IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Centrality-Guided Node Selection for Efficient Water Leak Detection

DANIELE UGO LEONZIO[1], (Member, IEEE), TIZIANA CATTAI[2], (Member, IEEE) PAOLO BESTAGINI[1], (Member, IEEE) STEFANIA COLONNESE[2], and STEFANO TUBARO [1], (Senior Member, IEEE)

[1]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy (e-mail: name.surname@polimi.it)
[2]Dipartimento di Ingegneria dell'Informazione, Elettronica e Telecomunicazioni, Sapienza Università di Roma, Rome, Italy

Corresponding author: Daniele Ugo Leonzio (e-mail: danieleugo.leonzio@polimi.it).

"This work was partially supported by the European Union - Next Generation EU under the Italian National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.3, CUP B53C22004050001, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART."

ABSTRACT Water Distribution Networks (WDNs) are critical infrastructures that require continuous monitoring to ensure efficient operation and mitigate water loss caused by leaks. The strategic placement of sensors within these networks is essential for timely leak detection and localization. Selecting the most effective nodes for sensor placement, however, remains a challenging problem. Centrality measures, widely used in network science to quantify the node importance, offer a promising approach to guide the sensor deployment. This study investigates the impact of various centrality measures on node selection for leak detection in WDNs. We address the leak-detection problem using a two-stage methodology. First, we derive compact and informative feature representations from pressure matrices using the encoder of an autoencoder. Then, we apply a binary classifier to identify anomalous embeddings. We apply our method to two topologically different WDNs, exploring different centrality measures for sparse sensor deployment. Results show that in looped, high-redundancy topologies, a reduced set of strategically placed sensors maintain or even improve detection accuracy. In contrast, more fragile tree-like networks experience performance loss when sensors are reduced. To explain this structural behavior, we introduce a metric that quantifies a network's structural vulnerability. Taken together, our findings offer practical guidance for water utilities seeking to balance leak-detection speed, accuracy, and cost through topology-based sensor placement.

INDEX TERMS Water distribution networks, graph modeling, leak detection, centrality metrics.

I. INTRODUCTION

Water Distribution Networks (WDNs) are complex, large-scale infrastructures designed to transport potable water from treatment facilities to end consumers through an interconnected system of pipes, pumps, valves, and reservoirs. Ensuring the structural integrity and hydraulic performance of WDNs is of paramount importance, as any failure or degradation, particularly in the form of leaks, can lead to significant water losses, elevated operational costs, deterioration of service quality, and even public health risks when water quality is compromised by pressure drops [1]. Estimates suggest that in many urban systems worldwide, Non Revenue Water (NRW) losses due to leaks and undetected bursts can account for 20–50% of the total volume supplied, thus imposing a substantial burden on water utilities and jeopardizing sustainability goals [2].

Leak detection in WDNs has evolved considerably over the past decades. Traditional approaches, such as acoustic monitoring [3], [4], transient analysis [5], and physical inspections, often require either a high density of specialized sensors or manual surveying of pipelines, methods that can be both time-consuming and cost-prohibitive in large or aging networks [6]. Furthermore, these methods typically focus on localizing individual leaks once they become sufficiently large to generate detectable signals, rather than enabling proactive, network-wide surveillance. As infrastructure ages and the frequency of small-to-medium leaks increases, there is a pressing need for more systematic, scalable, and data-driven leak detection frameworks.

In recent years, the representation of a WDN as a graph, where nodes correspond to junctions, reservoirs, or tanks, and edges represent pipes, has opened the door to novel monitoring, fault-detection strategies and IoT network sizing rooted in graph theory [7], [8]. Graph-based formulations exploit the

topological and hydraulic characteristics of the network to infer anomalies from pressure and flow measurements collected at a limited subset of nodes [9]. Notably, sensor placement guided by graph-theoretic metrics has proven effective in reducing the number of sensing locations required while still maintaining high detection rates and localization accuracy [10]. Within this paradigm, centrality measures play a key role in identifying influential nodes whose monitoring can provide maximal information about the network's overall state [11].

Among centrality metrics, betweenness, closeness and PageRank centrality are particularly noteworthy [12], and have been already applied in the context of WDNs. Despite their well-established theoretical basis, there is no consensus in the literature regarding which metric better trades-off detection performance and cost of sensing procedure, especially when leaks are spatially and temporally heterogeneous.

In this paper, we propose a two-step methodology for leak detection in the case of sparse sensor measurement available. After having deployed sensors according to a centrality metric to optimize their locations, we extract low-dimensional, representations from pressure time-series using a convolutional autoencoder. After being trained on leak-free data this stage encodes . This allows us to capture the typical dynamics of a leak free network. Then, we apply a one-class Support Vector Machine (SVM) to detect anomalies in the embedding space. This model detects abnormal patterns without requiring labeled leak data by thresholding smoothed Support Vector Machine (SVM) scores. It is important to note that while our framework is designed for general anomaly detection, in this specific study we simulate anomalies as leaks to validate the method's effectiveness in detecting leak events, which are the primary concern in water distribution networks. The framework enables both binary anomaly detection and estimation of the anomaly onset time. The framework enables both binary leak detection and estimation of the leak onset time.

We apply the entire pipeline, centrality based encoding and leak detection, on two topologically different networks: the looped, high-redundancy Modena system and the sparse, tree-like Pescara system. Our results demonstrate that, in the Modena topology, selecting just 20% of nodes according to closeness or PageRank not only preserves but in some cases exceeds full-network detection accuracy, whereas in the more fragile Pescara network reducing sensors leads to performance loss, and closeness and degree centrality prove the most robust of the four measures. To explain these differences, we introduce a new graph-theoretical measure, the medium articulation complexity metric, which quantifies the network's tendency to fragment under node removal and thus predicts the viability of topology-driven pruning.

Taken together, the key contributions of the paper are as follows:

1) We perform extensive experiments on two widely-used benchmark networks, which are different in both size and topology providing diverse settings for evaluating centrality-based strategies to optimally deploy sensors with the final goal of detecting leakages;
2) We introduce a novel graph-theoretic measure, namely the medium articulation complexity, able to measure the vulnerability to fragmentation and to explain experimental data;
3) We discuss practical implications of our findings for water utilities, highlighting scenarios where one centrality measure may be preferred over others and suggesting avenues for combining centrality insights with hydraulic knowledge in future work.

To validate the proposed approach we employ a leak detection algorithm previously developed in our prior work [13], which is described in Section IV for completeness, while the rationale for parameter selection is detailed in the original publication.

The rest of this paper is organized as follows. Section II provides an overview of the state-of-the-art in node selection in graphs and the leak detection problem in WDNs. In Section III, we formally define the problem addressed in this work. Section IV describes the proposed method for solving the task. Section V details the dataset, the configuration of the proposed approach, and the training setup used for the autoencoder model. Section VI presents the results obtained using the proposed method. Finally, Section VII concludes the paper.

II. BACKGROUND

A. NODE SELECTION - SENSOR PLACEMENT

Optimal monitoring of WDNs is critical to design leak detection strategies. WDNs are typically in deteriorated conditions due to inadequate maintenance and generally suffer from insufficient monitoring. Sensors deployed across the network are typically sparse, and their placement is often not the result of well-designed deployment strategies but rather guided by practical convenience or randomness. As a result, the majority of the monitoring and anomaly detection methods focuses on signal reconstruction algorithms, which may be either model-based [14] or data-driven [15], [16].

It has been shown that the optimal sensor placement significantly improve the reconstruction accuracy with the respect to random measurement points [10], [17], [18], it is typically necessary to optimally place the sensors to enable the reconstruction of network parameters from a limited subset of measurements. In this context, graph signal sampling is an important topic in graph signal processing. In the literature, many different strategies have been designed to obtain an optimal sampling of node signals, such as based on bayesian estimation [19], but they typically rely on the total signal and there is the hypothesis of bandlimited signal [17] Those approaches has been recently applied to higher topological spaces, such as cell-complex, that can be successfully used to model sensor networks [20], [21].

B. LEAK DETECTION

Maintaining the integrity of WDNs is essential to minimize water loss, preserve infrastructure, and ensure reliable service

to consumers. Leak detection techniques have evolved significantly, driven by advances in hydraulic modeling, data acquisition technologies, and machine learning. Based on the survey in [22], these techniques can be grouped into three broad categories: model-based, hybrid model/data-driven, and fully data-driven methods. Each category addresses specific operational challenges, from accurately representing physical network behavior to processing vast sensor datasets in real time.

1) Model-Based Techniques

Model-based leak detection relies on constructing a comprehensive hydraulic model of the WDN, which simulates pressures, flows, and nodal demands under various operating scenarios. Field measurements from pressure and flow sensors are compared against the model outputs: discrepancies may indicate the presence, location, and severity of leaks. Steffelbauer et al. [23] present an integrated workflow that alternates between calibrating nodal demands and pipe roughness, using data from Automatic Meter Reading (AMR), and detecting leaks through a dual-model formulation. Virtual components such as reservoirs or valves convert head differences at sensor locations into equivalent leak flows, which are then evaluated by a CUmulative SUM (CUSUM) algorithm and likelihood ratio tests [24] to flag anomalies. Pudar and Liggett [25] employ inverse modeling on synthetic network scenarios to systematically estimate leak positions by iteratively adjusting leak parameters to minimize divergence between simulated and observed pressures. Pérez et al. [26] deploy sensitivity analysis within a hydraulic model of Barcelona's network, revealing sensor placements that maximize detection accuracy while reducing installation costs. Marzola et al. [27] refine model calibration by aligning demand profiles across zones equipped with AMR devices, then apply engineering judgment on discrepancies between inflow and aggregated demand to infer leakage events.

