GRAPH THEORY AND COMMUNITY DETECTION FOR ELEMENTARY DMA DESIGN

STEFANO MAMBRETTI[1,2], ANITA RAIMONDI[1] & FERNANDO F. STROPPA[3]
[1]Politecnico di Milano, Italy
[2]Wessex Institute of Technology, UK
[3]Università di Pisa, Italy

ABSTRACT
In this work, with the objective of designing elementary district metered areas (eDMAs), graph theory methods and community structured algorithms are applied to the case of the water distribution network of Milano. This network is highly looped and very complex, serving about two million people. Initially, the network is subdivided into 27 eDMAs, which are the number of pumping stations in operation, obtaining three different divided networks (one from community structured algorithms and two from graph theory). Unlike other procedures based on traditional techniques to sectorize networks, the maximum size of each eDMA was not previously restricted, thus generating significantly larger districts. Afterwards, a simple algorithm was applied to aggregate the eDMAs into dynamic DMAs to reduce the energy consumption and overall pressures in the network for the hour of lowest water consumption. Results show improvement in both aspects, thus justifying the study. Further studies are recommended in order to propose other eDMAs designs and other dynamic DMAs configurations.
Keywords: water distribution networks, district metered areas, elementary districts, graph theory, community detection.

1 INTRODUCTION
Water distribution networks (WDN) represent one of the main infrastructure components of society, and, therefore, their operation and management (O&M) have direct impact in energy and water savings [1]. District Metered Areas (DMA) are created from the sectorization of a WDN into subnetworks by installing gate valves or flow meters into the boundary pipes [2]. A permanent pressure control system is facilitated by DMAs, thus reducing the pressure and pressure variation in the WDN, thus also reducing the background leakage [3] and energy consumption [2].

As WDS may consist of thousands to tens of thousands of elements [4], most of which are buried, O&M can be complex [5] and often sub-optimized. In the past, the planning stage for the design of sectors included an evaluation of ground contours, available hydraulic data, existing boundaries, and natural features, combined with a mathematical model [3]. Nowadays, however, with the enhancement of software technology and artificial intelligence, sectorization can be optimized while still considering operational and physical constraints and objectives [5]. The definition of DMAs is usually divided in two phases: clustering, aimed to establish the shape, number of clusters and dimensions of subnetworks, and partitioning, aimed to select the pipes in which the flow meters or gate valves should be installed [5]–[7]. Several methodologies have been proposed to cluster feasible DMAs [5], such as multi-agent approach [8], [9], spectral graph algorithms [10], [11], modularity-based algorithms [12], [13], community structure algorithms [13], [14], graph theory [15]–[17] and multilevel graph partitioning [18]. The partitioning phase is usually done by with evolutionary computation methodologies [5], with single or multi-objective functions designed to optimize hydraulic and/or economic conditions [19]–[22]. The water network partition (WNP) can present some disadvantages, such as a potential negative impact on

water quality, a decrement on energy efficiency and a reduction on the overall hydraulic redundancy of the network [5], [6].

Adaptatively reconfiguring the WDNs is the idea of dynamically changing the DMAs with the objective of reducing these negative impacts [13], [23], [24]. The WNP aimed to enable dynamically reconfigurable DMAs should follow three basis steps:

1. Clustering: Definition of elementary DMAs, which can be dynamically aggregated/segregated in order to achieve a specific goal (e.g. aggregating DMAs in large sectors during high demand periods and segregating DMAs in small sectors for leakage detection during low demand periods [23]).
2. Partitioning: Choice of the pipes in which flow meters, pressure relief valve (PRV) or gate valves should be installed.
3. Iterative refining procedure 1 and 2 to find optimized elementary DMAs.
4. O&M: Dynamically change the gate valves and/or PRVs configurations.

In this paper, the cluster stage of WNP will be applied to the WDN of Milano to select elementary DMAs using graph theory and community structure algorithms. Afterwards, the elementary DMAs will be joint in order to find the *best* sectors in terms of energy optimization.

