

# Water distribution network sectorisation using graph theory and many-objective optimisation

Saeed Hajebi, Ehsan Roshani, Nicolás Cardozo, Stephen Barrett, Aidan Clarke and Siobhán Clarke

## ABSTRACT

The water distribution network (WDN) sectorisation problem is characterised by structural and hydraulic requirements that make existing graph partitioning techniques inadequate to find a good solution. Specifically, sector isolation and direct access to at least one source for each sector are not addressed. This study proposes a method to address structural requirements of water network sectorisation with minimum negative impact on the hydraulic requirements. This paper first elaborates the sectorisation problem and discusses the requirements of water network sectorisation. Then, it proposes a novel method, called WDN-PARTITION, which applies a new heuristic structural graph partitioning algorithm, combined with a many-objective optimisation procedure, to find near-optimal arrangements of nodes into sectors. The criteria of optimisation and their priorities can be specified for each case. The outcome of the method is a set of non-dominated sectorisation solutions, ranked lexicographically based on their values for the chosen criteria and their priorities, from which the final decision can be made by the domain experts. WDN-PARTITION has been implemented and integrated with a hydraulic network simulator. The simulation-based evaluation results demonstrate that WDN-PARTITION generally achieves its design objectives to partition a water network into isolated sectors with a minimal negative impact on the hydraulic performance criteria of the network.

**Key words** | district metered area (DMA), graph partitioning, isolated district metered area (iDMA), many-objective optimisation, water distribution network (WDN), water network sectorisation (WNS)

**Saeed Hajebi** (corresponding author)

**Nicolás Cardozo**

**Stephen Barrett**

**Siobhán Clarke**

Lero, The Irish Software Research Centre,  
School of Computer Science and Statistics,  
Trinity College,

Dublin,

Ireland

E-mail: [hajebis@tcd.ie](mailto:hajebis@tcd.ie)

**Ehsan Roshani**

HydraTek & Associates,

Toronto,

Canada

**Aidan Clarke**

IBM Software Ireland Lab,

Dublin,

Ireland

## INTRODUCTION

A water distribution network (WDN) is the infrastructure that supplies drinking water to homes and businesses, and links water sources to consumers. Such networks are typically complex and dynamic, consisting of thousands of nodes (reservoirs, tanks and consumption nodes), interconnected by different types of links (pipes, pumps and valves).

Partitioning a WDN into smaller sub-networks is a strategy to manage its complexity, as advised by the International Water Association (IWA) (Morrison *et al.* 2007). Each sub-network, called a district metered area (DMA) (Burrows *et al.* 2000), is defined as a discrete area

of a distribution system, in which the quantity of water entering and leaving an area is metered. DMAs are usually created by closure of valves or completely disconnecting pipes (Water Research Centre 1980; Morrison *et al.* 2007).

The main goal of WDN partitioning is to achieve better control over the distribution of water (Chambers *et al.* 2004). Some of the benefits of partitioning a water distribution network include: enhanced leakage and burst detection and management, establishment of a permanent pressure control system (pressure zones) by providing different pressure levels, improved contamination spread control (which is

directly associated with water security) and enhanced rehabilitation and work planning (Herrera Fernández 2011). Partitioning a water network is also a major step towards a smart water grid (Hajebi *et al.* 2013a), which aims at the transformation of the existing water supply systems with ICT capabilities, to provide improved monitoring and control capacity over the network.

Recently, water security has gained a great deal of attention (National Research Council 2007; Porco 2010). Water security requires the division of the water distribution system into isolated sectors (also called isolated DMAs or *i*DMAs) ensuring no flow exchange between sectors. All water entering such a sector is consumed within it. In the event of a contamination incident, these sectors limit exposure to threat and minimise the number and the length of pipes that need to be decontaminated (Murray *et al.* 2009). Sector isolation implies that each sector must have direct access to at least one source (i.e., the path from the sector to a water source must not contain any nodes in other sectors), so that it can have access to water without having flow exchange with other sectors.

The process of partitioning a WDN into a set of isolated sectors is referred to as water network sectorisation (WNS) (Di Nardo *et al.* 2013b). This paper explores the WNS problem and proposes a novel WDN partitioning technique, called WDN-PARTITION, to solve this multifaceted partitioning problem. WDN-PARTITION applies a novel structural graph partitioning technique combined with a many-objective optimisation to find good sectorisation solutions. The key characteristics of the proposed method are as follows:

- Each sector has direct access to a water source, so the path from the sector to a source does not contain any nodes in other sectors.
- The identified sectors are isolated from each other, so there is no flow exchange between different sectors.
- The sizes of the identified sectors are within a predefined boundary.
- The candidate solutions are hydraulically analysed, and the best ones are chosen using a many-objective optimisation procedure.
- The criteria of optimisation and their priorities can be specified for each case.
- It works well for both small and large networks.

- It does sectorisation without adding any new component to the network.

This paper is organised as follows: first, the issues related to *i*DMA design are discussed and a set of WNS requirements are proposed. Then, based on the discussed requirements, the state of the art is reviewed briefly. Next, the WNS problem is formulated as a constrained many-objective optimisation problem. The next section explains how the proposed solution (i.e., WDN-PARTITION) addresses the requirements and challenges of the problem. The proposed method is evaluated in the next section. Finally, the last section concludes the paper and discusses future work.

## WNS REQUIREMENTS

Designing highly looped networks is the strategy applied in water distribution system design to guarantee high reliability of the network (Walski *et al.* 2003). In contrast, by cutting pipes or using isolation valves in the network re-design process to create DMAs (or *i*DMA), the loop structure of the network is reduced, which results in a decrease in network reliability. Additionally, water network partitioning causes more energy to be dissipated in the network, which may cause problems to ensure pressure requirements at some nodes, and therefore, may affect network efficiency (Alvisi & Franchini 2014). Furthermore, water may remain at dead-ends, which are created by disconnecting pipes and/or closing valves. This may cause water quality problems, especially water age issue, which is a surrogate for water quality (Walski *et al.* 2003).

Nonetheless, if partitioning is performed correctly, it can bring advantages in terms of water security and leakage detection, without highly compromising network reliability or water quality (Murray *et al.* 2009; Di Nardo *et al.* 2012, 2015). To perform partitioning correctly, the requirements of partitioning should be taken into account. In other words, a set of criteria and metrics should be defined in advance that reflect WNS requirements. These requirements should be considered in the sectorisation process, and the outcome of sectorisation should be evaluated using the defined metrics.

Despite the importance of defining WNS requirements, each of the existing approaches focuses on some particular

requirements and does not consider the other ones. To have a basis for this study, a general set of requirements is needed. To come up with a general set of WNS requirements, the following methodology has been used:

- First, the IWA guidelines on DMA design (Morrison et al. 2007) are considered: size (geographical area and number of customer connections), elevation of nodes, pressure requirements, number of pipes to be cut, number of meters to be installed, infrastructure conditions and minimum number of mains crossed by sectors.
- Second, the initial list of criteria has been complemented with that discussed on the state of the art in WDN partitioning.
- Finally, the list of criteria has been discussed and modified with experts and practitioners in WDN partitioning.

The outcome of this process is the following set of WNS requirements, which are as comprehensive as possible within the scope of this work.

- *Water security* requirements.
- *Isolation of sectors*: i.e., no flow exchange between different sectors is allowed.
- *Direct access* to water source for each sector must be guaranteed: i.e., the path from a sector to at least one source must not contain any nodes in other sectors.
- *Connectedness*: i.e., the partitioned sectors in the WDN should be connected to guarantee that all the nodes have access to water. There must be a path from the source to all the nodes in an identified sector, and there should be no isolated node in the network after partitioning.
- *Conservation of mass* and *conservation of energy*. Two fundamental hydraulic constraints that govern the physics of hydraulic networks and should be considered in all water distribution network design and re-design problems. *Conservation of mass* can be regarded as continuity of flow constraint, which dictates that the fluid mass entering any pipe will be equal to the mass leaving the pipe (since fluid is typically neither created nor destroyed in hydraulic systems). *Conservation of energy* can be regarded as hydraulic equilibrium equations, which imposes that the difference in energy between

two points in a network must be the same irrespective of flow path (Walski et al. 2003).

- *Pressure requirements*: i.e., the pressure at nodes should not exceed certain limits, i.e., a minimum of 28 (m) and a maximum of 100 (m) (US Army Corps of Engineers 2004). Additionally, minimum nodal pressure requirements should be guaranteed. Both of these requirements should consider the following scenarios (Walski et al. 2003; Morrison et al. 2007):
  - average day demand (ADD),
  - minimum hourly demand (MHD),
  - peak hourly demand (PHD),
  - maximum day demand (MDD),
  - fire demand (FD).

It should be noted that from the hydraulic point of view, whenever a system is able to satisfy the MDD it will be able to satisfy MHD, PHD and ADD as well. For water quality modelling mainly ADD is used. Therefore, {MHD, PHD, MDD + FD} cover all the five scenarios, as FD is modelled with MDD to ensure the capability of the system to provide fire demand at the maximum day demand, and a system with such a capacity will definitely satisfy ADD.

- *Sector sizes* must be balanced and within a predefined boundary (Morrison et al. 2007). Boundaries like (300–2,500), (500–3,000) and (500–5,000) customer connections can be found in the literature (Morrison et al. 2007; Di Nardo & Di Natale 2011; Herrera Fernández 2011); however, (500–5,000) is the most common one.
- *Customer demand* satisfaction: i.e., enough water should be supplied to satisfy different customer demands at all times during the different consumption scenarios (Diao et al. 2013).
- *Elevation* of the nodes in a DMA should be within a specific range (Morrison et al. 2007).
- *Limited water velocity*: i.e., the minimum and maximum velocity of water in pipes should be within certain limits (Gomes et al. 2012). Typically, a minimum of 1 m/s (3 ft/s) for large zones and a maximum of 3 m/s (10 ft/s) are recommended (Walski et al. 2003).
- *Minimum cut size*: i.e., the number of pipes that must be cut (and equipped with flow meters and valves) should be minimised (Morrison et al. 2007).
- *Minimum cut weight*: i.e., sum of diameters of the pipes that must be cut should be minimised.