Hybrid Model/Data-Driven Methods

Hybrid methods leverage hydraulic simulations to generate extensive labeled datasets offline, enabling the training of machine learning or statistical algorithms. In real-time operations, these algorithms detect leaks by processing sensor data without invoking the full hydraulic model, thus reducing computational burden and mitigating calibration uncertainties. Cantos et al. [28] construct a risk-based framework that synthesizes a hydraulic simulation database spanning various leak scenarios; statistical analysis then produces spatio-temporal leak likelihood indices from live flow measurements, facilitating rapid localization within the network. Soldevilla et al. [29] frame District Metered Area (DMA) inlet-flow monitoring as a change-point detection problem, implementing a two-layer sequential algorithm that validates detected shifts and estimates both leak size and occurrence time [30]. Additional studies combine physics-based insights with machine learning, for example, Gao et al. (2021) integrate hydraulic model residuals as features in ensemble classifiers to improve detection under noisy conditions, and Li and Huang (2022) pre-train deep networks on simulated pressure transients to enhance robustness when field data are limited.

2) Data-Driven Techniques

Fully data-driven approaches bypass explicit hydraulic modeling, instead exploiting patterns in historical and real-time sensor streams [31], [32]. Early efforts employed Artificial Neural Networkss (ANNs) to classify leak events from time-series flow data [33], [34], with time-delay architectures outperforming static models. Researchers have since applied self-organizing maps and mixture density networks for anomaly clustering and density estimation [35], [36], support vector machines and Bayesian networks to capture nonlinear relationships and uncertainty in burst localization [37], [38], regression and Kalman filtering for efficient leakage flow estimation [39], principal component analysis and expectation-maximization frameworks for unsupervised outlier detection in high-dimensional sensor spaces [40], and nonlinear Kalman filters to enable rapid detection of small leaks under dynamic conditions [41]. Beyond these classical techniques, pressure-based frameworks have been developed: Geelen et al. [42] introduce a real-time pressure anomaly detection system, Wu et al. [43] apply clustering to identify bursts within DMAs, and Xing et al. [44] analyze high-frequency pressure transients to uncover subtle leak signatures.

Deep learning further advances the state of the art: Long Short-Term Memory (LSTM) networks capture temporal dependencies in leak signals [45], Convolutional Neural Network (CNN)–SVM hybrids extract localized features for classification [46], and variational autoencoder, CNN pipelines detect sub-litre leaks with low false-positive rates [47]. Denoising techniques such as Improved Spline-Local Mean Decomposition (ISMLD) combined with AlexNet and Least Square Support Vector Machine (LSSVM) classifiers enhance sensitivity in noisy environments [48], while DenseNet-inspired architectures simplify pressure signal processing by using linear connections to reduce computational complexity [49], [50]. More recently, transformer-based models, have been introduced to capture long-range temporal or frequency dependencies [51].

III. PROBLEM FORMULATION

We consider the problem of node selection and leak detection by analyzing the time-series signals acquired by the pressure sensor distributed along a WDN.

Let us consider a WDN composed by K nodes, each one equipped with a pressure sensor. The pressure measured at the k-th node is a time series sampled at regular intervals. The n-th sample at the k-th node is defined as $p_k(n)$.

We defined with S the subset of nodes selected with a centrality measure, with $S \subset K$.

The presence of a leak modifies the pressure p_s with a drop of the pressure value at the sample $N^\star$ corresponding to the

leak starting time $T^\star$, with a reduction that is proportional to the leak size.

With this setup in mind, we can define the two main goals of this paper as follows

- *To detect if a leak starts in a WDN*: this means attributing to the WDN under analysis a label $\hat{c}$ equals to 1 if a leak is detected or 0 otherwise.
- *To determine the leak starting time*: this means to compute $\hat{T}$ which is an estimate of $T^\star$.

IV. PROPOSED METHOD

We address the centrality based leak-detection problem by employing a two-stage methodology: first, we derive compact, informative feature representations from pressure matrices via the encoder portion of an autoencoder, and second, we leverage a one-class SVM to identify anomalous embeddings. The autoencoder is a specialized type of neural network designed to encode input into a compressed and meaningful representation, then decode it back so that the reconstructed input closely resembles the original [52]. Building on the architecture proposed in the work [53], the autoencoder is trained exclusively on leak-free data, ensuring that its encoder learns to capture the intrinsic structure of normal operating conditions. Each input pressure matrix is thus mapped to a low-dimensional feature vector. A one-class SVM is then fitted to these embeddings, establishing a decision boundary that encompasses the manifold of normal operation. Since the SVM is exposed solely to representations of leak-free scenarios, any embedding generated from a leak-affected matrix will fall outside this boundary and be flagged as an anomaly. By combining autoencoder-based dimensionality reduction with one-class SVM classification, our approach effectively mitigates the challenges posed by the high dimensionality and complexity of raw pressure data, yielding robust detection of leaks in WDNs.

To solve the node-selection problem, we introduce a strategy grounded in graph-theoretic centrality measures. We rank candidate sensor locations according to various centrality indices, such as betweenness, closeness,PageRank and degree centrality, and evaluate each ranking by its effect on leak-detection performance. Through this comparative analysis, we identify the centrality measure that achieves leak-detection metrics comparable to those obtained using all nodes, while substantially reducing the number of sensors under analysis and thereby lowering the computational cost of the algorithm.

A complete overview of the proposed method is presented in Figure 1, which illustrates the two main phases: the first involves node selection, while the second focuses on leak detection based on the pressure time-series recorded at the selected nodes.

A. NODE SELECTION

In this block, we defined the subset S as a portion of the set K based on the node ranking results of the different centrality measures. The nodes are ranked in descending order and we select the most important nodes according to the centrality measure selected.

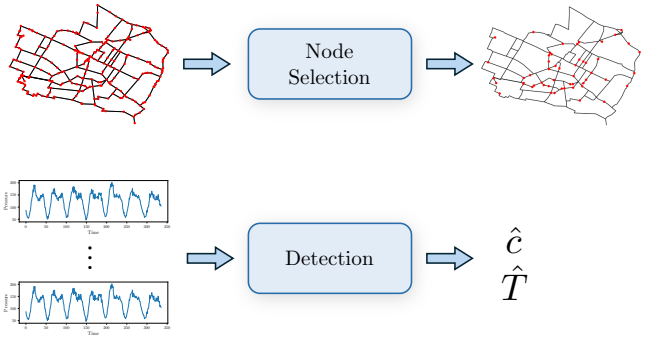


FIGURE 1. Proposed method pipeline.

Formally, let $K = \{1, 2, \ldots, N\}$ denote the full set of network nodes and let

$$c : K \longrightarrow \mathbb{R}$$

be the chosen centrality measure. We sort the nodes in descending order of their centrality scores:

$$c(i_1) \geq c(i_2) \geq \cdots \geq c(i_N), \quad \{i_1, i_2, \ldots, i_N\} = K.$$

For a prescribed subset size s, or percentage $\alpha = s/N$, we then define

$$S = \{ i_k : k = 1, 2, \ldots, s\} \subset K,$$

so that S contains the top ranked nodes according to the selected centrality measure.

Several centrality measures can be used to assess the importance of nodes in the network, each capturing a different notion of centrality. Below, we describe four widely used centrality metrics Betweenness, Closeness, PageRank and Degree which guide the selection of the subset S in our framework.

a: Betweenness Centrality:

$$c_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}},$$

where σ_{st} is the total number of shortest paths from s to t and $\sigma_{st}(v)$ is the number of those paths passing through v.

b: Closeness Centrality:

$$c_C(v) = \frac{N-1}{\sum_{u \in K} d(v,u)},$$

with $d(v,u)$ the shortest-path distance between nodes v and u.

c: PageRank Centrality:

$$c_P(i) = \alpha \sum_{j \in \mathcal{N}_i} \frac{c_P(j)}{\deg^+(j)} + \frac{1-\alpha}{N},$$

where α is the damping factor and $\deg^+(j)$ the out-degree of node j.

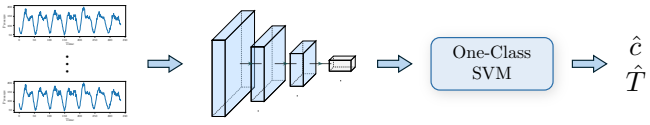


FIGURE 2. Proposed detection pipeline.

d: Degree Centrality:

$$c_D(v) = \deg(v) = \sum_{u \in K} A_{vu},$$

where A_{vu} is the adjacency matrix entry indicating connection between nodes v and u, and $\deg(v)$ denotes the degree (number of connections) of node v. For directed networks, one may distinguish in-degree $\deg^-(v)$ and out-degree $\deg^+(v)$.

B. DETECTION

The detection phase is composed of three main phases, as shown in Figure 2, that are pre-processing, encoder and one-class SVM and detection at the end. In the following, we describe each step in detail.

1) Preprocessing

Before encoding, raw pressure readings must be arranged in a fixed-size window for the encoder. At each time step n, the preprocessing unit acquires a new sample $p_s(n)$ from every node $k = 1, \ldots, S$, with $S \subset K$ and maintains the most recent L measurements per node in a circular buffer. Specifically, when $p_s(n)$ arrives, the oldest entry $p_s(n-L)$ is discarded and the window shifts to include $p_s(n-L+1), \ldots, p_s(n)$.

