


## Leak detection in water distribution networks based on graph signal processing of pressure data

Daniel Bezerra Barros <sup>\*</sup>, Rui Gabriel Souza, Gustavo Meirelles and Bruno Brentan

Engineering School, Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil

\*Corresponding author. E-mail: danielbezerrab@gmail.com

 DBB, 0000-0002-3664-8701

### ABSTRACT

Leakages in water distribution networks (WDNs) affect the hydraulic state of the entire or a large part of the network. Statistical correlation computed among pressure sensors monitoring network nodes aids the detection and localization of such leaks. This opens the possibility to work with water network databases, where graph signal processing (GSP) tools aid in understanding changes in pressure signals due to leakages in the hydraulic system. This paper presents a methodology for time-varying pressure signals on graph structures. The core of this methodology is based on changing of pressure, due to leaks, that modifies the graph structure. Computing for each time step a new topology of the graph and applying centrality analysis based on PageRank, it is possible to identify the presence of new leaks at the water system. A confusion matrix evaluates the precision of the proposed methodology on defining where and when such leakages start and end. Seven leaks are used to validate the process, which presented 86% in accuracy terms. The results show the benefits of the method in terms of speed, computational efficiency, and precision in detecting leakages.

**Key words:** graph signal processing, graph theory, leak detection, PageRank metric, water distribution networks

### HIGHLIGHTS

- A novel data analysis methodology for leak detection in water distribution networks.
- Correlation of pressure data for the creation of temporal graphs.
- Analysis of water networks as graphs and vertices ranking by the PageRank metric.

### INTRODUCTION

The water distribution network (WDN) is an essential infrastructure responsible for supplying citizens with drinkable water. Therefore, damage to pipes has an important impact on the water distribution process and also on water quality and network hydraulics (Hu *et al.* 2021a, 2021b). In addition, a large portion of the treated water is lost in the WDN, due to leaks, theft, measurement failures, and several other factors (Silva *et al.* 2021). According to Fan *et al.* (2021) in the UK around 3,281 Megalitres of water were lost between 2009 and 2011, and in the US around 15% of the treated water supplied is lost annually. In Brazil, it is estimated that in 2019 there were losses of about 38% in water distribution (Oliveira *et al.* 2021). This shows that even with research works focused on loss reduction and leak detection, the methodologies proposed still require further development, mainly aimed at fast and effective leak detection and easy application in WDNs.

Leaks are linked to losses in the water distribution process and are classified as reported, unreported, or background leakage (Adedeji *et al.* 2017). Reported leaks are visible on the ground and are easily detected by the public or employees of network administrators (Chan *et al.* 2018). On the other hand, unreported leaks are similar to those reported, but they do not emerge from the ground. Also finally, background leakage is small and difficult to detect by normal methods, and often goes unnoticed for a long time, resulting in significant losses (Abdulshaheed *et al.* 2017). Different methods are applied to detect leaks, and they can be classified as visual inspection, transient-based approach, model-based approach, and data-driven approach (Chan *et al.* 2018).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

Visual and sensor-based strategies require the use of mobile inspection equipment linked to optical, electromagnetic, or acoustic sensors. However, it is an expensive and time-consuming process, and often, the acoustic signals in particular are influenced by the soil type and pipe material (Fan *et al.* 2021). Transient-based approaches analyze and evaluate the pressure transient wave caused by hydraulic changes in the system to detect leaks. This wave can quickly travel through the entire network and affect flows, pressures, contract or expand pipes, and affect many other parameters to the system (Ayati *et al.* 2019). As the wave speed is usually high, from 400 m/s for PVC pipes achieving 1,200 m/s for iron pipes, transient sensors are more expensive and transient responses decay rapidly, making it difficult for large applications (Fan *et al.* 2021). Model-based leak detection methods include, for example, the use of sensitivity matrices created in normal situations and with known leaks (Geng *et al.* 2019; Salguero *et al.* 2019) and calibration approaches (Sophocleous *et al.* 2019). These methods are capable of detecting leaks; however, they require calibrated hydraulic models, with user-demand data, pipe conditions, and pressure behaviours, leading to results very sensitive to modelling and measurement errors (Fan *et al.* 2021; Hu *et al.* 2021a, 2021b). Finally, data-based methods include feature classification methods (Sun *et al.* 2019), prediction classification methods (Laucelli *et al.* 2016), mathematical–statistical methods (Quiñones-Grueiro *et al.* 2018), and unsupervised clustering (Geelen *et al.* 2019). These methods use network monitoring data, such as pressure, flow, and reservoir levels. Hu *et al.* (2021a, 2021b) conclude that data-based detection methods do not require deep WDN comprehension, but require large amounts of data, being more suitable when there is a large amount of historical network data under analysis.

