Di Nardo et al. Applied Network Science (2017) 2:19
DOI 10.1007/s41109-017-0033-4

Applied Network Science

RESEARCH Open Access

Weighted spectral clustering for water distribution network partitioning

Armando Di Nardo[1,2,3*], Michele Di Natale[1,2], Carlo Giudicianni[1], Roberto Greco[1,2] and Giovanni Francesco Santonastaso[1,2]

* Correspondence: armando.dinardo@unina2.it
[1]Dipartimento di Ingegneria Civile Design Edilizia e Ambiente, Università degli Studi della Campania 'L. Vanvitelli', via Roma 29, Aversa 81031, Italy
[2]Action Group CTRL + SWAN of the European Innovation Partnership on Water EU, via Roma 29, Aversa 81031, Italy
Full list of author information is available at the end of the article

Abstract

In order to improve the management and to better locate water losses, Water Distribution Networks can be physically divided into District Meter Areas (DMAs), inserting hydraulic devices on proper pipes and thus simplifying the control of water budget and pressure regime. Traditionally, the water network division is based on empirical suggestions and on 'trial and error' approaches, checking results step by step through hydraulic simulation, and so making it very difficult to apply such approaches to large networks. Recently, some heuristic procedures, based on graph and network theory, have shown that it is possible to automatically identify optimal solutions in terms of number, shape and dimension of DMAs. In this paper, weighted spectral clustering methods have been used to define the optimal layout of districts in a real water distribution system, taking into account both geometric and hydraulic features, through weighted adjacency matrices. The obtained results confirm the feasibility of the use of spectral clustering to address the arduous problem of water supply network partitioning with an elegant mathematical approach compared to other heuristic procedures proposed in the literature. A comparison between different spectral clustering solutions has been carried out through topological and energy performance indices, in order to identify the optimal water network partitioning procedure.

Keywords: Laplacian spectrum, Spectral clustering, k-means, Water network partitioning

Introduction

Civil engineering networks regard different infrastructures (e.g. transport, energy, phone, internet, water, gas, logistic). Water Distribution Networks (WDNs) are among the most important civil networks, because they deliver drinking and industrial water to metropolitan areas. From a topological perspective, a WDN with multiple interconnected elements may be represented essentially as a link-node planar weighted spatially organized graph for which pipes (and valves) correspond to links m, and nodes/junctions (such as pipe intersections, water sources and nodal water demands) correspond to graph nodes n. Planar graph have vertices whenever two edges cross, whereas non-planar graphs can have edges crossing but not forming vertices (Boccaletti et al., 2006). WDN belong to the class of networks with nodes occupying precise positions in two or three-dimensional Euclidean space, edges being real physical connections, and strongly constrained by their geographical embedding (Boccaletti et al, 2006), like other spatially organized urban infrastructure systems (Carvalho et al, 2009; Newman, 2003).

Springer Open

© The Author(s). 2017 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In an abstract modelling context, a mathematical graph can be used to express the relationships between groups of linked nodes. An important aspect of spatial networks is that node degrees are constrained, as the number of possible connections to a single node is physically limited. Furthermore, in WDN it is unlikely to have direct connections between very distant nodes, so that significant limitations to the small-world behaviour of such networks arise (Boccaletti et al, 2006). In particular, little variability is observed in the connectivity patterns of the nodes in WDN, no hubs (nodes with much more connections than the others) are present, and most of the nodes have very low degree (usually two or three, and mostly less than five), so in general they present a fairly homogeneous degree distribution (Di Nardo et al, 2015a). Furthermore, such networks are also equally sensitive to random or malicious failures (Barthelemy and Flammini, 2008).

WDN can be considered as complex networks for many reasons (Mays, 2000): they are often very large (up to tens of thousands nodes and links); they are buried underground, and thus are not easily accessible for monitoring and maintenance; they are strongly looped; their modelling includes non-linear equations requiring sophisticated numerical resolution methods; they often present severe water losses. Compared to other civil networks (e.g. gas, electricity, transport, telephone, internet), some of these WDN characteristics are peculiar, and make their management arduous, with many operational problems (such as water and energy losses). For all these reasons, in the last decades, the scientific community has proposed different approaches to improve WDNs management, without compromising their main function, i.e. providing water to end users ensuring a minimum level of service.

In this context, the implementation of the paradigm of "divide and conquer" in a WDN allows simplifying the management, defining sub-systems named District Meter Areas (DMAs), by inserting gate valves and flow meters along network pipes, properly selected, in order to define a Water Network Partitioning (WNP). In this way, it is possible to improve water losses identification (Water Industry Research Ltd, 1999), control district pressure (Alonso et al, 2000), and protect users from accidental and intentional contamination (Di Nardo et al, 2015b), because these activities are simpler to achieve if the network is divided in sub-systems. By dividing the water network in DMAs, implementing innovative Information and Communications Technology (ICT) remote-controlled devices and big data analysis, it is possible to change the traditional approach to the management of WDN, transforming the water systems into modern Smart Water Network (SWAN) (Di Nardo et al, 2016a), considered as part of Smart Cities.

It is important to underline that, to define a good WNP, it is necessary to satisfy two crucial major requirements for the optimal functioning of a WDN: 1) network connectivity, i.e. each demand node of the water network must be connected to at least one water source, and 2) nodal minimum pressure, i.e. each node must have a pressure equal or higher than the minimum level of service that allows satisfying the water demand of the users. Therefore, the design of a WNP, as any problem of network subdivision, is a complex challenge for operators, because the permanent partitioning changes the original topological layout of water systems. Indeed, network partitioning, achieved by pipe closures, reduces the overall pipe section availability, with the consequent decrease of network water pressure, especially during peak hours, worsening the level of service offered to users.

