2, are described below.

1. Time series of viral RNA concentrations detected in a specific period, at the inlet of ten WWTPs, were collected. For each WWTP, the number of samplings is variable, as shown in the last column of the Table 1, depending on (i) the number of inhabitants of the basin served (two weekly samplings for basins serving more than 50,000 inhabitants and 1 weekly sampling for basins serving fewer than 50,000 inhabitants); (ii) the possibility of carrying out scheduled sampling; (iii) the period in which the monitoring began.
2. The concentration time series were transformed into visibility graphs through visibility algorithms, so that each infection curve was visualized and studied through its associated visibility graph instead of time series. Each visibility graph is characterized by an adjacency matrix, which provides important information on its connectivity and represents the basis for evaluating CNT metrics.
3. Several CNT metrics were evaluated for each visibility graph, in order to define both the degree distributions and network features. More specifically, for the purposes of this study, nodal degree, average degree, clustering coefficient, diameter and average path length were evaluated.

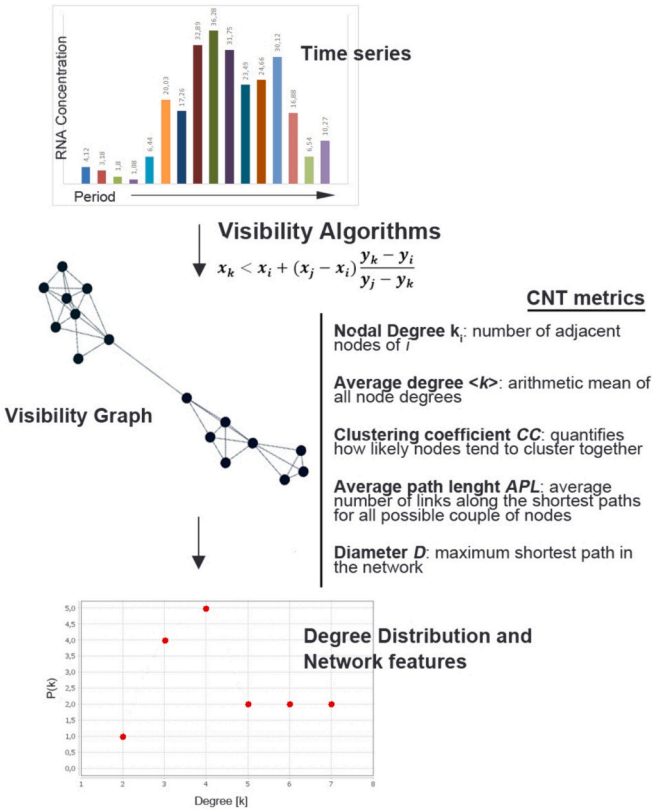

Fig. 2. Outline of the proposed methodology.

The results obtained for each WWTP were comparatively assessed to evaluate the incidence of housing contexts and population habits in the spread of the virus.

Table 1
Relevant Data for the WWTPs analysed.

ID	WWTP Location	Province	Inhabitants	Sampling period	Number of samples/ week	Total number of samples
1	Eboli	Salerno	30,000	02/15/22–09/27/2022	1	31
2	Manocalzati	Avellino	140,000	11/11/21–09/29/22	2	44
3	Napoli Ovest ex Camaldoli	Napoli	250,000	10/05/21–09/29/22	2	94
4	Nocera Sup	Salerno	299,121	10/05/21–09/29/22	2	93
5	Area Casertana	Caserta	370,769	10/05/21–09/29/22	2	93
6	Area Nolana	Napoli	400,000	10/05/21–09/29/22	2	91
7	Villa Literno	Caserta	631,714	12/02/21–09/27/2022	2	75
8	Salerno	Salerno	700,000	12/02/21–09/29/2022	2	76
9	Napoli Ovest	Napoli	950,000	10/05/21–09/27/22	2	93
10	Napoli Est	Napoli	1,750,000	10/05/21–09/29/22	2	93

4. Results and discussion

The present section discusses the results coming from the application of the proposed methodology to the time series of viral RNA concentrations obtained from the monitoring of the WWTPs reported in Table 1.

The first part of the section describes the application of the proposed methodology to the smallest case study, corresponding to the WWTP of Eboli: this serves the sewer network catchment area with the least number of inhabitants. Since the basin serves a population of less than 50,000 inhabitants, as per national directive, ARPAC planned to carry out only one sampling per week during the reference period.

In the second part of this section, the methodology is extended to the remaining WWTPs, for which two weekly samplings were instead carried out, because referred to WWTP serving basins with a population exceeding 50,000 inhabitants. The issue regarding the reliability of the monitored data is also evident concerning the sampling frequency, which should have been higher compared to what was scheduled, particularly regarding aspects related to flow variability, especially under dry period conditions. It is important to note that the data considered for the purpose of this analysis are those officially collected and disseminated by ARPAC. However, the issue of the number of samples was considered in the interpretation of the results, as discussed later in the text.

4.1. Study of the visibility graph obtained for the WWTP of Eboli

The steps supporting the visibility-based approach within the proposed methodology are shown in Fig. 3 with reference to the time series obtained for the WWTP of Eboli. Panel 3-a plots the starting time series and the corresponding visibility graph; red vertical lines represent the viral RNA concentrations detected on specific days from the 15th of February to the 27th of September 2021 and the blue lines indicate the visibility between the various concentrations/nodes. The corresponding visibility graph is characterized by many poorly connected nodes and few very connected nodes, the latter representing the hubs of the system.

Once transformed the time series into the visibility graph, it is possible to evaluate the CNT metrics previously defined and to plot the degree distribution. Table 2 reports the metrics evaluated for the visibility graph obtained from the time series.

Fig. 3-b reports the clustering coefficient for all nodes of the visibility graph. It is possible noting that all nodes with the maximum value of the clustering coefficient (equal to 1) correspond to minimum points in the time series, such as node 4 (see Fig. 3-a). Indeed, low values are surely surrounded (left and right) by higher values, which are visible to each other since they are not hindered by any intermediate values. This visibility supports the formation of triangles, thus guaranteeing high values of the clustering coefficient. Conversely, nodes with the minimum values of the clustering coefficient generally correspond to maximum points in the time series because, having very high ordinates, they do not allow visibility between their neighbour nodes (left and right), thus preventing the formation of triangles. Therefore, the clustering coefficient can be used as a parameter to evaluate the variability of the time

Table 2
Relevant Data and CNT metrics for the visibility graph of the WWTP of Eboli.