The leak detection process using mathematical–statistical tools has been successfully used in research related to WDNs. Brentan *et al.* (2021) used time series of hydraulic data (pressure, flow, and reservoir levels) and applied the fast Independent Component Analysis (fastICA) algorithm to separate the hydraulic data into independent components. The authors analyze the independent components with a statistical control algorithm to detect abrupt changes to identify cyber-attacks in WDNs. Gao *et al.* (2014) use a Blind Signal Separation (BSS) process on flow and pressure data to detect leaks and separate the leak flow from the node total flow. Okeya *et al.* (2014) used historical demand data, a network hydraulic model and a modified Kalman Filter method to detect leaks in networks. The authors predict hourly demand through historical data, apply the predicted data in the hydraulic model to estimate flows and pressures, and apply the Kalman Filter to calculate corrected demands at the current time step driven by the difference between predicted and observed data (Zanfei *et al.* 2022).

Computational mathematical methods successfully used in other research fields can be explored and applied to reduce losses in water distribution systems. In this sense, complex networks can be an effective tool for this analysis, since this theory models and identifies interactions through the relationship between objects and has been widely used in network analysis (Tanaka *et al.* 2020). Complex networks are linked to the graph theory, but with a more irregular, complex structure that can evolve dynamically over time. The application of this theory began with the effort to define new concepts and measures to characterize the topology of graphs and thus united a series of principles and statistical properties (Boccaletti *et al.* 2006).

Systems modelled as a graph are represented by set of vertices and edges (Stanković *et al.* 2019). Turning to the data acquired by the WDN monitoring sensors, it is possible to statistically correlate these data for the creation of graphs. As this monitoring occurs at each defined time steps, the graph is also updated with each time step. Thus, events in the physical WDN can affect the structure of the graph and consequently the metrics that evaluate such graphs. In this perspective, studying the relationship and interactions of hydraulic data (e.g. flow, pressure, and water quality parameters) can help in the leak detection and location. Methods linked to the graph theory have been successfully used to detect anomalies in radar images (Pham *et al.* 2015), detect connectivity patterns in human brain networks (Farahani *et al.* 2019; Wright *et al.* 2021), detect anomalies in the integrity of structures (Kaveh *et al.* 2022), etc. These studies create graphs from the relationship between the monitored data and as soon as new data are issued, the verification of some graph parameters happens, such as amount of information (Pham *et al.* 2015) and information paths (Farahani *et al.* 2019; Kaveh *et al.* 2022).

Since the application of data signal processing on water distribution systems is still incipient, this work presents a methodology for leak detection based on graph signal processing of pressure data. For this, the pressure data of each monitored node is used to create a graph structure and then, a vertex ranking metric is used. This process is performed with the historical data to recognize the pattern of vertex classifications. When the centrality metric deviates from the historical pattern behaviour, an anomaly is then detected and associated with leaks in the network. The proposed methodology proved to be fast, effective, and required low computational effort for leak detection.

## METHODS

This section presents the methodology for leak detection based on graph signal processing. The process starts considering the historical pressure data and as they are made available for the graphs creation. For this, the data will be treated in a Python programming environment and the graph signal processing package (PyGSP) (Defferrard *et al.* 2017) is used to calculate the correlation between the data. The graphs will be created with the correlation calculated at each time step using the Network Analysis package in Python (NetworkX) (Hagberg & Conway 2020). After creating the graph, the PageRank node ranking metric is employed. This metric assigns a rank value to the vertices based on the edge structures. After classifying the nodes with the historical data, as more data are monitored, the new classification value of the vertices is determined, and, if this value does not follow the historical trend, a leak in the network is indicated.

The first step for the development of this methodology is the leak simulation. For this, the Water Network Tool for Resilience (WNTR) package is used (Klise *et al.* 2017). After the simulations, the pressure data at the monitoring points of the network are correlated and used in the creation of graphs. Every process follows the flowchart shown in Figure 1.

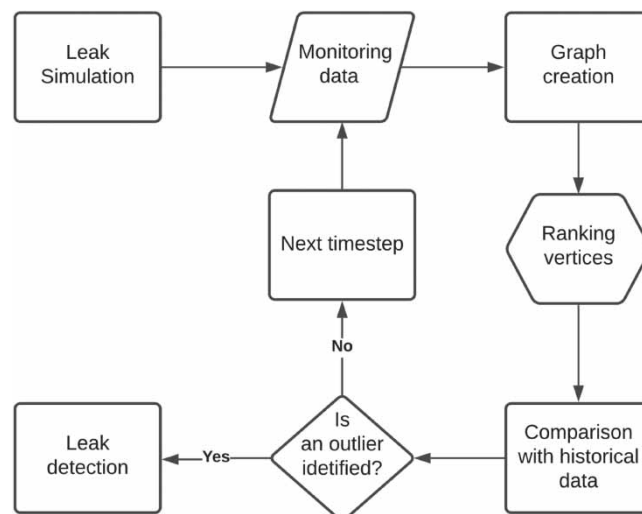
### Graph theory applied to WDS

To model a graph ( $G$ ), in general, its objects are determined, called vertices ( $V$ ) and the interaction between these vertices that is called edges ( $L$ ), being then the graph representation as  $G = (V, L)$ . The graph and its interactions can be expressed as an adjacency matrix ( $A$ ), which is a Boolean matrix with columns and rows indexed to vertices. Thus, when there is an interaction between the vertices, the edge is represented as value 1 in the matrix  $A$ . A weighted graph can be created by adding values to the edges, which are called edge weights. This results in a matrix of weights ( $W$ ), now no longer Boolean, because it contains values other than 0 and 1 (Stanković *et al.* 2019).