- *Tank level* limitations: i.e., the level of water in tanks should be within a boundary. The maximum allowable level is typically the top of the tank, and the minimum allowable water level is typically above the bottom of the tank to provide some residual storage for potential fire defeat events. Additionally, the difference between the water level at the start and the end of simulation period should be limited (usually less than 10% of the tank height).
- Maximum *network reliability* (Murray et al. 2009): i.e., the reliability of the network should be maximised. Resilience (Todini 2000) is a surrogate for network reliability, defined as the inherent capacity of a water network to manage unexpected failures. Network resilience is measured as 'the ratio between the surplus of power delivered to users and the maximum power that can be dissipated in the network when meeting exactly the design criteria' (Grayman et al. 2009).
- Maximum *network efficiency*: i.e., the energy efficiency of the network should be maximised. *Dissipated power* is a surrogate for energy efficiency (Di Nardo et al. 2013b), which is the power dissipated in the pipes during water flow from the source to the users; i.e., the total available power at the entrance of the distribution network minus the power that is delivered to the users. Dissipated power should be minimised to have the maximum energy efficiency in the network.
- *Water quality* considerations (Murray et al. 2009). *Water age* is usually considered as a replacement for water quality, which should be minimised.
- Minimum *background leakage*. Background leakage is characterised by small non-visible leaks that occur mostly at joints and fittings, not generating sufficient noise to be detected by existing equipment (Delgado 2008; WDSA 2014 team). It should be minimised in the network.
- *Sectors should cross as few mains as possible* (Morrison et al. 2007). The sector boundaries should follow the natural geographic and hydraulic boundaries. The final goal is to minimise the costs, both CAPEX and OPEX.
- Minimum *costs*. CAPEX (capital expenditure) is an expense incurred for WDN partitioning, e.g., expenditure on assets like pipes, valves, meters and installations. OPEX (operational expenditure) is the required day-to-day operation cost of the network, like valve operations, pump operations, maintenance and repairs. The total

running cost of the sector creation and operation (CAPEX + OPEX) should be minimised.

The above list can be divided into three categories:

- Structural requirements, which are related to the structure of the network after partitioning, including sector isolation, direct access, connectedness, sector size, sector size balance, minimum cut-set size, minimum cut-set weight and, finally, sectors should cross as few mains as possible. Of these structural requirements, sector isolation and direct access are specific to WNS (i.e., to address water security); while the other ones are general in WDN partitioning, including sectorisation.
- Hydraulic requirements, which are hydraulic constraints and objectives that need to be maintained in water networks, including conservation of mass, conservation of energy, limited pressure at nodes, limited water velocity in pipes, limited tank level, customers demand satisfaction or guaranteeing nodal pressure requirements, network reliability, energy efficiency, minimum nodal elevation differences within the sectors, water quality, and background leakage. Violation of one of these requirements could result in damage to the network infrastructure or create problems in service provisioning. These requirements must be satisfied in all types of water distribution network design and re-design, including partitioning and sectorisation.
- Economic requirements, that deal with the total cost of sectorisation, i.e., CAPEX + OPEX.

Although some of the mentioned requirements are basic hydraulic requirements of a water distribution system, they should be taken into account for WDN partitioning and should be re-examined after the process. In some sources (Charalambous 2005; Murray et al. 2009; Grayman et al. 2009), the process of DMA design is referred to as network re-design. The structure of the network will be modified and some of the hydraulic requirements may not hold after partitioning. Therefore, it is important to assess all the requirements of network design again.

The priorities for these requirements may depend on the local situations, policies and legislations. However, as this work focuses on water network sectorisation, water security requirements (sector isolation and direct access) are highly prioritised.

## RELATED WORK

Graph theory can be used to address the structural requirements of WNS. The WNS problem can be reduced to three different classic graph theory problems (Hajebi et al. 2014a): balanced graph partitioning (Andreev & Racke 2006), constraint graph clustering (Davidson 2005) and graph edge deletion (Yannakakis 1978). However, existing techniques are not suitable for handling water security related requirements in graph theory, i.e., sectors' isolation and direct access to sources for each sector.

On the other hand, existing solutions for WDN partitioning do not address all the requirements of water network sectorisation. In particular, water security is not addressed in most of the existing approaches; therefore, they cannot handle WNS. Three approaches address water security; however, they cannot handle large networks and/or do not address other WDN partitioning requirements. Herrera Fernández (2011) uses a multi-agent approach to divide a WDN into a collection of isolated sectors. It addresses sector isolation and direct access to water sources for each sector. However, no hydraulic requirement is considered in this work. Di Nardo et al. (2013a) identify isolated sectors and address direct access to water source for each sector; however, a number of hydraulic requirements are not addressed, including pressure requirements, minimum nodal elevation differences, sector size, limited water velocity in pipes, crossing few water mains, and water quality. More importantly, in both Herrera Fernández (2011) and Di Nardo et al. (2013a), as the number of sectors that can be created equals the number of main sources, for large networks, some of the resulting sectors may be larger than the maximum allowable size, especially if the size of the network divided by the number of sources is larger than the maximum allowable sector size. Ferrari et al. (2014) and Savic & Ferrari (2014) address sector isolation, direct access to a source for each sector, and sector size limitations to some extent; however, they do not address some WDN partitioning requirements. For example, pressure requirements for some scenarios, energy efficiency, limited water velocity in pipes, and minimum nodal elevation differences within the sectors are not considered in this work.

## PROBLEM STATEMENT

Water network sectorisation involves assigning all nodes in the network into different sectors, and deciding how many and which links should be closed to define the isolated sectors (Ferrari et al. 2014). This can be achieved by identifying and closing sector boundary links. In other words, the status of links (here, pipes or valves) being open or closed determines sector boundaries. Therefore, WNS can be formulated as a least-cost optimisation problem with a selection of link statuses as the decision variables. The links' layout and their connectivity, nodal demand, nodal elevation and minimum head requirements are assumed to be known.

In a general form, the WNS problem can be stated as: find the best combination of link statuses to partition a given WDN into a collection of isolated sectors, optimising the related objectives while satisfying the corresponding constraints. Objective functions reflect the metrics that are derived from WNS requirements and quantify them. Constraints are the conditions that must be held to have a feasible solution, also derived from the WNS requirements. There are three categories of objectives derived from WNS requirements: structural objectives, hydraulic objectives and economic objectives. Similarly, there are three categories of constraints in this problem: structural constraints, hydraulic constraints and explicit bound constraint that limits the domain of the decision variables: link statuses are binary variables, bounded to  $\{0, 1\}$ .

Mathematically, the problem can be formulated as:

Given  $G = (\text{nodes}, \text{links})$  a graph modelling a water distribution network, where  $\text{nodes} = \{n_1, n_2, \dots, n_N\}$ ,  $\forall n \in \text{nodes}$ ,  $\text{lable}(n) \in \{\text{source}, \text{consumer}\}$  and  $\text{links} = \{l_1, l_2, \dots, l_M\}$ ,  $\forall l \in \text{links}$ ,  $\text{label}(l) \in \{\text{pipe}, \text{valve}, \text{pump}\}$  such that  $l_j = \langle p, q \rangle$  where  $p, q \in \text{nodes}$ ,  $1 \leq j \leq M$ . A partition of  $G$  into connected components is defined as DMAs =  $\bigcup_{i=1}^k \text{dma}_i$  where  $\text{dma}_i = (\text{nodes}_i, \text{links}_i)$ , is a sub-graph of  $G$ , obtained by deleting links from  $G$  such that  $\forall i \in \{1, \dots, k\} : \text{nodes}_i \subset \text{nodes}$ ,  $\text{links}_i \subset \text{links}$  and also  $\text{node}_{s_i} \cap \text{nodes}_j = \emptyset$ ,  $\text{links}_i \cap \text{links}_j = \emptyset$ .

The goal is to find the optimal partition of  $G$  considering the following objectives and constraints:



**Objectives:**

- Minimise cut size:  $CS = |\text{Cuts}|$  (1)
- Minimise cut weight:  $CW = \sum_{l \in \text{Cuts}} \text{diameter}(l)$  (2)
- Minimise sector size imbalance:  $SSI = 1 - \frac{\min_i(\text{size}(dma_i))}{\min_j(\text{size}(dma_j))}$  (3)
- Minimise average customer exposure:  $AExp = \text{mean}(\text{size}(dma))$  (4)
- Minimise maximum customer exposure:  $MExp = \max(\text{size}(dma))$  (5)
- Minimise average pipe length exposure:  $ALnkExp = \text{mean}_{dma_i \in dma} \left( \sum_{\forall l \in dma_i} \text{length}(l) \right)$  (6)
- Minimise maximum pipe exposure:  $MLnkExp = \text{mean}_{dma_i \in dma} \left( \sum_{\forall l \in dma_i} \text{length}(l) \right)$  (7)
- Minimise pressure requirements violations (da Conceição Cunha & de Oliveira Sousa 2010):  $PV = \sum_{s \in \text{scen}} \sum_{j=1}^N C_{s,j}^{\text{pen}} \times \max(0, H_{s,j}^{\text{min}} - H_{s,j})$  (8)
- Minimise dissipated power (Di Nardo et al. 2013b):  $DP = \gamma \sum_{j=1}^{m_s} q_i \Delta H_i$  (9)
- Minimise elevation difference within sectors:  $ED = \sum_{d=1}^k \sigma_d$  (10)  
 where  $\sigma_d = \sqrt{\frac{1}{n} \sum_{j=1}^n (z_j - \bar{z})^2}$ ,  $\forall dma_d \in \text{DMAs}$
- Maximise resilience index (Todini 2000):  $ANR = \text{mean}_T \left( \frac{\sum_{i=1}^N D_i (H_i - H_i^*)}{\sum_{r=1}^{n_r} Q_r H_r + \sum_{j=1}^{n_p} P_j / \gamma - \sum_{i=1}^N D_i H_i^*} \right)$  (11)
- Minimise average of water age in the network in last 24 hours:  $WA = \frac{\sum_{i=1}^N \sum_{t=LT-24}^{LT} w a_i^t}{24}$  (12)
- Minimise background leakage:  $SL = \sum_{l=1}^M q_l^{\text{leaks}} (P_{l,\text{mean}})$  (13)
- Subject to (constraints):
- Connectedness of the partitioned network (PN):  $\forall u, v \in \text{nodes}$ : there is a  $u - v$  path in the PN (14)
- Direct access to a source for each sector:  $\forall dma_i \in \text{DMAs}$ ,  $\forall u \in \text{node } s_i \exists s \in \text{sources}$  (15)  
 such that there is a  $s - u$  path in the PN,  
 where  $\forall v$  in the  $s - u$  path,  $\nexists dma_j$  such that  $v \in dma_j$ .
- Isolation of sectors: Given  $u \in \text{nodes}_i$ ,  $v \in \text{node } s_j$ ,  $\nexists$  a  $u - v$  flow path in the PN (16)
- Sector size constraint:  $\forall dma \in \bar{G}: dmaMinSize < \text{size}(dma) < dmaMaxSize$  (17)
- Conservation of mass (Walski et al. 2003):  $-\sum Q_{in}^t + \sum Q_{out}^t + \text{demand}_C^t = 0$  (18)
- Conservation of energy (Walski et al. 2003):  $z_1 + \frac{P_1}{\gamma} + \frac{V_1^2}{2g} + \sum h_p = z_2 + \frac{P_2}{\gamma} + \frac{V_2^2}{2g} + \sum h_L + \sum h_M$  (19)
- Pressure constraints:  $P_s^{\text{min}} \leq P_{i,s}^t \leq P_s^{\text{max}} \quad \forall i \in \text{Nodes}, t \in T$  (20)
- Water velocity constraint:  $V_{p,s}^t \leq V_{\text{max}} \quad \forall p \in \text{links}, t \in T$  (21)