	VISIBILITY GRAPH – EBOLI					
ID	#nodes	#links	Average degree < k>	Diameter D	Average Clustering coefficient C	Average Path Length APL
1	31	90	5.806	5	0.783	2.615

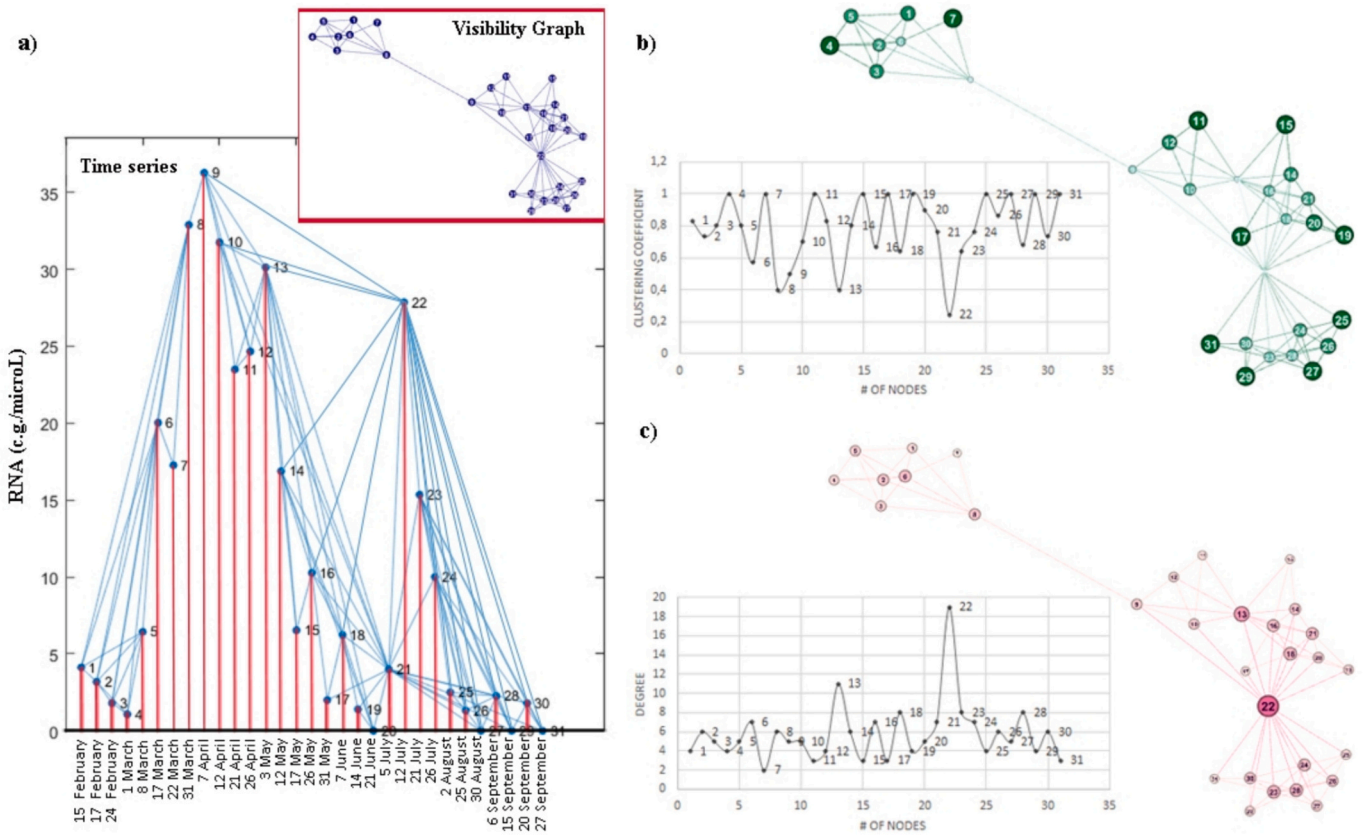

Fig. 3. RNA concentration time series and corresponding Visibility Graph for the WWTP of Eboli (a). Local clustering coefficient (b) and local degree (c) for all nodes of the visibility graph of Eboli.

series, since the formation of triangles is possible only in the presence of fluctuations. In the case of the Eboli network, the average clustering coefficient (0.783) indicates that the phenomenon is quiet variable, but still tending to high values.

The local degree (Fig. 3-c) shows that the most connected nodes in the visibility graph do not necessarily correspond to the maximum concentration values in the time series, as it might be assumed when thinking about the concept of visibility. Node 9, for example, corresponds to the maximum concentration value in the time series but it has a degree equal to 5 in the visibility graph (the higher degree in the graph is $k_{22} = 19$), which is close to the average value of the degree distribution of the graph (Fig. 4). This can be attributed to the fact that, although node 9 has the maximum concentration value, it is surrounded by nodes with similar concentration values, thus hindering the visibility of a large part of nodes with lower concentrations. It could be asserted that the behaviour in this time interval is homogeneous, with close concentration values for the various samplings. The situation is different when referring to nodes 13 and 22; in fact, in addition to presenting high concentration values, they are surrounded by nodes with much lower concentration values. Therefore, for these nodes, it is possible to expand the radius of visibility, which implies an increase in their number of connections, which is much higher than that of the other nodes, so much so that they are defined as hubs of the system. These observations highlight how the inflection points, which allow greater/less visibility (influencing the degree) and interconnection (influencing the formation of triangles) represent a key aspect in the analysis.

This outcome also supports the fact that highly connected nodes (hubs) do not have, generally, high values of the clustering coefficient because it is either difficult or improbable that all their neighbourhood nodes are connected to each other. At the same time, the presence of highly connected nodes (hubs) shortens the distances in paths between nodes in the visibility graph and influences the average short path length (APL = 2.615), which indicates the paths that allow connecting pairs of nodes in the graph with the fewest number of steps.

The degree distribution of the visibility graph of Eboli, reported in Fig. 4-a, has the characteristics of a Poisson curve, with most of the values concentrated near the mean value ($<k \geq 5.806$), thus deviating from the power law distribution, typical of scale-free models, and suggesting a behaviour more similar to random and small-world networks. The high value of the average clustering coefficient coupled with both the low value of the APL and the presence of few hubs in a predominantly random distribution, support the hypothesis that the network better corresponds to a small-world model, whereby most of the nodes can be reached through a low number of steps. This, in turn, perfectly responds to the Milgram's principle of the six degrees of separation [26], characterizing social networks. This trend is further confirmed by the adjacency matrix, reported in Fig. 4-b, which shows how most of the nodes (non-zero values) are grouped near the main diagonal, indicating that connections between nodes occur mainly over short distances.

More specifically, the spread of the epidemic, as modelled, identifies the analysed graph as responding to characteristics of spatial network, meaning that the space where people live influence their habits. Furthermore, the spatial component makes it probable that the contamination curve is repeated with the increase in the number of infections, whose average value will be a function of the concentration levels of viral RNA.

Spatial networks (e.g., communication networks, biological networks, neural networks, etc.) are systems in which the position of the elements plays a very important role because their distance strongly influence the system behaviour. In the spread of epidemics, the spatial component has been analysed considering various network configurations, from contact networks [27], to social networks [13] to the network of movements between individuals [28,29].

The degree distribution provides relevant information to characterize these systems, since it informs on the potential of each single individual to be infected and to cause further infections. Obviously, the greatest is the number of connections among individuals, the greatest is the probability that they are close to an individual who is either already infected [30] or likely to get easily infected. With reference to SARS-CoV-2, several authors demonstrated how socio-demographic and environmental characteristics significantly influenced the spatial patterns of transmission [31,32]. Similarly, the lack of spatial stationarity has been observed in the associations between environmental characteristics and transmission of SARS-CoV-2 [33], as these correlations vary depending on the space within which they develop [34]. The characteristics of the built environment (e.g., houses, schools, offices), the socio-economic activities (e.g., shopping centers, restaurants) and the services offered (e.g., transport) significantly influence viral transmission and incidence rates as they promote travels and/or meetings of people. In other terms, infectious diseases spread more rapidly in active social contexts, where frequent interactions increase the likelihood of connections between individuals. For this reason, pandemic control measures such as the use of face masks, social distancing, and travel restrictions proved to be useful for containing the spread of the epidemic [35]. However, relying solely on the number of inhabitants as the main discriminant may not have been sufficient.

