AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2016WR018944

A proposal of optimal sampling design using a modularity strategy

A. Simone[1], O. Giustolisi[1], and D. B. Laucelli[1]

[1]Department of Civil Engineering and Architecture, Technical University of Bari, Bari, Italy

Key Points:
- Performing the optimal location of pressure meters using the sampling-oriented modularity index
- An optimization strategy minimizes pressure meters cost maximizing the sampling-oriented modularity
- The methodology is discussed using the Apulian and Exnet networks

Correspondence to:
A. Simone,
antonietta.simone@poliba.it

Citation:
Simone, A., O. Giustolisi, and D. B. Laucelli (2016), A proposal of optimal sampling design using a modularity strategy, *Water Resour. Res.*, *52*, doi:10.1002/2016WR018944.

Received 17 MAR 2016
Accepted 14 JUL 2016
Accepted article online 22 JUL 2016

Abstract In real water distribution networks (WDNs) are present thousands nodes and optimal placement of pressure and flow observations is a relevant issue for different management tasks. The planning of *pressure observations* in terms of spatial distribution and number is named *sampling design* and it was faced considering model calibration. Nowadays, the design of system monitoring is a relevant issue for water utilities e.g., in order to manage background leakages, to detect anomalies and bursts, to guarantee service quality, etc. In recent years, the optimal location of *flow observations* related to design of optimal *district metering areas* (DMAs) and leakage management purposes has been faced considering optimal network segmentation and the modularity index using a multiobjective strategy. Optimal network segmentation is the basis to identify network modules by means of optimal conceptual cuts, which are the candidate locations of closed gates or flow meters creating the DMAs. Starting from the WDN-oriented modularity index, as a metric for WDN segmentation, this paper proposes a new way to perform the *sampling design*, i.e., the optimal location of *pressure meters*, using newly developed *sampling-oriented* modularity index. The strategy optimizes the pressure monitoring system mainly based on network topology and weights assigned to pipes according to the specific technical tasks. A multiobjective optimization minimizes the cost of pressure meters while maximizing the *sampling-oriented* modularity index. The methodology is presented and discussed using the Apulian and Exnet networks.

1. Introduction

Infrastructural, information and technological systems can be described as complex networks, defined by nodes and arcs to label their topology. The analysis and management of large networks has always presented many problems related with nonhomogeneous behaviour of the network. Over time, several solutions have been proposed to manage this problem, and the most commonly used and studied is the community detection strategy, i.e., the division of the network into smaller modules.

Complex network theory (CNT) classify water distribution networks (WDNs) as infrastructures networks. Community detection strategies of CNT allow dividing WDNs in modules/segments, which can facilitate the hydraulic system analysis and management. In fact, the application of the segmentation (also named partitioning) to WDNs divides the system in a number of smaller portions (named districts, segments or modules) bounded by installed devices (flow meters and closed valves), allowing to define district metering areas (DMAs), which are useful for different technical purposes: demand and background leakages management, burst detections, rehabilitation works, model calibration, etc.

The effectiveness of a WDN segmentation is then a relevant issue for water utilities to ensure adequate service for customers, increasing the benefits of the planned investments [*Heskett*, 1986]. Moreover, the advances of information communication technology (ICT) in the water sector are motivating water utilities and researchers to exploit the available information for the WDN management purposes, ranging from the conceptual segmentation of networks to the actual DMAs through the installation of devices.

Therefore, over time, the research community proposed several approaches to segmentation [*Jacobs and Goulter*, 1988; *Yang et al.*, 1996; *Walski*, 1983; *Davidson et al.*, 2005; *Deuerlein*, 2008; *Perelman and Ostfeld*, 2011; *Alvisi and Franchini*, 2014; *Yazdani and Jeffrey*, 2012; *Scibetta et al.*, 2013; *Diao et al.*, 2013; *Albert and Barabasi*, 2002; *Newman*, 2004, 2006a *Di Nardo et al.*, 2015], i.e., algorithms and metrics to identify the optimal division of the network with respect to topology and WDN characteristics, such as pipes length and diameter, nodal elevations, leakages, etc.

CNT studies proposed several metrics for network segmentation, each of them showing various advantages and drawbacks [*Steinhaeuser and Chawla*, 2010]. The modularity index [*Newman*, 2004] 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 [*Steinhaeuser and Chawla*, 2010]. 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. Many heuristic approaches were introduced to find the maximum value of the modularity in order to identify the community structure [*Chen et al*., 2014]. These approaches include greedy algorithms [*Newman*, 2004, 2006a], spectral methods [*Newman*, 2006b], extremal optimization [*Duch and Arenas*, 2005], simulated annealing [*Medus et al*., 2005], sampling technique [*Sales-Pardo et al*., 2007], but also nonheuristic approaches such as mathematical programming was proposed [*Agarwal and Kempe*, 2008].

The greedy algorithms allow detecting the community structure of very large networks in a reasonable time [*Newman*, 2004; *Newman and Girvan*, 2004; *Newman*, 2006a; *Clauset et al*., 2004; *Wakita and Tsurumi*, 2007; *Blondel et al*., 2008]. They solve the community detection problem using heuristic rules such as divisive (i.e., dividing the network in modules based on inter-community links), and agglomerative (i.e., merging similar nodes/communities recursively). The application of the heuristic rules at each stage is a locally optimal choice, which does not guarantee a global optimal solution.

Several researchers [e.g., *Scibetta et al*., 2013; *Diao et al*., 2013, *Fortunato*, 2010] used the maximization of modularity index for segmentation of WDNs, i.e., infrastructure networks, although its original formulation [*Barthélemy*, 2011] was proposed for immaterial networks. For this reason, *Giustolisi and Ridolfi* [2014a] tailored the original modularity index in order to obtain a WDN-oriented modularity index. They (i) considered the "conceptual cuts" segmenting the network close to nodes instead of the middle of pipes; (ii) introduced the pipe weights in the WDN-oriented modularity index and (iii) defined a different modularity index aimed at dividing the network in modules having an internal similarity of an assumed pipe attribute. The modified modularity index allows the division into module having internal similar attribute such as diameters or elevations, as opposite to the original formulation dividing the network in modules that are similar to each other with respect to a specific weight.

Afterward, *Giustolisi and Ridolfi* [2014b] proposed the *infrastructure modularity* index to overcome the resolution limit of the original modularity [*Fortunato and Barthélemy*, 2007] inherited by the WDN-oriented modularity. The segmentation problem [*Giustolisi and Ridolfi*, 2014a,b] was solved using a specific multiobjective 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 networks studied in complex network theory) 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 useful as decision support system in order to select the best solution according, for example, to the available budget for devices to be installed. In fact, the modularity index represents ultimately a driver to divide a WDN into modules by means of "conceptual cuts," while, in DMA design problem, such cuts are the candidate locations of closed gates or flow measurement devices. Nonetheless, system observations include also pressure measurements and the design of the spatial distribution and number of pressure gauges is named *sampling design*.