2 CASE STUDY: THE MILANO WATER SUPPLY NETWORK

2.1 The water supply network

The water supply system of Milano is composed of several reservoirs located at ground level, from which the water is pumped directly highly looped network [1]. The total length of the pipes in the WDN is 2228 km and there are 27 stations in operation, each one equipped with 2–4 pumps [1]. A simplified model of the WDN was built from the available data (such as geometry, mean daily water demand, pumping schedule, etc.) with the well-known software EPANET. Only diameters larger than 300 mm were considered for building the model, and a "fairly accurate reconstruction of the network was reached" [1]. The EPANET file was then transformed in an undirected graph, which consisted of 4994 nodes and 5333 edges. Each node is a graph vertex and represents a reservoir, a tank, or a junction, while each edge represents a pipe, a valve, or a pump. It was assigned a weight for each pipe, based on the estimated conductance defined as [16]:

$$weight = \left(\frac{\lambda}{D.2gA^2}L\right)^{-1}. \tag{1}$$

where λ, D, g, A and L are respectively the resistance coefficient, the diameter, the gravity of Earth, the cross-section area, and the length of the pipe.

2.2 Number of DMAs

The number of DMAs in a WDN has direct impact in the cost of creating them, because when the DMAs sizes decrease, more valves and flow meters are required to be installed and maintained. The *DMA Management Guidance Notes* [3] suggests that each DMA should serve between 500–3,000 properties, and where very poor infrastructure is present, decreasing that value to less than 500 could be ideal. The Guidance Notes also refers that, in DMAs serving more than 5,000 properties, it is difficult to discriminate small bursts during low demand periods. However, there is no imposed limit for a DMA size, as the hydraulic, practical, and economic factors should determine their sizes [3]. Moreover, the important

topic to be considered is the reason why the districts are built, which will drive to different definitions and design.

Considering that in the past only traditional methods and empirical approaches were used to O&M of the WDNs, the usage of small DMAs may no longer be required considering the new artificial intelligence approaches. To test this hypothesis, the number of elementary DMAs (eDMAs) to be found should be equal to the number of reservoirs, with one reservoir in each DMA. It is important to notice that during the O&M stage, because of aggregation of elementary DMAs, the number of actual DMAs will be reduced, and their sizes will be increased even further. Hydraulic analysis with EPANET will determine the efficiency of the system.

3 ELEMENTARY DISTRICTS CREATION

3.1 Graph theory

3.1.1 Breadth-first search algorithm
The breadth-first search (BFS) algorithm searches or traverse tree or graph data structures [25]. For the search in graphs, it starts at an arbitrary node and explores all the neighbour's nodes (other nodes directly connected to the present node) at the present depth prior to exploring the nodes on further depths. As the WDN of Milano is highly looped, the source nodes could reach far more than 2 other nodes, thus reducing the total depth during the search. For each station, a list of ordered junctions is generated as illustrated in Fig. 1, which shows an example of the algorithm in the WDN.
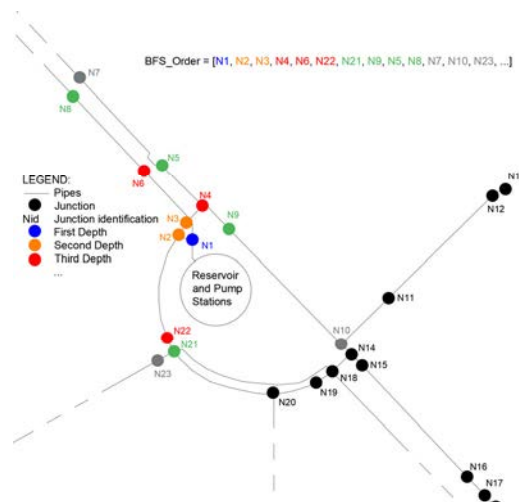

Figure 1: Illustrated BFS applied to a station located in the south-east of the WDN.

