

Water Distribution Network Clustering: Graph Partitioning or Spectral Algorithms?

A. Di Nardo^{1,2,3(✉)}, M. Di Natale^{1,2}, C. Giudicianni¹, R. Greco^{1,2},
and G. F. Santonastaso^{1,2}

¹ Università degli Studi della Campania “Luigi Vanvitelli”, Aversa, Italy
{`armando.dinardo,michele.dinatale,carlo.giudicianni,roberto.greco,`
`giovannifrancesco.santonastaso`}@unicampania.it

² Action Group CTRL+SWAN of European Innovative Partnership on Water,
Aversa, Italy

³ CNR Istituto Sistemi Complessi, Aversa, Italy

Abstract. Water Network Partitioning (WNP) is among the most attractive and studied strategies for the improvement of the Water Distribution Network (WDN) management. The proper definition of sub-regions (called clusters or districts) with high link density between nodes in the same group, and a relatively low link density between nodes in different groups, is a crucial aspect for the partitioning of a water system. If on one hand the definition of these monitored sub-areas, called District Metered Areas (DMAs), allows simplifying the water balance, the pressure control, the water leaks identification and the water quality protection, on the other hand it may worsen the hydraulic performance and the reliability of the system. In this paper, two clustering algorithms, graph partitioning based on a Multi-Level Recursive Bisection and Spectral Clustering, were used to define the districts. Some of the major geometrical and hydraulic characteristics of the network has been adopted as weights in the partitioning procedure. A comparison between the two clustering methods was made for a real water network of the South Italy, Parete, evaluating some clustering quality and hydraulic indices, in order to define the algorithm and the weight which work better for the definition of the optimal clustering layout.

1 Introduction

Clustering is the process of dividing a set of given elements into groups (or communities) such that elements in the same group are somewhat similar to each other, while elements from different groups are dissimilar [1]. The purpose and, at the same time, the key strength of community detection is to identify the groups and their organization, by only using the topological information [2]. The grouping is based on similarity measures defined for the elements without relying on any a priori information on how the classification should be done [3]. A survey of the most popular clustering methods used in pattern recognition is provided in [4]. Cluster analysis has been applied to many areas: gene analysis

[5]; natural language processing [6]; galaxy formation [7]; image segmentation [8]; global air transportation network [9]; logistics [10]; water distribution networks [11,12].

In all cases, clustering is used for modeling and predicting the functioning, behavior and evolution of a set of objects, identifying substructures to improve the analysis of the system and the detection of similarities and anomalies. Due to the wide range of applications, many clustering algorithms have been proposed. Generally, clustering algorithms can be roughly divided into two main groups [13]: (a) hierarchical and (b) partitioning algorithms, according to the solution representation and the algorithmic approach for generating clusters.

The hierarchical algorithms form clusters through agglomerations (agglomerative algorithms) or divisions (divisive algorithms), creating a tree structure where supergroups (or subgroups) of clusters are gradually defined.

Partitioning algorithms split the n elements of a set into $k \leq n$ clusters by moving them between clusters until a stop criterion is verified. Each solution is assessed by a given objective function.

Generally, graphs are structures formed by a set of n vertices (nodes) and a set of m edges (links) connecting vertices. Also Water Distribution Networks (WDNs) can be modeled as a graph, in particular as link-node planar spatially organized weighted graphs, for which pipes and valves correspond to links, while junctions, water sources and water demand points correspond to nodes [14]. WDNs belong to the class of networks strongly constrained by their geographical embedding [15], for which it is unlikely to find connections between distant nodes, due to obvious physical constraints. Furthermore, WDNs can be considered as complex networks [16], since they are often constituted of thousands of nodes and links, they are strongly looped and topologically irregular, as they follow the shape of the road network of the city they serve. These characteristics make WDNs management arduous, with many operational problems, such as water and energy losses [17].

In this regard, in the last years, Water Network Partitioning (WNP) has become one of the most attractive and studied strategies for the improvement of WDN management, such as leakage [18] and pressure management [19], monitoring of water quality [20], speeding up of repairing interventions [21]. This organization strategy is based on defining areas partially isolated from the rest of the network through the insertion of gate-valves and flow-meters along some pipes. Anyway, the partitioning of WDNs, besides increasing the energy consumption and the economic investment, could deteriorate the hydraulic performances and the reliability of the system due to the closure of pipes. The target of a partitioning layout design is to balance the negative and positive aspects described above [21], ensuring the fulfillment of the minimum required nodal pressure, so to satisfy the water demand of the users and save network resilience [21]. This is achieved by defining the proper number, shape and dimension of clusters, by minimizing the number of boundary pipes and then by optimally locating gate-valves and flow-meters. In this regard, WNP is usually carried out in two different phases [22,23]: (a) *clustering*, aimed to define the shape and

the dimension of the network subsets, balancing the number of nodes of each cluster and minimizing the number of edge-cuts and (b) *dividing*, aimed to the physical partitioning of the network, by selecting pipes along which flow meters or gate valves are to be inserted, minimizing the economic investment and the hydraulic performance deterioration. It is evident that the definition of a WNP is a complex challenge for operators, depending on the design choices of both clustering and dividing phases.