In the last years, different procedures have been proposed in the literature for finding an optimal WNP layout (reviews are given in Di Nardo et al, 2013a; Perelman et al, 2015), essentially based on heuristic algorithms and optimization procedures. Generally, they consist of two different phases:

a) **clustering**, aimed at defining the shape and the dimensions of the network subsets, based on different theories, among which: *graph theory algorithms*, obtaining the number of independent sectors through connectivity analysis, (Tzatchkov et al, 2006); identifying the pipes along which to insert hydraulic devices by searching minimum dissipated power paths using graph theory principles (Di Nardo et al, 2013a; Alvisi and Franchini, 2014); with an optimization model solved by a simulated annealing algorithm with an objective economic function (Gomes et al, 2012); based on shortest path search with dissipated power weight on pipes and refining through an objective function of the Genetic Algorithm based on network mean pressure (Di Nardo et al, 2013b); *spectral approach* with spectral clustering algorithm applied to adjacency matrix with different supply constraints (Herrera et al, 2010) or recursive bipartition of the graph through weighted graph Laplacian matrix (Di Nardo et al, 2017); *multi-agent approach* taking into account multiple interacting agents of WDN (Izquierdo et al, 2011); *community structure*, based on social network theory and graph partitioning algorithms (Di Nardo et al 2015a) or with an automatic identification of boundaries on the basis of the property that density of edges within communities should be higher than between them (Diao et al, 2013);
b) **dividing**, aimed at physically partitioning the network, by selecting pipes for the insertion of flow meters or gate valves: based on *recursive bisection* procedure and an algorithm for graph traversal to verify the reachability of each district from the water source and node connectivity (Ferrari et al, 2014); on *genetic algorithms* implementing an automatic heuristic optimization technique for DMAs definition with minimum hydraulic deterioration (Di Nardo et al., 2015c, 2016b), with the objective of identifying the optimal layout that minimises the economic investment and the hydraulic performance deterioration.

Generally, such a two steps approach allows simplifying the water network partitioning, as, once the optimal node clustering is identified, then it becomes the starting point of the subsequent dividing phase. It is worth to highlight that the proposed procedures can be more effective if the clustering phase takes into account some hydraulic features of the network (i.e., energy, geometry), as reported in other studies (Di Nardo et al., 2013a, 2016a) depending on the adopted clustering algorithm. To such aim, in this work, the most important energy parameters are taken into account for the clustering stage.

This paper, extending a previous basic work (Di Nardo et al, 2017), aims at investigating the feasibility of adopting weighted spectral clustering to identify the optimal sub-graphs layout, comparing different weights of pipes and different spectral methods (von Luxburg, 2007), and then, subsequently, to define the optimal water network partitioning not only from a topological but also from a hydraulic point of view.

Methodology

As described above for other approaches, the proposed procedure consists of two distinct phases (Di Nardo et al, 2016a), separately described in the following subsections.

Phase 1: water network clustering

As known, considering a simple graph $G = (V,E)$, where V is the set of n vertices v_i (or nodes) and E is the set of m edges e_l (or links), a k-way graph clustering problem consists in partitioning V vertices of G into k subsets, $P_1, P_2,..., P_k$ such that:$\cup_1^k P_k = V$ (the union of all clusters P_k must contain all the vertices V_i), $P_k \cap P_t = \emptyset$ (each vertex can belong to only one cluster P_i), $\emptyset \subset P_k \subset V$ (at least one vertex must belong to a cluster and no cluster can contain all vertices) and $1 < k < n$ (the number k of clusters must be different from one and from the number n of vertices). Clustering is usually defined in terms of weighted, undirected graphs, where weights correspond to either similarity scores, or distances, or, more generally, they express the strength of the link between elements in order to define sub-graphs which take into account proximity and/or similarity between elements.

Graph clustering can be achieved with many procedures aimed to define the optimal layout of each cluster, finding community structures minimizing or maximizing an objective function that emphasizes one of the clustering aims. In literature (wide reviews are provided in Boccaletti et al, 2006; Fortunato, 2010), several procedures were proposed: k-means; Markov cluster algorithm; spectral methods (as optimization algorithm of the cut problem, such as min-cut, ratio-cut, normalized-cut); hierarchical clustering; modularity; multi-level-recursive algorithm, Girvan and Newman algorithm and some other methods.

In recent years, spectral clustering, based on eigenvectors and eigenvalues of the graph Laplacian matrices (defined hereinafter), has become one of the most popular clustering algorithms (Chung, 1997; Saerens et al., 2004; von Luxburg, 2007), because it can be solved by standard linear algebra software developed by the authors in MATLAB™ (SimuLink Reference Books 2006) and so it is easy to implement. So, in this paper, the clustering phase to define sub-graph for the subsequent dividing phase has been achieved with different weighted spectral clustering techniques, investigating the effectiveness of this approach and the optimal choice of weights. As known, the main tools for spectral clustering are graph Laplacian matrices and, in the following, G is assumed as an undirected, weighted graph with weight matrix W_ω, where $w_{ij} = w_{ji} \geq 0$. In particular, as explained above, different weights have been adopted for the pipes to investigate which of them provides the best results. The choice of pipe weights is crucial, as different weights lead to significantly different layouts of the districts. As aim of the partitioning is to identify a balanced layout of the districts (i.e. districts with similar dimensions) least affecting the hydraulic performance of the network (i.e. minimising the unavoidable increase of head losses), pipe characteristics related to hydraulic resistance have been here tested as weights.

Given a graph $G = (V, E)$, the adjacency nxn matrix A (in the following indicated as W_A and corresponding to the no-weight matrix) expresses the connectivity of the graph, where elements $a_{ij} = a_{ji} = 1$ indicate that there is a link between nodes i and j and $a_{ij} = a_{ji} = 0$ otherwise.

Three spectral clustering methods have been tested. The first one, which solves relaxed versions of the RatioCut problem (von Luxburg, 2007), is based on the eigenvalues of the unnormalized graph Laplacian L, defined as:

$$L = D_k - W_\omega \tag{1}$$

where D_k = diag(K_i) and K_i is the degree of a node i.

The other two methods, both solving relaxed versions of the NCut problem (Shi and Malik, 2000), belong to normalized spectral clustering, as they use the eigenvalues of normalized graph Laplacian. In particular, the normalized spectral clustering according to Shi and Malik (2000) is based on the normalized Laplacian L_{rw}, closely related to a random walk (von Luxburg, 2007) and defined as:

$$L_{rw} = D_k^{-1} L \tag{2}$$

The third tested method is the normalized spectral clustering proposed by Ng et al. (2001), based on eigenvectors of the normalized Laplacian L_{sym}, a symmetric matrix defined as:

$$L_{sym} = D_k^{-1/2} L D_k^{-1/2} \tag{3}$$

The above mentioned three spectral clustering algorithms have been applied to identify the optimal clusters in a WDN. Namely, the tested W_ω matrices have been: W_A (i.e.no weights are given to the pipes, so to take into account only the connectivity); W_D (weight equal to pipe diameter D, related to pipe hydraulic resistance in formulas with exponent close to -5); $W_{1/L}$ (weight equal to the inverse of pipe length, linearly related to pipe hydraulic resistance); W_C (weight equal to pipe conductance, here assumed as proportional to D^5/L, under the simplifying hypothesis that all the pipes in the network share the same roughness coefficient); W_F (weight equal to pipe flow, indirectly related to both pipe hydraulic conductance and water demand distribution at nodes).