3.1.2 Dijkstra's algorithm
The Dijkstra algorithm searches the shortest paths between nodes in a graph [26]. Several variants of the algorithm have been developed along the years. A version that fixes a single node (a junction leaving each station) as the source node and produces a shortest-path tree by finding the shortest paths from the source to all other nodes in the graph will be applied

to the WDN. The conductance of the pipes introduced at point 2.1 will be used as the weight of the edges. As in point 3.1.1, a list of ordered junctions is generated for each station, as illustrated in Fig. 2.
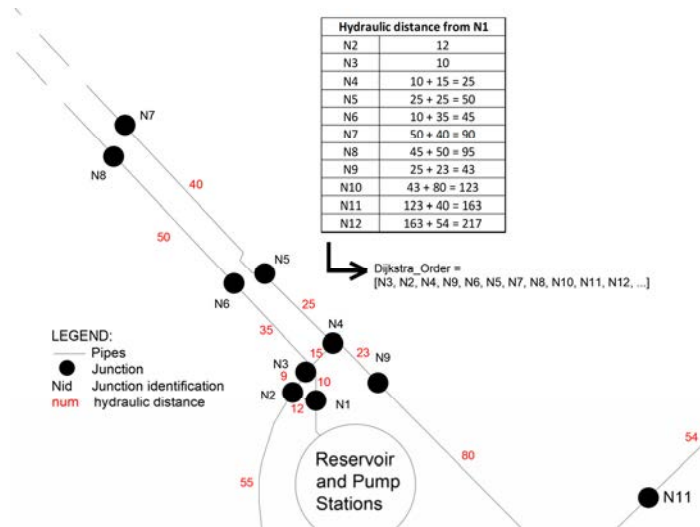

Hydraulic distance from N1	
N2	12
N3	10
N4	10 + 15 = 25
N5	25 + 25 = 50
N6	10 + 35 = 45
N7	50 + 40 = 90
N8	45 + 50 = 95
N9	25 + 23 = 43
N10	43 + 80 = 123
N11	123 + 40 = 163
N12	163 + 54 = 217

Dijkstra_Order = [N3, N2, N4, N9, N6, N5, N7, N8, N10, N11, N12, ...]

LEGEND:
— Pipes
● Junction
Nid Junction identification
num hydraulic distance

Figure 2: Illustrated Dijkstra applied to a station located in the south-east of the WDN.

3.1.3 Creating elementary districts from Dijkstra and BFS

The proposed algorithm searches for eDMAs with approximate equally demand in each district. Because of the highly complex looped structure of the WDN of Milano, this condition is not always feasible as the reservoirs are usually not spaced regularly along the WDNs. The algorithm is divided in the following steps:

1. Creation of the lists of ordered junctions by BFS and Dijkstra, following Steps 3.1.1 and 3.1.2.
2. Creation of an object structure that stores the currently nodes in each eDMAs, initially containing only the junction leaving of each station, also called control junction.
3. Creation of an object structure that stores the currently total demand of each eDMAs.
4. Selection the eDMAs with the lowest demand.
5. From the list defined in point 1, the next node not yet assigned to an eDMAs is selected if this node is adjacent to the selected eDMA, and:
    a. This node is added to the selected eDMA
    b. This node demand is added to the currently demand of the eDMA.
6. Steps 4 and 5 are repeated until all nodes of the network have been assigned to eDMAS.

3.2 Community structure algorithm

3.2.1 Girvan–Newman algorithm

A community in a network is a structure-property where internal nodes of the community are strongly connected among the group while loosely connected between other communities [27]. The Girvan–Newman algorithm is divisive method used to detect communities in complex systems by progressively removing edges from the original network. The method

chooses the edge with the highest betweenness centrality value, which, in the case of a WDN, represents the amount of influence the edge has over the WDN. These edges typically serve as bridges between communities. The algorithm can be summarized as the following:

1. All the values of betweenness centrality of the edges are calculated.
2. The edge with highest betweenness centrality is removed.
3. The values of betweenness centrality are recalculated for edges affected by the previous removal.
4. Steps 2 and 3 are repeated until there are no edges remaining or an established condition is satisfied.

3.2.2 Modified Girvan–Newman algorithm

As in the original Girvan–Newman method does not take into consideration the position of the reservoirs and the total demand of the communities, a modified method is proposed to achieve the goal of attaining 1 reservoir per eDMA. The betweenness centrality of the edges are calculated considering the weight as defined on point 2.1 and the algorithm is summarized as the following:

1. Application of the Steps 1–3 of Girvan–Newman Algorithm
2. Steps 2 and 3 of Girvan–Newman Algorithm is repeated until a new community emerges.
3. When a new community emerges, the following condition is analysed:
    a. If it contains one or more reservoirs: All not removed edges are marked as available.
    b. Else: The edges removed on Steps 1 and 2 are readded and the first edge of Step 2 of Girvan–Newman Algorithm is marked as unavailable.
4. While the number of the communities is lower than the number of reservoirs, repeat Steps 1–3.

