



Graph attention neural network for water network partitioning

Kezhen Rong¹ · Minglei Fu^{1,2} · Yangyang Huang¹ · Ming Zhang¹ · Lejin Zheng³ · Jianfeng Zheng³ · Miklas Scholz^{4,5,6} · Zaher Mundher Yaseen⁷

Received: 28 June 2022 / Accepted: 17 October 2022 / Published online: 15 November 2022
© The Author(s) 2022

Abstract

Partitioning a water distribution network into several district metered areas is beneficial for its management. Partitioning a network according to its node features and connections remains a challenge. A recent study has realized water network partitioning based on node features or pipe connections individually. This study proposes an unsupervised clustering method for nodes based on a graph neural network, which uses graph attention technology to update node features based on the connections and a neural network to cluster nodes. The similarity between nodes located in each area and the balance of the total water demand between areas are optimized, and the importance of the boundary pipes is calculated to determine the installation position of flowmeters and valves. Three water distribution networks with different structures and sizes are used to verify the proposed model. The results show that the average location differences (LocDiffs) within the areas of the three networks completed by partitioning are 0.12, 0.07, and 0.06, and the total demand differences (DemDiffs) between areas are 0.13, 0.27, and 0.29, respectively. The LocDiff and DemDiff of the proposed method decreased by 6% and 55%, respectively, when compared to the traditional clustering method. Additionally, the proposed method for calculating the importance of boundaries provides an objective basis for boundary closure. When the same number of boundaries are closed, the comprehensive impact of the proposed method on the pipe network decreases by 17.1%. The proposed method can be used in practical applications because it ensures a highly reliable and interpretive water distribution network partitioning method.

Keywords Deep learning · Unsupervised clustering · Graph attention · Water network partitioning

✉ Minglei Fu
fuml@zjut.edu.cn

✉ Zaher Mundher Yaseen
zaheryaseen88@gmail.com; z.yaseen@kfupm.edu.sa

Kezhen Rong
18268735665@163.com

Yangyang Huang
1365154450@qq.com

Ming Zhang
cim2046@zjut.edu.cn

Lejin Zheng
raymond.zheng@laisontech.com

Jianfeng Zheng
jianfeng.zheng@laisontech.com

Miklas Scholz
m.scholz@salford.ac.uk

² College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China

³ Hangzhou Laison Technology Co., Ltd, Hangzhou 310063, China

⁴ Directorate of Engineering the Future, School of Science, Engineering and Environment, The University of Salford, Newton Building, Salford M5 4WT, UK

⁵ Department of Civil Engineering Science, School of Civil Engineering and the Built Environment, University of Johannesburg, Kingsway Campus, Auckland Park, P.O. Box 524, Johannesburg 2006, South Africa

⁶ Department of Town Planning, Engineering Networks and Systems, South Ural State University, 76, Lenin Prospekt, Chelyabinsk 454080, Russia

⁷ Civil and Environmental Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

¹ College of Sciences, Zhejiang University of Technology, Hangzhou 310023, China

Introduction

Water distribution networks (WDNs) are composed of demand nodes and transport pipes that provide water to a city's residents and industrial facilities (Alvisi 2015). They are susceptible to unexpected water consumption because of pipe bursts, connection leakages, water theft, and other factors (Puust et al. 2010; Bhagat et al. 2019). The British Water Industry Association proposed the concept of water network partitioning (WNP), which divides a WDN into several district metered areas (DMAs) to manage the network more effectively (Wu et al. 2016). It has proven to be immensely beneficial to the management of WDNs, which include reducing losses (Nicolini and Zovatto 2009), limiting pollution flow (Grayman et al. 2009), and optimizing pressure (Gomes et al. 2011). Therefore, how to partition the water network based on the network structure and the points features can improve the detection efficiency and response speed of the water network for anomalies in the practical phase (Gupta et al. 2020; Sharma et al. 2022). But it is still a challenge for Hydraulic Utilities and academic circles.

The common WNP problem can be formulated into two steps: clustering and division. The clustering stage refers to the clustering of water demand nodes into areas based on their features and connections. The division stage refers to the installation of flowmeters or valves in boundary pipes between areas to achieve an independent meter (Sela Perelman et al. 2015). In the last decade, various methods for clustering water demand nodes and division of boundary pipes have emerged. In the clustering stage, several studies have been conducted to investigate and identify the factors affecting WNP. The features of water demand nodes in a network (such as water demand, horizontal and vertical positions, and elevation) affect the WNP (Twort et al. 2000). For example, if the nodes with similar spatial locations are divided into the same DMA, the potential energy loss of water flow is reduced, which makes the network pressure distribution more uniform (Grayman et al. 2009; Gajghate et al. 2021). Simultaneously, the close connection of water network within the DMA implies these nodes are less connected with other DMAs; therefore, there are fewer boundary pipes between DMAs, which helps simplify the determination of boundaries (Ciaponi et al. 2016). Additionally, connection between nodes is an important factor affecting WNP (Giudicianni et al. 2018). The correlation of nodes is not a simple spatial relationship. Nodes are connected through pipes, which implies the nodes at both ends of the pipes should be clustered as a DMA. Several data-based algorithms have been used to partition water networks. Graph theory mainly clusters the nodes based on a network structure, which includes depth-first and breadth-first searches (Tzatchkov and Yamanaka 2008; Di Nardo and Di

Natale 2011; Lifshitz and Ostfeld 2018); the community-structure algorithm divides a water network according to network density (Diao et al. 2013; Ciaponi et al. 2016); the modularity-based algorithm considers the features of nodes based on the former two methods (Giustolisi and Ridolfi 2014); the multilevel partitioning algorithm is used to evenly distribute the pressure, water demand, and structure of a pipe network (Di Nardo et al. 2013; Saldarriaga et al. 2019); spectral graph algorithms cluster the nodes using the vector information of a network (Herrera et al. 2010; Di Nardo et al. 2018); and the multi-agent approach divides networks by constructing competing node groups (Izquierdo et al. 2009; Herrera et al. 2012; Hajebi et al. 2014). While, in the division stage, the methods include single-objective optimization (Farmani et al. 2006), multi-objective optimization (Di Nardo et al. 2015), iterative approach (Pesantez et al. 2019), and adaptive sectorization (Giudicianni et al. 2020). Generally, these methods can achieve WNP according to different basics, but remain following problems:

- Not considered demand nodes self and connection relationship comprehensively.
- Not have proper way to avoid fall into local optimization.
- Not avoided the negative impact of trial-and-error partition on the actual network.

The development of artificial intelligence technology and its combination with traditional fields provide possibilities for innovative WNP methods (Choudhary et al. 2021; Rakhshani et al. 2021; Yang et al. 2022). Based on the analysis of node features and connections in networks, this study proposes a graph attention network to design DMAs (GAN-DMA) to realize WNP. In the proposed method, each node in the network aggregates the features of the surrounding nodes based on the connection, and the nodes are clustered using an unsupervised neural network. Simultaneously, the importance of the boundary pipes formed in the clustering stage is evaluated and ranked, and the pipes for installing flowmeters and valves are determined by calculating the influence degree and scope indices.

Contributions of this study are as follows:

- Graph attention of node features realizes the comprehensive consideration of node features and pipeline connections, making the input of neural network more reasonable.
- Unsupervised neural network realizes highly reliable node clustering and avoids the dilemma that traditional clustering is easy to fall into local optimization.
- Assessment of Boundary Pipes provides the objective basis for boundary division and avoids the negative impact of trial-and-error division on real pipe network.

