Two different approaches for monitoring planning in sewer networks: topological vs. deterministic optimization

Antonietta Simone [iD] [a,*], Alessandra Cesaro[b], Cristiana Di Cristo[b], Oreste Fecarotta[b] and Maria Cristina Morani[b]

[a] Department of Engineering and Geology, University 'G. D'Annunzio' of Chieti Pescara, Viale Pindaro 42, Pescara 65127, Italy
[b] Department of Civil, Building and Environmental Engineering (DICEA), University Federico II, Via Claudio 21, Naples 80125, Italy
*Corresponding author. E-mail: antonietta.simone@unich.it

[iD] AS, 0000-0001-8400-5201

ABSTRACT

Monitoring of sewer networks (SNs) is an important task whose planning can be related to various purposes, for example contaminant detection and epidemiological studies. This paper proposes two different approaches for the identification of a monitoring system in SNs. The first one proposes the identification of the best monitoring points starting from the knowledge of the hydraulic behavior of the system with respect to specific sensor threshold values through an optimization procedure that maximizes the reliability in detecting a contaminant. A new mathematical model is developed and a global optimization solver is employed to perform the optimization procedure. The second approach is based on the complex network theory (CNT) tools, adopting the in-relevance-based harmonic centrality, and does not require any hydraulic simulation. The metric is evaluated for each node of the network and provides a range of nodes, classified with respect to their importance, useful to identify suitable locations for sensors. With reference to both a benchmark and a real SN, the comparison between the results achieved by both strategies indicates that the two approaches provide comparable solutions in terms of sensor location.

Key words: complex network theory tools, deterministic optimization, harmonic centrality, monitoring, sensor location, sewer network

HIGHLIGHTS

- Comparison between two different approaches for monitoring planning in sewer networks.
- Use of a global optimization solver to find the optimal location of sensors maximizing the network reliability.
- Use of the in-based harmonic centrality as a tool to identify suitable locations for sensors referring to only topological information.
- Despite being based on different concepts, the two approaches provide comparable results.

GRAPHICAL ABSTRACT
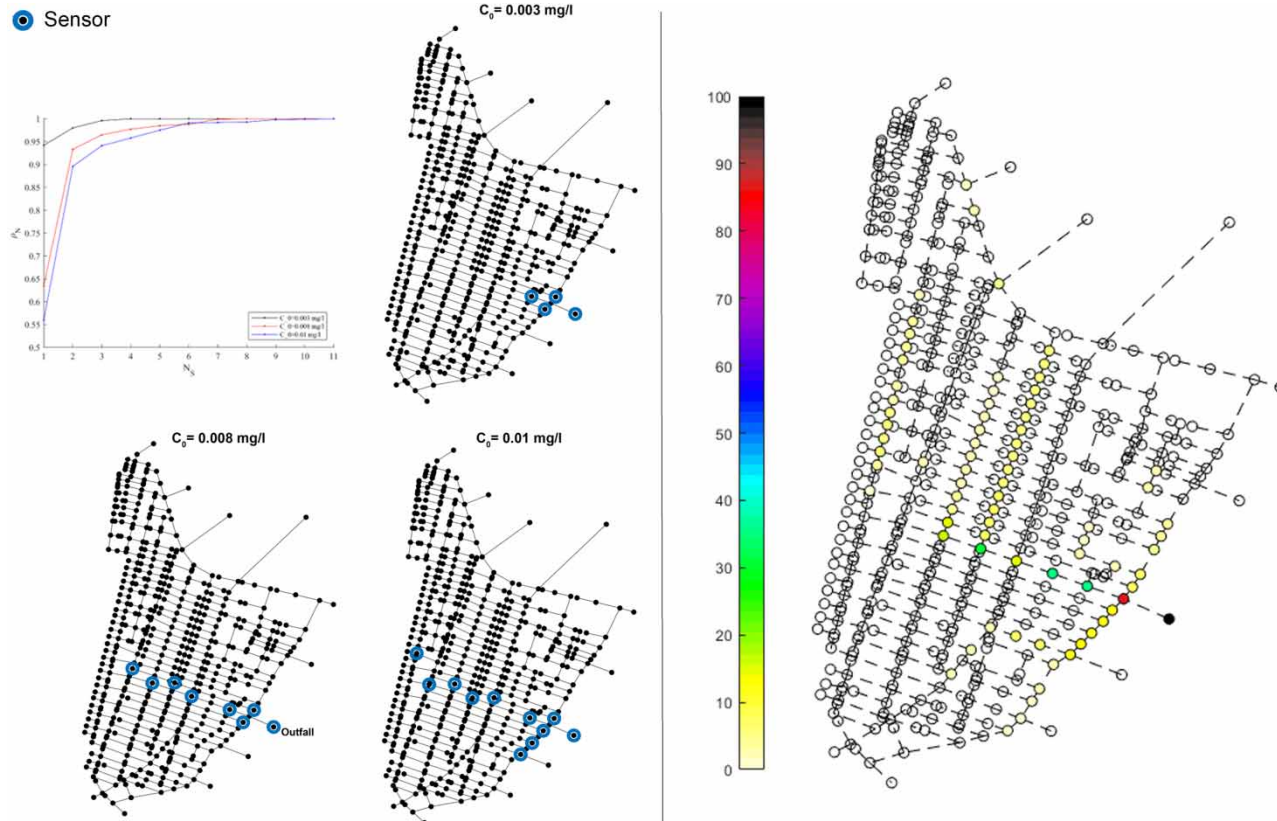
Deterministic approach *vs.* topological approach

1. INTRODUCTION

Sewer networks (SNs) are complex systems aimed at the collection and transport of wastewater (sanitary systems) or wastewater and rainwater (combined systems) up to the treatment plant (Simone *et al.* 2022b).

In recent decades, the need to propose new monitoring strategies for such complex systems has become increasingly urgent, mainly for factors related to the identification of illicit intrusions, as well as to the control of specific contaminants and to the limitation of the potential environmental impact (Gromaire *et al.* 2001). More recently, SNs monitoring has been proposed to support epidemiological studies (Larsen & Wigginton 2020).

The sensor location problem solved with optimization algorithms has been the focus of attention of many researchers, mainly relying on heuristic approaches. Banik *et al.* (2015a, 2015b) proposed a source identification methodology, based on a pre-screening aimed at reducing the network size, for identifying the set of candidate nodes for sensors. Banik *et al.* (2017a, 2017b) proposed optimization-based strategies aimed at the optimal location of sensors in sewer systems. Tinelli *et al.* (2017) investigated the optimal location of sensors with evolutionary algorithms, selecting the contamination events based on information about the network in terms of topology and operation. The problem of the optimal location of sensors was also addressed by Yazdi (2018), who proposed the use of a new procedure relying on evolutionary algorithms and entropy theory. Sambito *et al.* (2020) developed an approach for positioning water quality sensors based on the Bayesian decision network in order to facilitate the early isolation of illicit intrusions, focusing on soluble conservative pollutants, such as metals. The approach identifies the optimal sensor location gaining advantage from additional information, such as topology, thus reducing the computational effort needed to obtain the solution.

