Pareto-optimal design of water distribution networks: an improved graph theory-based approach

Mohsen Hajibabaei [iD][a], Sina Hesarkazzazi[a], Amin Minaei[a,b], Dragan Savić IWA[c,d] and Robert Sitzenfrei [iD][a,*]

[a] Unit of Environmental Engineering, Department of Infrastructure Engineering, University of Innsbruck, Innsbruck, Austria
[b] Austrian Academy of Sciences (ÖAW), Vienna, Austria
[c] KWR Water Research Institute, Nieuwegein, The Netherlands
[d] Centre for Water Systems, University of Exeter, Exeter, UK
*Corresponding author. E-mail: robert.sitzenfrei@uibk.ac.at

[iD] MH, 0000-0002-0047-9715; RS, 0000-0003-1093-6040

ABSTRACT

One of the main drawbacks of using evolutionary algorithms for the multi-objective design of water distribution networks (WDNs) is their computational inefficiency, particularly for large-scale problems. Recently, graph theory-based approaches (GTAs) have gained attention as they can help with the optimal WDN design (i.e., determining optimal diameters). This study aims to extend a GTA to further improve the quality of design solutions. The GTA design is based on a customized metric called 'demand edge betweenness centrality', which spatially distributes nodal demands through the weighted edges of a WDN graph and provides an estimation of water flows. Assigned edge weights can be constant (i.e., static) or modified iteratively (i.e., dynamic) during the design process, leading to different flow estimations and alternative design options. Three hydraulic-inspired dynamic weights are developed in this study to better reproduce hydraulic behavior and, consequently, find better solutions. Additionally, this work proposes a framework for the optimal design of multi-source WDNs and provides guidelines for obtaining near-optimal solutions in such networks. A comparative study between GTAs and evolutionary optimizations confirms the efficiency of the improved GTA in providing optimal/near-optimal solutions, especially for large WDNs, with a runtime reduction of up to seven orders of magnitude.

Key words: complex network analysis, multi-objective evolutionary algorithms, Pareto fronts, resilience

HIGHLIGHTS

- An improved graph theory-based approach for determining optimal diameters is developed.
- The proposed approach uses hydraulic-inspired dynamic weights to better reproduce hydraulic behavior.
- The newly developed framework is highly efficient, particularly for large-scale networks with thousands of pipes.

GRAPHICAL ABSTRACT
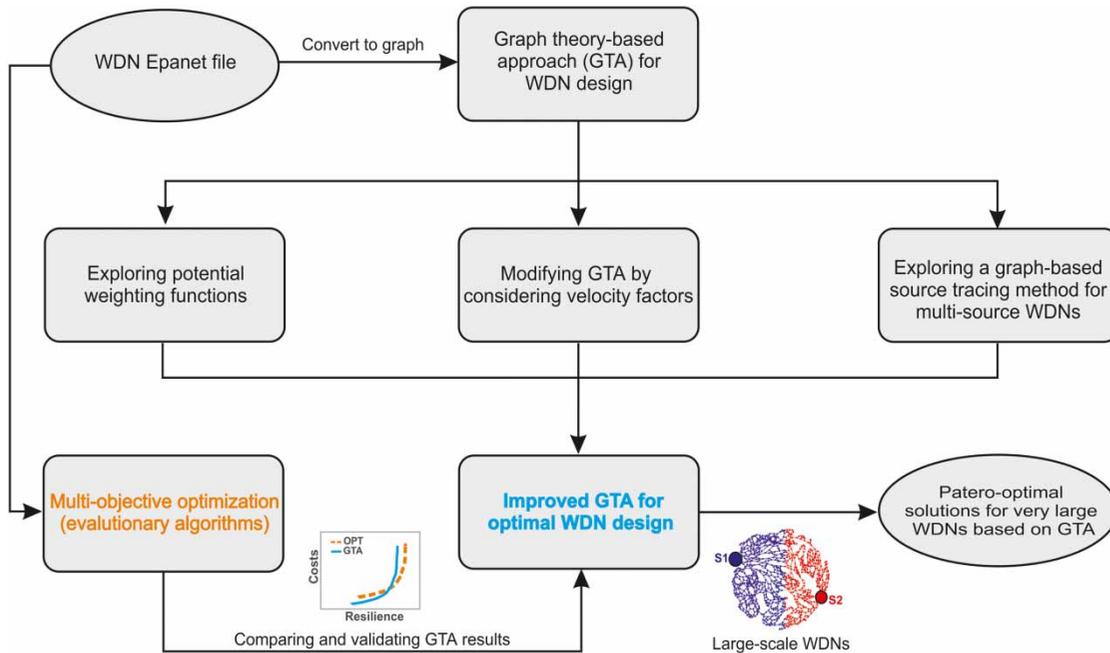
ACRONYMS

DP	demand parcel
EBC	edge betweenness centrality
EBC^Q	demand edge betweenness centrality
EBC^Q_S	demand edge betweenness centrality with static weights
EBC^Q_D	demand edge betweenness centrality with dynamic weights
GTA	graph theory-based approach
GTA*	graph theory-based approach with velocity factors
GTA* (S)	graph theory-based approach with velocity factors and static weights
GTA* (D)	graph theory-based approach with velocity factors and dynamic weights
OPT	Pareto solutions obtained based on evolutionary optimization
SPL	shortest path length
Tr	threshold
WDN	water distribution network

1. INTRODUCTION

Water distribution networks (WDNs) are crucial urban infrastructures with the primary aim of delivering required water in terms of quality and quantity to consumers (Hajibabaei *et al.* 2019). These infrastructures need to be designed and operated to such a degree that they can fulfill the primary aim while providing reliable performance with a reasonable investment. This is where optimization has emerged as a decision-making tool in different WDNs' phases, such as layout, design, and operation, to ensure maximum benefits (e.g., resilience) with minimum costs (De Corte & Sörensen 2013). In the layout phase, the focus is on network structure, involving system connectivity and topology, to determine the optimal locations of components. During the design phase, the types of pipes, valves, and pumps, as well as pipe diameters, are selected. The operation phase mainly focuses on the daily functioning of network components to ensure that required water is delivered (De Corte & Sörensen 2013). Among the aforementioned phases, design is one of the most challenging ones, particularly when it comes to the optimal selection of pipe diameters (Sitzenfrei *et al.* 2021).

Determining optimal diameters in WDNs is a nonlinear and NP-hard problem, commonly solved using nature-inspired evolutionary algorithms with conflicting objectives (e.g., minimizing costs and maximizing resilience) (Wang *et al.* 2017).

The advantage of applying evolutionary algorithms to pipe sizing is their ability to systematically search for possible combinations of design solutions, taking numerous decision variables into account (Diao *et al.* 2022). However, the major drawback of these algorithms is their computational inefficiency, especially for large-scale systems, where extensive execution times are required due to the exponential growth of the search space (Sitzenfrei *et al.* 2020). Several attempts have been made to improve the efficiency of evolutionary algorithms by reducing the search space in optimization models (Minaei *et al.* 2020; Diao *et al.* 2022). Nevertheless, achieving optimal design solutions for complex WDNs remains a computationally prohibitive process since numerous function evaluations with multiple simulations are still needed (Wang *et al.* 2017). Therefore, an efficient design procedure is required to overcome the drawbacks of evolutionary algorithms and find the best approximation of optimal solutions in a rational time frame. To this end, graph theory-based approaches (GTAs) are of interest since they can give valuable insights into complex network behavior through the analysis of topological characteristics. In GTAs, WDNs are presented as mathematical graphs consisting of sets of vertices (nodes) and edges (pipes), where the connections between elements of complex networks are better characterized. In addition, GTAs can be efficiently applied to real-world and large-scale WDNs with thousands of pipes in a relatively short computational time (Sitzenfrei *et al.* 2021).

GTAs have been employed for different purposes in WDNs, such as isolation valve systems analysis (Giustolisi *et al.* 2022), water quality assessment (Sitzenfrei 2021), resilience analysis (Pagano *et al.* 2019), missing diameters reconstruction (Hajibabaei *et al.* 2023), and district metered areas creation (Diao *et al.* 2013). These studies laid a reasonable theoretical foundation for the application of GTAs in WDN analysis, considering various topological characteristics. GTAs have also recently received attention for the multi-objective optimization of WDNs. For example, network characteristics and layout patterns of Pareto-optimal solutions were explored by Sitzenfrei *et al.* (2019) using several graph metrics. They indicated that employing a certain topological centrality metric (edge betweenness centrality) with a proper weighting function could describe the layout of low-cost optimal solutions. Topological metrics were also effectively used as surrogate measures for hydraulic metrics in WDN optimization to enhance seismic robustness and reduce optimization runtime (Pudasaini & Shahandashti 2020). In another study, Diao *et al.* (2022) used topological (tailored edge betweenness) and hydraulic (pipe flow and head loss) metrics to narrow down the search space for pipe re-sizing optimization. Sitzenfrei *et al.* (2020) applied GTAs for multi-objective optimization of pipe diameters. They developed a computationally efficient method for WDN design, where water flow is spatially distributed through the shortest path between demand nodes and a source node. After estimating the water flow in each pipe, different diameter options can be obtained by varying velocities. To identify the shortest path, weights are assigned to the edges (pipes) of the WDN graph. These weights can be constant (static) or modified iteratively in the course of the process, referred to herein as dynamic weights.

An initial study on a dynamic weighting function has highlighted its potential to provide a promising representation of WDN hydraulic features (Hajibabaei *et al.* 2022). However, no systematic attempt has been made to fully discover the ability of different dynamic weighting functions to mimic hydraulic behavior during the design procedure. Furthermore, previous studies on GTAs applied uniform velocities across the network in each design step (Sitzenfrei *et al.* 2020, 2021), while in reality, flow velocities change in different pipe classes. In addition, a detailed analysis of GTAs for the design of multi-source WDNs has not been performed since the focus was mainly on the design of single-source networks.