Several approaches to sampling design exists in technical-scientific literature, mostly driven by the need for model calibration. For instance, *Walski* [1983] proposed the location of pressure measurements near the high-demand locations and on the perimeter of the network. *Bhave* [1988] started from the number of available pressure sensors and divided the network into the corresponding number of portions. Others works based the analysis upon simulation methods, highlighting the importance of the sensitivity matrix and the coefficient of roughness in the network, in order to ensure a good WDN model calibration [*Shamir and Howard*, 1968; *Rahal et al*., 1980; *Gofman and Rodeh*, 1981; *Ormsbee and Wood*, 1986; *Boulos and Wood*, 1991]. About multiobjective optimization methods for sampling design, *Carrera et al*. [1984] proposed a method for selecting optimal locations from a discrete set of possible measurement points, based on nonlinear programming. *Yu and Powell* [1994] proposed a method maximizing the estimated accuracy and minimizing the total metering cost. *Piller et al*. [1999] formulated the sampling design problem using a greedy algorithm, minimizing the influence of measurement errors in the state vector estimation with the constraint that the Jacobian matrix should be of maximum rank. *De Schaetzen et al*. [2000] suggested three new sampling design approaches, the first two are based on the shortest path algorithm and rank potential measurement locations, and the third solves the optimization problem based on maximization of Shannon's entropy. *Kapelan et al*. [2003, 2005] formulated a two-objective optimization problem, maximizing the calibrated model accuracy versus the minimization of the relevant uncertainties and of the total cost. *Behzadian et al*. [2008] formulated a multiobjective optimization problem under calibration parameter uncertainty, maximizing the calibrated model accuracy versus the minimization of the number of sampling devices as a surrogate of sampling design cost.

This work proposes a novel multiobjective optimal sampling design method based on a topological analysis, which is not mainly intended for model calibration, but it can be integrated with the model calibration strategies in an effective way as reported in the text. The novelty of the strategy stems from using a new metric obtained by tailoring the WDN-oriented modularity index and infrastructure modularity index for sampling design, i.e., *sampling-oriented* modularity index and the consequent division of the WDN in modules bounded by pressure meters, named "pressure DMAs." For clarity, the WDN-oriented modularity and infrastructure modularity will be here named *segmentation-oriented* modularity.

To the purpose of defining *sampling-oriented* modularity, the *segmentation-oriented* modularity is modified to consider the positioning of nodal pressure measurements introducing the concept of "conceptual removal of nodes" which is consistent with the idea of "conceptual cuts" of the classic segmentation procedure. Consequently, the optimization strategy finds the best trade-offs between sampling-oriented modularity and cost of new devices to be installed.

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 gauges. This fact guarantees the information about pressure status at the boundary nodes of each pressure district of the network. In other words, "pressure DMAs" are connected each other by nodes where pressure devices are installed. The pressure readings on those nodes, which are the boundary conditions for internal pressures, are similar to flow measurements of classic DMAs for internal demand components.

The proposed approach, based on topological analysis of the network and, possibly, on pipe weights used in the modularity strategy [*Giustolisi and Ridolfi*, 2014a], is reliable with respect to the uncertainty of the hydraulic behavior of the WDN. In fact, the WDN topology and the pipe weights generally relate to the asset characteristics and are easy available information. Furthermore, the network topology has a relevant influence on the hydraulic behavior of the system, although other hydraulic objectives (e.g., related to accuracy of the calibration or burst detection) could be integrated into the proposed approach as additional drivers for sampling design.

The novel approach allows positioning other pressure gauges at *central* node of each "pressure DMA" in order to increase the information collected by the pressure monitoring system. The central node within the various modules is that with the maximum value of the *betweenness centrality* [*Freeman*, 1977] of the specific "pressure DMA." The use of betweenness centrality allows the selection of nodes which are characterized by the maximum number of shortest paths (possibly assuming pipe weights for the specific technical task) passing through it. This makes the selected nodes relevant with respect to the hydraulic behavior of the WDN.

Finally, the proposed approach using a modularity-based strategy of dividing the WDN in "pressure DMAs" is consistent with the recent approaches [*Giustolisi and Ridolfi*, 2014a] to design DMAs allowing for mass balance control. This fact hints an easy and effective integration of the two design strategies for the location of flow and pressure observations, i.e., the integrated planning of the monitoring system.

A small size test case, the Apulian network, will allow presenting the newly developed sampling-oriented modularity metrics for sampling design and clarifying the concept of "pressure DMA." A real medium size test case, the Exnet network, will allow testing and discussing the proposed strategy for sampling design.

2. Brief on Segmentation-Oriented Modularity-Based Metrics

Newman and Girvan [2004] firstly proposed the network segmentation through the maximization of the modularity index, i.e., the identification of networks communities (modules).

As mentioned in the introduction, the original modularity index is conceived for immaterial networks [*Barthélemy*, 2011], and the use of its formulation in WDNs is not advisable because they are infrastructure networks [*Giustolisi and Ridolfi*, 2014a]. For this reason, the original formulation of the modularity index was tailored for WDNs using the general topological incidence matrix, commonly adopted in hydraulic modeling, and developing a cut position-sensitive metric [*Giustolisi and Ridolfi*, 2014a].

The formulation of the WDN-oriented (here named segmentation-oriented) modularity index is:

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

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) 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 equation (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 and WDN-oriented modularity indexes present a resolution limit [*Fortunato and Barthélemy*, 2007], 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 [2014b] analyzed the resolution limit of the segmentation-oriented modularity 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 [2014a] also proposed a segmentation-oriented metric, called *attribute*–based segmentation index, measuring the similarity into each module with respect to a specified *attribute*, which is not length-based:

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

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 $\bar{a}(\mathrm{M_m})$ is the mean value of the pipe attributes in $\mathrm{M_m}$. Function δ limits the summation of the pipe attributes to the elements belonging to the same module.

Giustolisi et al. [2015] extended the infrastructure modularity index to *attribute–based infrastructure segmentation* index:

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

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

3. Tailoring the Segmentation-Oriented Modularity to Sampling Design

This paper proposes a novel optimal *sampling design* approach, based on a multiobjective optimization including novel metrics tailored for WDN sampling design and inspired by the *segmentation-oriented* modularity (class comprising the attribute-oriented and the infrastructure modularity solving the resolution limit) of the *segmentation design*. The proposed metrics for sampling design is named *sampling-oriented* modularity.

The *segmentation-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 **L** [*Brualdi and Ryser*, 1991] instead of the adjacency matrix **A** of equation (1).

The identification of the modules, for a given set of nodal observations, can be achieved constructing the matrix **L** from the arc-node incidence matrix reduced to junction nodes $\mathbf{A}_{pn}$ (also known in hydraulic modeling as general topological matrix of the network) by nulling the values in the columns corresponding to observation nodes and performing the matrix product $\mathbf{L}=\mathbf{A}_{pn}$ x $\mathbf{A}_{np}$. The connectivity analysis of **L** allows the identification of the modules.

Once the modules related to "conceptual removal of the nodes" are identified, the *segmentation-oriented* modularity formulations in equations (1–4) are extended to achieve the *sampling-oriented* modularity indices due to the similar conceptual basis:

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

where the number of pipes connecting modules of the network, i.e., the number of "conceptual cuts" n_c in equations (1–4), is replaced by the number of removed nodes n_{obs} (where pressure gauges will be installed) and the number of pipes n_p is replaced by the number of nodes n_n. The number of modules n_m, identified by removing nodes, has the same meaning of the segmentation-oriented formulations, 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 equation (5) is unchanged with respect to equations (1–4) because it refers to the characteristics of modules, which are now created by removed nodes.

Therefore, the equation (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 equation (5) have the same properties of equations (1–4) with respect to modules, i.e., the first two of equation (5) divide the network in modules similar to each other with respect to the assumed pipe weights, while the last two of equation (5) divide the network in modules having similar internal characteristics with respect to the assumed pipe weights.

4. Technical Perspective of Pressure DMAs

The proposed optimal sampling design recalls classic segmentation because of the partition of the hydraulic network in districts, in the first case having pressure meters at the boundary nodes of "pressure DMAs" in the latter case flow meters or closed gate valves at the boundary pipes of "classic DMAs."