The modelling of the WDN as graphs is possible by considering the network physical topology, with the vertices corresponding to the nodes (junction nodes, tanks, and reservoirs) and the edges representing the links (pipes, pumps, and valves) (Di Nardo & Di Natale 2011). Another way to model the WDN as a graph is by the correlation between the monitored data. In this case, the vertices represent the monitoring nodes and the edges are the interactions between these nodes (Bezerra *et al.* 2022).

### Graph creation via pressure data correlation

The graphs are created using pressure data at each time step. For this, the research done by Kalofolias (2016) is used. The authors present a process of creating graphs from temporal data, calculating distance between pairs of vertices and penalizing edges with high weights. The authors use an  $X$  data matrix, where the columns ( $x_{v1}, x_{v2}, \dots, x_{vi}$ ) represent the sensors and the



**Figure 1** | Flowchart of the methodological process.

rows represent the time signals. The method calculates the distances between data pairs, creating a  $Z$  distance matrix:

$$Z_{ij} = \|X_{vi} - X_{vj}\|^2 \quad (1)$$

denoting the distance between the temporal data of vertices  $v_i$  and  $v_j$ . This method assumes that the smoothness between  $Z$  and the edge weight matrix ( $W$ ) is small:

$$\sum_{v_i} \sum_{v_j} W_{v_i v_j} Z_{v_i v_j} \quad (2)$$

and create a graph that minimizes very high weights on  $W$ , so that:

$$\min_{W \in W_*} \|W \circ Z\|_{1,1} + f(W) \quad (3)$$

where  $W_*$  is a set of adjacency matrices (determined by the variation of weights over time). The matrix function  $f(W)$  prevents the matrix  $W$  from a value of zero, controlling its sparsity and imposing a data-dependent structure.

This methodology calculates the distance between data pairs to create the matrix  $W$ , being implemented by the package PyGSP. The data are separated into a square matrix ( $X$ ) where the columns represent the pressure data acquired by a set of sensors ( $N$ )<sub>s</sub> in the network and the rows are the temporal data ( $t$ ). A graph will be generated as monitoring data become available. However, the first graph will have the amount of data equal to the number of sensors, in order to generate a square matrix of weights  $W$ . Once the new data are available, the oldest data are removed, always keeping the matrix  $W$  square. Finally, the matrix  $W$  is used in the graph creation and the vertices are ranked with each new graph.

### Leak detection using ranking vertex

To measure the importance of web pages, Page *et al.* (1999) proposed a method to rank each page based on the web structure. The PageRank algorithm can be described as a way of evaluating the web page importance, based on the links quantity and quality that direct to it. In a classic discrete-time finite-state random walk model, Yao *et al.* (2012) denoted by  $P$  in  $n \times n$  transition matrix, where  $n$  is the number of states and  $P_{gh}$  is the probability of transition from state  $g$  to state  $h$ . A stationary distribution value  $s$  is defined as:

$$s^T = s^T P \quad (4)$$

$$s.t. \quad s_g \geq 0 \quad \text{and} \quad \sum_{g=1}^n s_g = 1 \quad (5)$$

where  $s^T$  is the eigenvector of  $P$  that corresponds to the eigenvalue 1.

PageRank ranking modifies the random walk model to:

$$s^T = \alpha s^T P + (1 - \alpha)t_v^T \quad (6)$$

where  $t_v$  is a column vector named teleport vector, which satisfies  $\sum_{g=1}^n t_{vg} = 1$  and  $\alpha$  is a scalar called the damping factor. The teleport vector and the damping factor are considered so that there is communication between pages (vertices) in case of graph not connected. Thus, Equation (6) corresponds to the union of two random walks, one with the matrix  $P$  and the other that is the transition from state  $g$  to any other state with probability  $t_v$ , and has  $\alpha$  as a mixing parameter. In the initial iteration, all vertices are considered equal with  $t_{vg} = 1/n$  and  $\alpha = 0.85$  (Yao *et al.* 2012), but over the interactions the weights of vertices and edges are being updated. This makes the convergence rate limited and empirically mimics the behaviour of web users.

The PageRank values of vertices are given in simplified form by:

$$PR(u) = \alpha \sum_{c \in B(u)} \frac{PR(c)}{N_c} \quad (7)$$

where  $u$  represents the web page (or the vertex in a graph),  $B(u)$  is the set of vertices that point to  $u$ ,  $PR(u)$ ,  $PR(c)$  are classification scores of vertices  $u$  and  $c$ ;  $N_c$  is the number of edges leaving vertex  $c$  (Xing & Ghorbani 2004).

To detect leaks, we used the maximum values determined by the PageRank metric. The maximum score does not refer to a specific vertex, but the highest value achieved by all vertices. The data of the first 7 days that are simulated without leaks are used as maximum base values and are compared with the data of the next days of simulation. If this comparison exceeds a threshold, it will be considered a data anomaly. The determination of this stipulated value will be determined by a sensitivity analysis and the percentage that presents the best result is chosen. Finally, a confusion matrix will be generated between two sets of data, the first considering the simulated leaks, where the moments of beginning and end are known, and the second the times with anomaly detections.