Problem Statement

Preliminary description

First, the number of DMAs formed must be determined. Ferrari et al. suggested that each DMA should connect 500–5000 customers, and the total number of customers of the entire network is based on the total water demand of the network and per capita water demand (Ferrari et al. 2014). Smaller DMAs facilitate loss detection but represent higher initial costs and management difficulties (Laucelli et al. 2016). In this study, the number of DNAs is determined using Eq. (1).

$$N_{DMAs} = \frac{D_t}{D_{av}N_c} \tag{1}$$

where D_t is the total water demand, D_{av} is the per capita water demand, and N_c is the average number of customer connections in each DMA. The number of customer connections in each DMA is an ambiguous value; therefore, the number of DMAs formed is variable, and the effect of the number of DMAs on the results will be examined in Section 4.4. Additionally, node clustering is determined by node features, including location and demand features (Herrera et al. 2010). The location features contain the longitude, dimension, and altitude data, whereas the demand feature refers to the water demand of the nodes. These features are collected and aggregated for node clustering. Owing to the aggregation based on network-connection information, each node in the network creates a list of its neighbor nodes and stores it as a sparse adjacency matrix, which is used in the following steps.

Problem description

In this study, the partitioning of a WDN into DMAs is defined as an optimization problem that can be automatically solved using the model, and the definition of objectives directly affects the model results. Several studies have described the objectives of WNP (Brentan et al. 2018; Liu and Han 2018). Firstly, the internal stability of each DMA is the most important problem to be considered in WNP. Internal stability means that the features of nodes in each DMA maintain a high degree of correlation, such as longitude, latitude, and altitude. This can be interpreted as intra-area correlation facilitating intra-area management and reducing the difficulty and workload of subsequent water loss location. Secondly, the balance between areas should also be considered. The balance between areas mainly means that the

water demand of each area should be kept close, which can optimize the number of DMAs, reduce the impact of WNP on the original water network, and facilitate the maintenance and detection of the water network. Finally, the original performance of the pipe network should not be ignored, and the WNP should not affect the daily water use of water units. In fact, due to the closure of some boundary pipe networks, the impact of DMAs was significant, which mainly refers to the decline of water node pressure head caused by the closure of boundary (Taha et al. 2021; Tiernan and Hodges 2022).

Problem formulation

An ideal model generally requires a lower average location difference (LocDiff) and total demand difference (DemDiff) between areas, and fewer boundary pipes, which facilitate subsequent management. Moreover, the network water pressure decreases after WNP because a part of the boundary pipes of the formed DMAs must be closed and valves must be installed. The pressure head of each demand node must be higher than the minimum pressure head to satisfy the daily water-pressure demand (Di Nardo et al. 2015). Therefore, the objectives and constraints of the model are as follows:

$$\min LocDiff = \frac{1}{N_{DMAs}} \sum_{i=1}^{N_{DMAs}} \left(\frac{1}{P_i} \sum_{j=1}^{P_i} \left(\sqrt{(L_j - \bar{L})^2} / \bar{L} \right) \right) \tag{2}$$

$$\min DemDiff = \frac{1}{N_{DMAs}} \sum_{i=1}^{N_{DMAs}} \frac{\sqrt{(D_i - \bar{D})^2}}{\bar{D}} \tag{3}$$

$$\min N_B \tag{4}$$

$$s.t. \forall h \geq h^* \tag{5}$$

where N_{DMAs} denotes the number of DMAs, P denotes the number of surrounding nodes, L denotes the location features, \bar{L} denotes the average location features, D denotes the total water demand of the DMA, \bar{D} denotes the average total water demand of the DMAs, N_B is the number of boundary pipes, h is the pressure head, and h^* is the minimum pressure head. The objectives and constraints of the model are set to realize an automatic clustering of demand nodes, which is implemented using the GANDMA method and divided into two steps in the clustering stage: feature aggregation and unsupervised clustering. Subsequently, in the division stage, the importance of the boundary pipes formed by clustering is calculated and used to evaluate the installation of flowmeters or valves.

Methodology

Method overview

In this study, the GANDMA method is used for WNP and maintains the daily operational status of the WDN. The method consists of three steps (Fig. 1). In the first step, feature sampling and aggregation are performed for each node (Section 3.2). In the second step, the aggregated node features are used for unsupervised clustering using a neural network (Section 3.3). Finally, the importance of the boundary pipes is calculated and used to determine whether they need to be closed (Section 3.4).

graph and the node j_1, j_2, \dots, j_K connected to it through the pipes, the aggregation index α between two nodes is based on the feature difference and pipe length. α represents the degree of influence of node j on node i . This process can be described by Eq. (6).

$$\alpha_{ij} = \frac{D_j \sum_{f=1}^{N_f} \left(1/\sqrt{(F_i^f - F_j^f)^2 + 1} \right)}{\sum_{k=1}^{K+1} \left(D_k \sum_{f=1}^{N_f} \left(1/\sqrt{(F_i^f - F_k^f)^2 + 1} \right) \right)} \quad (6)$$

where D is water demand, N_f is the type of feature, F is the feature of the node, and K is the number of surrounding nodes. The α values of all nodes adjacent to node i and the

Algorithm 1. Pseudo code of the GANDMA

Inputs: the feature of the nodes in the water distribution network: F
 the structure of the water distribution network: G
 the parameters of the neural network:
training epochs(E), *network layers and numbers of neurons in each layer*
for $i = 1$ to N_{nodes}
 for $j = 1$ to $N_{surrounding\ nodes}$
 Calculate α_{ij} using Eq. (6)
 Update the feature of the node i using Eq. (7)
 end
end
for $n = 1$ to E
 Obtain the node classification result: *model*(F)
 Calculate *LocDiff*, *DemDiff* using Eq. (1-2)
 Optimize the neural network parameters
end
 Calculate the I_{ids} of boundary pipes using Eq. (9)
 Rank the importance of boundary pipes
for $n = 1$ to N_{close}
 if $I_r \geq 0.3$ && $h \geq h^*$
 Update partition result
 else
 Output partition result
end
Output: water network partitioning result

3.2.

Feature Aggregation

The water distribution network can be regarded as a graph, G_0 , that contains several demand points and connection pipes. Each node adjacent to others through the pipes usually is similar in spatial position and water demand behavior (Zhang et al. 2021). Shorter pipes imply that two points have stronger similarities and are more suitable for division into a DMA (Giudicianni et al. 2018). Therefore, for a node i in the

i are calculated, and the features of the surrounding nodes are weighted and aggregated based on α to realize the feature update of node i . This process can be described by Eq. (7).

$$F_i^{(0+1)} = \sum_{k=1}^{K+1} \alpha_{ij} F_k \quad (7)$$

The same operations are performed on all nodes in the network to make each of them aggregate the location

features of its surrounding nodes, and these updated nodes form a new graph layer, G_1 . We repeat this step r times to obtain the r -th layer graph, G_r , which implies each node in the graph contains the features of nodes adjacent to its r -order.

Unsupervised Clustering

The aggregated node location features are used for unsupervised clustering, and the process is automatically realized by a neural network, which includes the input, hidden, and output layers (Oyedele et al. 2021). The input layer contains the features of each node; the hidden layer contains several layers with several neurons; the output layer is the clustering result of nodes. The weight and bias of each neuron are optimized through training to satisfy the expected objective, which is to maintain the similarity of nodes located in each area and the balance of total water demand between areas (Eqs. 2 and 3). Training of the neural network is implemented using the chain derivative optimization rule under the PyTorch framework, and the step size of the optimization is set to 0.02. The structure of the neural network has a significant effect on unsupervised clustering, which manifests in the model efficiency and clustering results. More complex networks can achieve better results; however, the workforce under the same condition requires a longer running time. Therefore, the influence of the model layers and neural numbers on clustering is discussed in Section 4.2.