This study proposes an improved GTA for the optimal design of WDNs (i.e., determining optimal diameters) to address the abovementioned knowledge gaps. For this purpose, three novel dynamic weighting functions are developed, and their variables are systematically tailored according to the hydraulic features of WDNs. Moreover, specific parameters (called velocity factors) are defined to provide appropriate velocities for different diameter classes in each design step, making the design procedure more realistic. Further, a structured framework based on the improved GTA is presented to design large-scale WDNs with multiple sources. The improved GTA is tested on five case studies, including two real, one semi-real and two benchmark WDNs, and the outcomes are compared with optimal results obtained with evolutionary approaches. Leveraging the methods proposed in this paper leads to a move toward high-quality optimal/near-optimal solutions, especially for complex and large-scale networks, in a short time frame.

2. METHODS AND MATERIALS

Figure 1 outlines the workflow considered in this paper. As shown in the figure, the Epanet file (network format) is used as an input file for the multi-objective optimization model and GTA. Epanet, as the widely recognized software for WDN analysis, includes advanced algorithms to solve network equations (Rossman 2000), making it a powerful tool to incorporate into
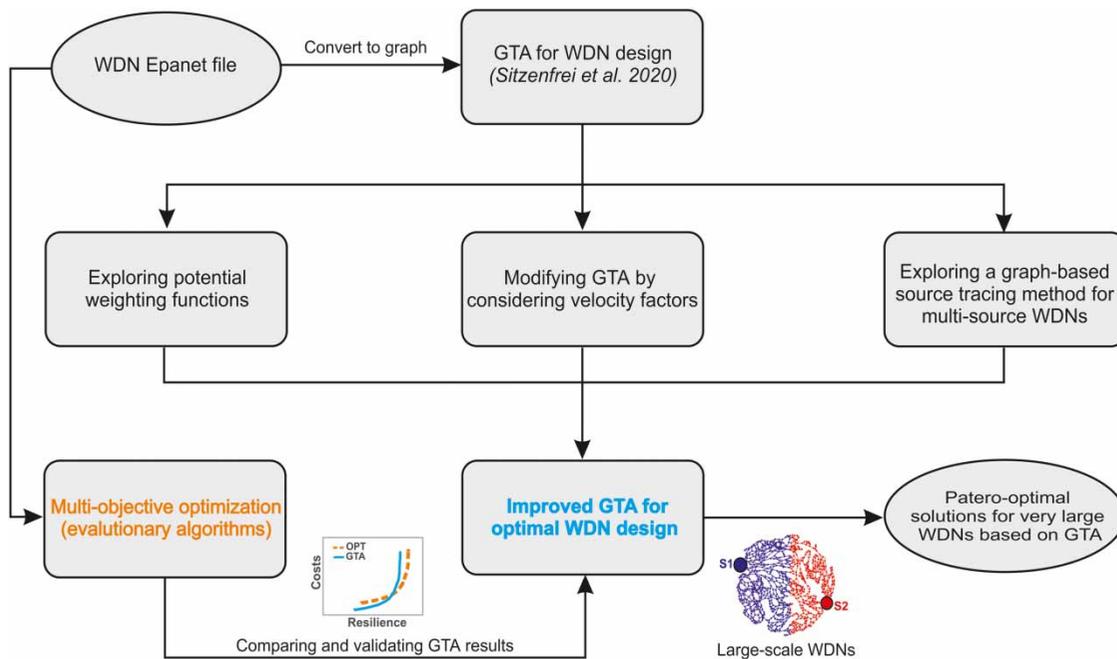
Figure 1 | Workflow of the methodology considered in this study.

optimization engines. Using Epanet's hydraulic calculations, the optimization model evaluates the performance of various design alternatives and guides the search toward optimal solutions. In addition, using the Epanet file for graph-based algorithms provides a standardized and structured representation of WDNs, facilitating the conversion of a network into a graph object. According to Figure 1, the optimization process starts with the Epanet input file converted to a graph object, wherein network information (e.g., pipe length, nodal demand) is stored as the graph characteristics. A GTA proposed for WDN design (Sitzenfrei *et al.* 2020) is further explored by systematically investigating various weighting functions and velocity factors in the design procedure. In addition, a graph-based source tracing method (Hajibabaei *et al.* 2022) is customized to achieve optimal solutions for multi-source WDN design. To accomplish these objectives, the GTA is applied to five case studies varying in size and complexity. Hence, sets of Pareto design solutions are subsequently obtained and examined to cherry-pick appropriate parameters, yielding an improved GTA for WDN design. The results of the improved GTA are then compared and validated with optimal solutions derived from evolutionary optimization approaches. Finally, the improved GTA is employed to optimally design large-scale multi-source WDNs, which is challenging to tackle with evolutionary optimizations.

2.1. GTA for WDN design

2.1.1. Graph metrics used for GTA

Urban water systems can be described using a branch of mathematics called graph theory. Accordingly, the relation between the components of a WDN can be represented as a graph G, including a set of graph nodes N (e.g., tanks, reservoirs) interconnected by a set of edges E (e.g., pipes, pumps). Source(s) and demand nodes are a subset of the graph nodes shown by S and D, respectively (i.e., $S \subseteq N$; $D \subseteq N$). In a weighted graph of WDN, different weighting functions (W_e) can be assigned to each edge e ($e \subseteq E$) depending on the analysis aim. For instance, pipe length (L_e) divided by velocity (V_e) can be used as a proper weighting function for water quality-related analysis (Sitzenfrei 2021). In addition, weighting functions can be static or dynamic. A constant value is utilized for edge weights in the static weighting approach, while the dynamic weighting function implies that weights are modified during an evaluation.

To obtain optimal solutions based on GTA, particular graph-based procedures are required. These procedures serve two primary purposes. Firstly, they assist in identifying the most efficient paths in WDNs from the source nodes (reservoirs or

tanks) to the sinks (demand nodes). Secondly, they are utilized to distribute nodal demands through efficient paths, enabling ideal flow estimation for optimal design. Thus, to achieve these two purposes, we selected the following two graph metrics:

1. *Shortest path length (SPL$_{i,j}$)* is the first metric utilized to determine the shortest 'distance' between two nodes (i and j), where 'distance' refers to the sum of (positive) edge weights in that path (Dijkstra 1959).
2. *Edge betweenness centrality (EBC)* is the second metric particularly modified for this study. In a WDN, source $EBC(e)$ counts the number of $SPL_{S,i}$ from S (Source) to every node i that go through the edge e. To provide an ideal flow estimation, Sitzenfrei *et al.* (2020) customized source EBC, specifically for WDN design, referred to as demand EBC (EBC^Q), which adds nodal demand $i \in D$ ($Q_i > 0$) to the $EBC^Q(e)$ values along $SPL_{S,i}$ instead of counting the number of $SPL_{S,i}$:

$$EBC^Q(e) = \sum_{i \in D} SPL_{S,i}(e) \cdot Q_i \tag{1}$$

In this work, both static and dynamic weighting functions are employed to calculate EBC^Q in the design process, denoted as (EBC^Q_S) and (EBC^Q_D), respectively. Different methods can be used for EBC^Q_D calculation. One approach could be increasing edge weights up to a specific threshold in each iteration. For instance, assuming pipe length as a weighting function, edge weights can be modified (increased) to a maximum threshold per iteration. This threshold (maximum lengthening) can be the function of demand nodes, maximum demand, or even a fixed value. Detailed information regarding the procedure of calculating static and dynamic weights for WDN design is described in the following section.

2.1.2. Design procedure

The value of $EBC^Q(e)$ in Equation (1) sums up demands routed through edge e, which can give an estimation of water flow $Q(e)$ in the edge (Sitzenfrei *et al.* 2020). Therefore, according to the volumetric flow rate in Equation (2), the required pipe diameter for e (DN_e) can be estimated by Equation (3):

$$Q(e) = V(e) \cdot \pi DN_e^2/4 \tag{2}$$

$$DN_e = \left\lceil \sqrt{\frac{4}{\pi} \cdot \frac{EBC^Q(e)}{V_{\text{design}} \cdot F(e)_v}} \right\rceil \quad \in \ DN_{\text{available}} \tag{3}$$

where DN_e is the commercially available diameter of pipe e (mm), V_{design} is the design velocity (m/s), and $F(e)_v$ is the velocity factor of pipe e (–).

In the design procedure, the design velocity (V_{design}) is varied from 0.5 to 2.5 m/s with the steps of 0.01 m/s, resulting in 201 different solutions for each pipe. Previous works (Sitzenfrei *et al.* 2020, 2021) used the same design velocity for all pipes in one design step. However, the current study considers the velocity factor (F_v) for each diameter class to explore the potential of having different velocities in each design step. Moreover, the potential dynamic weighting functions are investigated to better align the design approach with hydraulic behavior and improve the GTA design solutions.

An illustrative simple WDN is shown in Figure 2 to explain pipe sizing procedures based on static and dynamic weights. Note that for the simplicity in Figure 2, the velocity factor is assumed to be the same ($F_v = 1$) for all the pipes. Besides, pipe length (L_e) is used as the basis for the weighting functions ($W_e = L_e$). For EBC^Q_S (see Figure 2(a)), the procedure starts with a non-zero demand node (e.g., N2), and its demand is assigned to the corresponding shortest path ($SPL_{N2,S}$) to the source node. Moving to the next demand nodes, the procedure is conducted for all the nodes and EBC^Q_S value of each pipe is determined.

We know from the hydraulic behavior that water chooses the path that has the least hydraulic resistance. Hence, from the static weight perspective, one can realize that the least resistant path for EBC^Q_S design is always SPL. Consequently, EBC^Q_S values in Figure 2(a) are biased toward the right-hand branch (i.e., shortest path), neglecting the potential redundant capacity of the left-hand branch, as it belongs to a slightly longer path. This is where dynamic weighting functions can be utilized to avoid biasing EBC^Q values in just a few edges and consider the potential redundancy of alternative routes.

