Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 186 (2017) 559 – 566

Procedia Engineering

www.elsevier.com/locate/procedia

XVIII International Conference on Water Distribution Systems Analysis, WDSA2016

A proposal of optimal sampling design using infrastructure modularity

Antonietta Simone, Luigi Berardi, Daniele Laucelli, Orazio Giustolisi*

Politecnico di Bari, Dept. of Civil Engineering and Architecture via Orabona n.4, 70125, Bari, Italy

Abstract

Planning of pressure observations/meters in terms of spatial distribution and number is named sampling design. In the past, the hydraulic model calibration was the main driver of the sampling design. Today, water utilities are interested in system pressure monitoring for hydraulic system analysis and management with respect to other technical purposes as for example detection of anomalies (burst leakages and anomaly head losses) and service quality with respect to customers. In recent years, the optimal location of flow observations/meters, related to design of optimal district metering areas, has been faced considering optimal network segmentation and the modularity index using a multi-objective strategy. The original modularity index from the studies of the complex network theory was transformed to be WDN-oriented. Consistently, this paper proposes a new way to perform the sampling design using newly developed sampling-oriented modularity-based metrics. The strategy optimizes the location of the pressure meters based on network topology creating pressure district metering areas, i.e. it returns the optimal location of the nodal pressure meters defining "pressure DMAs". The multi-objective optimization problem minimizes the cost of newly installed meters while maximizing the sampling-oriented modularity metrics. The battle of background leakages assessment water network (BBLAWN) allows presenting and discussing the proposed sampling design methodology.

Peer-review under responsibility of the organizing committee of the XVIII International Conference on Water Distribution Systems

Keywords: Water distribution network; Sampling design; Modularity index; Segmentation; Multi-objective optimization

---

* Corresponding author. Tel.: +39-080-5963726; fax: +39-080-5963719.
E-mail address: orazio.giustolisi@poliba.it

Peer-review under responsibility of the organizing committee of the XVIII International Conference on Water Distribution Systems
doi:10.1016/j.proeng.2017.03.270

1. Introduction

The analysis and management of large water distribution networks (WDNs) has always presented many problems related with non-homogeneous behavior and the growing amount of field information (i.e. streams of data), that are available nowadays by ICT solutions, are increasing the complexity of WDNs analysis. Over time, several solutions have been proposed to manage this problem, and the most commonly used and studied is the community detection, i.e. the division of the network into smaller modules segmenting the hydraulic system [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11].

The community detection strategy [12] [13] allows the WDNs (infrastructure networks) division into modules (also named segmentation or partitioning) defining a number of smaller portions (named districts, segments or modules) bounded by installed devices, suited for different technical purposes including monitoring of background leakages, detection of anomalies (e.g. due to pipe bursts), etc. [14].

In complex network theory, several metrics have been proposed for network segmentation each of them showing various advantages and drawbacks. The modularity index [13] [15] is the most widely accepted and used metric to measure the propensity of the network division into modules (actually named communities, consistently with the earliest applications of the index). The modularity is a descriptive measure of topology and relies strictly on the network structure. The advantage of the modularity is the fact that it can be computed using only the adjacency matrix of the network, without requiring other information. For a given network, a higher value of the maximum modularity index indicates a better identification of communities; therefore, the maximum value of the modularity corresponds to the maximum degree of segmentation.

Several researchers [9] [10] used the maximization of modularity index for segmentation of WDNs, i.e., infrastructure networks, although its original formulation [16] was proposed for immaterial networks. For this reason, Giustolisi and Ridolfi [17] tailored the original modularity index in order to obtain a WDN-oriented modularity index.

They considered the “conceptual cuts” segmenting the network close to nodes instead of the middle of pipes, introducing the pipe weights and defining a different modularity index aimed at dividing the network in modules having an internal similarity of an assumed pipe attribute. For instance, a modified modularity index allows the division into modules having internal similar attribute such as diameters or elevations, as opposite to the original formulation dividing the network in modules, which are similar to each other with respect to a specific weight. Afterwards, Giustolisi and Ridolfi [18] proposed the infrastructure modularity index to overcome the resolution limit of both original [19] and WDN-oriented modularity.