Specifically, the clustering phase for the proposed water network partitioning consists of the following steps:

1. abstraction of the water supply network as a graph $G = (V, E)$;
2. definition of adjacency matrix and pipe weight matrices W_ω as defined above;
3. computation of the spectrum of unnormalized Laplacian matrix based on adjacency matrix in order to define the best number of clusters, k, according to the k-smallest eigenvalue, as explained below;
4. computation of the first k eigenvectors of unnormalized and of two normalized Laplacian matrices for all weight matrices W_ω;
5. definition, for all the weights and for the three spectral algorithms, of the matrix U_{nxk} containing the first k eigenvectors as columns;
6. clustering the nodes of the network into clusters $C_1,\ldots,C_k$ using the k-means algorithm applied to the rows of the U_{nxk} matrix;
7. check of the continuity of the obtained clusters C_k;
8. definition of the set of edge-cuts (or boundary pipes) N_{ec}.

The boundary pipes are links for which the start node and the end node belong to different clusters C_k.

It is important to highlight that in all three algorithms, an important aspect is to change the representation of the nodes n from Euclidian space to points of the matrix U_{nxk}, that enhances the cluster-properties in the data, so that clusters can be trivially detected in the new representation, in particular, through the simple k-means clustering algorithm (Tibshirani et al., 2001; von Luxburg, 2007).

Phase 2: water network dividing

Phase 1 provides the edge-cut between clusters, i.e. the set of N_{ec} boundary pipes along which gate valves or flow meters must be installed. First, the number N_{fm} of flow meters to be inserted in the network is chosen, so that the remaining boundary pipes $N_{bv} = (N_{ec}\text{-}N_{fm})$ are closed by inserting gate valves. In order to simplify the water budget computation, it is better to keep N_{fm} as low as possible (Di Nardo et al, 2016b). This problem can be assimilated to a valve placement problem in WDNs. This is a NP-hard problem (Bodlaender et al., 2010) and it requires heuristic algorithms to find optimal solutions (Tindell et al., 1992). In other terms, once defined all the e_{ij} boundary pipes between clusters, those that must be closed must be chosen among all the possible combinations N_{DL} of water network partitioning layouts, expressed by the binomial coefficient:

$$N_{DL} = \binom{N_{ec}}{N_{fm}} \tag{4}$$

It is important to underline that N_{DL} can be, already for a small water supply network and for a small number k of DMAs, such a huge number that it is often computationally impossible to investigate all the solution space.

However, closing pipes to divide the districts significantly changes the network layout, reducing the topological connectivity and the energy redundancy and, consequently, worsening the hydraulic performance.

Therefore, an optimization technique has been developed, in order to find, once fixed the number of flow meters N_{fm}, the optimal choice of the boundary pipes along which gate valves are to be inserted, by minimizing the alteration of the hydraulic performance and of the level of service for the users. This aim has been achieved by a heuristic procedure carried out with a Genetic Algorithm (GA) developed by the authors (Di Nardo et al, 2016a), maximizing the following objective function:

$$\max\left(\gamma\sum_{i=1}^{n}(z_i + h_i)\cdot Q_i\right) \tag{5}$$

corresponding to the total nodal power of the network (Di Nardo et al., 2013a), in which γ is the specific weight of water, z_i, h_i and Q_i are, respectively, the geodetic elevation, the pressure and the water demand at the i-th node. The GA parameters are the following: each individual of the population is a sequence of N_{ec} binary chromosomes corresponding to the pipes belonging to the edge-cut set; the l-th chromosome is set to 1 if a gate valve is inserted along the corresponding l-th pipe, while it is set to 0 if a flow meter is inserted. The GA has been carried out with 100 generations and with

a population consisting of 500 individuals with a crossover percentage equal to $P_{cross} = 0.8$.

In order to compute the objective function, hydraulic simulations are carried out in the GA. They are carried out using the freeware software EPANET2 (Rossman, 2000), that numerically solves the non-linear hydraulic equations of the water system.

Finally, after the dividing phase, hydraulic simulations are required to compute some performance indices (Di Nardo et al., 2015c) aimed at evaluating the hydraulic performance of WNP and so allowing to compare different layouts.

Case study

The effectiveness of the proposed procedure has been tested for the real case study of the water supply network of Parete, a town with 10,800 inhabitants, located in a densely populated area near Caserta (Italy). The water network has two sources and its main topological and energy characteristics are reported in Tables 1 and 2, respectively.

The network consists of $m = 282$ links and $n = 184$ nodes and, from a topological point of view, in agreement with most real systems, it is a sparse network, so it is not fully connected and its number of edges $m << n^2$, with a link density value (i.e. the ratio between the actual number of links and the number of links of a fully connected network with the same number of nodes) $q = 0.017$. As the number of edges that can be connected to a single node is limited by the physical space in spatial networks (Boccaletti et al, 2006), average node degree $\hat{K}$ =3.05 is small. The case study shows a small average path length $APL = 8.80$, presenting itself as a cohesive and robust network (Yazdani and Jeffrey, 2011) as well as the value of graph diameter $Dm = 20$ shows that the nodes are mutually and easily reachable and that the network is ordered in a decentralized fashion (Yazdani and Jeffrey, 2010), which is an important aspect for an efficient communication (the flow in the case of hydraulic networks). Concerning the main spectral measurements, the "spectral gap" $\Delta\lambda$ (the difference between the two largest eigenvalues of the adjacency matrix) is equal to 0.062 and the "algebraic connectivity" λ_2 (Fiedler, 1973) (the second smallest eigenvalue of the Laplacian matrix) is equal to 0.021. Both these values are small, showing that the graph arrangement can be decomposed into isolated parts (clusters or districts) (Estrada, 2006).

The hydraulic performances of the water supply network of Parete, reported in Table 2, is good in terms of maximum and mean nodal pressure heads, with h_{max} and h_{mean} higher than the design pressure head $h^* = 25$ m (the pressure head required to satisfy water demand at all nodes). Conversely, the minimum pressure head h_{min} is lower than h^*, indicating that in some nodes the design pressure requirement is not fulfilled. Consequently, the system shows little energy resilience and so a "low availability" of the water system to be partitioned without a decrease in hydraulic performance (Greco et al, 2012). In Table 2 the value of the input power P_A (a global performance index measuring the amount of energy entering the water system through the reservoirs and provided by pumps) is also reported.

