Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 89 (2014) 1216 – 1223

Procedia Engineering

www.elsevier.com/locate/procedia

16th Conference on Water Distribution System Analysis, WDSA 2014

Comparison of WDN Segmentations Based on Modularity Indexes

F. Boano[a], L. Berardi[b]*

[a]*Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino, Italy*
[b]*Department of Civil Engineering and Architecture, Politecnico di Bari, Italy*

Abstract

Water distribution networks (WDNs) are divided into smaller districts for different technical purposes (e.g., pressure control, leakage management) relying heavily on the technician's experience. A recently developed approach based on modularity metric allow to quantify the degree of uniformity of a particular segmentation including information on network properties (e.g., pipe diameters, flow rates, pressures, leakages) as weights in the matrix representation of the WDN. This study compares different segmentation solutions obtained by maximizing modularity entailing different weights by using the mutual information index coming from the information theory. This study aims at providing insights on modularity formulations to find optimal WDN segmentation for different management purposes.

Peer-review under responsibility of the Organizing Committee of WDSA 2014

*Keywords:*disinfection by-products; drinking water; water distribution network, chlorate;

1. Introduction

Water distribution networks (WDNs) are characterized by the presence of a large number of elements such as pipes, tanks, and pumps. The interactions among these elements are nontrivial and hinder the prediction of the response of the WDN hydraulic behavior to changes in topology (e.g., closure of valves) that are implemented for management purposes. Today, modern management approaches increasingly rely on the division of WDNs in smaller districts that are easier to control and manage [1]. Network segmentation can have different technical aims such as monitoring of water consumptions, pressure reduction, leakage identification and repairing. The decision on how to divide the WDN into segments is usually left to the water utilities and it is often performed relying heavily on the technician's experience. Although field knowledge of the hydraulic functioning of the peculiar system is fundamental to account

---

* Corresponding author.
E-mail address: luigi.berardi@poliba.it

Peer-review under responsibility of the Organizing Committee of WDSA 2014
doi:10.1016/j.proeng.2014.11.423

for the specific constraints of the WDN, there is a large degree of subjectivity, which might impair the effectiveness of segmentation solutions especially for large and growing urban WDNs. Thus, there is an urgent need for tools that can support water providers in the identification of the WDN districts in order to avoid suboptimal and ineffective segmentations [2, 3].

Recent studies have developed novel approaches for segmentation that employ the theory of complex networks, which describes the WDN as a set of interconnected nodes and relies on a compact matrix representation of the WDN topology [4]. In particular, a method has been recently proposed for WDN segmentation [3, 5, 6] that is based on the concept of modularity, a metric that quantifies the degree of uniformity of a particular segmentation [7]. An advantage of the modularity approach is that it allows easily including information on network properties (e.g., pipe lengths, diameters, pressures, leakages) as weights in the matrix representation of the WDN. However, no systematic study on the impact of the inclusion of these weights in the resulting segmentation solutions of the network have been produced so far.

In order to fill this gap, this work presents the comparison of different segmentations obtained by maximizing modularity based on different network properties as matrix weights. In more details, for a given choice of matrix weights, a multi-objective genetic algorithm (MOGA) approach is adopted to maximize modularity against segmentation costs (i.e. number of pressure/flow meters to be installed at segment boundaries) [5]. Segmentation solutions obtained with different matrix weights are then compared using the mutual information metric, an entropy-based index that quantifies the similarity among two segmentations of the same network. The analysis is performed for two benchmark WDNs. By providing a synthetic measure of similarity between couples of district structures, the mutual information enhances the comparison of the usually large number of segmentation solutions determined by the global optimization search (performed by MOGA), and allows to identify the role of different formulations of modularity in assigning WDN pipes to different segments.

2. Methods

Two benchmark WDNs have been analyzed in the present work. Both WDN models are implemented in the WDNetXL 3.0 system (www.hydroinformatics.it), which is also used for segment design runs. The first one is Town-C [8], which is composed of 444 pipes and 396 nodes. Water is supplied to the network by a first pumping station that is connected to the source point in the southeastern part of the WDN. The network includes other four pumping stations that provide the head required by WDN users in different parts of the network. A total of 44 measurement devices for pressure or flow rate are already installed close to pumps, tanks and valves. Pipe diameters range between 51 mm and 610 mm, with minimum and maximum lengths are equal to 4.3 m and 1280 m, respectively. According to a snapshot simulation with base customer demand this WDN is characterized by a strong spatial heterogeneity of pressures, which range between a minimum value of 3 m and a maximum value of 102 m.