In this context, the behaviour of the phenomenon at Eboli WWTP, with the degree distribution characterized by only one important peak settled around low asymptotic values, could be justified by the fact that it serves a small social context, mainly characterized by exchanges at both family and work levels. Although it is a touristic city, there is never an overpopulation that could affect the transmission of the virus, nor significant industrial and commercial centres. These characteristics proved to be favourable during the state of emergency, as they preserved the city context from important diffusion of the virus and, therefore, peaks in terms of infections. This conclusion, although realistic, could be obvious at times, as low virus transmission is expected in residential contexts with small populations. Nonetheless, the strategy improves this information by quickly identifying the level of risk (compared to the actual number of hubs and distribution curves of the phenomenon) or

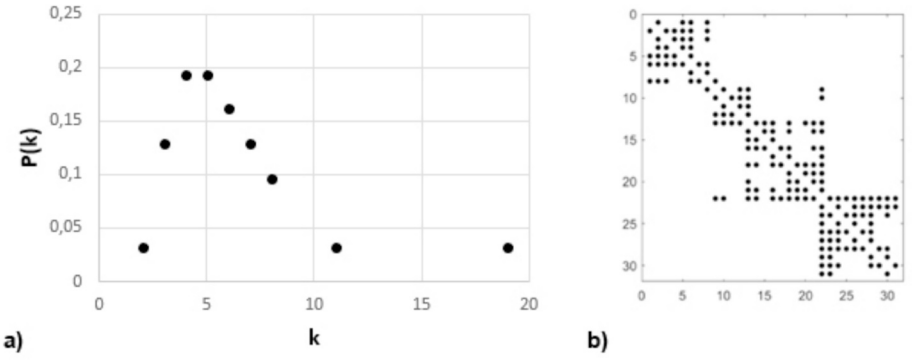

Fig. 4. Degree distribution P(k) (a). Adjacency matrix. Nonzero values are coloured while zero values are white (b).

situational awareness (suggesting specific preventive measures over others) solely from wastewater measurements, precisely for effectiveness in analysing non-linear phenomena. This information could represent a valuable support in times of health emergency, or even for monitoring specific substances.

4.2. Study of the visibility graphs obtained for the other WWTPs

The visibility-based approach was also applied to the time series of the others WWTPs reported in Table 1. The visibility graphs characteristics and the corresponding CNT metrics are reported in Table 3. A first analysis is possible only considering the elements characterizing each visibility graph. By correlating the number of nodes (corresponding to the number of samplings in the time series for each WWTP analysed) with the number of links for each visibility graph, they result highly correlated ($r = 0.965$), i.e., their visibility follows the same trend. Although it is obvious that as the number of nodes increases, the number of links increases, it is not equally obvious that this increase is so consistent for all graphs. This result highlights the fact that, regardless of the number of samplings performed for each basin, the concentration values in each time series are such as to globally develop a similar behaviour in terms of visibility, minimum, maximum and inflection points. The average degree $<k>$ is not a function of the number of inhabitants, but rather of the characteristics of the basins. In fact, it assumes higher values, indicating a greater interconnection (greater infections, peaks and inflection points and greater visibility for specific nodes) in correspondence to metropolitan areas and tends to decrease in correspondence to mostly residential centres.

Moreover, although not directly related, it generally increases as the number of inhabitants increases, precisely to indicate that in large cities, with greater probability of spreading the virus, there are greater possibilities of visibility/connection between nodes of the graph. In this case, however, the number of inhabitants to refer to is those who carries out their own activities in the basin considered. In effect, one of the problems in the correlation between the number of inhabitants served by each WWTP and the metrics of the corresponding visibility graph lies in the fact that the inhabitants considered (residents) are not those who actually live that area on a daily basis, due to the countless movements for study, work, tourism, etc. It follows that, inevitably, metropolises will be overcrowded during the day, leaving most residential contexts almost deserted. However, this fact did not limit their contamination, due to their inevitable interactions with people returning from metropolises. It follows that the information on the number of inhabitants of each basin, although partially correlated with some of the CNT metrics evaluated (e.g., $r = 0{,}60$ with $<k>$) for the visibility graphs, loses relevance. This means that utilizing indicators that track movement networks between cities (for purposes such as tourism, work, and study) can provide a clearer picture of the number of people arriving in or departing from a city over a specific period. From this point of view, this study aims to take the first step towards defining a multidisciplinary approach, which incorporates not only engineering and environmental aspects but also social and health dimensions.

The values of the diameter of the visibility graphs, in the range [4 – 6], once again leads such networks to small-world models, with maximum distance between pairs of nodes, in terms of number of steps, less than or equal to 6. Obviously, visibility graphs with a diameter equal to 4 characterize networks in which the distances between pairs of nodes are very short, due to the presence of many hubs or few highly connected hubs, to which also the maximum values of the average degree $<k>$ is generally associated. That is, in these basins, people are more likely to be infected.

As observed for the diameter, the APL values are smaller for large graphs, because the more hubs that characterize these systems shorten the distances between pairs of nodes. The lowest APL value corresponds to ID10, which certainly exceeds the other visibility graphs in size (number of pipes), and for which the highest value of $<k>$ and clustering coefficient occur. Therefore, this system is the most interconnected (high $<k>$ and Average C) and the one that best allows contact between nodes even over long distances (low APL and diameter). Graphs ID3 and ID5 have the same metric values but the corresponding WWTPs serve different number of inhabitants (i.e., 250,000 vs. 370,769), thus supporting the statement that the population does not necessarily influence the spread of the epidemic, or at least not as much as the spatial constraints.

From this point of view, any new restriction should focus more on the social components of the urban centres rather than on the dimensional ones.

On the other hand, the comparison between ID5 and ID6 shows that, although the two sewer network catchment areas serve a similar number of inhabitants, they are differently interconnected (ID5 has higher $<k>$ and Average C values) and organized (ID5 has lower diameter and APL). This difference leads, once again, to the social difference that characterizes the cities served by the two sewer network catchment areas; ID5 is much more densely inhabited, commercial and touristic than ID6, which nonetheless has a prestigious logistics centre.

The graphs that cover longer distances in few steps are Napoli Est (ID10) and the Area Casertana (ID5), probably due to the presence of very important hubs. Both the WWTPs corresponding to these visibility graphs serve sewer network catchment areas characterized by commercial and touristic cities, where an infection diffusion is expected.