Tanks level constraints:

$$L_s^{\min} \leq L_{i,s}^t \leq L_s^{\max} \quad \forall i \in \text{Tanks}, t \in T \quad (22)$$

$$L_{i,s}^{\text{end}} - L_{i,s}^{\text{start}} \leq \text{MLD}_i \quad \forall i \in \text{Tanks}, t \in T \quad (23)$$

Decision variables domain:

$$X = (x_1, x_2, \dots, x_M), x_i \in \{0, 1\} \text{ is the status of link } i \quad (24)$$

where in (1) and (2),  $\text{Cut} = \bigcup_{j=1}^k \text{Cut}_j$  such that  $\text{Cut}_j = l_j$ , for  $l_j \subset$  links with exactly one end-node in  $\text{dma}_j$ . In (2) and (17),  $\text{size}(\text{dma}_i) = \max_t \sum_{C \in \text{nodes}_i} \text{demand}_C^t$  or  $\text{size}(\text{dma}_i)$  is the number of customer connections in  $\text{dma}_i$ . In (3), (4), (5), (6) and (7)  $\text{dma} \in$  DMAs. In (6) and (7),  $l$  is a link in  $\text{dma}_i$  which is in  $\text{dma}$ . In (8),  $s \in \text{scen} = \{\text{MHD}, \text{PHD}, \text{MDD} + \text{FD}\}$ , a set of possible demand scenarios;  $N =$  number of nodes;  $C_{s,j}^{\text{pen}}$  is the penalty cost of pressure deficit for node  $j$  in scenario  $s$ ;  $H_{s,j}^{\min}$  is the minimum admissible pressure at node  $j$  in scenario  $s$ ; and  $H_{s,j}$  is the pressure for node  $j$  in scenario  $s$  obtained from simulations results. In (9),  $\gamma$  is the specific weight of water,  $m_s = M - y$  where  $y$  is the number of boundary links and  $M$  is the total number of links in the network,  $q_i =$  flow rate,  $\Delta H_i =$  head loss for each link  $i$  in the network (Di Nardo et al. 2013a). In (10),  $k$  is the number of identified DMAs,  $n$  is the number of nodes in  $\text{dma}_d$ ;  $\sigma_d$  is the standard deviation of elevations of nodes in DMA  $d$ ;  $z_j$  is the elevation of node  $j \in \text{dma}_d$ ; and  $\bar{z}$  is the mean of elevation of all nodes in  $\text{dma}_d$ . In (11),  $T$  is the operational time period (the mean is calculated over all the operational time period),  $N$  is the number of Nodes,  $D_i$  is the demand at node  $i$  in ( $\text{m}^3/\text{sec}$ ),  $H_i$  and  $H_{i,*}$  are available and minimum required head at node  $i$  in (kPa), respectively;  $n_r$  is the number of reservoirs;  $Q_r$  and  $H_r$  are the discharge in ( $\text{m}^3/\text{sec}$ ) and head in (kPa) at reservoir  $r$ , respectively;  $n_p$  is the number of pumps in the system,  $P_j$  is the power introduced into the network by the  $j$ -th pump in (kW) and  $\gamma$  is the specific weight of water in ( $\text{N}/\text{m}^3$ ) (Todini 2000). In (12),  $i$  is the subscript of the  $i$ -th node in the network,  $LT$  is the last time step in the operational time period and  $wa_i^t$  is the water age at node  $i$  at time  $t$ . In (13), if  $P_{l,\text{mean}} > 0$  then  $d_l^{\text{leaks}}(P_{l,\text{mean}}) = \beta_l L_l P_{l,\text{mean}}^{\alpha_l}$ , and it is zero otherwise;  $l$  is the subscript of the  $l$ -th pipe;  $P_{l,\text{mean}}$  is the model mean pressure along the  $l$ -th pipe in (m);  $d_l^{\text{leaks}}$  is background leakages outflow along the  $l$ -th pipe in ( $\text{m}^3/\text{sec}$ ),  $\alpha_l$  and  $\beta_l$  are the model parameters,  $\alpha$  is unitless and  $\beta$  is in ( $\text{m}^{2-\alpha}/\text{sec}$ ); and  $L_l$  is the length of the  $l$ -th pipe in (m) (WDSA 2014 team).

In (14), Sources are the nodes in the network that supply water. In (16), a  $u$ - $v$  path is a finite sequence of links which connect node  $u$  to node  $v$ . A flow path is defined as a path which respects flow directions in the network. In (17),  $\text{size}(\text{dma}_i) = \max_t \sum_{C \in \text{nodes}_i} \text{demand}_C^t$  or  $\text{size}(\text{dma}_i)$  is the number of customer connections in  $\text{dma}_i$ . In (18),  $\text{demand}_C^t$  is the demand at demand node  $C$  at time  $t$ ; and  $Q_{\text{in}}^t, Q_{\text{out}}^t$  are the flows entering and leaving the node at time  $t$ , respectively. In (19),  $Z_1$  and  $Z_2$  are elevations at points 1 and 2, respectively, in (m);  $P_1$  and  $P_2$  are pressures at points 1 and 2, respectively, in ( $\text{N}/\text{m}^2$ );  $\gamma =$  fluid (water) specific weight, in ( $\text{N}/\text{m}^3$ );  $V_1$  and  $V_2$  are flow velocities at points 1 and 2, respectively, in (m/s);  $g =$  acceleration due to gravity, in ( $\text{m}/\text{sec}^2$ );  $h_p =$  pumping head gain, in (m);  $h_L =$  head loss in pipes, in (m); and  $h_M =$  head loss due to minor losses, in (m) (US EPA 2005). In (20),  $P_s^{\min}$  and  $P_s^{\max}$  are the maximum and minimum pressure requirements for each scenario  $s$ ;  $P_{i,s}^t$  is the service pressure in node  $i$  at time  $t$  in scenario  $s$ .

In (21),  $V_{p,s}^t$  is the water velocity in pipe  $p$  at time  $t$  in scenario  $s$  in (m/s);  $V_{\max}$  is the maximum water velocity allowed in (m/s). In (22),  $L_{i,s}^t$  is the level of tank  $i$  at time  $t$  in scenario  $s$  in (m).  $L_i^{\min}$  and  $L_i^{\max}$  are minimum and maximum admissible levels for tank  $i$ , respectively, in (m). In (23),  $L_{i,s}^{\text{end}}$  and  $L_{i,s}^{\text{start}}$  are the levels of the tank  $i$  at the end and the beginning of the simulation in the scenario  $s$  in (m), and  $\text{MLD}_i$  is the maximum level difference for tank  $i$  in (m).

## METHOD

As discussed in the section 'Problem statement', WNS is a multifaceted problem embracing graph-theoretic, hydraulic and economic aspects. The graph-theoretic aspects of the problem may be handled using different algorithms for graph partitioning, clustering and community detection

among others. However, as the goal is to partition a WDN into isolated sectors (*i*DMAs), the current methods are inadequate, and there is a need to develop an algorithm that takes the structural constraints of the problem into account. Additionally, besides the structural constraints, WNS has other objectives and constraints, therefore, it is a multi-objective optimisation problem.

There are four different approaches to deal with optimisation problems: mathematical approaches, brute force search, meta-heuristic techniques and heuristic techniques. As can be seen in 'Problem statement', objective functions and constraints are not functions of the decision variables. Therefore, it is not possible to use mathematical approaches in this context. Brute force search is not practical in this case, as the search space of the WNS problem is very large and increases exponentially by increasing the number of links. The possible combinations of link statuses that make the search space is  $2^n$ , for  $n$  being the number of links in the network. Three categories of meta-heuristic techniques have been used in different stages of this study (i.e., multi-agent systems (Hajebi *et al.* 2013b), Tabu search (Hajebi *et al.* 2014a) and genetic algorithms (NSGA-II) (Hajebi *et al.* 2014b)). However, the solutions generated by these techniques usually do not satisfy the structural requirements of the problem, as they are typically general search techniques and do not consider the specificities of the problem. This study develops a heuristic technique, which is problem-specific, adapted to the problem at hand, and takes full advantage of the particularities of the problem.

## WDN-Partition

Based on valuable insights from existing solutions (Ferrari *et al.* 2014; Savic & Ferrari 2014), this paper proposes a novel heuristic WDN partitioning technique (WDN-PARTITION), to partition a water distribution network into (sub)-optimal isolated sectors, using a many-objective optimisation procedure. Optimisation 'is the act of obtaining the best result under given circumstances' (Rao 2009); in other words, it is 'the process of determining the best design' (Parkinson *et al.* 2013). To achieve the best results under the given circumstances explained in the problem statement (cf. the section 'Problem statement'), WDN-PARTITION first satisfies water security related constraints and

generates a collection of feasible solutions (i.e., solutions that address structural requirements of sectorisation, namely, isolation of sectors, direct access to at least one source for each sector, sector size limitations, and connectedness of the partitioned network) using a novel heuristic graph partitioning technique. Then, the best solutions are identified considering the values for their objective functions (i.e., the metrics for evaluation). Therefore, the efficiency of the optimisation method has been increased radically by first reducing the search space (by generating a collection of structurally feasible solutions), and then finding the best solutions among the feasible ones, which entails radically reduced number of hydraulic simulations.

