




Improved community detection for WDN partitioning in the dual topology based on segments and valves

Enrico Creaco ^{*}, Carlo Giudicianni  and Amirabbas Mottahedin 

Department of Civil Engineering and Architecture, University of Pavia, 27100 Pavia, Italy

*Corresponding author. E-mail: creaco@unipv.it

 EC, 0000-0003-4422-2417; CG, 0000-0002-4082-2557; AM, 0000-0001-7057-8770

ABSTRACT

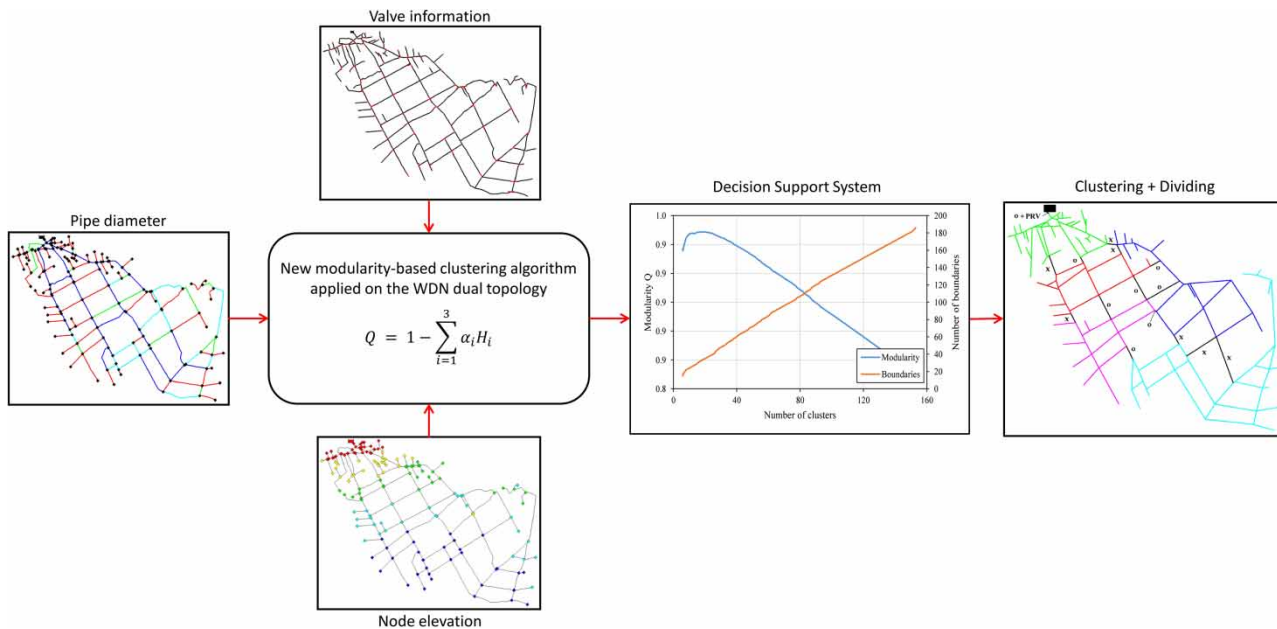
This paper proposes a formulation of modularity tailored to the dual water distribution network (WDN) topology based on segments and valves, to be conveniently adopted for the partitioning into district-metered areas (DMAs). Notably, it allows considering both properties to be made uniform across DMAs, such as water demand or total pipe length, and properties to be made uniform inside each DMA, such as nodal ground elevations or pipe age for the sake of pressure regulation or maintenance easiness, respectively. This paper also proposes a new algorithm for the identification of the optimal clustering of WDN segments into any desired number of DMAs. Taking as a starting point any WDN clustering solution, i.e., the solution obtained with Newman's fast algorithm for community detection, the novel algorithm operates by exploring changes in the community of belonging to segments lying in the boundary between adjacent communities, by applying an optimization inspired by the simulated annealing technique. The applications of the novel modularity formulation and optimization algorithm to two case studies yield well-performing clustering solutions in terms of engineering judgment criteria, such as the low number of inter-DMA boundary pipes, uniformity of DMAs and hydraulic performance.

Key words: community, optimization, partitioning, simulated annealing, water distribution networks, WDN

HIGHLIGHTS

- The clustering of water distribution networks into district-metered areas (DMAs) is tackled in an innovative way.
- An improved modularity function using the dual network topology based on segments and valves is proposed.
- The uniformity of properties across and inside DMAs is considered.
- A novel optimization algorithm inspired by the simulated annealing approach proves effective and efficient.

GRAPHICAL ABSTRACT



1. INTRODUCTION

One of the most explored topics in the recent scientific literature of water distribution networks (WDNs) is partitioning, which enables subdividing large networks into sufficiently small parts, typically called district-metered areas (DMAs), which can be easily monitored and managed by water utilities in terms of users' consumption and leakage (Morrison *et al.* 2007). This also results in fast detection of anomalies, e.g., unexpectedly high consumption and pipe bursts, and in more effective maintenance campaigns. WDN partitioning has also been implemented in commercial software over the last decade (e.g., WaterGEMS).

In most cases, the partitioning is carried out in two phases, i.e., clustering and dividing (e.g., Di Nardo *et al.* 2018; Giudicianni *et al.* 2020). The clustering is typically performed by graph theory techniques and aims to identify the size and extension of DMAs, as well as the location of boundary pipes. As the final aim of partitioning is the real-time knowledge of water exchanges between DMAs to facilitate water mass balance at the WDN level, the dividing follows to identify the boundaries to be closed or fitted with a flow meter, to have known water exchanges equal to 0 or measured water exchanges, respectively.

A widely adopted concept in the context of WDN partitioning is modularity, which was first defined by Clauset *et al.* (2004) in the context of complex network theory to assess the goodness of network subdivision into communities, associated with balanced cluster sizes and few boundary links. While the concepts of modularity and community detection have been adopted in numerous algorithms for WDN partitioning (Campbell *et al.* 2016; Zhang *et al.* 2017; Brentan *et al.* 2018; Rahmani *et al.* 2018; Creaco & Haidar 2019; Creaco *et al.* 2019; Yao *et al.* 2021; Sharma *et al.* 2022), Giustolisi & Ridolfi (2014a, 2014b) pointed out that the original formulation of modularity falls short of the representation of hydraulic properties for WDNs, e.g., water demand, pressure, pipe length, and background leakage, thus proposing a revisited one. Another formulation of modularity tailored to WDNs has recently been proposed by Shao *et al.* (2022), in which the modularity component associated with boundary pipes is independent from that associated with the water demand to be made uniform across DMAs. Though being significant contributions to this research field, the revisited formulations of modularity proposed so far fail to consider the actual distribution of isolation valves, which subdivides the WDN into segments. Incidentally, segments are the smallest WDN parts that can be disconnected for maintenance purposes. As Santonastaso *et al.* (2019) pointed out, DMAs must originate from the merging of WDN segments and therefore WDN partitioning algorithms must be formulated based on the dual topology (Walski 1993) based on segments and valves present in the WDN.

To bridge this research gap, the present work aims to propose a new formulation of modularity specifically tailored to WDNs and based on the dual segments/valves topology of the WDN. As an additional improvement from the previous works, the new formulation can also deal with properties to be made uniform inside DMAs, rather than only across DMAs. This is the case with nodal ground elevations or pipe age, the uniformity of which inside each DMA is a desired condition to facilitate pressure regulation/leakage mitigation or pipe maintenance, respectively. Furthermore, an optimization algorithm inspired by the simulated annealing (Laarhoven & van Peter 1987) approach is proposed for the clustering of WDN segments into the desired number of DMAs.

The following sections describe the methodology and its applications to two case studies. The paper ends with the discussion and the conclusions.

2. METHODOLOGY

The following subsections describe the methodological elements of the paper (Figure 1). First, the dual topology based on segments and isolation valves and its significance in the clustering are described, followed by the presentation of the novel formulation of modularity. Then, the fast greedy algorithm (FGA) is presented for deriving the initial clustering solution of the first attempt. Finally, the optimization algorithm for the maximization of the modularity is described.