The monitoring of pollutants and pathogens in sewers has also been tackled with approaches based on backtracking algorithms, which represent a technique for evaluating the presence of contaminants and their diffusion process through a

backward process, from downstream to upstream. This process is aimed at supporting the identification of candidate positions to host monitoring measures. Rodríguez-Alarcón & Lozano (2019) proposed an approach based on complex network theory (CNT) tools and a backtracking algorithm to evaluate potential spill or contaminant release with respect to all nodes of the system. Simone et al. (2022a) proposed a strategy based on network topology and a backtracking algorithm to model the diffusion of pollutants along the Urban Drainage Network (UDN), by using only information derived from Horton's hierarchy. Chachula et al. (2021) proposed a backtracking strategy by using the time-series of sensor measurements in an array of localization of contamination events providing good information on contaminant events with few sensors. Guadagno et al. (2023) proposed a backtracking methodology, based on the calculation of the impact coefficient related to the dilution and decay of contaminants in SNs, to locate sensors with only one steady hydraulic simulation.

Most of the presented approaches are very effective but also very cumbersome, both for the hydraulic simulations they require and for the expensive calculation times. Furthermore, it could also happen, particularly with very large and complex systems, that the lack of information (e.g., flow) makes such analyses difficult or even unreliable. In this context, the CNT is proposed as a useful approach for the analysis of complex real systems, which is gaining momentum. With reference to SNs, various works have been proposed to evaluate the vulnerability, resilience, and operability of such systems (Reyes-Silva et al. 2020; Hesarkazzazi et al. 2022; Simone 2023) but also aspects related to monitoring. Simone et al. (2022b) proposed the use of CNT-tailored centrality metrics in the analysis of sewers for vulnerability/resilience assessment, optimal monitoring design and spread of contaminants. They highlighted the importance of topology in the study of sewer systems, showing how their behavior is the result of the interaction between the role of nodes and network topology. Zuluaga et al. (2020) used the network theory together with differential equations to model and simulate water quality parameters in a hydrological network. García-Usuga et al. (2020) used PageRank's centrality to identify well-monitored nodes in UDNs more susceptible to contamination. Halverson & Fleming (2015) proposed the use of CNT tools for systems of streamflow gauges. Their goal was to evaluate whether this approach could be meaningful when applied to hydrometric data, and, more specifically, whether it may help to guide decisions in stream gauge placement. Results showed that the betweenness metric is effective in identifying key points for sensors' placements, especially in bridges between communities.

Considering the potential of the CNT tools for the analysis of real systems and the performance of optimization algorithms, the aim of this paper is to model the sensor placement problem, i.e., determining the candidate positions to host monitor sensors for detecting the presence of contaminants or pathogens in the system, using both an optimization procedure and a complex network theory approach (Simone et al. 2023).

The obtained optimal solutions, with respect to a specific objective are compared with CNT results to evaluate the effectiveness of the topological approach. In fact, obtaining promising comparison would mean validating the topological approach for several preliminary and complementary applications relative to the analysis and management of SNs, while reducing the computation effort.

The first strategy is based on an optimization procedure aimed at searching for the best location of a fixed number of water quality sensors, in order to maximize the reliability of the whole system to detect a target substance. This strategy relies on the development of a mathematical model, which is solved using a global optimization solver. The proposed optimization procedure allows one to obtain the global optimum of the problem, depending on the accuracy of the mathematical model.

The second one uses a CNT centrality metric tailored accounting for both information on the connectivity structure and on the intrinsic relevance of nodes of the system (Giustolisi et al. 2020; Simone et al. 2020). The strategy considers the different roles of nodes (e.g., inlet nodes, connection nodes, and outfall nodes) embedding the information about their intrinsic relevance as inflows and the presence of spatial constraints (e.g., slope). The in-relevance-based harmonic centrality is computed for all nodes of the direct graph of the SN to support the monitoring system planning.

The main goal of the paper is to understand whether the simpler and less computationally expensive CNT-based approach furnishes comparable results in terms of sensor positioning with respect to the optimization solver.

It is important to highlight that, despite providing very promising results with short computational effort, the topological approach requires a deep knowledge of the used tools, in order to be able to implement the procedure in the most effective way and to guarantee a reliable interpretation of the results, as well as a full competence of the hydraulic systems in order to validate the consistency of the achieved results.

The paper is organized as follows. The next section reports a description of the two methodologies, the third section presents and compares the results of the analyses applied both to a benchmark SN and to a real one. Concluding remarks are drawn in the last section.

2. METHODS

Determining the suitability of a node for the detection of a contaminant in SNs is essential for planning an efficient monitoring system. This paper presents two modeling approaches for optimal sensor placement aimed at detecting target substances in SNs.

The first methodology proposes an optimization problem based on one single objective function, that is, the maximization of the network reliability when a spill of contaminant occurs within the system.

The second methodology investigates the sensor monitoring problem from a more topological perspective, using a metric proposed by the CNT and adapted to infrastructural systems like SNs.

2.1. The optimization methodology

The first strategy is deterministic optimization, relying on the analytical properties of the problem to generate a sequence of points converging to a global optimum or an approximately global optimum. Deterministic methods include linear programming (LP), non-linear programming (NLP), mixed-integer linear programming (MILP), and mixed-integer non-linear programming (MINLP) (Belotti *et al.* 2013; Morani *et al.* 2023). Most of the available MILP/MINLP solvers can achieve global optima only in convex problems (Belotti *et al.* 2009). The only solvers managing to find global optima in both convex and non-convex problems are the global optimization solvers.

In this study, the optimization procedure is performed by the SCIP (Solving Constraint Integer Programs) (Vigerske & Gleixner 2018) solver, which is a global optimization solver implementing a spatial branch and bound and various heuristics.

Given a network consisting of L links and N nodes, the aim of the optimization procedure is to find the best location of a fixed number of sensors to maximize the network reliability (R_N) in case of pollutant introduction. In this study, the optimization procedure has been decoupled by the hydraulic and quality modelling of the network, thus flow velocity (V) and pollutant concentration (C) are preliminarily computed by means of the external software SWMM (Rossman 2017).

The network can be modelled as a directed graph so that each link l ($l = 1, \dots, L$) has a proper direction and the discharge flows from the initial node (i) to the final node (f). In this study, each node of the network is a possible candidate for the installation of a sensor, and it has been modelled by means of a binary variable (I_n), which is equal to 1 if the device is installed, and 0 otherwise. According to the proposed procedure, the sensors should be located in order to maximize the reliability of the network, which means maximizing the number of detected points when a pollutant is inserted within the network. In this study, the pollutant entrance is assumed in only one node at a time and the network reliability is assessed over several scenarios, differing by the node assumed as an inlet point.

A further variable of the optimization is represented by the pollutant detection ($\delta_{i,s}$), which is defined for each node i ($i = 1, \dots, N$) and for each scenario s ($s = 1, \dots \sigma$). Such a binary variable has been properly modelled by means of a set of constraints, written for each link l ($l = 1, \dots, L$) of the network:

$$\delta_{i,s} = I_i B_i^s + I_f B_f^s + (1 - I_f)(1 - I_i)(\delta_{f,s} B_i^s) \tag{1}$$

where B_i^s is a parameter equal to 1 if the concentration at the ith node ($C_{i,s}$) is greater than a threshold concentration (C_0), 0 otherwise. The parameter B_i^s results from the concentration values obtained by the quality simulation performed by the software SWMM.

It is worth underlining that the outfall (ω) is not constrained by Equation (1), since it cannot be the initial node of any links. In order to properly define the pollutant detection at the outfall node, the following constraint has been defined:

$$\delta_{\omega,s} = I_\omega B_\omega^s \tag{2}$$

According to Equation (2), the pollutant at the outfall is detected only in case of device installation (i.e., I_ω equal to 1) and concentration value greater than the threshold value (i.e., B_ω^s equal to 1).