The proposed dynamic approaches focus more on identifying the desired/ideal flow distribution within WDNs. The flow distribution can be considered ideal when friction losses are balanced in loops by utilizing alternative routes rather than solely relying on the shortest path. This becomes particularly important when dealing with specific large demands in
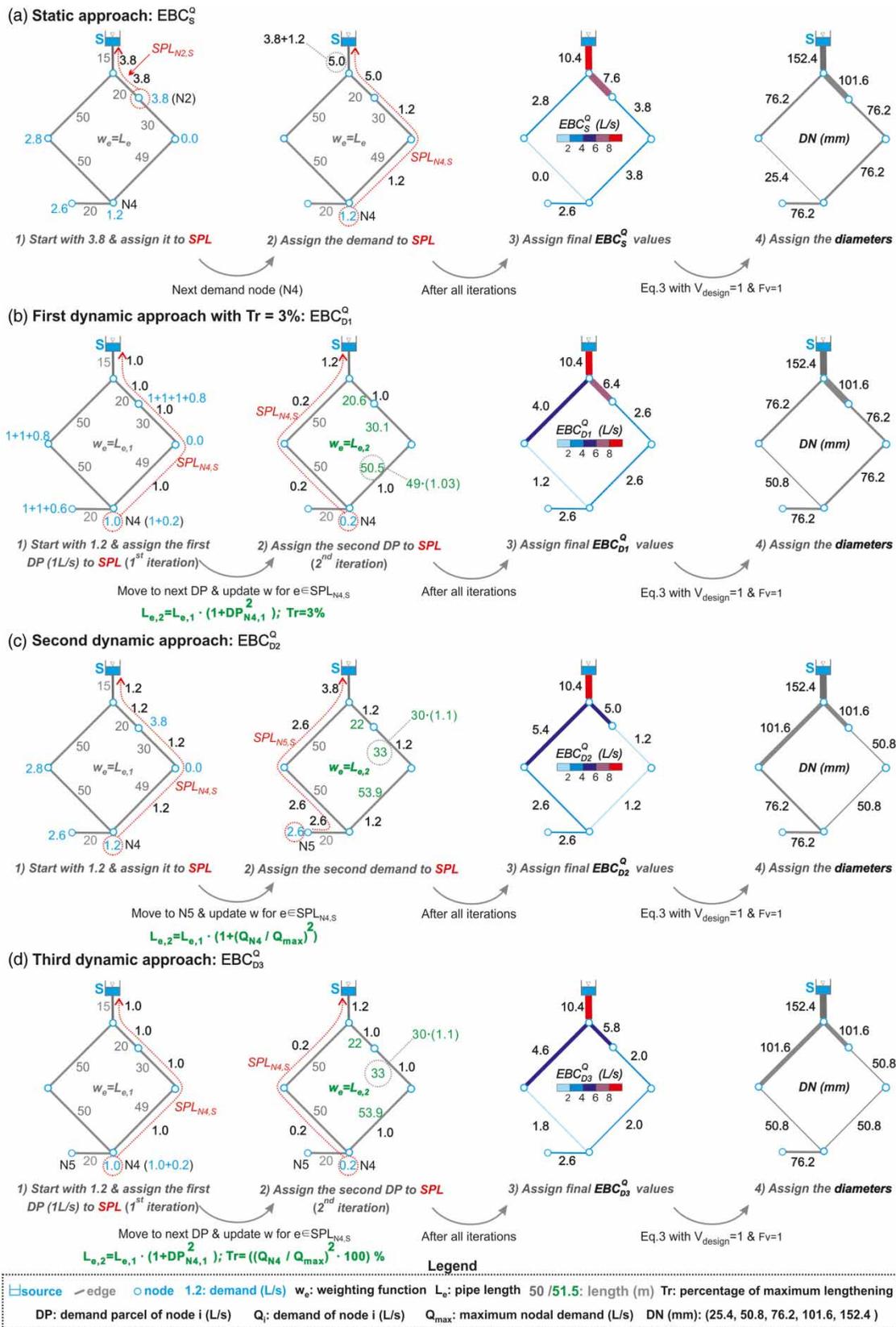
Figure 2 | Design procedure based on static and dynamic weights. Please refer to the online version of this paper to see this figure in colour: https://dx.doi.org/10.2166/hydro.2023.091.

WDNs. Under such circumstances, the ideal flow distribution from the source to the nodes with large demands utilizes, if available, multiple paths fulfilling the energy balance. To capture this real-world behavior of flow division in WDNs, our first dynamic approach (EBC_{D1}^{Q}) is developed (shown in Figure 2(b)). According to this approach, we need to split nodal demands into smaller demand parcels (DPs) to imitate the flow division process. For instance, two demand parcels ($DP_{N4} = DP_{N4,1} + DP_{N4,2}$) with the values of 1 and 0.2 L/s are considered for N4 with a total nodal demand of 1.2 L/s. The process of the first dynamic approach is initiated with the node that has the minimum demand (N4), and its first DP (1 L/s) is assigned to the edges where SPL ($W_e = L_e$) is minimal. Once the first parcel is routed from N4 to S, the friction increases along the red-colored path. Therefore, to balance friction losses within the WDN, the second DP (0.2 L/s) should be routed through the left-hand branch as an alternative route. In order to reproduce this hydraulic behavior for the GTA design, the edge weights related to the red-colored path in the first iteration are updated/increased by $L_{e,2} = L_{e,1} \cdot (1 + DP_{N4,1}^{2})$ up to a fixed threshold value (Tr), so that the second DP can select the left-hand branch. Tr is the percentage of maximum lengthening per iteration, which prevents edge weights from the excessive increase. For instance, Tr in Figure 2(b) is considered 3%, which means for DP < 0.173 L/s, the weights are updated with $L_{e,1} \cdot (1 + DP^2)$, and for DP $\geq$ 0.173 L/s, the factor of $L_{e,1} \cdot (1.03)$ is used as the maximum lengthening. Different values can be considered for Tr, resulting in activating various paths in the EBC_{D1}^{Q} design process. A sensitivity analysis is conducted in this paper to identify proper Tr values based on the characteristics of WDNs. It is worth mentioning that variations in the flow division could somewhat impact the overall flow distribution within WDNs. This inherent variability within loops can be a source of uncertainty. However, our investigations on the potential variations in flow division across multiple WDNs revealed that the proposed approach (Figure 2(b)) provides more reliable results for achieving ideal flow distribution.

The second dynamic approach (EBC_{D2}^{Q}) is also initiated with a minimum demand (1.2 L/s); however, instead of demand division in each node, the total nodal demand is routed through the edges. In this approach, no threshold is needed to be defined for the maximum lengthening. As indicated in Figure 2(c), the updated edge weights in the second iteration are a function of the nodal demand (Q_{N4}) and the maximum demand (Q_{max}), calculated as $L_{e2} = L_{e,1} \cdot (1 + (Q_{N4}/Q_{max})^2)$. This lengthening function implies that the larger the nodal demand, the more capacity for alternative paths is activated. It is worth mentioning that the idea of applying these lengthening functions is derived from the Darcy–Welsbach equation, where there is a quadratic relationship between friction losses and water flow. Using these dynamic weights results in creating loops in the course of the design procedure and consequently assigns pipes with more capacity to the left-hand branch of the WDN in Figure 2(b) and 2(c) compared to those in Figure 2(a).

The third dynamic approach (EBC_{D3}^{Q}) is the combination of the first and second approaches. As illustrated in Figure 2(d), $L_{e,2} = L_{e,1} \cdot (1 + DP_{N4,1}^{2})$ is used to update the weights in the second iteration; however, its maximum value is limited to $L_{e,1} \cdot (1 + (Q_{N4}/Q_{max})^2)$. In other words, while the first approach (Figure 2(b)) uses a fixed value (3%) for Tr in all iterations, the third approach (Figure 2(d)) adjusts Tr in each iteration based on $((Q_i/Q_{max})^2 \cdot 100)$%, which is proportional to the nodal demands and changing when moving from one demand node to another. For example, starting from N4 in Figure 2(d), Tr is set to 10%, and for the next node (N5), it changes to 47% as N5 has a larger demand.

The proposed dynamic weighting approaches are systematically investigated by applying them to different single-source WDNs. Thereafter, suggestions for selecting appropriate weighting functions based on network size and characteristics are provided. The selected weights are then utilized to get the best approximation for the Pareto design of multi-source WDNs.

The dimensionless velocity factor F_v (–) is another term of Equation (3) that needs to be determined to consider the effect of different velocities in each design step. For this purpose, the economic flow velocity V_{eco} recommended by standards codes is used as a suggestion for F_v values. Table 1 shows the recommended V_{eco} and optimal water flows (Q_{opt}) for diameter classes (Baur *et al.* 2019). The value of $F(e)_v$ (velocity factor of edge e) is equal to the magnitude of V_{eco} corresponding to Q_{opt} that is immediately higher than the calculated $EBC^Q(e)$. For instance, to determine F_v for the main pipe of the toy example in Figure 2, its EBC^Q value (10.4 L/s) is compared with Q_{opt} values in Table 1, and V_{eco} (0.85) corresponded to the next larger Q_{opt} (15.5 L/s) can be assigned as the velocity factor (i.e., $F_v = 0.85$).

2.1.3. Source tracing for multi-source WDNs

In multi-source WDNs, demand nodes are supplied by a source(s) based on nodal head values. Therefore, in order to design such networks based on GTA, supplied areas by reservoirs (sources) need to be identified through the estimation of nodal heads and decomposing WDN into subnetworks. Afterward, every subsystem can be designed individually regarding the proposed static and dynamic weighting functions.

Table 1 | Recommended economic flow velocity (V_{eco}) for different pipe diameters (Baur *et al.* 2019)

D (mm)	76.2	101.6	152.4	203.2	254	304.8	355.6	406.4	508	609.6	711.2	812.8	914.4
V_{eco} (m/s)	0.80	0.80	0.85	0.90	0.95	1.00	1.05	1.10	1.20	1.30	1.40	1.50	1.60
Q_{opt} (L/s)	3.6	6.4	15.5	29.1	48.1	73	104.3	142.7	243.2	379.4	556.1	778.3	1,050

Figure 3 shows a simple WDN with two sources, where node colors are assigned according to their corresponding source. To estimate the nodal head $H_{j,S}$ and decompose the network, the head loss $hf_{j,S}$ along the path from S to j is subtracted from the source head H_S (i.e., $H_{j,S} = H_S - hf_{j,S}$). From the hydraulic point of view, the head loss hf_e along pipe e is estimated with the Darcy–Weisbach equation as follows (Rossman 2000) (SI units):

$$hf_e = \frac{f_e L_e}{12.1 D_e^5} \cdot Q_e^2 \text{(m)} \tag{4}$$

where f_e is the friction factor, which depends on the flow regime, L_e is the pipe length, D_e is the pipe diameter, and Q_e is the water flow.