2.1. Dual topology based on segments and isolation valves

The new clustering algorithm is based on the dual topology based on segments and isolation valves (Walski 1993), instead of the traditional topology based on nodes and pipes. The disconnection of the generic segment can be obtained by closing the isolation valves that separate it from its adjacent segments. In this context, the reason for the choice of the dual topology is that its adoption guarantees that isolation valves are present in the boundaries between DMAs after the clustering, as was originally proven by Santonastaso *et al.* (2019) and later considered by other authors, such as Creaco *et al.* (2022). The use of the dual topology brings about the two following advantages, which make the clustering solutions obtained more feasible:

- In the boundaries to be closed during the dividing phase, there is no need for installing additional isolation valves.
- In the boundaries to be left open during the dividing phase, the valve vault can sometimes be used to host the flow meter, which must be installed for monitoring the inter-DMA exchange of water, entailing other potential money savings.

Figure 2 shows an example of WDN layout in the traditional and dual topology. The presence of 10 isolation valves causes the WDN to be subdivided into eight segments. The composition of these segments in terms of nodes and pipes is shown in Table 1. Table 2 shows the segments interconnected by each isolation valve, indicated as upstream and downstream segments.

To give evidence that any kind of clustering of segments on the dual topology always leaves isolation valves at the boundary, an example of segment clustering into four DMAs is provided in Figure 3. DMA1, which includes segments S1 and S2, is connected to DMA2, made up of segments S3 and S4, by means of valve V2. DMA2 borders with DMA3, made up of segments S5 and S7, with valves V5 and V6 as boundaries. Finally, valve V9 connects DMA3 to DMA4, which includes segments S6 and S8.

2.2. Modularity index

The modularity index Q was first proposed by Newman & Girvan (2004) to express the goodness of network subdivision into communities. In recent years, various authors have proposed formulations for tailoring it to WDNs (Giustolisi & Ridolfi 2014a, 2014b; Shao *et al.* 2022). However, they all fall short of meeting the dual topology. Furthermore, they only consider properties to be made uniform across DMAs, such as demand and pipe length, while neglecting properties to be made uniform inside DMAs, such as nodal ground elevations or pipe age for the sake of pressure regulation or maintenance's effectiveness, respectively. To solve these drawbacks, a new formulation of the modularity index considering the dual

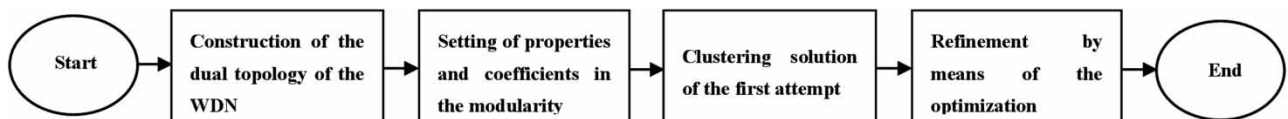


Figure 1 | Flowchart of the methodology.

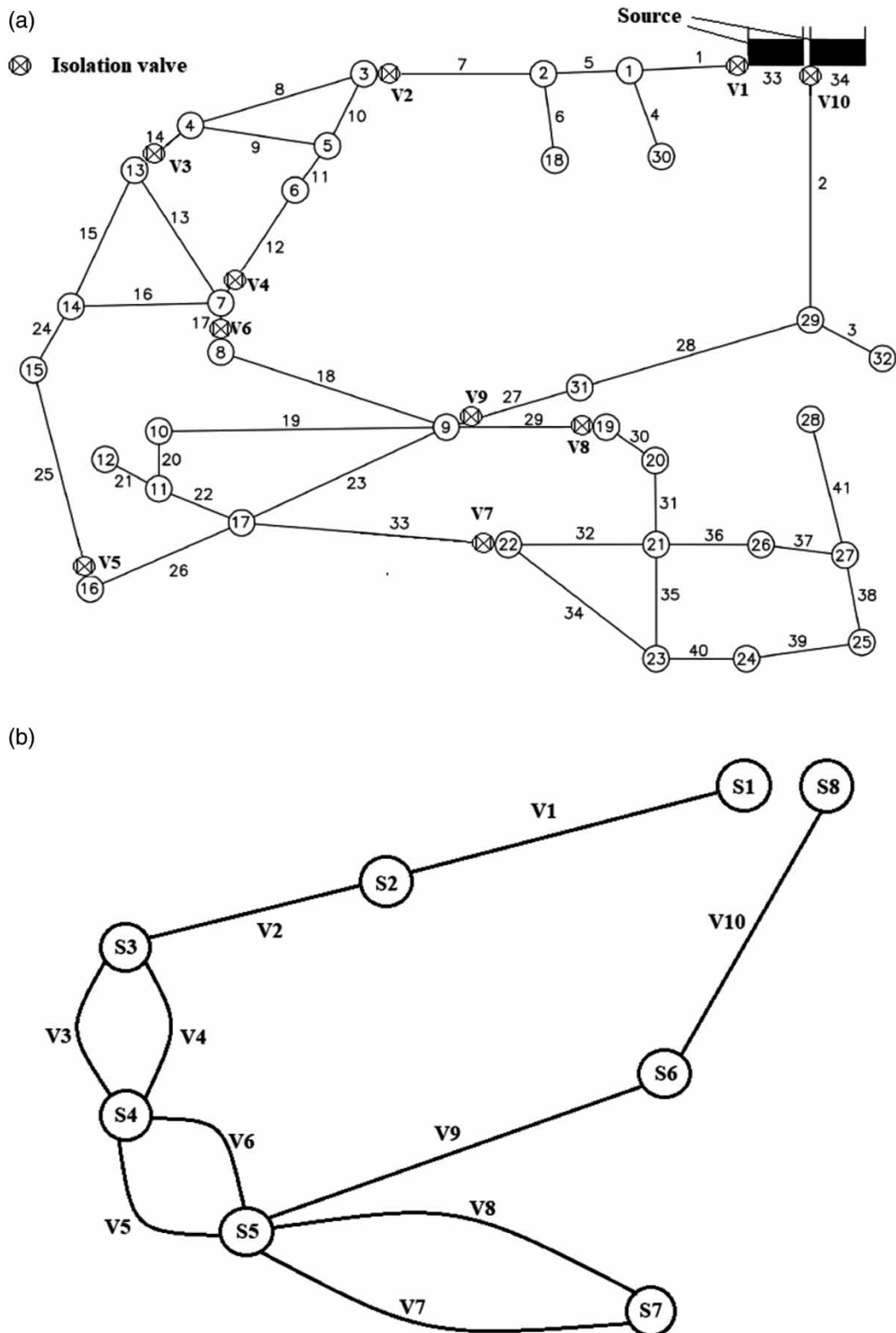


Figure 2 | Traditional (a) and dual (b) topology for the WDN of Santa Maria di Licodia, adapted from *Creaco et al. (2022)*. See text for description of the figure.

Table 1 | Composition of segments (for the WDN of Figure 2) in terms of nodes and pipes

Segment	S1	S2	S3	S4	S5	S6	S7	S8
Nodes	33	1, 2, 18, 30	3, 4, 5, 6	7, 13, 14, 15	8, 9, 10, 11, 12, 16, 17	29, 31, 32	19, 20, 21, 22, 23, 24, 25, 26, 27, 28	34
Pipes	/	1, 4, 5, 6, 7	8, 9, 10, 11, 12, 14	13, 15, 16, 17, 24, 25	18, 19, 20, 21, 22, 23, 26, 33	2, 3, 27, 28	30, 31, 32, 34, 35, 36, 37, 38, 39, 40, 41	/

Table 2 | Upstream and downstream segments (for the WDN of Figure 2) for each valve

Valve	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
Upstream segment	S1	S2	S3	S3	S4	S4	S5	S5	S5	S6
Downstream segment	S2	S3	S4	S4	S5	S5	S7	S7	S6	S8