After applying the structural graph partitioning technique, the structural, hydraulic and economic objectives and constraints should be considered. The challenge is that the objective functions of the problem are not formally defined based on the decision variables (as discussed in 'Problem statement', the decision variables are the statuses of the links with values either 0 or 1). The parameters of the objective functions and constraints are affected by the decision variables, however, their relationships are not trivial to formulate. A disaggregated methodology is used to solve this challenge. Although the decision variables do not appear in the objective functions and constraints, the results of changing them can be achieved indirectly using a network solver, i.e., a water distribution network hydraulic simulator, e.g., EPANET (Rossman 2000).

Each solution will be sent from the graph partitioning algorithm to the network solver to satisfy the hydraulic constraints. Once the network solver has solved the hydraulic equations and satisfied the hydraulic constraints, it will give the values for the parameters used in the objective functions and constraints (flow, head, pressure, etc.). Then, the values of these objective functions and constraints can be calculated and checked in the optimisation procedure.

Another issue is the optimisation method. WNS is a many-objective optimisation problem, so it might not be possible to have a single solution as the best one. The candidate solutions should be compared against each other in terms of different objectives, i.e., formulae (1) through (13). There are some multi-objective optimisation algorithms that perform well in dealing with two or three objectives (Coello *et al.* 2007); however, they cannot adequately



handle more than three objectives (Jain & Deb 2013). The number of objective functions in this problem is 13, so these methods are not suitable to handle the WNS problem. Recently, some algorithms were proposed to deal with many (more than three) objectives (Jain & Deb 2013; Yuan et al. 2014; Liu et al. 2014). However, because of the structural constraints of the WNS problem (especially sector isolation and direct access), they are incapable of generating structurally feasible solutions and/or modify the solutions in such a way that the new solutions are also structurally feasible.

To demonstrate the different steps of the proposed method, the WDN for the city of Novato, California is used. This WDN is the most complex example included in EPANET 2 (Rossman 2000) which covers an area of about 150 km<sup>2</sup>. It is a dual-source network, composed of 92 junctions, two reservoirs, three tanks, which are interconnected through 117 pipes and two pumps. This network is referred to as Novato network hereafter.

In a nutshell, WDN-PARTITION first identifies major flow paths, then identifies the groups of nodes that are connected to the major flow paths but are isolated from the rest of the network, then checks size of the identified groups, and partitions them if needed. Finally, the best solutions are selected. Figure 1 illustrates a flowchart of WDN-PARTITION, which consists of the following steps.

Step 1 (Initialisation): First of all, the network data are read from an EPANET (Rossman 2000) input file. An adjacency matrix  $A$  of the network is created using such data. The  $A(i,j)$  and  $A(j,i)$  entries in the matrix are set to the diameter of a link  $l$  if there is such a link  $l$  between two nodes  $i$  and  $j$  in the network; infinity if there is a pump or valve between  $i$  and  $j$ ; 0 if there is no link between  $i$  and  $j$ .

The user is asked for the following parameters during the initialisation phase: (1) *MainsSizeThreshold*, which is the minimum threshold size for the main pipes, (2) *dmaMinSize* and (3) *dmaMaxSize*, which are the minimum and maximum size of a sector, (4) *MaxIter*, which is the maximum number of iterations in the loop in Step 5.1.2, and (5) a list of required criteria for the specific network at hand (objective functions) and their priorities (if more than one is chosen).

Then, two parameters are set based on the network data: *LinkCount* which is the number of links in the network, and a solution template (called *position*), which is a bit vector of

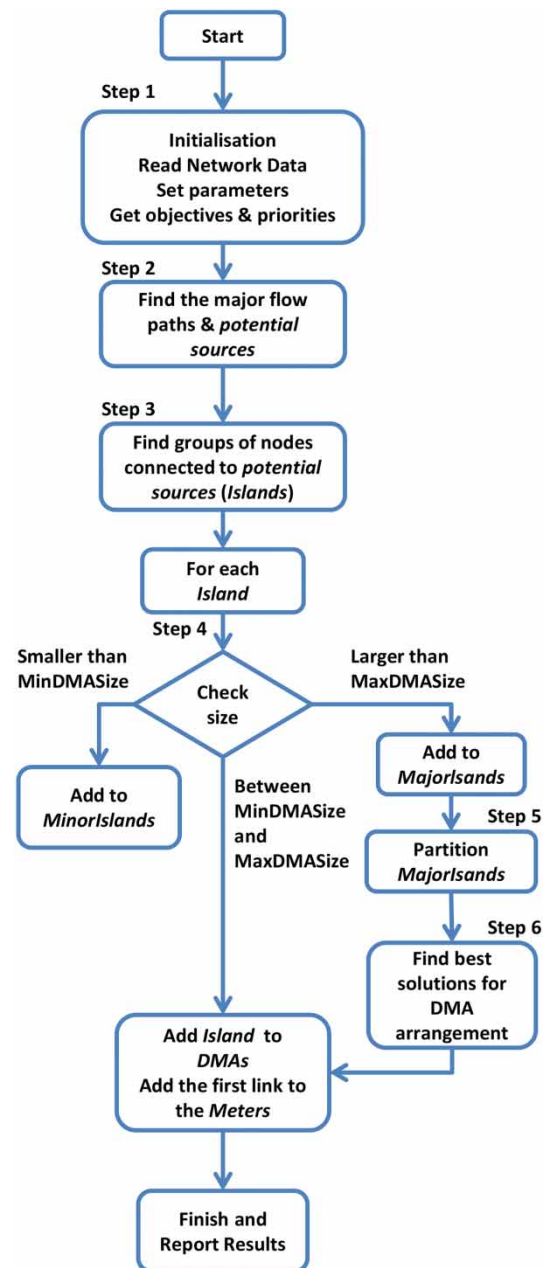


Figure 1 | WDN-Partition.

length *LinkCount* and is used as a template for possible solutions. Each bit in the *position* vector uniquely represents a link in the network; 0 represents a closed link and 1 represents an open link. The *position* vector is initialised by 1 in all places.

Finally, two sets are defined during initialisation: *Meters*, which is the set of links that should be equipped with meters

(the links that connect the DMAs to large water mains) and *Valves*, which is the set of links that should be closed off (i.e., the neighbouring links in the boundaries of different DMAs identified after partitioning).

**Step 2 (Finding major flow paths and potential sources):** Once the network data are read and the corresponding adjacency matrix is defined, it is possible to explore the network. First of all, WDN-PARTITION finds the large water mains (major flow paths) from the main sources (reservoirs) and considers nodes in these paths as new potential sources. Starting from the main sources in the network and using a modified depth-first search (DFS) (Tarjan 1972) algorithm, nodes are added to a list called *PotentialSources*. In this process, the nodes are identified in the direction of flow while their connecting links are larger than a specified threshold (*MainsSizeThreshold*). The sectors then can be identified starting from these potential sources to guarantee direct access to a source as these nodes are directly connected to a source, and would not be part of any sectors. Figure 2 shows the result of applying Step 2 on the Novato network.

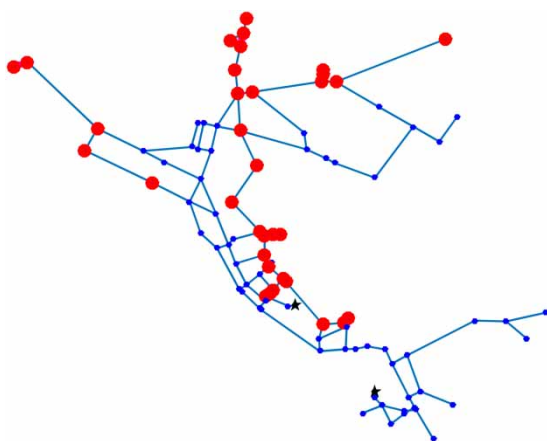
**Step 3 (Identifying Islands):** In this step, the groups of nodes which are connected to each *PotentialSources* are identified and added to a list called *Islands*. To this end, each node in *PotentialSources* is selected and a breadth-first search (BFS) (Pohl 1969) algorithm is started to find all its reachable nodes. In the first step (for the first-level neighbours), the left-hand side nodes are separated from the right-hand side nodes (or the upper-side nodes are

separated from the lower-side nodes), considering the large water mains as a reference line. This separation guarantees that the identified groups of nodes (and so, the identified sectors in the next steps) do not cross the large water mains. This is one of the advantages of WDN-PARTITION compared to the other approaches (Herrera Fernández 2011; Di Nardo et al. 2013a; Ferrari et al. 2014).

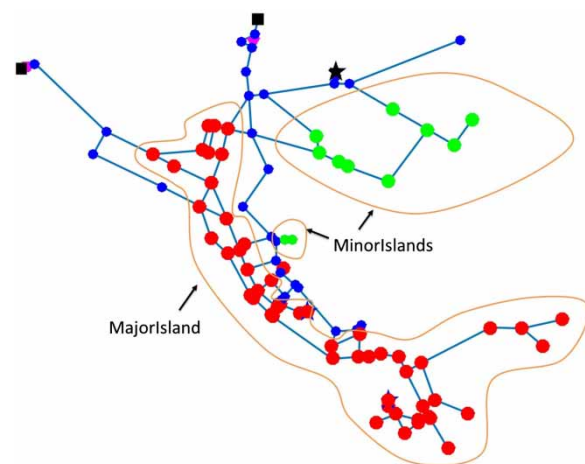
After identifying the groups connected to large water mains, the method may end up with some equal groups of nodes. This may happen for the groups of nodes that are connected to the large water mains by more than one link. Redundant groups are removed and just one of them is kept. Figure 3 shows the result of applying this step on the Novato network.

**Step 4 (Checking the size of Islands):** In this step, the sector size constraint is examined. For each group in the *Islands* list, the number of customer connections (or sum of nodal demands) is assessed. There are three possible situations:

- If size is in the range of the allowed size of a sector (between *dmaMinSize* and *dmaMaxSize*), the *Island* will be considered as an *iDMA*, and the first link from the corresponding potential source(s) will be added to the *Meters* set.
- If size is less than the minimum allowable size of a sector, it is considered as a *MinorIsland*. No meter is defined for *MinorIslands* as the cost is not justified (Ferrari et al.



