



EPiC Series in Engineering

Volume 3, 2018, Pages 2417–2424

HIC 2018. 13th International  
Conference on Hydroinformatics



# Partition of Water Distribution Networks into District Metered Areas using a Graph Theoretical Approach

Jure Zevnik<sup>1</sup> and Daniel Kozelj<sup>1\*</sup>

<sup>1</sup>University of Ljubljana, Faculty of Civil and Geodetic Engineering, Ljubljana, Slovenia  
daniel.kozelj@fgg.uni-lj.si

## Abstract

We present a method for partitioning Water Distribution Networks (WDNs) into District Metered Areas (DMAs) by using a spectral graph partitioning algorithm. The effectiveness of DMA design was tested for selected edge weights and multiple numbers of established DMAs. The presented method includes a novel graph theoretic approach to determine and evaluate only relevant combinations of DMA connection. It was tested on a real-life case study for which several different solutions were generated and evaluated against their hydraulic performance. The optimal solution, i.e. design of DMAs, was selected regarding the quality of partition and the cost of WDN segmentation, since hydraulic adequacy was met for all cases where partitioning resulted in connected subgraphs.

## 1 Introduction

One of the major challenges in the management of Water Distribution Networks (WDNs) is the high percentage of non-revenue water, wherein the largest contributor are real water losses. In order to decrease the volume of the non-revenue water, we can design District Metered Areas (DMAs), for which hydraulic parameters such as pressure and inflow/outflow are measured [1]. Dividing WDNs into different DMAs enables water utilities for faster identification of water losses and effective control over the whole water supply system. In this paper our goal was to test a spectral graph partitioning method for DMA design on a real-life WDN and to find solutions that are hydraulically equivalent in comparison to the original WDN. A large part of the previously developed methods for automatic DMA design also uses graph theoretical principles and algorithms. Tzatchkov et al. [2] were the first to address this topic. They implemented breadth-first search (BFS) and depth-first search (DFS) algorithm to find appropriate solutions.

---

\* Corresponding author

Later multiple graph theoretic heuristic techniques were developed which also consider hydraulic parameters in the process of DMA design. Ferrari et al. [3] used recursive bisection coupled with BFS algorithm, Alvisi and Franchini [4] implemented BFS and Dijkstra's shortest path algorithm, Galdiero [5] formed DMAs by using Floyd-Warshall algorithm to determine the shortest paths between pairs of vertices. Hajebi et al. [6] modified DFS and BFS algorithms to find potential flow paths between source and demand nodes and to divide WDNs into smaller sections. Kozelj et al. [7] proposed weighted spectral partitioning for DMA design, which also included hydraulic parameters in the partitioning process. Three different weighted spectral clustering methods were compared by Di Nardo et al. [8]. Opposing to the previous research, their results go in favour of unweighted partitioning. Later, Di Nardo et al. [9] compared a multilevel and a spectral clustering method and showed that while the former returns more balanced solutions, the latter results in hydraulically and economically superior DMA layouts.

## 2 Material and methods

The structure of a WDN consists of nodes and pipes which is similar to the definition of a graph, i.e. nodes as vertices and pipes as edges between vertices. Therefore we are able to represent and analyse WDNs using a graph theoretical approach. Other elements of WDNs, e.g. pumps, water sources, tanks, are also considered as vertices or edges of a graph. There are many algorithms in graph theory which can be applied to graph partitioning. The spectral graph partitioning algorithm of Lim et al. [10] was proposed by Kozelj et al. [7] as a promising method to generate DMAs in complex WDNs.

The applied spectral partitioning algorithm is valid for undirected, connected graphs with nonnegative values in the adjacency matrix. A graph is defined as  $G = (V, E)$ , where  $V$  denotes a set of vertices and  $E$  a set of edges between these vertices. In the implemented procedure, two objectives are taken into account for each graph partition  $P$ : the minimization of cost function  $C(P)$  and maximal allowed number  $l$  of vertices in a subgraph. The first objective selects partitions with minimal cut values, i.e. a partitions in which the sum of the weights on edges between different subgraphs (inter-cluster edges) is minimal, while the second objective prefers solutions where the obtained subgraphs have similar node multiplicities i.e., solutions that have equally distributed sizes of all subgraphs. Both objectives can be joined in a single cost function known as ratio cut [11].