The second WDN is Exnet [9], a network that serves approximately 400,000 users and is composed of 1894 nodes and 2467 pipes. Exnet is characterized by a much more complex topology compared to Town-C and it is hence suited to verify the performance of the modularity-based segmentation approach for highly looped WDNs. Two source points provide water to the WDN from its northern section, and a total of 6 observation devices are already installed in the network. Pipe diameters, lengths and pressures range between 50 to 1073 mm, 1 to 2530 m, and 4 to 65 m, respectively.

In this work we have employed the concept of extended modularity [5] to divide the WDN into districts. The extended modularity Q is defined as

$$Q = 1 - \frac{n_c}{n_p} - \sum_{m=1}^{n_m} \left[\sum_{i=1}^{n_p} \frac{w_i \delta\left(M_m, M_i\right)}{W} \right]^2 \tag{1}$$

where n_p is the number of pipes, n_c is the number of cuts (i.e., points of separations between adjacent districts), n_m is the number of modules, w_i (i=1, 2, …n_p) is the weight of the i-th pipe, W is the sum of all pipe weights, M_i is the identifier of the district to which the i-th pipe is assigned, and δ denotes the Kronecker delta function. This definition of extended modularity comes from the observation that WDNs are technological infrastructures and pipes represent the physical elements that are grouped into segments, while common definitions of modularity consider nodes as the relevant elements for segmentation.

A multi-objective approach is adopted to solve the optimization problem, and equation (1) represents the first objective function that is maximized in. Additionally, a second objective function is evaluated to account for the costs required to implement the desired segmentation. Since the divisions of the WDN into segments requires the installation of devices such as valves or flow meters in pipes identified as cut locations, the number of measurement devices (which is related to the number of cuts n_c) is used as a proxy for the total cost of WDN segmentation. Segmentation solutions are developed starting from the existing measurement devices, consistently with the practical approach of using equipment already installed in a real WDN. The segmentation obtained by maximizing the modularity defined by equation (1) depends on the choice of the weights w_i adopted to describe pipe properties. For the sake of brevity, in the present work, the following options have been analyzed:

- $w_i = 1$. This choice leads to the topologic modularity, in which all pipes are treated in the same way regardless of their characteristics;
- $w_i = p_i$, where p_i are average pipe pressures;
- $w_i = p_i L_i$, where the product between pipe pressure p_i and length L_i represents pipe propensity to leakages.

Although using the modularity index is known to be affected by resolution limits [10], the number of districts identifiable for real-sized WDNs is quite large and might put decision makers in a quandary while selecting the management actions (i.e. location of measurement devices to install).

A synthetic metric is used to compare different couples of segmentations and to promptly evidence the level of similarity between them. The chosen metric is the mutual information, which is commonly employed in information theory to determine the amount of shared information between two sets of grouped elements. Mutual information I is defined as [11]

$$I = H(X) - H(H|Y) \tag{2}$$

where X and Y are the two segmentations to be compared, $H(X)$ is the Shannon entropy of X, and $H(X|Y)$ is the entropy of X conditioned on Y. The latter are calculated as

$$H(X) = -\sum_{i=1}^{n_m^X} P_X(i)\log_{10} P_X(i) \tag{3}$$

$$H(H|Y) = -\sum_{i=1}^{n_m^X}\left[\sum_{j=1}^{n_m^Y} P_{X,Y}(i,j)\log_{10}\frac{P_{X,Y}(i,j)}{P_Y(i)}\right] \tag{4}$$

where $P_X(i)$ is the probability of finding a pipe in the i-th district of segmentation X, n_m^X and n_m^Y are the number of segments in X and Y, respectively, and $P_{X,Y}(i,j)$ is the joint probability of a pipe to belong to both the i-th district of segmentation X and to the j-th segment of Y. The normalized mutual information, whose value is bounded between 0 and 1, is finally defined as

$$I_{norm} = \frac{I}{0.5\left[H(X) - H(Y)\right]} \tag{5}$$

and used in this work to compare couples of segmentation solutions.

3. Results

The segmentations obtained for Town-C in the non-weighted (topologic) case are compared to the pressure-weighted case in Fig. 1a, which displays the values of I_{norm} versus the number of measurement devices installed at the segment boundaries. The comparison is made between solutions obtained with different weighting criteria (i.e. w_i) and for the same cost (i.e., number of observation devices). This approach is consistent with the technical purpose of assisting water providers in designing segments (and installing relevant flow/pressure meters) which can be optimal for different management purposes for a given budget. The first point in Fig. 1a corresponds to the case of the 44 observation devices that are already present in the original configuration of Town-C and that divide the WDN into 9 districts. Since these districts are already present before the optimization, all pipes are assigned to the same districts for both the non-weighted and pressure-weighted case, and the normalized mutual information for this starting configuration is obviously unitary because the two segmentations are identical and share exactly the same amount of information.