**Figure 2** | Identifying large water mains and the nodes on them (*PotentialSources*) in the Novato network. The identified nodes are highlighted. These nodes are used in the later steps to identify groups of isolated nodes that are connected to a source while their path to the source does not contain nodes in other groups.



**Figure 3** | Identifying *Islands* in the Novato network. Nodes that are highlighted are *Islands*. The left-side group is a *MajorIsland*, while the two smaller groups are *MinorIslands*.

2014). However, they cannot cause any water security risk for the rest of the network as there is no flow path from these islands to the other parts of the network.

- If size is greater than the maximum allowable size of a sector, it is considered as a *MajorIsland*, and must be partitioned into a set of isolated sectors all having direct access to the large water mains, and consequently, to a water source. Partitioning of major islands is done in Step 5.

It can be seen in Figure 3 that two *MinorIslands* and one *MajorIsland* are identified in the Novato network.

Step 5 (Partitioning *MajorIslands*): In this step, which is the central part of WDN-PARTITION, *MajorIslands* are partitioned into smaller sub-networks of proper size, each of them with direct access to the water sources. This is done by growing graphs from the nodes of the *MajorIslands* that are directly connected to large water mains (*PotentialSources*). Therefore, the resulting sub-graphs, and so the resulting sectors, are guaranteed to be directly connected to the large water mains and so to the main sources (this is another advantage of WDN-PARTITION compared to the other approaches (Ferrari et al. 2014)). If sizes of all the resulting sub-graphs are within the proper boundary, the arrangement is considered as a feasible solution for the *MajorIsland*. Figure 4 illustrates the flowchart of this process. As can be seen in the flowchart, the following sub-steps are taken for each *MajorIsland*.

Step 5.1 (Partition a *MajorIsland* into  $K$  groups): First, it is examined to see whether there are enough nodes to serve the method as *seeds* to grow graphs from. Let us consider  $ptnSrcs4MI = PotentialSources \cap MajorIsland$ ,  $K_{max} = \lfloor |MajorIsland|/dmaMinSize \rfloor$  and  $K_{min} = \lceil |MajorIsland|/dmaMaxSize \rceil$ . If the number of the nodes in  $PtnSrcs4MI$  is less than  $K_{max}$ , the number of seed nodes is not enough, so new nodes need to be found and used as seeds. In such a case, Step 5.1.1 should be taken, otherwise Step 5.1.2.

Step 5.1.1 (Find new potential sources): In this stage, the method finds new nodes in the *MajorIsland* that are connected to large water mains with large pipes (which are definitely smaller than the *MainsSizeThreshold* by some order, as all the nodes that are connected to large pipes with links of size *MainsSizeThreshold* or larger

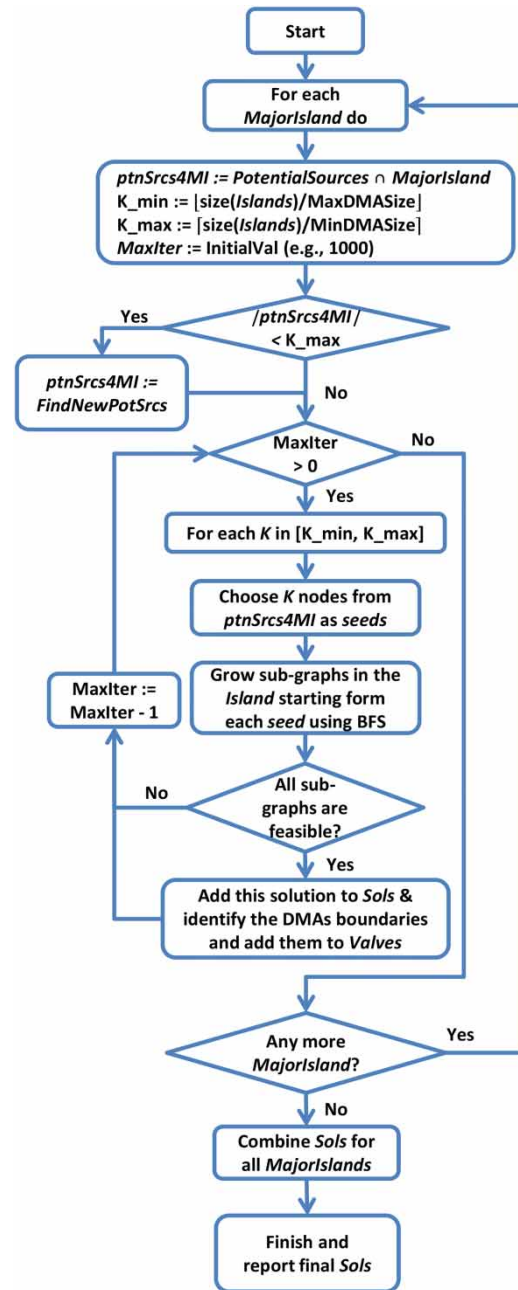


Figure 4 | Partition of *MajorIslands* (Step 5).

were previously identified and added to *PotentialSources*). To this end, for each node in  $PtnSrcs4MI$  a DFS algorithm is used to identify the neighbours in *MajorIsland* which are connected with links of size  $MainsSizeThreshold - ReductSteps$ . The identified nodes are added to the  $PtnSrcs4MI$  set. The method starts by  $ReductSteps = 1$

and continues increasing it until at least  $K_{\max}$  nodes are in *PtnSrcs4MI*.

Step 5.1.2 (Select seeds and grow graphs): Once there are enough seed nodes,  $K$  is set to be the number of partitions; starting from  $K = K_{\min}$  and continuing to  $K_{\max}$  (for example if size of *MajorIsland* is 10,000 and the proper sector size is between 500 and 5,000, any number between 2 and 20 partitions is acceptable; therefore,  $K$  starts from 2 and continues to increase up to 20). Then, in a loop of *MaxIter* (e.g., 100) times,  $K$  nodes are selected randomly from *PtnSrcs4MI* (selection by replacement) to serve as *seeds*. Then using a BFS algorithm, the method grows  $K$  graphs from the selected seeds.

The graph growing algorithm works as the following: starting from each  $K$  seed node in parallel, all the neighbouring nodes that are in *MajorIsland* are added to a group corresponding to the seed node and will be removed from *MajorIsland*. This process continues until all the nodes in *MajorIsland* are visited and removed from it and assigned to different groups. The resulting arrangement of nodes into identified groups could be a feasible solution for partitioning this *MajorIsland*, if sizes of all the resulting groups are within the limits of a proper sector size (larger than *dmaMinSize* and smaller than *dmaMaxSize*). If this condition holds, the resulting graphs are added to the sectors corresponding to the solution; the first link from the selected *seed* is added to *Meters* set; the neighbouring links are identified and added to the *Valves* set, and the corresponding bits are set to 0 in the *position* vector of the solution (the bit string representing the status of links in a solution). The neighbouring links are the links with exactly one end in a group. This process is repeated *MaxIter* times for each  $K$ , and all the feasible solutions are identified.

This is the end of Step 5.1 (Partition a *MajorIsland* into  $K$  groups). After that, the resulting solutions of partitioning different *MajorIslands* should be combined to create overall solutions, which cover all the network.

Step 5.2 (Combine solutions for different *MajorIslands*): As discussed, if there is more than one *MajorIsland* in the network, all the resulting components of partitioning different *MajorIslands* should be reflected in the overall solution which covers the entire network. Therefore, the solutions related to different *MajorIslands* should be combined into overall solutions. This is done

by doing bitwise ‘and’ operations on *positions* of partial solutions and combining the corresponding DMAs. Note that the solutions related to each *MajorIsland* deal with a part of the network, so they are called *partial solutions*. Each partial solution corresponds to a specific set of links and represents the links that should be closed (the positions with 0) or remain open (the positions with 1), by doing bitwise ‘and’ the overall positions of the links that should be closed is identified. As there might be multiple solutions for partitioning each *MajorIsland*, all possible combinations of all solutions for different *MajorIslands* should be considered. Finally, the identified sectors in this step should be combined with the identified sectors from Step 4 to make the final candidate solutions.

After this stage, if there is at least one *MajorIsland* in the network, it is possible to end up having a couple of different solutions, as there might be more than one solution for partitioning each *MajorIsland*. Therefore, the best solution(s) should be identified. This is done in the next stage.

Step 6 (Selecting the best solutions): In this step, the solutions that do not have any advantage over the other solutions, with respect to their objective functions, are removed. As explained in the section ‘Problem statement’, WDN partitioning is a many-objective optimisation problem, so a single solution cannot be chosen as the best one. Thus, a set of Pareto optimal solutions should be identified, which make the Pareto front (Deb 2014). Pareto optimal solutions (in short, Pareto solutions) are the *non-dominated* ones. To compare candidate solutions, the dominance (Deb & Pratap 2002) concept is used. Solution A is said to dominate solution B if and only if A is no worse than B in terms of all the objective functions, and A is exactly better than B in terms of at least one objective function. All solutions are compared against each other, and if one solution dominates another one, the dominated solution will be removed. Therefore, all the non-dominated solutions will remain.

Each candidate solution’s costs is calculated using the formulae (1) through (13). In order to obtain the values for the parameters involved in the objective functions (8) through (13) (e.g., flow, pressure, hydraulic head, tank level, water velocity in pipes and water age), an extended period simulation of the candidate solutions should be run. To do so, the resulting networks (the network in

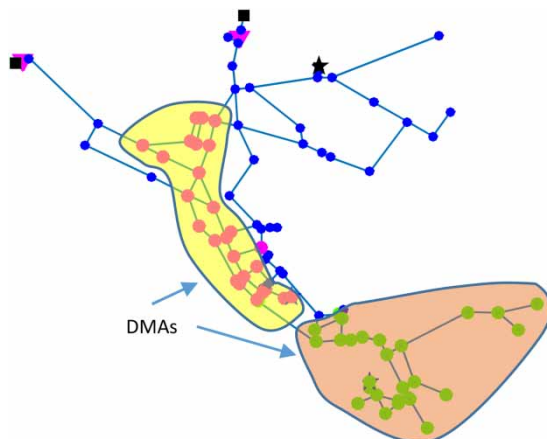


which statuses of links are set based on the *position* vector) are sent to the hydraulic solver (EPANET 2) to solve the network hydraulic equations and to satisfy constraints (18) and (19). Once the required information is returned from the hydraulic solver, the objective functions can be calculated, and the constraint violations can be examined for the candidate solutions.