From a technical point of view, the *pressure DMAs* are conceived to allow checking the network pressure status by partitioning the hydraulic system, while *classic DMAs* are conceived to allow the same for mass balance. They refer to the two aspects of the hydraulics of the networks, head losses (momentum equations) and pipe flows (mass balance equations).

However, it is worth noting that the implementation of the pressure DMAs does not involve the modification of the network topology, i.e., the hydraulic status of the WDN, while the implementation of classic DMAs alters the network topology because of the installation of closed gate valves, i.e., the hydraulic status of the WDN results modified. Consequently, the number of classic DMAs implemented in a WDN is generally lower than the number of pressure DMAs for two main reasons: the cost of flow meters is greater than pressure meters and the initial hydraulic capacity with respect to service requirements of the WDN influences the possibility of installing closed gate valves.

Therefore, the integration of pressure DMAs with classic ones is not a trivial task because it depends on many characteristics of the WDN such as network size, initial hydraulic capacity with respect to service requirements, level of background leakages, available budget, etc. Furthermore, the simple installation of pressure meters in the nodes close to the "conceptual cuts" of the segmentation is a starting simplification, which does not solve some technical aspects: (i) the positioning of pressure meters is not optimal from a topological perspective; (ii) closed gate valves locally disconnect the network and consequently local pressure meters are not observations at the boundary of districts; (iii) for large size WDNs, it is possible to conceive pressure DMAs internal to larger classic DMAs; and (iv) sometimes in medium/small size WDNs, classic districtualization is not planned, while a network of pressure meters is.

However, independently on previous notes, the proposed optimal sampling can have practical and technical effectiveness, since it is related to the network topology, and thus conceived independently on the hydraulic of the WDN, whilst a useful integration for the model calibration procedures in medium and large size WDNs as will be reported later. In fact, network of pressure meters, placed into the network based on topology using the paradigm of complex network theory, should be effective for zooning the detection of burst leakages and other anomalies, because the pressure meters are rationally located at the boundary among pressure DMAs. Furthermore, the pressure status checking using such a network of sensors is easy to plan because based on the network topology, which is the first information available for a WDN.

Finally, about the possibility of integrating the pressure DMA concept with the classic model calibration procedures for medium and large size WDNs, the starting point is the first equation of the global gradient algorithm [*Todini and Pilati*, 1988; *Giustolisi et al.*, 2008]

$$\begin{gathered}\mathbf{A}_{nn}\mathbf{H}_n=\mathbf{F}_n \\ \mathbf{A}_{nn}=\left[\mathbf{A}_{np}\left(\mathbf{D}_{pp}\right)^{-1}\mathbf{A}_{pn}\right] \\ \mathbf{F}_n=\mathbf{A}_{np}\left(\mathbf{Q}_p-\frac{\mathbf{Q}_p}{n}\right)-\mathbf{d}_n-\mathbf{A}_{np}\left(\mathbf{D}_{pp}\right)^{-1}\left(\mathbf{A}_{p0}\mathbf{H}_0\right)\end{gathered} \tag{6}$$

where $\mathbf{Q}_p$ is the $[n_p,1]$ column vector of pipe flows, $\mathbf{H}_n$ is the $[n_n,1]$ column vector of unknown nodal heads, $\mathbf{H}_0$ is the $[n_0,1]$ column vector of known nodal heads, $\mathbf{d}_n$ is the $[n_n,1]$ column vector of nodal demands, $\mathbf{A}_{pn}=\mathbf{A}_{np}$ and $\mathbf{A}_{p0}$ are topological incidence sub-matrices of size $[n_p,n_n]$ and $[n_p,n_0]$, respectively, and $\mathbf{D}_{pp}$ is a diagonal matrix whose elements are the derivatives of the head loss function (having exponent n) with respect to $\mathbf{Q}_p$.

The matrix $\mathbf{A}_{nn}$ is the Jacobian reduced to nodes, i.e., the sensitivity matrix of the WDN model. It is the same matrix that *Sanz and Pérez* [2015] adopted for a demand calibration methodology using the singular value decomposition in order to better condition the inverse problem. They named $\mathbf{A}_{nn}$ complete sensitivity matrix and wrote the following equations linking the variation Δ of the hydraulic variables ($\mathbf{H}_n$ and $\mathbf{Q}_p$) to the variation of the nodal demands ($\mathbf{d}_n$),

$$\begin{gathered}\Delta\mathbf{H}_n=\left(\mathbf{A}_{nn}\right)^{-1}\Delta\mathbf{d}_n \\ \Delta\mathbf{Q}_p=\left(\mathbf{D}_{pp}\right)^{-1}\mathbf{A}_{pn}\left(\mathbf{A}_{nn}\right)^{-1}\Delta\mathbf{d}_n\end{gathered} \tag{7}$$

It is to note that the condition number of $\mathbf{A}_{nn}$ increases with network size, i.e., the inverse problem related to calibration becomes worse and worse conditioned. The readings at the pressure meters in the boundary positions of pressure DMAs, plus the nodal elevations, can be considered as known heads. From a hydraulic modeling point of view, this means to move those nodal heads in the column of known terms and resize the incident matrix ($\mathbf{A}_{pn}$) removing the columns related to the nodal pressure observations. Assuming x nodal observations dividing the network in y districts (pressure DMAs), it can be stated that the number of columns of the incident matrix is reduced by x. Furthermore, the incidence matrix results divided into y components, each one corresponding to a pressure DMA. Consequently, the Jacobian matrix $\mathbf{A}_{nn}$ has a reduced order by x and can be divided into y components. This way, the model calibration can be faced considering the sub-models having a lower size, which makes better conditioned the inverse problem.

Therefore, the proposed sampling design strategy is connected with model calibration of large-medium size WDNs, for which the network can be divided in "calibration districts," i.e., pressure DMAs. Then, it is possible to apply the classic optimal sampling designs for calibration to each districts.

5. Multiobjective 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 multiobjective optimization using the novel *sampling-oriented* modularity metrics.

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

$$\begin{cases} [\mathrm{M}, n_{obs}, n_{act}] = connectivity(I_c, \mathbf{L}) \\ f_1 = \max\left\{IQ(\mathbf{w}_p)\right\} = \max\left\{1 - \dfrac{n_{obs} - (n_{act} - 1)}{n_n} - \displaystyle\sum_{m=1}^{n_m}\left[\sum_{k=1}^{n_p}\frac{(\mathbf{w}_p)_k\,\delta(\mathrm{M}_m, \mathrm{M}_k)}{W}\right]^2\right\} \\ f_2 = \min\limits_{I_c}\left\{n_{obs}\right\} \end{cases} \tag{8}$$

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 [*Giustolisi and Ridolfi*, 2014a, 2015b] 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, could be considered as constraints.

The optimization problem of equation (8) assumes as f_1 the *sampling-oriented infrastructure modularity*, although any of *sampling-oriented metrics* of equation (5) can be used. Therefore, the strategy optimizes the pressure monitoring system mainly based on network topology, also considering weights $\mathbf{w}_\mathrm{p}$ assigned to pipes according to the specific technical tasks. For example, pipe lengths are general purpose weights, because they significantly influence the hydraulic behavior having a technical meaning also with respect to background leakages [*Giustolisi et al.*, 2008] of the system and probability of pipe bursts [*Berardi et al.*, 2008]. Consequently, the use of pipe length could be effective for sampling design related to the burst detection task, but also model calibration as reported in the previous section.