Before the graph partitioning procedure is executed, the weights in the adjacency matrix must be determined. The obtained adjacency matrix can be either unweighted or weighted, in order to include adequate criteria into DMAs design. The weights can be selected to represent hydraulic parameters or characteristics of nodes and links in a WDN. By the results of reviewed studies, it is not yet clear whether the weighted or unweighted adjacency matrices generally produce better results for WDN partitioning. Kozelj et al. [7] presented a case of applying combined weights from two different parameters: pipe diameter and downstream nodal head. Di Nardo et al. [8, 9] showed that the unweighted spectral partitioning outperformed tested weights for presented case studies by producing solutions with fewer inter-cluster links. The proposed weight cases from previous studies are the following:  $U$  (unweighted),  $D$  (pipe diameter),  $1/L$  (reciprocal value of pipe length),  $D^5/L$  (pipe hydraulic conductance),  $L/D^5$  (pipe hydraulic resistance),  $Q$  (pipe flow),  $H_{ds}$  (downstream nodal head). In this study we present results for the following cases:  $U$ ,  $D$ ,  $1/L$ ,  $D^5/L$ ,  $Q$ , and for three newly proposed weights:  $\Delta H$ ,  $C_{min}$ , and  $C_{avg}$ . Weight  $\Delta H$ , which represents head loss for each pipe, was chosen due to the fact that it already includes multiple hydraulic weights ( $D$ ,  $1/L$ ,  $D^5/L$ , and  $Q$ ) as can be seen from the Darcy-Weisbach equation.  $C_{min}$  and  $C_{avg}$ ,

$$C_{min} = \min(\text{cost}_{gv}, \text{cost}_{fm}) \quad \text{and} \quad C_{avg} = \frac{\text{cost}_{gv} + \text{cost}_{fm}}{2}, \quad (1)$$

represent the minimum and the average expected implementation costs for each pipe and were selected because the cost plays an important role in the selection of the final solution. The diameter dependent information for gate valve and flow meter costs was taken from Di Nardo et al. [8].

Inputs to the spectral partitioning algorithm are adjacency matrix  $A$ , number of partitions  $k$  and the number of repetitions  $r$ . Graph partitioning is carried out on eigenvectors corresponding to the  $k$  largest eigenvalues of normalized weighted adjacency matrix  $A$  by  $k$ -means ++ algorithm [12], using cosine distance for calculation of cluster centroids. By running the graph partitioning algorithm an optimal solution is found and all nodes are assigned to each of the  $k$  subgraphs. Connectivity within each generated subgraph is tested to prevent disconnected subgraphs. The affiliation of nodes is fixed and an analysis can be performed to assign links to the corresponding subgraphs (DMAs). Links are either inside of a DMA or connecting two different DMAs. The former are called the intra-cluster links and the later the inter-cluster links.

To connect all DMAs inside a WDN, the concept of a spanning tree [13] was used, which ensures overall connection in a WDN with the lowest possible number of opened inter-cluster links. Therefore only  $k - 1$  out of all  $N_{IC}$  inter-cluster links need to remain opened, while others can be closed. Minimising the number of opened inter-cluster pipes is preferred to achieve lower implementation costs, reduce flow meter observational errors and to improve the accuracy of water balance calculation. Theoretically, there are  $\binom{N_{IC}}{k-1}$  possible combinations for connecting DMAs with the minimal number of opened inter-cluster links. Since the numbers of possible combinations rise dramatically with larger values of  $N_{IC}$ , we present a novel graph theoretic approach that identifies only topologically valid solutions by ensuring the original WDN connection. This significantly reduces the number of possible combinations of opened/closed inter-cluster links. Comparison between theoretical and relevant numbers of combinations is presented in Table 5.