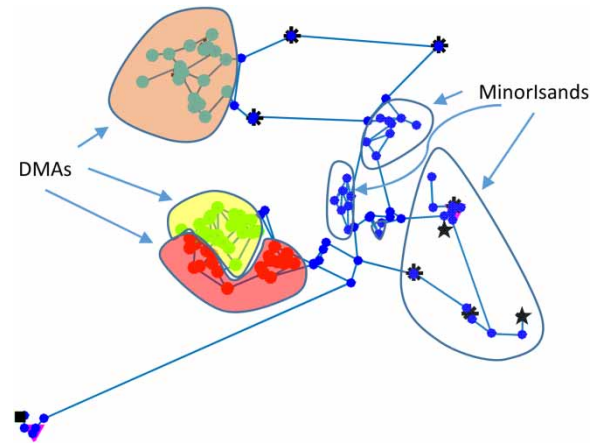
The final results including the identified *DMAs*, *Meters* set, *Valves* set, and the related costs of each non-dominated solution will be reported at the end of the process. The domain experts should choose the best solution(s) based on other requirements of the problem (e.g., financial requirements, infrastructure constraints, etc.) and their experience.

Figure 5 illustrates the identified *iDMAs* in the Novato network. The identified *DMAs* contain 28 and 27 nodes.

The results of applying WDN-PARTITION on another benchmark network are shown in Figure 6. The network comprises 126 nodes, one constant head source, two tanks, 168 pipes, two pumps, eight valves, and is subject to four variable demand patterns. The system was simulated for a total extended period interval of 96 hours. This network is used as a test bed for a number of modelling exercises, including the Battle of the Water Sensor Networks competition (Ostfeld et al. 2008). The network EPANET input file can be downloaded from the Exeter Centre for Water Systems (<http://www.exeter.ac.uk/cws/bwsn>). This network is referred to as BWSN\_Network1 hereafter.



**Figure 5** | Identified *DMAs* in the Novato network. The total number of closed pipes is one of size 8 for this arrangement (the pipe between the two *DMAs*). The identified *DMAs* contain 28 and 27 nodes, total pipe length of 40.7 and 38.8 km. There is also no pressure violations in the partitioned network.



**Figure 6** | Identified *DMAs* in BWSN\_Network1. The top left-hand side *DMA* is identified in Step 4. The two *DMAs* on the bottom left-hand side are identified after partitioning a *Majorisland* containing both (Steps 5 and 6). The *Minorislands* are highlighted as well. The total number of closed pipes is six for this arrangement.

## RESULTS

WDN-PARTITION has been applied on some real-world water distribution networks including the two networks used for illustration of the method in the ‘Method’ section. However, as those two networks are small, they cannot sufficiently reflect the benefits of partitioning. Additionally, no other approach has used these networks for partitioning, so there is no baseline for evaluation on these two networks. Therefore, to evaluate WDN-PARTITION, a large network, which is used as a benchmark in other base line approaches (Murray et al. 2009; Diao et al. 2013; Ferrari et al. 2014; Savic & Ferrari 2014) was chosen, and the results are compared with the results of other approaches, based on the metrics defined in ‘Problem statement’ as objective functions.

The studied WDN comprises 12,523 nodes, two constant head sources, two tanks, 14,822 pipes, four pumps, five valves, and is subject to five variable demand patterns. It serves about 150,000 people and is a good example of a relatively large water distribution system. The system was simulated for a total period of 48 hours. This WDN is used as a test bed for a number of modelling exercises, including the Battle of the Water Sensor Networks competition (Ostfeld et al. 2008). The network is a real water distribution system, but it is modified to preserve its anonymity. However, these modifications do not affect the connectivity and the hydraulic behaviour of the network.



The network's EPANET input file can be downloaded from the Exeter Centre for Water Systems. Hereafter, this network is referred to as BWSN\_Network2. As there are no exact data for the number of customer connections per node in the BWSN\_Network2, the average customer connection per node is used instead. The total number of customer connections is 77,916, so to have the boundary of (500–5,000) customer connections per DMA, 80–800 nodes per DMA on average are considered. Additionally,  $MainsSizeThreshold = 14$  (inch) is considered for this network.  $MaxIter$  is set to 100 (as it proved to be a good trade-off between execution time and the probability of finding different solutions; however, if execution time is not an issue, larger numbers are preferred). As there is no detailed information for each node separately,  $H^*$  is fixed to 28 (m) for all the demand nodes (Ferrari et al. 2014).

As mentioned in the 'Method' section, the method is designed and implemented in such a way that a list of requirements and their priorities (from the list of objective functions stated in 'Problem statement') are input, so for each network or each run, the user chooses what criteria are important and in what order. In running WDN-PARTITION on BWSN\_Network2, the cost and background leakage are not chosen as criteria, as there are no exact data to calculate them. (It should be noted that to calculate background leakage, there should be data about the model parameters, i.e.,  $\alpha_l$  and  $\beta_l$ , for different sectors, which are

not available for the current case. For costs, the exact costs of partitioning including the costs of cutting pipes and installing valves and meters and operating them is dependent on the exact situation of the network at hand, which are not also available for the current case.) Minimum pressure violation, maximum network reliability and water quality are chosen as high priority criteria, so the results are lexicographically ordered based on these requirements.

WDN-PARTITION identifies three *MajorIslands*. After partitioning them and combining the solutions, the method recognises 81 feasible candidate solutions. After removing the dominated ones, the method ended up with 78 non-dominated solutions, i.e., Pareto solutions. The Pareto solutions have between 28 and 48 sectors of sizes between 1,515 and 2,425 customer connections per sector. The total number of closed pipes vary from 66 to 161, and the total number of meters varies from 56 to 78 in different solutions.

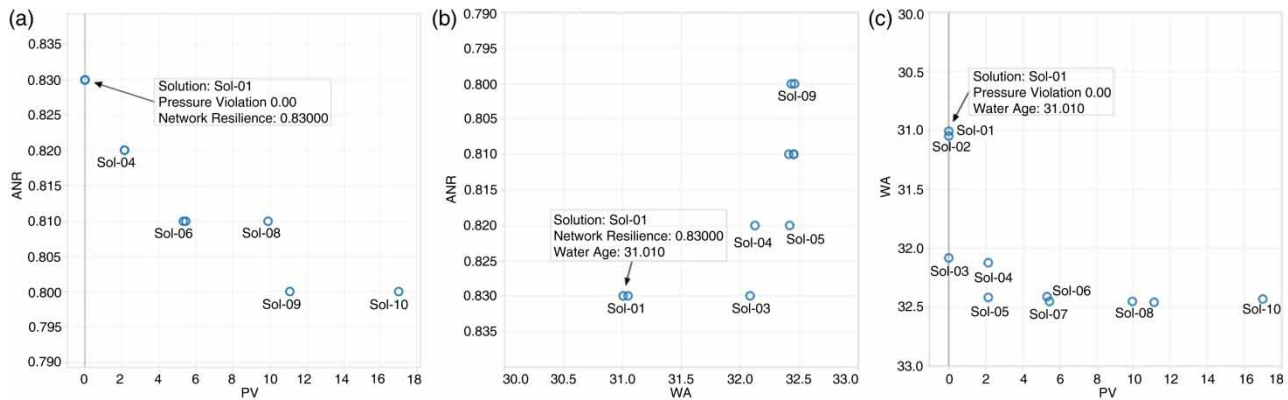
The values of objective functions for 10 different Pareto solutions are reported in Table 1. These 10 solutions are selected from the set of 78 non-dominated solutions after sorting them based on the minimum pressure violations, maximum network reliability and minimum water age (as a surrogate measure for water quality).

Figure 7 illustrates the Pareto front of 10 lexicographically ordered non-dominated solutions and how they behave against each other in terms of pressure violations, network resilience and water age objective functions.

**Table 1** | The first 10 non-dominated solutions for partitioning BWSN\_Network2, sorted lexicographically based on least pressure violations, average network resilience and average water age during the last 24 hours of simulation

	DMA <sub>s</sub>	CS	CW	SSI	Aexp	MExp	ALnkExp	MLnkExp	PV	DP	ED	ANR	WA
Sol-01	28	66	494	0.879	2,423	4,844	$1.55 \times 10^5$	$3.25 \times 10^5$	0	$3.83 \times 10^7$	330.14	0.83	31.01
Sol-02	30	79	624	0.860	2,263	4,206	$1.56 \times 10^5$	$3.23 \times 10^5$	0	$3.82 \times 10^7$	329.49	0.83	31.05
Sol-03	30	78	618	0.860	2,262	4,206	$1.38 \times 10^5$	$3.34 \times 10^5$	0	$3.88 \times 10^7$	362.53	0.83	32.08
Sol-04	30	87	662	0.880	2,261	4,906	$1.66 \times 10^5$	$3.27 \times 10^5$	2.11	$1.10 \times 10^8$	317.74	0.82	32.12
Sol-05	30	87	662	0.880	2,262	4,906	$1.66 \times 10^5$	$3.26 \times 10^5$	2.13	$1.10 \times 10^8$	317.66	0.82	32.42
Sol-06	30	90	629	0.882	2,262	4,969	$1.42 \times 10^5$	$2.93 \times 10^5$	5.32	$8.53 \times 10^7$	356.80	0.81	32.41
Sol-07	30	90	639	0.882	2,263	4,969	$1.42 \times 10^5$	$2.93 \times 10^5$	5.44	$8.44 \times 10^7$	356.56	0.81	32.45
Sol-08	30	84	588	0.875	2,263	4,719	$1.56 \times 10^5$	$3.33 \times 10^5$	9.92	$8.33 \times 10^7$	329.31	0.81	32.45
Sol-09	30	82	562	0.879	2,262	4,850	$1.56 \times 10^5$	$3.86 \times 10^5$	11.1	$8.65 \times 10^7$	330.67	0.80	32.46
Sol-10	31	94	673	0.859	2,190	4,175	$1.34 \times 10^5$	$3.02 \times 10^5$	17	$8.28 \times 10^7$	365.31	0.80	32.43

DMA<sub>s</sub>: number of identified sectors; CS: cut size; CW: cut weight; SSI: sector size imbalance; AExp: average number of customer connections per sector; MExp: maximum number of customer connections in a sector; ALnkExp: average pipe length per sector; MLnkExp: maximum pipe length in a sector; PV: pressure violations; DP: dissipated power; ED: sum of the standard deviations of elevation differences within sectors; ANR: mean network resilience; WA: average water age during the last 24 hours of the simulation.