A confusion matrix is generally used in classification problems between two sets of data in four combinations: true positive (TP), true negative (TN), false positive (FP), and false negative (FN) (Berry *et al.* 2021). Table 1 shows how the classifications are arranged in the confusion matrix.

Between the two datasets in the confusion matrix, there is one with the real observations, in the case of this research they are the known leaks, and another with the observations predicted by the method. In other words, in this work, the observed data are the time that the known leaks start and the estimated data are the time that PageRank values vary significantly from the normal trend, higher than a marginal error.

To evaluate the confusion matrix, the accuracy metric, which is the correct proportion of the method, is used. For this, the equation follows:

$$Ac = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (8)$$

The results are between 0 and 100, with 100 being a perfect prediction result (Berry *et al.* 2021).

### Case study

The WDN presented by Bragalli *et al.* (2012) is adapted and a set of leaks are simulated for the application of the methodology presented in this work. This network is based on the distribution network of the Italian city Modena, in the Emilia-Romagna region, and it is composed of 268 nodes, 4 reservoirs, and 317 pipes and does not have valves and pumps, as shown in Figure 2.

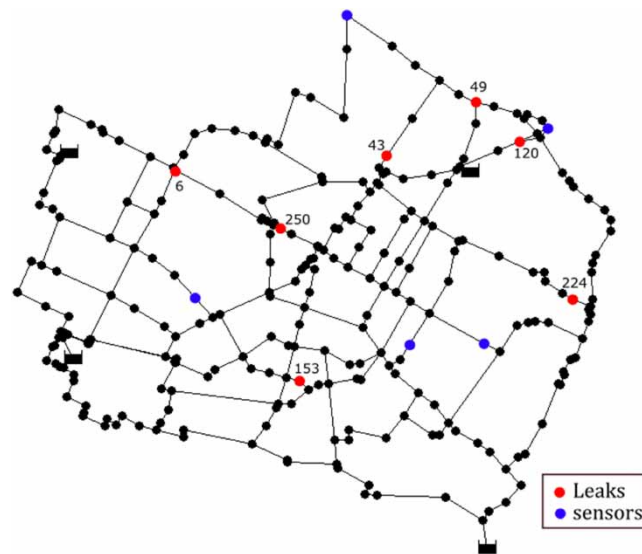
The research by Mankad *et al.* (2022) uses this network to place pressure sensors to detect and locate leaks. The authors placed five pressure sensors (exposed in Figure 2) and only the data from these sensors will be considered in the process of graphs generation. This means that the  $X$  data matrix (Equation (1)) will have five columns and will result in graphs with five vertices. A leak simulation process is also required due to lack of leakage scenarios in this network. Therefore, seven nodes were selected with leak points and these are shown in Figure 2.

### Simulation process

Since there is no real monitoring data for this network, a leak simulation process will be performed using the WNTR package (Klise *et al.* 2017). Leaks will be considered as an additional demand flow at the network nodes, with the leakage flow rate

**Table 1** | Confusion matrix example

		Observed	
		Positive	Negative
Estimated	Positive	TP	FP
	Negative	FN	TN



**Figure 2** | Modena network.

( $q_L$ ) being determined by the orifice equation:

$$q_L = C_d A \sqrt{2gP} \quad (9)$$

where  $C_d$  is the discharge coefficient,  $A$  is the orifice area,  $g$  is the gravity acceleration, and  $P$  is the pressure at the leak point. As the leak flow affects the node pressure and the pressure determines the leak flow, an iterative process must be carried out until the pressure and the leak flow rate stabilize and then this flow is used in addition to the demand. The  $C_d$  utilized is equal to 0.6, following the research by Van Zyl (2014). The area values will be chosen randomly between 0.00012 and 0.00050 m<sup>2</sup>, and the values are used in the research by Van Zyl & Cassa (2014).

Another factor with random choice will be the start times of the leaks and their durations. The entire simulation process is done for 40 days, with the first 7 days without leaks to obtain a reference historical data. After that date, leaks are included. During the simulation process, the base demand and the hourly multiplication factor are multiplied by a randomness factor, in this study considered between 0.9 and 1.1, making the process closer to reality. The sensors have an acquisition frequency of 1 h, meaning that a graph will be created and PageRank will be calculated every hour.

## RESULTS AND DISCUSSION

Seven leak scenarios are performed at different locations on the network. The leaks are simulated with different magnitudes and behaviours: in some cases, there is an increase in flow over time and in others there was an abrupt appearance of leaks. Figure 3 shows the behaviour of leak flows for each node where leaks are simulated. It can be observed that there were leaks occurring in a spaced way and also a situation of simultaneous leaks in different locations (nodes 120 and 224).

The entire simulation process considered that the sensors emitted monitoring signals every hour, resulting in a dataset with 960 measurements. In this sense, there are the same amount of graph creations and nodes rankings by the PageRank metric, since each measurement is used in the process. The graphs change as pressure data changes at monitoring points. This change can be observed in the edge weights in Figure 4.