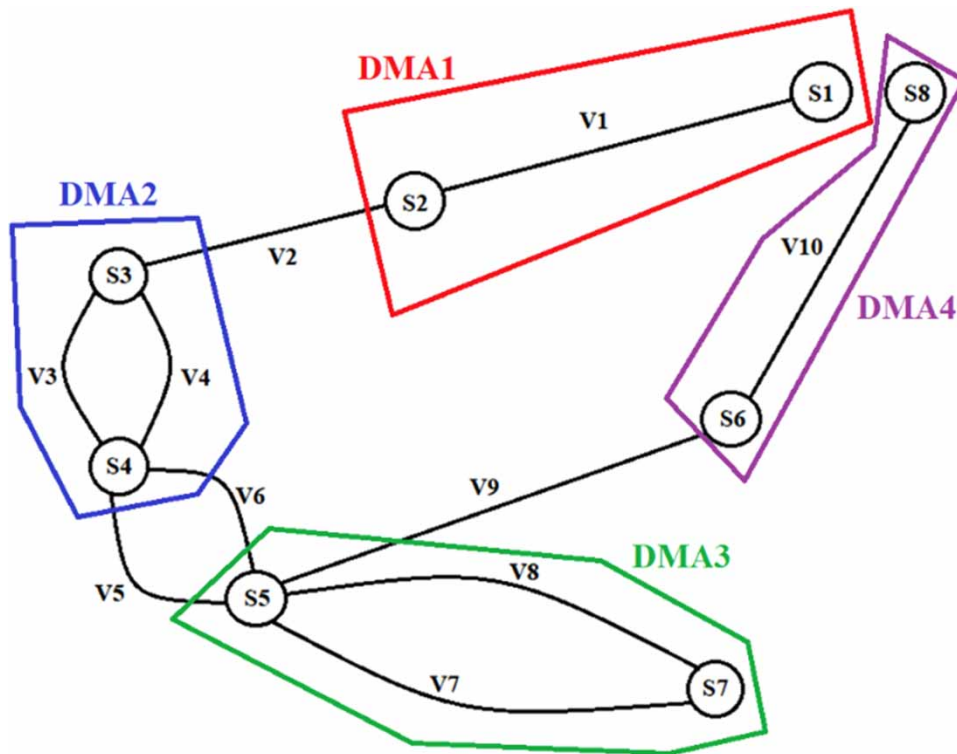


Figure 3 | For the WDN in Figure 2, clustering of segments into four DMAs.

topology based on segments and isolation valves is proposed as:

$$Q = 1 - \sum_{i=1}^3 \alpha_i H_i \tag{1}$$

As Equation (1) shows, Q is calculated by subtracting three contributions H from 1, each of which is multiplied by a positive coefficient α . The contribution H_1 is associated with the boundary pipes. Considering a total number n_v of isolation valves in

the WDN and a total number n_b of isolation valves at the boundaries between DMAs, it can be expressed as:

$$H_1 = \frac{\sum_{i=1}^{n_b} p_i}{\sum_{i=1}^{n_v} p_i} \quad (2)$$

where p_i is a generic attribute of the i -th isolation valve, as an example, the diameter of the valve. If p_i is assumed to be equal to 1 for all isolation valves, H_1 becomes the ratio of n_b to n_v . In this case, the lower n_b , the smaller H_1 . H_1 grows in the presence of numerous boundary isolation valves.

The second contribution H_2 is a property heterogeneity to be minimized across DMAs. Its minimization enables clustered DMAs to feature a similar value for the property, e.g., similar total demand, total number of users or total pipe length, and so forth. If U is the generic property to be made uniform across the DMAs and if the WDN is subdivided into M DMAs, H_2 can be calculated as:

$$H_2 = \frac{1}{U_{\text{tot}}^2} \sum_{i=1}^M (U_i)^2 \quad (3)$$

In Equation (3), U_i is the property value in the generic i -th DMA, obtained as the sum of the $n(i)$ segment values inside, and $U_{\text{tot}} = \sum_{i=1}^M U_i$ is the total value of the property in the WDN. As an example, by considering the DMA demand D_i and the total WDN demand D_{tot} in place of U_i and U_{tot} , respectively, the property to be made uniform is the demand. The lower the difference in the demand between the DMAs, the lower the value of H_2 . When all the M DMAs feature the same value of demand, the value of H_2 is the smallest. It grows, instead, when the values of DMA demands are dissimilar.

H_3 relates, instead, to a property heterogeneity inside each of the DMAs. Its minimization enables all segments inside each of the DMAs to feature a similar value for the property, e.g., pipe age or nodal elevation. If u is the generic property to be made uniform inside each of the M DMAs, H_3 can be calculated as follows:

$$H_3 = \frac{\sum_{i=1}^M \sum_{j=1}^{n(i)} [(u_{i,j} - \bar{u}_i)^2 / n(i)]}{(u_{\text{max}} - u_{\text{min}})^2}, \quad (4)$$

where $u_{i,j}$ is the average property value in the j -th segment belonging to the i -th DMA; \bar{u}_i is instead the average property value in the i -th DMA, estimated as the average of the property values in the $n(i)$ belonging segments. Furthermore, u_{max} and u_{min} are the maximum and minimum segment property values in the WDN, respectively. As an example, by considering the segment ground elevation $z_{i,j}$, the average elevation \bar{z}_i in the i -th DMA and the maximum and minimum segment elevations z_{max} and z_{min} , in place of $u_{i,j}$, \bar{u}_i and u_{max} and u_{min} , respectively, the property to be made uniform inside each DMA is the ground elevation. Like H_2 , H_3 is small when the segments inside the generic DMA have similar elevations. Otherwise, it tends to increase. Equation (4) can also be applied to pipe age, the uniformity of which inside each DMA facilitates maintenance campaigns or to other variables of interest.

The three contributions subtracted from 1 in Equation (1) are all positive. Therefore, index Q cannot exceed 1. It will be close to 1 in the case of good clustering, in which the three contributions H are small. Otherwise, it will decrease from 1. Finally, it must be remarked that coefficients α play the role of weights, to give leverage to the three elements of modularity in the context of the clustering. Indeed, though all values can be assigned to the weights, the maximization of the modularity in an optimization framework is only affected by their mutual relationship. Hereinafter, weight coefficient values with a sum equal to 2 are considered. Since the generic weight can then range from 0 to 2, a low value close to 0 will be assigned to decrease the leverage of the corresponding contribution of the modularity. A high value close to 2 will, instead, be used to increase its leverage. As general rules for assigning these values do not exist, a sensitivity analysis must be carried out to analyze the clustering solutions obtained in each case study for different triplets of weights. Even if the number M of DMAs can be preventively assumed, based on the desired partitioning of the whole demand, the sensitivity analysis can also be extended to assess the impact of this parameter. In comparison with most methodologies present in the scientific literature, which are

based on fixed parameters, the variable parameters considered in the present methodology can give flexibility to the clustering, while accounting for various issues such as those related to topology, demands and ground elevations. The various clustering configurations obtainable as the values of weights α and parameter M change must be assessed based on topological, hydraulic and engineering judgment criteria, leading to the choice of the ultimate solution. While the results of the clustering can be of assistance to water utility managers, the ultimate solution may be edited to better match the presence of geographic/natural constraints related to, as an example, water bodies, major limited access highways, railroad tracks and parks/golf courses present in the territory.

2.3. Initial clustering solution of the first attempt

The clustering algorithm proposed in the present work enables refining a clustering solution of the first attempt featuring the desired number M of DMAs. A good solution of the first attempt can be obtained by applying the FGA of Newman & Girvan (2004) to the dual segment/valve topology (e.g., see Figure 2(b)). Initially, the number of DMAs is therefore equal to the number N_s of segments, in other words, each segment represents a DMA of its own. Then, all possible merging actions between adjacent DMAs are considered in step 1. These actions are easily identified based on the topological connections in the dual segment/valve topology. As an example, Figure 2(b) shows that segments S1 and S2 can be merged because they are adjacent. S1 cannot be merged, instead, with segment S8 because no valve connects these two segments. For each possible merging in the network, the associated variation ΔQ is calculated. After testing all possible merging actions, the one yielding the largest positive value of ΔQ is applied at the end of step 1, producing a new WDN clustering into a number of DMAs equal to $N_s - 1$. Starting from this new clustering configuration, step 2 explores all potential merging actions, to identify that yielding the largest positive variation ΔQ , which will be the basis for step 3, and so forth. The steps are repeated $N_s - M$ times until the number of DMAs equals M . Although being very effective, FGA does not reach the optimal WDN clustering into M DMAs in terms of Q . In fact, considering the largest ΔQ value at each step does not guarantee that the highest possible value of Q is reached at the $(N_s - M)$ -th step. As there is no guarantee that a sequence of optimal steps results in the optimal final value of Q , the solution obtained with FGA lends itself to being refined, as is explained below in Section 2.4.