Table 1 Topological characteristics of the water distribution network of Parete

m	n	q	K	APL	Dm	λ_2	$\Delta\lambda$
282	184	0.017	3.05	8.80	20	0.021	0.062

Table 2 Hydraulic characteristics of the water distribution network of Parete

h*	h_{min}	h_{mean}	h_{max}	P_A
[m]	[m]	[m]	[m]	[kW]
25.00	21.36	31.05	50.47	12122.11

h* = 25 m (the pressure head required to satisfy water demand at all nodes)

Following the steps of the proposed methodology for water network partitioning, WDN first can be seen as a graph (step 1) $G = (V, E)$ in which V is the set of n vertices v_i (the junction, the delivering nodes and the reservoirs) and E is the set of m edges e_l (the pipes connecting the nodes). Then, all the above defined five weight matrices W_ω are computed (step 2) and used to choose the most appropriate number k of clusters; this is a common problem in all clustering algorithms. The tool designed for spectral clustering, the eigengap heuristic (von Luxburg, 2007), is applied to all three graph Laplacian matrices, choosing the number of clusters k such that all eigenvalues $\lambda_1,\ldots,\lambda_k$ assume small and similar values, while λ_{k+1} is relatively larger (step 3). According to this criterion, as shown in Fig. 1 for the case study of the water network of Parete, relatively to no-weight Laplacian matrix, where the first 10 smallest eigenvalues are plotted, the most appropriate number of clusters is three or four. It is worth to note that, as explained in von Luxburg, (2007), the eigengap heuristic works well only if the clusters in the data are very well pronounced, i.e. the more overlapping the clusters are, the less clear is the detection of the number of clusters. However, such a method gives in any case a useful preliminary indication.

Once fixed the number $k = 4$ of clusters into which the network is subdivided, the first phase of the proposed partitioning procedure provides the spectral clustering of the water supply network of Parete. A total number of clustering layouts $N_{CL} = 15$ is obtained (three algorithms for five weight matrices), as reported in Table 3. The result of the partitioning of the graph is represented in Fig. 2, without loss of generality, for the case of pipe diameter as weight and L_{rw}

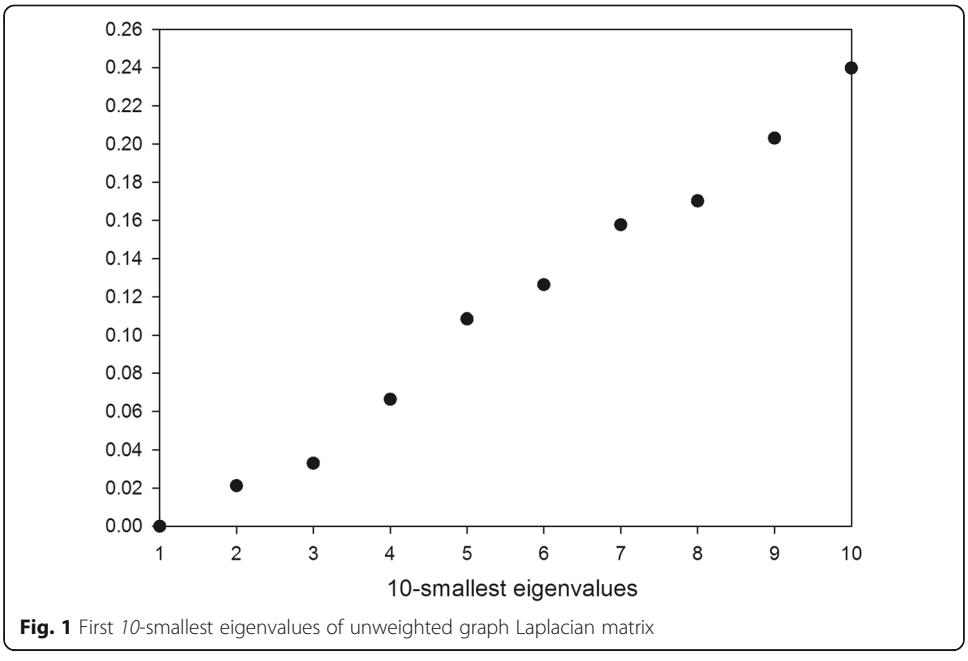

Fig. 1 First *10*-smallest eigenvalues of unweighted graph Laplacian matrix

Table 3 Characteristics of the clusters obtained with the three tested algorithms and the five adopted pipe weights

Laplacian Matrix	Weight	n° nodes DMA 1	n° nodes DMA 2	n° nodes DMA 3	n° nodes DMA 4	I_b	N_{ec}
L	W_A	48	48	43	45	2.45	16
	W_D	53	46	43	42	4.97	17
	$W_{1/L}$	59	56	40	29	14.07	20
	W_C	84	65	19	16	33.83	26
	W_F	76	64	42	2	32.54	15
L_{rw}	W_A	48	48	45	43	2.45	16
	W_D	49	47	45	43	2.58	17
	$W_{1/L}$	54	45	44	41	5.60	19
	W_C	82	40	33	29	24.43	21
	W_F	78	43	36	27	22.32	24
L_{sym}	W_A	48	48	45	43	2.45	16
	W_D	49	47	45	43	2.58	17
	$W_{1/L}$	55	47	42	40	6.68	19
	W_C	126	32	15	11	54.10	-
	W_F	86	46	28	24	28.33	-

Laplacian matrix. The network nodes are plotted in the eigenspace of the first three non-constant eigenvectors. The division into four sub-regions is evident, as the points result clearly arranged into four distinct groups. The results in terms of topological metrics are reported, for each Laplacian matrix and for each weight combination in Table 3, which gives: the number of nodes n_k of each cluster, the balanced node index I_b (standard deviation of the total number of nodes of the four clusters), the number N_{ec} of pipes of the edge-cut set.