The eDMAs are showed in Fig. 3.

4 AGGREGATING ELEMENTARY DISTRICTS

The main idea of aggregating eDMAs is to reduce the energy consumption and to balance the pressure in the WDN. The following equation proposes a simple objective function that take in consideration both factors and can be used as target function to be minimized. *MinPres* is the minimum acceptable pressure in the WDN, which is equal to 25 m for Milano, *CompPres* is the computed pressure in a specific junction and *AvgCompPres* is the average of the computed pressures of the control junctions, which for Milano are the closest junctions to the reservoirs.

$$Objective = EnergyCons + \sum_{n=1}^{NumStations}(MinPres - AvgPres - MinPres)^2. \quad (2)$$

where Objective, EnergyCons, NumStations, are respectively the objective value to be retrieved, the total hourly energy consumption by the pumps and the number of stations presents in the WDN.

A simple algorithm is proposed to aggregate eDMAs into bigger dynamic DMAs during the O&M stage. The aggregation follows the steps.

1. A random percentual of demand and a random number of districts are calculated.
2. All stations are marked as not visited, regardless of the current state of the pumps.
3. A not visited station is selected and marked as visited.
4. One by one, the nearest stations are aggregated to the selected station until the percentual demand is reached. All these stations are marked as visited.
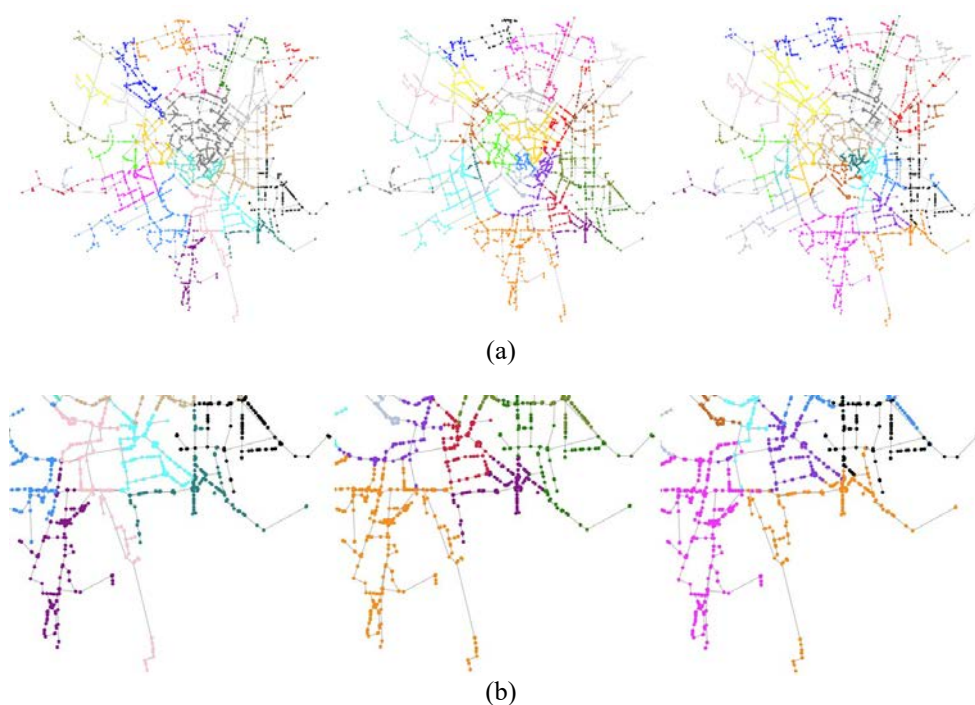
(a)

(b)

Figure 3:   Created eDMAs in the WDN of Milano. Left: Modified Girvan–Newman. Middle: Graph theory using BFS. Right: Graph theory using Dijkstra. (a) Overall WDN; and (b) Southeast of the WDN.