The behaviour is also analysed with reference to the degree distribution of the various visibility graphs. Fig. 5 reports the results obtained considering the time series of Manocalzati WWTP (ID2). The size of the nodes is sorted according to the number of connections, so that highly connected nodes are larger in size than less connected ones (Fig. 5-a). Similarly, Fig. 5-b reports the visibility graph with node size sorted with respect to the clustering coefficient values. Many nodes have a high clustering coefficient, i.e., the formation of many triangles between neighbouring nodes is favoured. Once again, nodes with high degree have a low clustering coefficient, probably due to the impossibility of

Table 3
Relevant Data and CNT metrics for the time series/ visibility graph analysed.

	TREATMENT PLANT		VISIBILITY GRAPH					
ID	Name	#Inhab.	#nodes	#pipes	Average degree $<k>$	Diameter D	Average Clustering Coefficient C	Average Path Length APL
1	Eboli	30,000	31	90	5.806	5	0.783	2.615
2	Manocalzati	140,000	44	124	5.636	5	0.764	2.584
3	Napoli Ovest Ex Camaldoli	250,000	94	394	8.383	4	0.818	2.166
4	Nocera Sup	299,121	93	344	7.398	5	0.788	2.550
5	Area Casertana	370,769	94	394	8.383	4	0.818	2.166
6	Area Nolana	400,000	91	371	8.154	5	0.796	2.419
7	Villa Literno	631,714	75	236	6.293	6	0.787	2.944
8	Salerno	700,000	76	292	7.684	6	0.802	3.120
9	Napoli Ovest	950,000	93	350	7.527	5	0.803	2.469
10	Napoli Est	1,750,000	93	434	9.333	4	0.830	1.977

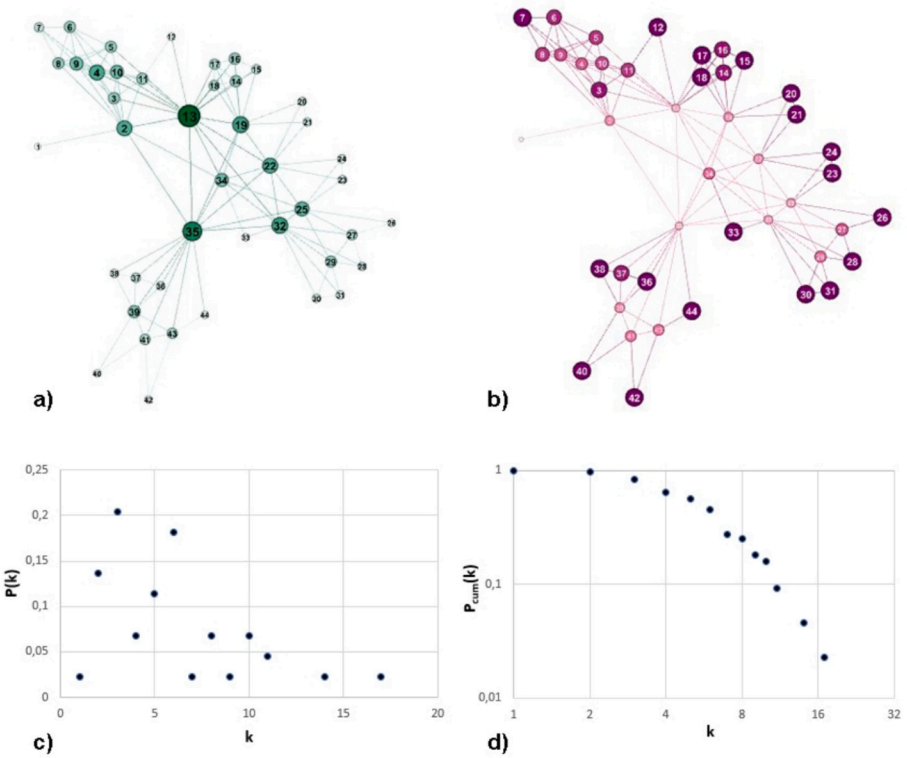

Fig. 5. Analysis of the time series obtained for the WWTP of Manocalzati. Visibility graph with degree information (a). Visibility graph with clustering coefficient information (b). Degree distribution (c). Cumulative degree distribution in the logarithmic scale (d).

forming triangles with immediately adjacent nodes.

The degree distribution is reported in Fig. 5-c. As already observed, the nodal degree is randomly distributed around a mean value. This result can be interpreted in terms of a progressive reduction of infections over the reference period and suggests a lower probability of future infections, while not excluding the possibility of small epidemics. The fact that the curve follows the Poisson distribution is also highlighted in Fig. 5-d, which reports the cumulative distribution in the logarithmic scale.

The treatment plant of Manocalzati serves 19 municipalities possessing very different characteristics and habits, varying from agricultural areas to touristic sites, shops and industrial poles.

Fig. 6 reports the visibility graphs obtained for the other eight time series corresponding to the treatment plants serving the larger sewer network catchment areas, for which information about degree distribution, cumulative degree distribution and adjacency matrix is reported in Fig. 7. It is possible to observe, in all graphs, the presence of many nodes with few connections and few hubs, i.e., nodes with many connections, facilitating connections between different parts of the network in a small number of steps. Sometimes, these hubs result so significant in relation to the nature of the phenomenon analysed that they may suggest possible errors in the original time series, as will be detailed later. Overall, the results obtained show that all the graphs, being subjected to spatial constraints, tend to develop neighbourhood rather than distant connections. This means that the concentration values detected in the network, although randomly distributed, are dependent on the values of

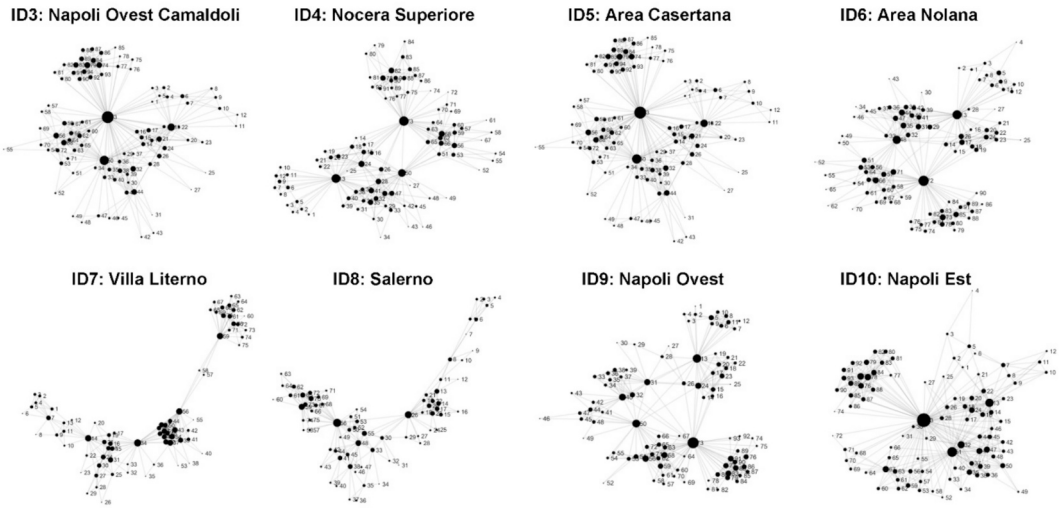

Fig. 6. Visibility graphs for the remaining eight WWTP.