2.4. Optimization of the clustering

The clustering is carried out to maximize the modularity expressed with Equation (1), that is:

$$\max(Q) \quad (5)$$

The decision variables of the problem are the DMAs the WDN segments belong to. If C_i , with $i = 1, \dots, N_s$ number of segments in the WDN, is the DMA index the i -th segment belongs to, C_i can range between 1 and M , with M being the total number of DMAs considered in the clustering. The following constraints are considered in the optimization:

$$\forall \text{ DMA } j = 1, \dots, M \quad \sum_{i=1}^{N_s} \delta(C_i, j) \geq 1 \quad (6)$$

$$\forall \text{ segment } i, \text{ segment } j : \delta(C_i, C_j) = 1, \quad \text{shortest path } (i, j) < \infty \quad (7)$$

in which δ represents the Kronecker delta function, equal to 1 or 0 when its arguments are equal or different, respectively. The constraints in Equations (6) and (7) enforce that the generic DMA contains at least one segment and that the segments inside each DMA are physically interconnected, respectively.

The optimization algorithm considered in the present work is inspired by the simulated annealing (Laarhoven & van Peter 1987) approach and operates by refining the WDN clustering solution obtained with FGA, or alternatively a generic configuration of the first attempt indicated as C_0 , in which the constraints (6) and (7) hold. In the context of the present methodology, a boundary segment is a segment connected to a valve lying in the boundary between two adjacent DMAs. At the generic iteration, the algorithm i) explores the potential exchanges of boundary segments between adjacent DMAs, while ii) preventing the transfer of segments from DMAs made up of only one segment. The actions (i) and (ii) enable constraints (6) and (7) to hold and the number M of DMAs to be unaltered during the whole optimization process.

The following subsections explain the operation of the optimization algorithm at the generic iteration and the framework of iterations, respectively.

2.4.1. Operation of the optimization algorithm at the generic iteration

At the generic n -th iteration of this algorithm, all potential exchanges of boundary segments starting from the clustering configuration C_{n-1} resulting from the previous $(n - 1)$ -th iteration are explored. This consists of exploring all neighbors of C_{n-1} . In this context, only DMAs including more than one segment can give out one boundary segment, as the transfer of one segment from a one-segment DMA would cause its disappearance, thus violating the above-mentioned constraint on M . As an example, looking at the DMA configuration with $M = 4$ DMAs in Figure 3, there are six potential exchanges of boundary segments, i.e., S2 from DMA1 to DMA2, S3 from DMA2 to DMA1, S4 from DMA2 to DMA3, S5 from DMA3 to DMA2, S5 from DMA3 to DMA4 and S6 from DMA4 to DMA3 (see Figure 4).

The boundary segments to be exchanged can be easily identified in the dual segment/valve topology by looking at the isolation valves actively separating the DMAs. As in Figure 4, Table 3 links active isolation valves to the potential segment exchanges between the DMAs.

In most cases, the transfer of one segment from a giving DMA does not create any disconnections in this DMA, which then stays feasible. However, in some cases, the transfer of one segment from a giving DMA may cause this DMA to remain

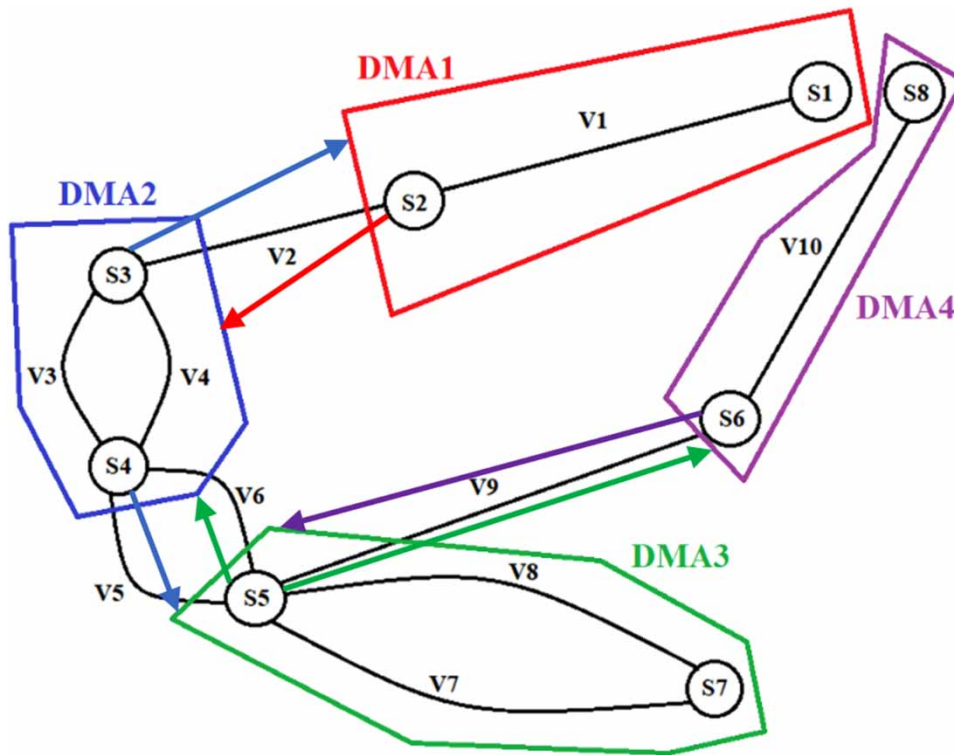


Figure 4 | For the WDN in Figure 2 and clustered into four DMAs as is shown in Figure 3, all potential exchanges of boundary segments.

Table 3 | Relationship between active isolation valves and segment exchanges

Reference isolation valve(s)	Connected DMAs	Segment exchange(s)
V2	DMA1, DMA2	S2 from DMA1 to DMA2, S3 from DMA2 to DMA1
V5 and V6	DMA2, DMA3	S4 from DMA2 to DMA3, S5 from DMA3 to DMA2
V9	DMA3, DMA4	S5 from DMA3 to DMA4, S6 from DMA4 to DMA3

subdivided into $n_r > 1$ remaining parts. As considering these n_r parts as DMAs of their own would cause the increase in M beyond the prefixed value, only the largest of the n_r parts in terms of number of segments is kept in the giving DMA; the other $n_r - 1$ parts are transferred, instead, to the receiving DMA.

The generic k -th of the N_e possible segment exchanges will be associated with a variation $\Delta Q_{r,k}$ in the modularity Q evaluated with Equation (1), with the subscript r standing for the word refinement.

Considering the set of all N_e possible variations $\Delta Q_{r,k}$, the selection of the segment exchange to consider for the n -th iteration is made in a probabilistic way. Instead of deterministically selecting the segment exchange yielding the largest positive $\Delta Q_{r,k}$, the potential variations $\Delta Q_{r,k}$ are sorted in ascending order and their cumulative frequency F_k for the selection is calculated. Then, a random number *rand* is generated between 0 and 1 and the segment exchange corresponding to the $\Delta Q_{r,k}$ value with the smallest cumulative frequency F larger than *rand* is selected, leading to the new WDN clustering configuration C_n . In comparison with C_{n-1} , C_n may be better or worse, as variations $\Delta Q_{r,k}$ may be either positive or negative.

In the present work, the cumulative frequency F_k for the sampling is calculated between 0 and 1 as a function of k by means of the following relationship:

$$F_k = \max\left(\frac{k - k_{\text{val}}}{N_e - k_{\text{val}}}, 0\right), \quad (8)$$

in which k_{val} is a parameter ranging from 0 to $N_e - 1$ and regulated across the iterations, as will be explained in the following subsection. The impact of k_{val} can be explained by means of the graph in Figure 5, which was constructed for $N_e = 100$. As the graph shows, the growth of k_{val} from 0 to $N_e - 1$ shifts rightward the pattern of $F_k(k)$, thus favoring the selection of the segment exchanges yielding the large $\Delta Q_{r,k}$ improvements. In fact, for $k_{\text{val}} = 0$, all segment exchanges have the same probability of selection, equal to 0.01. For $k_{\text{val}} = 50$, the segment exchanges with $k \leq 50$, i.e., the 50 worst segment exchanges, have a probability of selection equal to 0. Those with $51 \leq k \leq 100$ have a probability of selection equal to 0.02. Finally, for $k_{\text{val}} = 99$, all the segment exchanges have a probability of selection equal to 0, except for the last one associated with $k = 100$. This last segment exchange has a probability of selection equal to 1.

Summing up, the calculation of F_k based on Equation (8) can modulate the probabilistic selection of segment exchanges. Indeed, for low values of k_{val} , the approach is fully probabilistic with no privilege for the good segment exchange solutions. In this case, a worse WDN clustering solution C_n than C_{n-1} in terms of Q is likely to be obtained. The privilege for good segment exchanges grows with k_{val} increasing. For high values of k_{val} , the probability of obtaining a worse C_n than C_{n-1} decreases. For $k_{\text{val}} = 99$, the exchange becomes fully deterministic leading to the choice of the best exchange solution for the growth of Q (steepest ascent).