After the inter-cluster links are identified we can model the interconnection of the obtained DMAs with a different graph, in which the nodes represent the DMAs and the edges the links between them. This yields an undirected and unweighted multigraph  $G_1$ , whose spanning trees represent all topologically valid combinations of the DMA connection with the least possible number of the installed flow meters. To find all spanning trees in an undirected multigraph, we first construct an auxiliary undirected simple graph  $G_2$  by replacing all the multiple edges in  $G_1$  by a simple one. Therefore graph  $G_2$  is a minor of graph  $G_1$ . Then all spanning trees in  $G_2$  are determined with one of the available algorithms [14]. Finally, given all spanning trees in  $G_2$  we can obtain all spanning trees in  $G_1$  by taking all possible combinations for the choice of the particular edge among all the multi-edge connected two vertices in  $G_1$ .

All topologically valid DMA layouts were evaluated for their hydraulic performance. We looked at three performance indices: generalised resilience index  $I_r$  [15], available power index  $API$ , and pipe hydraulic resilience index  $PHRI$  [16]. All three indices were calculated and compared for all possible options. Selection of the best layouts at different  $k$ -DMAs was based on hydraulic performance, implementation costs and quality of partition.

### 3 Results and discussion

The presented method of graph partitioning was tested on a real-life WDN. The WDN consists of 812 nodes, 1072 links, 3 water sources and 2 tanks. Water source locations are monitored and have fixed measurement stations. The case study is illustrated in Figure 2. Hydraulic parameters were obtained by means of steady-state simulations at average demand loading conditions. The creation of DMAs for the case study was performed for different values of  $k$ , i.e. number of DMAs. The graph partitioning algorithm was implemented in MATLAB 2015a [17] with inclusion of EPANET-

MATLAB toolkit [18]. EPANET 2.0 software [19] was used for reviewing solutions. All simulations were carried out for values of  $k$  between 2 and 12.

Table 1 summarises graph clustering results for the tested weights. By comparing the standard deviations of DMA sizes, i.e. the number of nodes, we can determine that the unweighted ( $U$ ) case generally outperforms other tested weights in regard to the equality of DMA sizes. As can be seen, pipe hydraulic conductance ( $D^5/L$ ) and pipe flow ( $Q$ ) produced poor results for lower values of  $k$  and resulted in unconnected (-) DMAs later on. Di Nardo et al. [8] also reported that the same two weights resulted in unconnected DMAs in certain cases. Therefore we conclude that pipe hydraulic conductance and pipe flow are not suited for spectral graph partition and we exclude them from further examination.

$k$	Avg. size	Standard deviation							
		$U$	$D$	$1/L$	$D^5/L$	$Q$	$\Delta H$	$C_{min}$	$C_{avg}$
2	406	12	103	182	341	359	269	133	104
3	271	109	140	144	279	206	198	189	140
4	203	104	153	103	222	184	73	155	152
5	162	80	81	81	-	179	81	82	81
6	135	71	73	71	185	135	72	72	73
7	116	37	72	79	-	-	43	73	73
8	102	48	40	48	-	-	48	75	39
9	90	44	46	45	-	-	45	46	46
10	81	41	42	40	-	-	41	40	40
11	74	34	35	37	-	-	43	33	33
12	68	26	27	39	-	-	35	35	35
$\Sigma$ (sum over all $k$ )		608	811	868	-	-	949	932	815

**Table 1:** Average DMA sizes and standard deviations of DMA sizes for the tested weights

Table 2 shows a comparison between the number of inter-cluster edges ( $N_{IC}$ ) and implementation costs which were obtained in the DMA connecting procedure. It seems that although the number of inter-cluster edges strongly correlates with the final cost ( $R = 0,95$ ), there are still noticeable differences in the final cost for seemingly very similar partitions at  $k = 5$  and  $k = 6$ .