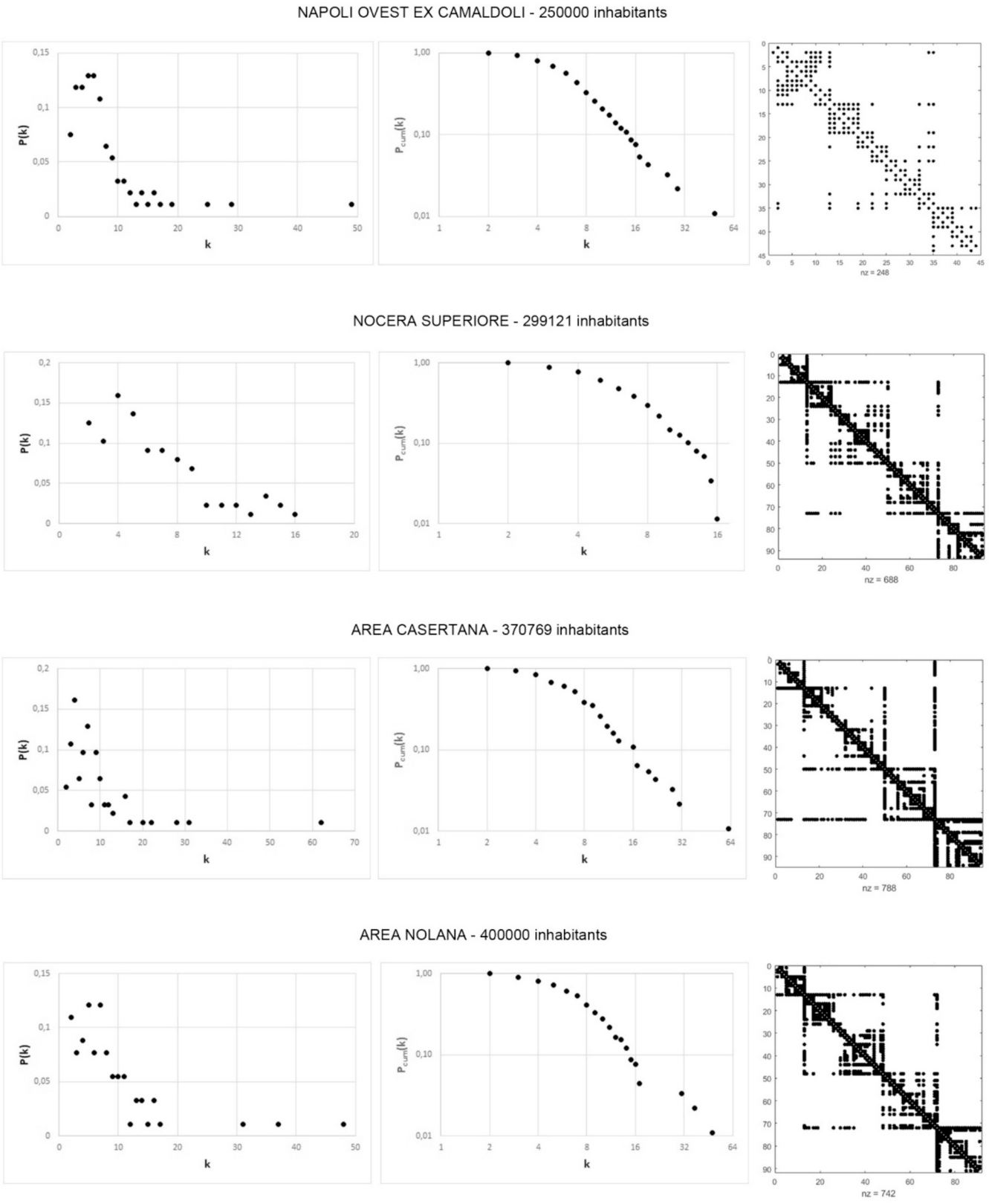

Fig. 7. Degree distribution (left), the cumulative degree distribution in the logarithmic scale (center) and the adjacency matrix (right) for the visibility graphs of Fig. 6.

their neighbours. Therefore, future connections, created based on present and past infection values, can randomly connect to the network resulting in increases or decreases of infections.

The degree distributions for all visibility graphs follows a Poisson distribution. The spatial component, which favours connections between nodes over short distances, is well highlighted in the adjacency

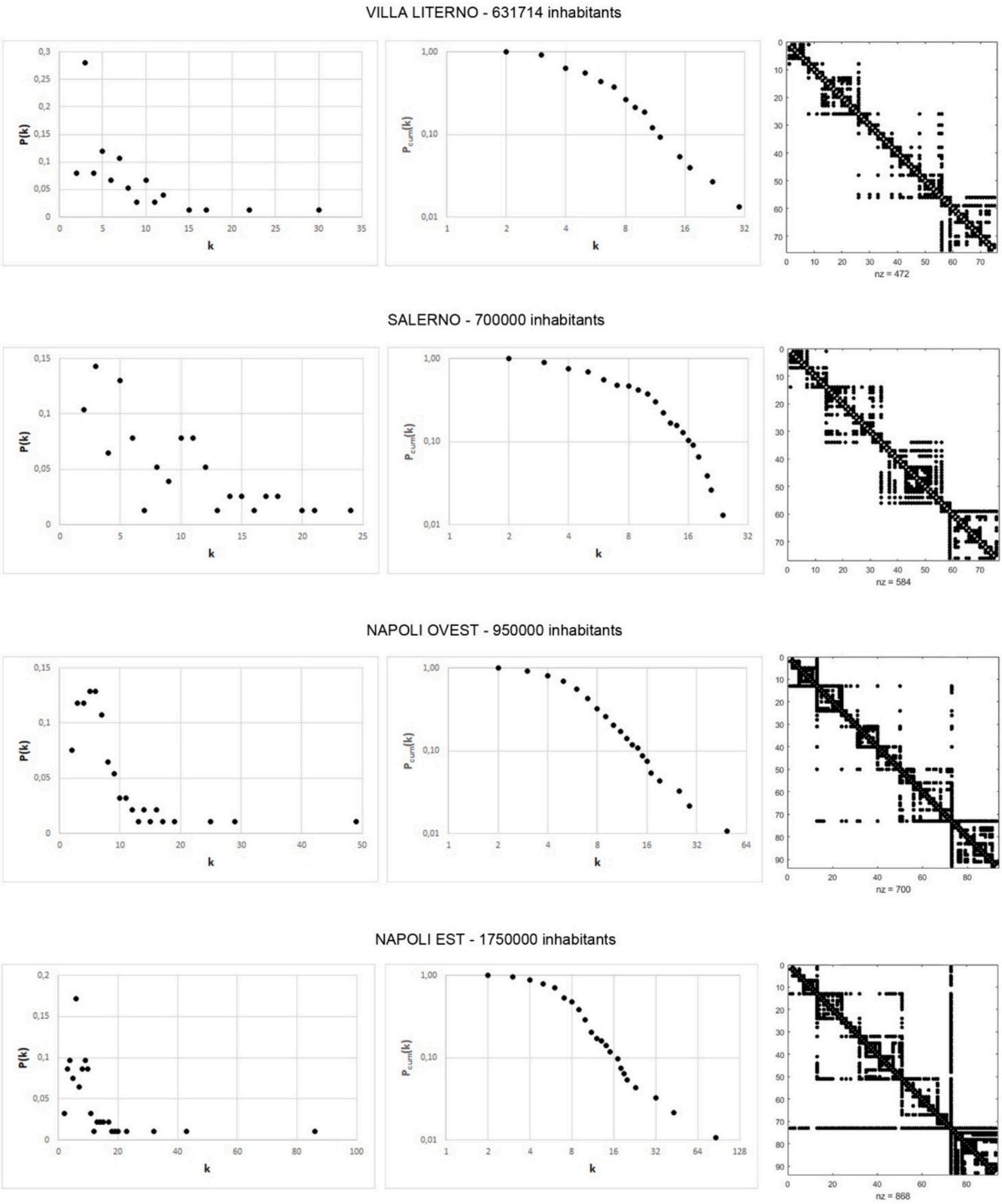

Fig. 7. (*continued*).