2.4.2. Framework of the iterations

A maximum number N_{max} of iterations must be set (e.g., $N_{\text{max}} = 2,000$).

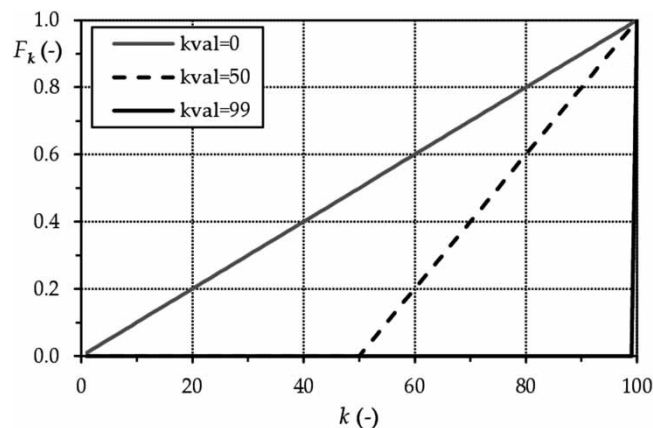


Figure 5 | F_k as a function of k for three values of k_{val} .

During the iterations, the value of k_{val} to be used in Equation (8) is changed to modulate the probabilistic nature of segment exchange selection, based on the following relationship:

$$k_{\text{val}} = \min\left(N_e \frac{n - n_{\text{stag}}}{K}, N_e - 1\right), \quad (9)$$

with K (–) being a factor regulating the speed at which k_{val} reaches its threshold value of $N_e - 1$ (e.g., $K = 50$). In Equation (9), n_{stag} is the last iteration in which the algorithm has reached a stagnation point, i.e., when the value of Q has stopped increasing.

The overall operation of the algorithm can then be explained as follows. When the algorithm is run, it starts with $n = 0$ and $n_{\text{stag}} = 0$. As n grows, k_{val} grows according to Equation (9). The algorithm then changes smoothly from fully probabilistic to fully deterministic (e.g., see Figure 5). This change in behavior enables first exploring the research space, consisting of modified WDN clustering configurations in comparison with the one of the first attempt, and then searching for a local maximum in a certain research space area. When Q increases no more, n_{stag} is set to n . Then, starting from the next iteration, i.e., when $n = n_{\text{stag}} + 1$, the algorithm becomes fully probabilistic again. Again, the growth of n causes the nature of the research space exploration to progressively change from probabilistic to deterministic, until a new local maximum of Q is found, in correspondence to which n_{stag} is set to n . The iterations proceed up till $n = N_{\text{max}}$, with the ultimate WDN clustering solution being the one featuring the highest value of Q during all the N_{max} iterations, i.e., the maximum of all the local maxima of Q encountered during the iterations.

3. APPLICATIONS

3.1. Case studies

Two case studies are considered in the present work. The first case study is the WDN of Ferrara (Alvisi *et al.* 2011), with two source nodes with a fixed head of 30 m, 536 demand nodes and 825 pipes. Overall, 969 isolation valves are present in the WDN, plus two isolation valves at the exit of the sources, leading to the formation of 682 segments. The first case study was used to search for WDN clustering configurations featuring uniform demands, as the uniformity of nodal elevations inside each DMA is *per se* guaranteed by the flat topography of the WDN area.

The second case study is the WDN of Saadnayel in Lebanon (Creaco & Haidar 2019), made up of one source node with a fixed head of 962 m, 152 demand nodes and 186 pipes. Each pipe is assumed here to be equipped with an internal isolation valve, leading to a total of 153 segments, each made up of a single node. Nodal ground elevations lie between 874.3 and 911.8 m above sea level (asl). Due to the larger diversity of ground elevations than the WDN of Ferrara, the second case study was then used to search for WDN clustering configurations featuring uniform demands across the DMAs and uniform ground elevations inside each DMA.

More information about the two case studies can be found in the referenced works. Besides the application of the clustering methodology presented in this paper, the following subsections also report some results of the dividing phase, in which the inter-DMA boundaries are either closed or fitted with a flow meter. In the end, the flow meters present in the WDN enable estimating the overall demand (user consumption + leakage) of each DMA. When all DMAs are supplied in parallel by single flow meters (one for each DMA), each flow meter reading represents the overall demand of one DMA. Otherwise, trivial addition/subtraction calculations must be carried out on the readings to estimate the overall demand for intermediate DMAs.

3.2. Results

The applications of this work were carried out in the Matlab[®] 2022b environment by using a single unit of a 12th Gen Intel(R) Core(TM) i9-12900H 2.50 GHz processor. To speed up calculations, the algorithm can also be applied with parallel computation to WDNs of any size, with no restraints on the number of nodes and pipes. H_1 in Equation (2) was calculated as n_b/n_v . H_2 and H_3 were calculated considering demands and ground elevations, respectively.

In the first case study, the clustering into $M = 8$ DMAs was initially considered, like in Creaco *et al.* (2022). The performance of the novel algorithm in comparison with FGA is evaluated in terms of values of Q , H_1 , H_2 , number n_b of boundary pipes and coefficient of variation $C_{v,d}$ for DMA demands. Incidentally, $C_{v,d}$ is a dimensionless coefficient calculated as the ratio of mean to standard deviation, which expresses the heterogeneity of DMA demands: the larger $C_{v,d}$, the more

heterogeneous the DMA demands. The choice of reporting this coefficient in the analysis is due to its presence in previous works in the scientific literature (Creaco *et al.* 2019, 2022).

In the case of $M = 8$, FGA and the novel optimization algorithm were applied considering six different triplets of values for the weights in Equation (1). In all triplets, α_3 was set to 0 since the uniformity of ground elevations was not investigated in the first case study. While the sum of α_1 and α_2 was kept equal to 2 in all triplets, various combinations of α_1 and α_2 were considered, as reported in Table 4. This was done to analyze the extent to which the WDN clustering configurations react to the increase in the weight α_1 from 0.1 to 1.9 given to the first term of the modularity, associated with the boundary pipes, and then to the simultaneous decrease in the weight α_2 from 1.9 to 0.1 given to the second term of the modularity, associated with the demand. Table 4 shows that, in every triplet of α_1 , α_2 and α_3 , the optimization algorithm applied to the solution of the first attempt obtained by means of FGA was beneficial in that it yielded increases in Q , as well as decreases in H_1 , H_2 , n_b and $C_{v,d}$. These improvements are the evidence of the suboptimal performance of FGA, which lends itself to be improved by means of the optimization algorithm. Looking at the values of n_b and $C_{v,d}$ obtained after the optimization algorithm, which are more meaningful from the engineering viewpoint, the increase in α_1 at the expense of α_2 tends to produce WDN clustering configurations with fewer boundary pipes and more heterogeneous DMA demands. This is because the increase in α_1 and the decrease in α_2 give more weight to H_1 and less weight to H_2 in Equation (1).

Figure 6 reports the WDN clustering configuration obtained with $\alpha_1 = 0.1$, $\alpha_2 = 1.9$, $\alpha_3 = 0.0$. This configuration of DMAs has similar performance to that obtained by Creaco *et al.* (2022), which featured $n_b = 64$ and $C_{v,d} = 0.008$. However, it must be remarked that the combination of FGA + optimization algorithm has a computation time of about 90 min, whereas the algorithm of Creaco *et al.* (2022) requires about 1 day of computational time in this case study. Furthermore, the eight DMAs in the clustering solution in Figure 6(a) feature demand values ranging from 36.1 to 37.3 L/s, attesting to an almost perfect uniformity (Figure 6(b)). From the engineering viewpoint, an additional increase in uniformity beyond this threshold is not meaningful. As far as the size of the DMAs is concerned, DMAs 1, 3, 5, 6 and 7 are quite balanced, being about 10 km long (Figure 6(b)). DMA 8 is sensibly smaller than the others due to the presence of demand hotspots inside whereas the external DMAs 2 and 4 with a lower concentration of demand are sensibly larger. These results are consistent with the distribution of demands over WDN pipes shown in Figure 6(c).