The goal of graph-based source tracing is to simplify the head loss equation to utilize it as a weight for the graph measures. Pipe length L_e is the only information from Equation (4) that is available in the GTA design procedure. Therefore, if we simplify Equation (4), assuming constant velocity and the same diameter for pipes, the following term is derived:

$$hf_e \approx C \cdot L_e \text{(m)} \tag{5}$$

where the coefficient C (m/km) can be interpreted as a linear hydraulic gradient (friction slope) in WDNs.

According to Equation (5), the head loss can be estimated using the term $C \cdot L_e$ in conjugation with the weighting function of the shortest path. Therefore, the graph-based head loss for the flow path between S and j is calculated using $SPL_{j,S}$ with the weight of $w = C \cdot L_e$ (i.e., $hf_{j,S} \approx SPL_{j,S}$). Consequently, the nodal head $H_{j,S}$ is estimated by subtracting $SPL_{j,S}$ from the head of the corresponding source (i.e., $H_{j,S} \approx H_S - SPL_{j,S}$). After comparing the calculated nodal heads based on each source, node j is assigned to the source where it has a higher $H_{j,S}$. The WDN is then decomposed into subnetworks based on the source tracing results, and each subnetwork is designed using the proposed GTA with static or dynamic weights. This method is applied to one benchmark and two large WDNs, and recommendations are given for the coefficient C selection.
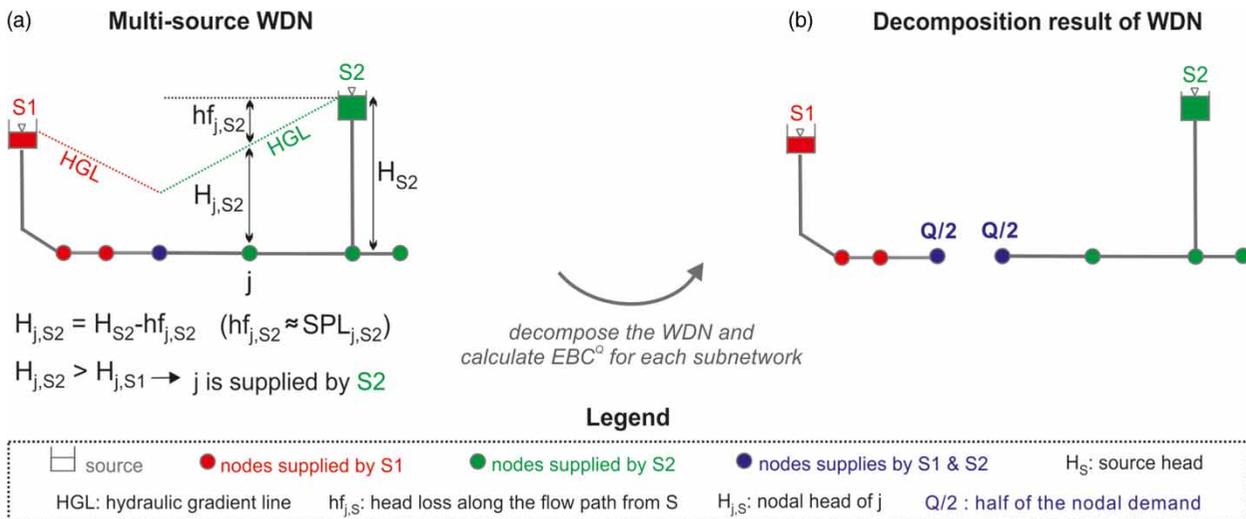
Figure 3 | Graph-based source tracing for multi-source WDNs.

2.2. Multi-objective optimization for WDN design

The obtained GTA solutions from the real and benchmark case studies are compared with the results of a multi-objective optimization procedure to evaluate the accuracy and efficiency of the design solutions. Multi-objective design for real-world cases is conducted using the state-of-the-art approach called GALAXY (Wang *et al.* 2017). GALAXY was developed based on the framework of evolutionary algorithms to deal with the multi-objective and combinatorial design of WDNs. As a hybrid approach, it can explore the solutions efficiently and handle conflicting objectives effectively, making it an attractive method for WDN optimization. The methodology of GALAXY was introduced in detail by Wang *et al.* (2017). In addition, the best-known Pareto fronts in literature (Wang *et al.* 2015) are employed to validate the GTA results of benchmark cases. These best-known Pareto fronts were generated by Wang *et al.* (2015) using five state-of-the-art multi-objective evolutionary algorithms. These algorithms include two hybrid methods known as AMALGAM (Vrugt & Robinson 2007) and Borg (Hadka & Reed 2013), as well as three enhanced NSGA-based algorithms, namely NSGA-II (Deb *et al.* 2002), ε-NSGA-II (Kollat & Reed 2006), and ε-evolutionary algorithms (Deb *et al.* 2005). Comparing GTA results with the best-known Pareto fronts enables us to measure the competitiveness of GTA against the state-of-the-art evolutionary algorithms.

Two conflicting objects are used to formulate the multi-objective design for all the cases. The first object is to minimize the total cost, and the second one is to maximize resilience I_n (Prasad & Park 2004), both of which are calculated as follows:

$$\text{Minimize Cost} = \sum_{e=1}^{\#E} U_{\text{cost}}(DN_e) \cdot L_e \tag{6}$$

$$\text{Maximize } I_n = \frac{\sum_{j=1}^{\#N} U_j \cdot Q_j \cdot (H_j - H_j^{\min})}{\left[\sum_{r=1}^{\#S} Q_r \cdot H_r + \sum_{i=1}^{\#P} P_i/\gamma \right] - \sum_{j=1}^{\#N} Q_j \cdot H_j^{\min}} \tag{7}$$

In the cost formula (Equation (6)), $\#E$ is the number of pipes; U_{cost} is the unit pipe cost which is the function of discrete diameter D_e, and L_e is the pipe length.

In the resilience equation (Equation (7)), $\#N$ is the number of nodes; U_j, Q_j, H_j, and $H_j^{\min}$ are the uniformity, nodal demand, actual nodal head, and minimum head of j, respectively; $\#S$ is the number of sources (reservoirs); Q_r and H_r are the discharge and the head of source node r, respectively; $\#P$ is the number of pumps; P_i is the pump power of i; and γ is the unit weight of water.

Additionally, U_j considers the uniformity of $\#Ej$ pipes with diameters DN_{ej} connected to node j:

$$U_j = \frac{\sum_{e=1}^{\#Ej} DN_{ej}}{\#Ej \cdot \max(DN_{ej})} \tag{8}$$

The selected resilience index I_n is a widely used measure for WDN resilience analysis (Assad & Bouferguene 2022). This index offers a comprehensive assessment of network resilience by considering the structural and hydraulic aspects, including connectivity, redundancy, and capacity. It has been extensively utilized in multiple studies focusing on multi-objective WDN design problems (Wang *et al.* 2015). Therefore, the selected resilience index allows a meaningful comparison of the GTA results with those reported in the existing literature

It is worth noting that the parameters of optimization problems (e.g., diameter classes, pressure constraints) are selected according to the characteristics of case studies, which are described in the next section.

2.3. Case studies

Five WDNs are selected as case studies to investigate the capability of the improved GTA in providing optimal design solutions (see Figure 4). The considered case studies include two benchmarks (TLN and Modena), two real (medium and large), and one very large semi-real WDN (virtRome). Detailed information regarding the size and characteristics of the case studies are provided in the Supplementary Material.

For the benchmark cases (small TLN and Modena), the best-known Pareto front resulting from multi-objective evolutionary algorithms in the literature are used as benchmark values for optimization (Wang *et al.* 2015). GALAXY is used for the
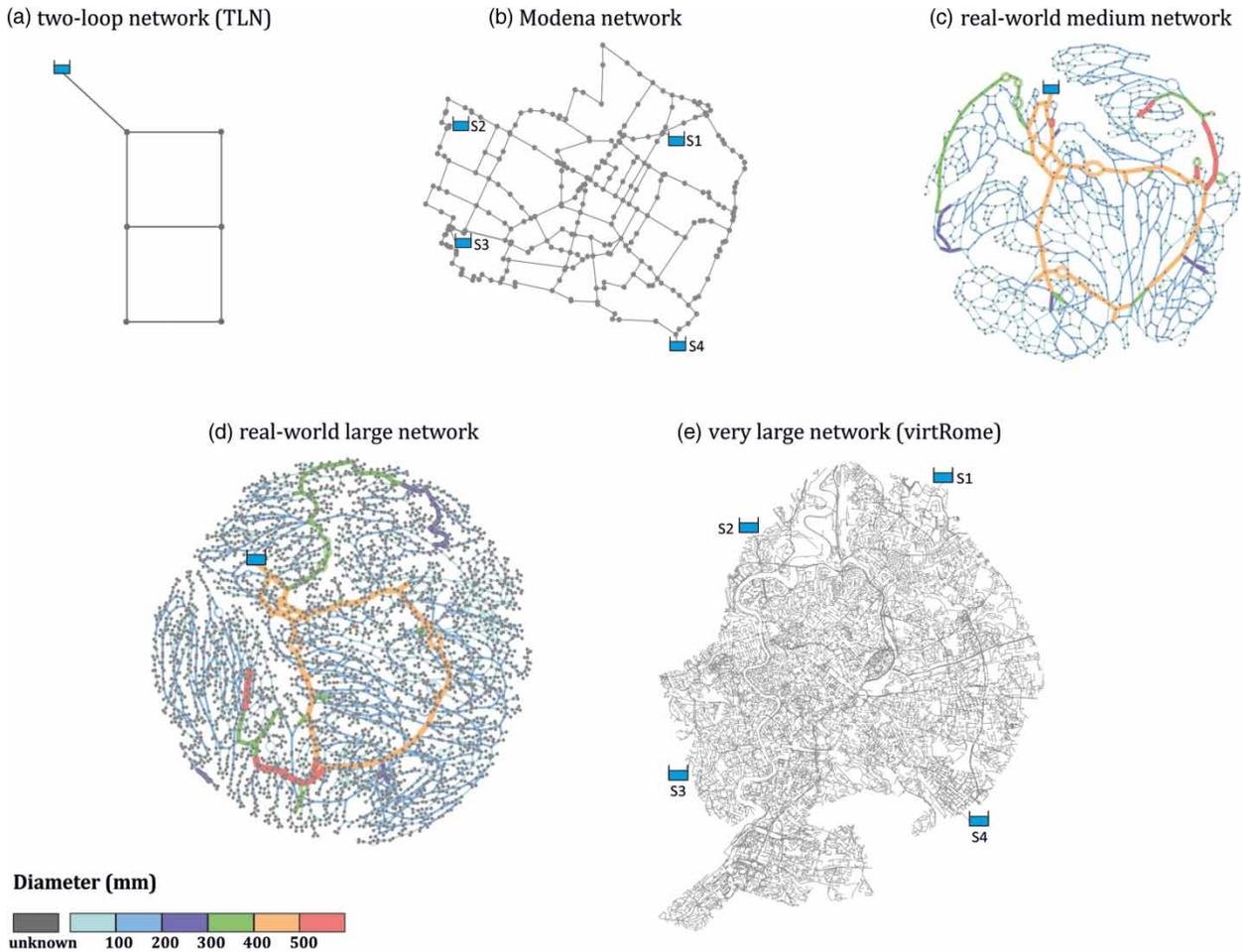
Figure 4 | Layouts of considered case studies.