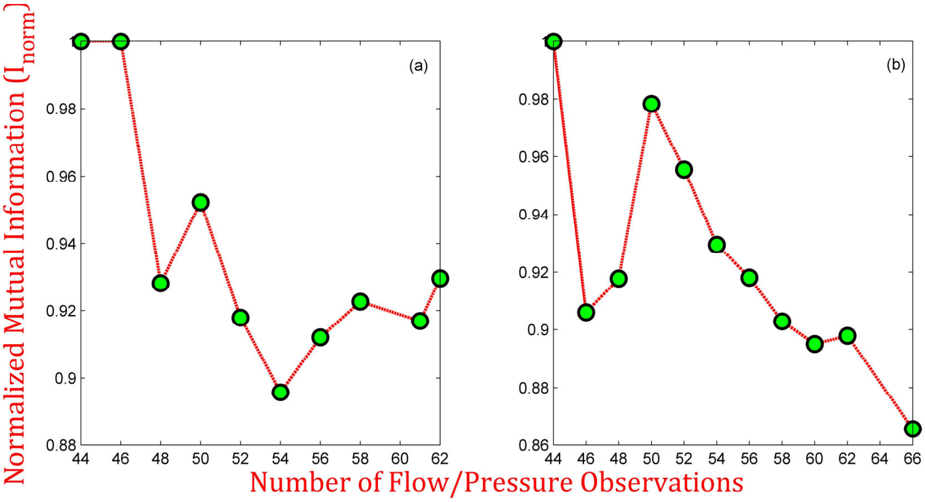

Fig. 1. Normalized mutual information (I_{norm}) between segmentations of Town-C obtained for non-weighted (topologic) case and (a) pressure-weighted case and (b) leakage-propensity-weighted case against number of installed observation devices.

Fig. 1a shows that I_{norm} generally decreases with increasing number of observation points, which seems to suggest that segments tend to be less similar for configurations with a higher number of devices, and hence of segments. However, the values of I_{norm} are always very high (>0.89). These values suggest that use of pipe pressures as weights in equation (1) leads to segmentations that are very similar to those obtained with non-weighted modularity. The I_{norm} curve in Fig. 1a also exhibits a minimum at 54 observation devices, which corresponds to 14 segments. This point denotes two solutions that are relatively more different than those obtained with similar number of observation points, although the solutions are still quite similar ($I_{norm} > 0.89$). The actual configurations portrayed in Fig. 2 (non-weighted case) Fig. 3a (pressure-weighted case) and Fig. 3b (leakage-propensity-weighted case). The comparison shows that when modularity is weighted with pipe pressures (i.e. segments contain pipes with similar average pressure), the two segments in the southern part of the WDN (Fig. 2, yellow and green) are grouped into a larger district (Fig. 3a, yellow), while the two small northwestern segments (Fig. 2, blue and green) are grouped into a larger one (Fig. 3a, blue) because differences in pipe pressures among these districts were limited. Additionally, some pipes in one of the grouped segment (Fig. 2, yellow) are detached to form a new segment (Fig. 3a, red). Despite these differences, the other segments in the WDN are mostly equal in both Figures, which explains the large value of I_{norm} and implies that pipe pressures provide little additional information compared to those contained in the topological structure of the WDN. The use of leakage propensity as pipe weight (Fig. 3b) also leads to a segmentation ($I_{norm} > 0.89$) that is generally similar to the non-weighted case (Fig. 2), with some minor differences in the shape of some districts due to the influence of the chosen weight.

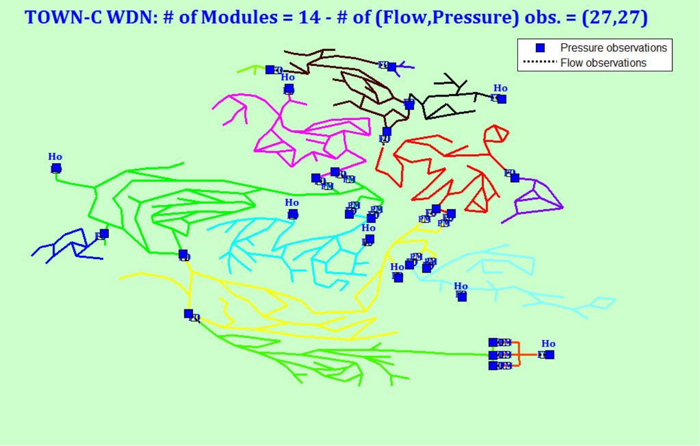

Fig. 2. Segmentation of Town-C for non-weighted (topologic) case. The configuration includes 14 segments delimited by 54 observation points.