Formally, the buffer for node s at time n is

$$\mathbf{b}_s^n = \begin{bmatrix} p_s(n-L+1) & p_s(n-L+2) & \cdots & p_s(n) \end{bmatrix}^\top,$$

and the update rule is

$$\mathbf{b}_s^n = \left[\mathbf{b}_s^{n-1}(2:L),\ p_s(n) \right].$$

Stacking these windows for all S nodes yields the data matrix

$$\mathbf{W}^n = \begin{bmatrix} p_1(n-L+1) & p_1(n-L+2) & \cdots & p_1(n) \\ p_2(n-L+1) & p_2(n-L+2) & \cdots & p_2(n) \\ \vdots & \vdots & \ddots & \vdots \\ p_S(n-L+1) & p_S(n-L+2) & \cdots & p_S(n) \end{bmatrix} \in \mathbb{R}^{S \times L},$$

where row s collects the last L pressure samples from node s. This sliding-window matrix $\mathbf{W}^n$ is then supplied to the encoder for subsequent feature extraction.

2) Feature Extraction

The second stage of our pipeline involves extracting a compact and informative representation from the input matrix $\mathbf{W}^n$ using a dedicated feature extractor. This extractor corresponds to the encoder portion of a convolutional autoencoder, which is trained to reconstruct pressure measurements under leak-free conditions. By training exclusively on normal data, the encoder learns to capture the underlying structure of typical system behavior without being influenced by anomalous patterns. Details regarding the autoencoder architecture and training procedure are provided in [53].

The input to the encoder is the matrix $\mathbf{W}^n \in \mathbb{R}^{S \times L}$, which is processed through a series of convolutional layers to produce a low-dimensional feature vector $\mathbf{f}^n$. Specifically, the encoder consists of the following components:

- A 1D convolutional layer with 64 filters of size 7 and stride 2, followed by a Rectified Linear Unit (ReLU) activation function;
- A second 1D convolutional layer with 32 filters of size 7 and stride 2, also followed by a ReLU activation.

We can write the first convolutional layer's output at filter f as

$$h_f^{(1)} = \mathrm{ReLU}\left(\mathbf{W}^n * w_f^{(1)} + b_f^{(1)} \right),$$

where $w_f^{(1)} \in \mathbb{R}^{1 \times 7}$ is the fth kernel and $*$ denotes 1D convolution with stride 2. The second layer is analogous, yielding the feature vector

$$\mathbf{f}^n = \left[h_1^{(2)}, h_2^{(2)}, \ldots, h_d^{(2)} \right]^\top.$$

The output of the encoder is the feature vector $\mathbf{f}^n$, which serves as a compressed representation of the input matrix $\mathbf{W}^n$, capturing the essential spatiotemporal patterns relevant for subsequent anomaly detection.

$$\mathbf{f}^n = \mathcal{E}(\mathbf{W}^n). \tag{1}$$

3) One-Class SVM

The final stage of the proposed framework is the detection phase, which employs a One-Class SVM to identify potential anomalies in the feature space. One-Class SVM is a well-established technique for anomaly detection, particularly suited to cases where only data from the nominal class, i.e., normal operating conditions, are available during training [54]. The model learns to encapsulate the distribution of the normal data in a high-dimensional space by defining a hypersurface that separates the majority of training instances from potential outliers. During inference, any test point falling outside this learned boundary is classified as anomalous.

In our approach, the one-class SVM is trained on the embeddings extracted from leak-free scenarios, the same dataset used to train the autoencoder, ensuring consistency in feature representation. When a new embedding $\mathbf{f}^n$ is fed into the trained model, it produces a soft anomaly score $\hat{y}^n$, where higher values suggest a greater likelihood of an abnormal condition.

To analyze anomaly scores over time, we construct a time series vector $\hat{\mathbf{y}} = [\hat{y}^1, \hat{y}^2, \ldots, \hat{y}^N]$, aggregating the predictions across consecutive windows. To suppress transient noise and enhance robustness, we apply a moving average filter to this vector, yielding the smoothed signal $\hat{\mathbf{y}}_{\text{smoothed}}$. Anomalies are then identified by thresholding the smoothed scores using

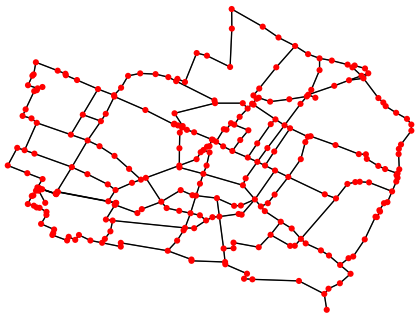


FIGURE 3. Modena network topology. Each dot represents a node. Lines represent connections between nodes.

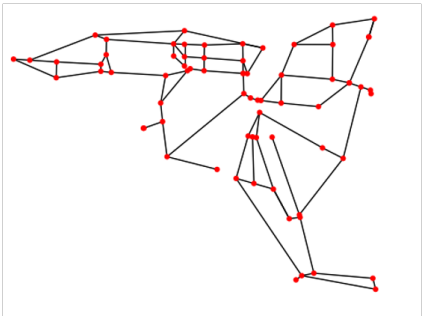


FIGURE 4. Pescara network topology. Each dot represents a node. Lines represent connections between nodes.

a predefined threshold γ. Specifically, we define the binary decision function as:

$$\hat{c}^n = \begin{cases} 1, & \text{if } \hat{y}^n_{\text{smoothed}} \geq \gamma, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where $\hat{c}^n$ indicates the presence or absence of an anomaly at time instant n.

If at least one anomaly is detected over the observation period, i.e., $\hat{c}^n = 1$ for some n, we consider a leak to have occurred. The time index of the last pressure sample associated with the first detected anomaly is denoted as $\hat{N}$, and the corresponding time instant is labeled $\hat{T}$, which we interpret as the estimated leak onset time.

V. EXPERIMENTAL SETUP

A. DATASET

To design and evaluate our proposed algorithm, we simulated various WDN topologies using the EPANET toolkit [55]. In particular, we considered two representative networks with differing structural characteristics: the Modena WDN, which exhibits a circular topology, and the Pescara WDN, which has a predominantly longitudinal layout.

The Modena network is a widely adopted benchmark in the literature [56]–[58]. It comprises 268 junction nodes and 317 pipes, with no pumps or valves included in the hydraulic model. Each pipe is characterized by parameters such as diameter, length, and roughness. The pipe lengths range from 1 m to 1094.73 m, while diameters vary from 100 mm to 400 mm. The Pescara network, in contrast, is composed of 68 junctions, 3 reservoirs, and 99 pipes, forming a layout more typical of elongated urban pipeline infrastructures.

In Figure 3 and Figure 4 we report the topology of the Modena network used in our work.

In each scenario, we have a simulation of a WDN sampled every 30 minutes for one year, so every 30 minutes a new sample is added and the buffer is shifted. The node number K is equal to 268 for all scenarios. For every node we have the pressure time series.

B. TRAINING SETUP

In this section, we provide further details on the training procedure for our approach. To obtain a compact yet informative representation of the pressure time-series data, we construct an autoencoder and use its encoder component as our feature extractor. This strategy, including the complete architectural specifications and training procedures, is adopted from our prior work [13], where the methodology is comprehensively documented. In the current paper, we adapt this established autoencoder framework to operate on the reduced input size resulting from the preceding centrality-based node selection stage, which constitutes the primary contribution of this work.

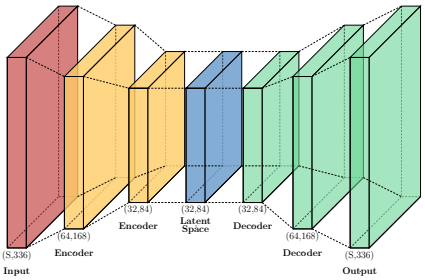


FIGURE 5. Autoencoder architecture adopted in the proposed method.

The autoencoder architecture is shown in Figure 5. The variable S denotes the number of nodes, i.e., the pressure time-series, used as input to the autoencoder.

The autoencoder consists of two parts: the *Encoder* ($\mathcal{E}$) and the *Decoder* ($\mathcal{D}$).

1) Encoder

The encoder receives an input matrix $\mathbf{W}^n \in \mathbb{R}^{K\times S}$ and maps it to a lower-dimensional latent representation. Its layer structure is detailed in Section IV-B2.

2) Decoder

The decoder takes this compact encoding and reconstructs an approximation $\hat{\mathbf{W}}^n \in \mathbb{R}^{K\times S}$. It comprises three 1D transposed convolutional layers:

- A transposed convolution with 32 filters of size 7 and stride 2, followed by a Rectified Linear Unit.
- A transposed convolution with 64 filters of size 7 and stride 2, followed by a Rectified Linear Unit.
- A transposed convolution with 32 filters of size 7 and stride 1.

TABLE 1. Autoencoder Architecture

Layer	Type	Filters	Kernel	Stride	Activation
Encoder					
Input	-	-	-	-	$\mathbf{W}^n \in \mathbb{R}^{S\times L}$
E1	Conv1D	64	7	2	ReLU
E2	Conv1D	32	7	2	ReLU
Output	-	-	-	-	$\mathbf{f}^n$
Decoder					
D1	Conv1D^T	32	7	2	ReLU
D2	Conv1D^T	64	7	2	ReLU
D3	Conv1D^T	32	7	1	-
Output	-	-	-	-	$\hat{\mathbf{W}}^n \in \mathbb{R}^{S\times L}$

Both encoder and decoder architectures were chosen via grid search to balance reconstruction loss and computational cost.