Boundary Division

The water demand nodes are clustered into several areas, and each area has multiple pipe connections to others on the boundary. These boundary pipes were equipped with flowmeters or valves to facilitate management. The installation of flowmeters is costly, and the installation of valves prevents the flow of water and worsens water quality by extending its age (Herrera et al. 2016). *Todini et al.* proposed the use of resilience index (I_r) to evaluate the comprehensive impact of valve installation on daily network operation (Todini 2000). This is widely used to evaluate network reliability in the event of an emergency and can be expressed as follows:

$$I_r = \frac{\sum_{i=1}^{n_n} Q_i(h_i - h^*)}{\sum_{r=1}^{n_r} Q_r H_r - \sum_{i=1}^{n_n} Q_i h^*} \tag{8}$$

where n_n and n_r are the numbers of demand nodes and reservoirs, Q and h are the water demand and pressure head, respectively; and h^* is the design minimum pressure head. The GANDMA method is used to evaluate the importance and influence of boundary pipes, which in turn is based on α and distance Dt of the surrounding node, and can be described as the influence degree and scope index as follows:

$$I_{ids} = \sum_{i=1}^{N_s} \sum_{j=1}^{K_i} \frac{\alpha_{ij}}{Dt_i} \tag{9}$$

where N_s denotes the number of selected nodes. A higher I_{ids} indicates the boundary pipe has a higher impact on areas on both sides; therefore, the pipe is not suitable for closure and the installation of a flowmeter should be prioritized.

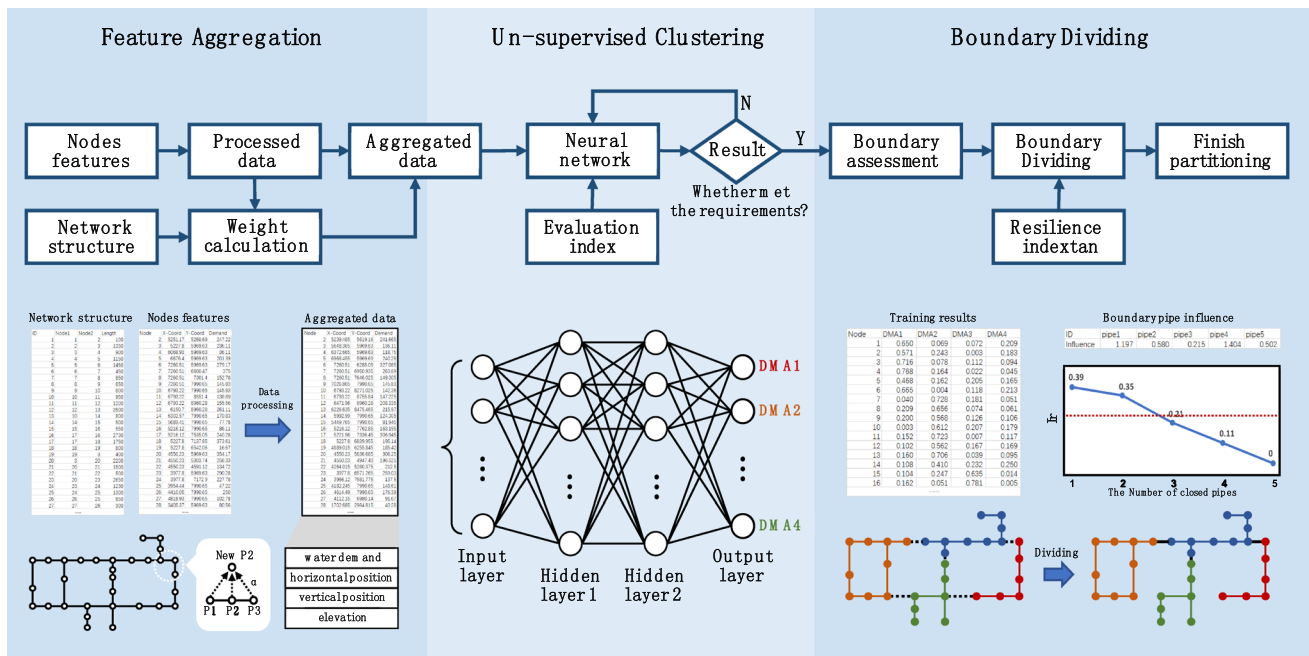


Fig. 1 Steps adopted for partitioning water networks

Conversely, valves should be installed in boundary pipes with lower I_{ids} . Furthermore, the influence of different flowmeter and valve ratios in the network is evaluated, as described in Section 4.6.

Experiment and Evaluation

Description of the data

GANDMA was used in three WDNs with different sizes and network structures: Hanoi, C-town, and E-town (Pesantez et al. 2019), which represent small, medium, and large WDNs, respectively (Fig. 2). The Hanoi network is a part of the total Hanoi water distribution network, which transports water by gravity and consists of one water source, 31 water demand nodes, and 50 pipes. The C-town network was proposed in the Battle of Water Calibration Networks and

imitates a real medium-sized water distribution network consisting of one water source, 388 water demand nodes, and 429 pipes. The E-town network is a large water distribution network for supplying water to 400,000 people. It consists of five water sources, 11,063 water demand nodes, and 13,896 pipes. The features of each water demand node and the connection between pipes in the three networks were recorded and used to cluster and divide the networks. These features include spatial location and water demand.

Determination of network structure

The structure of a neural network directly affects the efficiency and results of the node clustering. Generally, a more complex network can achieve better results; however, a workforce under the same conditions requires a longer running time. This study analyzed the influence of different network layers and numbers of neurons in each layer on the final results. The