Finally, a further constraint has been considered in order to set the total number of sensors (N_s) installed within the network:

$$\sum_{i=1}^{n} I_n = N_s \tag{3}$$

As an objective function of the optimization, the system reliability has been considered and defined as follows:

$$maximize \left(\rho_N = \frac{\sum_{s=1}^{\sigma} \sum_{i=1}^{N} \delta_{i,s}}{\sum_{s=1}^{\sigma} \sum_{i=1}^{N} B_i^s} \right) \tag{4}$$

2.2. In-relevance-harmonic based methodology

Centrality is one of the most studied concepts in CNT. Several centrality metrics have been proposed (Freeman 1978; Borgatti 2005; Newman 2010) to evaluate the most central element in real systems with respect to both different physical phenomena and the way information flows in the network. The proposed approach recalls the concept of harmonic centrality and investigates the possibility of using this metric to efficiently plan a quality monitoring system in SNs. Its application is justified because the harmonic centrality assumes that information through the network moves only along the shortest possible paths, just as happens in SNs, for which the shortest paths between the various pairs of nodes are uniquely determined by the slope of the system.

The harmonic centrality represents a metric that detects the ability of each node to disseminate and receive information in the system based on the concept of proximity. The original formulation of the harmonic centrality, H_i^C, is mathematically expressed as (Rochat 2009):

$$H_i^C = \sum_{j=1}^{N} \frac{1}{d_{ij}} \tag{5}$$

where d_{ij} is the distance from node i to node j in the network.

The standard metric is extended to embed the information about the intrinsic relevance of the nodes (outfall, connection nodes, etc.) through the function $f(R_i, R_j)$, obtaining the relevance-harmonic centrality (Giustolisi *et al.* 2020) mathematically expressed as:

$$H_i^C = R_i \sum_{j=1}^{N} \frac{1}{d_{ij}} \tag{6}$$

The intrinsic relevance of the ending nodes, R_i and R_j, of each link, is a piece of information depending on the type of network and analysis to perform.

Several functions $f(R_i, R_j)$ have been proposed to account for the intrinsic relevance of the analysis:

$$\begin{cases} f(R_i, R_j) = R_i \cdot R_j \\ f(R_i, R_j) = \dfrac{(R_i + R_j)}{2} \\ f(R_i, R_j) = R_i \\ f(R_i, R_j) = \max[R_i, R_j] \\ f(R_i, R_j) = \Sigma Rpath \\ f(R_i, R_j) = \Pi Rpath \end{cases} \tag{7}$$

The strategy here proposed is performed using the function $f(R_i, R_j) = (R_i \cdot R_j)$, corresponding to the product of the intrinsic relevance of the ending nodes i and j of each pipe l. It is considered as the one that best highlights the role of intrinsic relevance in the connections between nodes for the specific considered problem. The intrinsic relevance of nodes is assumed equal to the inflow for each node, proportional to the quantity of contaminant introduced.

It is important to note that, since the SNs are direct networks, it is necessary to use a specific version of the harmonic centrality, i.e., the in-harmonic (Simone *et al.* 2023), which, in the specific case, corresponds to the one that collects the information entering the nodes.

Therefore, the in-relevance harmonic centrality is evaluated for each node of the systems considered and the analysis provided a node importance ranking, normalized in the range [0, 100]. Nodes with higher values of the metric represent the points where most of the information disseminated on the network is concentrated and are, most likely, the best candidate to host sensors/sampling points.

3. RESULTS AND DISCUSSION

3.1. Benchmark SN

The benchmark SN reported in Figure 1 (Simone *et al.* 2023) is considered to perform the two approaches. The network model is composed of 77 nodes (manholes), 79 links (sewer pipes), and 1 outfall. The flow directions are imposed by the slope of the pipes.

3.1.1. Application of the optimization methodology

For the analysed case study, each node of the network has been assumed as the possible inlet of the pollutant. Therefore, the total number of analysed scenarios (σ) is equal to the total number of nodes (N). Regarding the inserted pollutant, the inflow concentration has been assumed equal to 1 mg/l. The pollutant has been considered to decay with a first-order kinetic, with
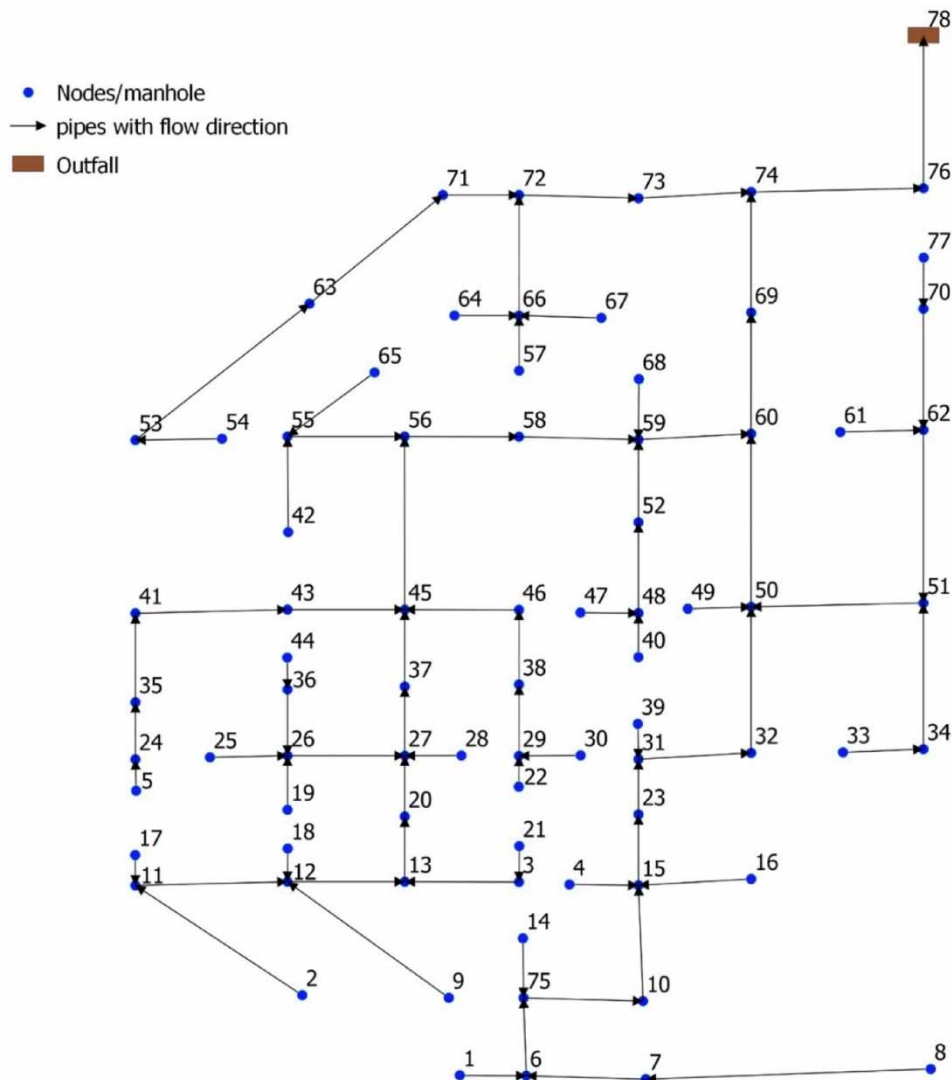

Figure 1 | Benchmark SN layout with flow direction.

the decay coefficient set as $0.2\,\text{h}^{-1}$, consistent with the study of Hart & Halden (2020) targeting SARS-COVID-19 as a contaminant.