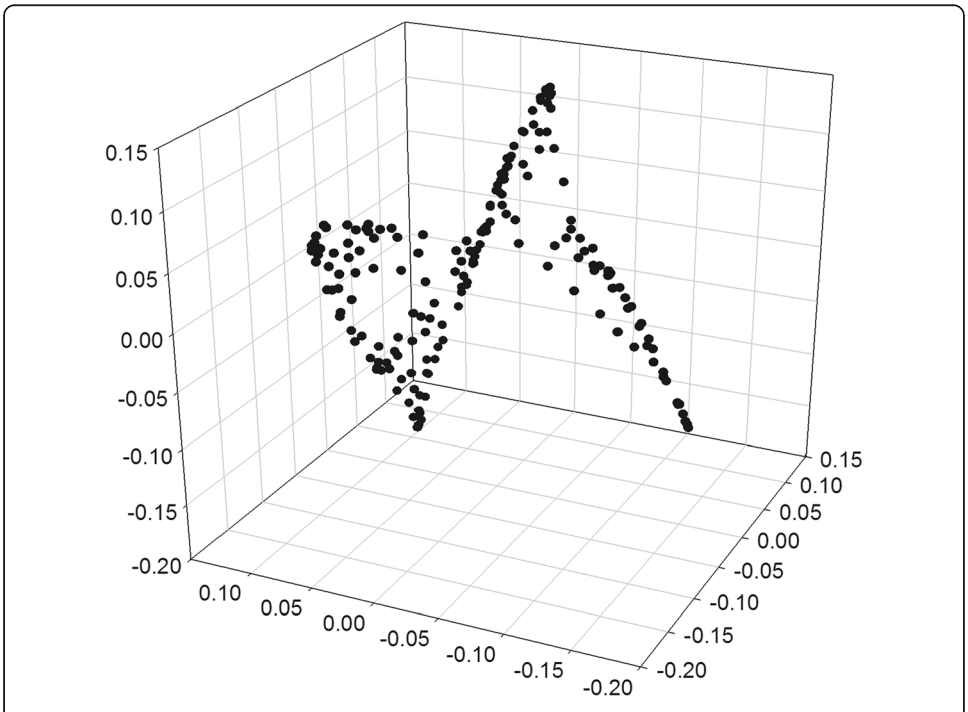

Fig. 2 Node coordinates in the eigenspace of the first 3-smallest eigenvectors of the diameter-weighted graph L_{rw} Laplacian matrix

For the two last solutions in Table 3 (W_C and W_F weight matrices with L_{sym} as Laplacian matrix), the continuity of the network is not ensured. For the continuity check, it has been exploited another important property of the Laplacian matrix, namely the multiplicity m_a of its zero eigenvalue, that is equal to the number of connected sub-graphs of the network. So, the multiplicity of the zero eigenvalue of the unweighted unnormalized Laplacian matrix of the graph subdivided into four clusters has been evaluated. It resulted $m_a = 4$, indicating that the obtained sub-graphs result internally connected, while, if $m_a > 4$, it would mean that the network has been divided into more than four sub-graphs.

It is also evident that the most balanced layouts (i.e. clusters with similar numbers of nodes) correspond to the $W_{A,}W_D$ and $W_{1/L}$ for all three Laplacian matrices: as reported in Table 3, they lead to the lowest values of I_b. As expected, the most balanced layout corresponds to no-weight matrix. In fact, without any weight, the spectral clustering leads to sub-graphs containing similar numbers of links that, in a WDN like Parete, implies also similar numbers of nodes (indeed the number of links connected to a node varies only slightly throughout the network). Conversely, when weights are given to pipes, the sum of the weights of the pipes belonging to the clusters is balanced, which does not necessarily imply that the clusters contain similar numbers of nodes. The last step of the clustering phase is the definition of the edge-cut set. This phase must be achieved with the aim of minimizing, in the subsequent dividing phase, the network perturbation and the investment related to the insertion of hydraulic devices. In this respect, it is reasonable that edge-cut sets containing a small number N_{ec} of intra-cluster pipes would be preferable. Also for this index, the optimal solutions correspond to $W_{A,}W_D$ and $W_{1/L}$ for all three Laplacian matrices (they lead to the lowest values of N_{ec}). Even if the W_F with L Laplacian matrix provides the lowest N_{ec}, such a solution should not be considered, as it corresponds to a very unbalanced cluster layout.

The results given in Table 3 should be interpreted considering that, while unweighted spectral clustering provides the edge-cut set with the minimum N_{ec} (compatible with the requirement of obtaining balanced clusters), weighted spectral clustering minimises the sum of the weights of the edges constituting the edge-cut set. Hence, it was expected that such a minimization would have not necessarily led to small values of N_{ec}.

In this respect, it is interesting to note that the application of weights resulted in edge-cut sets less suitable to carry out the following dividing phase (achieved with EPANET software embedded in the GA) with little disturbance to the hydraulic performance of the network, if compared to the edge-cut set identified by the unweighted spectral clustering. In fact, for the dividing phase to affect as little as possible the hydraulic performance, the closed pipes should be the ones carrying small flows (pipes with small conductance, i.e. with small diameter), while the pipes remaining open after the dividing phase (i.e. the pipes along which water meters are installed) should be large, as they have to be capable of carrying also the discharge which, before the division, flew through the closed pipes. In other words, the edge-cut set should contain pipes with both small and large diameters (or hydraulic conductance). As reported in Table 4, which gives the diameters of the pipes belonging to the edge-cut sets obtained with different weights, in the case of the WDN of Parete, the edge-cut set provided by the unweighted clustering has such a feature. Conversely, the more the adopted weights

Table 4 Characteristics of the edge-cut set obtained with the three tested algorithms and the five assumed pipe weights

Laplacian Matrix	Weight	N_{ec}	Multiplicity of pipe diameters D [mm]						
			60	80	100	110	125	150	200
L	W_A	16	5	1	4	1	1	2	2
	W_D	17	6	1	4	1	1	2	2
	$W_{1/L}$	20	5	1	3	0	1	3	7
	W_C	26	20	0	2	1	1	1	1
	W_F	15	4	2	4	0	1	1	3
L_{rw}	W_A	16	5	1	4	1	1	2	2
	W_D	17	6	1	4	1	1	2	2
	$W_{1/L}$	19	7	1	3	1	1	2	4
	W_C	21	10	3	3	1	1	3	0
	W_F	24	10	1	4	1	1	2	6
L_{sym}	W_A	16	5	1	4	1	1	2	2
	W_D	17	6	1	4	1	1	2	2
	$W_{1/L}$	19	6	1	4	1	1	2	4
	W_C	-	-	-	-	-	-	-	-
	W_F	-	-	-	-	-	-	-	-

are effective (i.e. the weights assume very different values for the various pipes of the network), the more the pipes of the edge-cut tend to be homogeneous (for instance, when the hydraulic conductance is adopted as a weight, the edge-cut set is formed by only very small pipes).