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:

$$C_1 = \frac{\sum_{k=1}^{n_p}(\mathbf{L}_p)_k\,\delta(\mathrm{M}_m, \mathrm{M}_k)}{L} \geq x\% \tag{9}$$

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 equation (8) using a MOGA optimization strategy, as above specified. In fact, MOGA optimization strategies are efficient and flexible for combinatorial multiobjective problems, able to ensure a good sub-optimal Pareto set of solutions. Furthermore, the multiobjective 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.

6. From Conceptual Division into Modules to Classic and Pressure DMAs

The previous two sections introduced *sampling-oriented* modularity based metrics and a multiobjective optimization in order to divide the network in modules with respect to nodes similarly to *segmentation-oriented* aiming at dividing the network with cuts close to nodes. The similarity of the two approaches allowed tailoring the *segmentation-oriented* metrics to *sampling-oriented* metrics.

It is worth to note that the optimal "conceptual cuts" close to ending nodes of pipes is conceptually, computationally and practically different from "conceptual removal" of nodes generating "pressure DMAs." In fact, for optimal sampling design (which excludes terminal nodes) each pressure observation node, i.e., "conceptual node removal," can be at the boundary of two or more modules/segments, while for optimal segmentation design (which excludes terminal pipes) each pipe cut, i.e., "conceptual cuts," always is at the boundary of two modules/segments. Obviously, nodes having high centrality are more candidate to divide in modules with respect to cuts close to those nodes.

In both cases, the segmentation is conceptual, i.e., the network is conceived as immaterial. In fact, the "conceptual removing of nodes" identifies the modules and the *sampling-oriented* modularity metrics allow the optimal division in modules. In the same way the "conceptual cuts" identify the modules and the *segmentation-oriented* modularity metrics allow the optimal division in modules [*Giustolisi and Ridolfi*, 2014a,b].

It is important to clarify that the multiobjective optimization, minimizing the number of "conceptual cuts" or "conceptual removed nodes" versus the maximization of the *segmentation-oriented* or *sampling-oriented* metrics, aims at providing "conceptual scenarios" of network division, which are the basis for the design of actual metering areas.

In the case of segmentation by cuts, these are the candidate positions (of the assumed "conceptual scenarios") for closed valves or flow meters. Therefore, a classic DMA design process consists in (i) selecting one segmentation scenario and (ii) deciding which pipes to close, as paths for water transmission, or where to locate the flow measurements in order to achieve the control of the mass balance internal to the specific DMA. Consequently, there is a (i) conceptual and a (ii) practical step, moving from selected segmentation scenario to DMAs. The steps involve considerations on hydraulic capacity and background leakages reduction [*Ferrari and Savic*, 2015] related to closed valves, installation and management costs related to the specific type of device and its location (e.g., pipe diameter).

In the case of the segmentation by nodal removal, they are the candidate position for pressure meters. Differently from the classic DMA, a "pressure DMA" directly descends from the selected "conceptual scenario" of segmentation, because the installation of pressure measurements does not alter the hydraulic behavior of the network, not altering its topology. Furthermore, the number of removed nodes to obtain the modules can surrogate the total cost of "pressure DMA" design, because the cost of the single pressure meter does not generally depend on its location through the asset characteristics. Consequently, only the practical step, i.e., the installation of devices, for passing from a segmentation scenario to a "pressure DMA" does exist.

7. Apulian Case Study

In the present work, the first case study relates to a small network, named Apulian [*Giustolisi et al*., 2008], having one reservoir feeding by gravity all nodes, 23 nodes and 34 pipes, whose layout is in Figure 1.

This simple network is helpful to show and discuss the main features of the proposed sampling design strategy, where the "pressure DMAs" are easily visible. The analysis considers the pipe length as weight (attributes) and the segmentation solutions descend from the two-objective optimization, where the selected *sampling-oriented* modularity is maximized with the minimum number of pressure meters (observations).

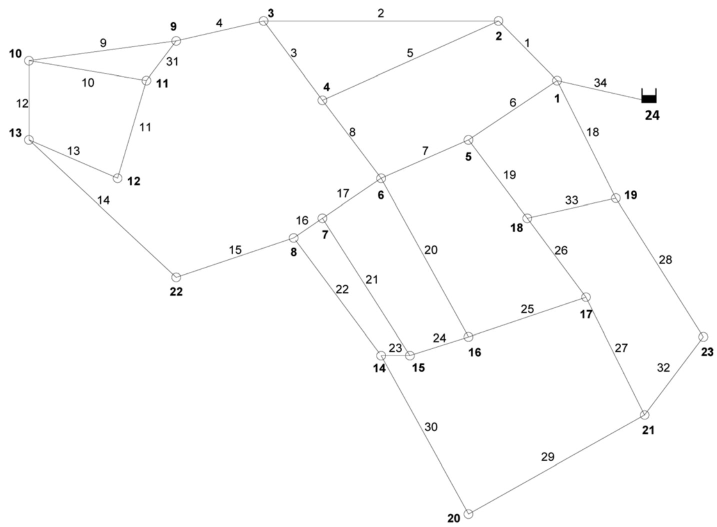

Figure 1. Apulian Layout.

The strategy uses the first metric of equation (5), i.e., the *sampling-oriented modularity* Q_s, allowing identifying modules similar to each other with respect to their total length. The resolution limit affects the selected modularity, although, as previously said, it is not significant for small networks.

The optimization problem provides the Pareto set of solutions reported in Figure 2. The seven black circles represent the optimal tradeoffs between the minimization of the number of installed pressure meters (*x*-axis) on the border of each "pressure DMA" versus the maximization of the sampling-oriented modularity (*y*-axis). All optimal "conceptual segmentation" configurations, based on the objective functions in equation (8) with $f_1 = \max\{Q_s\}$, are characterized by a number of "conceptual removing of nodes," i.e., the number of pressure meters on *x*-axis, which are located in nodal positions maximizing the value of the sampling-oriented modularity.

Therefore, Figure 3 shows the sampling design configurations starting from the trivial configuration (a), i.e., the original network without pressure meters. The pressure measure close to the reservoir is actually assumed by the procedure supposing the existence of level observations for water storages.

Then, Figures 3b–3g show the nontrivial configurations where the black squares represent the locations of pressure meters while the pipes with different grey scale represent the modules, i.e., the "pressure DMA" designed by each solution of the Pareto set (Figure 2). The number of "pressure DMA" increases with the number of observations and the meters are located in the peripheral of the modules. For instance, the first *sampling-oriented* modularity of equation (5) allows obtaining a maximum number of modules equal to 7 with 8 pressure meters because of the resolution limit. Each solution is the best considering the number of pressure observations and the specific pipe length-based sampling-oriented modularity, and the fact that in same solutions two or more pressure observations are close to each other is related to the fact that in small networks not many solution are provided. Finally, all solutions are characterized by a pressure observation in node 1, which is significant being the entrance of the network.

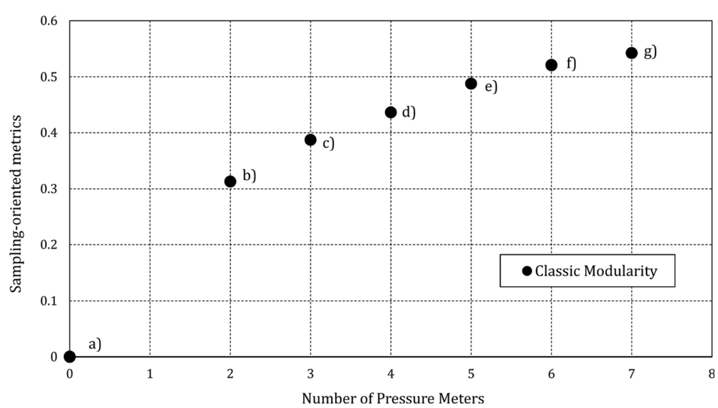

Figure 2. Pareto set of optimal solutions, i.e., optimal sampling configurations.