Once a threshold concentration value (C_0) and a number of installed sensors (N_s) are fixed, the optimization process is accomplished by the solver SCIP in less than 1 s. The achievement of the global optimum is extremely fast since SCIP is a very high-performance solver for MILP problems.

It is worth underlining that, by increasing the number of installed sensors, the value of the objective function (i.e., the network reliability) progressively increases up to the maximum value (i.e., 1). In particular, Figure 2 reports the value of the objective as a function of the number of sensors for the case $C_0 = 0.008$ mg/l, showing that four sensors are sufficient for reaching the maximum value.

Figure 3 reports the selected optimal positions for the four sensors. It is possible to note that one of the sensors is installed in correspondence with the outfall, as foreseeable, and the other three are located in points that receive important flow rates within the network. Moving upstream, i.e., where the dilution effect is mild due to the smaller flows, the installation of additional sensors is worthless, since the coverage of this part of the network in terms of pollutant detection is already ensured by the downstream sensors.

3.1.2. In-relevance-harmonic-based methodology

The in-relevance-harmonic centrality is applied to the case study to understand whether it can provide useful information about the sensor location. The intrinsic relevance is set equal to the lateral inflow for each node, which is proportional to the concentration of contaminant introduced as mg/l. The obtained in-relevance-harmonic values are graphically shown in Figure 4. The maximum value of the metric occurs for node 78, i.e., the system outfall. This result indicates that the outfall is the most suitable node to detect the information of other nodes and, thus, it can be considered a good candidate for sensor placement, consistently with the well-known knowledge of SN functioning. Increasing the number of sensors to be installed in the network, it is possible to refer to the metric values and select the nodes immediately less central than the outfall. The
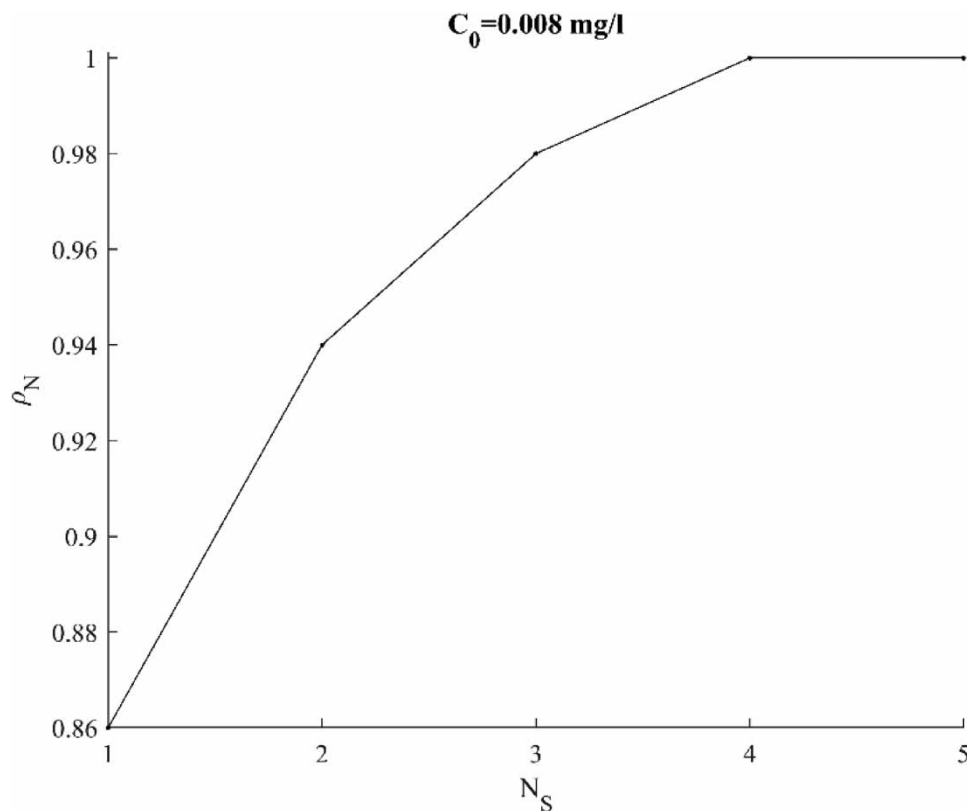

Figure 2 | Objective function variability depending on the number of installed sensors for $C_0 = 0.008$ mg/l.
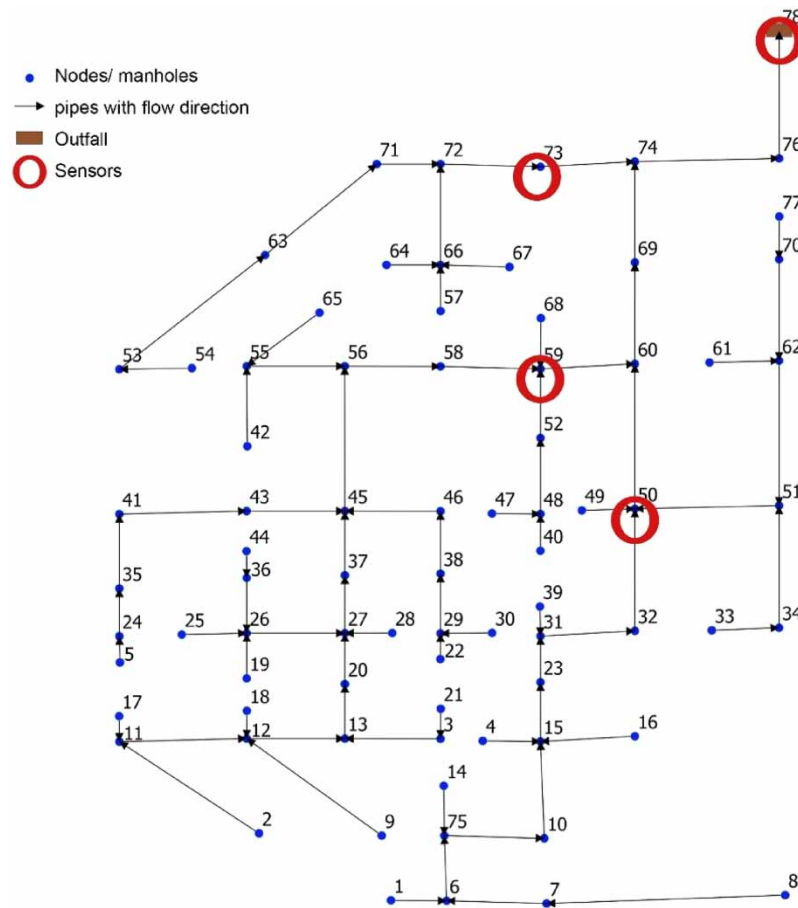
Figure 3 | Network layout with indication of the optimal positions for the sensors according to the optimization procedure.

results indicate nodes 76 and 59 as good information receptors. It is evident that node 76 is an alternative to node 78 since they are connected to each other. The analysis suggests that node 76 receives a lot of information about the system, being close to the outfall, and therefore represents an excellent position to host a sensor, but at the same time its proximity to the outfall makes it an alternative, to be considered, perhaps, as a sensor backup and for step-testing during monitoring campaigns.

Therefore, the analysis suggests placing two sensors in nodes 78 and 59. As the number of sensors to be positioned increases, the ranking of metric values allows for immediately identifying the most suitable position to host them. According to such an interpretation of the metric, nodes 50 and 27 follow as candidate nodes for hosting sensors.