Following the clustering, the dividing was carried out, which consists of deciding on the boundaries to be closed or fitted with a flow meter, to prove that the WDN clustering configurations obtained with the optimization are competitive from the hydraulic viewpoint. The closure of a boundary causes loss of reliability while the installation of a flow meter results in a cost. Therefore, the dividing problem was tackled in the first case study by using the same approach as Creaco *et al.* (2022), that is by solving a bi-objective optimization to obtain trade-off solutions between installation costs and reliability expressed by means of the generalized resilience failure (GRF) index (Creaco *et al.* 2016), to be minimized and maximized, respectively. In this optimization run, the decision variables are as numerous as the inter-DMA boundaries, to decide alternatively on valve closure or flow meter installation. The optimization was run for all the clustering solutions obtained by applying the optimization (see Table 4). The Pareto fronts produced by the multi-objective optimization are shown in Figure 7. Globally, these fronts show that, despite the high number of boundary pipes present in the clustering solutions (up to 59) due to the highly interconnected and looped nature of the WDN, the GRF grows rapidly when the number of boundaries fitted with a flow

Table 4 | Performance of FGA and improvements with the novel optimization algorithm in terms of Q , H_1 , H_2 , n_b and $C_{v,d}$ in the first case study with number of DMAs $M = 8$

Weights	FGA					Optimization				
	Q	H_1	H_2	n_b	$C_{v,d}$	Q	H_1	H_2	n_b	$C_{v,d}$
$\alpha_1 = 0.1, \alpha_2 = 1.9, \alpha_3 = 0.0$	0.74	0.093	0.130	91	0.207	0.76	0.061	0.125	59	0.013
$\alpha_1 = 0.4, \alpha_2 = 1.6, \alpha_3 = 0.0$	0.76	0.068	0.132	66	0.237	0.78	0.053	0.125	51	0.048
$\alpha_1 = 1.0, \alpha_2 = 1.0, \alpha_3 = 0.0$	0.81	0.059	0.128	57	0.160	0.82	0.051	0.127	49	0.119
$\alpha_1 = 1.3, \alpha_2 = 0.7, \alpha_3 = 0.0$	0.83	0.055	0.136	53	0.291	0.85	0.047	0.131	46	0.214
$\alpha_1 = 1.6, \alpha_2 = 0.4, \alpha_3 = 0.0$	0.86	0.047	0.168	46	0.589	0.87	0.040	0.154	39	0.483
$\alpha_1 = 1.9, \alpha_2 = 0.1, \alpha_3 = 0.0$	0.91	0.018	0.605	17	1.959	0.92	0.021	0.435	20	1.575

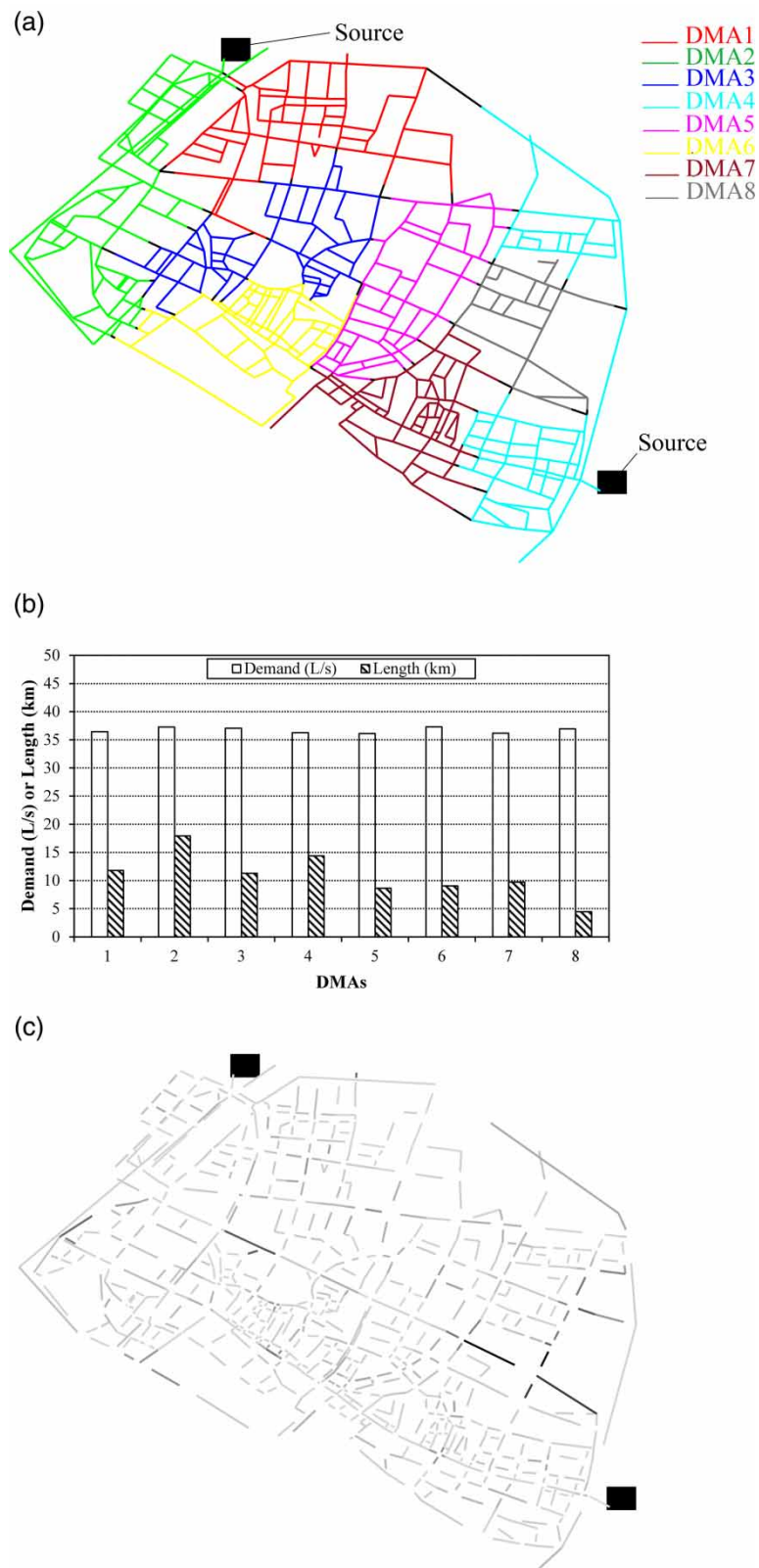


Figure 6 | (a) WDN clustering solution obtained in the first case study for $\alpha_1 = 0.1$, $\alpha_2 = 1.9$, $\alpha_3 = 0.0$; (b) total length and demand of the DMAs; (c) contour plot of pipe demands with isolation valves removed; the darker the color of the pipe, the larger its demand.

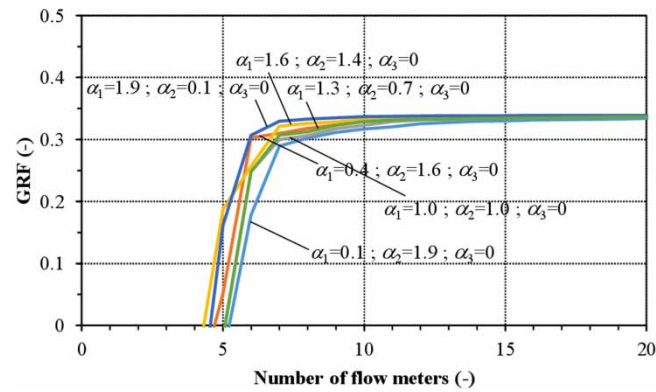


Figure 7 | Pareto fronts reporting the values of the GRF index as a function of the number of installed flow meters during the dividing phase in the first case study, for the different weight triplets considered.

meter grows. In this case study, a number of flow meters equal to 10 at the inter-DMA boundaries is always sufficient to guarantee a resilience value close to that of the un-partitioned WDN (equal to 0.34). It must be remarked that this number of flow meters, to which two flow meters must be added, i.e., one at the exit of either source, is reasonable for a WDN partitioned into eight DMAs (namely, only the 17% of the boundary pipe set when $n_b = 59$). The closure of the isolation valve in the remaining boundaries does not jeopardize the performance of the WDN in terms of service pressure, attesting to the hydraulic effectiveness of WDN partitioning for all weight triplets considered in the clustering (see Table 4). Some ultimate partitioning solutions are shown in the supplementary material.

Though neglected in the present work, water quality implications may arise from closing valves. Additional water quality simulations could be carried out to identify potentially negative issues, suggesting changes in the boundaries to be closed or fitted with a flow meter.