The segmentation problem [17] [18] was solved using a specific multi-objective evolutionary optimization strategy, based on the use of genetic algorithms (MOGA). The approach proved to be effective because WDNs are not large size networks (if compared with other typical immaterial complex networks) and the division into modules needs to be performed just few times in WDN service life. In addition, the MOGA strategy allows searching for the optimal trade-off between the minimization of the segmentation cost versus its effectiveness (i.e., the maximization of value of the WDN-oriented or infrastructure modularity indexes), which is the Pareto front of solutions.

Starting from the work on segmentation of Giustolisi and Ridolfi [17] a novel sampling-oriented modularity indices and a MOGA optimal sampling design is here proposed. The original WDN-oriented modularity [17] is modified to consider the positioning of nodal pressure meters as “conceptual removal of nodes” instead of “conceptual cuts” of the original segmentation procedure. Consequently, the optimization strategy finds the best trade-offs between sampling-oriented modularity index and cost of the newly installed devices, minimizing the number of conceptual removal nodes (i.e. pressure meters) while maximizing the sampling-oriented modularity metric. This strategy aims at providing conceptual scenarios of network division, which are the basis for the design of actual pressure metering areas. In the case of the segmentation by nodal removal, the nodes that are conceptually removed are the candidate position for pressure meters.

The originality of such approach relates to the fact that the “conceptual removal of nodes” divides the networks into “pressure DMAs”. This way, each “pressure DMA” results bounded by a sub-set of nodal pressure meters, which guarantees the information about pressure status at the boundary of each pressure district of the network. In other words, “pressure DMAs” are connected each other by nodes where pressure devices are installed, which are the boundary conditions for internal pressures, similarly to flow measurements of classic DMAs for all demand components. The methodology is presented and discussed using the battle of background leakages assessment water network (BBLAWN).

2. Brief on WDN-oriented Modularity based metrics

The modularity index measures the strength of a network to be divided into communities [15]. Newman and Girvan [20] first proposed the network segmentation through the maximization of the modularity index, i.e., the identification of networks communities (modules).

Afterwards Giustolisi and Ridolfi [17] tailored the formulation of the modularity index for infrastructural networks (WDNs), obtaining the WDN-oriented modularity index formulation, reported below:

$$Q(\mathbf{w}_p) = \left\{ 1 - \frac{n_c}{n_p} \right\} + \left\{ -\sum_{m=1}^{n_m} \left[\sum_{k=1}^{n_p} \frac{(\mathbf{w}_p)_k \, \delta(\mathrm{M}_m, \mathrm{M}_k)}{W} \right]^2 \right\} = Q_1 + Q_2 \tag{1}$$

where n_c is the number of pipes linking modules of the infrastructure, namely the number of “conceptual cuts” in the network (i.e. the decision variables of the WDN segmentation problem); n_p is the number of pipes and n_m is the number of network modules. The summation inside the square brackets is related to pipe weights stored in the vector $\mathbf{w}_p$, whose sum is W, and Kronecker's δ function makes that the sum refers only to the weights of pipes belonging to the m-th module (i.e. $\delta = 1$ if $\mathrm{M}_m = \mathrm{M}_k$ and $\delta = 0$ otherwise).

It is worth to note that the term Q_1 of Eq. (1) decreases with the number of cuts, while Q_2 generally increases with the number of modules. For a given number of modules, Q_2 increases with the similarity of modules to each other with respect to the assigned weights.

Both original [20] and WDN-oriented modularity indexes present a resolution limit [19], increasing with network size, which prevents to further increasing of the metric value when Q_1 starts dominating Q_2. In other words, a number of modules exists for which a further single cut decreases Q_1 more than any optimal identification of a further module, i.e. smaller modules cannot be identified.

Giustolisi and Ridolfi [18] analyzed the resolution limit of the WDN-oriented modularity index and proposed a new infrastructure modularity index to overcome such limit:

$$IQ(\mathbf{w}_p) = 1 - \frac{n_c - (n_{act} - 1)}{n_p} - \sum_{m=1}^{n_m} \left[\sum_{k=1}^{n_p} \frac{(\mathbf{w}_p)_k \, \delta(\mathrm{M}_m, \mathrm{M}_k)}{W} \right]^2 \tag{2}$$

where n_{act} is the actual number of modules satisfying given constraints (e.g. the minimum length of the modules, the minimum number of pipes, etc.). Accordingly, the same authors demonstrated that the infrastructure modularity resolves the resolution limit, but might require the definition of technical constraints to avoid a resolution of the segmentation beyond the required by specific technical tasks.