The analysis provides null values of the metric for all extremal/ head nodes because no information arrives at these points. This result implies that these points are not suitable candidates for hosting measurements.

The results of the analysis show that the in-relevance-harmonic centrality is effective in evaluating the ability of the nodes to receive contaminants and therefore it could be effective in the study related to the spread of contaminants and the planning of monitoring systems.

3.1.3. Optimization procedure vs. in-relevance-harmonic centrality

Comparing the results of the proposed methodologies, the nodes identified through the in-relevance-harmonic centrality (Figure 4) are very close to the ones individuated by the optimization method (Figure 3), by considering a threshold value of $C_0 = 0.008$ mg/l. Both approaches indicate, indeed, that nodes 78, 59 and 50 as the best candidates for either the positioning of sensors or the collection of samples to detect target pollutants, showing that, despite being based on different concepts and approaches, the two strategies provided quite comparable results.
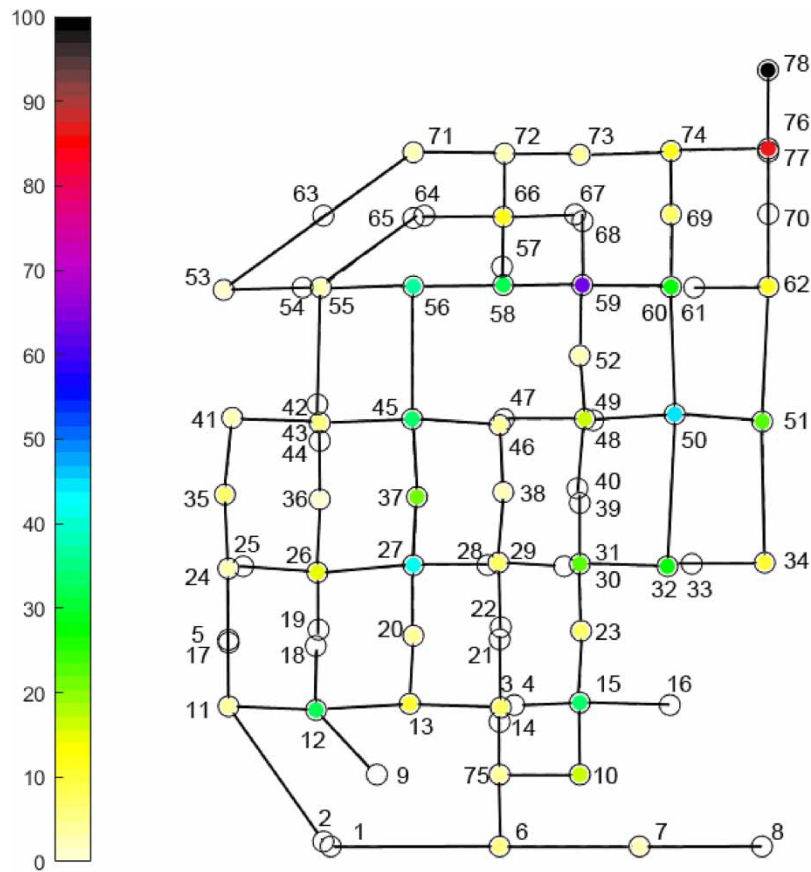
Figure 4 | Network layout with the in-relevance-harmonic centrality values for all nodes.

Although this result confirms the validity of the topological approach in addressing the study of SNs, it is important to underline that such analysis requires technicians who are able to implement and adequately interpret the CNT tools with respect to the SNs operation, as well as to evaluate the feasibility of the results with respect to their hydraulic behavior.

3.2. Real SN

A further comparison between the two approaches is performed with a larger and real SN, proposed in Hesarkazzazi *et al.* (2022), and redrafted to consider dry weather conditions setting the lateral inflows and excluding the sub-catchments. The network model is composed of 531 nodes, 530 pipes, and 1 outfall. The flow directions are imposed by the slope of the pipes (Figure 5).

3.2.1. Application of the optimization methodology

As for the benchmark network, each node of this real network has been assumed as a possible inlet of the pollutant, with an input concentration equal to 1 mg/l. The pollutant has been considered to decay with a first-order kinetic with a decay coefficient set as 0.2 h^{-1}. Despite its larger size compared to the benchmark network, once the threshold concentration value (C_0) and the number of installed sensors (N_s) are fixed, the optimization process is performed in a fraction of a second.

Figure 6 shows the trend of the objective function with the number of sensors for three different values of threshold concentration. According to this figure, given a value of threshold concentration, the network reliability increases progressively up to the maximum value, i.e., 1. By increasing the threshold concentration, the total number of sensors ensuring the total coverage of the network (i.e., $\rho_N = 1$) increases accordingly. Indeed, given a threshold concentration equal to 0.003 mg/l, the coverage of the network is ensured by installing 4 sensors. Increasing the threshold concentration to 0.008 and 0.01 mg/l, the total number of sensors increases to 8 and 11, respectively.
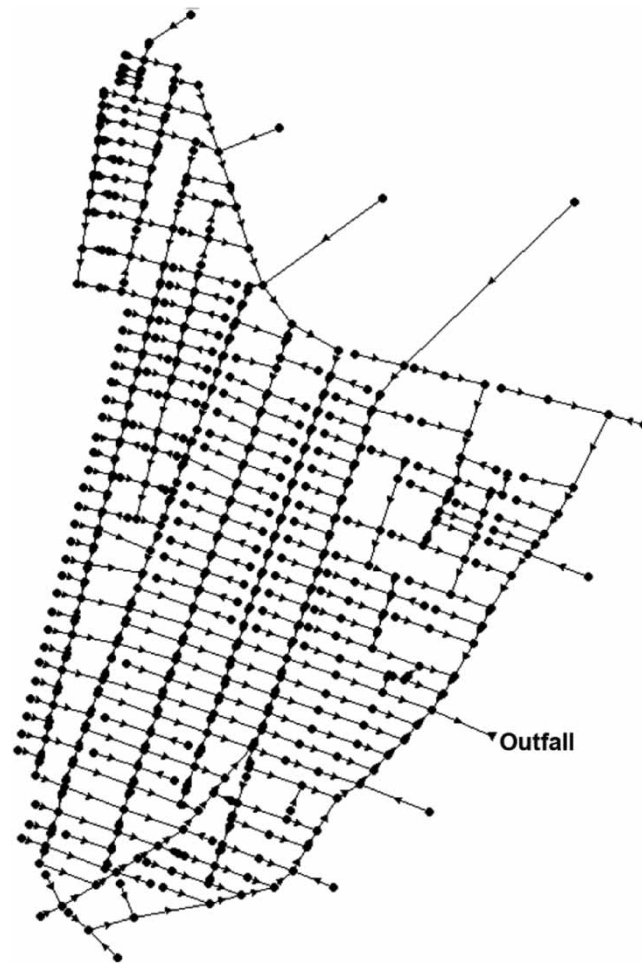
Figure 5 | Real network layout with flow direction.

Moreover, with reference to the threshold concentration equal to 0.008 mg/l, Figure 7 reports the selected optimal positions for the required eight sensors.

3.2.2. In-relevance-harmonic-based methodology

The in-relevance-harmonic centrality is applied to the real case study and the results are graphically shown in Figure 8. As for the previous case study, the intrinsic relevance for each node is set equal to the lateral inflow, which is proportional to the concentration of contaminant introduced as mg/l. The collection capacity varies from node to node, and for all head nodes is null, because they spread the contaminant in the network, but they do not collect it.