3) Training

We train the autoencoder on leak-free data using the Mean Squared Error (MSE) loss. Denoting the full autoencoder mapping by $\mathcal{AE}$, the reconstruction is

$$\hat{\mathbf{W}}^n = \mathcal{AE}(\mathbf{W}^n), \tag{3}$$

which can be equivalently expressed by separating encoder and decoder:

$$\hat{\mathbf{W}}^n = \mathcal{D}\big(\mathcal{E}(\mathbf{W}^n)\big). \tag{4}$$

We minimize the MSE between the input and its reconstruction:

$$\mathcal{L}_{\text{AE}} = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{W}_i^n - \mathcal{AE}(\mathbf{W}_i^n)\|_F^2, \tag{5}$$

where $\{\mathbf{W}_i^n\}_{i=1}^N$ are the leak-free tres, $\|\cdot\|_F$ denotes the Frobenius norm, and $\mathcal{AE} = \mathcal{D}(\mathcal{E}(\cdot))$.

We train a One-Class SVM on the encoded representations $\mathbf{f}^n = \mathcal{E}(\mathbf{W}^n)$. The One-Class SVM solves:

$$\min_{\mathbf{w},\rho,\xi} \ \frac{1}{2}\|\mathbf{w}\|^2 - \rho + \frac{1}{\nu N}\sum_{i=1}^{N}\xi_i, \tag{6}$$

$$\text{s.t. } (\mathbf{w}^\top \mathbf{z}_i) \geq \rho - \xi_i, \quad \xi_i \geq 0, \quad i = 1,\ldots,N, \tag{7}$$

where $\nu \in (0,1]$ controls the fraction of outliers and the margin, ρ is the offset, and ξ_i are slack variables.

At inference time, a new sample $\mathbf{W}^n$ is first mapped to $\mathbf{f}^n = \mathcal{E}(\mathbf{W}^n)$; then, the One-Class SVM decision function

$$\text{y}(\mathbf{f}) = \mathbf{w}^\top \mathbf{f} - \rho \tag{8}$$

is used to flag anomalies, samples above the threshold γ are classified as anomalous.

Both models leverage the same leak-free dataset, ensuring consistent feature extraction and anomaly detection boundaries.

C. PARAMETERS SELECTION

In this section we report additional details about the parameters we adopt in our pipeline.

We set the window length to cover exactly one week of measurements, i.e. $L = 336$ samples at our simulation's sampling rate. This choice ensures that each sliding window naturally captures both daily and weekly periodicities, while remaining robust to slower seasonal variations. The initial "setup time", the period required to accumulate enough data to fill the first window, is therefore also one week.

The second parameter is the threshold γ reported in Section IV-B3. The γ value has been chosen comparing the results obtained with different values of False Positive Rate (FPR), on a validation set formed by 50 simulations. In our method we decide to adopt a FPR of 10% which gave as a threshold value of 7.44.

For each WDN, nodes are ranked by centrality and we monitor the top 20%. Restricting attention to one-fifth of the nodes significantly reduces computational load while maintaining strong coverage of critical junctions, including both high-degree hubs and essential bridge nodes, thus preserving the network's structural diversity in the detection analysis.

VI. RESULTS

A. TOPOLOGY COMPLEXITY

This section evaluates the structural resilience of WDNs using a novel graph-theoretic measure: *medium articulation complexity*. This metric quantifies a network's vulnerability to fragmentation, with direct implications for operational robustness and leak detection reliability.

Traditional articulation point analysis identifies critical nodes whose removal disconnects a graph [59]. However, WDNs require a more nuanced assessment of *average nodal criticality*. To address this, we define medium articulation complexity for a graph $G = (K, E)$ as:

$$\mathcal{A}_{\text{med}}(G) \ = \ \frac{1}{|K|}\sum_{v\in K}\big(\kappa(G\setminus\{v\}) - \kappa(G)\big), \tag{9}$$

where $\kappa(G)$ denotes the number of weakly connected components. The metric computes the *average increase in component count* when individual nodes are removed, normalized by network size.

$\mathcal{A}_{\text{med}}(G)$ characterizes systemic redundancy rather than isolated critical nodes:

- **Low values**: Indicate high robustness, with multiple redundant paths maintaining connectivity despite node removals. The network lacks single points of failure.
- **High values**: Reveal structural fragility, where individual nodes disproportionately influence connectivity. Removal of such nodes fragments the network.

For WDNs, weak connectivity ensures the metric reflects physical redundancy in pipe-junction topologies rather than operational flow constraints. This is critical for assessing supply continuity risks, as infrastructure damage disrupts physical connections regardless of nominal flow directions.

TABLE 2. Medium Articulation Complexity metric results.

WDN	Medium Articulation
Modena	0.049
Pescara	0.22

Table 2 compares $\mathcal{A}_{\text{med}}$ for the Modena and Pescara WDNs:

- **Modena**: The low $\mathcal{A}_{\text{med}}$ value signifies a resilient loop-dominated topology. Node removals cause minimal fragmentation, preserving hydraulic connectivity for leak detection across diverse sensor placements.
- **Pescara**: The high $\mathcal{A}_{\text{med}}$ highlights a branch-dominated structure. Critical junctions act as choke-points; their removal creates numerous disconnected components, amplifying leak detection variability depending on sensor locations.

B. NODE SELECTION

This section analyzes sensor placement strategies for leak detection in WDNs, comparing the four centrality-based node selection methods adopted in this study: betweenness, closeness, PageRank, and degree. Figure 6 and Figure 7 visualize the top-ranked nodes, highlighted in red, for the Modena and Pescara networks, respectively, using these metrics.

For **Modena**, the *low* medium articulation complexity $\mathcal{A}_{\text{med}} = 0.049$ reflects its **loop-dominated topology**:

- The presence of many alternative paths means that removing a single node rarely disrupts connectivity.
- **Betweenness** highlights nodes near major flow intersections or reservoirs, since they frequently lie on shortest paths between different network regions.
- **Closeness** unexpectedly favors peripheral nodes. In a dense, looped network, even boundary nodes can reach the rest of the system with only slightly longer paths, making them appear ''centrally located'' from a distance perspective.
- **PageRank** tends to emphasize highly connected hubs, overlapping partly with betweenness selections, but with additional weight on nodes connected to other important nodes.
- **Degree** identifies nodes with the most direct connections, typically major junctions where multiple pipes converge. In Modena's dense topology, these high-degree nodes are distributed throughout the network rather than concentrated at specific critical points, achieving competitive but slightly lower performance than closeness and PageRank.

Overall, Modena's redundancy reduces reliance on a small set of critical nodes, even peripheral junctions remain well integrated. The strong performance of closeness and PageRank suggests that in highly looped networks, global reachability metrics effectively identify nodes with comprehensive network coverage through multiple paths.

For **Pescara**, the *moderate* medium articulation complexity $\mathcal{A}_{\text{med}} = 0.22$ reflects its **branch-like structure**:

- Connectivity relies on a hierarchical structure where certain nodes serve as critical junctions between branches, and their removal can significantly impact network integrity.
- **Betweenness** identifies bridge nodes between different branches, since they concentrate most shortest paths. However, in a sparse topology, these nodes may not coincide with the highest-connectivity junctions, resulting in suboptimal leak detection performance.
- **Closeness** spreads selections across both central and peripheral areas, but the sparse connectivity and longer path lengths result in less effective coverage compared to other metrics.
- **PageRank** emphasizes nodes near critical branches, as their structural position gives them influence over connectivity, but performs less effectively than degree in capturing local hydraulic information flow.
- **Degree** proves particularly effective in this topology, identifying nodes where multiple branches converge. These high-connectivity junctions serve as natural aggregation points for hydraulic information (pressure and flow signals), making them ideal monitoring locations. The superior performance of degree in Pescara demonstrates that local connectivity is crucial when alternate paths are limited.

Thus, unlike Modena, Pescara's limited structural redundancy makes sensor placement highly sensitive to topology. Degree centrality's success reveals that in tree-like networks, monitoring high-connectivity junctions provides comprehensive coverage by capturing information from multiple branches at critical convergence points, compensating for the lack of alternate paths that would allow path-based metrics to excel.

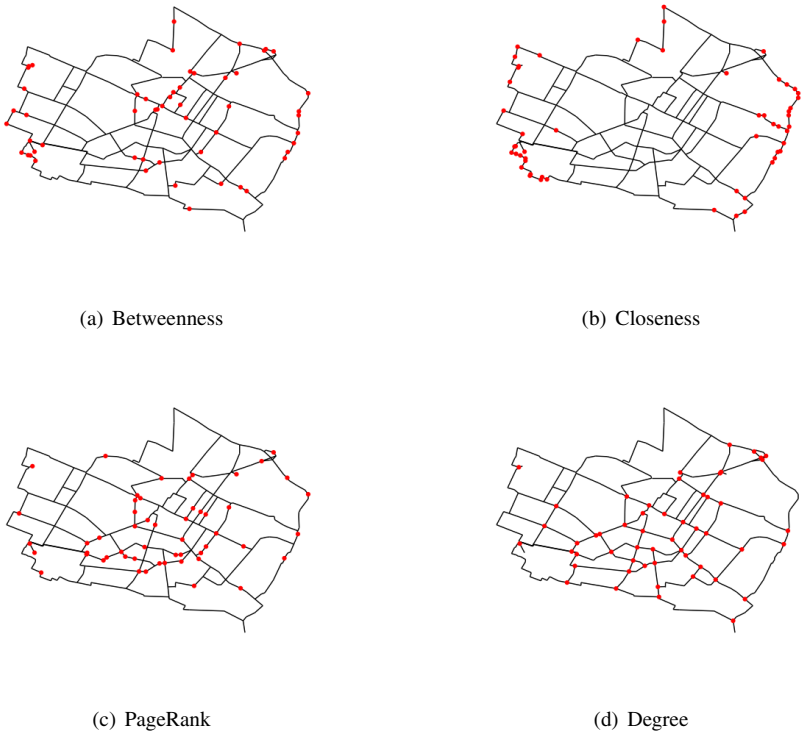
C. DETECTION

In this section we present the result for the leak detection task making a comparison between the different centrality measure used to select which node we used in the analysis. To measure the detection performances we used two metrics that are the True Positive Rate (TPR) and the Detection Delay (DD). The TPR measure how many leak that effectively occurred in the different scenario we were able to detect, whereas the DD is state-of-the-art metric in the water leak domain and it measure how fast we are able to detect the leak with respect to the actual leak starting time. The DD is defined as

$$\text{DD} = \hat{T} - T^{\star}, \tag{10}$$

were we assume that $\hat{T} \geq T^{\star}$ as we consider the detected point before the leak happens as false positive. The DD is expressed in hours.