5. Steps 3–4 are repeated until the number of districts is reached.
6. All stations are marked as not visited, and the Step 5 is repeated in order to select all stations in Step 3.
7. Steps 1–6 are repeated for a predetermined number of times and the best DMAs is retrieved based on the proposed objective function.

The proposed algorithm presents two major flaws. Firstly, it can suffer from the *curse of dimensionality* as there are too many possible combinations to be analysed, thus it is much likely that the algorithm will not be able to find the best solution. Secondly, due to the high number of random generations along with the high number of possible combinations, the algorithm is computationally expensive. Further works are recommended to optimize the aggregation of eDMAs.

For achieving the results, firstly an evolutionary computation algorithm developed in previous works [1] was applied to optimize the pumps speeds and status, then the proposed aggregating algorithm was applied for the 3 WDN with different eDMAs to determine what valves should be closed, and finally, the evolutionary computation algorithm was applied once again to optimize the pumps with the established dynamic DMAs.

5 RESULTS
Simulations have been performed for the lowest demand hour of the day (3.00 a.m.). Results show that dividing the WDN of Milano in two or three dynamic DMAs (which is far greater

than 3,000 properties) have the possibility to be already highly beneficial regarding energy saving. Fig. 4 illustrates the three dynamic DMAs generated and Fig. 5 shows the pressure in each control junction (only junctions with more than 10% of pressure difference are shown).
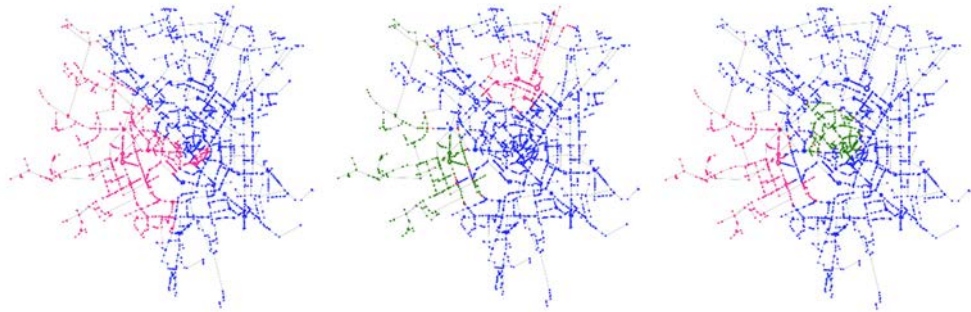

Figure 4:   Created dynamic DMAs in the WDN of Milano. Left: Modified Girvan–Newman. Middle: Graph theory using BFS. Right: Graph theory using Dijkstra.
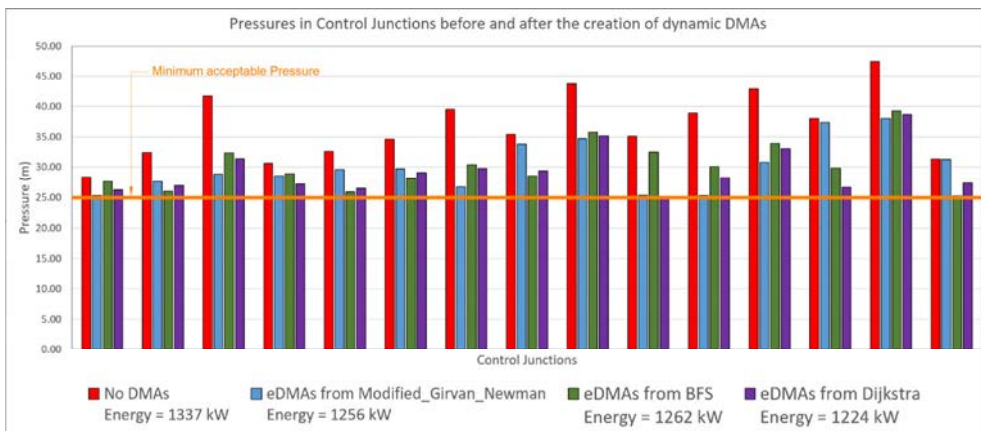

Figure 5: Pressures in control junctions and total energy.