The maximum value of the metric is at the outfall, i.e., node 531, which makes it, as expected, the most suitable node to host a sensor. The ranking of Figure 8 is indicative of the ability of each node to collect information derived from other nodes in the system. In particular, it is possible to note that the metric identifies several main paths in the network, characterized by the presence of nodes with high intrinsic relevance (yellow nodes), corresponding to endpoints of separate branches, and in any case, connected to other well-connected nodes. Obviously, each node can receive information only from the nodes adjacent to it, and the ability to collect information from some nodes is greater than that of the others due to their topological relevance. Most of the nodes with high values of the metric are located near the only outfall, along the main path directed towards this point. A second important path intersects the main one close to the outfall (in the red node), while another path of medium importance is located in the northwest area of the network. The remaining part of the nodes takes on very low or even zero metric values.
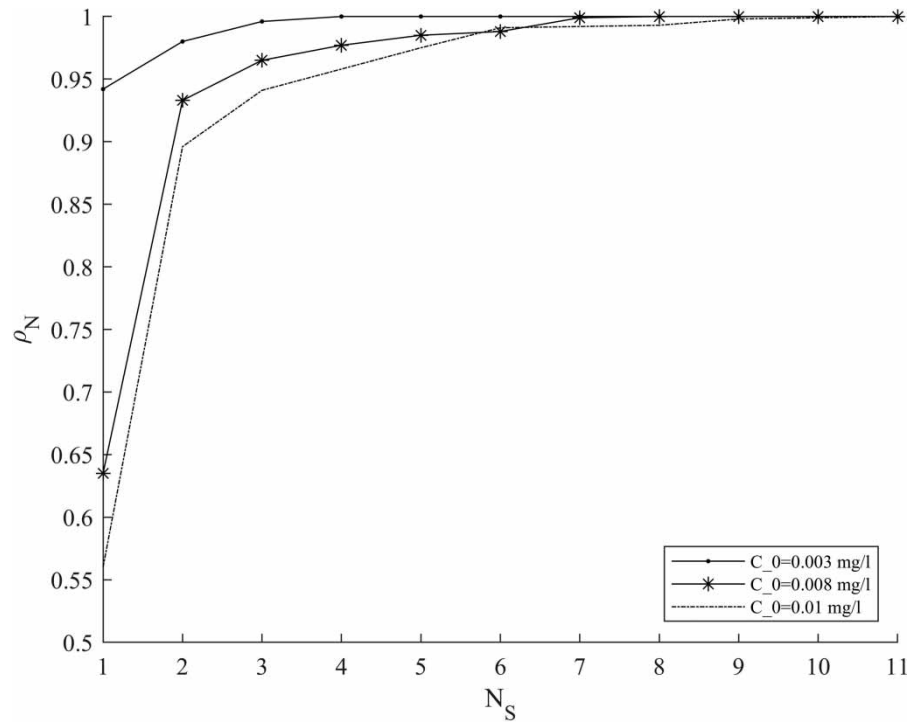
Figure 6 | Objective function variability depending on the number of installed sensors for three different values of threshold concentration.
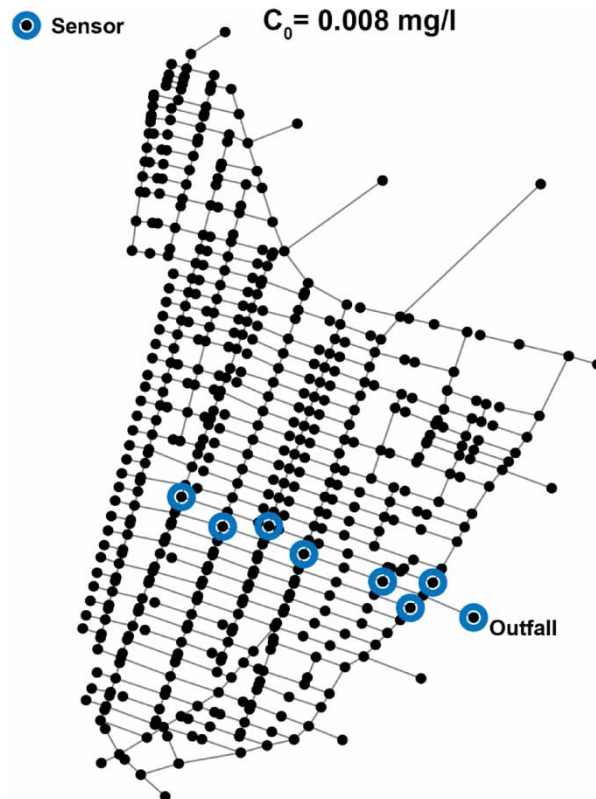

Figure 7 | Position of installed sensors for $C_0 = 0.008$ mg/l using the optimization procedure.
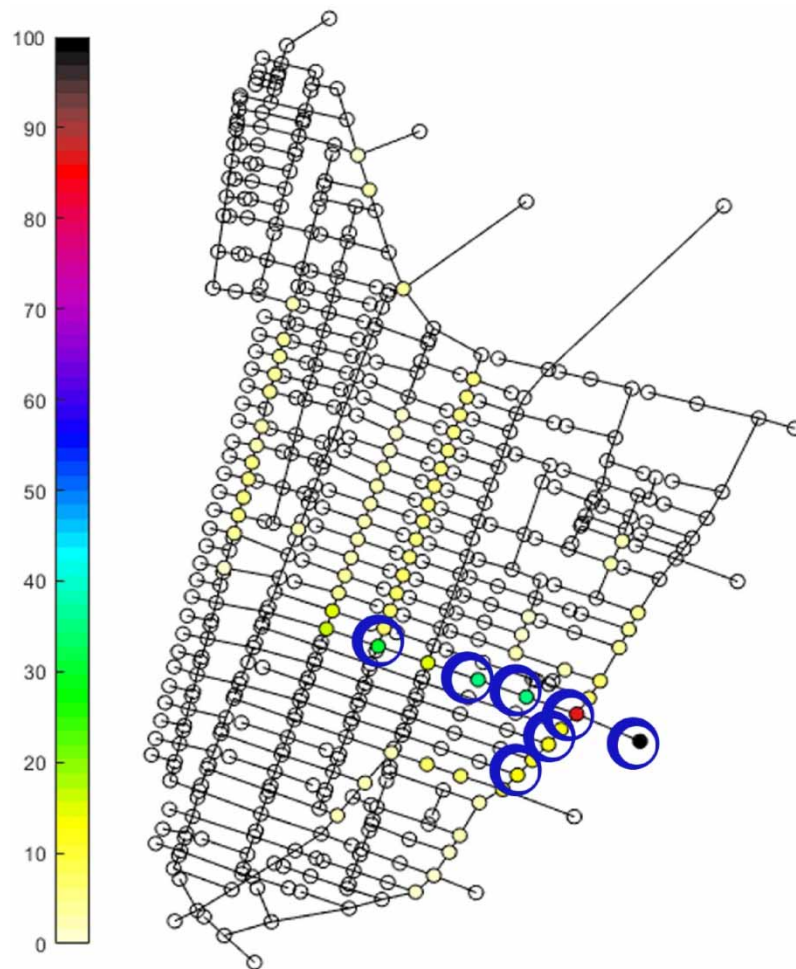
Figure 8 | In-relevance-harmonic centrality values for all nodes of the real case study network.

The selection of the nodes to be indicated as candidate positions for the installation of sensors should consider metric values in the range [20, 100]. This criterion leads to the identification of seven nodes as candidate positions for sensors, obviously belonging to the main routes mentioned before.