Giustolisi and Ridolfi [17] also proposed a WDN-oriented metric, called attribute–based modularity index, measuring the similarity into each module with respect to a specified attribute, which is not length-based:

$$Q_a = 1 - \frac{n_c}{n_p} - \sum_{m=1}^{n_m} \left[\frac{\sum_{k=1}^{n_p} \left| \mathbf{a}_p - \bar{a}(\mathrm{M}_m) \right|_k \delta(\mathrm{M}_m, \mathrm{M}_k)}{\sum_{k=1}^{n_p} \left| \mathbf{a}_p - \bar{a}(\mathrm{N}) \right|_k} \right]^2 \tag{3}$$

where $\mathbf{a}_p$ is the vector of pipe attributes, $\bar{a}(\mathrm{N})$ is the mean value of the pipe attributes of the network N, i.e. of $\mathbf{a}_p$, and $\bar{a}(\mathrm{M}_\mathrm{m})$ is the mean value of the pipe attributes in M_m. Function δ limits the summation of the pipe attributes to the elements belonging to the same module.

Giustolisi et al. [12] extended the infrastructure modularity index to attribute–based infrastructure modularity index:

$$IQ_a = 1 - \frac{n_c - (n_{act} - 1)}{n_p} - \sum_{m=1}^{n_m} \left[\frac{\sum_{k=1}^{n_p} \left| \mathbf{a}_p - \bar{a}(\mathrm{M}_m) \right|_k \delta(\mathrm{M}_m, \mathrm{M}_k)}{\sum_{k=1}^{n_p} \left| \mathbf{a}_p - \bar{a}(\mathrm{N}) \right|_k} \right]^2 \tag{4}$$

Eq. (4) solves the resolution limit that might occur also for the attribute-based metric.

3. Tailoring the WDN-oriented Modularity to Sampling Design

This paper proposes a novel optimal sampling design approach, based on a multi-objective optimization including novel metrics tailored for WDN sampling design and inspired by both WDN-oriented modularity and infrastructure modularity of the segmentation design. The proposed metrics for sampling design is named sampling-oriented modularity index.

The WDN-oriented and sampling-oriented modularity metrics differ for the approach of identifying modules in WDNs. The first segments the network considering pipes, i.e., by means of "conceptual cuts", the second considering nodes, i.e., by means of "conceptual removal of nodes".

Therefore, the sampling-oriented modularity can be seen as a dual way for the segmentation, allowing using the connectivity matrix of the edges/pipes [21] instead of the adjacency matrix.

Once the modules related to "conceptual removal of the nodes" are identified, the WDN-oriented modularity formulations in Eqs. (1-4) are extended to achieve the sampling-oriented modularity due to the similar conceptual basis:

$$\begin{aligned}
Q_s(\mathbf{w}_p) &= 1 - \frac{n_{obs}}{n_n} - \sum_{m=1}^{n_m} \left[\sum_{k=1}^{n_p} \frac{(\mathbf{w}_p)_k \, \delta(\mathrm{M}_m, \mathrm{M}_k)}{W} \right]^2 \\
IQ_s(\mathbf{w}_p) &= 1 - \frac{n_{obs} - (n_{act} - 1)}{n_n} - \sum_{m=1}^{n_m} \left[\sum_{k=1}^{n_p} \frac{(\mathbf{w}_p)_k \, \delta(\mathrm{M}_m, \mathrm{M}_k)}{W} \right]^2 \\
Q_{a-s}(\mathbf{a}_p) &= 1 - \frac{n_{obs}}{n_n} - \sum_{m=1}^{n_m} \left[\frac{\sum_{k=1}^{n_p} \left| \mathbf{a}_p - \bar{a}(\mathrm{M}_m) \right|_k \delta(\mathrm{M}_m, \mathrm{M}_k)}{\sum_{k=1}^{n_p} \left| \mathbf{a}_p - \bar{a}(\mathrm{N}) \right|_k} \right]^2 \\
IQ_{a-s}(\mathbf{a}_p) &= 1 - \frac{n_{obs} - (n_{act} - 1)}{n_n} - \sum_{m=1}^{n_m} \left[\frac{\sum_{k=1}^{n_p} \left| \mathbf{a}_p - \bar{a}(\mathrm{M}_m) \right|_k \delta(\mathrm{M}_m, \mathrm{M}_k)}{\sum_{k=1}^{n_p} \left| \mathbf{a}_p - \bar{a}(\mathrm{N}) \right|_k} \right]^2
\end{aligned} \tag{5}$$