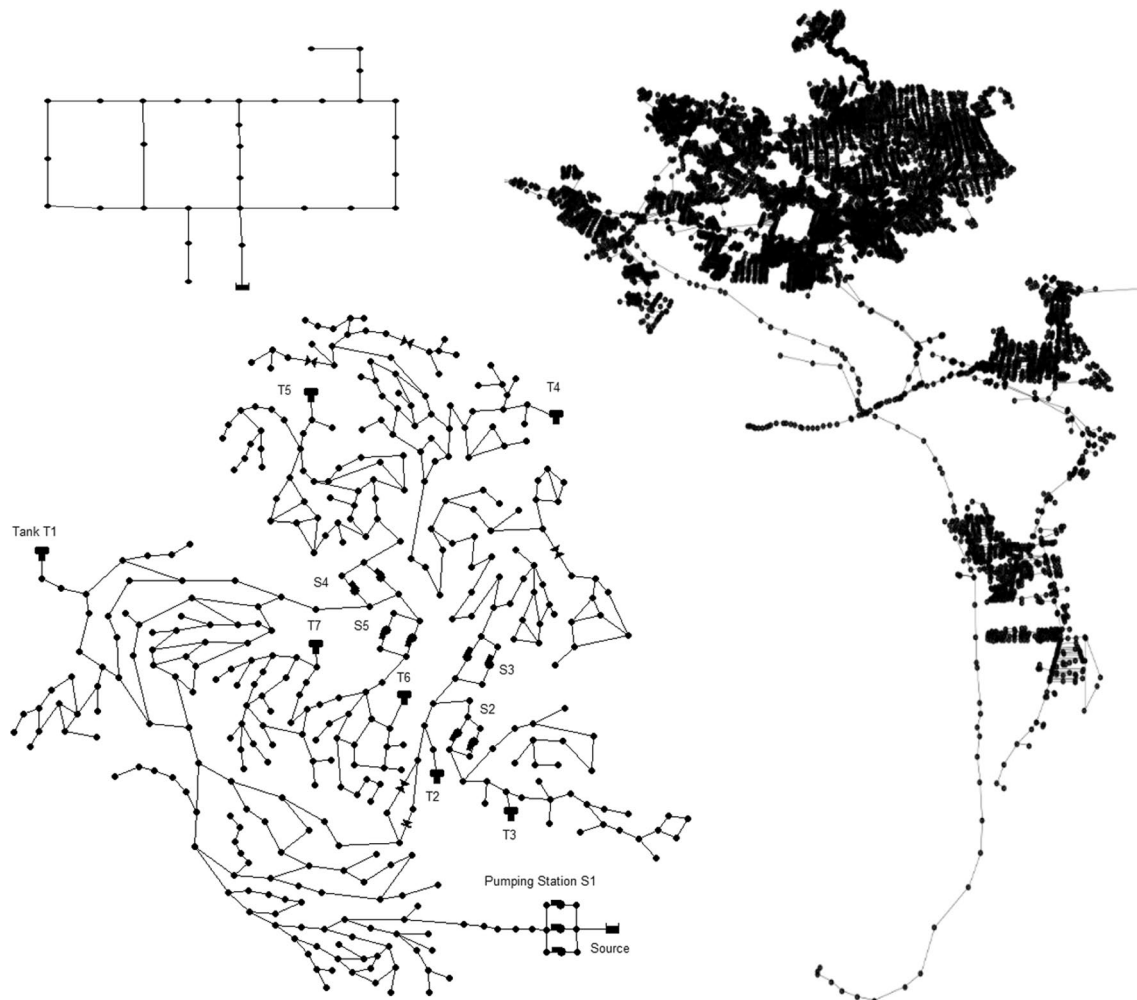


Fig. 2 Water distribution networks of Hanoi (upper left), C-town (lower left), and E-town (right)

effect was measured using LocDiff and DemDiff. Fig. 3 shows that under the 100 training epochs, the structure with two layers and 20 neurons in each layer achieved the best effect; the LocDiff and DemDiff were 0.07 and 0.31, respectively, which are in line with the expected results (Alvisi and Franchini 2014). Therefore, this structure was used in the clustering of demand nodes. A further analysis showed that the largest difference was more than 15% based on the different models, which indicates the model structure has a significant impact on node clustering (Table 1).

Performance of node clustering

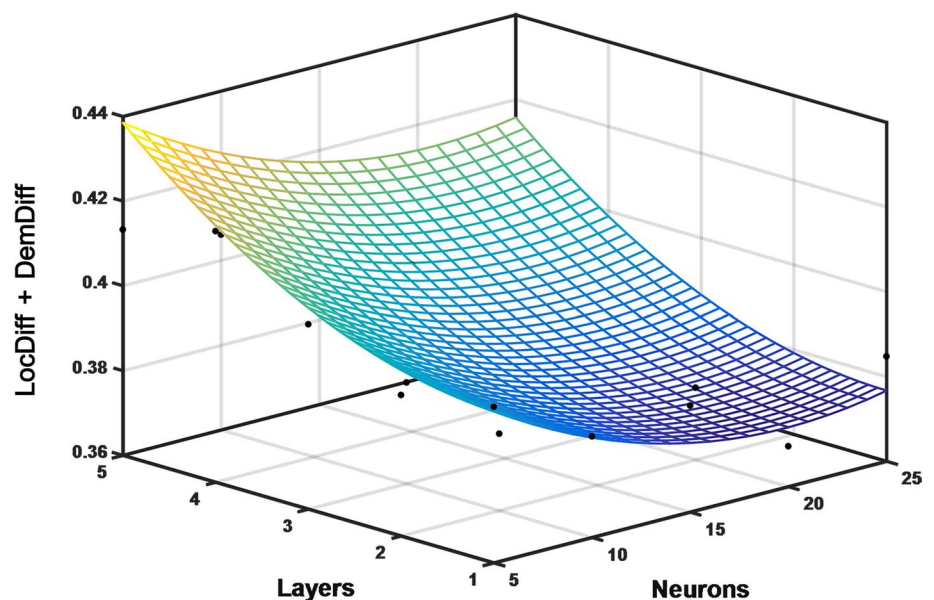
The GANDMA method was applied to the cluster nodes of three independent WDNs, and the clustering results were evaluated using LocDiffs and DemDiffs (Fig. 4). First, the number of DMAs in the three networks were determined to be 4, 7, and 160, and the aggregation range was 5 (the influence of the number of DMAs and aggregation range on the results will be discussed in Sections 4.4 and 4.5). The neural network structure was then used to cluster the demand nodes. Table 2 shows the GANDMA method's excellent performance for all network sizes. For the Hanoi network, the LocDiff and DemDiff values were 0.12 and 0.13, respectively; for the C-town network, the LocDiff and DemDiff values were 0.07 and 0.27, respectively; and for the E-town network, the LocDiff and DemDiff values were 0.06 and 0.29, respectively. Compared to the K-means and spectral graph methods (Wu et al. 2016), the GANDMA method significantly enhanced the similarity of nodes located in each area and maintained the balance of total water demand between areas. Compared to the K-means method, the average LocDiff and DemDiff values in the three areas formed using the GANDMA

method decreased by 6% and 55%, respectively. Compared to the spectral graph method, the average LocDiff and DemDiff values decreased by 3% and 36%, respectively. This is because the GANDMA method allows each node to include the location features of multiple nearby nodes by aggregation, and the new graph formed by these multi-information nodes significantly enriches the information from the original graph. Subsequently, the neural network was used for automatic clustering to address the limitation of the traditional method, which is only suitable for single-objective clustering with low-dimensional information. It realizes multi-objective node clustering with a high stability. The GANDMA method achieved a balance of water demand between areas by maintaining similar node location features within each area.

Influence of number of DMAs

This study evaluated the influence of the number of DMAs on clustering results (Fig. 5). The C-town and E-town networks were used for this purpose because the Hanoi network is small and the allowable number of DMAs is limited. For C-town, LocDiffs and DemDiffs were recorded when the demand nodes were clustered into 3–11 DMAs. Table 3 shows that with an increase in the number of DMAs, the LocDiff gradually decreased and stabilized at approximately 0.07 when the number of DMAs was more than seven. Additionally, the DemDiff gradually decreased and stabilized at approximately 0.32 when the number of DMAs was more than eight. After a comprehensive consideration, the most appropriate number of DMAs was chosen to be seven (LocDiff = 0.07, DemDiff = 0.32, and $N_B = 22$). While the results of E-town were similar to that of C-town, the DemDiff gradually decreased and stabilized at approximately

Fig. 3 Grid-search results based on model effectiveness for the number of neural network layers and neurons in each layer