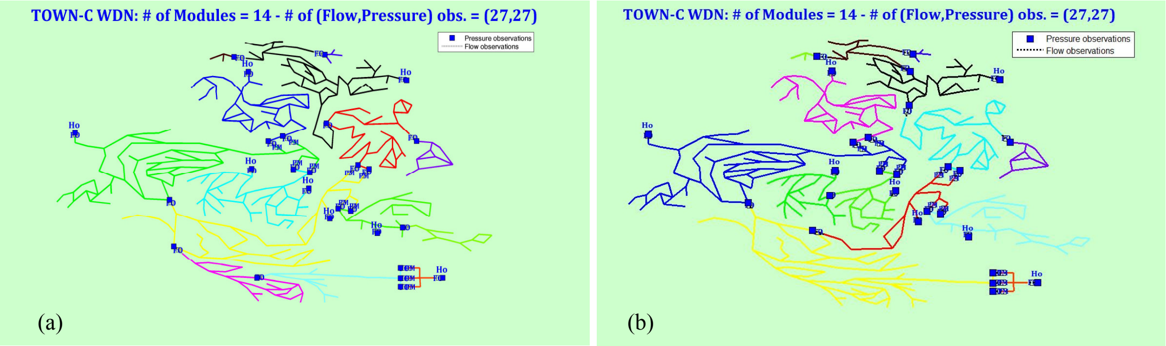

Fig. 3. Segmentation of Town-C for (a) pressure-weighted case and (b) leakage-propensity weighted case. The configuration includes 14 segments delimited by 54 observation points.

The effect of the application of leakage propensity p_iL_i as pipe weight is shown in Fig. 1b, which displays the values of mutual information between the topologic case and the weighted case. Again, the high values of dimensionless mutual information ($I_{norm} > 0.86$) imply that the use of this weight has a limited impact on the resulting configuration of segments. These values further confirm that topology is the WDN feature that exerts the strongest control on the segmentation of Town-C.

The same kind of comparison has also been performed for Exnet WDN, and the resulting normalized mutual information values are shown in Fig. 4. The two curves compare the behavior of I_{norm} between non-weighted modularity and pressure-weighted and leakage-propensity-weighted modularities, respectively. In both cases, the values of I_{norm} exhibit strong fluctuations between high and low values when the number of observation devices, and hence of districts, is low. Since these districts are relatively large, some of them include a high number of pipes with heterogeneous properties (e.g., pipe pressure), and when these properties are used as pipe weights the optimization of equation (1) can assign them to different districts to maximize the value of weighted modularity.

As the number of installed segmentation devices is increased, the curves of normalized mutual information in Fig. 4 progressively shift toward a more stable behavior with relatively high value ($I_{norm} > 0.7$). This behavior is caused by the progressive increase in the number of segments which become internally more homogenous in terms of pipe properties. As a result, the choice of different weights tends to have a more limited impact on the segmentation that is mainly dominated by WDN topology, which strongly drives the segmentation according to specific pipe properties (i.e. average pressure). An example of these segmentations is shown in Fig. 5, which displays the district configurations obtained with 35 observation devices using non-weighted (Fig. 5a) and pressure-weighted (Fig. 6b) modularities, respectively. The comparison between the Fig.s evidence some differences, such as for the large southern district (Fig. 5a, red) whose pipes are partly reassigned to other segments when the pressure-weighted modularity is employed (Fig. 5b, yellow and blue). The spatial distribution of pressure in the WDN, reported in Fig. 6, shows how the reassignment of these pipes results in districts are much more homogeneous in terms of pressure than those obtained with non-weighted modularity. However, the remaining districts are almost identical for both segmentations, coherently with the high value of $I_{norm} = 0.8$. The similarity among segments determined using different pipe weights suggests that even in complex looped WDNs such as Exnet the same segmentation may be applied for different technical aims (e.g. pressure control and leakage identification) provided that the value of I_{norm} is sufficiently high and that districts are not excessively large.