$k$	Number of inter-cluster links ( $N_{IC}$ )						Final cost [1000 EUR]					
	$U$	$D$	$1/L$	$\Delta H$	$C_{min}$	$C_{avg}$	$U$	$D$	$1/L$	$\Delta H$	$C_{min}$	$C_{avg}$
2	4	4	2	1	6	3	11,1	8,8	8,5	2,6	11,0	8,5
3	4	5	3	2	7	4	13,6	13,6	10,9	7,7	14,0	13,2
4	6	4	6	7	4	4	22,8	16,6	22,5	21,2	20,9	16,6
5	5	5	5	4	5	5	23,8	22,7	19,0	17,9	20,1	22,7
6	5	5	5	5	5	5	25,8	19,5	21,3	22,1	18,4	19,5
7	8	7	6	9	7	8	31,6	25,3	27,4	34,4	22,8	22,7
8	10	11	10	10	8	11	38,0	33,9	40,9	38,0	28,9	33,9
9	12	12	12	12	12	12	42,4	40,5	42,0	41,2	37,1	39,5
10	15	17	16	15	15	15	47,1	44,8	47,5	46,7	42,5	44,8
11	17	20	17	16	21	19	50,3	50,1	52,1	52,6	46,4	48,4
12	18	22	18	19	22	20	55,1	55,1	56,6	56,8	54,5	50,6
$\Sigma$	104	112	100	100	112	106	361	331	349	341	317	320

**Table 2:** Comparison of the number of inter-cluster links and the final costs for the tested weights

At higher values of  $k$ , cost oriented weights outperformed others in regard to the final cost even though they did not necessarily provide partitions with the least inter-cluster edges. Since minimising

implementation costs is one of the main final objectives we decided to further examine and present the results for two selected cases: the unweighted case  $U$ , which resulted in the best DMA size distributions and low  $N_{IC}$ , and the  $C_{min}$  weight case, which overall resulted in the most economical final solutions. Although both cost based weight cases ( $C_{min}$  and  $C_{avg}$ ) performed similarly, the former was chosen because it provided economically favourable or similar solutions for the most relevant values of  $k$ , between 5 and 11.

The presented method of graph partitioning tries to balance node multiplicities of DMAs and does not directly take into account other important aspects of DMA formation, such as total user demand and network length of each DMA. Table 3 shows correlation values between these parameters for selected weights. Given strong overall correlations we can conclude that DMA user demands and network lengths are already indirectly balanced through node multiplicity.

$k$	$N$ vs. $q$		$N$ vs. $L$		$L$ vs. $q$	
	$U$	$C_{min}$	$U$	$C_{min}$	$U$	$C_{min}$
2	1,00	1,00	-1,00	1,00	-1,00	1,00
3	0,92	0,95	1,00	0,99	0,94	0,98
4	0,68	0,96	1,00	1,00	0,67	0,98
5	0,85	0,86	0,99	0,99	0,91	0,92
6	0,85	0,85	0,98	0,98	0,92	0,92
7	0,63	0,93	0,88	0,97	0,78	0,99
8	0,72	0,94	0,94	0,98	0,78	0,98
9	0,85	0,88	0,94	0,95	0,90	0,92
10	0,83	0,84	0,91	0,95	0,90	0,88
11	0,75	0,74	0,91	0,87	0,85	0,82
12	0,75	0,77	0,84	0,89	0,87	0,86
Overall	0,94	0,96	0,99	0,99	0,94	0,97

**Table 3:** Correlation coefficients between: number of nodes ( $N$ ) and DMA user demand ( $q$ ), number of nodes and DMA network length ( $L$ ), DMA network length and DMA user demand for the selected weights

Further examination of partitions for the selected weights can be seen in Table 4. There two measures for inter-cluster link evaluation are shown. The first one, the number of inter-cluster links ( $N_{IC}$ ), presents the unweighted cut values for each of the  $k$ -partitions, and the second one, the cut cost ( $Cut\ cost^*$ ), presents the cut values according to the weighted case  $C_{min}$ . The latter represents the minimum expected implementation cost for a given partition.