In order to have a better comparison among the different centrality measure, we decided to present the results with two plots. The first plot represent the variation of the TPR as a function of the maximum delay accepted as a leak. This



(a) Betweenness (b) Closeness

(c) PageRank (d) Degree

FIGURE 6. **Comparison of node selection results on the Modena network using the different centrality measures evaluated in this study.**

means that we state that a leak a started only if our prediction has a value in hours equal or less that the maximum delay accepted. The second plot instead shows the DD as a function of the TPR. We report three version for each plot, at different FPR used to select the threshold for the detection step.

Furthermore we compared the results achieved with the three centrality measure against a random node selection. In order to have a fair comparison for the random node selection case we have computed the results among 5 different realization and the averaged the results obtained.

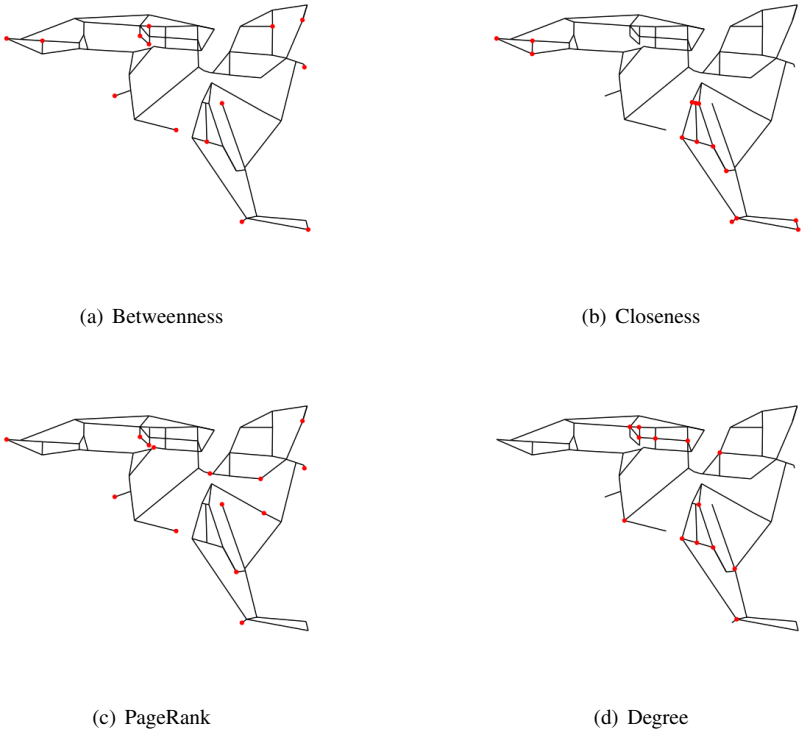
1) Baseline comparison

To validate the choice of the AutoEncoder (AE) as the feature extraction component of our pipeline, we conducted a comparative analysis against Principal Component Analysis (PCA), a standard linear dimensionality reduction technique widely used for correlated time-series data. PCA offers significant advantages in terms of interpretability and computational efficiency, making it a natural baseline for comparison. We implemented a PCA-based pipeline using the same experimental setup employed for the AE approach. Specifically, we applied PCA for feature extraction followed by One-Class SVM for anomaly detection, maintaining identical preprocessing procedures, training/test splits, and evaluation protocols to ensure a fair comparison. The PCA component was configured to retain a latent representation of comparable dimensionality to that produced by the AE encoder. The comparative results are presented in Table 3. The AE-based approach achieves substantially superior performance across both evaluation metrics. Specifically, the AE attains a TPR of 92%, representing an 8% improvement over PCA's 84%. More critically, the AE reduces the average DD by more than half, achieving 40.2 hours compared to PCA's 92.5 hours. This dramatic reduction in detection delay is particularly significant for real-world water distribution networks, where early leak detection is crucial for minimizing water loss and preventing infrastructure damage. The results are summarized in Table 3.

TABLE 3. **Comparison between PCA and AE feature extraction methods**

Method	TPR [%]	DD [hours]
PCA + One-Class SVM	84.0	92.5
AE + One-Class SVM	92.0	40.2

The superior performance of the AE can be attributed to its ability to capture temporal dependencies and nonlinear patterns in hydraulic time-series data, which are not adequately modeled by PCA's linear assumptions. Water distribution networks exhibit complex dynamics, including periodic consumption patterns, transient effects, and nonlinear pressure propagation phenomena. The convolutional architecture of the autoencoder is specifically designed to extract hierarchical temporal features that better represent



(a) Betweenness (b) Closeness

(c) PageRank (d) Degree

FIGURE 7. Comparison of node selection results on the Pescara network using the different centrality measures evaluated in this study.

these complex hydraulic behaviors. While we acknowledge that the AE operates as a "black box" compared to PCA's interpretable linear transformations, the substantial performance gains, particularly the 52.3-hour reduction in detection delay, provide strong empirical justification for the additional complexity. In critical infrastructure applications, the trade-off between interpretability and detection speed must be carefully evaluated; in this context, the ability to detect leaks more than twice as fast can translate to significant reductions in water loss and operational costs over the system's lifetime.

2) Modena WDN

The results obtained on the Modena WDN are shown in Figure 8 and Figure 9. In this case, the Closeness and PageRank centrality metrics outperform all other strategies, including the scenario where all nodes are used, i.e., no selection is applied, as well as the Betweenness-based degree-based and random node selection strategies. Specifically, at the selected threshold of 10% FPR, Figure 8(b) and Figure 9(b), we observe a TPR exceeding 80% with a DD of approximately 25-28 hours for Closeness and PageRank. This represents a substantial improvement over the other methods, which exhibit detection delays of 45-50 hours at comparable TPR levels, demonstrating nearly a two-fold reduction in delay.

This performance can be explained by considering the value of $\mathcal{A}_{\text{med}}(G)$ for the Modena WDN, which is 0.049 Table 2. Such a low value suggests that the network topology is highly resilient to node removals and can maintain its connectivity even when only a small subset of nodes is selected. As a result, centrality-based methods that prioritize structurally important nodes, such as Closeness and PageRank, are especially effective in this context.

Further promising results are reported in Figure 8(a) and Figure 9(a), which show the performance of the proposed method using a stricter threshold of 5% FPR. Even under this more conservative condition, Closeness and PageRank selections continue to achieve high TPR values (approaching 90%) while maintaining detection delays of approximately 40-48 hours, compared to 50-60 hours for other methods. These findings are particularly valuable, as they suggest that the method can operate effectively under conservative settings, significantly reducing the number of false alarms during the testing phase while preserving high detection performance. Simultaneously, we are also reducing the computational cost of the algorithm by analyzing only a small subset of the WDN nodes.

Analyzing the results obtained using a threshold corresponding to an FPR of approximately 15%, Figure 8(c) and Figure 9(c), we observe that all node selection strategies converge to comparable performance. At this less stringent threshold, the differentiation between methods diminishes significantly, with detection delays clustering in the 30–40 hour range at high TPR values. Although this setting results in

somewhat improved DD values relative to the lower-threshold scenarios, the associated false positive rate is excessively high, rendering it impractical for real-world applications due to the elevated risk of frequent false alarms.

3) Pescara WDN

For the Pescara WDN, the results are presented in Figure 11 and Figure 10. In contrast to the Modena case, these outcomes reveal more nuanced patterns in node selection performance. While the all-nodes approach generally performs well, degree centrality consistently achieves comparable results, demonstrating that strategic node selection based on degree can be as effective as full network monitoring. Closeness centrality also shows strong performance, particularly at lower FPR thresholds.

This behavior can be explained by examining the structural properties of the Pescara network. The $\mathcal{A}_{\text{med}}(G)$ value for this network is 0.22, as shown in Table 2, indicating sensitivity to node removal. However, the network's topology is such that high-degree nodes play a critical role in maintaining connectivity and information flow. By prioritizing these well-connected nodes, degree-based selection preserves the network's essential structural properties, enabling effective leak detection even with partial observations.

At the 10% FPR level, shown in Figure 11(b) and Figure 10(b), both degree and closeness centrality achieve performance nearly indistinguishable from the full-node scenario. Degree centrality reaches approximately 91% TPR with an average detection delay of around 43 hours, while closeness achieves roughly 77% TPR with a delay of approximately 55 hours. The all node approach yields about 91% TPR with a delay of 37 hours, demonstrating that degree selection captures most of the monitoring capability while using only a subset of sensors.