Overall, such a metric trend allows for tracing the detection of contaminants in the network and highlighting the points that are most characterized by this process.

3.2.3. Optimization procedure vs. in-relevance-harmonic centrality

The comparison of the results obtained by the in-relevance-harmonic centrality and the optimization procedure performed assuming a threshold concentration equal to 0.008 mg/l highlighted that the two procedures provide comparable results since the points selected for either sensor installation or sampling are very close. In fact, the nodes contained in the range chosen for the topological metric enclose the ones provided by the optimization or indicate adjacent positions. In both cases, the candidate nodes to host the sensors are located along the same path.

However, it is worth noting that the solution achieved by the optimization procedure varies depending on the threshold concentration. Indeed, assuming worse quality sensors (i.e., increasing the threshold concentration), the number of installed sensors ensuring the full coverage of the network increases accordingly. The solution provided by the topological metric, instead, remains unchanged, being only based on the network topology.

In order to evaluate whether the topological approach is still efficient even considering different sensors, i.e., different threshold values, Figure 9 shows the optimization solutions obtained by using two different threshold concentrations, i.e., assuming smaller and larger values than the one considered in Figure 7, set as 0.003 mg/l (a) and 0.01 mg/l (b), respectively.
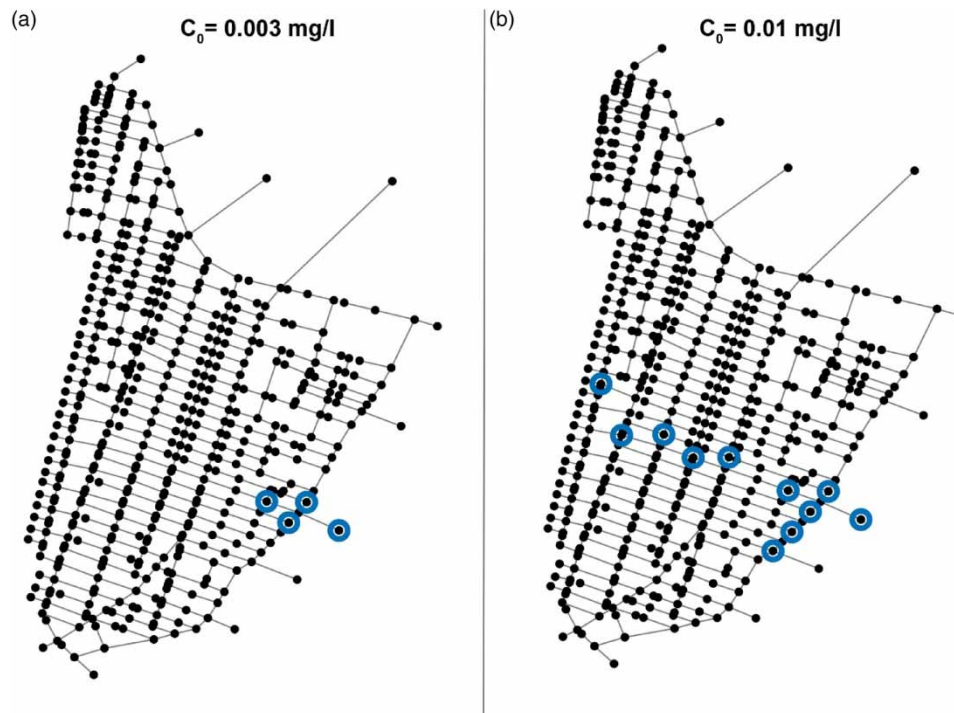
Figure 9 | Sensor installation according to the optimization procedures performed for threshold concentration equal to 0.003 mg/l (a) and 0.01 mg/l (b).

Results highlight that the nodes selected by the optimization solver still lie within the range of nodes with high metric values. Particularly, considering the optimization performed using a threshold value equal to 0.003 mg/l (Figure 9(a)), the analysis suggests installing only four sensors, specifically in the neighborhood where the topological approach provides higher values of the metric (Figure 8), which corresponds to the first sensors selected in the case $C_0 = 0.008$ mg/l (Figure 7). Again, considering the optimization performed using a threshold value equal to 0.01 mg/l, the suggested locations of the additional sensors are reported in Figure 9(b). The position of the new devices defines a further path of monitoring in the network, just as occurs for the topological metric (Figure 8). They still include the eight locations selected in the test $C_0 = 0.008$ mg/l (Figure 7), demonstrating good robustness of the optimal solution with respect to the threshold value.

Overall, it is evident that the analysis with the topological metric cannot be as exhaustive and complete as the results provided by the optimization for the various threshold values, and this is not the objective. However, the single solution provided by the topological metric encloses important information that allows us to proceed progressively to an efficient position of sensors in the network without a more complex optimization method application. This result is surprising considering that the used metric is applicable from the first stages of the study, requiring only data relating to the system topology without performing any simulation.

As already explained, the topological approach provides very promising results using only information on the system topology. However, it is worth underlining the importance of knowledge requirements for technicians implementing the topological approach, who need to have full competence in the hydraulic operation of the SNs, in order to best implement the procedure and evaluate the reliability of the results coherently with the behavior of the system. Conversely, the deterministic approach allows for obtaining the global optimum of the problem once a mathematical model is developed and the hydraulic behavior of the system is known. The assessment of the found optimal in terms of quality cannot be made when the optimization is performed by means of heuristic procedures, in which the relation between input parameters and output variables is not conclusively determined. However, with reference to deterministic optimization, the achieved optimum is global with respect to the search space associated with the developed mathematical model, i.e., the effectiveness of the optimum depends on the accuracy of the mathematical model and its efficiency in interpreting the behavior of the considered system.

Finally, several advantages may result from coupling the proposed approaches in one single procedure, where the topological approach first provides the user with a preliminary selection of the most suitable nodes for sensor location, among which the deterministic optimization searches for the best solution based on both hydraulic and mathematical constraints. Indeed, in large-size problems (i.e., large SNs), the implementation of a deterministic optimization may require high computational effort, thus the integration of both approaches in a two-step procedure may allow for a faster convergence of the solver without affecting the quality of the found solution.

4. CONCLUSIONS

The present paper addresses the problem of the optimal location of sensors within SNs based on two different approaches. The first approach relies on deterministic optimization aiming at finding the best location and number of sensors to maximize the reliability of the network for the detection of a target substance. The second approach uses a relevance-based CNT centrality metric, which does not require any simulation, but only the knowledge of the topological scheme and the flow inputs. In particular, the in-relevance-harmonic centrality, which evaluates the ability of each node to disseminate and receive information, is used to sort the nodes of the network based on their importance. The two approaches were first tested on a benchmark network, and then a real SN was assumed as a case study. According to the results, the two approaches provide comparable solutions in terms of sensor location within the networks. Indeed, the nodes selected by the optimization procedure lie within the range of the nodes with the highest values of the topological metric.

Considering the promising results achieved by the topological approach, the proposed metric can be used as a very efficient complementary tool for the design of SN monitoring systems in complex schemes.

Indeed, when large-size networks are considered, the optimization procedure can be significantly demanding and a preliminary selection of the most suitable nodes for sensor location can be crucial to tackling the computational complexities affecting the problem. Therefore, an integrated procedure could be investigated in future studies, based on the use of the CNT centrality to reduce the research space and speed up the optimization procedure.