8. Exnet Case Study

The optimal sampling design strategy based on *sampling-oriented* modularity has been further performed for the Exnet network [*Farmani et al.*, 2004; *Giustolisi et al.*, 2008] a real infrastructure network of a medium size hydraulic system, whose layout is reported in Figure 4.

The network is composed of 1,894 nodes (reservoir comprised) and 2,471 pipes. It is a powerful test for the proposed sampling design strategy using

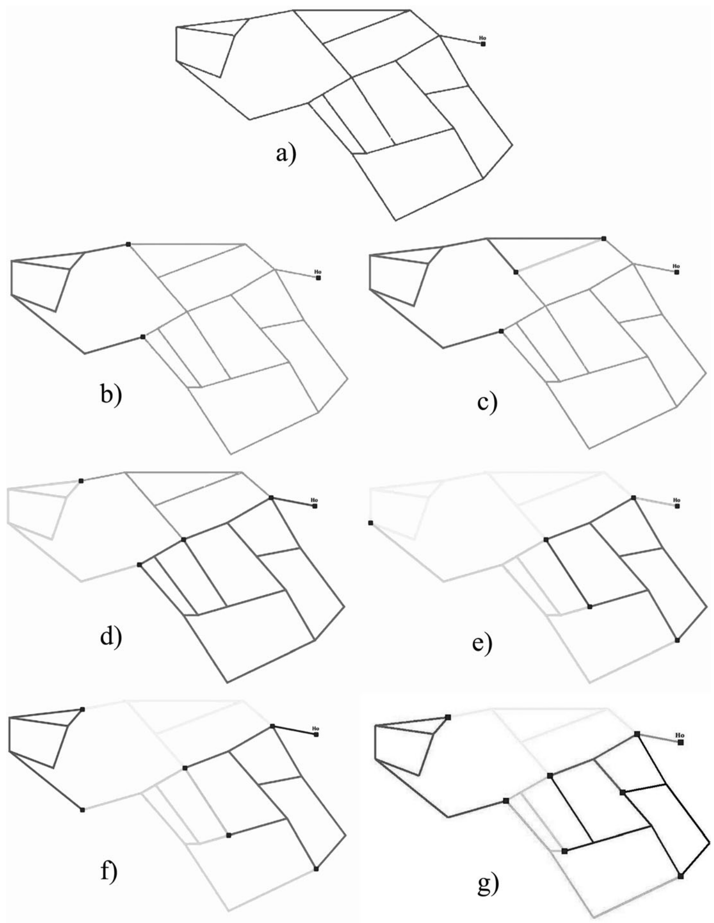

Figure 3. Sampling design configurations corresponding to the Pareto set. The black squares represent the locations of pressure meters.

the first two metrics of equation (5) i.e., *classic* versus *infrastructure* to solve the resolution limit. As for the Apulian network case, the optimization is conceived to identify modules similar each other with respect their total lengths, which are assumed as pipe weights, see equation (9). The two-objective optimization problem of equation (8) using for f_1 the *sampling-oriented* modularity in the classic (Q_s) and infrastructure (IQ_s) versions, was solved using the MOGA optimization strategy as implemented in a specific function of WDNetXL [*Giustolisi et al*., 2011]. Note that a single optimization run takes about 3 h, which is an acceptable computational time from a technical standpoint, since it has to be performed in planning phase only. Figure 5 reports the two Pareto sets corresponding to the solution of the two-optimization procedures. The circles represent the optimal tradeoffs between the number of installed pressure meters (*x*-axis) on the border of each "pressure DMA" versus the maximization of the *classic* (grey circles) and *infrastructure* (black circles) *sampling-oriented* metrics (*y*-axis).

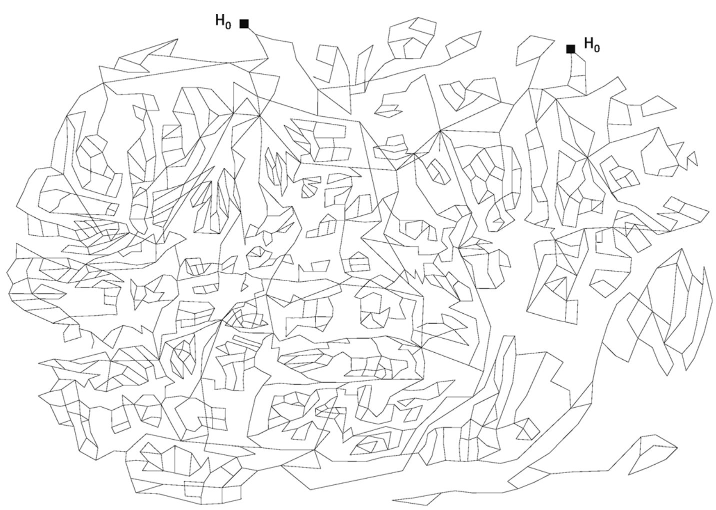

Figure 4. Layout of Exnet network.

From a practical standpoint, water utilities can decide the best trade-off resolution of "pressure DMA" versus budget. It is to bear in mind that for example the active leakage detection and the natural decreasing of the pressure meter cost and management will move the trade-off toward a higher number of "pressure DMAs."

Figure 5 shows the maximum value of pressure meters for the sampling-oriented modularity. The number is equal to 44 and 101 for the classic and infrastructure modularity,

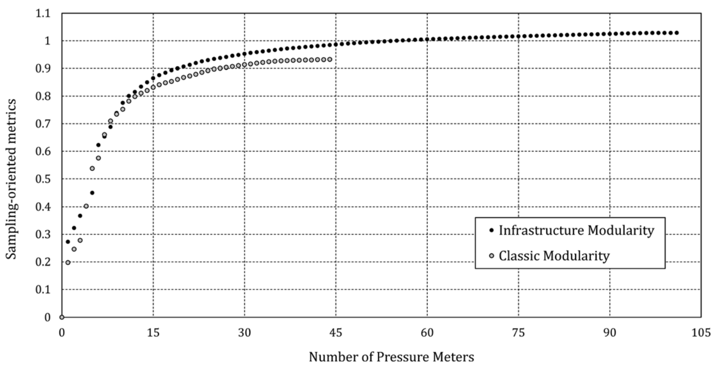

Figure 5. Pareto set of optimal solutions, i.e., optimal sampling configurations, corresponding to the use of the first and second sampling-oriented modularity of equation (5).

respectively. Figures 6 and 7 report the configurations of pressure meters (black squares) corresponding to the maximum value of the sampling-oriented metrics. Note that the 44 pressure meters determine 45 modules, i.e., "pressure DMAs," while the 101 pressure meters determine 205 "pressure DMAs."

Therefore, Figures 5–7 demonstrate using a medium-size network that the infrastructure modularity is helpful to overcome the resolution limit, allowing determining a number of modules more than four time greater than classic metric case and about doubling the number of pressure meters. Note that being the decision variables related to new pressure meters to be installed, the already existing pressure measurements corresponding to control valves, pumps, tanks and reservoirs could be considered as constraints. The bias due to the existence of those nonoptimal constraints is a further reason of introducing the infrastructure modularity index, which solving the resolution limits is not sensitive to constraints as better reported in *Giustolisi and Ridolfi* [2014b].