Figure 4 shows four distinct situations of the weights in the graph. Figure 4(a) exposes the graph in a leak-free situation. Also, Figures 4(b)–4(d) show the graphs generated at the peak moments of the leaks at the 250, 120, and 49 nodes, respectively. Thus, Figure 5 shows these values throughout the simulation, including the moments when there are leaks.

It can be observed in Figure 5 that the maximum PageRank values change significantly when there are leaks in the network, not having much relation with the intensity of the leaks. It can be seen, for example, that leakage at node 250, even with a high value of leakage, does not significantly impact the behaviour of PageRank values, especially when compared to leakage



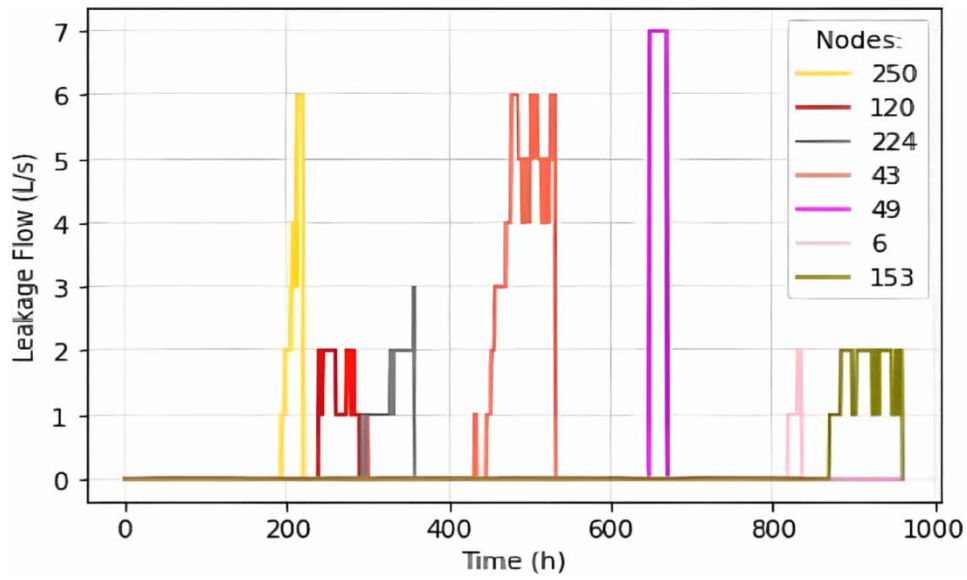


Figure 3 | Leakage flow behaviour.

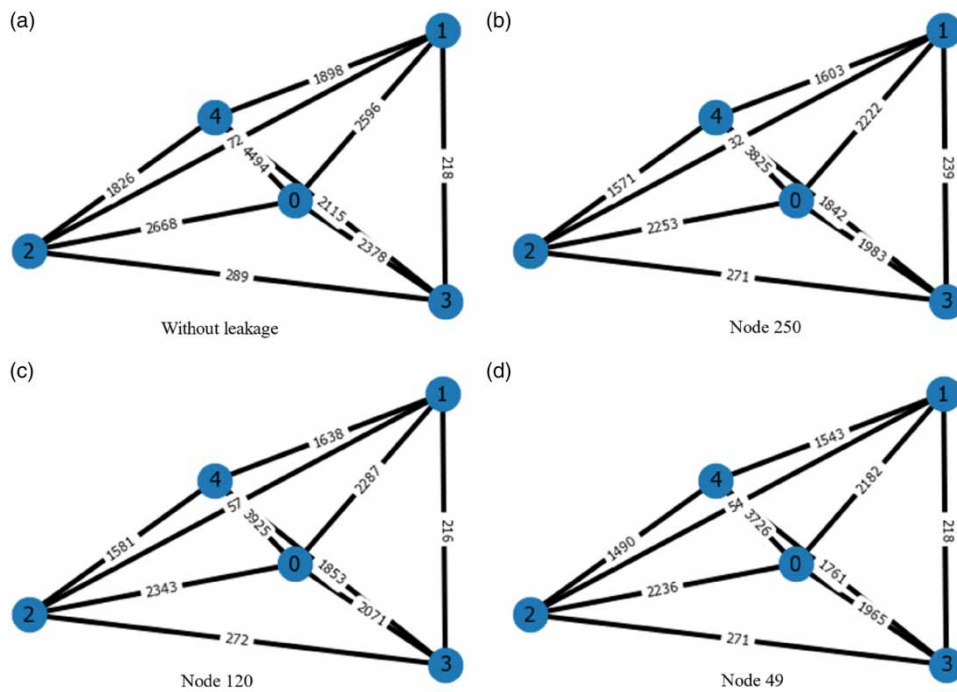
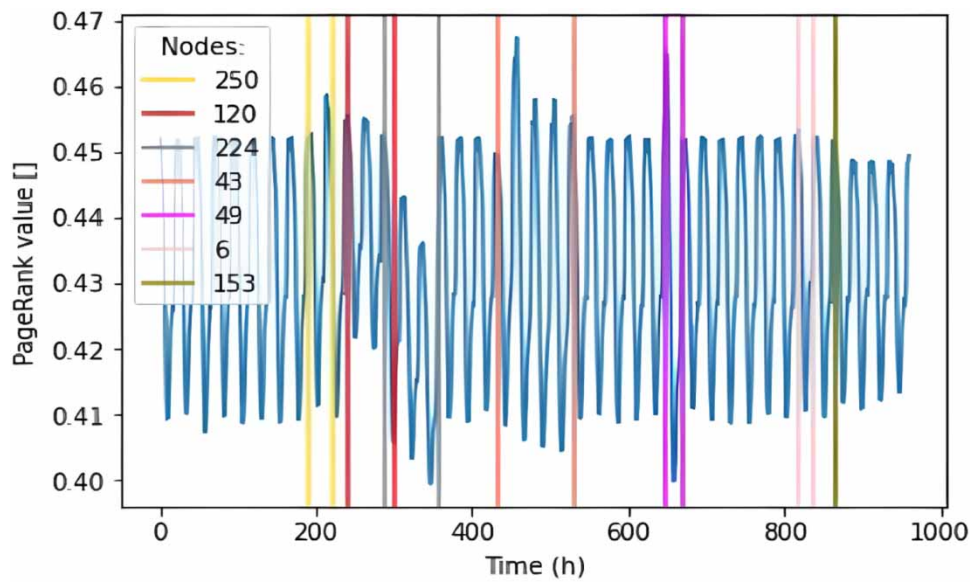