The number of possible clusters and gate-valves/flow-meters combinations grows enormously with the network dimension [24], making necessary the recourse to complex mathematical and computational algorithms.

This paper is focused on the first phase of the partitioning of water networks, i.e. the definition of the clusters, since a good subdivision of the network certainly affects positively the subsequent phase of device insertion, in terms of economic investment and hydraulic performances. In this respect, the aim of this work is to investigate the feasibility of two of the most adopted clustering algorithms, a graph partitioning algorithm with multi-level recursive bisection, implemented in *Metis* software [25,26], and a spectral clustering based on normalized cut algorithm *NCut* [27], in order to establish which of them works better for the definition of proper grouping (districts) for water distribution networks.

The comparison is made for a real water network which serves Parete, a city near Naples, Italy. The network graph was considered either un-weighted, or weighted with some of the major geometric/hydraulic characteristics, in order to take into account also different operator purposes. The network graph was clustered into $k = 2$ to $k = 10$ clusters with both the clustering algorithms and with the different chosen weights. Finally, some quality and hydraulic indices were used to compare the different obtained clustering layouts.

2 Clustering Algorithms

A Water Distribution Networks can be modelled as a 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_{ij} (or links).

A k -way graph clustering problem consists in grouping the set of vertices V of G into k subsets, P_1, P_2, \dots, P_k such that:

- $\bigcup_{c=1}^k P_c = V$ (the union of all clusters P_k must contain all the vertices v_i);
- $P_c \cap_t P_t = \emptyset$ (each vertex can belong to only one cluster P_c);
- $\emptyset \subset P_c \subset V$ (at least one vertex must belong to a cluster and a cluster cannot contain all vertices);
- $1 < k < n$ (the number k of clusters must be different from one and from the number n of vertices).

The graphs are considered undirected and weighted. Link weights express the strength of the links between elements, in terms of proximity and/or similarity, indicated with non-negative weight $\varepsilon_{ij} > 0$ if i and j are linked, $\varepsilon_{ij} = 0$ otherwise. Node weights express some node characteristics (i.e. age, demand, etc.),

indicated with positive weight ϖ_i , with $i \in V$. The edge-cut of a clustering, denoted with N_{EC} , is equal to the sum of weights (ε_{ij}) of edges (e_{ij}) whose incident vertices belong to different clusters:

$$N_{EC} = \sum_{i \in P_c \Rightarrow j \notin P_k} e_{ij} \text{ OR } \sum_{i \in P_c \Rightarrow j \notin P_k} \varepsilon_{ij} \tag{1}$$

In this paper two algorithms were used to cluster the graph of the water distribution network of Parete: (a) *Metis*, based on a graph partitioning technique, and (b) *NCut*, based on spectral clustering.

(a) *Metis*. The graph partitioning algorithm is based on Multi-Level Recursive Bisection (MLRB) proposed in *Metis* software by Karypis and Kumar [25, 26, 28]. *Metis* belongs to the class of multi-level partitioning techniques. Graph clustering starts with constructing a sequence of successively smaller (coarser) graphs, and a bisection of the coarsest graph is applied. Subsequently, a finer graph is generated in the next level based on the previous bisections. At each level, an iterative refinement algorithm is used to further improve the bisection.

The goal of *Metis* is to compute a k -way graph partitioning by minimizing the N_{ec} (Eq. 1) with the constraint: $I_B \leq 1 + \varepsilon$, where ε is a small positive number and I_B is the balance index defined as follows:

$$I_B = k \cdot \max(d_p) / \sum_i^n \varpi_i \tag{2}$$

where can be the maximum number of nodes or the maximum sum of the vertex-weights among the subsets P_i obtained by the k -way partitioning algorithm.