This was achieved assuming the minimum module length as using $x\% = 0.05\%$ of the total Exnet pipeline length, i.e., see equation (9). It is important to remark that although the *sampling-oriented infrastructure* metric solving the resolution limit is much powerful, it does not necessary mean that the configuration corresponding to the maximum number of "pressure DMAs" should be selected. Nevertheless, the number of solutions obtained by using the *sampling-oriented infrastructure* metric are much greater (see Figure 5) than the *sampling-oriented classic* metric; therefore, it is preferable from a decision-making support standpoint.

9. Refining the Sampling Design

The proposed sampling design strategy locates nodal pressure meters creating "pressure DMAs," which can be the basis of a positioning of further meters internal to modules.

For instance, a specific technical task can ask for locating a pressure meter in the central node of "pressure DMAs" in order to select the "most relevant" position with respect to the hydraulic behavior. In Figure 8, the grey circles represent the "candidate positioning" for central nodes. The proposed metric to determine the positions is the *betweenness centrality*, measuring how much the node is central with respect to water paths. In fact, *betweenness centrality* is maximum for the node that is central because located between many other nodes, i.e., many shortest paths connecting couples of nodes pass through it. It is important to note that this is a refinement, based on graph theory, for the proposed sampling design strategy. This means that other measures of centrality can be used [*Wang et al.*, 2008] to support the proposed strategy.

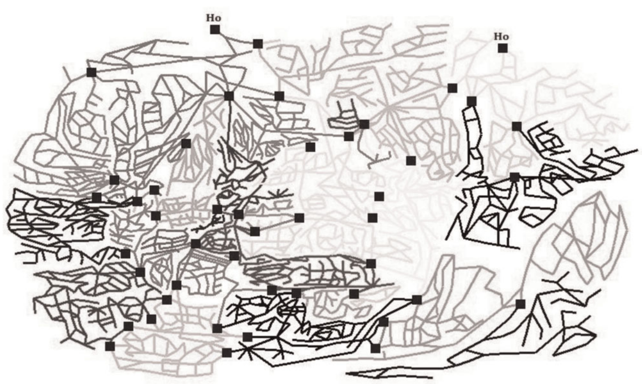

Figure 6. Sampling design configuration corresponding to the maximum value of the sampling-oriented modularity. The black squares represent the locations of pressure meters.

The option of selecting the central node of one specific module can depend on the technical task (e.g., monitoring of head losses, demands, background leakages, or burst detections, rehabilitation works, model calibration, etc.) or other topologic characteristics such as size, connectivity, etc.

In the same line of the selection of central nodes, the proposed strategy

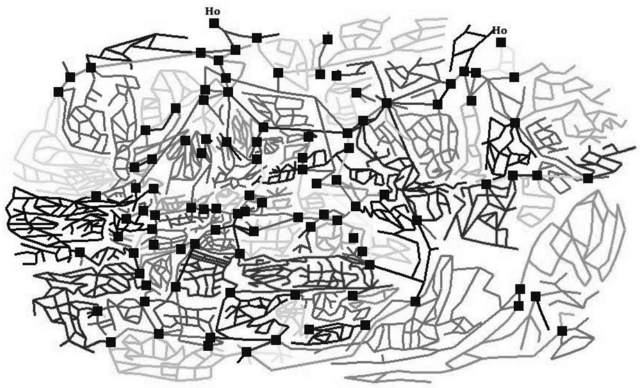

Figure 7. Sampling design configuration corresponding to the maximum value of the sampling-oriented infrastructure modularity. The black squares represent the locations of pressure meters.

allows considering nodes with high demands and/or nodes on the perimeter of the network [*Walski*, 1983] for the peripheral pressure DMAs. It is worth noting that the proposed strategy is helpful to locate further pressure meters [*Walski*, 1983] based on engineering judgment because the decision is simplified from WDN dimension to the "pressure DMA" level.

9.1. Additional Remarks

It is worth to remark now that *equation* (5) assume that n_{obs} stays for conceptually "removed nodes" and consistently the denominator is the number of nodes n_n. This fact means that one node can generate more than one district and the maximum value of the *sampling-oriented* metrics is greater than one and depends on the network topology, as showed in Figure 5 for the infrastructure modularity.

Therefore, a slightly different formulation for *sampling-oriented* metrics can be proposed:

$$\begin{aligned}
Q_s(\mathbf{w}_p) &= 1 - \frac{n_{\deg}}{2n_p} - \sum_{m=1}^{n_m}\left[\sum_{k=1}^{n_p}\frac{(\mathbf{w}_p)_k\ \delta(\mathrm{M}_m,\mathrm{M}_k)}{W}\right]^2 \\
IQ_s(\mathbf{w}_p) &= 1 - \frac{n_{\deg}-(n_{act}-1)}{2n_p} - \sum_{m=1}^{n_m}\left[\sum_{k=1}^{n_p}\frac{(\mathbf{w}_p)_k\ \delta(\mathrm{M}_m,\mathrm{M}_k)}{W}\right]^2 \\
Q_{a-s}(\mathbf{a}_p) &= 1 - \frac{n_{\deg}}{2n_p} - \sum_{m=1}^{n_m}\left[\frac{\sum_{k=1}^{n_p}|\mathbf{a}_p-\bar{a}(\mathrm{M}_m)|_k\ \delta(\mathrm{M}_m,\mathrm{M}_k)}{\sum_{k=1}^{n_p}|\mathbf{a}_p-\bar{a}(\mathrm{N})|_k}\right]^2 \\
IQ_{a-s}(\mathbf{a}_p) &= 1 - \frac{n_{\deg}-(n_{act}-1)}{2n_p} - \sum_{m=1}^{n_m}\left[\frac{\sum_{k=1}^{n_p}|\mathbf{a}_p-\bar{a}(\mathrm{M}_m)|_k\ \delta(\mathrm{M}_m,\mathrm{M}_k)}{\sum_{k=1}^{n_p}|\mathbf{a}_p-\bar{a}(\mathrm{N})|_k}\right]^2
\end{aligned} \tag{10}$$

where, n_{deg} is summation of the degrees of the "removed nodes" and consistently $2n_p$ is summation of nodal degrees. Consequently, the same tests on Exnet network were performed using the first two metrics of equation (10). Figures 9 and 10 show the sampling configurations related to the maximum value of the *sampling-oriented* modularity, classic and infrastructure respectively.

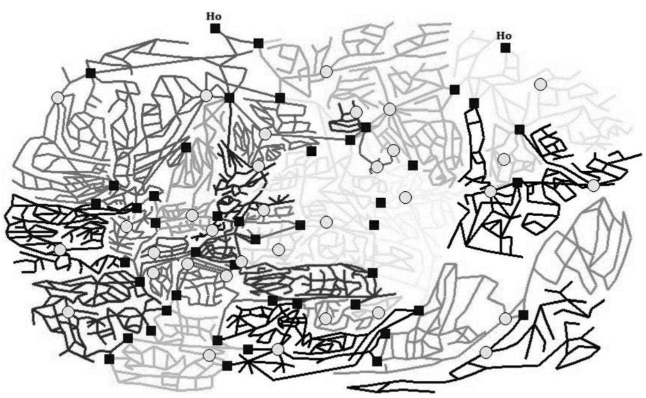

Figure 8. Sampling design configuration corresponding to the maximum value of the sampling-oriented modularity. The grey circles represent the locations of pressure meters positioned by means of betweenness centrality.