optimization of medium and large networks. For the medium case, 10,000 generations with an individual population of 100 are considered. For the large case, optimization is solved with both 100,000 and 500,000 generations. Diameter classes and the related costs for TLN and Modena are considered based on the literature's values (Wang *et al.* 2015). For other WDNs, 15 pipe options in the range of 76.2–914.4 mm (3–36 inches) with unit costs ranging from 8 to 1,200 \$/m are considered.

Using evolutionary algorithms to optimize virtRome as a very large-scale network is challenging. This process demands intensive execution efforts due to the large search space size of $DN^{\#E}$ (i.e., $15^{157,044}$), which is impossible to address in a reasonable time. Therefore, only GTA is applied for the design of this WDN. Additionally, virtRome with only one reservoir (i.e., S3 with 250 m elevation) is used along with other single-source networks to identify proper dynamic weights and velocity factors proposed in GTA. In this study, the minimum pressure is set to 10 m for virtRome, 20 m for Modena, and 30 m for the other cases (Sitzenfrei *et al.* 2020).

3. RESULTS AND DISCUSSION

3.1. Role of velocity factors in design solutions

The impacts of velocity factors (F_v) on the design solutions are investigated in this section. To accomplish this, pipes of the single-source case studies were designed using static weights. Each colored dot in Figure 5 indicates one solution obtained with the corresponding design velocity, denoted as GTA. The same design velocity is used for all diameter classes in each design step of the GTA solutions, implying that F_v in Equation (3) is always set to 1. On the other hand, asterisks points denoted as GTA* represent solutions where different values are assigned to F_v (i.e., F_v is set to V_{eco} based on Table 1), imply-ing different velocities in each design step. Both GTA and GTA* results are found based on static weights (indicated by (S))
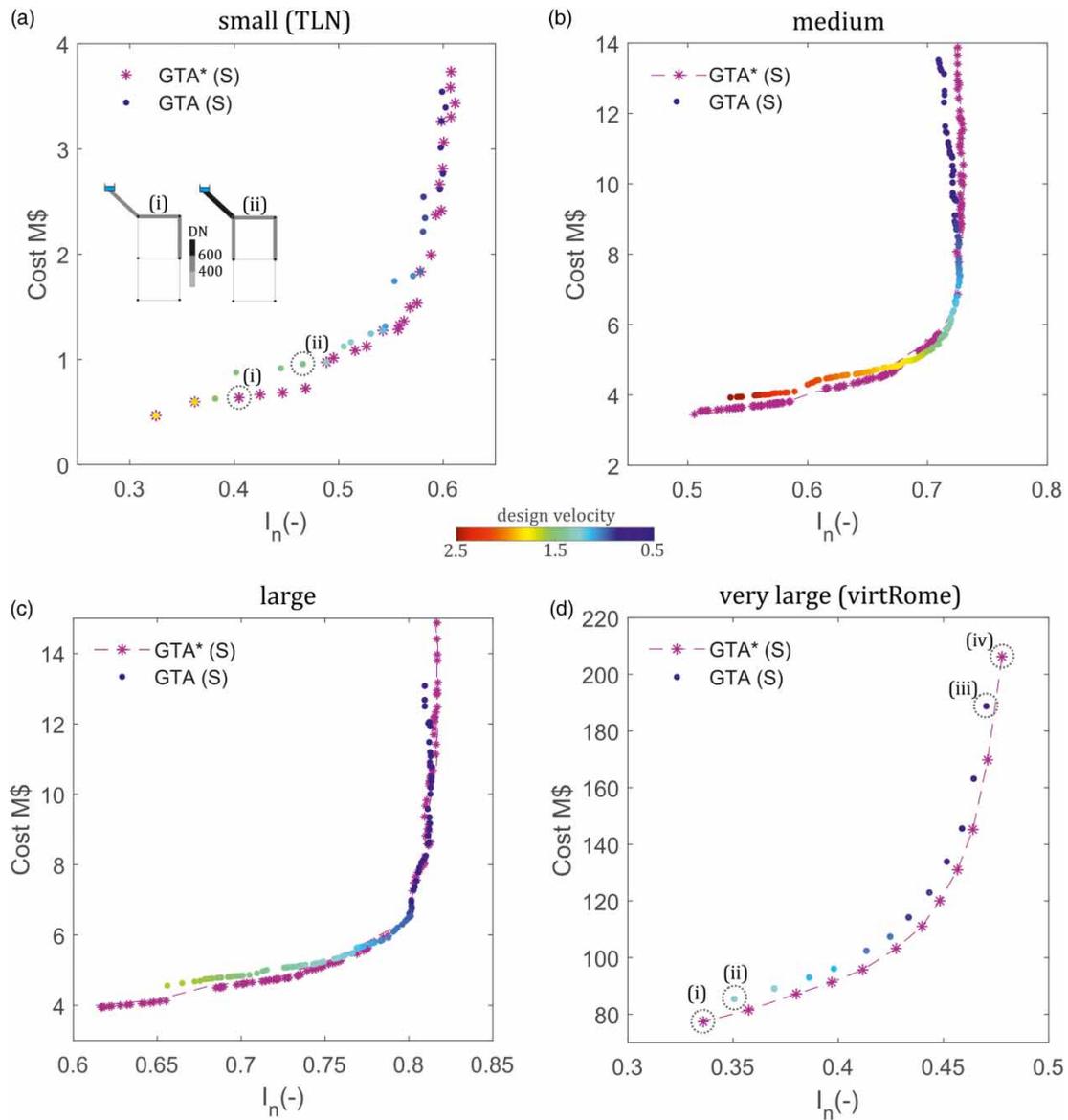
Figure 5 | Design solutions based on GTA ($F_v = 1$) and GTA*(with different F_v) using static weights.

and varying design velocities from 0.5 to 2.5 m/s with the steps of 0.1 m/s for virtRome and 0.01 m/s for the other cases. All final solutions were examined with Epanet2, and those not fulfilling the minimum required pressure were excluded. For instance, Figure 5(a) shows that the pressure constraint for TLN could not be met by the GTA solutions with a design velocity above 1.6 m/s. Due to the relative simplicity and the small number of pipes in TLN, only 24 unique GTA solutions are found for this network (e.g., solutions obtained with design velocities of 1.2 and 1.21 m/s have the same values for cost and resilience). However, when it comes to the larger WDN, like the medium one, the level of complexity increases, resulting in 201 unique solutions.

Figure 5 shows that GTA* results provide a better trade-off between resilience and cost compared to those obtained by GTA. It can be observed in Figure 5 that applying different F_v to a low-resilient GTA solution results in a new cheaper solution with lower resilience (GTA*). For example, in Figure 5(a), two solutions, (i) and (ii) with the same design velocity are indicated, where (i) is a network with smaller diameters (i.e., cheaper) and lower resilience compared to (ii). The same behavior can be seen in other case studies (e.g., (i) and (ii) in virtRome). On the other hand, when different F_v values are applied to a

highly resilient GTA solution, a GTA* solution is obtained with higher cost and resilience (see (iii) and (iv) in Figure 5(d)). Thus, it can be concluded that employing proper velocity factors can lead to higher-quality Pareto solutions and a better trade-off between the objectives.

3.2. Exploring dynamic weighing functions for WDN design

The proposed dynamic approaches, along with appropriate F_v values (identified in the previous section), are applied to the single-source WDNs to identify suitable weighting functions for each case study. Figure 6 shows the Pareto solutions obtained based on the first (D1), second (D2), and third (D3) proposed dynamic weighting functions.