(b) *NCut*. The minimum-cut criterion for graph clustering refers to a class of techniques which divide a graph into subgraphs, such that the number of boundary edges is minimized. To avoid an unnatural bias towards splitting small-sized subgraphs based on the minimum-cut criterion, Shi and Malik [27] proposed the Normalized Cut (*NCut*), to compute the cut cost as a fraction of the total edge connections to all the nodes in a graph. A generalized eigenvector decomposition was used to speed up computation time. For this reason, these graph clustering algorithms, that rely on the eigenvector decomposition of a Laplacian matrix *Lxn* [29], are also called spectral clustering.

The Laplacian matrix is the difference between the diagonal matrix D_K , with the connectivity degrees of the nodes ($D_K = \text{diag}(K_i)$ and K_i is the degree of a node v_i), and the adjacency matrix A (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). The Normalized Cut exploits the property of the Normalized Laplacian Matrix defined as $L_{rw} = D_K^{-1}L$. The goal of *Ncut* algorithm is carried out by minimizing the following relationship:

$$\frac{N_{EC}}{\sum_{c=1}^k \text{vol}(P_c)} \tag{3}$$

where $\text{vol}(P_c)$ is the sum of the degrees or weighed degrees of all nodes that belong to the i -th cluster and N_{ec} is the edge-cut set between the clusters. The

minimum of Eq. (3) is achieved if all $vol(P_i)$ coincide or, in other words, if $vol(P_1) \cong vol(P_2) \cong \dots \cong vol(P_k)$; in this way the *NCut* algorithm tries to obtain k balanced clusters.

It is worth to highlight that both algorithms were performed with different weights of graph edges. Specifically, taking into account the strength of the connections between links provides certainly different cluster layout [30] which can meet different needs of the operators, offering specific solutions according to the peculiarities of the problem to solve. In this way, introducing weights for the pipes, the adjacency matrix A is replaced by the weight matrices W to calculate the Laplacian matrix and so the corresponding spectrum. The performances of the two clustering techniques as the weight changes are evaluated.

Several metrics were used to compare the performance of the two different adopted clustering techniques; specifically: two clustering quality indices (a), and three hydraulic indices (b) were evaluated for each cluster layout combination:

- (a) clustering quality indices
 - N_{EC} : it represents the total number of boundary pipes;
 - I_B : the *balance index* (as described above);
- (b) hydraulic indices
 - C_{EC} : it represents the sum of the ratios d/l of all boundary pipes, and can be considered as a proxy of pipe conductance;
 - R_{EC} : it represents the sum of the hydraulic resistances l/d^5 of all boundary pipes;

3 Case Study

The two described clustering algorithms were compared for a real water distribution network, which serves the city of Parete, near Naples, Italy, with 10,800 inhabitants. The water network has two sources, $m = 282$ links and $n = 184$ nodes; its main topological characteristics are reported in Table 1.

Table 1. Topological characteristics of the WDN of Parete

M	N	q	s	APL	λ_2	$\Delta\lambda$
[-]	[-]	[-]	[-]	[-]	[-]	[-]
282	184	0.017	3.05	8.80	0.021	0.062

From a topological point of view, in agreement with most large-scale real networks, it is a sparse network with a link density value $q=0.017$. Furthermore, the average node degree $K = 3.05$ is small, since in a WDN the number of edges that can be connected to a node is limited by physical constraints [16,17]. The small average path length $APL=8.80$ shows that the graph has a cohesive and robust behaviour [27], an important aspect for an efficient water flow. Regarding

the main spectral measurements, the spectral gap $\Delta\lambda= 0.062$ and the algebraic connectivity $\lambda_2= 0.021$, they assume low values, showing that the graph arrangement can be easily decomposed into isolated parts (clusters) [31,32].

The results of the analysis for all weight/algorithm combinations are reported, respectively, in Table 2 for *Metis* and in Table 3 for *NCut*. The graph of the network of Parete was subdivided in $k = 2$ to $k = 10$ clusters, analyzing the performances of the two clustering techniques as the number of clusters changed. The clustering phase was obtained considering the network graph un-weighted ($\varepsilon_{ij} = 1$), as well as weighted with some of the major geometric/hydraulic characteristics: namely, the ratio between diameter and length of pipes ($\varepsilon_{ij} = d/l$) and the hydraulic resistance ($\varepsilon_{ij} = l/d^5$). In order to better compare the two algorithms, *Metis* was forced to balance the clusters in a similar way as *NCut*, by using the following node weights:

- node degree $\varpi_i = K_i = \sum_{j=1}^n a_{ij}$, for the un-weighted link;
- weighted node degree $\varpi_i = K_{\varpi i} = \sum_{j=1}^n d_{ij}/l_{ij}$, coupled with the link weight $\varepsilon_{ij} = d_{ij}/l_{ij}$;
- weighted node degree $\varpi_i = K_{\varpi i} = \sum_{j=1}^n l_{ij}/d_{ij}^5$, coupled with the link weight $\varepsilon_{ij} = l_{ij}/d_{ij}^5$.