**Figure 7** | Pareto front of 10 lexicographically ordered non-dominated solutions. (a) Pressure violation vs. average network resilience. (b) Average water age during the last 24 hours vs. average network resilience. (c) Pressure violations vs. average water age during the last 24 hours.

The results demonstrate that WDN-PARTITION generally achieves its design objectives to partition a water network into isolated sectors with a minimal impact on hydraulic requirements. The sizes of all identified sectors are within the minimum and maximum acceptable sector size. Additionally, all the identified sectors are isolated from each other and all have direct access to a water source. The arrangement of one of the Pareto solutions (Sol-01) is illustrated in Figure 8, which identifies 28 sectors of sizes between 588 and 4,844 nodes (2,423 on average). This solution demonstrates 1% decrease in network reliability and less than 1% increase in water age, which are acceptable considering the benefits of sectorisation.

Table 2 illustrates the results of partitioning BWSN\_Network2 using WDN-PARTITION in comparison with other approaches.

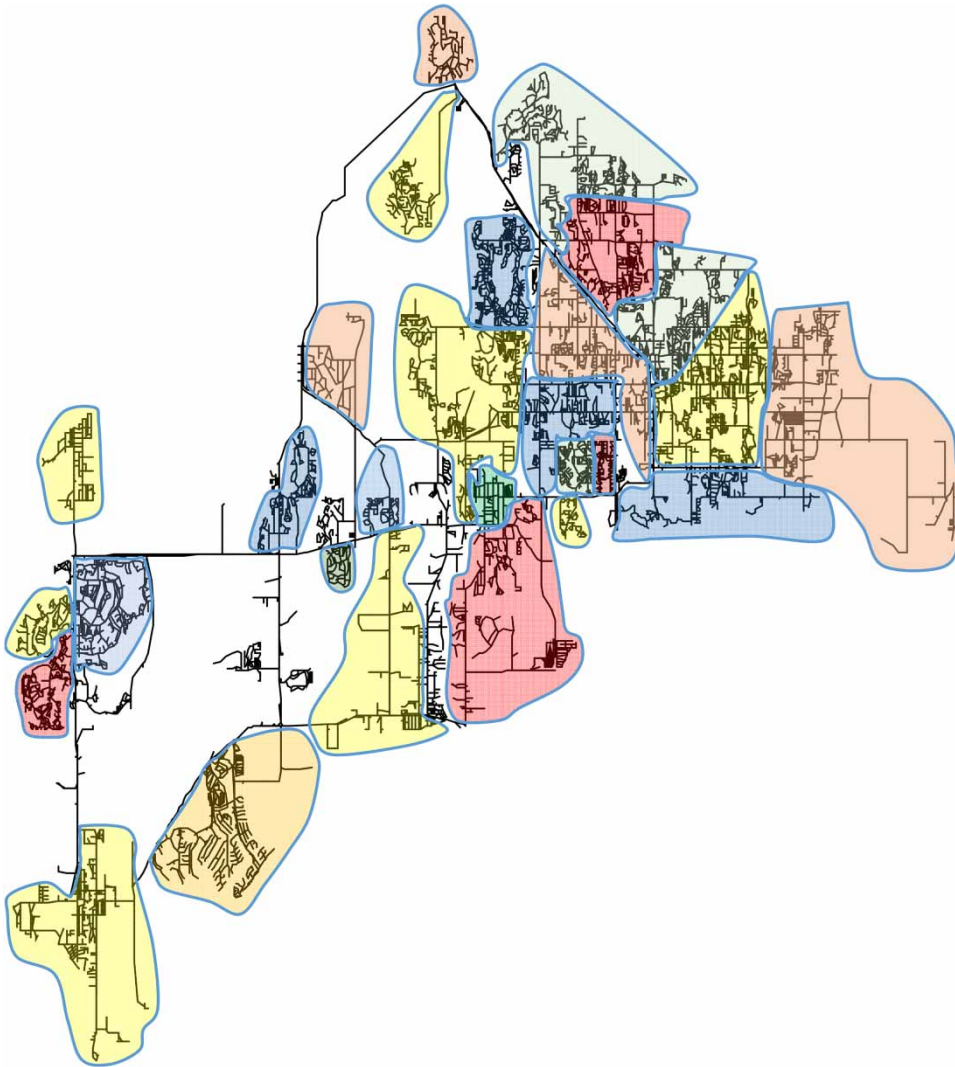
Murray et al. (2009) manually partition BWSN\_Network2 into 43 sectors using engineering principles with trial and error. They modified the network by adding 11 pipes to have at least two feed lines for each sector to ensure sufficient redundancy within the system. The average sector size is 1,996 connections. Three sectors are smaller than 500, and one is slightly larger than 5,000. Additionally, one sector is not directly connected to a source. Total number of closed pipes and meters are 163 and 53, respectively. The average water age during the last 24 hours of simulation is 31.51 hours. The other metrics are not reported in this work.

Diao et al. (2013) applied community detection for the automated identification of sector boundaries in

BWSN\_Network2. This approach identifies 41 sectors, but as the focus of the method is not on identifying *i*DMAs, a number of the identified sectors are not connected directly to a source. Two sectors are larger than the maximum allowable size, and one sector is smaller than the minimum allowable sector size. The average sector size is 2,044 customer connections, and the average water age during the last 24 hours of simulation is 32.01 hours. The other metrics are not reported in this work.

Ferrari et al. (2014) and Savic & Ferrari (2014) applied graph theory to partition BWSN\_Network2 into isolated sectors. As this method applied a multi-objective optimisation method to find the sub-optimal solutions, a set of non-dominated solutions results, comprising solutions having 32 to 43 sectors. The reported sub-optimal solution identifies 36 sectors, of which, at least four sectors are not directly connected to a source (this can be seen in the arrangement illustrated in the paper). One sector is larger than the maximum acceptable size, and three sectors are smaller than the minimum acceptable sector size. The average sector size of the reported solution is 1,863 customer connections per sector, and the total number of closed pipes is between 53 and 132 for different Pareto solutions. The average reported resilience index is between 0.81 and 0.82 for different solutions, while the average reported water age during the last 24 hours of simulation is between 31.04 and 31.62 hours for different solutions. The other metrics are not reported in this work.

WDN-PARTITION identifies between 28 and 48 sectors in different Pareto solutions. All the identified sectors are



**Figure 8** | Identified sectors in BWSN\_Network2. The arrangement illustrated in this figure is one of the Pareto solutions (Sol-01 in Table 1), which identifies 28 sectors, of sizes between 588 and 4,844 customer connections (2,423 on average).

**Table 2** | Comparison of the results of partitioning BWSN\_Network2

	DMA <sub>s</sub>	ND	LT5000	ST500	CS	Meters	AExp	ANR	WA	PA
Original network	–	–	–	–	–	–	–	0.84	30.71	–
Murray et al. (2009)	43	1	1	3	163	53	1,996	0.802	31.51	11
Diao et al. (2013)	41	1	2	1	NA	NA	2,044	NA	32.01	0
Ferrari et al. (2014)	32–43	4	1	3	53–132	NA	1,863	0.82–0.81	31.04–31.62	0
WDN-PARTITION	28–48	0	0	0	66–161	56–78	1,415–2,423	0.83–0.72	31.01–33.14	0

DMA<sub>s</sub>: number of identified sectors; ND: number of sectors not directly connected to source; LT5000: number of sectors larger than 5,000 customer connections; ST500: number of sectors smaller than 500 customer connections; CS: cut size; Meters: number of meters that should be installed; AExp: average size; Meters: number of meters that should be installed; AExp: average number of customer connections per sector; ANR: average network resilience; WA: average water age during the last 24 hours; PA: number of pipes added to the network.

isolated from each other and have direct access to a source. No sector has less than 500 or more than 5,000 customer connections. The average sector size varies from 1,515 to 2,425 customer connections for different solutions. The total number of closed pipes is between 66 and 161 for different solutions. The average resilience index is between 0.83 and 0.72, while the average water age during the last 24 hours of simulation is between 31.01 and 33.14 hours for different Pareto solutions. The total number of closed pipes varies from 66 to 161, and the total number of meters varies from 56 to 78 in different Pareto solutions.

WDN-PARTITION is implemented in MATLAB 2014b, and integrated with EPANET 2. The execution time for BWSN\_Network2 on a laptop with an Intel Core i5 processor and 8 GB of memory is about 15 hours (the execution times for Novato network and BWSN\_Network2 are 69 and 153 seconds, respectively).

## CONCLUSION

WDN-PARTITION addresses water security requirements of water network sectorisation, i.e., isolation of sectors and direct access to at least one source for each sector, with minimum negative impact on the other structural (connectivity of the network and sector size limitations) and hydraulic requirements (customer demand satisfaction, pressure requirements, network reliability, energy efficiency, limited water velocity in pipes, minimum nodal elevation differences within the sectors and water quality). It partitions a water network into isolated sectors without adding any new component to the network, and works well for both small and large networks, as the number of sources is not a limiting factor for the number of sectors. The criteria of optimisation and their priorities can be specified for each case, which makes the method a general purpose one. Simulation-based results show WDN-PARTITION generally achieves its design objectives with a minimal decrease in the performance criteria of the network.

Despite its advantages, WDN-PARTITION has limitations. Similar to every heuristic method, finding the global optimal solution is not guaranteed. However, the results are structurally and hydraulically feasible. Moreover, WDN-PARTITION,

like other methods, can only follow a predefined set of steps. There might be room for improving the results by experts. Finally, in WDN-PARTITION only closure of links (pipes and valves) is considered. It might be useful in some situations to add new components to the network; however, it is beyond the scope of this study.

Some improvements are planned for future work. For example, in the seed selection stage (Step 5.1.2), it is possible to consider some criteria to select better seeds (e.g., the diameter and/or flow of the connecting links, the demand of the seed nodes, and the distance of the seed nodes from each other and/or from the large water mains). Additionally, in Step 5.1.2, after graph growing, it is possible to optimise the boundaries of the identified sub-graphs considering sub-graph size, nodal elevation difference in each sub-graph, the number and the diameter of the pipes that should be cut, distances of the nodes in a group, nodal demands, etc. Finally, in some cases, it might be possible to merge nearby *Minor Islands* with each other to create a sector, or merge them with the neighbouring sectors.

