
Journal of Hydroinformatics

General metrics for segmenting infrastructure networks

O. Giustolisi, L. Ridolfi and L. Berardi

ABSTRACT

The classic modularity index for community detection in complex networks was recently tailored to water distribution networks (WDNs) and extended in order to be cut-position sensitive. Next, the WDN-oriented modularity index was enhanced in order to overcome the resolution limit of the classic modularity. Nonetheless, the modularity-based metrics developed so far allow the networks to be segmented into modules/segments that are similar to each other according to specific pipe characteristics (e.g., pipe lengths, distributed demand, background leakages, etc.). The present work extends and proves the strategy to overcome the resolution limits focusing on an infrastructure index that drives WDN segmentation toward modules that are internally similar with respect to given attributes (e.g., pipe diameters, average pipe pressures, average pipe elevations, etc.), since this aim is suitable for several practical purposes. The introduction of the *attribute*-based infrastructure index permits a comprehensive discussion and comparison of the metrics for infrastructure network segmentation through simple examples. Finally, the practical implications of increasing the resolution of internally similar modules are demonstrated on a well-known benchmark WDN considering various pipe attributes.

Key words | infrastructure networks, modularity index, network segmentation, resolution limit

O. Giustolisi
L. Berardi (corresponding author)
Department of Civil and Environmental Engineering and Architecture,
Politecnico di Bari,
Via E. Orabona, 4,
70126,
Bari,
Italy
E-mail: *luigi.berardi@poliba.it*

L. Ridolfi
Department of Environmental, Land, and Infrastructure Engineering,
Politecnico di Torino,
Turin,
Italy

INTRODUCTION

Water distribution networks (WDNs) are essential for all human activities in urban areas. The complexity in analyzing, managing and planning works on such infrastructures stems from their large size (up to thousands of pipes), the underlying hydraulics, as well as the alteration of asset conditions from their original installation. A pragmatic approach to understand WDN real behavior and support effective decisions resorts to segmenting the system into smaller portions (named *districts*, *segments* or *modules*) suited for different technical purposes including monitoring (e.g., district metering areas), control (e.g., pressure control zones) or even WDN modeling (e.g., calibration). Nonetheless, segmenting a real WDN is not a trivial task due to its size, the looped topology, its hydraulic functioning and because cuts that separate modules are actually costly devices (e.g., valves, flow/pressure gauges) to be installed at technically feasible locations (i.e., in accessible vaults/manholes at pipes end nodes). In addition, a WDN segmentation designed to match a technical purpose might not be adequate for another different scope.

The interest in this topic is documented by many contributions where segmentation is analyzed for various purposes including reliability analysis (e.g., Jacobs & Goulter 1988; Yang *et al.* 1996), location of isolation valves (e.g., Walski 1983), and analysis of contaminant spread (e.g., Davidson *et al.* 2005). Other studies exploited concepts from graph theory to identify the main structure in WDN for monitoring and control goals (e.g., Deuerlein 2008; Perelman & Ostfeld 2011; Alvisi & Franchini 2014) like model calibration, metering water consumption, early contaminant detection, control of pressure/leakages, and network vulnerability analysis (Yazdani & Jeffrey 2012).

A recent approach (Scibetta *et al.* 2013; Diao *et al.* 2013) faced the problem of segmenting WDNs by applying the concepts of community detection (e.g., Fortunato 2010) from complex network theory (Albert & Barabasi 2002;

doi: 10.2166/hydro.2015.102

Newman 2010). The initial contributions in this area analyzed the application of the classic modularity concept (Newman & Girvan 2004) to identify WDN modules. However, it was observed by some authors (e.g., Barthélemy 2011; Giustolisi & Ridolfi 2014a) that strong differences exist between the immaterial networks (e.g., food web, trade, World Wide Web), for which the classical modularity concept was conceived, and infrastructure networks (e.g., gas, electricity, water). In particular, WDNs have material links (pipes), spatial constraints (two dimensionality, urban layout, location of water sources/demands, etc.), and material devices installed along links (e.g., valves, pumps, meters, etc.).

Giustolisi & Ridolfi (2014a) introduced a WDN-oriented modularity index that accounted for the WDN infrastructural peculiarities. In short, WDN-oriented modularity was (i) sensitive to position of cuts (since devices in WDN are installed close to nodes instead of middle of links, as originally assumed for immaterial networks); (ii) its formulation was based on the number of pipes into the modules instead of nodal degree; and (iii) pipe features were introduced in terms of *weights* of network links to drive the segmentation process.