Figure 4 | Graphs in anomalous situations.

at nodes 120 and 224, which are of lower intensity, but which cause considerable changes in PageRank values. On the other hand, the leak at node 6 does not cause a change in behaviour, probably because it is a short time, low-flow leak, which the sensors may not have been able to monitor.

To define the marginal error, a sensitivity analysis is performed, in which the process varied between 1 and 10% the hourly PageRank values calculated during historical data (7 days of data) looking for the highest TP values. This analysis showed that, considering a margin of error higher than 5%, the TP values are very low (134–13%, times detect and percentage of

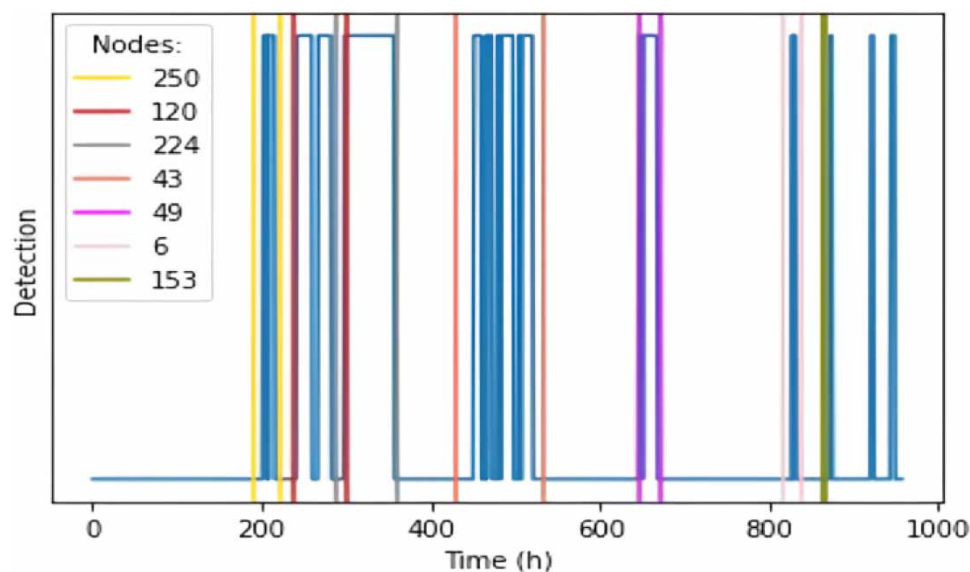


**Figure 5** | PageRank maximum values.

total). However, if the marginal error is 1%, the detection has a high value of FP (202–21%). Thus, it indicates a marginal error of 2%, both in this application and in future applications, as it is the percentage that achieves the best TP values. The detection moments using the margin of error equal to 2% are shown in [Figure 6](#).

It can be seen in [Figure 6](#) that some leaks are detected throughout their duration (nodes 120, 224, and 49). As for the leaks at nodes 250, 43, and 6, the detection takes place in certain time steps, but during a large part of the leaks' duration. However, the leak at node 153, which lasts until the end of the simulation, is detected in just a few moments. For a better understanding of the results, a confusion matrix is created and is shown in [Table 2](#).

Evaluating the confusion matrix in terms of accuracy, the detection method obtained an 86% score. This means, according to [Chicco et al. \(2021\)](#), that there is high accuracy in leak detection. This shows that the leak detection process proposed in this research guarantees a good detection rate using monitoring data already performed based on the application. The proposed methodology is easy and quick to execute, proving to be an effective approach to leak detection.