matrices reported in Fig. 7. To simplify the comparison between the behaviours of the various sewer network catchment areas, the cumulative degree distributions in logarithmic scale of all visibility graphs of Fig. 6 are reported in Fig. 7 (center).

The logarithmic scale representation highlights how all graphs exhibit similar behaviours attributable to the same phenomenon. However, some graphs demonstrate greater efficiency in information transfer and contact between points (e.g., graphs with more hubs or highly connected hubs) compared to others. Additionally, some graphs deviate from the expected behaviour, suggesting the need for further

investigation into data quality. Therefore, the degree distribution could provide valuable insights both for analysing the phenomenon (e.g., trends, epidemic containment measures) and for assessing the quality of the data (e.g., detection of outliers).

Analysing the distribution of ID3 (Napoli Ovest Ex Camaldoli) and ID9 (Napoli Ovest), it is possible to note that their trend corresponds. Although these two basins are characterized by significantly different numbers of inhabitants (250,000 vs. 950,000), these have similar habits and spatial constraints. Even with respect to the analysis of the curves, therefore, the result of the analysis remains unchanged.

Comparing graphs with similar size (number of nodes and links), such as ID4 (Nocera Superiore) and ID9 (Napoli Ovest), from the Table 3 it is possible observing that the ID9 is more interconnected than ID4, also featuring a higher value of average clustering coefficient, as consequence of the greater variability, and therefore of triangles, in the visibility graph. In this regard, ID9 acts much better in terms of interconnectivity than ID4, probably because of the different social context. This behaviour also emerges by comparing their degree distribution, where it is possible to note that, although with similar number of connections between nodes (pipes), the connective structure of the two graphs is very different, highlighting the clearly homogeneous character of ID4, with nodes having very similar degree and hub quite in line with the behaviour of the analysed phenomenon ($k = 16$), compared to that of ID9, which, although preserves the predominantly homogeneous character of the phenomenon, presents very relevant hubs ($k = 49$). This further consideration also suggests a more in-depth analysis regarding the presence of such relevant hubs in phenomena with almost homogeneous behaviour, which could result in anomalous values in the time series.

In fact, some distributions present anomalous values towards the final part of the curve because of main hubs in the corresponding visibility graphs, which tend to shift the behaviour from random to scale-free, as shown in Fig. 8. These points could both indicate a strong component of interaction that anomalous concentration values, which should be cautiously evaluated.

From a careful analysis of both the time series and the results, this anomaly in the distribution is evidently due to the presence of an outlier within the starting time series. In fact, as can be seen from the graph in Fig. 9, the starting time series corresponding to four WWTPs are characterized by the presence of a very high concentration value detected on the 07/05/2022. The value can even be 10 times higher than the second highest value in the same series. This fact could be due to a series of factors not attributable to the spread of the virus, such as heavy rain after weeks of drought. In this case, the dry period would have limited the transport of the viral RNA towards the treatment plant, while the rain event would have conveyed to the WWTP the material settled in the drainage system, leading to the detection of a greater value on a single day. Although this hypothesis should be better considered in the light of the viral RNA concentration decay kinetics, it is evident that the accuracy of the data becomes very important, as well as the introduction of amplifying/reducing factors when necessary.

In light of these considerations, however, it is possible to assert that the methodology also resulted reliable in identifying spurious elements, giving the opportunity to improve more and more the various steps involved in the process, and that the same presence of outliers, although affecting the connective structure of the graphs, does not affect the overall result of the analysis.

The analysis performed is useful to state that the main variable in determining the behaviour of the phenomenon is the spatial one. Indeed, the fact that similar behaviours occur for graphs serving sewer network catchment areas with different number of inhabitants (299,121 for ID4 and 950,000 for ID9) confirms that the spread of the epidemic does not only depend on the number of inhabitants, but that habits and spatial limits govern the process of the virus spreading, which is only amplified by the increasing number of inhabitants. Similar outcomes were obtained also in other works on the topic [36,37], based on different strategies, which agree that epidemiological networks are strongly influenced by spatial factors, also confirming that contacts between individuals generally prefer short distances, thus occurring mainly on a local scale.

5. Conclusions

The present work proposes a novel, non-invasive approach aimed at producing a spatial analysis of the trend of SARS-CoV-2. The visibility algorithms were used to convert the time series of viral RNA concentrations, previously identified by using the WBE-based approach, into visibility graphs, which are then analysed with CNT tools. The visibility graphs analysis allowed to extract nontrivial information on the phenomenon.

Monitoring and control measures have so far favoured basins serving many inhabitants, while the present study has demonstrated that the

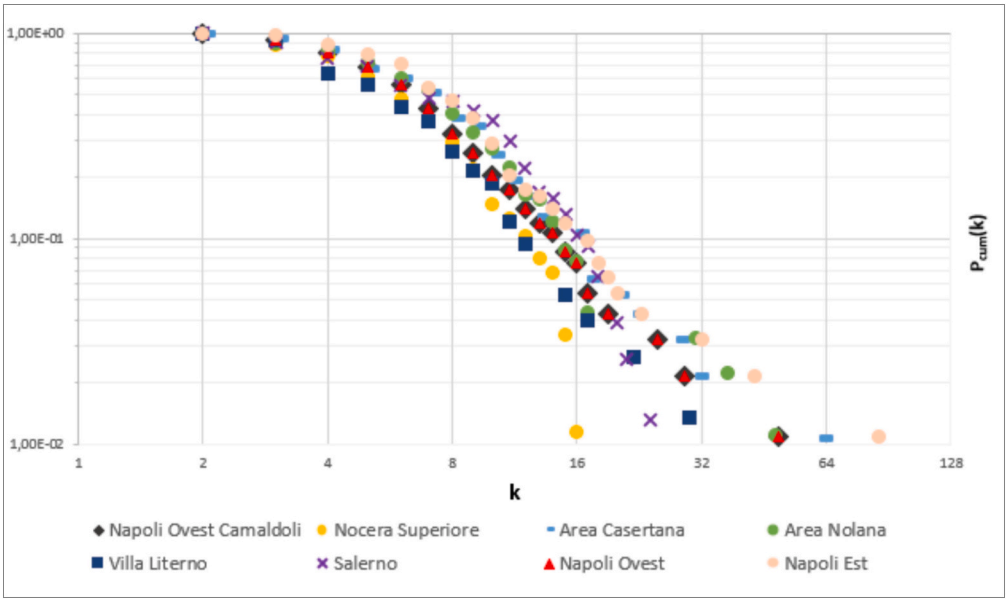

Fig. 8. Cumulative degree distribution corresponding to the visibility graphs of Fig. 6.