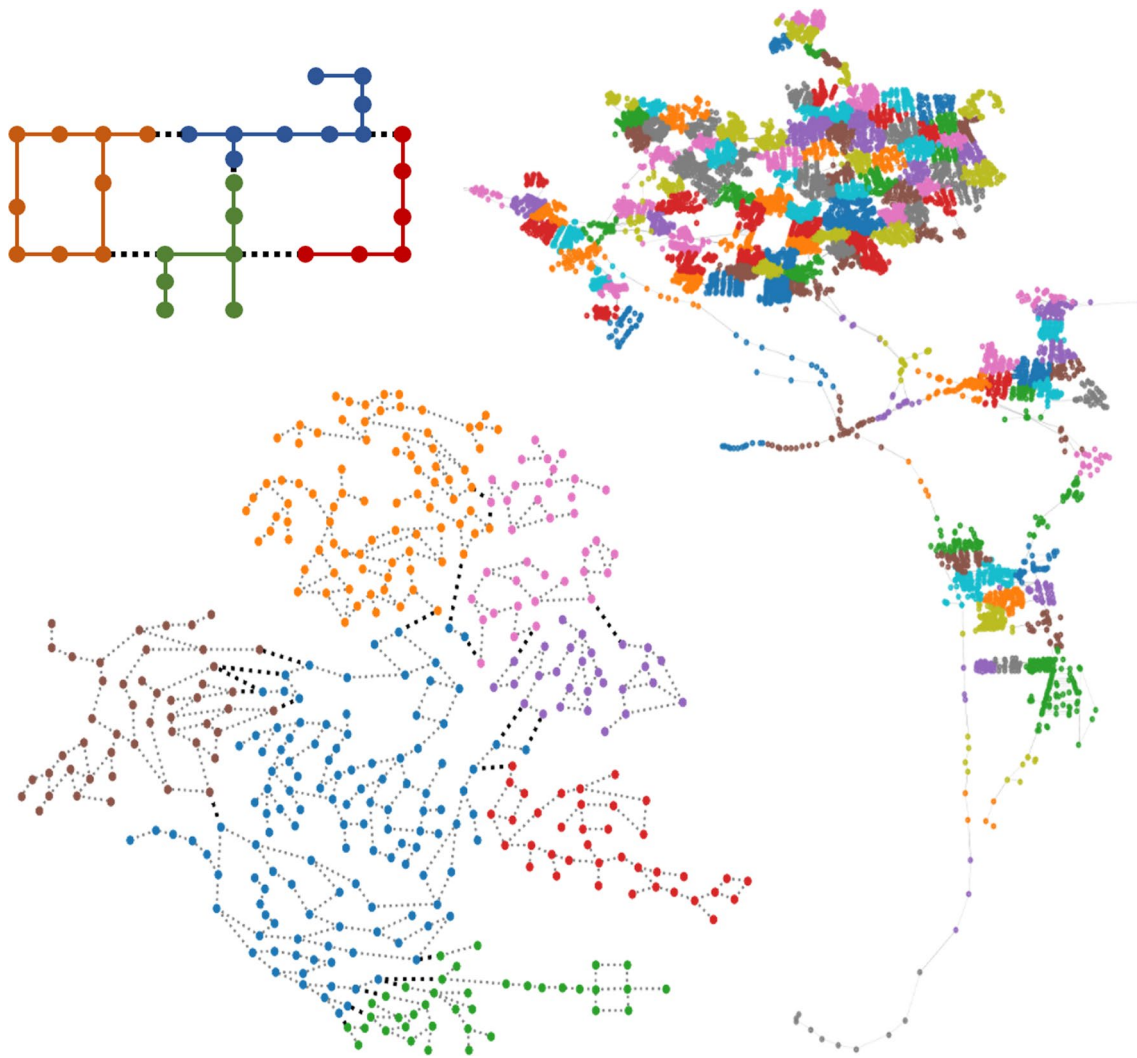


Fig. 4 Clustering results of Hanoi (upper left), C-town (lower left), and E-town (right)

0.06 when the DMA was more than 160, and at approximately 0.39 when the DMA number was more than 170. After a comprehensive consideration, 160 was chosen as the most appropriate number of DMAs for E-town (LocDiff = 0.06, DemDiff = 0.40, and $N_B = 1348$). The results of C-town and E-town show that each DMA should connect 2000–2500 consumers as the optimal choice, although this is based on the limited experimental data quantity. This indicates that an appropriately sized DMA optimizes the integral properties of the network and prevents the negative effects of numerous boundary pipes on subsequent divisions (Morrison et al. 2007).

Influence of aggregation range

The influence of aggregation range on clustering was evaluated (Fig. 6). A significant characteristic of the GANDMA method is that it allows each node to aggregate the features

of its neighbors; this step was repeated several times to make each node aggregate the features of its multi-order neighbors. Thus, the effect of the aggregation range in the GANDMA method is significant and should be evaluated. For the three networks, the DMA numbers were set to 4, 7, and 160, respectively. The aggregation ranges were set to 1–8 to study the changes in clustering results following the changes in aggregation range. Table 4 shows that for the Hanoi network, both LocDiff and DemDiff values first decrease and remain stable with the increase in the aggregation range; the optimal LocDiff and DemDiff values were obtained in the aggregation ranges 4 and 5. An aggregation range of five is considered the most appropriate choice (LocDiff = 0.12, DemDiff = 0.13, and $N_B = 5$). Similarly, in C-town and E-town, both LocDiffs and DemDiffs first decrease and remain stable. After a comprehensive consideration, the optimal aggregation range of C-town is five (LocDiff = 0.07, DemDiff = 0.27, and $N_B = 22$), and that of E-town is also five (LocDiff =

Table 1 Description of the data

Network	Nodes ID	X-Coord	Y-Coord	Demand	Surrounding nodes
Hanoi	1	5251.17	4485.98	179	2
	2	5251.17	5268.69	247.22	1,14,32
	3	5227.8	5969.63	236.11	2,4
	4	6068.93	5969.63	36.11	3,5
	5	6676.4	5969.63	201.39	4,6,10
	6	7260.51	5969.63	279.17	6,7
...
C-town	1	-247075.89	147770.08	1.076586	2
	2	-247536.76	147694.06	0.5797212	1,3
	3	-247265.94	148064.65	0.1740539	2,4,5
	4	-247508.25	148007.64	0.3090042	3,7
	5	-247641.28	148045.65	0.1294744	3,6
	6	-247042.63	147988.63	0.7212659	5,7,8,11
...
E-town	1	8099.52	17569.15	0	2
	2	6073.94	19530.63	0.011977	1,3
	3	6122.22	19538.3	0.047910	2,4
	4	2328.87	15769.23	0.131753	3,5
	5	2317.52	15772.81	0	4,6
	6	3066.97	16038.31	0.08384	5,7
...

Table 2 Performance of node clustering with different methods in three networks

Network	Method	LocDiff	DemDiff	N_B
Hanoi	GANDMA	0.120667	0.129542	5
	K-Means	0.126084	0.290129	5
	Spectral graph	0.125874	0.22028	5
C-town	GANDMA	0.066669	0.266762	22
	K-Means	0.070306	0.316714	27
	Spectral graph	0.068932	0.31796	21
E-town	GANDMA	0.056741	0.287505	1,126
	K-Means	0.061445	0.350991	1,287
	Spectral graph	0.057961	0.338763	1,187

0.06, DemDiff = 0.29, and $N_B = 1,126$). The results indicate that the optimal aggregation range of the GANDMA method is always five. This is because an extremely small aggregation range makes it difficult to aggregate the features of the surrounding nodes; however, an excessively large aggregation range leads to similar features of irrelevant nodes, blurring the boundaries of clustering (Hamilton et al. 2017).

Performance of boundary division

The clustering stage refers to clustering the demand nodes into some areas, and the division stage refers to the installation of flowmeters or valves in the boundary pipes of the

areas to form DMAs. The locations of the flowmeters and valve installations directly affect the actual operating capacity of the DMAs (Di Nardo et al. 2014). The GANDMA method determines a location for installing flowmeters and valves by evaluating the importance of boundary pipes; a flowmeter is preferentially installed at an important boundary. The importance of the boundary pipes of the three networks was evaluated (Fig. 7). The unimportant boundary pipes were closed successively, and the variation rule of the I_r was recorded and analyzed.

These simulations are performed in EPANET, the water level of the tank must be equal to or higher than the initial water level at the end of the simulation. If the water level is extremely low, a discarded solution is not required; however, pump or valve controls must be adjusted to maintain the water level (Herrera et al. 2010). The results show that when 40% of the boundary pipes of each DMA were closed in Hanoi, the pressure heads h of the 100% water demand nodes were at a reasonable level, and the I_r of the network was 0.35. When the same proportion of pipes was closed in C-town and E-town, the pressure heads h of 97% and 95% of demand nodes were at a reasonable level, and the I_r of networks were 0.37 and 0.41, respectively. Therefore, a 40% closure of boundary pipes is suitable for most networks. Compared to the heuristic method (Zhang et al. 2017), the GANDMA method reduced the damage of I_r by 17.1% when 40% of the boundary pipe network of each DMA was closed. Notably, the GANDMA method clarifies the sequence of