Under the stricter 5% FPR threshold, shown in Figure 11(a) and Figure 10(a), degree centrality emerges as the most effective selection strategy alongside all node. Degree achieves approximately 88% TPR with a mean detection delay around 59 hours, outperforming closeness which reaches about 87% TPR with a delay of 57 hours. The other selection policies, pagerank, random, and betweenness, struggle significantly, with TPR values not exceeding 61%. This result suggests that in the Pescara network, nodes with high connectivity are more informative for leak detection than nodes selected based on betweenness or random sampling.

When a 15% FPR is permitted, as shown in Figure 11(c) and Figure 10(c), degree and closeness both perform well and approach the all node benchmark. Degree centrality achieves approximately 84% TPR with delays around 33 hours at high detection thresholds, while closeness reaches similar TPR levels with delays of about 46 hours. Remarkably, at very high detection thresholds (exceeding 150 hours), both strategies can slightly exceed the performance of the full-node approach. However, as noted, the high FPR substantially limits practical applicability due to excessive false alarms.

The strong performance of degree centrality in the Pescara network can be attributed to its alignment with the network's topology. High-degree nodes serve as critical junctions where pressure and flow information from multiple branches converge. Monitoring these nodes provides comprehensive coverage of the network's hydraulic state, making degree-based selection particularly effective. This contrasts with betweenness centrality, which prioritizes nodes on shortest paths but may miss important local connectivity patterns, explaining its relatively poor performance in this case.

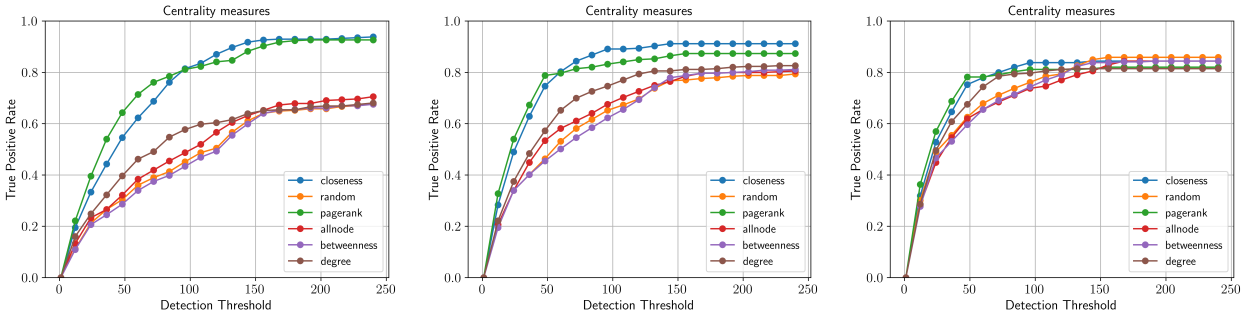
VII. CONCLUSION

In this work, we proposed a centrality-guided framework for selecting sensor nodes in WDNs to enable efficient and robust leak detection. By benchmarking degree, betweenness, closeness, and PageRank centrality measures on two contrasting networks, Modena's highly looped, redundant topology and Pescara's sparse, tree-like layout, we have demonstrated that network structure fundamentally determines which centrality metric will yield the greatest leak-detection benefit.

In Modena, where multiple alternate paths preserve connectivity, closeness and PageRank emerge as the top-performing strategies, not only matching but occasionally outperforming full-network monitoring and allowing an 80% reduction in sensor count without loss of accuracy. Degree centrality achieves competitive performance, reaching approximately 87-88% TPR at saturation, though slightly below closeness and all node. In contrast, Pescara reveals a striking reversal: degree centrality becomes the most effective strategy, consistently achieving performance comparable to full-network monitoring by prioritizing high-connectivity nodes that serve as critical information junctions. At the 5% FPR threshold, degree reaches 88% TPR with a 59-hour detection delay, outperforming all other selection strategies including closeness (87% TPR, 57-hour delay).

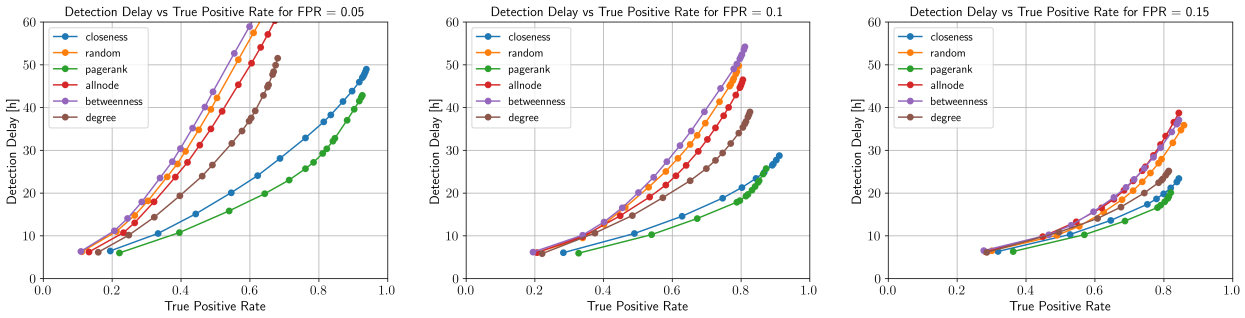
To interpret these divergent results, we introduced the medium articulation complexity metric, which quantifies how readily a network fragments under node removal. The contrast between Modena's low complexity ($\mathcal{A}_{\text{med}}(G) = 0.049$) and Pescara's moderate complexity ($\mathcal{A}_{\text{med}}(G) = 0.22$) directly explains why different centrality measures excel in each network. In highly looped networks with abundant redundancy, path-based centralities (closeness, PageRank) excel by capturing global connectivity patterns and identifying nodes with optimal reachability across the entire network. In sparser, tree-like topologies with limited redundancy, local connectivity measures, degree, prove more effective by identifying nodes where hydraulic information from multiple branches naturally converges, providing comprehensive coverage despite limited alternate paths. By providing a single scalar measure of structural fragility, this metric can guide utilities in selecting the appropriate centrality measure for their specific network topology.

The contrasting outcomes reveal a fundamental principle for sensor placement in WDNs: the optimal centrality measure must align with the network's structural characteristics.



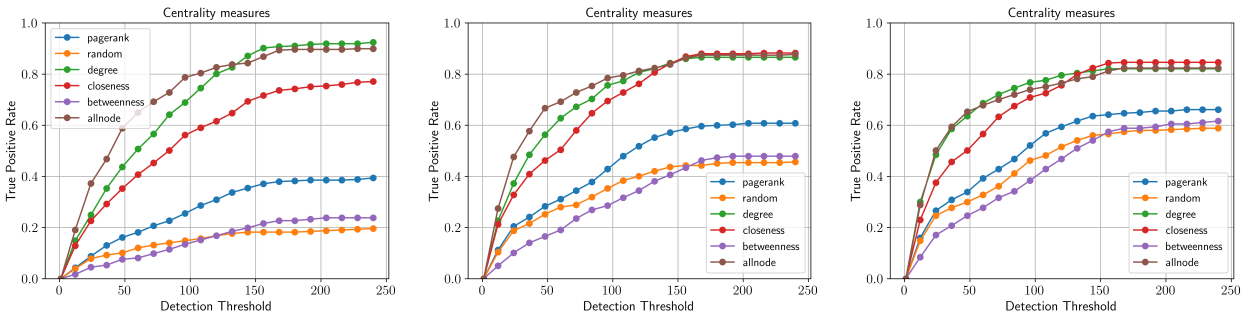
(a) FPR 0.05 (b) FPR 0.1 (c) FPR 0.15

FIGURE 8. TPR as a function of the maximum accepted detection delay for different FPR values in detection selection for Modena WDN.



(a) FPR 0.05 (b) FPR 0.1 (c) FPR 0.15

FIGURE 9. DD as a function of TPR for different FPR values in detection selection for Modena WDN.

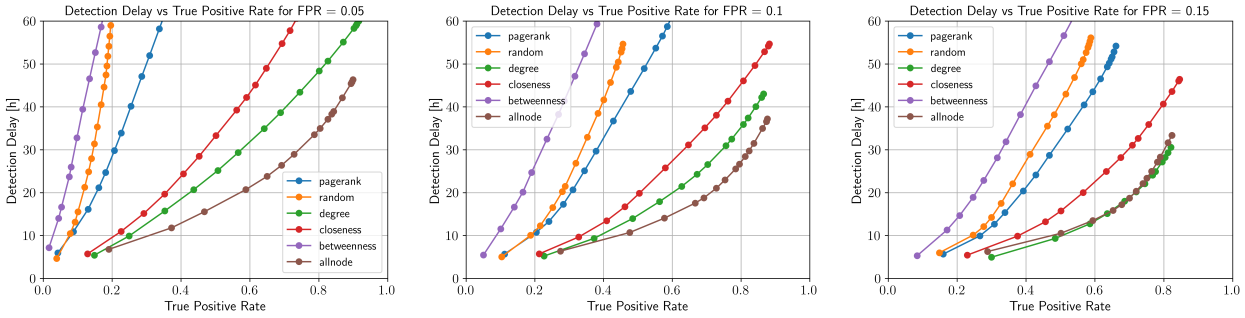


(a) FPR 0.05 (b) FPR 0.1 (c) FPR 0.15

FIGURE 10. TPR as a function of the maximum accepted detection delay for different FPR values in detection selection for Pescara WDN.