As shown in Table 5, which gives the main hydraulic performance indices of the network after the dividing phase (evaluated through the EPANET software), the number of

Table 5 Hydraulic performance indices after the dividing phase for the three tested algorithms and the five adopted pipe weights

Laplacian Matrix	Weight	N_{ec} [-]	N_{bv} [-]	N_{fm} [-]	P_D [kW]	P_N [kW]	h_{min} [m]	h_{mean} [m]	h_{max} [m]
L	W_A	16	11	5	1831.64	10290.47	22.78	30.46	50.07
	W_D	17	12	5	1875.29	10246.82	22.09	29.82	50.16
	$W_{1/L}$	20	15	5	1890.23	10231.88	22.23	29.93	49.87
	W_C	26	21	5	1954.84	10167.27	14.04	28.68	50.03
	W_F	15	10	5	1716.15	10405.96	21.58	30.99	50.41
L_{rw}	W_A	16	11	5	1831.64	10290.47	22.78	30.46	50.07
	W_D	17	12	5	1875.29	10246.82	22.09	29.82	50.16
	$W_{1/L}$	19	14	5	1863.96	10258.15	22.49	30.24	49.98
	W_C	21	16	5	1895.35	10226.76	20.81	29.13	50.61
	W_F	24	19	5	1720.86	10401.25	20.79	30.81	50.50
L_{sym}	W_A	16	11	5	1831.64	10290.47	22.78	30.46	50.07
	W_D	17	12	5	1875.29	10246.82	22.09	29.82	50.16
	$W_{1/L}$	19	14	5	1877.49	10244.62	22.36	30.09	50.11
	W_C	-	-	-	-	-	-	-	-
	W_F	-	-	-	-	-	-	-	-

flow meters has been fixed for all combinations to $N_{fm} = 5$, which is the minimum possible number that guarantees the hydraulic performance of the network and, at the same time, simplifies the computation of the water budget (Water Industry Research Ltd, 1999), allowing an easier identification of water losses. Clearly, the number of gate valves is in all cases equal to the difference $N_{bv} = N_{ec}$-N_{fm}.

After the first clustering phase, the dividing phase has been carried out, computing all the hydraulic performance metrics reported in Table 5, namely: the dissipated power P_D; the total nodal delivered power $P_N = P_A$-P_D (Di Nardo et al., 2013a); the minimum, mean and maximum pressure h_{min}, h_{mean} and h_{max}.

Obviously, the results for the two last clustering solutions (W_C and W_F weight matrices with L_{sym} Laplacian matrix) are not reported, because the continuity of the network is not respected and so the hydraulic simulation needed for the evaluation of the hydraulic performance could not be carried out.

It is important to highlight that the reported results are the optimal for each weight-Laplacian combination, meaning that, for the fixed number of flow-meters $N_{fm} = 5$ and within the investigated solution space, the power P_D dissipated by the system is minimized, and, consequently, the total nodal power P_N is maximized.

As expected, the results in terms of hydraulic performance indicate that the best solutions correspond to the $W_{A,}W_D$ and $W_{1/L}$ for all three Laplacian matrices, as they lead to the smallest numbers of closed pipes. Also in this phase, even if the clusters layouts obtained with W_F by means of both L and L_{rw} Laplacian matrix seem to correspond to the lowest value of dissipated power P_D, they cannot be considered as good solutions, because they are very unbalanced.

For the presented case study, it is clear that, from both topological and hydraulic point of views, the best solutions have been achieved with the normalized L_{rw} Laplacian matrix. Indeed, at the same time it provides the most balanced clusters solutions, an edge-cut set with few pipes (Table 3), the lowest dissipated power, and the highest minimum pressure head. With reference to the weight choice, it looks clear that, even if there is not a great difference between $W_{A,}W_D$ and $W_{1/L}$, the best solution was achieved with the unweighted matrix, regardless of the adopted Laplacian matrix. In this respect, it is worth to note that this result cannot be generalized, as it depends on the peculiar distribution of pipe diameters of the analysed WDN. In fact, unweighted clustering takes into account only the topological structure of the network, without using any information related to the hydraulic characteristics of the pipes. However, the obtained results are good in terms of nodal pressure and confirm the suitability of spectral clustering for water network partitioning. Further investigation about the choice of the weights is required to define a spectral clustering approach of general validity for the definition of DMAs.

Finally, Fig. 3 shows the WNP of the network of Parete corresponding to the best solution in terms of minimum pressure $h_{min} = 22.78$, obtained with unweighted matrix and L_{rw} Laplacian. In particular, in the left pane of Fig. 3, the first clustering phase is reported, highlighting the edge-cut set (dashed lines). In the right pane, the second dividing phase is illustrated, highlighting the optimal positioning of devices, which ensures the minimum hydraulic performance deterioration. For comparison, in Fig. 4, the WNP of Parete obtained with conductance-weighted adjacency matrix and L_{rw} Laplacian is reported, highlighting the less balanced obtained layout. The different

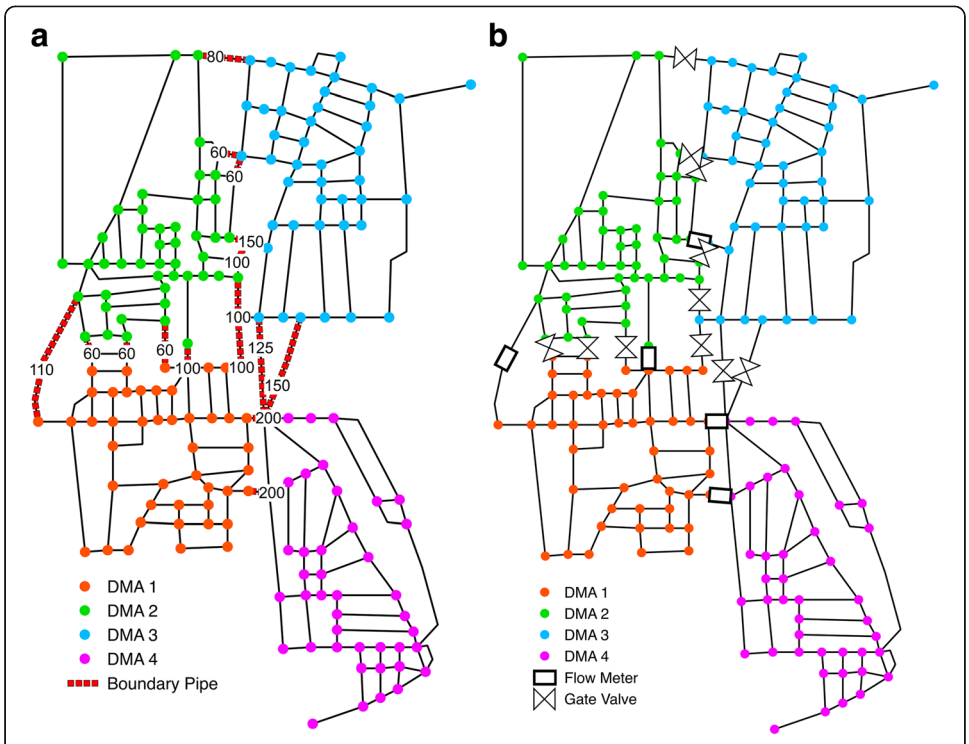

Fig. 3 Parete WSN partitioning in *4*-DMA for the unweighted graph adjacency matrix: clustering phase (**a**) and dividing phase (**b**)