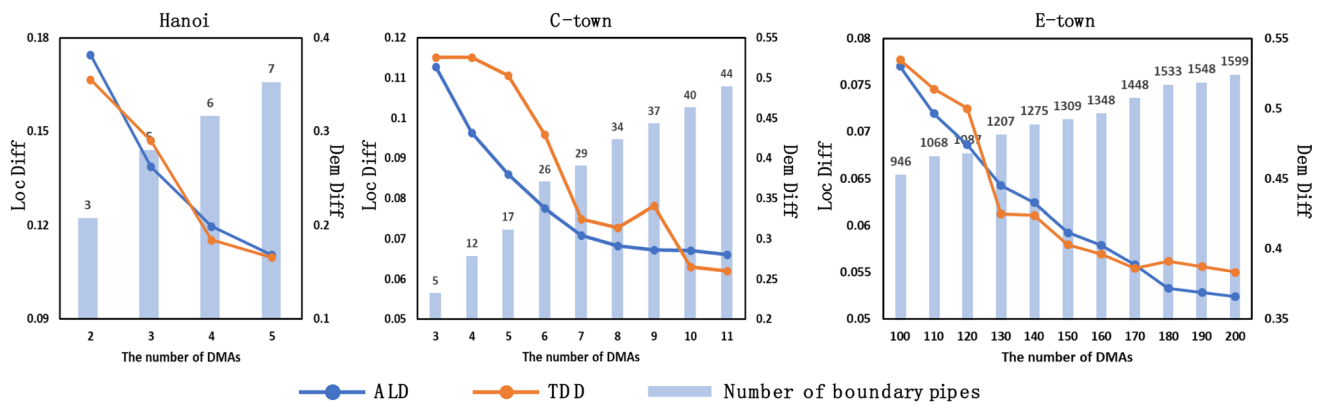


Fig. 5 Influence of number of DMAs on LocDiff, DemDiff, and boundary pipe number in the networks of Hanoi (left), C-town (middle), and E-town (right).

closing boundary pipes, prevents the rationality loss of previous methods and provides a foundation for dynamic boundary control.

Study implications

Clustering of demand nodes and division of boundary pipes to form DMAs is an effective method for managing WDNs (Charalambous 2008). Therefore, this study employs a deep

learning method to optimize the DMAs construction and proposes a network distribution method based on GANDMA for clustering demand nodes and determining the importance of boundary pipes. Specifically, a reasonable clustering is critical and some boundary pipes should be shut down for the overall management because the demand node features are diverse and the connections between nodes are complex. Moreover, a clear evaluative criterion and selection of closed pipes for the establishment of DMAs are indispensable (Brentan et al. 2017). GANDMA aggregated the location feature for each node based on the connection between pipes, and the features were input into the neural network for an automatic clustering to ensure the similarity of nodes located in each area and the balance of total water demand between areas. The importance of the boundary pipes after clustering was evaluated, and unimportant pipes were prioritized for closure. The GANDMA-based node clustering and boundary-division method not only realizes automated WNP with a high reliability but also realizes the importance of boundary pipe evaluation. This addresses a limitation of the traditional method: it does not provide a basis for the opening and closing of boundary pipes. Moreover, it provides reliable conditions for the subsequent dynamic control of DMAs (Giudicianni et al. 2020). In practical application, the structure of the water network and the features of the demand nodes are recorded as data, and the suitable positions of the valve installation are simulated by computer. The workers will install valves or flowmeters to realize partition control of the water network.

Conclusion

WNP to form DMAs is an effective method of water distribution network management, and the rationality of node clustering and division of boundary pipes directly affects the daily operations of DMAs. However, its implementation

Table 3 Effect of number of DMAs on clustering results (aggregation range is set to 0)

Network	DMAs number	LocDiff	DemDiff	N_B
C-town	3	0.112712	0.525631	5
	4	0.096386	0.525438	12
	5	0.085993	0.503174	17
	6	0.077584	0.429723	26
	7	0.070827	0.324558	29
	8	0.068216	0.31376	34
	9	0.067178	0.340748	37
	10	0.067061	0.26529	40
	11	0.066044	0.259774	44
	E-town	100	0.077035	0.534821
110		0.071977	0.514157	1,068
120		0.068638	0.500115	1,087
130		0.064288	0.424952	1,207
140		0.062436	0.424026	1,275
150		0.059247	0.403108	1,309
160		0.057883	0.39627	1,348
170		0.055823	0.386188	1,448
180		0.053298	0.391349	1,533
190		0.052832	0.387483	1,548
200		0.052404	0.383595	1599

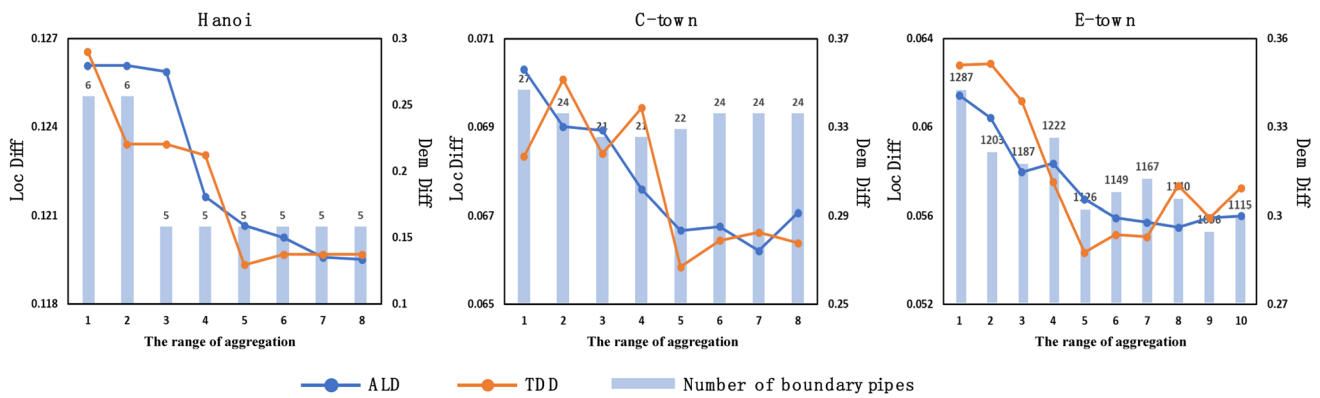


Fig. 6 Influence of aggregation range on LocDiff, DemDiff, and boundary pipe number in the networks of Hanoi (left), C-town (middle), and E-town (right).

is difficult because of the diversity of node features and the complexity of the connections between nodes. Therefore, we proposed the GANDMA method to achieve an automatic

and reasonable clustering of water demand nodes and discover the importance of evaluating and dividing boundary pipes. The main conclusions are as follows:

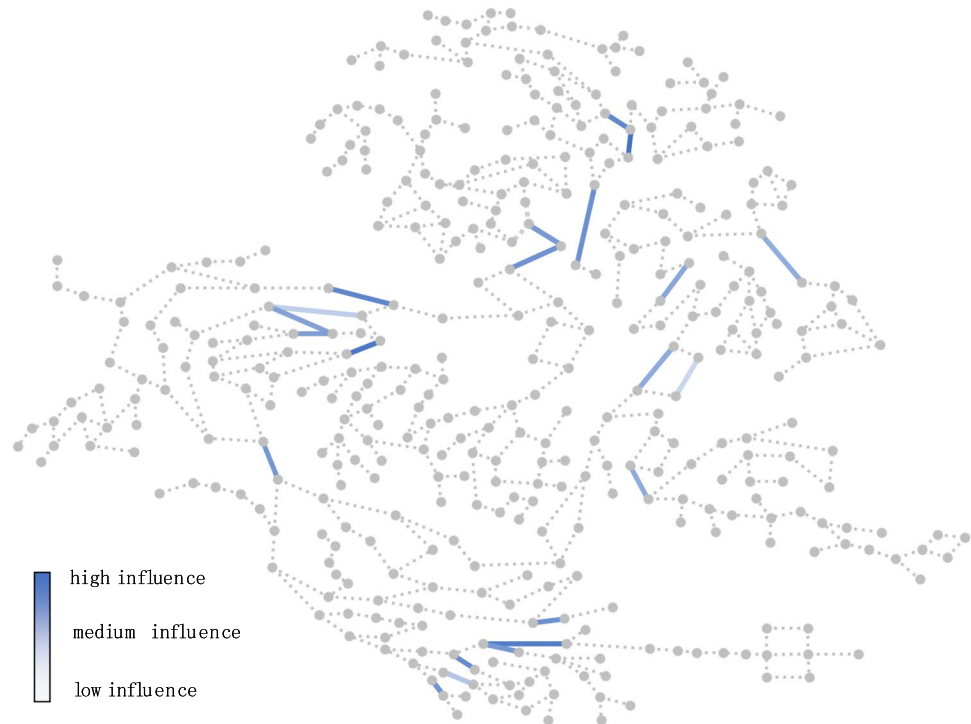
Table 4 Effect of aggregation range on clustering results (DMA numbers to 4, 7, and 160)