As shown in Figure 6, dynamic weights applied to the small (TLN) and medium networks result in a similar pattern in some respects. However, due to the small size of TLN, only a few alternative paths exist from the dynamic weights perspective, and therefore, the differences between the Pareto designs are not significant in Figure 6(a). According to D1, the best Pareto design for highly resilient solutions is obtained when $Tr = 2\%$ is used for the small and medium cases. No further significant changes are observed for $Tr > 2\%$. This means that identical alternative routes would be activated utilizing the first dynamic approach (D1) with Tr values greater than 2%. Figure 6(b) shows that employing D1 with low Tr values (e.g., 0.1%) results in a
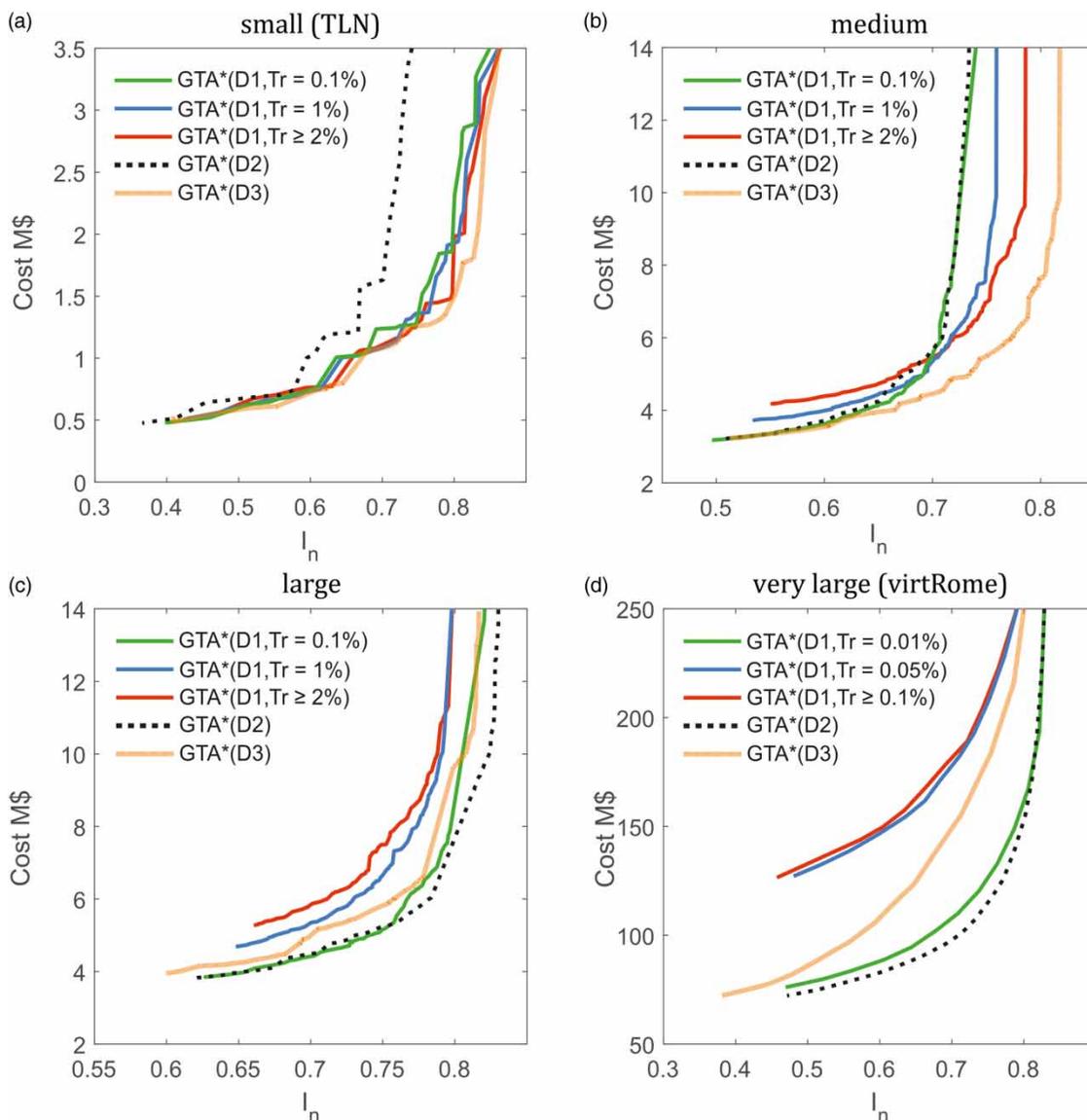
Figure 6 | Pareto solutions based on GTA* using the first (D1), second (D2), and third (D3) dynamic weighting approaches.

similar outcome as D2. This can be explained by the design procedure of D2, where nodal demands are not divided into smaller demand packages (i.e., DP in Figure 2). This consequently avoids excessive edge lengthening and acts similarly to D1 with small Tr values. Visual comparison of all the Pareto fronts indicates that the most optimal solutions based on dynamic weights for the small and medium WDNs are achieved with D3, wherein Tr in each iteration is proportional to nodal demands and maximum network demand.

Figure 6(c) and 6(d) show that increasing Tr values for the large and virtRome (with one source) yields Pareto designs with higher cost and lower resilience values. Large-scale WDNs consist of thousands of edges and nodes; therefore, numerous alternative routes exist in each design step from the dynamic weight point of view. The increase of threshold values (Tr) for these networks lengthens the edges excessively, leading to incorrect identification of the shortest path (SPL); hence a sub-optimal route is selected during the design process. Therefore, the larger the network, the smaller value should be chosen for Tr. For instance, suitable results can be obtained for the large and very large WDNs when Tr value is set to 0.1 and 0.01%, respectively. However, Figure 6 indicates that the best Pareto design for large-scale networks (especially virtRome) is achieved when D2 is applied. As guidance, using D2 or D1 (with small Tr values) for large-scale networks is recommended to obtain the best Pareto design solutions regarding dynamic weights. Besides, for small-to-medium-sized networks, D3 or D1 (with large Tr values) are a better choice.

Nevertheless, one can realize that using D2 for large-scale WDNs such as VirtRome with 150,630 nodes needs less computational time than D1 since no demand division is required. In addition, for the small and medium WDNs, the integrated parameters into the weighting functions of D3 can reproduce hydraulic behavior better than the weights used in D1. Therefore, D3 and D2 are selected in this work as ideal dynamic approaches for the small-to-medium-sized and large-scale networks, respectively.

3.3. Comparing GTA results with multi-objective optimization

After identifying appropriate velocity factors and dynamic weighting functions in the previous selection, the GTA solutions are compared with those derived from multi-objective optimization. Note that this comparison is conducted for the WDNs with a single source. In addition, only GTA results are shown for virtRome since the optimization of this large network with evolutionary approaches cannot be tackled in an acceptable time frame. Figure 7(a) shows that obtained solutions (colored dots) based on static weights, i.e., GTA (S), are inferior to the gray line obtained with optimization (OPT). The reason is that applying GTA (S) to such a network with equal pipe length cannot provide a realistic estimation of water flow in the course of design, as EBC_S^Q values of specific pipes are always equal to zero (e.g., two pipes in TLN). This issue implies that GTA (S) solutions do not benefit from the potential redundancy in the network. In contrast, using GTA with the appropriate velocity factors and dynamic weights, i.e., GTA* (D3), can significantly improve solutions, yielding competitive results with OPT.

Figure 7(b) indicates that promising design solutions are found using GTA* (D3) for the medium network. GTA* (D3) results clearly outperform those obtained with GTA (S) and OPT in a wide range of I_n values (from 0.5 to 0.8). However, GTA* (D3) cannot further employ redundant capacity for the solutions with $I_n > 0.8$, and therefore, these solutions cannot enter into competition with OPT.

For the large network, Figure 7(c) shows that the design obtained with GTA* (D2) can outperform OPT solutions after 100,000 generations with resilience values ranging from 0.62 to 0.79. Moreover, GTA* (D2) results can also outperform OPT results after 50 million function evaluations ($g = 500,000$) in a specific resilience range from 0.62 to 0.72. One can observe that GTA* (D2) and GTA (S) encompass a smaller range of design solutions compared to OPT. The reason is that the hydraulic and topographic information of the WDNs are not involved in the GTA solutions, as they are determined solely based on topology and the spatial distribution of nodal demands using EBC^Q. Nonetheless, Figure 7(c) shows that GTA* (D2) and GTA (S) cover the knee area of OPT, which is often of greater interest for decision-makers. It is worth mentioning that the resilience index used for the objective of the evolutionary approach is pressure-based, which means not all the OPT solutions may be desirable when other aspects of WDNs are considered. For instance, water quality analysis for OPT with 100,000 generations conducted by Sitzenfrei (2021) showed that the water age in particular solutions (marked with red circular) exceeds 7 d, making them less desirable for decision-makers.