In redundant networks, global reachability metrics identify nodes that provide comprehensive network coverage through multiple paths, leveraging the network's inherent robustness. In sparse networks, local connectivity metrics identify critical junctions where monitoring captures the maximum hydraulic information, compensating for the lack of alternate paths. Degree centrality's strong performance in Pescara demonstrates that high-connectivity nodes act as natural aggregation points for pressure and flow signals, making them particularly valuable for monitoring in tree-like topologies. This insight suggests that a topology-adaptive approach, automatically selecting between degree-based and path-based centralities based on network complexity metrics, could provide robust performance across diverse WDN configurations.

As future work, we plan to enrich our framework along several complementary axes. First, we will weight centrality scores with hydraulic parameters, flow rates, pressure gradients, and pipe diameters, to produce hydraulically informed rankings that more faithfully reflect leak propagation dynamics and complement topology-based selection. This is particularly promising for degree centrality, where hydraulic weighting could distinguish between high-degree nodes with high-capacity pipes versus those with low-capacity connections. Second, we will move beyond static rankings by developing adaptive re-ranking schemes that respond to seasonal demand shifts, valve operations, or detected anomalies,



(a) FPR 0.05 (b) FPR 0.1 (c) FPR 0.15

FIGURE 11. DD as a function of TPR for different FPR values in detection selection for Pescara WDN.

keeping sensor layouts optimally tuned over time. Third, we will explore hybrid centrality measures that combine multiple metrics (e.g., degree and closeness weighted by network complexity) to leverage the strengths of different approaches across varying network conditions.

A limitation of this work is its reliance on simulated hydraulic data rather than real field measurements. While our simulations incorporate realistic demand patterns from operational networks, field validation would provide stronger evidence of practical applicability. Future work should prioritize collaboration with water utilities to obtain real pressure measurements and validate our centrality-based framework under actual operating conditions, incorporating sensor noise, measurement uncertainties, and real leak scenarios. Finally, we intend to validate our approach in field settings by collaborating with water utilities on pilot deployments, incorporating realistic sensor noise, failure modes, and maintenance schedules so as to assess operational robustness and refine our methods in light of real-world feedback.

REFERENCES

[1] Sam Fox, Will Shepherd, Richard Collins, and Joby Boxall. Experimental quantification of contaminant ingress into a buried leaking pipe during transient events. *Journal of Hydraulic Engineering*, 142(1):04015036, 2016.

[2] Samer El-Zahab and Tarek Zayed. Leak detection in water distribution networks: an introductory overview. *Smart Water*, 4(1):5, 2019.

[3] Harris Fan, Salman Tariq, and Tarek Zayed. Acoustic leak detection approaches for water pipelines. *Automation in construction*, 138:104226, 2022.

[4] Daniele Ugo Leonzio, Paolo Bestagini, Marco Marcon, Gian Paolo Quarta, and Stefano Tubaro. Water leak detection and classification using multiple sensors. In *IFIP Networking Conference*, 2024.

[5] S Meniconi, C Capponi, M Frisinghelli, and Bruno Brunone. Leak detection in a real transmission main through transient tests: Deeds and misdeeds. *Water Resources Research*, 57(3):e2020WR027838, 2021.

[6] Elias Farah and Isam Shahrour. Water leak detection: A comprehensive review of methods, challenges, and future directions. *Water*, 16(20), 2024.

[7] Giacomo Vittori, Yelizaveta Falkouskaya, Daniel Mauricio Jimenez Gutierrez, Tiziana Cattai, and Ioannis Chatzigiannakis. Graph neural networks to model and optimize the operation of water distribution networks: A review. *accepted on Journal of Industrial Information Integration*, -:–, 2025.

[8] Antonino Pagano, Domenico Garlisi, Fabrizio Giuliano, Tiziana Cattai, and Francesca Cuomo Sapienza. Swi-feed: Smart water iot framework for evaluation of energy and data in massive scenarios. In *2024 IFIP Networking Conference (IFIP Networking)*, pages 583–585. IEEE, 2024.

[9] Garðar Örn Garðarsson, Francesca Boem, and Laura Toni. Graph-based learning for leak detection and localisation in water distribution networks*. *IFAC-PapersOnLine*, 55(6):661–666, 2022. 11th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS 2022.

[10] Tiziana Cattai, Stefania Colonnese, Domenico Garlisi, Antonino Pagano, and Francesca Cuomo. Graphsmart: a method for green and accurate iot water monitoring. *ACM Transactions on Sensor Networks*, 20(6):1–32, 2024.

[11] Antonino Pagano, Domenico Garlisi, Fabrizio Giuliano, Tiziana Cattai, Redemptor Jr Laceda Taloma, and Francesca Cuomo. Introducing and evaluating swi-feed: A smart water iot framework designed for large-scale contexts. *Computer Communications*, 237:108146, 2025.

[12] Douglas J Klein. Centrality measure in graphs. *Journal of mathematical chemistry*, 47(4):1209–1223, 2010.

[13] Daniele Ugo Leonzio, Paolo Bestagini, Marco Marcon, and Stefano Tubaro. Enhanced Water Leak Detection with Convolutional Neural Networks and One-Class Support Vector Machine. In *IFIP Networking Conference*, 2025.

[14] Tiziana Cattai, Stefania Sardellitti, Stefania Colonnese, Francesca Cuomo, and Sergio Barbarossa. Physics-informed topological signal processing for water distribution network monitoring. *arXiv preprint arXiv:2505.07560*, 2025.

[15] Gergely Hajgató, Bálint Gyires-Tóth, and György Paál. Reconstructing nodal pressures in water distribution systems with graph neural networks. *arXiv preprint arXiv:2104.13619*, 2021.

[16] Garðar Örn Garðarsson, Francesca Boem, and Laura Toni. Graph-based learning for leak detection and localisation in water distribution networks. *IFAC-PapersOnLine*, 55(6):661–666, 2022.

[17] Mikhail Tsitsvero, Sergio Barbarossa, and Paolo Di Lorenzo. Signals on graphs: Uncertainty principle and sampling. *IEEE Transactions on Signal Processing*, 64(18):4845–4860, 2016.

[18] Stefania Colonnese, Mauro Biagi, Tiziana Cattai, Roberto Cusani, Fabrizio De Vico Fallani, and Gaetano Scarano. Green compressive sampling reconstruction in iot networks. *Sensors*, 18(8):2735, 2018.

[19] Luiz FO Chamon and Alejandro Ribeiro. Greedy sampling of graph signals. *IEEE Transactions on Signal Processing*, 66(1):34–47, 2017.

[20] Sergio Barbarossa and Stefania Sardellitti. Topological signal processing over simplicial complexes. *IEEE Transactions on Signal Processing*, 68:2992–3007, 2020.

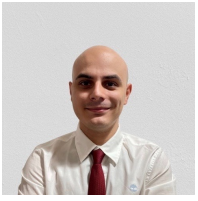
[21] Tiziana Cattai, Stefania Sardellitti, Stefania Colonnese, Francesca Cuomo, Sergio Barbarossa, et al. Leak detection in water distribution networks using topological signal processing. In *33rd European Signal Processing Conference (EUSIPCO) 2025*, 2025.

[22] Luis Romero-Ben, Débora Alves, Joaquim Blesa, Gabriela Cembrano, Vicenç Puig, and Eric Duviella. Leak detection and localization in water distribution networks: Review and perspective. *Annual Reviews in Control*, 55:392–419, 2023.

[23] David B. Steffelbauer, Jochen Deuerlein, Denis Gilbert, Edo Abraham, and Olivier Piller. Pressure-leak duality for leak detection and localization in water distribution systems. *Journal of Water Resources Planning and Management*, 148(3):04021106, 2022.

[24] Michèle Basseville and Igor Nikiforov. *Detection of Abrupt Change Theory and Application*, volume 15. PTR Prentice-Hall, 04 1993.