Consistently with classic modularity concept, the *weight*-based WDN-oriented modularity index allows identification of modules that are similar to each other. From a technical perspective, it was reported to be of direct relevance for length-based pipe characteristics (e.g., total water demand distributed along pipes, propensity to pipe background leakages, etc.). Indeed, maximizing such WDN-oriented modularity index is likely to return modules that are suited for technical purposes such as, for example, water consumption metering and leakage monitoring.

The problem of designing WDN modules was formulated as a multi-objective problem where the WDN-oriented modularity index should be maximized with the minimum number of cuts (i.e., costly devices to be installed).

Unfortunately, both classic and WDN-oriented modularity indexes are known to suffer from a resolution limit (Fortunato & Barthélemy 2007), which prevents further maximizing the value of the metric by increasing the number of cuts, after a threshold number of modules in the network is reached. Giustolisi & Ridolfi (2014b) analyzed the resolution limit of the *weight*-based modularity and proposed a new infrastructure modularity index to overcome such limit.

In addition, the first work by Giustolisi & Ridolfi (2014a) reported an *attribute*-based variant of the WDN-oriented modularity index that was conceived to maximize the similarity of pipes within each module. From a technical perspective, the *attribute*-based index was conceived to exploit pipe features not strictly related to pipe length (e.g., pipe diameter, average elevation, average pipe pressure, etc.). Thus, modules identifiable by maximizing the *attribute*-based index are better suited for other practical purposes such as, for example, WDN model calibration, pressure control, or leakage control.

Although the referenced works provided relevant innovations on the modularity-based approach for WDN segmentation, the framework of the WDN-oriented modularity indices need to be completed and explicitly framed from a technical perspective.

The present work aims at filling this gap by introducing and discussing the *attribute-oriented infrastructure index* that extends the strategy for mitigating the resolution limit to the *attribute*-based index.

A comprehensive framework of the segmentation metrics (directly based on modularity index or simply recalling the structure of that index) is presented along with a discussion on practical implications from WDN management perspectives.

Simple examples clarify the key concepts and the differences among the infrastructure segmentation metrics and provide thoughtful criteria for practitioners to select the best one according to specific technical purposes. Finally, the TOWN-C (Ostfeld *et al.* 2012) water distribution network is used to discuss the practical implications of increasing the resolution of modules by using the *attribute-oriented infrastructure segmentation index*, considering diameters or average elevation as pipe attributes.

BRIEF ON WDN-ORIENTED MODULARITY INDEX

The recently proposed WDN-oriented metrics for optimal segmentation by Giustolisi & Ridolfi (2014a, b) are modularity-based indexes tailored and modified in order to be much more effective for infrastructure systems, starting from the modularity index as defined in Newman & Girvan (2004)

$$Q = \frac{1}{2n_p}\sum_{ij}(A_{ij} - P_{ij})\delta(\mathrm{M}_i, \mathrm{M}_j) = \frac{1}{2n_p}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2n_p}\right)\delta(\mathrm{M}_i, \mathrm{M}_j) \quad (1)$$

where n_p is the number of network links/pipes, A_{ij} are the elements of the adjacency matrix, P_{ij} is the expected fraction of pipes between vertices/nodes i and j in the null/random network (i.e., the expected number of pipes in the network if they were randomly distributed), M_i is the identifier of network modules, δ is the Kronecker's delta function to apply the summation to the elements of the same module (i.e., $\delta = 1$ if M_j=M_i and $\delta = 0$ otherwise), and summation runs on all the possible node couples (i,j), with $i \neq j$. In Equation (1), the expected fraction P_{ij} is computed using node degree k_i (k_j) of the i-th (j-th) node, i.e., the number of pipes incident in the node. The metric behind the modularity index measures the strength of a network or graph division into communities/modules. Hereinafter, we will use the word 'module' as it is more usual for infrastructure networks.

The proposed formulation of the WDN-oriented modularity in Giustolisi & Ridolfi (2014a) is

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

where n_c is the number of pipes linking modules of the infrastructure, namely the number of '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 delta function δ makes that the sum refer only to the weights of pipes belonging to the m-th module (i.e., $\delta = 1$ if M_m=M_k and $\delta = 0$ otherwise). The WDN-oriented modularity is cut-position sensitive and counts pipes belonging to modules instead of the nodal degrees of the classic modularity. In fact, in WDN the pipe cuts occur in the actual position of the devices segmenting the network instead of in the middle of links, differently from what is assumed in classic modularity index when immaterial networks are considered.