Table 2. Performance Indices for *Metis* algorithm considering the graph unweighted, and weighted with d/l and l/d^5 , and for $k = 2$ to $k = 10$ clusters

k	Unweighted				d/l weighted				l/d ⁵ weighted			
	N _{EC}	I _B	R _{EC}	C _{EC}	N _{EC}	I _B	R _{EC}	C _{EC}	N _{EC}	I _B	R _{EC}	C _{EC}
	[-]	[-]	[m ⁻⁴]	[-]	[-]	[-]	[m ⁻⁴]	[-]	[-]	[-]	[m ⁻⁴]	[-]
2	8	1.01	281	0.008	10	1.02	1430	0.004	19	1.05	158	0.049
3	10	1.01	367	0.011	16	1.08	1665	0.015	22	1.17	481	0.038
4	18	1.02	1327	0.022	20	1.24	1517	0.014	35	1.11	1389	0.064
5	24	1.03	1178	0.040	31	1.22	3919	0.019	30	1.22	1604	0.043
6	27	1.04	937	0.040	36	1.14	4417	0.025	40	1.37	2225	0.068
7	32	1.07	2151	0.047	47	1.37	5975	0.028	40	1.33	1997	0.070
8	40	1.04	2816	0.067	49	1.35	5983	0.041	45	1.61	3430	0.075
9	34	1.03	1678	0.052	54	1.32	7158	0.036	53	1.57	1939	0.115
10	39	1.09	2954	0.058	59	1.25	7744	0.052	51	1.36	3202	0.098

First, the number of the edge-cuts N_{EC} for all groupings from $k = 2$ to $k = 10$ clusters considering the graph un-weighted and weighted with d/l and l/d^5 are shown in Fig. 1. As expected, the number of boundaries increases as the

number of clusters increases for both algorithms, even if the increasing trends are clearer for *NCut* than *Metis*.

In general, *Metis* provides clustering solutions with higher number of boundaries than spectral *NCut*, and this difference between the two algorithms grows as the number of clusters increases for the un-weighted and d/l -weighted graph. For the l/d^5 -weighted graph this difference is evident for any value of k , reaching about 71% for $k = 4$.

Table 3. Performance Indices for the spectral *NCut* algorithm considering the graph unweighted and weighted with d/l and l/d^5 and for $k = 2$ to $k = 10$ clusters

k	Un-weighted				d/l weighted				l/d ⁵ weighted			
	N _{EC}	I _B	R _{EC}	C _{EC}	N _{EC}	I _B	R _{EC}	C _{EC}	N _{EC}	I _B	R _{EC}	C _{EC}
	[-]	[-]	[m ⁻⁴]	[-]	[-]	[-]	[m ⁻⁴]	[-]	[-]	[-]	[m ⁻⁴]	[-]
2	4	1.54	391	0.006	6	1.33	391	0.004	5	1.57	2	0.013
3	10	1.29	338	0.011	10	1.30	511	0.009	7	2.36	2	0.019
4	16	1.04	494	0.028	20	1.22	1510	0.017	10	3.15	3	0.031
5	21	1.33	725	0.029	23	1.33	1853	0.023	18	2.04	48	0.053
6	23	1.37	1437	0.032	29	1.34	2201	0.025	29	2.15	92	0.088
7	26	1.45	2350	0.038	32	1.64	3331	0.028	31	2.51	160	0.096
8	27	1.30	1782	0.048	38	1.48	4550	0.029	35	2.57	247	0.101
9	28	1.37	1738	0.050	39	1.37	4539	0.031	34	2.93	204	0.098
10	32	1.36	1091	0.054	45	1.52	5380	0.034	37	3.10	534	0.099