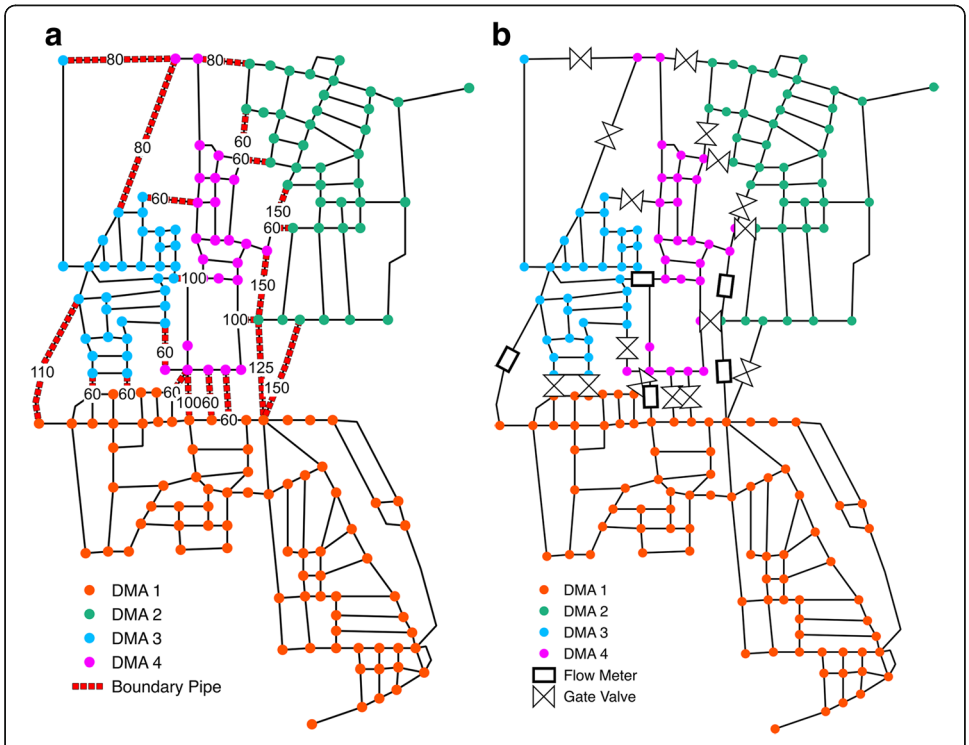

Fig. 4 Parete WSN partitioning in *4*-DMA for the conductance-weighted graph adjacency matrix and L_{nw} Laplacian: clustering phase (**a**) and dividing phase (**b**). Flow meters are represented by rectangles while gate valves with the double triangle

shapes and dimensions of the obtained clusters, as well as the greater number of edge-cuts are also evident. In both cases (unweighted WNP and conductance-weighted WNP), the gate-valves (closing pipes) have been located along the pipes with smallest diameters by means of the *GA* algorithm, so to reduce hydraulic performance deterioration.

Conclusions

The division of a WDN into DMAs aims at improving water supply network management and so, consequently, leakage detection and system safety. At the same time, the closure of pipes with gate valves to define the DMAs unavoidably increases the hydraulic head losses, leading to lower pressure at the water delivery nodes, compared to non-clustered layout. So far, although the design of optimal DMA layout is a problem deeply studied in the scientific literature, there is not an established procedure to solve it. In this regard, the paper presents an application to a real WDN of weighted spectral clustering for water network partitioning.

Spectral clustering is based on the eigenvalues of the graph Laplacian matrix of the network, for which three different formulations have been tested. Five different weights have been adopted, chosen among the major characteristics of the pipes: adjacency (in this case the partitioning is based only on the topology of the network), diameter, length, conductance and flow. Aim of the application is to understand which of the considered characteristics provides the best clustering layout, in terms of minimizing the edge-cuts and simultaneously balancing the dimensions of the clusters.

Compared to other heuristic methodologies, weighted spectral clustering allows to take into account either topological, geometrical, or hydraulic information about the system, within the framework of an elegant mathematical formalism.

Simulation results for the analysed case study, carried out with a number of DMAs $k = 4$, defined through the analysis of the eigenvalues of the unweighted Lapalcian matrix, confirm the effectiveness of the procedure, providing balanced clustering layouts and small numbers of intra-cluster boundary pipes. The latter result may favour the following heuristic dividing phase, consisting in the choice of the positions of flow meters and gate valves along the pipes of the edge-cut. Indeed, the hydraulic performance of the network, measured with several indices, is satisfactorily preserved in most of the weight-Laplacian combinations.

In particular, in this study the best solution was found, with the spectral clustering algorithm, using unweighted matrices. This result is different from previous studies found in the literature, in which different clustering techniques were adopted, and weighted matrices provided the best results. This result directly depends on the distribution of pipe diameters within the considered network, and therefore cannot be considered of general validity. In fact, it can be ascribed to the fact that, when weights related to pipe geometry are minimized, the optimal edge-cut tends to be formed by pipes with similar characteristics. In the dividing phase, instead, the best hydraulic results are obtained by installing gate valves along pipes with small diameter, and water meters along pipes with large diameter. Therefore, further investigation is required to define a weighted spectral clustering approach of general validity for the definition of DMAs.

Acknowledgement
The authors would like to thank Action Group CTRL + SWAN of EIP on Water for supporting this research.

Funding
No funding was received.

Authors' contributions
The authors contributed equally to this manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Dipartimento di Ingegneria Civile Design Edilizia e Ambiente, Università degli Studi della Campania 'L. Vanvitelli', via Roma 29, Aversa 81031, Italy. [2]Action Group CTRL + SWAN of the European Innovation Partnership on Water EU, via Roma 29, Aversa 81031, Italy. [3]Istituto Sistemi Complessi (Consiglio Nazionale delle Ricerche), Via dei Taurini 19, Rome 00185, Italy.