[25] Ranko S. Pudar and James A. Liggett. Leaks in pipe networks. *Journal of Hydraulic Engineering*, 118:1031–1046, 1992.
[26] R. Pérez, V. Puig, J. Pascual, A. Peralta, E. Landeros, and L. Jordanas. Pressure sensor distribution for leak detection in barcelona water distribution network. *Water Science and Technology: Water Supply*, 9(6):715–721, 2009.
[27] Irene Marzola, Filippo Mazzoni, Stefano Alvisi, and Marco Franchini. Leakage detection and localization in a water distribution network through comparison of observed and simulated pressure data. *Journal of Water Resources Planning and Management*, 148(1):04021096, 2022.
[28] Wilmer P. Cantos, Ilan Juran, and Silvia Tinelli. Machine-learning–based risk assessment method for leak detection and geolocation in a water distribution system. *Journal of Infrastructure Systems*, 26(1):04019039, 2020.
[29] A. Soldevila, G. Boracchi, M. Roveri, S. Tornil-Sin, and V. Puig. Leak detection and localization in water distribution networks by combining expert knowledge and data-driven models. *Neural Computing and Applications*, pages 4759–4779, 2022.
[30] Enzo Galdiero, Francesco De Paola, Nicola Fontana, Maurizio Giugni, and Dragan A. Savić. Decision support system for the optimal design of district metered areas. *Journal of Hydroinformatics*, pages 49–61, 2016.
[31] Daniele Ugo Leonzio, Paolo Bestagini, Marcon Marco, Paolo Quarta, and Stefano Tubaro. Robust Water Leak Detection and Localization with Graph Signal Processing. In *IEEE Industrial Electronics, Control, and Instrumentation Conference (IECON)*, 2023.
[32] Daniele Ugo Leonzio, Paolo Bestagini, Marcon Marco, Paolo Quarta, and Stefano Tubaro. Water Leak Detection via Domain Adaptation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024.
[33] S.R. Mounce, A.J. Day, A.S. Wood, A. Khan, P.D. Widdop, and J. Machell. A neural network approach to burst detection. *Water Science and Technology*, 45(4-5):237–246, 2002.
[34] S.R. Mounce and J. Machell. Burst detection using hydraulic data from water distribution systems with artificial neural networks. *Urban Water Journal*, 3(1):21–31, 2006.
[35] K. Aksela, M. Aksela, and R. Vahala. Leakage detection in a real distribution network using a som. *Urban Water Journal*, 6(4):279–289, 2009.
[36] S.R. Mounce, J.B. Boxall, and J. Machell. An artificial neural network/fuzzy logic system for dma flow meter data analysis providing burst identification and size estimation, 2007.
[37] S.R. Mounce, R.B. Mounce, and J.B. Boxall. Novelty detection for time series data analysis in water distribution systems using support vector machines. *Journal of Hydroinformatics*, 13(4):672–686, 2011.
[38] M. Romano, Z. Kapelan, and D.A. Savic. Geostatistical techniques for approximate location of pipe burst events in water distribution systems. *Journal of Hydroinformatics*, 15(3):634–651, 2013.
[39] G. Ye and R.A. Fenner. Kalman filtering of hydraulic measurements for burst detection in water distribution systems. *Journal of Pipeline Systems Engineering and Practice*, 2(1):14–22, 2011.
[40] G. Ye and R.A. Fenner. Weighted least squares with expectation-maximization algorithm for burst detection in uk water distribution systems. *Journal of Water Resources Planning and Management*, 140(4):417–424, 2014.
[41] D. Jung and K. Lansey. Water distribution system burst detection using a nonlinear kalman filter. *Journal of Water Resources Planning and Management*, 141(5):04014070, 2015.
[42] C.V. Geelen, D.R. Yntema, J. Molenaar, and K.J. Keesman. Monitoring support for water distribution systems based on pressure sensor data. *Water Resources Management*, 33:3339–3353, 2019.
[43] Yanyan Wu, Shuming Liu, Xuesong Wu, Yang Liu, and Yingjie Guan. Burst detection in district metering areas using a data-driven clustering algorithm. *Water Research*, 100:28–37, 2016.
[44] Lin Xing and Liran Sela. Unsteady pressure patterns discovery from high-frequency sensing in water distribution systems. *Water Research*, 158:291–300, 2019.
[45] X. Wang, G. Guo, S. Liu, Y. Wu, X. Xu, and K. Smith. Burst detection in district metering areas using deep learning method. *Journal of Water Resources Planning and Management*, 146(6):04020031, 2020.
[46] J. Kang, Y.J. Park, J. Lee, S.H. Wang, and D.S. Eom. Novel leakage detection by ensemble cnn-svm and graph-based localization in water distribution systems. *IEEE Transactions on Industrial Electronics*, 65(5):4279–4289, 2017.
[47] Roya A. Cody, Bryan A. Tolson, and Jeff Orchard. Detecting leaks in water distribution pipes using a deep autoencoder and hydroacoustic spectrograms. *Journal of Computing in Civil Engineering*, 34(2):04020001, 2020.
[48] M. Zhou, Z. Pan, Y. Liu, Q. Zhang, Y. Cai, and H. Pan. Leak detection and location based on islmd and cnn in a pipeline. *IEEE Access*, 7:30457–30464, 2019.
[49] X. Zhou, Z. Tang, W. Xu, F. Meng, X. Chu, K. Xin, and G. Fu. Deep learning identifies accurate burst locations in water distribution networks. *Water Research*, 166:115058, 2019.
[50] J. Zhang, C. Lu, X. Li, H.J. Kim, and J. Wang. A full convolutional network based on densenet for remote sensing scene classification. *Mathematical Biosciences and Engineering*, 16(5):3345–3367, 2019.
[51] Juan Luo, Chongxiao Wang, Jielong Yang, and Xionghu Zhong. A transformer-based approach to leakage detection in water distribution networks. *Sensors*, 24(19):6294, 2024.
[52] D. Bank, N. Koenigstein, and R. Giryes. Autoencoders. In *Arxiv*, 2020.
[53] Daniele Ugo Leonzio, Paolo Bestagini, Marcon Marco, Paolo Quarta, and Stefano Tubaro. Water Leak Detection and Localization using Convolutional Autoencoders. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.
[54] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.
[55] L.A. Rossman. Epanet 2.0 user manual, 2000.
[56] Suribabu Conety, T.R. Neelakantan, Perumal Raja Sivakumar, and Diego Páez. Analysis of water distribution network under pressure-deficient conditions through emitter setting. *Drinking Water Engineering and Science*, 12, 2019.
[57] H Monsef, M Naghashzadegan, Ali Jamali, and Raziyeh Farmani. Comparison of evolutionary multi objective optimization algorithms in optimum design of water distribution network. *Ain Shams Engineering Journal*, 10:103–111, 2019.
[58] Marcos Quiñones-Grueiro, Marlon Ares Milián, Maibeth Sánchez Rivero, Antônio J Silva Neto, and Orestes Llanes-Santiago. Robust leak localization in water distribution networks using computational intelligence. *Neurocomputing*, 438:195–208, 2021.
[59] Liang Tian, Amir Bashan, Da-Ning Shi, and Yang-Yu Liu. Articulation points in complex networks. *Nature communications*, 8:14223, 2017.



DANIELE UGO LEONZIO (Member, IEEE) was born in Foggia, Italy, in 1997. He received the M.Sc degree in Music and Acoustic Engineering from the Politecnico di Milano, Italy, in 2021 and the PhD in Information Technology in 2025. He is currently a Post-doc Researcher with the Image and Sound Processing Lab (ISPL) at the Department of Electronics, Information and Bioengineering (DEIB) of the Politecnico di Milano. His research activity focuses on the study of multimedia signal processing techniques.

TIZIANA CATTAI (Member, IEEE) received the bachelor's degree (cum laude) in clinical engineering and the master's degree (cum laude) in biomedical engineering from the Sapienza University of Rome, Italy, and the joint Ph.D. degree from the Brain and Spine Institute, Sorbonne University, Paris, and the Sapienza University of Rome, with a thesis on brain connectivity networks to detect mental states. She is currently an Assistant Professor with the Department of Information Engineering, Electronics, and Telecommunications (DIET), Sapienza University of Rome. Her research interests include signal processing, graph signal processing applied to the IoT sensor networks, neuroscience, and extended reality.

PAOLO BESTAGINI (S'11-M'15) was born in Novara, Italy, in 1986. He received the M.Sc. degree in Telecommunications Engineering and the Ph.D. degree in Information Technology from the Politecnico di Milano, Italy, in 2010 and 2014, respectively. He is currently Associate Professor with the Image and Sound Processing Lab (ISPL) at the Department of Electronics, Information and Bioengineering (DEIB) of the Politecnico di Milano. His research interests focus on multimedia forensics and acoustic signal processing for microphone arrays. He has been scientific investigator for the European projects SCENIC and REWIND coordinated by the Politecnico di Milano. He has been co-principal investigator for the DARPA-funded MediFor project. He is co-principal investigator for the DARPA-funded SemaFor project. He is an elected member of the IEEE Information Forensics and Security Technical Committee for the second time. He serves as an Associate Editor for the IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) and the Elsevier Journal of Visual Communication and Image Representation (JVCI). He is Co-Organizer of the IEEE Signal Processing Cup 2018 and 2022.

STEFANIA COLONNESE (Senior Member, IEEE) is Associate Professor with the Department of Information Engineering, Electronics and Telecommunications (DIET), Sapienza University of Rome, Italy. She has coauthored more than a hundred journal articles and conference papers, two book chapters, and several ISO MPEG-4 contributing documents. Her research interests include statistical signal processing, image deconvolution and restoration, and biomedical signal processing to video encoding, processing, and networking. She serves as Associate Editor for IEEE Transactions on Multimedia, devoted to the topics of multimedia broadcasting, standardization, and quality of experience.

STEFANO TUBARO (SM'01). He received the M.Sc. degree in Electronic Engineering from the Politecnico diMilano, Milan, Italy, in 1982. He then joined the Department of Electronics, Information and Bioengineering (DEIB), Politecnico di Milano, first as a Researcher of the National Research Council, and then in 1991 as an Associate Professor. Since 2004, he has been appointed Full Professor in Telecommunications at the Politecnico di Milano. He coordinates the research activities of the Image and Sound Processing Lab (ISPL) at the Department of Electronics, Information and Bioengineering (DEIB), Politecnico di Milano. In this role he coordinated several Research Projects funded by the European Commission and other institutions. Moreover, he was the DEIB Chairman from 2016 to 2022. In the past few years, he has focused his research interest on the development of innovative techniques for audio, image and video tampering detection, for the detection of synthetically generated contents and, in general, for the blind recovery of the processing history of multimedia objects. He has been a member of the IEEE Multimedia Signal Processing Technical Committee and of the IEEE SPS Image Video and of the Multidimensional Signal Processing Technical Committee. Moreover, he has been an Associate Editor and Senior Area Editor of the IEEE Transactions on Image Processing (IEEE TIP) and of the IEEE Transactions on Information Forensics and Security (IEEE TIFS).