An additional analysis was carried out in the first case study to investigate the extent to which the change in the number of DMAs affects the performance of the clustering optimization. In this context, Table 5 shows the results obtained in terms of Q , H_1 , H_2 , n_b and $C_{v,d}$ when the number of DMAs is increased from eight to 13, using the same triplets α_1 , α_2 , α_3 as in Table 4. The analysis of Table 5 points out for $M = 13$ similar remarks to $M = 8$. Once more, the optimization algorithm applied to the starting solution obtained by means of FGA yielded increases in Q , as well as decreases in H_1 , H_2 , n_b and $C_{v,d}$, confirming the suboptimality of FGA and the possibility of improving the results with the optimization algorithm. With α_1 and α_2 increasing and decreasing, respectively, WDN clustering configurations with fewer boundary pipes and more heterogeneous DMA demands are obtained, because of the higher weight given to H_1 than to H_2 in Equation (1). Overall, the comparison of the optimization results in Tables 4 and 5 shows that the increase in M has detrimental effects in n_b and $C_{v,d}$. The triplets being the same, configurations with less uniform DMAs and more numerous boundary pipes are obtained when M increases.

Table 5 | Performance of FGA and improvements with the novel optimization algorithm in terms of Q , H_1 , H_2 , n_b and $C_{v,d}$ in the first case study with a number of DMAs $M = 13$

Weights	FGA					Optimization				
	Q	H_1	H_2	n_b	$C_{v,d}$	Q	H_1	H_2	n_b	$C_{v,d}$
$\alpha_1 = 0.1, \alpha_2 = 1.9, \alpha_3 = 0.0$	0.83	0.119	0.082	115	0.254	0.85	0.083	0.077	80	0.045
$\alpha_1 = 0.4, \alpha_2 = 1.6, \alpha_3 = 0.0$	0.84	0.089	0.080	86	0.198	0.85	0.073	0.078	71	0.120
$\alpha_1 = 1.0, \alpha_2 = 1.0, \alpha_3 = 0.0$	0.84	0.080	0.081	78	0.241	0.85	0.072	0.081	70	0.225
$\alpha_1 = 1.3, \alpha_2 = 0.7, \alpha_3 = 0.0$	0.84	0.073	0.092	71	0.450	0.85	0.067	0.091	65	0.423
$\alpha_1 = 1.6, \alpha_2 = 0.4, \alpha_3 = 0.0$	0.85	0.054	0.156	52	1.014	0.87	0.046	0.146	45	0.947
$\alpha_1 = 1.9, \alpha_2 = 0.1, \alpha_3 = 0.0$	0.89	0.026	0.584	25	2.568	0.90	0.025	0.565	24	2.518

In the second case study, the clustering into $M = 5$ DMAs was considered, like in Creaco & Haidar (2019). Two triplets were considered for the parameters in Equation (1), as is shown in Table 6. In the first triplet, $\alpha_1 = 0.2$, $\alpha_2 = 1.8$ and $\alpha_3 = 0.0$ were considered, i.e., no weight was given to the uniformity of nodal elevations inside each DMA. The second triplet, instead, was set to $\alpha_1 = 0.15$, $\alpha_2 = 0.15$ and $\alpha_3 = 1.7$ to take into account the impact of the uniformity of nodal elevations inside each DMA.

Then, the FGA and the optimization algorithm were applied to the second case study considering both triplets. In the application with the second triplet, the two algorithms were fictitiously applied considering $M = 6$, as the source, which has a very different elevation from the remainder of the WDN, was always identified as a whole DMA of its own.

Table 6 shows the results of the two algorithms in terms of Q , H_1 , H_2 , H_3 , n_b and $C_{v,d}$, confirming the capacity of the optimization algorithm to improve the features of the WDN clustering of the first attempt obtained with FGA, in terms of modularity Q . As expected, the values of the indicators H_2 and $C_{v,d}$ obtained by the optimization algorithm considering the second triplet are higher than those obtained with the first triplet, because the second triplet compromises the uniformity of DMA demands with the uniformity of nodal elevations inside each DMA. The values of DMA nodal heterogeneity H_3 (Equation (4)) obtained by applying the optimization algorithm with the first and second triplet are equal to 0.023 and 0.010, respectively, entailing that the second triplet yields more uniform ground elevations inside each DMA. Figure 8 compares the results obtained by means of the optimization algorithm considering the two triplets of weight coefficients, i.e., $\alpha_1 = 0.2$, $\alpha_2 = 1.8$, $\alpha_3 = 0.0$ (Figure 8(a)) and $\alpha_1 = 0.15$, $\alpha_2 = 0.15$, $\alpha_3 = 1.7$ (Figure 8(b)). These results confirm that the change in the weight coefficients significantly impacts on the clustering. In fact, the shape, size, and extension of the five DMAs are very different.

To prove the suitability of the WDN clustering obtained with $\alpha_1 = 0.15$, $\alpha_2 = 0.15$ and $\alpha_3 = 1.7$ for pressure management, the dividing phase and pressure reducing valve (PRV) installations were carried out on the configuration in Figure 8(b). Based on the results of an optimization run aimed at simultaneously minimizing leakage and number of flow meters and PRVs, the isolation valve was closed in seven boundary pipes while the eight remaining boundary pipes were equipped with the flow

Table 6 | Performance of FGA and improvements with the novel optimization algorithm in terms of Q , H_1 , H_2 , H_3 , n_b and $C_{v,d}$ in the second case study

Weights	FGA						Optimization					
	Q	H_1	H_2	H_3	n_b	$C_{v,d}$	Q	H_1	H_2	H_3	n_b	$C_{v,d}$
$\alpha_1 = 0.2, \alpha_2 = 1.8, \alpha_3 = 0.0$	0.594	0.086	0.216	0.030	16	0.28	0.625	0.070	0.201	0.023	13	0.05
$\alpha_1 = 0.15, \alpha_2 = 0.15, \alpha_3 = 1.7$	0.936	0.091	0.219	0.010	17	0.56	0.937	0.081	0.222	0.010	15	0.58

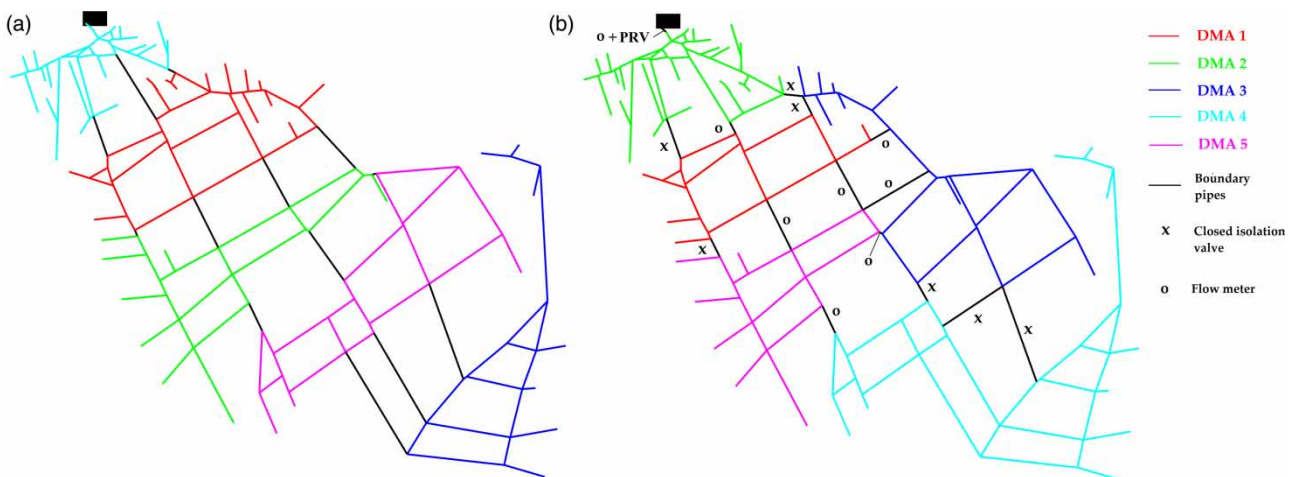


Figure 8 | WDN clustering solutions obtained in the second case study for (a) $\alpha_1 = 0.2$, $\alpha_2 = 1.8$ and $\alpha_3 = 0.0$ and (b) $\alpha_1 = 0.15$, $\alpha_2 = 0.15$ and $\alpha_3 = 1.7$. In (b), the results of the dividing and PRV installations are also shown.