Although the results show the benefits of dynamic DMAs, further works are recommended to determine a better algorithm that aggregates eDMAs, along with a detailed study including the performance during all hours of the day. Lastly, with the introduction of the DMAs, the previous points used for controlling the pressure could become obsolete and could not represent the new state of the WDN. A work to verify that hypothesis and to create an algorithm to find the best new points of pressure control is also recommended.

6 CONCLUSIONS

In the paper, the use of several algorithms has been applied to the creation of eDMAs in the WDN of the city of Milano. Afterwards, thousands of different dynamic DMAs combinations have been tried, led by the simple algorithm previously described, aiming to find optimal

configurations in which energy consumption and overall pressures are reduced. Results show that improvements are possible even when the dynamic DMAs are dramatic bigger than traditional DMAs. Although the proposed algorithm is extremely simple, it was able to find configurations where pressure in almost all control junctions were reduced in all the three eDMAs scenarios.

With the simplifications made to generate the model of the WDN in EPANET, field tests are required to study the actual effect on the WDN. Further research is suggested to improve the actual algorithm the generate dynamic DMAs. Further research is also suggested to find other methods to generate eDMAs for complexed looped WDNs.

REFERENCES

[1]   Wrachien, D., Mambretti, S. & Orsi, E., Optimization of pump operations in a complex water supply network: New genetic algorithm frameworks. *International Journal of Sustainable Development and Planning*, **12**, pp. 79–88, 2017. DOI: 10.2495/SDP-V12-N1-79-88.

[2]   Di Nardo, A., Di Natale, M., Giudicianni, C., Santonastaso, G.F., Tzatchkov, V. & Varela, J.M.R., Economic and energy criteria for district meter areas design of water distribution networks. *Water*, **9**, pp. 1–13, 2017. DOI: 10.3390/w9070463.

[3]   Morrison, J., Tooms, S. & Rogers, D., *DMA Management Guidance Notes*. IWA Publishing: London, 2007.

[4]   Lifshitz, R. & Ostfeld, A., Clustering for analysis of water distribution systems. *Journal of Water Resources Planning and Management*, **144**, 2018. DOI: 10.1061/(ASCE)WR.1943-5452.0000917.

[5]   Khoa Bui, X.S., Marlim, M. & Kang, D., Water network partitioning into district metered areas: A state-of-the-art review. *Water*, **12**(4), 2020. DOI: 10.3390/w12041002.

[6]   Giudicianni, C., Herrera, M., di Nardo, A. & Adeyeye, K., Automatic multiscale approach for water networks partitioning into dynamic district metered areas. *Water Resources Management*, **34**, pp. 835–848, 2020. DOI: 10.1007/s11269-019-02471-w.

[7]   Di Nardo, A., Di Natale, M., Giudicianni, C., Musmarra, D., Santonastaso, G.F. & Simone, A., Water distribution system clustering and partitioning based on social network algorithms. *Procedia Engineering*, **119**, pp. 196–205, 2015. DOI: 10.1016/j.proeng.2015.08.876.

[8]   Herrera, M., Izquierdo, J., Pérez-García, R. & Ayala-Cabrera, D. Water supply clusters by multi-agent based approach. *Water Distribution Systems Analysis*, pp. 861–869, 2010. DOI: 10.1061/41203(425)79.

[9]   Hajebi, S., Barrett, S., Clarke, A. & Clarke, S., Multi-agent simulation to support water distribution network partitioning, *Modelling and Simulation 2013 – European Simulation and Modelling Conference*, ESM, 2013.

[10] Di Nardo, A., Di Natale, M., Giudicianni, C., Greco, R. & Santonastaso, G.F., Weighted spectral clustering for water distribution network partitioning. *Applied Network Science*, **2**, 19, 2017. DOI: 10.1007/s41109-017-0033-4.

[11] Lui, J. & Han, R., Spectral clustering and multicriteria decision for design of district metered areas. *Journal of Water Resources Planning and Management*, **144**(5), 2018. DOI: 10.1061/(asce)wr.1943-5452.0000916.

[12] Brentan, B.M., Campbell, E., Meirelles, G.L., Luvizotto, E. & Izquierdo, J., Social network community detection for DMA creation: Criteria analysis through multilevel optimization. *Mathematical Problems in Engineering*, **2017**, 2017. DOI: 10.1155/2017/9053238.