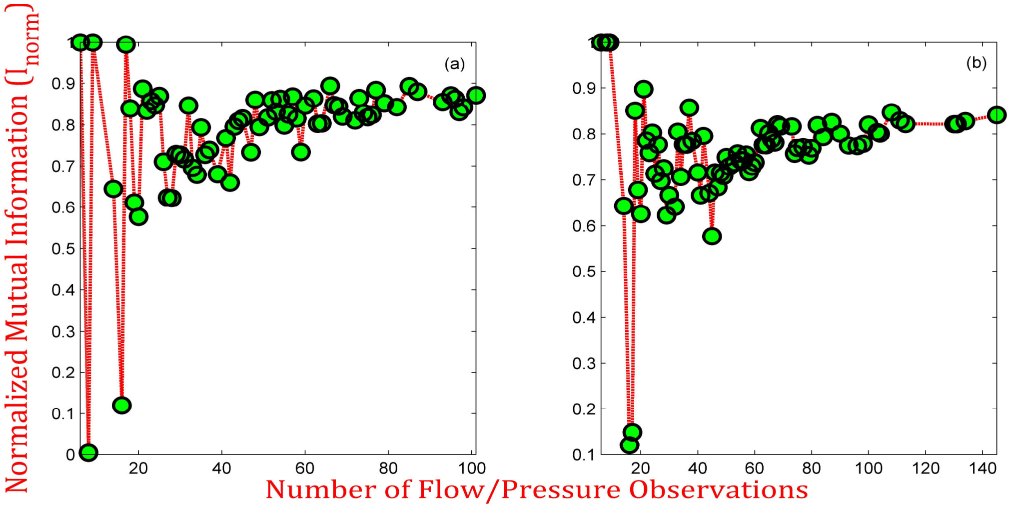

Fig. 4. Normalized mutual information (I_{norm}) between segmentations of Exnet obtained for non-weighted (topologic) case and (a) pressure-weighted case and (b) leakage-propensity-weighted case against number of installed observation devices.

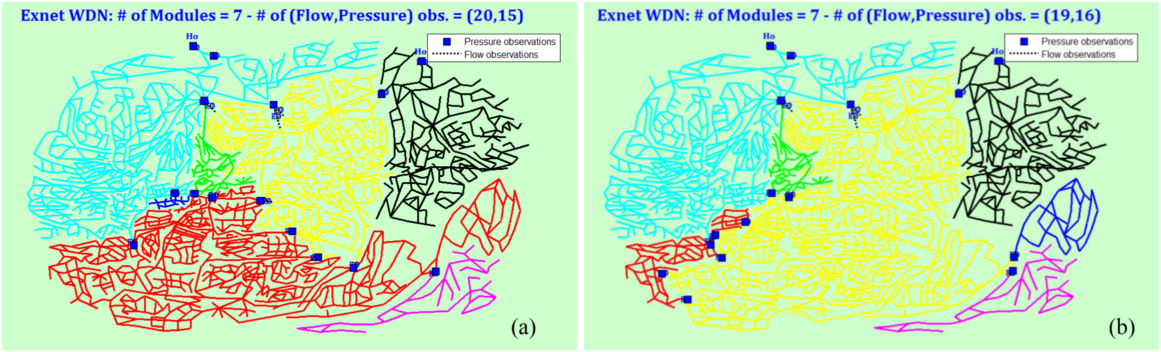

Fig. 5.Segmentation of Exnet for (a) non-weighted (topologic) case and (b) pressure-weighted case. The configuration includes 7 segments delimited by 35 observation points.

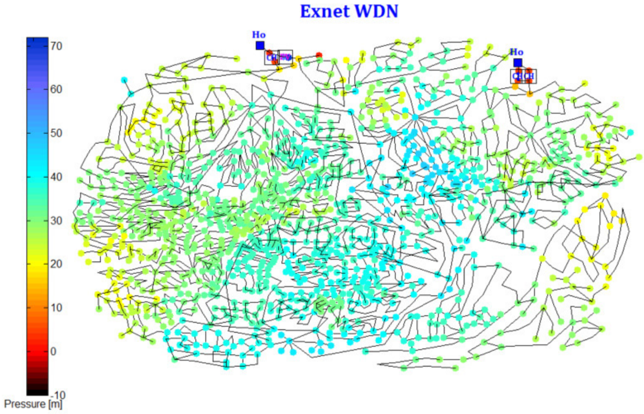

Fig. 6.Spatial distribution of pressures for snapshot simulation of Exnet WDN (returned by WDNetXL system).