Figure 7(d) indicates that solutions obtained with GTA* (D2) for virtRome can considerably outperform those found with GTA (S). The reason is that GTA*(D2) can activate alternative routes overlooked by static weights and better utilize the potential redundancy of these routes for demand distribution.
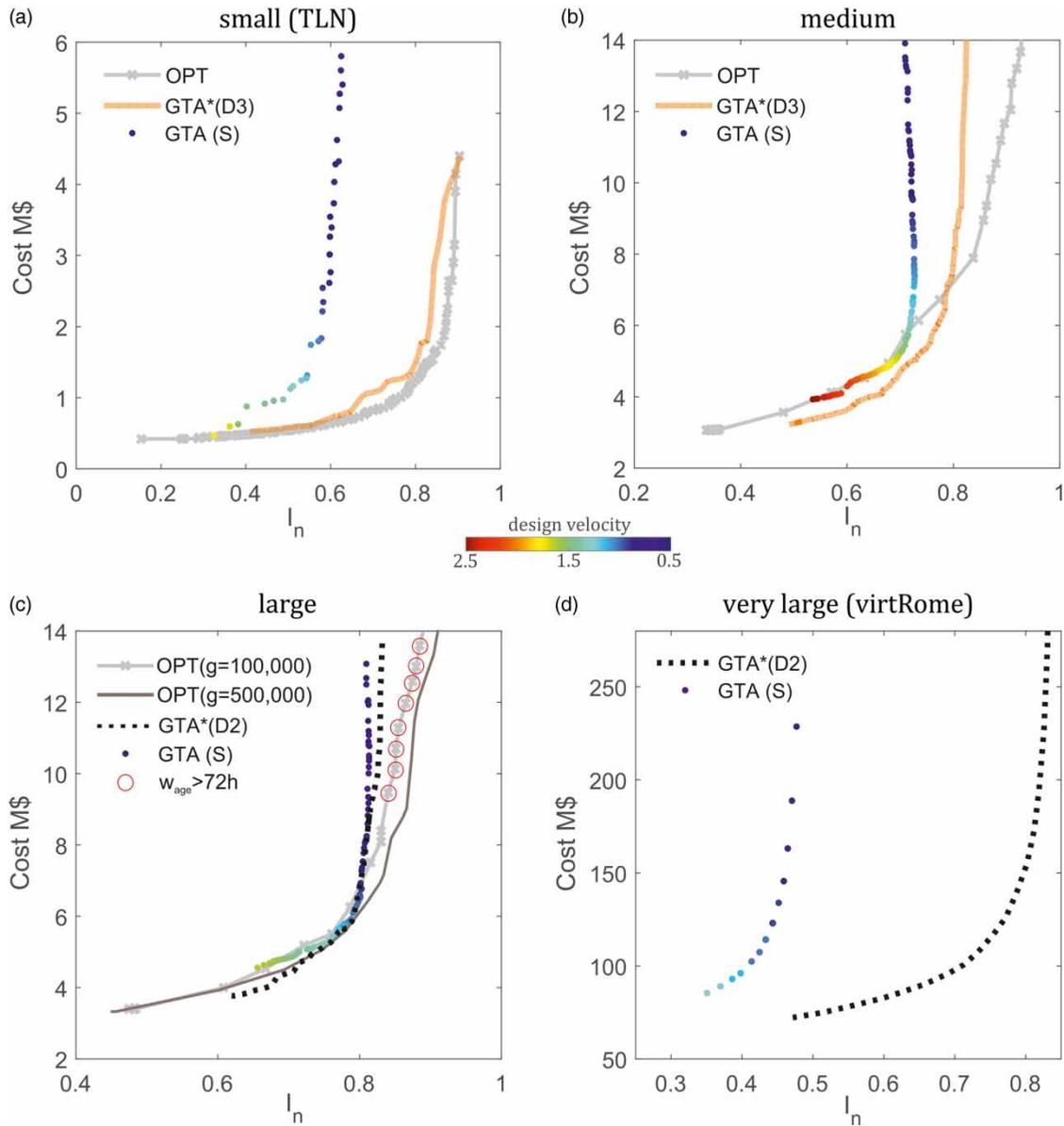
Figure 7 | Comparing GTA (S) and GTA*(D) Pareto solutions with those obtained based on optimization (OPT) with corresponding generation numbers (*g*). Please refer to the online version of this paper to see this figure in colour: https://dx.doi.org/10.2166/hydro.2023.091.

Evaluating the computational time of the approaches reveals that 37.5 h is needed to obtain OPT with 10,000 generations for the medium network. In comparison, GTA* (D3) requires only 15.3 s in a Matlab-based environment on a desktop PC (Intel® Core™ i7–8,700 CPU @ 3.2 GHz). The computational efficiency of GTA can be more highlighted for the large network, where OPT with 10 and 50 million function evaluations needs 8 and 35 weeks of execution time, respectively, and the design procedure based on GTA* (D2) is conducted only in 46.9 s. In addition, Pareto solutions for VirtRome with more than 157,000 edges are found in 71 min. This comparison confirms the high computational efficiency of GTA, which enables us to achieve a wide range of optimal (or at least near-optimal) design solutions within seconds/minutes.

3.4. Using GTA for multi-source WDNs

In this section, GTA is employed to design WDNs with multiple sources. For this purpose, GTA is first applied to the Modena network, and the results are compared with those obtained based on the evolutionary optimization (best-known Pareto front). Afterward, GTA is used to design the large-scale WDNs with multiple reservoirs (virtRome and large network). Note that due

to limited textual space, the results of applying GTA on the large network with two sources are provided in the Supplementary Material.

GTA (with static and dynamic weights) and optimization results of the Modena network are shown in Figure 8(a). Figure 8(b) illustrates the graph-based source tracing results based on different values of friction slopes (C), which are determined using SPL with the weight of $w = C \cdot L_e$ (see Section 2.1.3). In this work, three different C values are selected to investigate the role of friction slopes (linear hydraulic gradient) in designing multi-source WDNs. The source tracing results for Modena indicate that when a small value is assigned to C, source heads (HS) can play a more significant role compared to the head losses along the pipes (SPL_e). For instance, assuming $C = 1$ m/km, most nodes are supplied by S2 and S4 since they have a higher source head than S1 and S3 (see Figure 8(b)). This implies that source tracing is mainly conducted based on the source heads, and less attention is given to the losses occurring from each reservoir to the demand nodes. However, according to Figure 8(a), the best GTA solutions are found when a greater C value is selected, where the head losses (SPL_e) can also play a critical role.

Comparing GTA with OPT shows that although the GTA results cannot compete with the highly resilient OPT solutions, the outcome is still promising for the low-to-medium resilience values. As shown in Figure 8(a), using GTA with the C values of 10 or100 m/km covers part of the OPT solutions in the specific range of resilience values from 0.49 to 0.74. In terms of computational time, GTA (S) designs Modena's diameters in just 2.56 s, showcasing the advantage of GTAs in delivering optimal/near-optimal design solutions within seconds for multi-source WDNs.

The results of GTA and source tracing for virtRome are illustrated in Figure 9. Accordingly, assuming C to be 1 m/km, most nodes are supplied by S2 and S4. Similar to the results of other networks (i.e., Modena in Figure 8 and the large network in Supplementary Fig. S2), a visual comparison of Pareto fronts in Figure 9(a) confirms that a small C value is not an appropriate friction slope for GTA design. Instead, using a C value between 10 and 100 m/km results in higher-quality solutions in a specific range of resilience values, especially with the dynamic approach. Further, Figure 9(a) indicates that the Pareto fronts of GTA* (D2) outperform those derived based on static weights, again highlighting the potential of the proposed dynamic approaches for the optimal design of WDNs. Regarding Figure 9(a), the least-cost solution is achieved based on GTA* (D2) and using $C = 100$ m/km with a total cost of 69.45 M\$, which is almost half of the cost obtained in the previous studies (Sitzenfrei *et al.* 2021). Note that the suggested values for friction slopes in this work are case-specific and cannot be generalized to other WDNs with different levels of complexity and details. However, the GTA design and source tracing
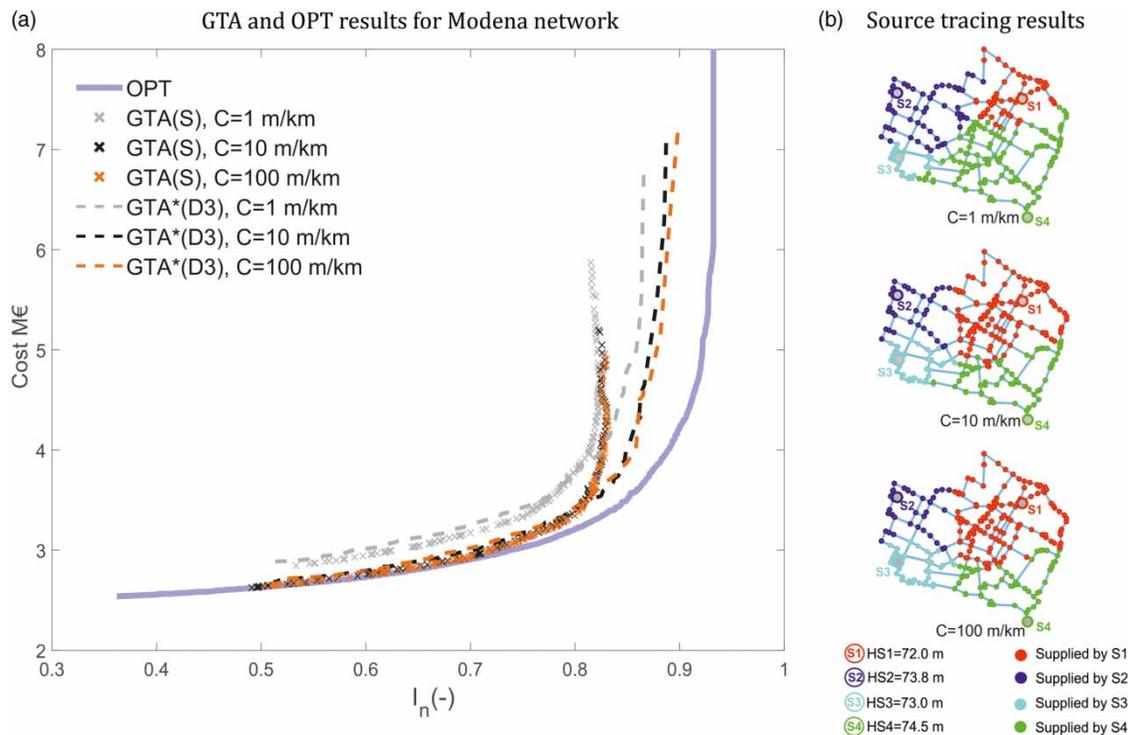
Figure 8 | (a) GTA and optimization (OPT) results of Modena WDN with four sources and (b) graph-based source tracing results.
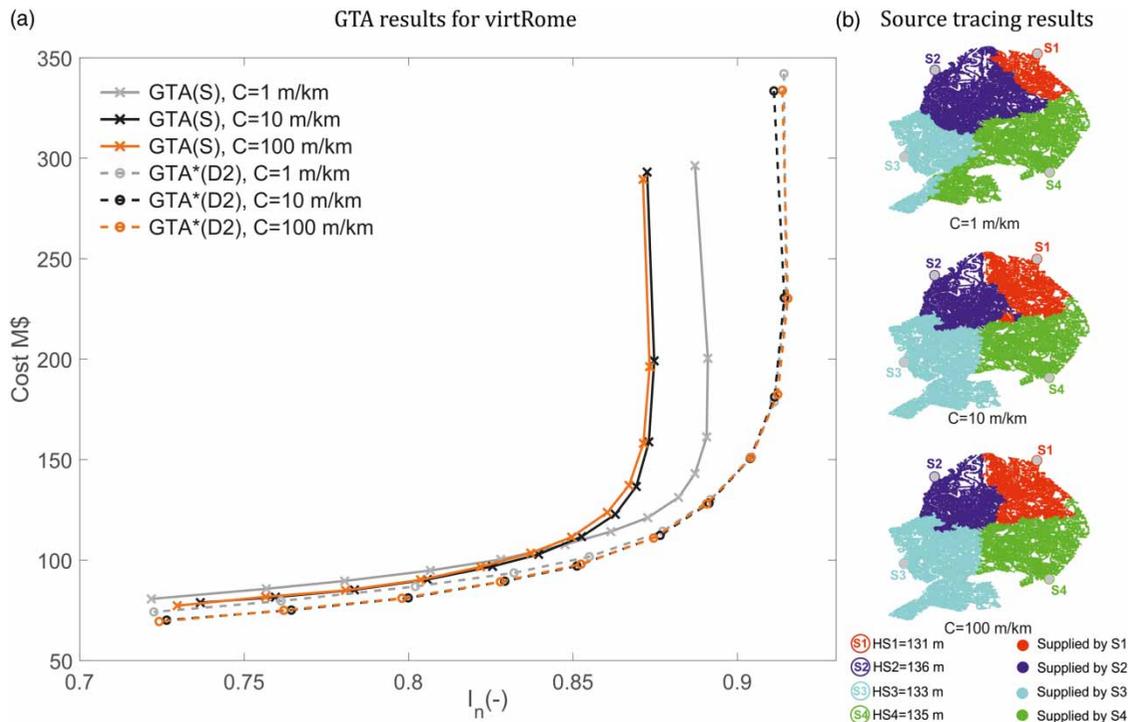
Figure 9 | (a) GTA results of virtRome with four sources and (b) graph-based source tracing results.