Network	DMA numbers	LocDiff	DemDiff	N_B
Hanoi	1	0.126084	0.290129	6
	2	0.126084	0.22028	6
	3	0.125874	0.22028	5
	4	0.121633	0.212022	5
	5	0.120667	0.129542	5
	6	0.120257	0.137211	5
	7	0.119585	0.137211	5
	8	0.119506	0.137211	5
C-town	1	0.070306	0.316714	27
	2	0.06901	0.351561	24
	3	0.068932	0.31796	21
	4	0.067595	0.338657	21
	5	0.066669	0.266762	22
	6	0.066758	0.278827	24
	7	0.066207	0.282487	24
	8	0.067066	0.277661	24
E-town	1	0.061445	0.350991	1,287
	2	0.060411	0.351546	1,203
	3	0.057961	0.338763	1,187
	4	0.058358	0.311454	1,222
	5	0.056741	0.287505	1,126
	6	0.055906	0.293643	1,149
	7	0.055702	0.292866	1,167
	8	0.055469	0.310068	1,140
	9	0.05559	0.299252	1,096
	10	0.055999	0.309362	1,115

- The GANDMA method effectively aggregated the features of demand nodes in water distribution networks and realized automatic node clustering. Experiments on three networks show that the average LocDiff and DemDiff values are 0.08 and 0.23, respectively, which are, respectively, 6% and 55% lower than those of the traditional clustering method.
- The effects of the number of DMAs and the aggregation range of the GANDMA method in the clustering stage were evaluated, and the results showed that both had a significant impact on LocDiff and DemDiff. The clustering result was optimal when the number of connected populations was approximately 2500 in each DMA, and the appropriate aggregation range was five.
- In the division stage, the experimental results showed that the GANDMA method is effective and objectively provides the order of importance of boundary pipes. When 40% of the relatively unimportant pipe were closed, the I_r decreased by 11.9% on an average, and 97% of the demand points satisfied the lowest pressure head. Compared to the traditional method, the GANDMA method reduced the impact of the network by 17.1%.

Although GANDMA performs well in WNP, some problems remain unresolved. Similar to the traditional models, the GANDMA model realizes water network partitioning by clustering and dividing in two stages, which reduces the search range of the optimal solution. Therefore, future studies should focus on the development of a one-step water network-partitioning method. Moreover, a few dynamic DMAs exist. GANDMA's boundary-importance evaluation provides a basis for setting dynamic DMAs, and this should be explored further in the future.

Fig. 7 The result of influence calculation of boundary pipes for C-town



Acknowledgement The authors would like to reveal their appreciation and gratitude to the respected reviewers and editors for their constructive comments. In addition, the authors would like to acknowledge the support received by the Natural Science Foundation of Zhejiang Province (Grant No. LGG20F030009). Finally, an admirable appreciation is keen to the King Fahd University of Petroleum and Minerals, for their technical support.

Declaration

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alvisi S (2015) A new procedure for optimal design of district metered areas based on the multilevel balancing and refinement algorithm. *Water Resour Manag* 29:4397–4409
- Alvisi S, Franchini M (2014) A procedure for the design of district metered areas in water distribution systems. *Procedia Eng* 70:41–50. <https://doi.org/10.1016/j.proeng.2014.02.006>
- Bhagat SK, Tiyasha Welde W et al (2019) Evaluating physical and fiscal water leakage in water distribution system. *Water* (Switzerland). <https://doi.org/10.3390/w11102091>
- Brentan BM, Luvizotto E, Montalvo I et al (2017) Near real time pump optimization and pressure management. *Procedia Eng* 186:666–675. <https://doi.org/10.1016/j.proeng.2017.06.248>
- Brentan B, Campbell E, Goulart T et al (2018) Social network community detection and hybrid optimization for dividing water supply into district metered areas. *J Water Resour Plan Manag.* [https://doi.org/10.1061/\(asce\)wr.1943-5452.0000924](https://doi.org/10.1061/(asce)wr.1943-5452.0000924)
- Charalambous B (2008) Use of district metered areas coupled with pressure optimisation to reduce leakage. *Water Supply* 8:57–62. <https://doi.org/10.2166/ws.2008.030>
- Choudhary M, Chouhan SS, Pilli ES, Vipparthi SK (2021) BerConvoNet: a deep learning framework for fake news classification. *Appl Soft Comput* 110:107614. <https://doi.org/10.1016/j.asoc.2021.107614>
- Ciaponi C, Murari E, Todeschini S (2016) Modularity-based procedure for partitioning water distribution systems into independent districts. *Water Resour Manag* 30:2021–2036. <https://doi.org/10.1007/s11269-016-1266-1>
- Di Nardo A, Di Natale M (2011) A heuristic design support methodology based on graph theory for district metering of water supply networks. *Eng Optim* 43:193–211. <https://doi.org/10.1080/03052151003789858>
- Di Nardo A, Di Natale M, Di Mauro A (2013) Case study: Monteruscio network. In: Di Nardo A, Di Natale M, Di Mauro A (eds) *Water supply network district metering*. Springer Vienna, Vienna, pp 39–85. https://doi.org/10.1007/978-3-7091-1493-3_4
- Di Nardo A, Di Natale M, Santonastaso GF et al (2014) Divide and conquer partitioning techniques for smart water networks. *Procedia Eng* 89:1176–1183. <https://doi.org/10.1016/j.proeng.2014.11.247>