Received: 8 March 2017 Accepted: 15 May 2017
Published online: 30 June 2017

References

Alonso JM, Alvarruiz F, Guerrero D, Hernàndez V, Ruiz PA, Vidal AM, Martìnez F, Vercher J, Ulanicki B (2000) Parallel computing in water network analysis and leakage minimization. J Hydraul Eng ASCE 126:251–260
Alvisi S, Franchini M (2014) A procedure for the design of district metered areas in water distribution systems. Procedia Eng 70:41–50
Barthelemy M, Flammini A (2008) Modelling urban street patterns. Phys Rev Lett 100:138702, doi: 10.1103
Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D-U (2006) Complex networks. Structure and dynamics. Phys Rep 424:175–308
Bodlaender HL, Hendriks A, Grigoriev A, Grigorieva NV (2010) The valve location problem in simple network topologies. INFORMS J Comput 22(3):433–442
Carvalho R, Buzna L, Bono F, Gutierrez E, Just W, Arrowsmith D (2009) Robustness of trans-European gas networks. Phys Rev E 80:016106, doi: 10.1103
Chung F. (1997) Spectral graph theory. CBMS Regional Conference Series in Mathematics 92:212
Di Nardo A, Di Natale M, Santonastaso G.F, Venticinque S (2013a) An automated tool for smart water network partitioning. Water Resour Manag 27:4493–4508
Di Nardo A, Di Natale M, Santonastaso G.F, Tzatchkov V.G, Alcocer Yamanaka V.H (2013b) Water Network Sectorization based on genetic algorithm and minimum dissipated power paths. J Water Sci Technol Water Supply 13:951–957
Di Nardo A, Di Natale M, Giudicianni C, Musmarra D, Santonastaso G.F, Simone A (2015a) Water distribution system clustering and partitioning based on social network algorithms. Procedia Eng 119:196–205
Di Nardo A, Di Natale M, Musmarra D, Santonastaso GF, Tzatchkov V, Alcocer-Yamanaka V.H (2015b) Dual-use value of network partitioning for water system management and protection from malicious contamination. J Hydroinf 17: 361–76
Di Nardo A, Di Natale M, Santonastaso GF, Tzatchkov VG, Alcocer-Yamanaka VH (2015c) Performance indices for water network partitioning and sectorization. Water Sci Technol Water Supply, 15:499–509
Di Nardo A, Di Natale M, Musmarra D, Santonastaso GF., Tuccinardi F.P., Zaccone G. (2016a). Software for partitioning and protecting a water supply network. Civ Eng Environ Syst, 33:55–69
Di Nardo A, Di Natale M, Giudicianni C, Santonastaso G.F, Tzatchkov V.G, Varela J.M.R, Yamanaka V.H.A (2016b) Water Supply Network Partitioning Based on Simultaneous Cost and Energy Optimization. Procedia Eng 162:238–245
Di Nardo A, Di Natale M, Giudicianni C, Greco R, Santonastaso GF (2017) Water supply network partitioning based on weighted spectral clustering, vol 693, Studies in computational intelligence: complex networks & their applications., pp 797–807, doi: 10.1007
Diao K, Zhou Y, Rauch W (2013) Automated creation of district metered area boundaries in water distribution systems. J Water Resour Plan Manag 139:184–190
Estrada E (2006) Network robustness to targeted attacks. The interplay of expansibility and degree distribution. Eu Phys J B 52:563–574
Ferrari G, Savic D, Becciu G (2014) A graph theoretic approach and sound engineering principles for design of district metered areas. J Water Resour Plan Manag 140:04014036, doi: 10.1061
Fiedler M (1973) Algebraic connectivity of graphs. Czechoslov Math J 23:298
Fortunato S (2010) Community detection in graphs. Phys Rep 486:75–174
Gomes R, Sá Marques A, Sousa J (2012) Identification of the optimal entry points at district metered areas and implementation of pressure management. Urban Water J 9:365–384
Greco R, Di Nardo A, Santonastaso GF (2012) Resilience and entropy as indices of robustness of water distribution networks. J Hydroinf 14:761–771
Herrera M, Canu S, Karatzoglou A, Pérez-García R, Izquierdo J (2010) In: Swayne DA, Yang W, Voinov AA, Rizzoli A, Filatova T (eds) Proceedings of international environmental modelling and software society(IEMSS), Ottawa, Canada, July 5-8

Izquierdo J, Herrera M, Montalvol P-GR (2011) Division of water distribution systems into district metered areas using a multi-agent based approach. Commun Comput Inf Sci 50:167–180
Mays W (2000) Water distribution systems handbook. McGraw-Hill, New York
Newman MEJ (2003) The structure and function of networks. SIAM Rev 45:167–256
Ng AY, Jordan MI, Weiss Y (2001) On Spectral Clustering: Analysis and an algorithm. Adv Neural Inf Process Syst, Dietterich TG, Becker S, Ghahramani Z (Eds.), vol. 14, MIT Press, Cambridge, USA
Perelman LS, Allen M, Preis A, Iqbal M, Whittle AJ (2015) Automated sub-zoning of water distribution systems. Environ Model Softw 65:1–14
Rossman LA (2000) EPANET2 users manual. US EPA, Cincinnati, Ohio
Saerens M, Fouss F, Yen L, Dupont P (2004) The principal components analysis of a graph, and its relationships to spectral clustering. In: Proceedings of the 15th European conference on machine learning (ECML)., pp 371–383
Shi J, Malik J (2000) Normalized cuts and image segmentation. IEEE Trans-actions Pattern Anal Mach Intell 22:888–905
MATLAB SimuLink Reference Books (2006) MathWorks,.Inc., Natick, MA,
Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a dataset via the gap statistic. J R Stat Soc 63(2):411–423
Tindell KW, Burns A, Wellings AJ (1992) Allocating hard real-time tasks: an NP-hard problem made easy. Real-Time Syst 4(2):145–165
Tzatchkov VG, Alcocer-Yamanaka VH, Ortiz VB (2006) Graph theory based algorithms for water distribution network sectorization projects. In: Buchberger SG, Clark RM, Grayman WM, Uber JG (eds) Proc. of 8th annual water distribution systems analysis symposium, Cincinnati, USA
Von Luxburg U (2007) A tutorial on spectral clustering. Stat Comput 17:395–416
Water Industry Research Ltd (1999) A manual of DMA practice. Water Industry Research, UK, London
Yazdani A, Jeffrey P (2010) Robustness and vulnerability analysis of water distribution networks using graph theoretic and complex network principles. In: Proceeding of water distribution system analysis, Tucson, Arizona, September 12–15
Yazdani A, Jeffrey P (2011) Complex network analysis of water distribution systems. Interdisciplinary J Nonlinear Sci Chaos 21:016111

Submit your manuscript to a SpringerOpen journal and benefit from:

► Convenient online submission
► Rigorous peer review
► Open access: articles freely available online
► High visibility within the field
► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com