where the number of removed nodes n_{obs} (where pressure gauges will be installed) replaces the number of pipes connecting modules of the network, i.e., the number of "conceptual cuts" n_c of Eqs. (1-4) and the number of nodes n_n. replaces the number of pipes n_p. The number of modules n_m, identified by removing nodes, has the same meaning of the WDN-oriented formulation, as well as the actual number of modules matching the technical constraints to avoid excessive resolution of the segmentation, n_{act}, of the "infrastructure" version. Note that the term Q_2 of Eqs. (5) is unchanged with respect to Eqs. (1-4) because it refers to the characteristics of modules, which are now generated by removed nodes.

Therefore, the Eqs. (5) define respectively the sampling-oriented modularity Q_s, sampling-oriented infrastructure modularity IQ_s, sampling-oriented attribute–based modularity $Q_{a\text{-}s}$ and sampling-oriented infrastructure attribute–based modularity $IQ_{a\text{-}s}$.

It is worth to note that Eqs. (5) have the same properties of Eqs. (1-4) with respect to modules, i.e., the first two of Eqs. (5) divide the network in modules similar to each other with respect to the assumed pipe weights, while the last two of Eqs. (5) divide the network in modules having similar internal characteristics with respect to the assumed pipe weights.

4. Multi-objective strategy for Optimal Sampling Design using sampling-oriented modularity

The problem of optimal sampling design, which is related to the definition of "pressure DMAs", is here solved as multi-objective optimization using the novel sampling-oriented modularity metrics.

Therefore, the two-objective optimization problem can be formulated as follows:

$$\begin{cases} [M, n_{obs}, n_{act}] = connectivity\left(I_c, \mathbf{L}\right) \\ f_1 = \max\left\{IQ\left(\mathbf{w}_p\right)\right\} = \max\left\{1 - \dfrac{n_{obs} - \left(n_{act} - 1\right)}{n_n} - \sum_{m=1}^{n_m}\left[\sum_{k=1}^{n_p}\dfrac{\left(\mathbf{w}_p\right)_k \delta\left(\mathrm{M}_m, \mathrm{M}_k\right)}{W}\right]^2\right\} \\ f_2 = \min\limits_{I_c}\left\{n_{obs}\right\} \end{cases} \tag{6}$$

where $\mathbf{L}$ is the edge/pipe adjacency matrix, I_c is the set of n_obs conceptually removed nodes in the network, the decision variables, corresponding to the new pressure meters to be installed [17] [18] in order to obtain "pressure DMAs", and connectivity($\mathbf{L}$, I_c) stands for component analysis of the graph with respect to edge, as reported in previous section. Note that being the decision variables related to new pressure meters to be installed, the already existing pressure measurements, corresponding to control valves, pumps, tank levels and reservoir levels, can be considered as priors.

The optimization problem of Eq. (6) assumes as f_1 the sampling-oriented infrastructure modularity, although any of sampling-oriented metrics of Eqs. (5) can be used.

The infrastructure modularity metrics allow identifying very small modules because the resolution limit is solved. Nonetheless, this poses some problems when the resolution is too high for supporting WDN management tasks. For this reason, n_{act} is here defined considering the following constraint C_1:

$$C_1 = \frac{\sum_{k=1}^{n_p}\left(\mathbf{L}_p\right)_k \delta\left(\mathrm{M}_m, \mathrm{M}_k\right)}{L} \geq x\% \tag{7}$$

where L is the sum of pipes lengths; the numerator is the module length (i.e., sum of length of module pipes $\mathbf{L}_p$) and x% is a threshold percentage to be assumed. C_1 is then a sort of "pressure" for the optimization to search for solutions characterized by modules having length larger than x% of the total network length L.

The sampling design is a WDN management activity that needs to be performed few times during the lifetime of the hydraulic system; therefore, it is effective to solve the problem of Eq. (6) using a MOGA optimization strategy, as above specified. In fact, MOGA optimization strategies are efficient and flexible for combinatorial multi-objective problems, able to ensure a good sub-optimal Pareto set of solutions. Furthermore, the multi-objective approach provides a Pareto set of solutions which is a decision making support for water utilities, e.g., considering the available budget for pressure meters.