Comparing Figures 6 and 9 concerning to the DMA configurations using the two *sampling-oriented classic modularity*, first metric of equations (5) and (10) respectively, it is possible to state that they work similarly. In fact, the number

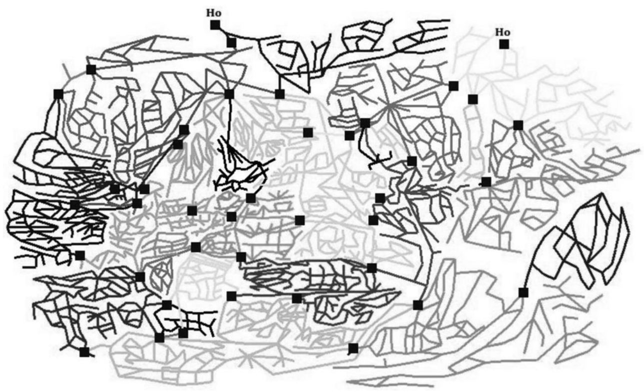

Figure 9. Sampling design configuration corresponding to the maximum value of the sampling-oriented modularity. The black squares represent the locations of pressure meters.

of pressure meters is 44 and 39 respectively, and the number of "pressure DMAs" is 45 and 43 respectively.

Comparing Figure 7 and 10 concerning to the DMA configurations using the two *sampling-oriented infrastructure modularity*, second metric of equations (5) and (10) respectively, it is evident that the formulation with n_{obs} is much more powerful than the formulation with n_{deg}, even though, Figure 7 shows a high number of the "pressure DMA" composed by a few pipes. In fact, the number of pressure meters is 103 and 58 respectively, and the number of "pressure DMAs" is 203 and 67 respectively.

The assumption that one removed node counts its degree with respect to $2n_p$ (or similarly half of its degree with respect to n_p) causes the fact that the infrastructure modularity of equation (10) is less powerful than that of equation (5). Indeed, *Giustolisi and Ridolfi* [2014b] introduced the infrastructure modularity in order to solve the dominance (given one cut with respect to n_p) of the decreasing of Q_1 with respect to the increasing of Q_2 when a limit number of modules (depending on the network size) is reached, see the reported discussion after equation (1). The second and forth modularity of equation (10) cannot solve the resolution limit because one removed node now counts half of its nodal degree (with respect to n_p) which is greater or equal than unit for internal nodes.

10. 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 multiobjective optimization problem, where a new segmentation metric, suited for sampling design, is maximized while minimizing the number of "conceptual removed nodes."

The multiobjective 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.

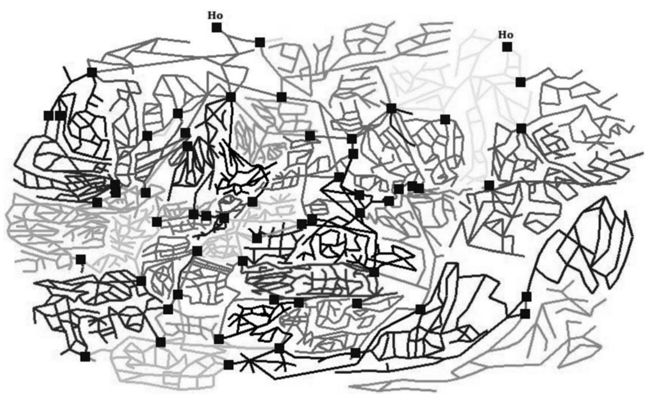

Figure 10. Sampling design configuration corresponding to the maximum value of the sampling-oriented infrastructure modularity. The black squares represent the locations of pressure meters.

It is worth noting that the proposed strategy is helpful to locate further pressure meters based on engineering judgment, e.g., high demands and/or nodes on the perimeter of the network [*Walski*, 1983], because the decision is simplified from WDN dimension to the "pressure DMA" level.

The novel approach allows also positioning other pressure gauges at *central* node of each "pressure DMA," using the maximum value of the *betweenness centrality* [*Freeman*, 1977] of the specific "pressure DMA," in order to increase the information collected by the pressure monitoring system. This makes the selected nodes relevant with respect to the hydraulic behavior of the WDN.

Notation

$\mathbf{A}_{pn}$	general topological matrix of the network ($A_{pn} = \mathbf{A}_{np}{}^{T}$).
$\mathbf{a}_{p}$,	vector of pipe attributes.
$\bar{a}$(N)	mean value of the pipe attributes in N.
C_1	constraint.
f_i	objective function.
I_c	set of n_{obs} conceptually removed nodes in the network.
k_i	node degree of the i-th node.
$\mathbf{L}$	connectivity matrix of the edges/pipes.
$(\mathbf{L}_p)_k$	vector of pipe length values in module k.
L	sum of pipes length values.
M_i	identifier of network modules.
N	network.
n_{act}	actual number of modules satisfying the constraints.
n_c	number of pipes linking modules of the infrastructure.
n_m	number of network modules.
n_p	number of network links/pipes.
Q	WDN-oriented modularity index.
IQ	infrastructure modularity index.
Q_a	attribute–based modularity index.
IQ_a	attribute–based infrastructure modularity index
Q_s	sampling-oriented modularity index.
IQ_s	sampling infrastructure modularity index.
$Q_{a\text{-}s}$	sampling-oriented attribute–based modularity index.
$IQ_{a\text{-}s}$	sampling infrastructure *attribute–based* modularity index
$\mathbf{w}_{p}$,	vector of pipe weights.
W	sum of pipe weights.
x%	threshold percentage of total network length for a module to be technically useful.
δ	Kronecker's delta function.

Acknowledgments
The proposed optimal sampling design strategy is implemented in WDNetXL Design Module. The system tool WDNetXL can be requested free of charge for students and research purposes at www.idea-rt.com. The network data used in WDNetXL can be obtained contacting Orazio Giustolisi (orazio.giustolisi@poliba.it). 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."

References

Agarwal, G., and D. Kempe (2008), Modularity-maximizing graph communities via mathematical programming, *Eur. Phys. J. B*, *66*(3), 409–418.
Albert, R., and A. L. Barabasi (2002), Statistical mechanics of complex networks, *Rev. Mod. Phys.*, *74*(1), 47–97.
Alvisi, S., and M. Franchini (2014), A heuristic procedure for the automatic creation of district metered areas in water distribution systems, *Urban Water J.*, *11*(2), 137–159.
Barthélemy, M. (2011), Spatial networks, *Phys. Rep.*, *499*(13), 1–101.
Behzadian, K., A. Ardeshir, Z. Kapelan and D. A. Savic (2008), Stochastic sampling design for water distribution model calibration, *IJCE*, *6*(1), 48–57.
Berardi, L., O. Giustolisi, Z. Kapelan, and D. A. Savic (2008), Development of pipe deterioration models for water distribution systems using EPR, *J. Hydroinformatics*, *10*(2), 113–126.
Bhave, P. R. (1988), Calibrating water distribution network models, *J. Environ. Eng.*, *114*(1), 120–136.
Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008), Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment*, *10*, 12 pp.
Boulos, P. F., and D. J. Wood (1991), An explicit algorithm for calculating operating parameters for water networks, *Civ. Eng. Syst.*, *8*, 115–122.
Brualdi, R. A., and H. J. Ryser (1991), *Combinatorial Matrix Theory*, Cambridge University Press, Cambridge.
Carrera, J., E. Usunoff, and F. Szidarovszky (1984), A method for optimal observation network design for groundwater management, *J. Hydrol.*, *73*(1-2), 147–163.
Chen M., K. Kuzmin, and B. K. Szymanski (2014), Community detection via maximization of modularity and its variants, *IEEE Trans. Comput. Soc. Syst.*, *1*(1), 46–65.
Clauset, A., M. E. J. Newman, and C. Moore (2004), Finding community structure in very large networks, *Physical Review E*, *70*, 066111.
Davidson, J., F. Bouchart, S. Cavill, and P. Jowitt, (2005), Real-time connectivity modelling of water distribution networks to predict contamination spread, *J. Comput. Civ. Eng.*, *19* (4), 377–386.
De Schaetzen, W., G. A. Walters, and D. A. Savic (2000), Optimal sampling design for model calibration using shortest path, genetic and entropy algorithms, *Urban Water*, *2*, 141–152.