It is worth noting that the last term in Equation (2) generally decreases with the number of components n_m (Giustolisi & Ridolfi 2014a) and the following constraint holds

$$\sum_{m=1}^{n_m}\left[\sum_{k=1}^{n_p}\frac{(\mathbf{w}_p)_k\delta(\mathrm{M}_m, \mathrm{M}_k)}{W}\right] = 1 \quad (3)$$

Equation (3) allows us to explain that the WDN-oriented modularity index, as well as the classic modularity index, measures the similarity of the modules to each other. In fact, if we write

$$Q = Q_1 + Q_2$$
$$Q_1 = 1 - \frac{n_c}{n_p}$$
$$Q_2 = -\sum_{m=1}^{n_m}a_m^2 = -\sum_{m=1}^{n_m}\left[\sum_{k=1}^{n_p}\frac{(\mathbf{w}_p)_k\delta(\mathrm{M}_m, \mathrm{M}_k)}{W}\right]^2 \quad (4)$$

it is evident that maximizing the metric Q means finding the best set of cuts (in terms of number and positions) that generates segments that maximize Q_2. Since Q_2 is the negative summation of the squares of n_m numbers whose sum is unitary (see Equation (3)), it is maximized if the modules are similar to each other as much as possible depending on the topological distribution of the pipe weights in the network (Giustolisi & Ridolfi 2014b).

The WDN-oriented modularity index in Equation (2) is known to suffer from the resolution limit that occurs because there is a bound of the metrics of Equations (1) and (2) to the identification of small size modules. In fact, the two components, Q_1 and Q_2 are conflicting with respect to the number of modules n_m and a mathematical dominance of Q_1 with respect to Q_2 (namely the value of Q_1 is always larger than Q_2), always occurs. This fact generates a sort of barrier for the identification of small modules whose value depends on the size of the network (Fortunato & Barthélemy 2007; Giustolisi & Ridolfi 2014b).

To overcome the resolution limit, Giustolisi & Ridolfi (2014b) proposed the *infrastructure modularity index*

$$IQ(\mathbf{w}_p) = 1 - \frac{n_c - (n_m - 1)}{n_p} - \sum_{m=1}^{n_m}\left[\sum_{k=1}^{n_p}\frac{(\mathbf{w}_p)_k\delta(\mathrm{M}_m, \mathrm{M}_k)}{W}\right]^2 \quad (5)$$

It is obtained starting from the WDN-oriented modularity index and adding the term $(n_m - 1)/n_p$ that represents the minimum theoretical fraction of pipes to be cut to obtain n_m modules. This biases the term Q_1 in Equation (4) by a function of the number of modules (n_m) and eliminates the mathematical dominance of Q_1 vs. Q_2. The *infrastructure modularity index* has further good features with respect to the WDN-oriented modularity index which makes it more effective for multi-objective segmentation design for WDNs, as discussed by Giustolisi & Ridolfi (2014b).

Finally, Giustolisi & Ridolfi (2014a) proposed a different metric for infrastructure network segmentation which is better suited to divide the hydraulic system into modules having pipe *attributes* which are similar inside each module. It yields

$$Q_a = 1 - \frac{n_c}{n_p} - \sum_{m=1}^{n_m} \left[\frac{\sum_{k=1}^{n_p} \left| \mathbf{a}_p - \bar{a}(\mathrm{M}_m) \right|_k \delta(\mathrm{M}_m, \mathrm{M}_k)}{\sum_{k=1}^{n_p} \left| \mathbf{a}_p - \bar{a}(\mathrm{N}) \right|_k} \right]^2 \quad (6)$$

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 M_m. Function δ limits the summation of the pipe attributes to the elements belonging to the same module.

Indeed, for some WDN management purposes, it is technically advisable to segment the network by searching for cuts (i.e., devices) which generate modules with similar internal pipe characteristics like pipe diameters, pipe average pressures/elevations, etc.

Thus, the *attribute*-based index in Equation (6) has only a similar mathematical expression as the modularity-based index of Equation (2), but different properties.

In summary, Equation (2) is the WDN-oriented modularity (*weight*-based) index *measuring the similarity of modules to each other* and Equation (5) is the infrastructure modularity index having the same feature but aimed at eliminating the resolution limit drawback. Equation (6) is a further WDN-oriented (*attribute*-based) index *measuring similarity within each module with respect to a specified attribute*. In this latter case, we use the word *attribute*, instead of *weight*, to indicate a specific pipe characteristic in order to stress the different aim of Equation (6) with respect to Equations (2) and (5).