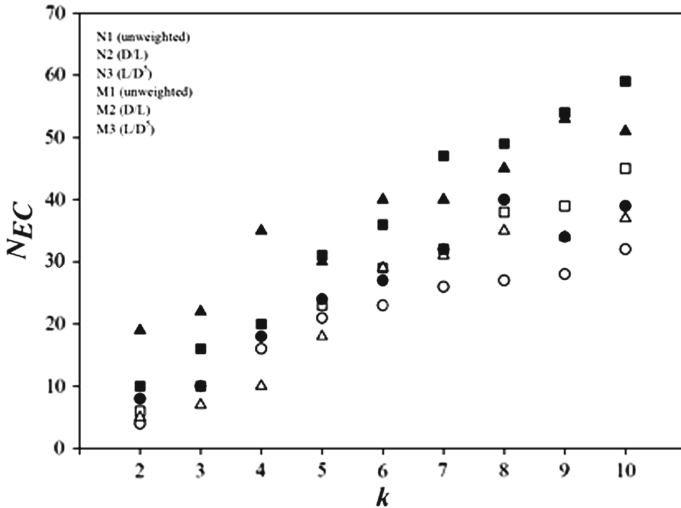


Fig. 1. Number of edge-cuts N_{EC} for *Metis* and *NCut* algorithms, considering the graph un-weighted and weighted with d/l and l/d^5 , and from $k = 2$ to $k = 10$ clusters.

Another important aspect is that the number of edge-cuts N_{EC} , for both *Metis* and *Ncut*, is generally lower if the unweighted graph is considered, except for $k = 3, 4$ and 5 . As the solutions with the smallest values of N_{EC} likely cause lower hydraulic deterioration and they reasonably will have a lower cost for device purchase and installation, they can be considered preferable. Anyway, it may happen that a small boundary set presents a narrow distribution of pipe diameters, so to make difficult the subsequent dividing phase, for which it is better to have both small pipes (to be closed) and large pipes (to be left open) [30].

The trends of the Balance Index I_B for all clustering from $k = 2$ to $k = 10$, considering the graph un-weighted and weighted with d/l and l/d^5 , are shown in Fig. 2.

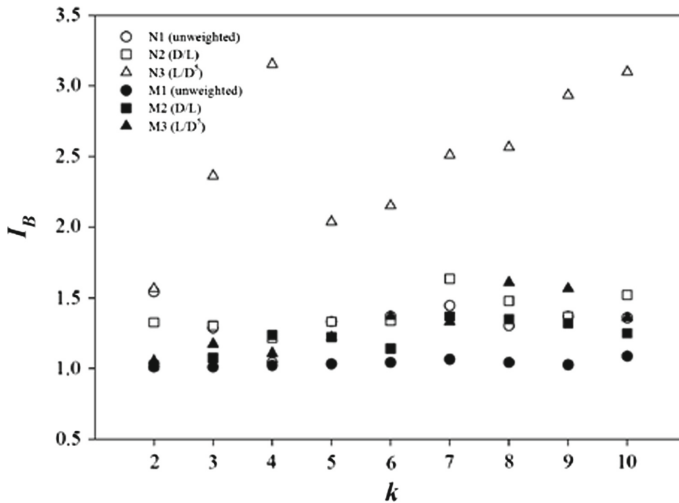


Fig. 2. Balance Index I_B for *Metis* and *Ncut* algorithms, considering the graph un-weighted and weighted with d/l and l/d^5 , and from $k = 2$ to $k = 10$ clusters.

Except for the *Ncut* with l/d^5 -weighted graph, the value of I_B ranges from 1,00 to 1,60, indicating that all the layouts are satisfactorily balanced. As expected, for both algorithms, when the graph is considered un-weighted, the clustering layouts result more balanced. *Metis* provides clustering solutions with lower values of I_B than *Ncut*, so clusters are more balanced. The least balanced layouts come from the *Ncut* when the graph is weighted with l/d^5 , and the difference between the two algorithms reaches the highest value (65%) for $k = 4$. The difference between the results of the two algorithms is clearly due to the fact that in *Metis* I_B is a constraint (it searches layouts in the space of solutions that minimize the number of edge-cuts, or the sum of their weights, in compliance with the constraint on I_B). Instead, *Ncut* provides solutions that simultaneously minimize the number of boundaries (or the sum of their weights) and balance the clusters.

Concerning the index C_{EC} , it is evident from Fig.3 that, for both algorithms, consistently with the minimized objective function, the lowest values are obtained when the graph is considered d/l -weighted. Conversely, the highest values are obtained considering the l/d^5 -weighted graph, as the minimization of the resistance of the boundary leads to select the pipes with greater values of the diameter as edge-cuts.