[13] Qingzhou, Z., Wu, Z., Zhao, M., Qi, J., Huang, Y. & Zhao, H., Automatic partitioning of water distribution networks using multiscale community detection and multiobjective optimization. *Journal of Water Resources Planning and Management*, **143**, 2017. DOI: 10.1061/(ASCE)WR.1943-5452.0000819.

[14] Campbell, E., Ayala-Cabrera, D., Izquierdo, J., Pérez-García, R. & Tavera, M., Water supply network sectorization based on social networks community detection algorithms. *Procedia Engineering*, **89**, pp. 1208–1215, 2014. DOI: 10.1016/j.proeng.2014.11.251.

[15] Di Nardo, A., Di Natale, M., Santonastaso, G.F., Tzatchkov, V.G. & Alcocer-Yamanaka, V.H., Water network sectorization based on graph theory and energy performance indices. *Journal of Water Resources Planning and Management*, **140**(5), pp. 620–629, 2014. DOI: 10.1061/(ASCE)WR.1943-5452.0000364.

[16] Alvisi, S. & Franchini, M., A procedure for the design of district metered areas in water distribution systems. *Procedia Engineering*, **70**, pp. 41–50, 2014. DOI: 10.1016/j.proeng.2014.02.006.

[17] Scarpa, F., Lobba, A. & Becciu, G., Elementary DMA design of looped water distribution networks with multiple sources. *Journal of Water Resources Planning and Management*, **142**, 2016. DOI: 10.1061/(ASCE)WR.1943-5452.0000639.

[18] Alvisi, S., A new procedure for optimal design of district metered areas based on the multilevel balancing and refinement algorithm. *Water Resources Management*, **29**, pp. 4397–4409, 2015. DOI: 10.1007/s11269-015-1066-z.

[19] Di Nardo, A., Di Natale, M., Greco, R. & Santonastaso, G.F., Ant algorithm for smart water network partitioning. *Procedia Engineering*, **70**, pp. 525–534, 2014. DOI: 10.1016/j.proeng.2014.02.058.

[20] Di Nardo, A., Di Natale, M., Santonastaso, G.F., Tzatchkov, V.G. & Alcocer-Yamanaka, V.H., Water Network Sectorization based on genetic algorithm and minimum dissipated power paths. *Water Science and Technology Water Supply*, **13**(4), pp. 951–957, 2013. DOI: 10.2166/ws.2013.059.

[21] Shao, Y., Yao, H., Zhang, T., Chu, S. & Liu X., An improved genetic algorithm for optimal layout of flow meters and valves in water network partitioning. *Water*, **11**(5), 2019. DOI: 10.3390/w11051087.

[22] Campbell, E., Izquierdo, J., Montalvo, I., Ilaya-Ayza, A., Pérez-García, R. & Tavera, M., A flexible methodology to sectorize water supply networks based on social network theory concepts and multi-objective optimization. *Journal of Hydroinformatics*, **18**(1), pp. 62–76, 2015. DOI: 10.2166/hydro.2015.146.

[23] Wright, R., Stoianov, I., Parpas, P., Henderson, K. & King, J., Adaptive water distribution networks with dynamically reconfigurable topology. *Journal of Hydroinformatics*, **16**(6), pp. 1280–1301, 2014. DOI: 10.2166/hydro.2014.086.

[24] Farah, E. & Shahrour, I., Leakage detection using smart water system: Combination of water balance and automated minimum night flow. *Water Resources Management*, **31**, pp. 4821–4833, 2017. DOI: 10.1007/s11269-017-1780-9.

[25] Pohl, I.S., Bi-directional and heuristic search in path problems. PhD thesis, Stanford Linear Accelerator Center, Stanford University, 1969.

[26] Dijkstra, E.W., A note on two problems in connection with graphs. *Numerische Mathematik*, **1**, 1959.

[27] Newman, M.E.J. & Girvan, M., Finding and evaluating community structure in networks. *American Physical Society*, *Physical Review E*, **69**(2), 2004. DOI: 10.1103/PhysRevE.69.026113.