Deuerlein, J. W. (2008), Decomposition model of a general water supply network graph, *J. Hydraul. Eng.*, *134*(6), 822–832.

Diao, K., Y. Zhou, and W. Rauch (2013), Automated creation of district metered area boundaries in water distribution systems, *J. Water Resour. Plann. Manage.*, *139*(2), 184–190.

Di Nardo, A., M. Di Natale, C. Giudicianni, D. Musmarra, G. F. Santonastaso, and A. Simone (2015), Water distribution system clustering and partitioning based on social network algorithms, *Procedia Eng.*, *119*, 196–205.

Duch, J., and A. Arenas (2005), Community detection in complex networks using extremal optimization, *72*, 027104, doi:10.1103/PhysRevE.72.027104.

Farmani, R., D. A. Savic, and G. A. Walters (2004), ''EXNET'' benchmark problem for multi-objective optimization of large water systems. Modelling and control for participatory planning and managing water systems, paper presented at IFAC workshop, Venice, Italy, 29 Sept.–1 Oct.

Ferrari, G., and D. A. Savic (2015), Economic Performance of DMAs in Water Distribution Systems, *Procedia Eng.*, *119*, 189–195.

Fortunato, S. (2010), Community detection in graphs, *Phys. Rep.*, *486*(3-5), 75–174.

Fortunato, S., and M. Barthélemy (2007), Resolution limit in community detection, *Proc. Natl. Acad. Sci. U. S. A.*, *104*, 36–41.

Freeman, L. (1977), A set of measures of centrality based upon betweenness, *Sociometry*, *40*(1), 35–41.

Giustolisi, O., and L. Ridolfi (2014a), A new modularity-based approach to segmentation of water distribution networks, *J. Hydraul. Eng.*, *140*(10), 1–14.

Giustolisi, O., and L. Ridolfi (2014b), A novel infrastructure modularity index for the segmentation of water distribution networks, *Water Resour. Res.*, *50*, 7648–7661, doi:10.1002/2014WR016067.

Giustolisi, O., D. A. Savic, and Z. Kapelan (2008), Pressure-driven demand and leakage simulation for water distribution networks, *J. Hydraul. Eng.*, *134*(5), 626–635.

Giustolisi, O., D. A. Savic, L. Berardi, and D. B. Laucelli (2011), An Excel-based solution to bring water distribution network analysis closer to users, in *Proceedings of Computer and Control in Water Industry*, vol. 3, edited by D. A. Savić, Z. Kapelan, and D. Butler, pp. 805–810.

Giustolisi, O., L. Ridolfi, and L. Berardi (2015), General metrics for segmenting infrastructure networks, *J. Hydroinformatics*, *17*(4), 505–517.

Gofman, E. and M. Rodeh (1981), Loop equations with unknown pipe characteristics, *J. Hydraul. Div. Am. Soc. Civ. Eng.*, *107*(9), 1047–1060.

Heskett, J. L. (1986), *Managing in the Service Economy*, Harvard Bus. Schl. Press, Boston.

Jacobs, P., and I. Goulter (1988), Evaluation of methods for decomposition of water distribution networks for reliability analysis, *Civ. Eng. Syst.*, *5*(2), 58–64.

Kapelan, Z., D. A. Savic, and G. A. Walters (2003), Multi-objective sampling design for water distribution model calibration, *J. Water Resour. Plann. Manage.*, *129*(6), 466–479.

Kapelan, Z., D. A. Savic, and G. A. Walters (2005), Optimal sampling design methodologies for water distribution model calibration, *J. Hydraul. Eng.*, *131*(3), 190–200.

Medus, A., G. Acuña, and C. Dorso (2005), Detection of community structures in networks via global optimization, *Physica A*, *358*(24), 593–604.

Newman, M. E. J. (2004), Fast algorithm for detecting community structure in networks, E 69, 066133, doi:10.1103/PhysRevE.69.066133.

Newman, M. E. J. (2006a), Modularity and community structure in networks, *Proc. Natl. Acad. Sci. U. S. A.*, *103*(3), 8577–8582.

Newman, M. E. J. (2006b), Spectral methods for network community detection and graph partitioning, *88*, 042822, doi:10.1103/PhysRevE.88.042822 2013.

Newman, M. E. J., and M. Girvan (2004), Finding and evaluating community structure in networks, *Phys. Rev. E*, *69*, 026113.

Ormsbee, L. E., and D. J. Wood (1986), Explicit pipe network calibration, *J. Water Resour. Plann. Manage.*, *112*(2), 166–182.

Perelman, L., and A. Ostfeld (2011), Topological clustering for water distribution systems analysis, *Environ. Modell. Software*, *26*(7), 969–972.

Piller, O., B. Bremond, and P. Morel (1999), A spatial sampling procedure for physical diagnosis in a drinking water supply network, paper presented at International Conference on Computing and Control for the Water Industry CCWI '99, Exeter, U. K.

Rahal, C. M., M. J. H. Sterling, and B. Coulbeck (1980), Parameter tuning for simulation models of water distribution networks, *Proc. Inst. Civ. Eng., Part 2*, *69*, 751–762.

Sales-Pardo M., R. Guimerà, A. A. Moreira, and L. A. N. Amaral (2007), Extracting the hierarchical organization of complex systems, *Proc. Natl. Acad. Sci. U. S. A.*, *104*(39), 15,224–15,229.

Sanz, G., and R. Pérez (2015), Sensitivity analysis for sampling design and demand calibration in water distribution networks using the singular value decomposition, *J. Water Resour. Plann. Manage.*, *141*(10), 04015020.

Scibetta, M., F. Boano, R. Revelli, and L. Ridolfi (2013), Community detection as a tool for complex pipe network clustering, *EPL*, *103*, 48001.

Shamir, U., and C. D. D. Howard (1968), Water distribution systems analysis, *J. Hydraul. Div. Am. Soc. Civ. Eng.*, *94*(1), 219–234.

Steinhaeuser, K., and N. V. Chawla (2010), Identifying and evaluating community structure in complex networks, *Pattern Recognition Lett. Arch.*, *31*(5), 413–421.

Todini, E., and S. Pilati (1988), A gradient method for the solution of looped pipe networks, in *Computer Applications in Water Supply*, vol. 1, pp. 1–20, John Wiley, London.

Wakita, K., and T. Tsurumi (2007), Finding community structure in mega-scale social networks [extended abstract], in *Proceedings of the 16th international conference on World Wide Web*, pp. 1275–1276, ACM, N. Y.

Walski, T. M. (1983), Technique for calibrating network models, *J. Water Resour. Plann. Manage.*, *109*(4), 360–372.

Wang, G., Y. Shen, and E. Luan (2008), A measure of centrality based on modularity matrix, *Prog. Nat. Sci.*, *18*(8), 1043–1047.

Yang, S.-L., N.-S. Hsu, P. W. F. Loule, and W. W.-G. Yeh (1996), Water distribution network reliability: Connectivity analysis, *J. Infrastructure Syst.*, *2*(2), 54–64.

Yazdani, A., and P. Jeffrey (2012), Water distribution system vulnerability analysis using weighted and directed network models, *Water Resour. Res.*, *48*, W06517, doi:10.1029/2012WR011897.

Yu, G., and R. S. Powell (1994), Optimal design of meter placement in water distribution systems, *Int. J. Syst. Sci.*, *25*(12), 2155–2166.