4. Conclusions

The present work has analyzed the application of a modularity-based approach for WDN segmentation, and it has compared how the identified segment structures are influenced by the use of different pipe properties as weights. The analysis has demonstrated that the application of a weighted modularity leads to different segmentation solutions depending on the characteristics of the network. Results have shown that the segmentation of relatively simple WDNs (e.g. Town-C) is mainly governed by WDN topology. In these networks, topology strongly controls the spatial distribution of pipe properties and the application of different weights of the extended modularity index lead to identification of very similar districts. This, in turns, would suggest water providers that many flow/pressure meters may be effective for different operational/management purposes.

In contrast, segmentation of complex highly-looped WDNs (e.g. Exnet) depends also on the number of districts to be created and hence on the budget available for the installation of segmentation devices. When these WDNs are divided into a small number of districts (i.e., with a low segmentation budget) the resulting segmentation is strongly dependent on the chosen pipe weights because pipes within the same district are characterized by remarkable variations in properties such as average pressures or background leakages. As the budget available for segmentation increases and the WDN is divided into a higher number of districts by installing a larger number of devices, the similarity among segments identified according different criteria (as measured in terms of modularity index) tends to reach a constant value. This, again, confirm that network topology is dominant regardless of the choice of pipe weights as the segmentation are more refined (i.e. requiring higher investment costs). From operational/management perspective this means that, solutions with sufficiently high number of districts permits to achieve network segmentation that are more effective and can be used for various technical purposes (e.g., pressure reduction, flow and pressure monitoring, leakage control). Such behavior suggests the opportunity of designing WDN segments of this kind of networks by using the abovementioned multi-objective optimization where costs and extended modularity index are simultaneously optimized. In fact, the initial optimization run can point out segmentation devices shared by solutions that have been obtained assuming different weighting criteria and that are hence suited for multiple technical purposes. Successive segment design runs will assume such devices as already installed in the network, while others will entail progressively refines configurations suited for peculiar technical purposes. After that, the analysis of similarities might provide again indication of the most similar solutions for similar investments required. From such perspective, the segment design approach based on extended modularity can be seen as a practical tool for supporting decision makers in a step-by-step procedure for deciding devices to be progressively installed.

Although WDN topology is always expected to be a first-order control on network hydraulics, the present study needs to be expanded in the future by considering a wider array of pipe weights. Moreover, the study has focused on the use of mutual information to quantify the similarities among segmentations obtained with different weighting criteria. In the future, the application of other metrics may provide deeper insights on the strengths and limitations of WDN segmentations determined with different criteria.

Acknowledgements

The research reported in this paper was founded by the Italian Scientific Research Program of National Interest PRIN-2012 “Tools and procedure for advanced and sustainable management of water distribution networks”.

References

[1] L.Perelman, A. Ostfeld, Topological clustering for water distribution systems analysis, Environ. Modell. Software, 26 (2011), 969–972.
[2] K. Diao, Y. Zhou, W. Rauch (2013), Automated creation of district metered area boundaries in water distribution systems, J. Water Res.Plann. Manage., 139 (2013) 184–190.
[3] M. Scibetta, F. Boano, R. Revelli, L. Ridolfi, Community detection as a tool for complex pipe network clustering, EPL - Europhysics Letters, 103 (2013) 48001.
[4] S. Fortunato, Community detection in graphs, Phys. Rep., 486 (2010) 75–174.
[5] O. Giustolisi, L. Ridolfi, New Modularity-Based Approach to Segmentation of Water Distribution Networks, J. Hydraul. Eng., 140 (2014) 04014049.
[6]O. Giustolisi, L. Ridolfi, A novel infrastructure modularity index for the segmentation of water distribution networks. Water Resources Research, 50(10) (2014), 7648–7661.
[7] M. E. J. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E, 69 (2004), 026113.
[8] A. Ostfeld et al., The battle of the water calibration networks (BWCN), J. Water Res. Plann. Manage., 138 (2012) 523–532.
[9] R. Farmani, G.A. Walters, D.A. Savic, Trade-off between total cost and reliability for Anytown water distribution network, J. Water Res. Plan. Manage., 131 (2005) 161-171.
[10] S. Fortunato, M. Barthelemy, Resolution limit in community detection, Proc., Natl. Acad. Sci. U.S.A., 104 (2007), 36–41.
[11] T. M. Cover, J. A. Thomas, Elements of Information Theory, New York, Wiley, 1991.