- Di Nardo A, Di Natale M, Giudicianni C et al (2015) Water distribution system clustering and partitioning based on social network algorithms. *Procedia Eng* 119:196–205. <https://doi.org/10.1016/j.proeng.2015.08.876>
- Di Nardo A, Di Natale M, Gargano R et al (2018) Performance of partitioned water distribution networks under spatial-temporal variability of water demand. *Environ Model Softw* 101:128–136. <https://doi.org/10.1016/j.envsoft.2017.12.020>
- Diao K, Zhou Y, Rauch W (2013) Automated creation of district metered area boundaries in water distribution systems. *J Water Resour Plan Manag* 139:184–190. [https://doi.org/10.1061/\(asce\)wr.1943-5452.0000247](https://doi.org/10.1061/(asce)wr.1943-5452.0000247)
- Farmani R, Walters G, Savic D (2006) Evolutionary multi-objective optimization of the design and operation of water distribution network: total cost vs. reliability vs. water quality. *J Hydroinformatics* 8:165–179. <https://doi.org/10.2166/hydro.2006.019b>
- Ferrari G, Savic D, Becciu G (2014) Graph-theoretic approach and sound engineering principles for design of district metered areas. *J Water Resour Plan Manag*. [https://doi.org/10.1061/\(asce\)wr.1943-5452.0000424](https://doi.org/10.1061/(asce)wr.1943-5452.0000424)
- Gajghate PW, Mirajkar A, Shaikh U et al (2021) Optimization of layout and pipe sizes for irrigation pipe distribution network using Steiner point concept. *Math Probl Eng*. <https://doi.org/10.1155/2021/6657459>
- Giudicianni C, Di Nardo A, Di Natale M et al (2018) Topological taxonomy of water distribution networks. *Water* 10:444. <https://doi.org/10.3390/w10040444>
- Giudicianni C, Herrera M, di Nardo A, Adeyeye K (2020) Automatic multiscale approach for water networks partitioning into dynamic district metered areas. *Water Resour Manag* 34:835–848. <https://doi.org/10.1007/s11269-019-02471-w>
- Giustolisi O, Ridolfi L (2014) A novel infrastructure modularity index for the segmentation of water distribution networks. *Water Resour Res* 50:7648–7661. <https://doi.org/10.1002/2014wr016067>
- Gomes R, Sá Marques A, Sousa J (2011) Estimation of the benefits yielded by pressure management in water distribution systems. *Urban Water J* 8:65–77. <https://doi.org/10.1080/1573062x.2010.542820>
- Grayman WM, Murray R, Savic DA (2009) Effects of redesign of water systems for security and water quality factors. In: *World environmental and water resources congress 2009. Great Rivers*, pp 1–11
- Gupta A, Bokde N, Kulat K, Yaseen ZM (2020) Nodal matrix analysis for optimal pressure-reducing valve localization in a water distribution system. *Energies* 13:1878
- Hajebi S, Temate S, Barrett S et al (2014) Water distribution network sectorisation using structural graph partitioning and multi-objective optimization. *Procedia Eng* 89:1144–1151. <https://doi.org/10.1016/j.proeng.2014.11.238>
- Herrera M, Izquierdo J, Pérez-García R, Montalvo I (2012) Multi-agent adaptive boosting on semi-supervised water supply clusters. *Adv Eng Softw* 50:131–136. <https://doi.org/10.1016/j.advengsoft.2012.02.005>
- Herrera M, Abraham E, Stoianov I (2016) A graph-theoretic framework for assessing the resilience of sectorised water distribution networks. *Water Resour Manag* 30:1685–1699. <https://doi.org/10.1007/s11269-016-1245-6>
- Herrera M, Canu S, Karatzoglou A, et al (2010) An approach to water supply clusters by semi-supervised learning
- Izquierdo J, Herrera M, Montalvo I, Pérez-García R (2009) Division of water supply systems into district metered areas using a multi-agent based approach. In: *International conference on software and data technologies*. Springer, pp 167–180
- Laucelli DB, Simone A, Berardi L, Giustolisi O (2016) Optimal design of district metering areas. *Procedia Eng* 162:403–410. <https://doi.org/10.1016/j.proeng.2016.11.081>
- Lifshitz R, Ostfeld A (2018) District metering areas and pressure reducing valves trade-off in water distribution system leakage management. In: *WDSA/CCWI joint conference proceedings*
- Liu J, Han R (2018) Spectral clustering and multicriteria decision for design of district metered areas. *J Water Resour Plan Manag*. [https://doi.org/10.1061/\(asce\)wr.1943-5452.0000916](https://doi.org/10.1061/(asce)wr.1943-5452.0000916)
- Morrison J, Tooms S, Rogers D (2007) DMA management guidance notes version 1. In: Brothers KJ, Eng P (eds) *Water loss task force, vol 1*. IWA, London, pp 25–36
- Nicolini M, Zovatto L (2009) Optimal location and control of pressure reducing valves in water networks. *J Water Resour Plan Manag* 135:178–187. [https://doi.org/10.1061/\(asce\)0733-9496\(2009\)135:3\(178\)](https://doi.org/10.1061/(asce)0733-9496(2009)135:3(178))
- Oyedele A, Ajayi A, Oyedele LO et al (2021) Deep learning and boosted trees for injuries prediction in power infrastructure projects. *Appl Soft Comput* 110:107587. <https://doi.org/10.1016/j.asoc.2021.107587>
- Pesantez JE, Berglund EZ, Mahinthakumar G (2019) Multiphase procedure to design district metered areas for water distribution networks. *J Water Resour Plan Manag*. [https://doi.org/10.1061/\(asce\)wr.1943-5452.0001095](https://doi.org/10.1061/(asce)wr.1943-5452.0001095)
- Puust R, Kapelan Z, Savic DA, Koppel T (2010) A review of methods for leakage management in pipe networks. *Urban Water J* 7:25–45
- Rakhshani H, Idoumghar L, Ghambari S et al (2021) On the performance of deep learning for numerical optimization: An application to protein structure prediction. *Appl Soft Comput* 110:107596. <https://doi.org/10.1016/j.asoc.2021.107596>
- Saldarriaga J, Bohorquez J, Celeita D et al (2019) Battle of the water networks district metered areas. *J Water Resour Plan Manag*. [https://doi.org/10.1061/\(asce\)wr.1943-5452.0001035](https://doi.org/10.1061/(asce)wr.1943-5452.0001035)
- Sela Perelman L, Allen M, Preis A et al (2015) Automated sub-zoning of water distribution systems. *Environ Model Softw* 65:1–14. <https://doi.org/10.1016/j.envsoft.2014.11.025>
- Sharma AN, Dongre SR, Gupta R, Ormsbee L (2022) Multiphase procedure for identifying district metered areas in water distribution networks using community detection, NSGA-III optimization, and multiple attribute decision making. *J Water Resour Plan Manag*. [https://doi.org/10.1061/\(asce\)wr.1943-5452.0001586](https://doi.org/10.1061/(asce)wr.1943-5452.0001586)
- Taha AF, Wang S, Guo Y et al (2021) Revisiting the water quality sensor placement problem: optimizing network observability and state estimation metrics. *J Water Resour Plan Manag*. [https://doi.org/10.1061/\(asce\)wr.1943-5452.0001374](https://doi.org/10.1061/(asce)wr.1943-5452.0001374)
- Tiernan ED, Hodges BR (2022) A topological approach to partitioning flow networks for parallel simulation. *J Comput Civ Eng*. [https://doi.org/10.1061/\(asce\)cp.1943-5487.0001020](https://doi.org/10.1061/(asce)cp.1943-5487.0001020)
- Todini E (2000) Looped water distribution networks design using a resilience index based heuristic approach. *Urban Water* 2:115–122. [https://doi.org/10.1016/s1462-0758\(00\)00049-2](https://doi.org/10.1016/s1462-0758(00)00049-2)
- Twort AC, Ratnayaka DD, Brandt MJ (2000) *Water supply*. Elsevier, Amsterdam
- Tzatchkov VG, Alcocer-Yamanaka VH, Bourguett Ortíz V (2008) Graph theory based algorithms for water distribution network sectorization projects. In: *Water distribution systems analysis symposium 2006*
- Wu Y, Liu S, Wu X et al (2016) Burst detection in district metering areas using a data driven clustering algorithm. *Water Res* 100:28–37. <https://doi.org/10.1016/j.watres.2016.05.016>
- Yang T, Yu X, Ma N et al (2022) An autonomous channel deep learning framework for blood glucose prediction. *Appl Soft Comput* 120:108636. <https://doi.org/10.1016/j.asoc.2022.108636>

Zhang Q, Wu ZY, Zhao M et al (2017) Automatic partitioning of water distribution networks using multiscale community detection and multiobjective optimization. *J Water Resour Plan Manag.* [https://doi.org/10.1061/\(asce\)wr.1943-5452.0000819](https://doi.org/10.1061/(asce)wr.1943-5452.0000819)

Zhang T, Yao H, Chu S et al (2021) Optimized DMA partition to reduce background leakage rate in water distribution networks.

J Water Resour Plan Manag. [https://doi.org/10.1061/\(asce\)wr.1943-5452.0001465](https://doi.org/10.1061/(asce)wr.1943-5452.0001465)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.