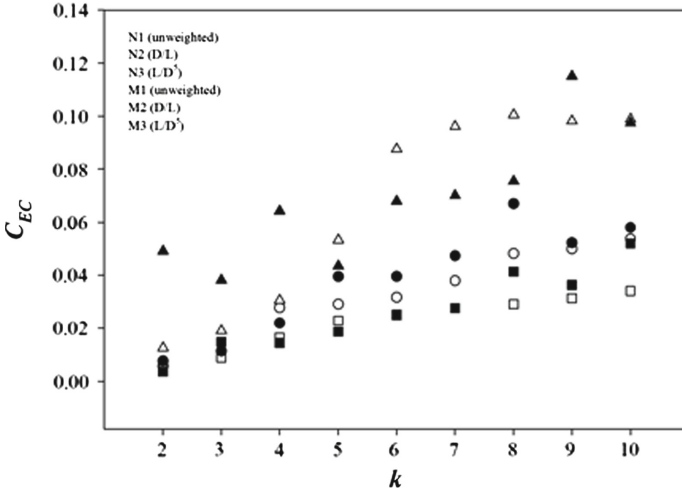


Fig. 3. The index C_{EC} for *Metis* and *NCut* algorithms, considering the graph unweighted and weighted with d/l and l/d^5 , and from $k = 2$ to $k = 10$ clusters.

In Fig. 3, it is apparent that the clustering layouts obtained with *Metis* and *Ncut* without weights show intermediate d/l values between lower and higher values obtained for weighted graphs. The lowest values correspond respectively for *Metis* to $k = 2, 4, 5$ and 6 , and for *NCut* to $k = 3, 7, 8, 9, 10$. Concerning the total resistance R_{EC} , it is evident from Fig. 4 that, as expected, the lowest value is obtained with the weight l/d^5 .

In this case, the difference between the two algorithms reaches the value of 93% for $k = 8$. The highest values are provided by both the algorithms for the d/l -weighted graph, as in this case the pipes with wider diameters are selected. The solutions with the lowest value of resistance are obtained with *NCut*, while the highest values are obtained with *Metis*. Also in this case, the clustering layouts obtained without weights show intermediate values compared to the weighted graphs and increasing trends are clear for *NCut* as a consequence of the trends of N_{EC} .

Finally, in the Fig. 5, the clustering layout of the network of Parete with $k = 5$ is reported, comparing the results obtained with the two clustering algorithms for the un-weighted graph. The Figure shows unequivocally that *Metis* and *Ncut* algorithms provide different clustering of the network, not only in

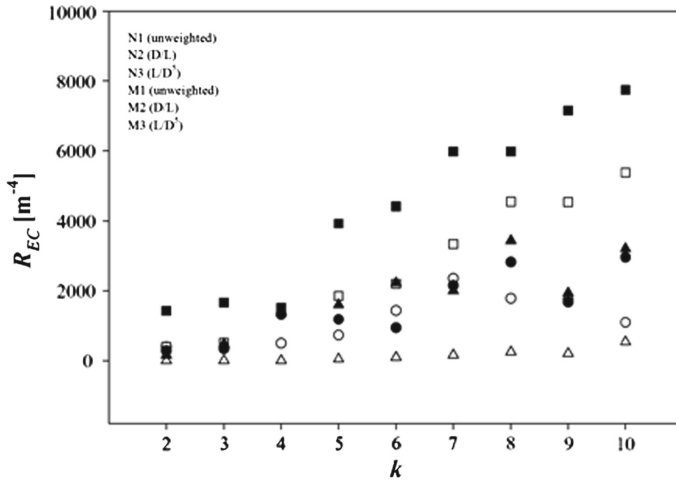


Fig. 4. The index R_{EC} for *Metis* and *NCut* algorithms, considering the graph un-weighted and weighted with d/l and l/d^5 , and from $k = 2$ to $k = 10$ clusters.

terms of indices N_{EC} , I_B , R_{EC} and C_{EC} , but also in terms of shape of each cluster and positioning of boundary pipes.

For the presented case study and taking into account the above described indices, it seems clear that *Ncut* algorithm allows to minimize the number of edge-cuts (both weighted and un-weighted) better than *Metis*, providing solu-