$k$	$N_{IC}$		$Cut\ cost^*$ [EUR]		$RC^*$	
	$U$	$C_{min}$	$U$	$C_{min}$	$U$	$C_{min}$
2	4	6	9270	9140	0,020	0,033
3	4	7	8812	10372	0,028	0,052
4	6	4	15160	13280	0,054	0,049
5	5	5	13324	9852	0,061	0,057
6	5	5	14512	7568	0,077	0,074
7	8	7	20822	11368	0,137	0,139
8	10	8	23994	14156	0,215	0,216
9	12	12	26780	21560	0,279	0,283
10	15	15	27534	26160	0,371	0,376
11	17	21	29010	28476	0,448	0,556
12	18	22	31968	33288	0,538	0,688

**Table 4:** Number of inter-cluster links ( $N_{IC}$ ), minimum cost weighted cut cost ( $Cut\ cost^*$ ) and adjusted ratio cut ( $RC$ ) for the selected weights

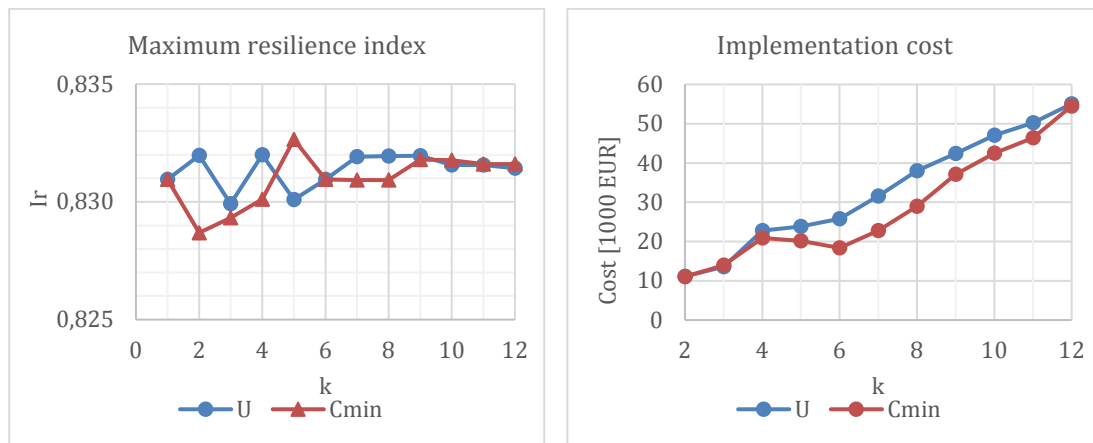
Comparison between both measures results in similar conclusions than before, which were, that weights that represent implementation costs give favourable results over others, while not necessarily yielding results with the least  $N_{IC}$ . The third presented measure is adjusted ratio cut ( $RC^*$ ), which was for the sake of comparison calculated using unweighted cut values for both cases. It can be seen that the unweighted case generally outputs better results, which is due to the fact that it returns more equally distributed DMA sizes. For values of  $k$  between 5 and 10 both cases have similar values of adjusted ratio cut.

The efficiency of the proposed DMA connection algorithm can be seen from Table 5, where we compare the number of all theoretically possible combinations and the number of all relevant combinations of DMA connection. It can be seen that the numbers start to vary strongly for values of  $k$  above 6. By using the proposed approach, we can therefore hydraulically and economically evaluate all the relevant combinations of DMA connection and select the optimal solution.

$k$	No. of theoretically possible comb.		No. of relevant combinations	
	$U$	$C_{min}$	$U$	$C_{min}$
2	4	6	4	6
3	6	21	3	6
4	20	4	4	2
5	5	5	2	2
6	1	1	1	1
7	28	7	3	2
8	120	8	4	2
9	495	495	8	8
10	5005	5005	20	24
11	19448	352716	48	240
12	31824	705432	96	240