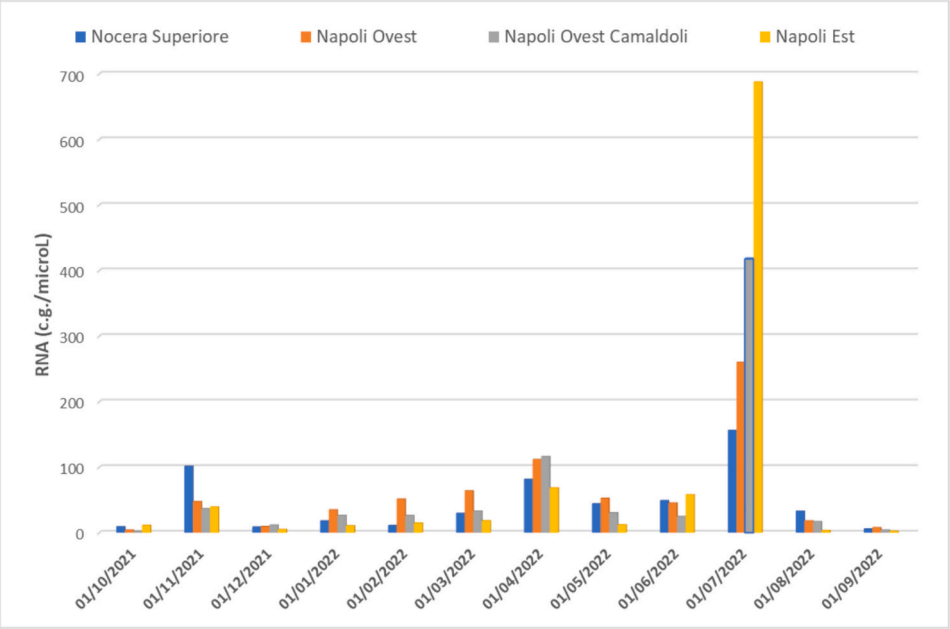

Fig. 9. Time series of viral RNA for four WWTP.

sole number of inhabitants is not indicative of the trend of the phenomenon. Somewhat, the transformation of the time series highlighted how the phenomenon is directly influenced by the spatial constraints and habits in the analysed basins, growing exponentially as the number of inhabitants increases. Therefore, the habits and the spatial development of the territory influenced the spread of the epidemic, which tends to be amplified by the number of inhabitants. This information represents a very important support to the monitoring actions because different measures could be devised according to the type of territory to be monitored. This means that the procedures aimed at preventing the spread of the virus could be much more effective if habits and spatial limits are considered as parameters instead of the number of inhabitants, also in creating an order of importance in either the closure or the control of urban centres in case of alert. In fact, the probability of diffusion is significantly higher in metropolitan areas, characterized by higher level of services for the community, excellent vehicles of the disease.

With respect to the reliability of the data used, it was highlighted how the approach manages to bring out possible outliers. In fact, despite the presence of anomalous data within the analysed time series, the emerging behaviour of the phenomenon remained evident.

In perspective, it would be useful to carry out more in-depth studies that consider other factors in addition to the size of the population, also considering a multidisciplinary approach, such as the level of commuting, land use, daily network of movement between cities, demographics, vaccine uptake rate, variant circulating, to have increasingly reliable results with respect to the trend of the phenomenon and any further measures to contain the epidemic.

The proposed methodology has evidently confirmed the importance of the WBE as an investigative tool. Furthermore, the visibility algorithms coupled with the WBE opens new perspectives, with actions and implications also in social and public health terms, such as, for example, in monitoring the variability in the use of drugs and medicines, in determining periods and areas with a high polluting impact, arranging, where necessary, ad hoc measures and interventions. The study also lays the foundations for further developments, both aimed at a well-founded forecast analysis of the trend of the virus and at the non-invasive investigation of other substances, with obvious implications both in the social and medical fields. More specifically, future efforts may use the proposed methodology to support the development of guidelines during states of emergency by identifying characteristics that may significantly influence the spread of epidemics.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

A. Simone: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **A. Cesaro:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Methodology, Data curation, Conceptualization. **G. Esposito:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used in this study are collected in a controlled access data acquisition and management dashboard, handled by Italian local authorities under the coordination of the Italian Health Institute (ISS - Istituto Superiore di Sanità). Therefore, raw data will remain confidential and will not be shared, unless differently stated by the Managing Authorities.

Acknowledgements

The authors wish to thank dr. Luigi Cossentino and dr. Renato Olivares from the Regional Agency for Environmental Protection in Campania (ARPAC) for providing the raw data on the concentrations of viral RNA collected at the wastewater treatment plants in the framework of the SARI project. The support of Elvira Oliva in the processing of the raw

data is deeply acknowledged as well.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jwpe.2024.106107.

References

[1] L. Alanagreh, F. Alzoughool, M. Atoum, The human coronavirus disease covid-19: its origin, characteristics, and insights into potential drugs and its mechanisms, Pathogens (2020), https://doi.org/10.3390/pathogens9050331.

[2] K.A. Prather, L.C. Marr, R.T. Schooley, M.A. McDiarmid, M.E. Wilson, D.K. Milton, Airborne transmission of SARS-CoV-2, Science (2020), https://doi.org/10.1126/science.abf0521.

[3] J.D. Kong, E.W. Tekwa, S.A. Gignoux-Wolfsohn, Social, economic, and environmental factors influencing the basic reproduction number of COVID-19 across countries, PLoS One (2021), https://doi.org/10.1371/journal.pone.0252373.

[4] L.Y.K. Nakada, R.C. Urban, COVID-19 pandemic: environmental and social factors influencing the spread of SARS-CoV-2 in São Paulo, Brazil, Environ. Sci. Pollut. Res. (2021), https://doi.org/10.1007/s11356-020-10930-w.

[5] J. Howard, et al., An evidence review of face masks against COVID-19, in: Proceedings of the National Academy of Sciences of the United States of America, 2021, https://doi.org/10.1073/pnas.2014564118.

[6] N. Gozzi, et al., Anatomy of the first six months of COVID-19 vaccination campaign in Italy, PLoS Comput. Biol. (2022), https://doi.org/10.1371/journal.pcbi.1010146.

[7] D. Polo, et al., Making waves: wastewater-based epidemiology for COVID-19 – approaches and challenges for surveillance and prediction, Water Res. (2020), https://doi.org/10.1016/j.watres.2020.116404.

[8] R. Gonzalez, et al., COVID-19 surveillance in Southeastern Virginia using wastewater-based epidemiology, Water Res. (2020), https://doi.org/10.1016/j.watres.2020.116296.