It should be noted that the constraint to unit of Equation (3), that is the driver for similarity among modules in Equations (2) and (5) does not hold for the *attribute*-oriented index in Equation (6). However, also the *attribute*-oriented index could be affected by the resolution limit drawback. Therefore, the aim of the next section is to extend the infrastructure index to Equation (6) and demonstrate that, although the resolution limit does not strictly exist for Equation (6), the modification of adding the term $(n_m - 1)/n_p$ is also beneficial for that *attribute*-oriented index.

This is of technical relevance since, according to the multi-objective strategy for WDN segment design, the increase of number of cuts (i.e., costly valves/devices) is justified by an increased value of the adopted metric.

ATTRIBUTE-BASED INFRASTRUCTURE SEGMENTATION INDEX

The resolution limit concerns the actual possibility to increase the value of the WDN-oriented modularity considering the increase of segmentation by one module using one cut, i.e., the minimum possible number of cuts, starting from n_m modules. Actually, this means to assume a sequential search of optimal cuts provided that, for generality of discussion, the segmentation with n_m modules is a global optimum. Thus, the question is if with one cut it is always possible to obtain $Q_a(n_c + 1, n_m + 1) > Q_a(n_c, n_m)$ assuming starting from a global optimal division in n_m modules.

This means to verify if

$$\begin{gathered} Q_a(n_c + 1, n_m + 1) > Q_a(n_c, n_m) \quad \Rightarrow \\ \sum_{m=1}^{n_m} \left[\frac{\sum_{k=1}^{n_p} \left| \mathbf{a}_p - \bar{a}(\mathrm{M}_m) \right|_k \delta(\mathrm{M}_m, \mathrm{M}_k)}{\sum_{k=1}^{n_p} \left| \mathbf{a}_p - \bar{a}(\mathrm{N}) \right|_k} \right]^2 \\ - \sum_{m=1}^{n_m+1} \left[\frac{\sum_{k=1}^{n_p} \left| \mathbf{a}_p - \bar{a}(\mathrm{M}_m) \right|_k \delta(\mathrm{M}_m, \mathrm{M}_k)}{\sum_{k=1}^{n_p} \left| \mathbf{a}_p - \bar{a}(\mathrm{N}) \right|_k} \right]^2 > \frac{1}{n_p} \end{gathered} \quad (7)$$

Since the constraint of Equation (3) does not hold for the present *attribute*-oriented metric, it is always possible that dividing one of n_m modules of the global optimal segmentation generates modules with the same attributes: i.e., $\mathbf{a}_p = \bar{a}(\mathrm{M}_m)$, thus null values of summation of distances $|\mathbf{a}_p = \bar{a}(\mathrm{M}_m)|_k$ in Equation (7). Therefore, Equation (7) does not strictly hold for *attribute*-oriented metric. In any case, it can be argued that Equation (7) is a 'soft barrier' to the identification of small modules also for the *attribute*-oriented metric.

Similarly to the WDN-oriented modularity in Giustolisi & Ridolfi (2014b), it is effective to define the *attribute*-oriented *infrastructure index* adding the term $(n_m - 1)/n_p$

$$IQ_a(\mathbf{a}_p) = 1 - \frac{n_c - (n_m - 1)}{n_p} - \sum_{m=1}^{n_m} \left[\frac{\sum_{k=1}^{n_p} \left|\mathbf{a}_p - \bar{a}(\mathrm{M}_m)\right|_k \delta(\mathrm{M}_m, \mathrm{M}_k)}{\sum_{k=1}^{n_p} \left|\mathbf{a}_p - \bar{a}(\mathrm{N})\right|_k} \right]^2 \quad (8)$$

In fact, Equation (7) becomes

$$\sum_{m=1}^{n_m} \left[\frac{\sum_{k=1}^{n_p} \left|\mathbf{a}_p - \bar{a}(\mathrm{M}_m)\right|_k \delta(\mathrm{M}_m, \mathrm{M}_k)}{\sum_{k=1}^{n_p} \left|\mathbf{a}_p - \bar{a}(\mathrm{N})\right|_k} \right]^2 - \sum_{m=1}^{n_m+1} \left[\frac{\sum_{k=1}^{n_p} \left|\mathbf{a}_p - \bar{a}(\mathrm{M}_m)\right|_k \delta(\mathrm{M}_m, \mathrm{M}_k)}{\sum_{k=1}^{n_p} \left|\mathbf{a}_p - \bar{a}(\mathrm{N})\right|_k} \right]^2 > 0 \quad (9)$$