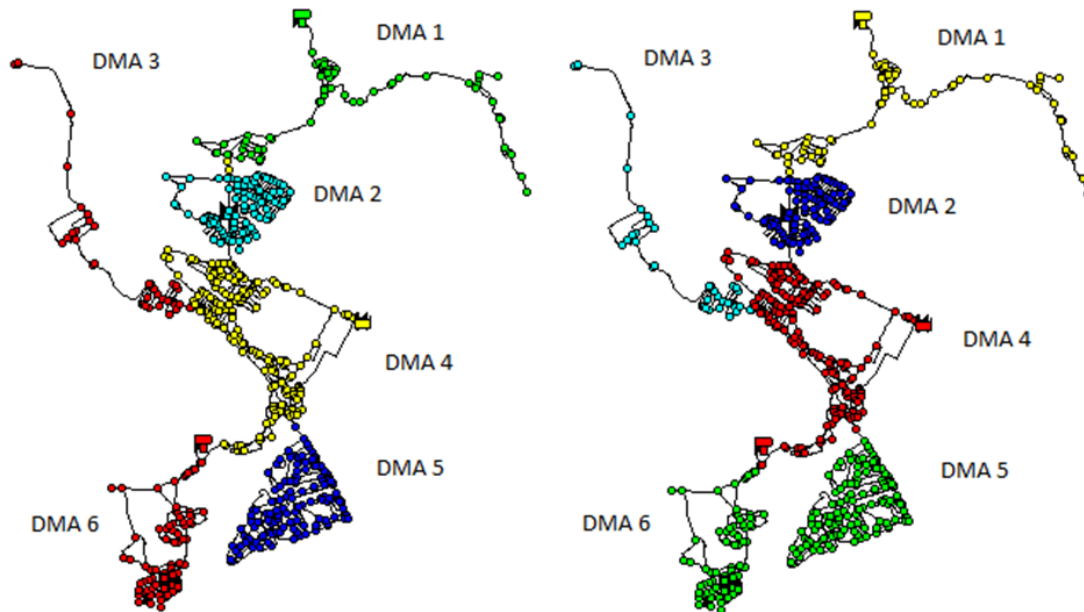
**Table 5:** Number of combinations for the selected weights

Interestingly, we were able to find hydraulically feasible solutions, with negligible hydraulic index ( $I_r$ ,  $API$ ,  $PHRI$ ) deviations according to the original WDN, for all the tested weight cases presented in Table 2. Figure 1 (left) shows final resilience indices for the two selected cases. Since we were able to find hydraulically almost identical solutions compared to the original WDN, the main objectives for the final selection became implementation costs (graph on the right-hand side in Figure 1) and partitioning quality.



**Figure 1:** Resilience index  $I_r$  (left) and implementation cost (right) for the selected solutions

The final selected partition divides the original WDN into six ( $k = 6$ ) DMAs. It was obtained with the  $C_{min}$  weight case and is presented on the left side of Figure 2, where it can be seen that this solution yields lower implementation costs due to the fact that no inter-cluster links have to be closed. Very similar partition for the same value of  $k$  was obtained by the unweighted case (Figure 2, right), also with the same  $N_{IC}$ , but with 40,3 % higher implementation costs. Table 6 summarises the general information for each DMA for the selected final solution.



**Figure 2:** The final selected solution of  $C_{min}$  (left) and similar partition for unweighted case  $U$  (right)

ID	No. of nodes [-]	No. of links [-]	Total user demand [l/s]	Total pipe length [km]	Mean node elevation [m]
DMA 1	76	94	14,1	11,6	312,2
DMA 2	138	188	6,25	14,5	301,5
DMA 3	48	58	5,45	8,8	321,3
DMA 4	274	348	33,3	28,4	315,8
DMA 5	152	203	23,7	19,9	309,9
DMA 6	124	167	15,4	14,1	312,8

**Table 6:** General information for each DMA for final selected solution  $C_{min}$

## 4 Conclusions

The aim of this study was to develop an efficient graph partitioning algorithm for DMA design and to test it on a real-life WDN. The ratio cut spectral partitioning method proved to be fast and adequate for WDN partitioning. Selection of weights, hydraulic performance criteria, and boundary link selection are among the most important parts of this method that were investigated. The quality of a given partition was determined by the standard deviation of DMA sizes, which correlates with DMA characteristics of interest, e.g. total DMA user demand. We selected two weight cases for further comparison, which showed promising results for the novel weight case  $C_{min}$ , since obtaining

economical solutions is one of the main objectives. Furthermore, the newly proposed DMA connection algorithm showed a substantial decrease in the number of combinations to evaluate in comparison to the traditional approach. This means that all combinations of connection can be evaluated and the optimal one can be selected. The final solution was selected, analysed and commented. Future research includes implementation of other spectral graph partitioning methods, along with comparison and testing on different complex WDNs. Additionally, we will improve the proposed DMA connection algorithm by also considering the water mains.