[9] K.G. Kuhn, et al., Predicting COVID-19 cases in diverse population groups using SARS-CoV-2 wastewater monitoring across Oklahoma City, Sci. Total Environ. (2022), https://doi.org/10.1016/j.scitotenv.2021.151431.

[10] C.G. Daughton, Monitoring wastewater for assessing community health: Sewage Chemical-Information Mining (SCIM), Sci. Total Environ. (2018), https://doi.org/10.1016/j.scitotenv.2017.11.102.

[11] J. Wen, et al., Stability and WBE biomarkers possibility of 17 antiviral drugs in sewage and gravity sewers, Water Res. (2023), https://doi.org/10.1016/j.watres.2023.120023.

[12] L.A. Meyers, B. Pourbohloul, M.E.J. Newman, D.M. Skowronski, R.C. Brunham, Network theory and SARS: predicting outbreak diversity, J. Theor. Biol. (2005), https://doi.org/10.1016/j.jtbi.2004.07.026.

[13] T.I. Vasylyeva, S.R. Friedman, D. Paraskevis, G. Magiorkinis, Integrating molecular epidemiology and social network analysis to study infectious diseases: towards a socio-molecular era for public health, Infect. Genet. Evol. (2016), https://doi.org/10.1016/j.meegid.2016.05.042.

[14] S. Supriya, S. Siuly, H. Wang, J. Cao, Y. Zhang, Weighted visibility graph with complex network features in the detection of epilepsy, IEEE Access (2016), https://doi.org/10.1109/ACCESS.2016.2612242.

[15] R. Bose, K. Samanta, S. Modak, S. Chatterjee, Augmenting neuromuscular disease detection using optimally parameterized weighted visibility graph, IEEE J. Biomed. Heal. Inform. (2021), https://doi.org/10.1109/JBHI.2020.3001877.

[16] G. Zhu, Y. Li, P.P. Wen, Epileptic seizure detection in EEGs signals using a fast weighted horizontal visibility algorithm, Comput. Methods Prog. Biomed. (2014), https://doi.org/10.1016/j.cmpb.2014.04.001.

[17] D. Tsiotas, L. Magafas, The effect of anti-COVID-19 policies on the evolution of the disease: a complex network analysis of the successful case of Greece, Phys (2020), https://doi.org/10.3390/physics2020017.

[18] Y. Hu, C. Chu, P. Wu, J. Hu, A linear time series analysis of carbon price via a complex network approach, Front. Phys. 10 (November) (2022) 1–14, https://doi.org/10.3389/fphy.2022.1029600.

[19] W. Zhang, Z. Guan, J. Li, Z. Su, W. Deng, W. Li, Chinese cities' air quality pattern and correlation, J. Stat. Mech. Theory Exp. (2020), https://doi.org/10.1088/1742-5468/ab7813.

[20] Y. Zou, M. Small, Z. Liu, J. Kurths, Complex network approach to characterize the statistical features of the sunspot series, New J. Phys. (2014), https://doi.org/10.1088/1367-2630/16/1/013051.

[21] R. John, M. John, Adaptation of the visibility graph algorithm for detecting time lag between rainfall and water level fluctuations in Lake Okeechobee, Adv. Water Resour. (2019), https://doi.org/10.1016/j.advwatres.2019.103429.

[22] G. Iacobello, S. Scarsoglio, L. Ridolfi, Visibility graph analysis of wall turbulence time-series, Phys. Lett. Sect. A Gen. At. Solid State Phys. (2018), https://doi.org/10.1016/j.physleta.2017.10.027.

[23] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, J.C. Nuño, From time series to complex networks: the visibility graph, Proc. Natl. Acad. Sci. USA (2008), https://doi.org/10.1073/pnas.0709247105.

[24] M. Newman, Networks: An Introduction, 2010, https://doi.org/10.1093/acprof:oso/9780199206650.001.0001.

[25] G. La Rosa, et al., Flash survey on SARS-CoV-2 variants in urban wastewater in Italy 1st report (investigation period: 04–12 July 2021), in: no. July, 2021, pp. 4–12, https://doi.org/10.1016/j.scitotenv.2020.139652.2.

[26] S. Milgram, An Experimental Study of the Small World Problem, Source Sociom, 1967.

[27] A. Endo, et al., Heavy-tailed sexual contact networks and monkeypox epidemiology in the global outbreak, 2022, Science (2022), https://doi.org/10.1126/science.add4507.

[28] I.M. Hall, J.R. Egan, I. Barrass, R. Gani, S. Leach, Comparison of smallpox outbreak control strategies using a spatial metapopulation model, Epidemiol. Infect. (2007), https://doi.org/10.1017/S0950268806007783.

[29] C. Viboud, O.N. Bjørnstad, D.L. Smith, L. Simonsen, M.A. Miller, B.T. Grenfell, Synchrony, waves, and spatial hierarchies in the spread of influenza, Science (2006), https://doi.org/10.1126/science.1125237.

[30] L. Danon, et al., Networks and the epidemiology of infectious disease, Interdiscip. Perspect. Infect. Dis. 1 (2011) 284909.

[31] R. Mogi, J. Spijker, The influence of social and economic ties to the spread of COVID-19 in Europe, J. Popul. Res. (2022), https://doi.org/10.1007/s12546-021-09257-1.

[32] M.A. Raifman, J.R. Raifman, Disparities in the population at risk of severe illness from COVID-19 by race/ethnicity and income, Am. J. Prev. Med. (2020), https://doi.org/10.1016/j.amepre.2020.04.003.

[33] M.P. Kwan, The stationarity bias in research on the environmental determinants of health, Health Place 70 (March) (2021) 102609, https://doi.org/10.1016/j.healthplace.2021.102609.

[34] J. Huang, M.P. Kwan, Z. Kan, M.S. Wong, C.Y.T. Kwok, X. Yu, Investigating the relationship between the built environment and relative risk of COVID-19 in Hong Kong, ISPRS Int. J. Geo-Inform. (2020), https://doi.org/10.3390/ijgi9110624.

[35] A. Wilder-Smith, D.O. Freedman, Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak, J. Travel Med. (2020), https://doi.org/10.1093/jtm/taaa020.

[36] M. Barthelemy, Spatial Networks, Oct. 2010, https://doi.org/10.1016/j.physrep.2010.11.002.

[37] N. Sharma, A.K. Verma, A.K. Gupta, Spatial network based model forecasting transmission and control of COVID-19, Phys. A Stat. Mech.Appl. (2021), https://doi.org/10.1016/j.physa.2021.126223.

[38] T. Phan, et al., A simple SEIR-V model to estimate COVID-19 prevalence and predict SARS-CoV-2 transmission using wastewater-based surveillance data, Sci. Total Environ. 857 (Pt 1) (2023) 159326, https://doi.org/10.1016/j.

[39] F. Wu, et al., SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases, mSystems 5 (4) (2020) e00614-20, https://doi.org/10.1128/msystems.

[40] A. Simone, C. Di Cristo, V. Guadagno, G. Del Giudice, Sewer networks monitoring through a topological backtracking, J. Environ. Manage. 346 (2023) 119015, https://doi.org/10.1016/j.jenvman.2023.