**Figure 6** | Anomaly detection.



**Table 2** | Confusion matrix

		Observed	
		Positive	Negative
Estimated	Positive	655 (68%)	14 (1.46%)
	Negative	113 (11.7%)	176 (1.3%)

## CONCLUSIONS AND RECOMMENDATIONS

This research presented a methodology for leak detection that makes use of the ranking of vertices in graphs by the PageRank metric. The proposed methodology creates graphs with the correlation between the pressure data emitted by monitoring sensors using graph creation and analysis packages in a Python programming environment. It was shown that the graphs undergo changes when leaks occur. Thus, there are also changes in the vertex ranking values, and these changes are used to detect leaks. A historical series of hourly pressure data is used to learn the vertices rank for the different days, and the leak is detected when the maximum PageRank value of the vertices is different from those learned from the historical data. The proposal achieved a score of 86% based on seven simulated leaks with the results evaluated in terms of the accuracy of a confusion matrix. The methodology proved to be easy and quick to apply.

Even so, there are still points to be improved. Some of these points can be addressed in future works, seeking, for example, to use new ways of creating graphs, such as methods that consider the topology of the network. New vertex classification metrics can also be used considering, for example, the centrality of the vertices or the number of degrees and also using other sources of monitoring data, such as flows, reservoir levels, and water quality. Another approach that can be followed in future is leak detection by methods linked to the graph theory, such as through the connection between the sensor nodes and the other network nodes.

## 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

- Abdulshaheed, A., Mustapha, F. & Ghavamian, A. 2017 A pressure-based method for monitoring leaks in a pipe distribution system: a review. *Renewable and Sustainable Energy Reviews* **69**, 902–911.
- Adedeji, K. B., Hamam, Y., Abe, B. T. & Abu-Mahfouz, A. M. 2017 Towards achieving a reliable leakage detection and localization algorithm for application in water piping networks: an overview. *IEEE Access* **5**, 20272–20285.
- Ayati, A. H., Haghighi, A. & Lee, P. 2019 Statistical review of major standpoints in hydraulic transient-based leak detection. *Journal of Hydraulic Structures* **5** (1), 1–26.
- Berry, M. W., Zain, J. M., Yap, B. W. & Mohamed, A. 2021 *Soft Computing in Data Science*. Springer, Singapore.
- Bezerra, D., Souza, R., Meirelles, G. & Brentan, B. 2022 Leak detection in water distribution networks based on graph signal processing of pressure data. In *International Joint Conference on Water Distribution Systems Analysis*. (Vol. 86, p. 14073).
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D. U. 2006 Complex networks: structure and dynamics. *Physics Reports* **424** (4–5), 175–308.
- Bragalli, C., D'Ambrosio, C., Lee, J., Lodi, A. & Toth, P. 2012 On the optimal design of water distribution networks: a practical MINLP approach. *Optimization and Engineering* **13** (2), 219–246.
- Brentan, B., Rezende, P., Barros, D., Meirelles, G., Luvizotto Jr., E. & Izquierdo, J. 2021 Cyber-attack detection in water distribution systems based on blind sources separation technique. *Water* **13** (6), 795.
- Chan, T. K., Chin, C. S. & Zhong, X. 2018 Review of current technologies and proposed intelligent methodologies for water distributed network leakage detection. *IEEE Access* **6**, 78846–78867.
- Chico, D., Tötsch, N. & Jurman, G. 2021 The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining* **14** (1), 1–22.
- Defferrard, M., Martin, L., Pena, R. & Perraudin, N. 2017 *Pygsp: Graph Signal Processing in Python*. Available from: <https://github.com/epfl-lts2/pygsp>.