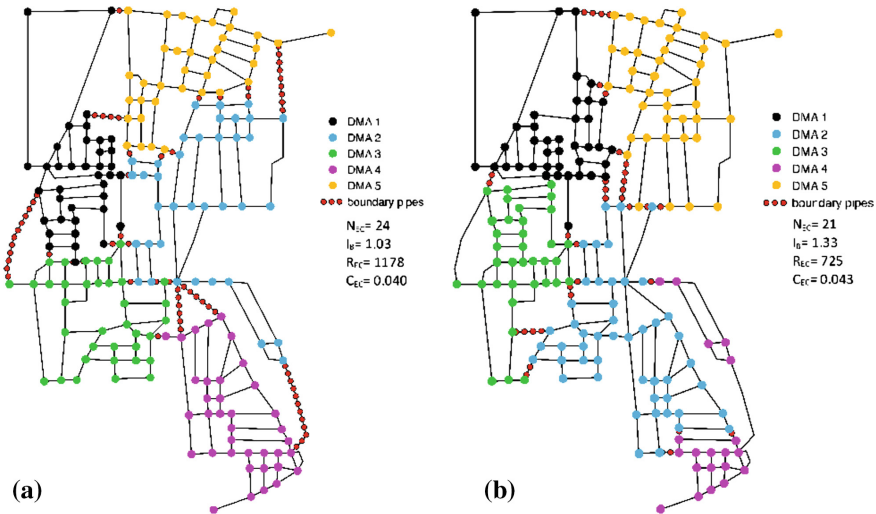


Fig. 5. WDN of Parete divided into $k = 5$ clusters, for un-weighted network graph, with (a) *Metis* clustering, (b) *NCut* clustering

tions with lower resistance and, probably, with lower cost for device purchase and installation (because N_{EC} is significantly smaller). Conversely, *Metis* provides solutions generally more balanced than *Ncut* and, consequently, showing more ability to identify cluster layouts with balanced number of nodes, water demand or other characteristics useful for the aims of WNP.

Anyway, this preliminary study was focused only on the first phase (clustering) of WNP. To fully compare the solutions found by *Metis* and *Ncut*, analyzing the consequences on the hydraulic behavior of the clustered network, also the second phase (dividing) should be studied, in which the locations of gate valves and flow meters are selected, affecting the hydraulic performance of the network.

4 Conclusions

The paper presents the comparison between two clustering algorithms, *Metis* and *NCut*, applied to the problem of water network partitioning. In fact, the first grouping phase is of crucial importance for determining the hydraulic performance of the system, since it establishes the shape and the dimension of the clusters and the number and the typology of the boundary between them. The comparison was developed for the case study of a real WDN, which graph was considered either un-weighted, or weighted with some of the major geometric/hydraulic characteristics of the pipes, aiming at establishing how the weight choice influences the clustering for both the tested algorithms, and which of them provides the optimal grouping layout. To evaluate the simulation results, four indices were used. Simulations, obtained with the two algorithms (*Metis* and *NCut*) and three pipe weights (no-weight, d/l and l/d^5) and for a different number of clusters (from 2 to 10), confirm the effectiveness of both the algorithms in providing good clustering layouts. In particular, *Metis* provides solutions more balanced in terms of number of nodes for each cluster than *NCut*, but the latter provides solutions with smaller edge cut sets, lower infra-cluster resistance and smaller cost, ensuring apparently clustering layouts more convenient from both hydraulic and economic point of view.

Anyway, only the application to other larger networks, by carrying out also the dividing phase, consisting in the insertion of flow meters and gate valves, can confirm these first results. Further studies will compare also other clustering techniques (i.e., Infomap, Louvain, Label propagation, etc.) in order to find the best one for the water distribution networks.

References

1. Gunter, S., Bunke, H.: Validation indices for graph clustering. *Pattern Recogn. Lett. Graph-based Represent. Pattern Recogn.* **24**, 1107–1113 (2003)
2. Fortunato, S.: Community detection in graphs. *Phys. Rep.* **486**, 75–174 (2010)
3. Schaeffer, S.E.: Survey: graph clustering. *Comput. Sci. Rev.* **1**, 27–64 (2007)
4. Dubes, R.: Cluster analysis and related issues. In: Chen, C., Pau, L., Wang, P. (eds.), *Handbook of Pattern Recognition and Computer Vision*, pp. 3–32. World Scientific Publishing Co. Inc. River Edge, NJ, USA, (1993)