5. BBLAWN Case Study

The WDN named Battle of Background Leakage Assessment for Water Networks (BBLAWN) is here used to show and discuss the main features of the proposed sampling design strategy. The network was designed as a competition held at the 16th Water Distribution Systems Analysis Conference, in Bari (Italy) in July 2014. The BBLAWN is composed of 445 pipes (9 closed), 396 nodes, 11 controlled pumps (PM), 1 controlled valves (CV), 1 check valve (CH), 25 pressure control valves (PCV), 7 variable level tanks and 1 reservoir.

The proposed optimal sampling design was performed considering as priors 19 pressure meters corresponding to 11 pumps, 7 tanks and 1 reservoir. The first two *sampling-oriented modularity indices* of Eqs. (5), Q_s and IQ_s, were used to perform the MOGA optimal design according to Eq. (6) using for f_1 the specific index using length as pipe weights $\mathbf{w}_p$. In the case of the infrastructure index (IQ_s), C_1 of Eq. (7) was set equal to 0.5% of total pipe length of the hydraulic system. Figure 2 shows the two Pareto fronts returned by the optimal sampling design using the *classic* and the *infrastructure* modularity index approaches.

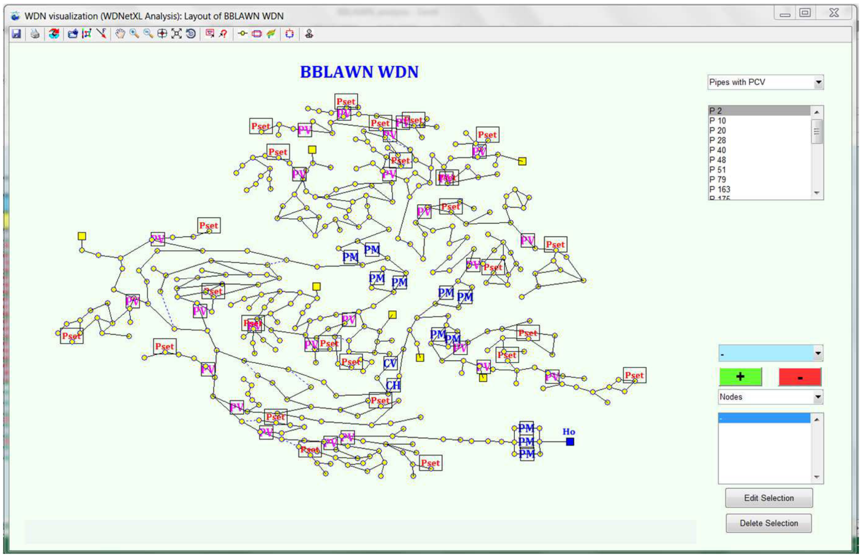

Fig. 1. Layout of the BBLAWN.

Both the Pareto fronts starts from 20 pressure meters because the 19 pressure meters corresponding to devices, as previously said, are assume existing and used as constraints to the optimization. Then, the classic form of the modularity index (Q_s) returns nine solutions while the infrastructure form of the modularity index (IQ_s) twenty-eight solutions. In fact, the constraint related to the existing devices exacerbate the resolution limit problem [19] because they are not optimally placed into the network [18].

Therefore, the infrastructure version of the *sampling-oriented modularity index* is much more effective of the classic version. In fact, the most extended Pareto front contains twenty-eight solution that can be evaluated considering the available budget.

Figure 3 reports some of those solutions of the *sampling-oriented infrastructure modularity index* corresponding to {9, 15, 21, 28} added pressure meters showing layouts of "pressure DMAs" (colored) with different resolutions. Furthermore, the optimization process can be driven to achieve solutions, which are nested. This way the uncertainty about the available budget over time results solved because of the flexibility achieved by the nesting characteristic of the solutions.