- Di Nardo, A. & Di Natale, M. 2011 A heuristic design support methodology based on graph theory for district metering of water supply networks. *Engineering Optimization* **43** (2), 193–211.
- Fan, X., Zhang, X. & Yu, X. B. 2021 Machine learning model and strategy for fast and accurate detection of leaks in water supply network. *Journal of Infrastructure Preservation and Resilience* **2** (1), 1–21.
- Farahani, F. V., Karwowski, W. & Lighthall, N. R. 2019 Application of graph theory for identifying connectivity patterns in human brain networks: a systematic review. *Frontiers in Neuroscience* **13**, 585.
- Gao, J., Qi, S., Wu, W., Li, D., Ruan, T., Chen, L., Shi, T., Zheng, C. & Zhuang, Y. 2014 Study on leakage rate in water distribution network using fast independent component analysis. *Procedia Engineering* **89**, 934–941.
- Geelen, C. V., Yntema, D. R., Molenaar, J. & Keesman, K. J. 2019 Monitoring support for water distribution systems based on pressure sensor data. *Water Resources Management* **33** (10), 3339–3353.
- Geng, Z., Hu, X., Han, Y. & Zhong, Y. 2019 A novel leakage-detection method based on sensitivity matrix of pipe flow: case study of water distribution systems. *Journal of Water Resources Planning and Management* **145** (2), 04018094.
- Hagberg, A. & Conway, D. 2020 *NetworkX: Network Analysis with Python*.
- Hu, X., Han, Y., Yu, B., Geng, Z. & Fan, J. 2021a Novel leakage detection and water loss management of urban water supply network using multiscale neural networks. *Journal of Cleaner Production* **278**, 123611.
- Hu, Z., Chen, B., Chen, W., Tan, D. & Shen, D. 2021b Review of model-based and data-driven approaches for leak detection and location in water distribution systems. *Water Supply* **21** (7), 3282–3306.
- Kalofolias, V. 2016 How to learn a graph from smooth signals. In: *The 19th International Conference on Artificial Intelligence and Statistics*, Cadiz, Spain, 9–11 May 2016. PMLR, pp. 920–929.
- Kaveh, A., Rahmani, P. & Dadras Eslamlou, A. 2022 A multistage damage detection approach using graph theory and water strider algorithm. *Iranian Journal of Science and Technology, Transactions of Civil Engineering* **46** (1), 33–54.
- Klise, K. A., Hart, D., Moriarty, D. M., Bynum, M. L., Murray, R., Burkhardt, J. & Haxton, T. 2017 *Water Network Tool for Resilience (WNTR) User Manual (No. SAND2017-8883R)*. Sandia National Lab.(SNL-NM), Albuquerque, NM, United States.
- Laucelli, D., Romano, M., Savić, D. & Giustolisi, O. 2016 Detecting anomalies in water distribution networks using EPR modelling paradigm. *Journal of Hydroinformatics* **18** (3), 409–427.
- Mankad, J., Natarajan, B. & Srinivasan, B. 2022 Integrated approach for optimal sensor placement and state estimation: a case study on water distribution networks. *ISA Transactions* **123**, 272–285.
- Okeya, I., Kapelan, Z., Hutton, C. & Naga, D. 2014 Online burst detection in a water distribution system using the Kalman filter and hydraulic modelling. *Procedia Engineering* **89**, 418–427.
- Oliveira, G., Scazufca, P., Sayon, P. & Oliveira, R. 2021 *Perdas de água 2021 (SNIS 2019): Desafios para disponibilidade hídrica e avanço da eficiência do saneamento básico*. São Paulo.
- Page, L., Brin, S., Motwani, R. & Winograd, T. 1999 *The PageRank Citation Ranking: Bringing Order to the web*. Stanford InfoLab, Stanford, CA.
- Pham, M. T., Mercier, G. & Michel, J. 2015 Change detection between SAR images using a pointwise approach and graph theory. *IEEE Transactions on Geoscience and Remote Sensing* **54** (4), 2020–2032.
- Quiñones-Grueiro, M., Verde, C., Prieto-Moreno, A. & Llanes-Santiago, O. 2018 An unsupervised approach to leak detection and location in water distribution networks. *International Journal of Applied Mathematics and Computer Science* **28** (2), 283–295.
- Salguero, F. J., Cobacho, R. & Pardo, M. A. 2019 Unreported leaks location using pressure and flow sensitivity in water distribution networks. *Water Supply* **19** (1), 11–18.
- Silva, M. A., Loureiro, D., Amado, C. & Arregui, F. 2021 Análise de dados de telemedição de consumos domésticos para apoio na avaliação de perdas aparentes e reais em sistemas de distribuição de água. *Ingeniería del Agua* **25** (1), 37–47.
- Sophocleous, S., Savić, D. & Kapelan, Z. 2019 Leak Localization in a Real Water Distribution Network Based on Search-Space Reduction. *Journal of Water Resources and Planning* **145** (7).
- Stanković, L., Daković, M. & Sejdić, E. 2019 Introduction to graph signal processing. In: *Vertex-Frequency Analysis of Graph Signals* (Stanković, L. & Sejdić, E., eds.). Springer, Cham, pp. 3–108.
- Sun, C., Parellada, B., Puig, V. & Cembrano, G. 2019 Leak localization in water distribution networks using pressure and data-driven classifier approach. *Water* **12** (1), 54.
- Tanaka, Y., Eldar, Y. C., Ortega, A. & Cheung, G. 2020 Sampling signals on graphs: from theory to applications. *IEEE Signal Processing Magazine* **37** (6), 14–30.
- Van Zyl, J. E. 2014 Theoretical modeling of pressure and leakage in water distribution systems. *Procedia Engineering* **89**, 273–277.
- Van Zyl, J. E. & Cassa, A. M. 2014 Modeling elastically deforming leaks in water distribution pipes. *Journal of Hydraulic Engineering* **140** (2), 182–189.
- Wright, L., De Marco, M. & Venneri, A. 2021 A graph theory approach to detect age and Alzheimer's disease related changes to structural neural networks. *Alzheimer's & Dementia* **17**, e050583.
- Xing, W. & Ghorbani, A. 2004 Weighted PageRank algorithm. In: *Proceedings. Second Annual Conference on Communication Networks and Services Research, 2004*, Fredericton, Canada, 21–21 May 2004. IEEE, New York, pp. 305–314.
- Yao, Z., Mark, P. & Rabbat, M. 2012 Anomaly detection using proximity graph and pagerank algorithm. *IEEE Transactions on Information Forensics and Security* **7** (4), 1288–1300.
- Zanfei, A., Menapace, A., Brentan, B. M., Righetti, M. & Herrera, M. 2022 Novel approach for burst detection in water distribution systems based on graph neural networks. *Sustainable Cities and Society* **86**, 104090.