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Banik, B. K., Alfonso, L., Torres, A. S., Mynett, A., Di Cristo, C. & Leopardi, A. 2015a Optimal placement of water quality monitoring stations in sewer systems: an information theory approach. *Procedia Engineering*. https://doi.org/10.1016/j.proeng.2015.08.956.

Banik, B. K., Di Cristo, C. & Leopardi, A. 2015b A pre-screening procedure for pollution source identification in sewer systems. *Procedia Engineering*. https://doi.org/10.1016/j.proeng.2015.08.896.

Banik, B. K., Alfonso, L., Di Cristo, C. & Leopardi, A. 2017a Greedy algorithms for sensor location in sewer systems. *Water (Switzerland)*. https://doi.org/10.3390/w9110856.

Banik, B. K., Alfonso, L., Di Cristo, C., Leopardi, A. & Mynett, A. 2017b Evaluation of different formulations to optimally locate sensors in sewer systems. *Journal of Water Resources Planning and Management* **143** (7), 1–12. https://doi.org/10.1061/(asce)wr.1943-5452.0000778.

Belotti, P., Lee, J., Liberti, L., Margot, F. & Wächter, A. 2009 Branching and bounds tighteningtechniques for non-convex MINLP. *Optimization Methods and Software*. https://doi.org/10.1080/10556780903087124.

Belotti, P., Kirches, C., Leyffer, S., Linderoth, J., Luedtke, J. & Mahajan, A. 2013 Mixed-integer nonlinear optimization. *Acta Numerica* **22**, 1–131. https://doi.org/10.1017/S0962492913000032.

Borgatti, S. P. 2005 Centrality and network flow. *Social Networks*.https://doi.org/10.1016/j.socnet.2004.11.008.

Chachula, K., Nowak, R. & Solano, F. 2021 Pollution source localization in wastewater networks. *Sensors (Switzerland)* **21** (3), 1–16. https://doi.org/10.3390/s21030826.

Freeman, L. C. 1978 Centrality in social networks conceptual clarification. *Social Networks*. https://doi.org/10.1016/0378-8733(78)90021-7.

García-Usuga, J. M., Olivar-Tost, G. & Mesa-Mazo, M. J. 2020 Application of the theory of networks to model a drainage network of a watershed: case study department of quindío colombia. *Hidrobiologica*. https://doi.org/10.24275/UAM/IZT/DCBS/HIDRO/2020V30N2/GARCIA.

Giustolisi, O., Ridolfi, L. & Simone, A. 2020 Embedding the intrinsic relevance of vertices in network analysis: the case of centrality metrics. *Scientific Reports* **10** (1). https://doi.org/10.1038/s41598-020-60151-x.

Gromaire, M. C., Garnaud, S., Saad, M. & Chebbo, G. 2001 Contribution of different sources to the pollution of wet weather flows in combined sewers. *Water Research*. https://doi.org/10.1016/S0043-1354(00)00261-X.

Guadagno, V., Del Giudice, G., Di Cristo, C., Leopardi, A. & Simone, A. 2023 Impact coefficient evaluation for sensor location in sewer systems. *Journal of Water Resources Planning and Management*. ASCE. doi:10.1061/JWRMD5/WRENG-6093.

Halverson, M. J. & Fleming, S. W. 2015 Complex network theory, streamflow, and hydrometric monitoring system design. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-19-3301-2015.

Hart, O. E. & Halden, R. U. 2020 Modeling wastewater temperature and attenuation of sewage-borne biomarkers globally. *Water Research*. https://doi.org/10.1016/j.watres.2020.115473.

Hesarkazzazi, S., Bakhshipour, A. E., Hajibabaei, M., Dittmer, U., Haghighi, A. & Sitzenfrei, R. 2022 Battle of centralized and decentralized urban stormwater networks: from redundancy perspective. *Water Research* **222** (July), 118910. https://doi.org/10.1016/j.watres.2022.118910.

Larsen, D. A. & Wigginton, K. R. 2020 Tracking COVID-19 with wastewater. In: *Nature Biotechnology*. https://doi.org/10.1038/s41587-020-0690-1.

Morani, M. C., Crespo Chacón, M., Garcia Morillo, J., McNabola, A. & Fecarotta, O. 2023 Exploring the optimal location of pumps as turbines within branched irrigation networks by global optimization. *Water Resources Research* **59** (3). https://doi.org/10.1029/2022wr033317.

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

Reyes-Silva, J. D., Zischg, J., Klinkhamer, C., Rao, P. S. C., Sitzenfrei, R. & Krebs, P. 2020 Centrality and shortest path length measures for the functional analysis of urban drainage networks. *Applied Network Science*. https://doi.org/10.1007/s41109-019-0247-8.

Rochat, Y. 2009 Closeness Centrality Extended To Unconnected Graphs : The Harmonic Centrality Index. *Application of Social Network Analysis (ASNA)*.

Rodríguez-Alarcón, R. & Lozano, S. 2019 A complex network analysis of Spanish river basins. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2019.124065.

Rossman, L. A. 2017 *Storm Water Management Model Reference Manual Volume II — Hydraulics*. EPA/600/R-17/111. USEPA, Cincinnati, OH.

Sambito, M., Di Cristo, C., Freni, G. & Leopardi, A. 2020 Optimal water quality sensor positioning in urban drainage systems for illicit intrusion identification. *Journal of Hydroinformatics*. https://doi.org/10.2166/hydro.2019.036.

Simone, A. 2023 Vulnerability assessment of urban drainage network using relevance-based centrality metrics. *River*, December 2022, 39–51. https://doi.org/10.1002/rvr2.30.

Simone, A., Ciliberti, F. G., Laucelli, D. B., Berardi, L. & Giustolisi, O. 2020 Edge betweenness for water distribution networks domain analysis. *Journal of Hydroinformatics*. https://doi.org/10.2166/hydro.2019.030.

Simone, A., Cesaro, A., Del Giudice, G., Di Cristo, C., Esposito, G. & Fecarotta, O. 2022a *The Spread of Contaminants in Urban Drainage Networks Based on a Topological Analysis*. 20. https://doi.org/10.3390/environsciproc2022021020.

Simone, A., Cesaro, A., Del Giudice, G., Di Cristo, C. & Fecarotta, O. 2022b Potentialities of complex network theory tools for urban drainage networks analysis. *Water Resources Research* **58** (8). https://doi.org/10.1029/2022WR032277.

Simone, A., Cesaro, A., Di Cristo, C., Fecarotta, O. & Morani, M. C. 2023 Monitoring planning for urban drainage networks. *IOP Conference Series: Earth and Environmental Science* **1136** (1). https://doi.org/10.1088/1755-1315/1136/1/012008.

Tinelli, S., Creaco, E. & Ciaponi, C. 2017 Sampling significant contamination events for optimal sensor placement in water distribution systems. *Journal of Water Resources Planning and Management*. https://doi.org/10.1061/(asce)wr.1943-5452.0000814.

Vigerske, S. & Gleixner, A. 2018 SCIP: global optimization of mixed-integer nonlinear programs in a branch-and-cut framework. *Optimization Methods and Software*. https://doi.org/10.1080/10556788.2017.1335312.

Yazdi, J. 2018 Water quality monitoring network design for urban drainage systems, an entropy method. *Urban Water Journal*. https://doi.org/10.1080/1573062X.2018.1424215.

Zuluaga, V., García, J. M., Mesa, M. J. & Acosta-Minoli, C. A. 2020 Use of network theory to model water quality parameters in a hydrological network. *Journal of Physics: Conference Series*. https://doi.org/10.1088/1742-6596/1448/1/012006.