## References

- [1] J. Morrison, D. Rogers, and S. Tooms, District Metered Areas Guidance Notes, Water Loss Task Force, IWA Publication, 2007.
- [2] V. G. Tzatchkov, V. H. Alcocer-Yamanaka, and V. J. Bourguett-Ortíz, Graph theory based algorithms for water distribution network sectorization projects, in Proc. 8th Annual Water Distribution Systems Analysis Symposium (WDNA 2006), Cincinnati, Ohio, USA, 2006.
- [3] G. Ferrari, D. A. Savić, and G. Becciu, Graph-Theoretic Approach and Sound Engineering Principles for Design of District Metered Areas, *J. Water Resour. Plann. Manage.* 140.12 (2014) 04014036.
- [4] S. Alvisi and M. Franchini, Heuristic procedure for the automatic creation of district metered areas in water distribution systems, *Urban Water Journal* 11.2 (2014) 137-159.
- [5] E. Galdiero, Multi-Objective Design of District Metered Areas in Water Distribution Networks, Department of Civil, Architectural and Environmental Engineering, University of Naples Federico II, Naples, Italy, 2015.
- [6] S. Hajebi, E. Roshani, N. Cardozo, S. Barrett, A. Clarke, and S. Clarke, Water distribution network sectorisation using graph theory and many-objective optimisation, *J. Hydroinform.* 18.1 (2016) 77-95.
- [7] D. Kozelj, M. Gorjup, and M. Kramar Fijavž, An application of spectral graph partition for designing district metered areas in water supply networks, *Acta hydrotechnica*, 30.53 (2017) 81-96.
- [8] A. Di Nardo, M. Di Natale, C. Giudicianni, R. Greco and G. F. Santonastaso, Weighted spectral clustering for water distribution network partitioning, *Applied Network Science* 2 (2017) 1-16.
- [9] A. Di Nardo, M. Di Natale, C. Giudicianni, R. Greco and G. F. Santonastaso, Water Distribution Network Clustering: Graph Partitioning or Spectral Algorithms?, *Complex Networks & Their Applications VI* (2018) 1197-1209.
- [10] C. Lim, S. Bohacek, J. P. Hespanha and K. Obraczka, Hierarchical Max-Flow Routing, in Proc. of the IEEE GLOBECOM, St. Louis, Missouri, USA, 2005.
- [11] Y. Wei and C. Cheng, Ratio Cut Partitioning for Hierarchical Designs, *IEEE Transactions on Computer-Aided Design* 10 (1991) 911-921.
- [12] D. Arthur and S. Vassilvitskii, k-means++: The Advantages of Careful Seeding. SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (2000) 1027-1035.
- [13] R. Diestel, *Graph Theory*, Berlin Heidelberg: Springer-Verlag, 2006.
- [14] M. Hotz, generateSpanningTrees(A), MATLAB Central File Exchange. Retrieved October 1, 2017, (<https://www.mathworks.com/matlabcentral/fileexchange/53787-generatespanningtrees-a->).
- [15] E. Creaco, M. Franchini and E. Todini, Generalized resilience and failure indices for use with pressure-driven modeling and leakage, *J. Water Resour. Plann. Manage.* 142.8 (2016) 04016019.
- [16] H. Liu, D. A. Savić, Z. Kapelan, E. Creaco, and Y. Yuan, Reliability Surrogate Measures for Water Distribution System Design: Comparative Analysis, *J. Water Resour. Plann. Manage.* 143.2 (2016).
- [17] MATLAB 2015a, The MathWorks, Inc., Natick, Massachusetts, United States.
- [18] D. G. Eliades, M. Kyriakou, S. Stelios Vrachimis in M. M. Polycarpou, EPANET-MATLAB Toolkit: An Open-Source Software for Interfacing EPANET with MATLAB, in 14th Int. Conf. on Computing and Control for the Water Industry (CCWI 2016), Amsterdam, Netherlands, 2016.
- [19] L. A. Rossman, Epanet2 Users Manual, US Environmental Protection Agency, National Risk Management Research Laboratory, Cincinnati, Ohio, USA, 2000.