## ACKNOWLEDGEMENTS

This work was supported, in part, by Science Foundation Ireland grant 10/CE/I1855 to Lero – the Irish Software Research Centre ([www.lero.ie](http://www.lero.ie)). This paper is an extension to the paper entitled ‘Water distribution network sectorization using structural graph partitioning and multiobjective optimization’, presented at the 16th Conference on Water Distribution System Analysis, WDSA 2014, Bari, Italy.

## REFERENCES

- Alvisi, S. & Franchini, M. 2014 A heuristic procedure for the automatic creation of district metered areas in water distribution systems. *Urban Water J.* **11** (2), 137–159.
- Andreev, K. & Racke, H. 2006 Balanced graph partitioning. *Theory Comput. Syst.* **39** (6), 929–939.
- Burrows, R., Crowder, G. S. & Zhang, J. 2000 Utilisation of network modelling in the operational management of water distribution systems. *Urban Water* **2** (2), 83–95.
- Chambers, K., Creasey, J. & Forbes, L. 2004 Design and operation of distribution networks. In: *Safe Piped Water: Managing*



- Microbial Water Quality in Piped Distribution Systems* (R. Ainsworth, ed.). IWA Publishing, London, UK.
- Charalambous, B. 2005 Experiences in DMA redesign at the Water Board of Lemesos, Cyprus. In: *International Water Association's Specialised Conference on Leakage, 12–14 September*, Halifax, Nova Scotia, Canada, pp. 1–11.
- Coello Coello, C. A., Lamont, G. B. & Van Veldhuizen, D. A. 2007 *Evolutionary Algorithms for Solving Multi-Objective Problems*, 2nd edn. Springer Science + Business Media LLC, New York, USA.
- da Conceição Cunha, M. & de Oliveira Sousa, J. J. 2010 [Robust design of water distribution networks for a proactive risk management](#). *J. Water Resour. Plann. Manage.* **136** (2), 227–236.
- Davidson, I. 2005 Clustering with constraints: Feasibility issues and the K-Means Algorithm. In: *Proceedings of the 2005 SIAM International Conference on Data Mining, 21–23 April*, Newport Beach, CA, USA.
- Deb, K. 2014 Multi-objective optimization. In: *Search Methodologies SE – 15* (E. K. Burke & G. Kendall, eds). Springer, New York, USA, pp. 403–409.
- Deb, K. & Pratap, A. 2002 [A fast and elitist multiobjective genetic algorithm: NSGA-II](#). *Evol. Comput. IEEE Trans.* **6** (2), 182–197.
- Delgado, D. M. 2008 Infrastructure Leakage Index (ILI) as a Regulatory and Provider Tool. <http://wsp.arizona.edu/sites/wsp.arizona.edu/files/uawater/documents/Fellowship200708/Delgado.pdf>.
- Di Nardo, A. & Di Natale, M. 2011 [A heuristic design support methodology based on graph theory for district metering of water supply networks](#). *Engineering Optimization* **43** (2), 193–211.
- Di Nardo, A., Di Natale, M., Guida, M. & Musmarra, D. 2012 [Water network protection from intentional contamination by sectorization](#). *Water Resour. Manage.* **27** (6), 1837–1850.
- Di Nardo, A., Di Natale, M., Santonastaso, G. F., Tzatchkov, V. G. & Alcocer-Yamanaka, V. H. 2013a [Water network sectorization based on graph theory and energy performance indices](#). *J. Water Resour. Plann. Manage.* **140** (5), 620–629.
- Di Nardo, A., Di Natale, M., Santonastaso, G. F., Tzatchkov, V. & Alcocer-Yamanaka, V. H. 2013b [Water network sectorization based on a genetic algorithm and minimum dissipated power paths](#). *Water Sci. Technol. Water Supply* **13** (4), 951–957.
- Di Nardo, A., Di Natale, M., Musmarra, D., Santonastaso, G. F., Tzatchkov, V. & Alcocer-Yamanaka, V. H. 2015 [Dual-use value of network partitioning for water system management and protection from malicious contamination](#). *J. Hydroinform.* **17** (3), 361–376.
- Diao, K., Zhou, Y. & Rauch, W. 2013 [Automated creation of district metered area boundaries in water distribution systems](#). *J. Water Resour. Plann. Manage.* **139** (2), 184–190.
- Ferrari, G., Savic, D. & Becciu, G. 2014 [Graph-theoretic approach and sound engineering principles for design of district metered areas](#). *J. Water Resour. Plann. Manage.* **140** (12), 04014036. doi:10.1061/(ASCE)WR.1943-5452.0000424.
- Gomes, R., Marques, A. S. & Sousa, J. 2012 [Decision support system to divide a large network into suitable district metered areas](#). *Water Sci. Technol.* **65** (9), 1667–1675.
- Grayman, W., Murray, R. & Savic, D. 2009 Effects of redesign of water systems for security and water quality factors. In: *World Environmental and Water Resources Congress 2009, 17–21 May*, Reston, VA, USA, pp. 504–514.
- Hajebi, S., Song, H., Barrett, S., Clarke, A. & Clarke, S. 2013a [Towards a reference model for water smart grid](#). *Int. J. Adv. Eng. Sci. Technol.* **2** (3), 310–317.
- Hajebi, S., Barrett, S., Clarke, A. & Clarke, S. 2013b [Multi-agent simulation to support water distribution network partitioning](#). In: *27th European Simulation and Modelling Conference – ESM<sup>2</sup>2013, 23–25 October*, Eurosis, Lancaster, UK.
- Hajebi, S., Barrett, S., Clarke, A. & Clarke, S. 2014a [Graph partitioning with structural constraints – Poster](#). In: *The 25th International Workshop on Combinatorial Algorithms (IWCA2014), 15–17 October*, Duluth, Minnesota, USA.
- Hajebi, S., Temate, S., Barrett, S., Clarke, A. & Clarke, S. 2014b [Water distribution network sectorisation using structural graph partitioning and multi-objective optimization](#). *Procedia Engineering* **89**, 1144–1151.
- Herrera Fernández, A. M. 2011 [Improving water network management by efficient division into supply clusters](#). PhD thesis, Universitat Politècnica de València, Spain.
- Jain, H. & Deb, K. 2013 [An improved adaptive approach for elitist nondominated sorting genetic algorithm for many-objective optimization](#). In: *Evolutionary Multi-Criterion Optimization SE – 25* (R. C. Purshouse, P. J. Fleming, C. M. Fonseca, S. Greco & J. Shaw, eds). Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 307–321.
- Liu, H.-L., Gu, F. & Zhang, Q. 2014 [Decomposition of a multiobjective optimization problem into a number of simple multiobjective subproblems](#). *IEEE Trans. Evol. Comput.* **18** (3), 450–455.
- Morrison, J., Tooms, S. & Rogers, D. 2007 *District Metered Areas Guidance Notes*. International Water Association (IWA), Water Loss Task Force, London, UK.
- Murray, R., Grayman, W. M., Savic, D. A., Farmani, R., Boxall, J. & Maksimovic, C. 2009 Effects of DMA redesign on water distribution system performance. In: *Proceedings of the Tenth International Conference on Computing and Control in the Water Industry, 1–3 September*, Sheffield, UK.
- National Research Council 2007 *Improving the Nation's Water Security: Opportunities for Research*. The National Academies Press, Washington, DC, USA.
- Ostfeld, A., Uber, J. G., Salomons, E., Berry, J. W., Hart, W. E., Phillips, C. A. & Watson, J.-P. 2008 [The battle of the water sensor networks \(BWSN\): A design challenge for engineers and algorithms](#). *J. Water Resour. Plann. Manage.* **134** (6), 556–568.
- Parkinson, A. R., Balling, R. & Hedengren, J. D. 2013 *Optimization Methods for Engineering Design*. Brigham Young University, Provo, Utah, USA.
- Pohl, I. S. 1969 *Bi-directional and heuristic search in path problems*. Stanford University, California, USA.
- Porco, J. W. 2010 Municipal water distribution system security study: Recommendations for science and technology investments. *J. AWWA* **102** (4), 30–32.



- Rao, S. S. 2009 *Engineering Optimization: Theory and Practice. Engineering Optimization*, 4th edn. John Wiley & Sons, New York, USA.
- Rossmann, L. A. 2000 *Epanet 2 Users Manual*. Cincinnati US Environmental Protection Agency National Risk Management Research Laboratory, Cincinnati, OH, USA.
- Savic, D. & Ferrari, G. 2014 Design and performance of district metering areas in water distribution systems. In: *16th Conference on Water Distribution System Analysis, WDSA 2014, 14–17 July*, Bari, Italy.
- Tarjan, R. 1972 *Depth-first search and linear graph algorithms*. *SIAM J. Computing* **1** (2), 146–160.
- Todini, E. 2000 *Looped water distribution networks design using a resilience index based heuristic approach*. *Urban Water* **2** (2), 115–122.
- US Army Corps of Engineers 2004 *Water Supply: Water Distribution*. UFC 3-230-10A, US Army Publications, Washington DC, USA.
- US EPA 2005 *Water Distribution System Analysis: Field Studies, Modeling and Management. A Reference Guide for Utilities*. United States Environmental Protection Agency, Washington, DC, USA.
- Walski, T. M., Chase, D. V., Savic, D. A. & Grayman, W. M. 2003 *Advanced Water Distribution Modeling and Management*. Haestad Press, Bentley Institute Press, Exton, PA, USA.
- Water Research Centre 1980 *Report 26 Leakage Control Policy & Practice*. UK Water Authorities Association, London, UK.
- WDSA 2014 team 2014 *Battle of Background Leakage Assessment for Water Networks (BBLAWN)*. In: *16th Water Distribution Systems Analysis Conference, July 2014*, Procedia Engineering, Bari, Italy.
- Yannakakis, M. 1978 *Node-and Edge-Deletion NP-Complete Problems*. In: *Proceedings of the Tenth Annual ACM Symposium on Theory of Computing – STOC '78, New York, USA*, ACM Press, New York, USA, pp. 253–264.
- Yuan, Y., Xu, H. & Wang, B. 2014 *An Improved NSGA-III Procedure for Evolutionary Many-Objective Optimization*. In: *Proceedings of the 2014 Conference on Genetic and Evolutionary Computation – GECCO '14, New York, USA*, ACM Press, New York, USA, pp. 661–668.

First received 15 December 2014; accepted in revised form 29 June 2015. Available online 7 August 2015