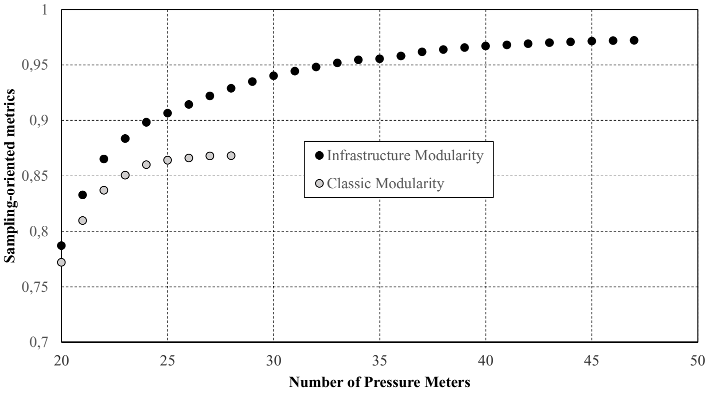

Fig. 2. Pareto Fronts returned by MOGA sampling design.

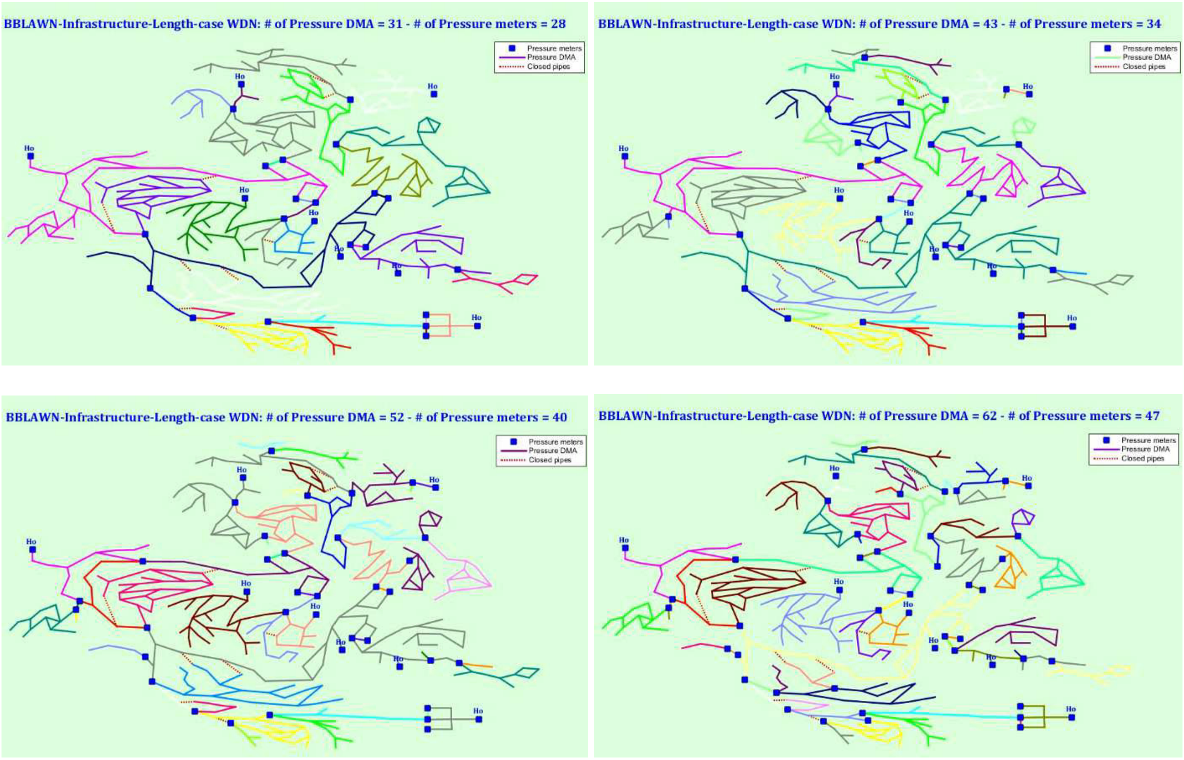

Fig. 3. Solution corresponding to {9, 15, 21, 28} added pressure meters.

6. Conclusions

Sampling design is of key importance for various WDN analysis and management tasks by means of analysis and monitoring of pressure status. Therefore, the present work proposes a novel methodology for designing a pressure sampling system, based on WDN topological analysis. The pressure sampling design is formulated as a multi-objective optimization problem, where a new segmentation metric, suited for sampling design, is maximized while minimizing the number of "conceptual removed nodes".