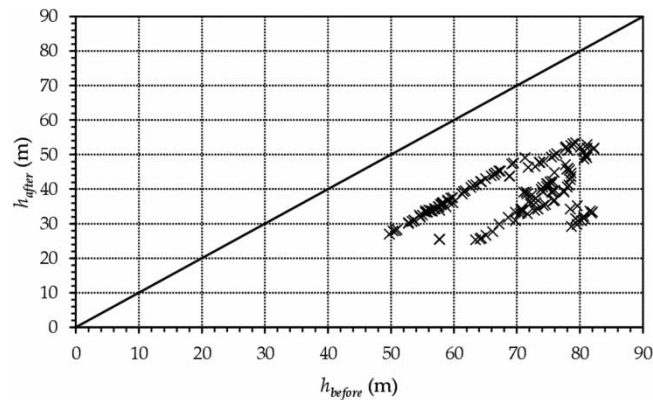


Figure 9 | For the WDN clustering solution in Figure 8(b), pressure heads before (h_{before}) and after (h_{after}) dividing phase and PRV installations.

meter. Then, one of these was also fitted with a PRV (with downstream pressure head setting of 27 m and downstream total head setting of 938.82 m). As Figure 9 shows, this enables significantly reducing the nodal pressure heads in the hydraulic model of the Saadnayel WDN, that is from the range 50–83 m to the range 25–52 m. Starting from the initial condition in which leakage adds up to about 37 L/s, i.e., 22% of the total outflow from the WDN, WDN partitioning and PRV installation enable attenuating leakage by 44%, with no nodes falling below the threshold of 25 m for full demand satisfaction.

4. CONCLUSIONS

In this work, a new formulation of modularity tailored to WDNs was proposed and implemented inside an effective optimization algorithm inspired by the simulated annealing approach, able to refine an initial clustering solution of the first attempt. As an improvement in comparison with the others present in the scientific literature, this formulation has the following two beneficial features for the sake of WDN clustering into DMAs:

- It is based on the dual topology based on segments and valves.
- It considers properties to be made uniform both across and inside DMAs.

As proven by the applications to the first case study of this paper, the use of the dual topology in the modularity enables obtaining configurations of DMAs that take advantage of the isolation valves actually present in the WDN. For the boundaries to be closed, there is then no need of installing additional isolation valves. Furthermore, the already available valve-vaults can sometimes be used to host the flow meters in the boundary pipes that are left open after the dividing phase. As proven by the applications to the second case study, the possibility of considering properties to be made uniform inside DMAs can be exploited to obtain clustering configurations with uniform nodal elevations, suitable for carrying out pressure management in the WDN.

The performance of the optimization algorithm for the maximization of the modularity in the clustering phase proved outstanding in the trade-off between numerical effectiveness and computational burden, taking as benchmark results obtained in previous works.

All the clustering solutions obtained perform well from the hydraulic viewpoint during the dividing phase, in which the decision of closing the valve or installing a flow meter must be taken at each boundary.

Furthermore, the methodology developed enables overcoming one of the main difficulties encountered by previous authors with the clustering, that is the possibility of considering nodal (instead of pipe) properties as an objective to pursue. In fact, while weighted coefficients are typically assigned to links, the new framework of this work can assign weights to both nodes and links.

Though the clustering applications of this work considered only the minimization of number of boundary pipes, heterogeneity of demand across DMAs and heterogeneity of nodal ground elevations inside DMAs, the framework proposed is very flexible. Therefore, other properties can be easily included, such as boundary pipe diameter, pipe length across DMAs and pipe age inside DMAs.

As the applications conducted in this paper concerned small to medium-sized networks, future work will be dedicated to extending the methodology to complex and large networks with numerous control devices (e.g., pumps and control valves).

ACKNOWLEDGEMENTS

Support from Italian MIUR and University of Pavia is acknowledged within the program Dipartimenti di Eccellenza 2023–2027.

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

- Alvisi, S., Creaco, E. & Franchini, M. 2011 Segment identification in water distribution systems. *Urban Water Journal* **8** (4), 203–217.
- Brentan, B., Campbell, E., Goulart, T., Manzi, D., Meirelles, G., Herrera, M., Izquierdo, J. & Luvizotto Jr., E. 2018 Social network community detection and hybrid optimization for dividing water supply into district metered areas. *Journal of Water Resources Planning and Management* **144** (5), 04018020.
- Campbell, E., Izquierdo, J., Montalvo, I. & Pérez-García, R. 2016 A novel water supply network sectorization methodology based on a complete economic analysis, including uncertainties. *Water* **8**, 179. <https://doi.org/10.3390/w8050179>.
- Clauset, A., Newman, M. E. J. & Moore, C. 2004 Finding community structure in very large networks. *Physical Review E* **70** (6), 066111.
- Creaco, E. & Haidar, H. 2019 Multiobjective optimization of control valve installation and DMA creation for reducing leakage in water distribution networks. *Journal of Water Resources Planning and Management* **145**, 04019046.
- Creaco, E., Franchini, M. & Todini, E. 2016 Generalized resilience and failure indices for use with pressure-driven modeling and leakage. *Journal of Water Resources Planning and Management* **142** (8), 04016019.
- Creaco, E., Cunha, M. & Franchini, M. 2019 Using heuristic techniques to account for engineering aspects in modularity-based water distribution network partitioning algorithm. *Journal of Water Resources Planning and Management* **145** (12), 04019062.
- Creaco, E., Zheng, F. & Pezzinga, G. 2022 Minimum transport driven algorithm for water distribution network partitioning. *Aqua-Water Infrastructure Ecosystems and Society* **71** (1), 120–138.
- Di Nardo, A., Giudicianni, C., Greco, R., Herrera, M. & Santonastaso, G. F. 2018 Applications of graph spectral techniques to water distribution network management. *Water* **10** (1), 45.
- Giustolisi, O. & Ridolfi, L. 2014 A novel infrastructure modularity index for the segmentation of water distribution networks. *Water Resources Research* **50** (10), 7648–7661.
- Giustolisi, O. & Ridolfi, L. 2014 New modularity-based approach to segmentation of water distribution networks. *Journal of Hydraulic Engineering* **140** (10), 04014049.
- Giudicianni, C., Herrera, M., di Nardo, A. & Adeyeye, K. 2020 Automatic multiscale approach for water networks partitioning into dynamic district metered areas. *Water Resources Management* **34** (2), 835–848.
- Laarhoven, P. J. M. & van Peter, J. M., 1987 In: *Simulated Annealing: Theory and Applications* (Aarts, E. H. L. ed.). D. Reidel, Dordrecht. ISBN 90-277-2513-6. OCLC 15548651.
- Morrison, J., Tooms, S. & Rogers, D. 2007 *District Metered Areas Guidance Notes*. Water Loss Task Force, IWA Publications, London, UK.
- Newman, M. E. J. & Girvan, M. 2004 Finding and evaluating community structure in networks. *Physical Review E* **69** (2), 026113.
- Rahmani, F., Muhammed, K., Behzadian, K. & Farmani, R. 2018 Optimal operation of water distribution systems using a graph theory-based configuration of district metered areas. *Journal of Water Resources Planning and Management* **144** (8), 04018018.
- Santonastaso, G. F., Di Nardo, A. & Creaco, E. 2019 Dual topology for partitioning of water distribution networks considering actual valve locations. *Urban Water Journal* **16** (7), 469–479.
- Shao, Y., Liu, J., Yao, H., Zhang, T., Lima Neto, I. E., Yu, T. & Chu, S. 2022 An improved hybrid community detection algorithm for partition of water distribution networks. *Engineering Optimization*. <https://doi.org/10.1080/0305215X.2022.2155148>.
- Sharma, A. N., Dongre, S. R., 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. *Journal of Water Resources Planning and Management* **148** (8), 04022040.
- Walski, T. M. 1993 Water distribution valve topology for reliability analysis. *Reliability Engineering & System Safety* **42** (1), 21–27.

WaterGEMS [Computer software]. Bentley Systems, Exton, PA.

Yao, H., Zhang, T., Shao, Y., Yu, T. & Lima Neto, I. E. 2021 Improved modularity-based approach for partition of water distribution networks. *Urban Water Journal* **18** (2), 69–78.

Zhang, Q., Wu, Z. Y., Zhao, M. & Qi, J. 2017 Automatic partitioning of water distribution networks using multiscale community detection and multiobjective optimization. *Journal of Water Resources Planning and Management* **143** (9), 04017057.

First received 1 December 2022; accepted in revised form 31 May 2023. Available online 12 June 2023