5. Huttenhower, C., Flamholz, A.I., Landis, J.N., Sahi, S., Myers, C.L., Olszewski, K.L., Hibbs, M.A., Siemers, N.O., Troyanskaya, O.G., Collier, H.A.: Nearest neighbor networks: clustering expression data based on gene neighborhoods. *BMC Bioinform.* **12**, 8–250 (2007)
6. Ushioda, A., Kawasaki, J.: Hierarchical clustering of words and application to nlp tasks. In: Ejerhed, E., Dagan, I. (eds.) *Fourth Workshop on Very Large Corpora*, pp. 28–41. Association for Computational Linguistics, Somerset, New Jersey (1996)
7. White, S.D.M., Frenk, C.S.: Galaxy formation through hierarchical clustering. *Astrophys. J.* **379**, 52–79 (1991)
8. Wu, Z., Leahy, R.: An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 1101–1113 (1993)
9. Guimerà, R., Mossa, S., Turtschi, A., NunesAmaral, L.A.: The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci. USA* **102**, 7794–7799 (2005)
10. Bar-Ilan, J., Kortsarz, G., Peleg, D.: How to allocate network centers. *J. Algorithm.* **15**, 385–415 (1993)
11. Di Nardo, A., Di Natale, M., Santonastaso, G.F., Tzatchkov, V.G., Alcocer Yamanaka, V.H.: Divide and conquer partitioning techniques for smart water networks. *Proc. Eng.* **89**, 1176–1183 (2014)
12. 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. *Proc. Eng.* **119**, 196–205 (2015)
13. Nascimento, M.C.V., de Carvalho, A.C.P.L.F.: Spectral methods for graph clustering-A survey. *Eur. J. Oper. Res.* **211**, 221–231 (2011)
14. Di Nardo, A., Di Natale, M., Giudicianni, C., Greco, R., Santonastaso, G.F.: Water supply network partitioning based on weighted spectral clustering. *Stud. Comput. Intell. Complex Netw. Appl.* **693**, 797–807 (2016)
15. Boccaletti, S., Latora, V., Morenod, Y., Chavezf, M., Hwanga, D.-U.: Complex networks Structure and dynamics. *Phys. Rep.* **424**, 175–308 (2006)
16. Mays, W.: *Water Distribution Systems Handbook*. McGraw-Hill, New York (2000)
17. Di Nardo A., Di Natale M., Giudicianni C., Greco R., Santonastaso G.F.: Complex network and fractal theory for the assessment of water distribution network resilience to pipe failures. *Water Sci. Technol. Water Suppl* (2017). <https://doi.org/10.2166/ws.2017.124>
18. Alonso, J.M., Alvarruiz, F., Guerrero, D., Hernández, V., Ruiz, P.A., Vidal, A.M., Martínez, F., Vercher, J., Ulanicki, B.: Parallel computing in water network analysis and leakage minimization. *J. Hydr. Eng. ASCE* **126**, 251–260 (2000)
19. Water Industry Research Ltd: *A Manual of DMA Practice*. UK Water Industry Research, London (1999)
20. Di Nardo, A., DiNatale, M., Musmarra, D., Santonastaso, G.F., Tzatchkov, V., Alcocer-Yamanaka, V.H.: Dual-use value of network partitioning for water system management and protection from malicious contamination. *J. Hydroinform.* **17**, 361–376 (2015a)
21. Campbell, E., Izquierdo, J., Montalvo, I., Pérez-García, R.: A novel water supply network securization methodology based on a complete economic analysis, including uncertainties. *Water* **8**, 179 (2016)
22. Di Nardo, A., Di Natale, M., Santonastaso, G.F., Venticinque, S.: An automated tool for smart water network partitioning. *Water Resour. Manage* **27**, 4493–4508 (2013a)

23. Perelman, L.S., Allen, M., Preis, A., Iqbal, M., Whittle, A.J.: Automated sub-zoning of water distribution systems. *Environ. Model Softw.* **65**, 1–14 (2015)
24. 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. *J. Water Sci. Technol. Water Supply* **13**, 951–957 (2013b)
25. Karypis, G., Kumar, V.: Multilevel k-way partitioning scheme for irregular graphs. *J. Parallel Distrib. Comput.* **48**, 96–129 (1998)
26. Karypis, G., Kumar, V.: A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM J. Sci. Comput.* **20**, 359–392 (1999)
27. Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 888–905 (2000)
28. Di Nardo, A., Di Natale, M., Greco, R., Santonastaso, G.F.: Antalgorithm for smart water network partitioning. *Proc. Eng.* **70**, 525–534 (2014)
29. Von Luxburg, U.: A tutorial on spectral clustering. *Stat. Comput.* **17**, 395–416 (2007)
30. Di Nardo, A., Di Natale, M., Giudicianni, C., Greco, R., Santonastaso, G.F.: Weighted spectral clustering for water distribution network partitioning. *Appl. Netw. Sci.* **2**, 19 (2017a)
31. Fiedler, M.: Algebraic connectivity of graphs. *Czech. Math. J.* **23**, 298 (1973)
32. Yazdani, A., Jeffrey, P.: Complex network analysis of water distribution systems. *Interdisc. J. Nonlinear Sci. Chaos* **21**, 016111 (2011)