The multi-objective strategy provides optimal solutions, which are "conceptual scenarios" of pressure gauges locations into the network. Each conceptual scenario is related to a different "pressure DMAs" configuration, i.e. DMAs bounded by a sub-set of nodal pressure gauges. Moreover, the proposed sampling design strategy is based on WDN topology and asset characteristics, which are generally available information.

Acknowledgements

The Italian Ministry of Education, University and Research (MIUR) has supported this research, under the Projects of Relevant National Interest "Advanced analysis tools for the management of water losses in urban aqueducts" and "Tools and procedures for an advanced and sustainable management of water distribution systems". This research was also partly funded by the project "Methodologies and tools for the sustainable management of urban water distribution networks in the Mediterranean area" funded by the "Future in Research" program – Apulian Region (Italy)

References

[1] Jacobs P., Goulter I., Evaluation of methods for decomposition of water distribution networks for reliability analysis. Civil Engineering Systems, 5 (1988) 58–64.
[2] Yang S.-L., Hsu N.-S., Loule P.W.F., Yeh, W.W.-G., Water distribution network reliability: connectivity analysis. J. of Infrastructure Systems, 2 (1996) 54-64.
[3] Walski T. M., Technique for calibrating network models, J. Water Res. Plan. and Manage., 109 (1983) 360–372.
[4] Davidson J., Bouchart F., Cavill S., Jowitt, P., Real-time connectivity modelling of water distribution networks to predict contamination spread, J. of Computing in Civil Engineering, 19 (2005) 377–386.
[5] Deuerlein J.W., Decomposition model of a general water supply network graph, J. Hydr. Eng., 134 (2008) 822–832.
[6] Perelman, L., Ostfeld A., Topological clustering for water distribution systems analysis. Environ. Modell. & Software, 26 (2011) 969–972.
[7] Alvisi S., Franchini M., A heuristic procedure for the automatic creation of district metered areas in water distribution systems, Urban Water J., 11 (2014) 137–159.
[8] Yazdani A., Jeffrey P., Water distribution system vulnerability analysis using weighted and directed network models, Water Resour. Res., 48 (2012).
[9] Scibetta M., Boano F., Revelli R., Ridolfi L., Community detection as a tool for complex pipe network clustering. EPL, 103 (2013) 48001.
[10] Diao K., Zhou Y., Rauch, W., Automated Creation of District Metered Area Boundaries in Water Distribution Systems, J. Water Res. Plan. and Manage., 139 (2013) 184–190.
[11] Di Nardo A., Di Natale M, Giudicianni C., Musmarra D., Santonastaso G.F., Simone A., Water distribution system clustering and partitioning based on social network algorithms Procedia Engineering, 119 (2015) 196–205.
[12] Albert R., Barabasi A.L., Statistical mechanics of complex networks, Rev. Mod. Phys., 74 (2002) 47–97.
[13] Newman M.E.J., Networks: An introduction, Oxford Univ. Press, Oxford (UK), 2010.
[14] Giustolisi O., Ridolfi, L., Berardi, L., General metrics for segmenting infrastructure networks, Journal of Hydroinformatics, 17 (2015) 505–517.
[15] Newman M.E.J., Fast algorithm for detecting community, structure in networks, E 69, 066133, (2004) doi: 10.1103/PhysRevE.69.066133.
[16] Barthélemy M., Spatial networks, Phys. Rep., 499 (2011) 101.
[17] Giustolisi O., Ridolfi L., A new modularity-based approach to segmentation of water distribution networks. J. Hydr. Eng., 140 (2014), 04014049-(1-14).
[18] Giustolisi O., Ridolfi L., A novel infrastructure modularity index for the segmentation of water distribution networks Water Resources Research, 50 (2014), 7648–7661.
[19] Fortunato S., Barthélemy M., Resolution limit in community detection, Proc. Nat. Acad. Sciences of the USA, 104 (2007) 36–41.
[20] Newman M.E.J., Girvan M., Finding and evaluating community structure in networks. Phys. Rev. E, 69 (2004) 026113.
[21] Brualdi, R.A., Ryser, H.J., Combinatorial matrix theory. Cambridge University Press, 1991.
[22] Giustolisi, O., Berardi, L., Laucelli, D., Savic, D.A., Kapelan, Z., Operational and Tactical Management of Water and Energy Resources in Pressurized Systems: Competition at WDSA 2014, J. Water Res. Plan. and Manage. 142(5), C4015001 (2016).