procedure is a fast task, facilitating the determination of the suitable friction slope for other WDNs within a reasonable timeframe.

4. SUMMARY AND CONCLUSION

This paper proposed an improved graph theory-based approach (GTA) for the optimal design of pipe diameters in WDNs, which is fast and computationally efficient. The improved GTA is based on a customized topological metric called demand edge betweenness centrality. According to this metric, a nodal demand can be routed through the shortest path, where the sum of Euclidean lengths as edge weights from a node to a source is minimal. Euclidean lengths (weights) are either constant (called static weights) or changed when iterating through all demand nodes, referred to as dynamic weights. Three hydraulic-inspired dynamic weights were suggested in this work, which could provide higher-quality design options compared to static weights, thanks to providing a better estimation of friction losses. The advantages of GTAs with dynamic approaches became more noticeable once employed for large-scale WDNs. For instance, when the improved GTA with the dynamic weights was applied to a real-world network with more than 4,000 pipes, a range of optimal solutions was found, some of which could outperform those derived with a state-of-the-art evolutionary optimization. In this case, the evolutionary optimization with 100,000 and 500,000 generations took 8 and 35 weeks, respectively, whereas the whole design process based on the improved GTA only required 47 s. This confirms the computational efficiency of the improved GTA, making it suitable for various applications, including extended design/rehabilitation problems, as well as studies requiring numerous design solutions for sensitivity and uncertainty analysis. Furthermore, it can be of interest to employ the improved GTA in combination with evolutionary approaches to reduce the search space of multi-objective optimization, leading to faster convergence to optimal solutions. In addition, the dynamic approaches proposed in this work can be adapted for other tasks of WDN analyses. For example, as recently shown by Hajibabaei *et al.* (2022), one of the dynamic approaches could be successfully adapted for pipe criticality analysis.

In this work, GTA was also used to optimize multi-source WDNs. For this purpose, the multi-source networks were first decomposed into subnetworks based on a graph-based source tracing method. Afterward, the improved GTA was employed to design each subnetwork using proper dynamic weighting functions. In the course of the source tracing, network

decomposition was highly dependent on a value selected for friction slope (linear hydraulic gradient). Therefore, a comprehensive analysis was conducted by exploring the impact of different friction slopes on three multi-source design problems to identify the most suitable values. Although some parameters and values proposed in this study are case-specific, they can be easily and efficiently obtained for other WDNs with different characteristics. This is possible as the improved GTA is a fast and straightforward procedure.

In future work, the improved GTA will be upgraded and combined with evolutionary methods to reduce the runtime of pipe resizing optimization. In addition, investigations will be conducted to incorporate topographic information alongside topological characteristics in customized graph metrics utilized for WDN design.

ETHICAL APPROVAL

Ethical responsibilities of authors are approved.

CONSENT TO PARTICIPATE

All the authors agreed with the content and participation.

CONSENT TO PUBLISH

All the authors agreed with submission and publishing.

AUTHORS' CONTRIBUTIONS

Conceptualization, R.S., M.H.; Methodology, M.H., R.S., D.S., S.H.; Software, M.H., R.S.; Original draft preparation, M.H.; writing, review and editing, M.H., R.S., S.H., A.M., D.S.; supervision, R.S. All authors have read and agreed to the published version of the manuscript.

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

Assad, A. & Bouferguene, A. 2022 Resilience assessment of water distribution networks – bibliometric analysis and systematic review. *J. Hydrol.* **607**, 127522.

Baur, A., Fritsch, P., Hoch, W., Merkl, G., Rautenberg, J., Weiß, M. & Wricke, B. 2019 *Mutschmann/Stimmelmayr Taschenbuch der Wasserversorgung.* Springer-Verlag, Wiesbaden, Germany.

Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. 2002 A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6**, 182–197.

Deb, K., Mohan, M. & Mishra, S. 2005 Evaluating the ε-domination based multi-objective evolutionary algorithm for a quick computation of Pareto-optimal solutions. *Evol. Comput.* **13**, 501–525.

De Corte, A. & Sörensen, K. 2013 Optimisation of gravity-fed water distribution network design: a critical review. *Eur. J. Oper. Res.* **228**, 1–10.

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.

Diao, K., Berardi, L., Laucelli, D. B., Ulanicki, B. & Giustolisi, O. 2022 Topological and hydraulic metrics-based search space reduction for optimal re-sizing of water distribution networks. *J. Hydroinf.* **24** (3), 610–621.

Dijkstra, E. W. 1959 A note on two problems in connexion with graphs. *Numer. Math.* **1**, 269–271.

Giustolisi, O., Ciliberti, F. G., Berardi, L. & Laucelli, D. B. 2022 A novel approach to analyze the isolation valve system based on the Complex Network Theory. *Water Resour. Res.* **58** (4), e2021WR031304.

Hadka, D. & Reed, P. 2013 Borg: an auto-adaptive many-objective evolutionary computing framework. *Evol. Comput.* **21**, 231–259.

Hajibabaei, M., Nazif, S. & Sitzenfrei, R. 2019 Improving the performance of water distribution networks based on the value index in the system dynamics framework. *Water* **11**, 2445.

Hajibabaei, M., Yousefi, A., Hesarkazzazi, S., Roy, A., Hummel, M., Jenewein, O., Shahandashti, M. & Sitzenfrei, R. 2022 Identification of critical pipes of water distribution networks using a hydraulically informed graph-based approach. In: *2022 World Environmental & Water Resources Congress (EWRI)*.

Hajibabaei, M., Hesarkazzazi, S., Dastgir, A., Minaei, A. & Sitzenfrei, R. 2023 Reconstruction of missing information in water distribution networks based on graph theory. In: *World Environmental and Water Resources Congress 2023*, pp. 1015–1026.

Kollat, J. B. & Reed, P. M. 2006 Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. *Adv. Water Resour.* **29**, 792–807.

Minaei, A., Sabzkouhi, A. M., Haghighi, A. & Creaco, E. 2020 Developments in multi-objective dynamic optimization algorithm for design of water distribution mains. *Water Resour. Manage.* **34**, 2699–2716.

Pagano, A., Sweetapple, C., Farmani, R., Giordano, R. & Butler, D. 2019 Water distribution networks resilience analysis: a comparison between graph theory-based approaches and global resilience analysis. *Water Resour. Manage.* **33**, 2925–2940.

Prasad, T. D. & Park, N.-S. 2004 Multiobjective genetic algorithms for design of water distribution networks. *J. Water Resour. Plan. Manag.* **130**, 73–82.

Pudasaini, B. & Shahandashti, M. 2020 Topological surrogates for computationally efficient seismic robustness optimization of water pipe networks. *Comput. Civ. Infrastruct. Eng.* **35** (10), 1101–1114.

Rossman, L. A. 2000 *EPANET 2 User Manual*. Cincinnati, OH, Natl. Risk Manag. Res. Lab. Environ. Prot. Agency.

Sitzenfrei, R. 2021 Using complex network analysis for water quality assessment in large water distribution systems. *Water Res.* **201**, 117359.

Sitzenfrei, R., Oberascher, M. & Zischg, J. 2019 Identification of network patterns in optimal water distribution systems based on complex network analysis. In: *World Environmental and Water Resources Congress 2019: Hydraulics, Waterways, and Water Distribution Systems Analysis*. American Society of Civil Engineers Reston, VA, pp. 473–483.

Sitzenfrei, R., Wang, Q., Kapelan, Z. & Savić, D. 2020 Using complex network analysis for optimization of water distribution networks. *Water Resour. Res.* **56**, e2020WR027929.

Sitzenfrei, R., Wang, Q., Kapelan, Z. & Savić, D. 2021 A complex network approach for Pareto-optimal design of water distribution networks. In *World Environmental and Water Resources Congress*, pp. 901–913.

Vrugt, J. A. & Robinson, B. A. 2007 Improved evolutionary optimization from genetically adaptive multimethod search. *Proc. Natl. Acad. Sci.* **104**, 708–711.

Wang, Q., Guidolin, M., Savic, D. & Kapelan, Z. 2015 Two-objective design of benchmark problems of a water distribution system via MOEAs: towards the best-known approximation of the true Pareto front. *J. Water Resour. Plan. Manag.* **141**, 4014060.

Wang, Q., Savić, D. A. & Kapelan, Z. 2017 GALAXY: a new hybrid MOEA for the optimal design of Water Distribution Systems. *Water Resour. Res.* **53**, 1997–2015.