Thus, the modification of the *attribute*-oriented index is beneficial for the identification of small modules and also has the same effective features for multi-objective optimal WDN segmentation discussed in Giustolisi & Ridolfi (2014b). (i) IQ_a is conflicting with the objective functions that increase monotonically with the number of cuts (modules), thus making it appropriate for providing a solution to the cost–benefit optimization problems of WDN segment design. (ii) Maximizing IQ_a results in the identification of modules that are actually nested in those achievable maximizing the *attribute*-based index Q_a (Equation (6)). (iii) The possibility of increasing the resolution starting from any number of existing modules makes IQ_a not influenced by existing (even non-optimal) segmentations.

SOME SIMPLE EXEMPLIFYING NETWORKS

To discuss the features of the infrastructure modularity metrics, it is helpful to report them all together as follows:

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

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

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

$$IQ_a = Q_a + \frac{n_m - 1}{n_p} = 1 - \frac{n_c - (n_m - 1)}{n_p} - \sum_{m=1}^{n_m} \left[\frac{\sum_{k=1}^{n_p} \left|\mathbf{a}_p - \bar{a}(\mathrm{M}_m)\right|_k \delta(\mathrm{M}_m, \mathrm{M}_k)}{\varepsilon + \sum_{k=1}^{n_p} \left|\mathbf{a}_p - \bar{a}(\mathrm{N})\right|_k} \right]^2 \quad \text{(d)} \quad (10)$$

where ε is a small value (e.g., the value of the precision of the 32 bit computing environment 2.2204×10^{-16}) which is useful in order to avoid a null denominator when the values in $\mathbf{a}_p$ are equal (i.e., $\mathbf{a}_p = \bar{a}(\mathrm{N})$).

Equation (10(a)) and (10(b)) define metrics allowing the division of the network into modules which are similar to each other according to the internal sum of a vector of pipe *weights* which are user-defined in $\mathbf{w}_p$. The metric of Equation (10(b)) overcomes the resolution limit barrier in identifying small size modules during optimal segmentation.

Differently, Equation (10(c)) and (10(d)) are metrics allowing the division of the network into modules whose internal *attributes* are similar to each other according to their distance from the mean value. The pipe *attributes* are user-defined in $\mathbf{a}_p$. The metric of Equation (10(d)) generally overcomes the resolution limit barrier in identifying small size modules during optimal segmentation search.

It is worth recalling that we here distinguish pipe *attributes* from *weights*. For example, the pipe lengths can be seen as *weights* when we sum them in the case of the metrics

Equation (10(a)) and (10(b)), while they become *attributes* when we use them in order to segment the network modules with the same internal characteristics using the statistical distance from the mean value.

We refer to the two simple networks shown in Figure 1. The system in Figure 1(a) is a linear network (i.e., fully branched) composed of 48 pipes (n_p) and 49 nodes (n_n) delivering water from a reservoir. The system in Figure 1(b) is derived from that in Figure 1(a) closing the odd pipes around a loop. It is composed of 24 couples of loops connected by one pipe, 192 pipes (n_p) and 145 nodes (n_n).

Case study I: *IQ* vs. *Q* using the linear network

The multi-objective segmentation is performed on the network in Figure 1(a) using the metrics defined in Equation (10(a)) and (10(b)) assuming vector $\mathbf{w}_p$ equal to the identity vector. Therefore, the segmentation is based on topology and the solution using Equation (10(c)) and (10(d)) provides the trivial segmentation, i.e., the network is already a module with the same internal, constant, attributes $\mathbf{a}_p = \mathbf{w}_p$.

In fact, the segmentation based on the maximization of IQ vs. minimization of the number of cuts n_c already provides the optimal values of Q because IQ is a metric shifted from Q by means of the term $(n_m - 1)/n_p$ depending on the number of modules n_m (Giustolisi & Ridolfi 2014b).

Figure 2 shows that the maximum resolution of the infrastructure modularity IQ corresponds to 48 modules, each composed of one pipe as in Figure 3(a), obtained with the minimum number of cuts, i.e., 47 $(=48-1)$, because the network is fully branched. The value of IQ (=0.979) depends on the number of pipes (i.e., modules). It is interesting to note that the segmentation considering

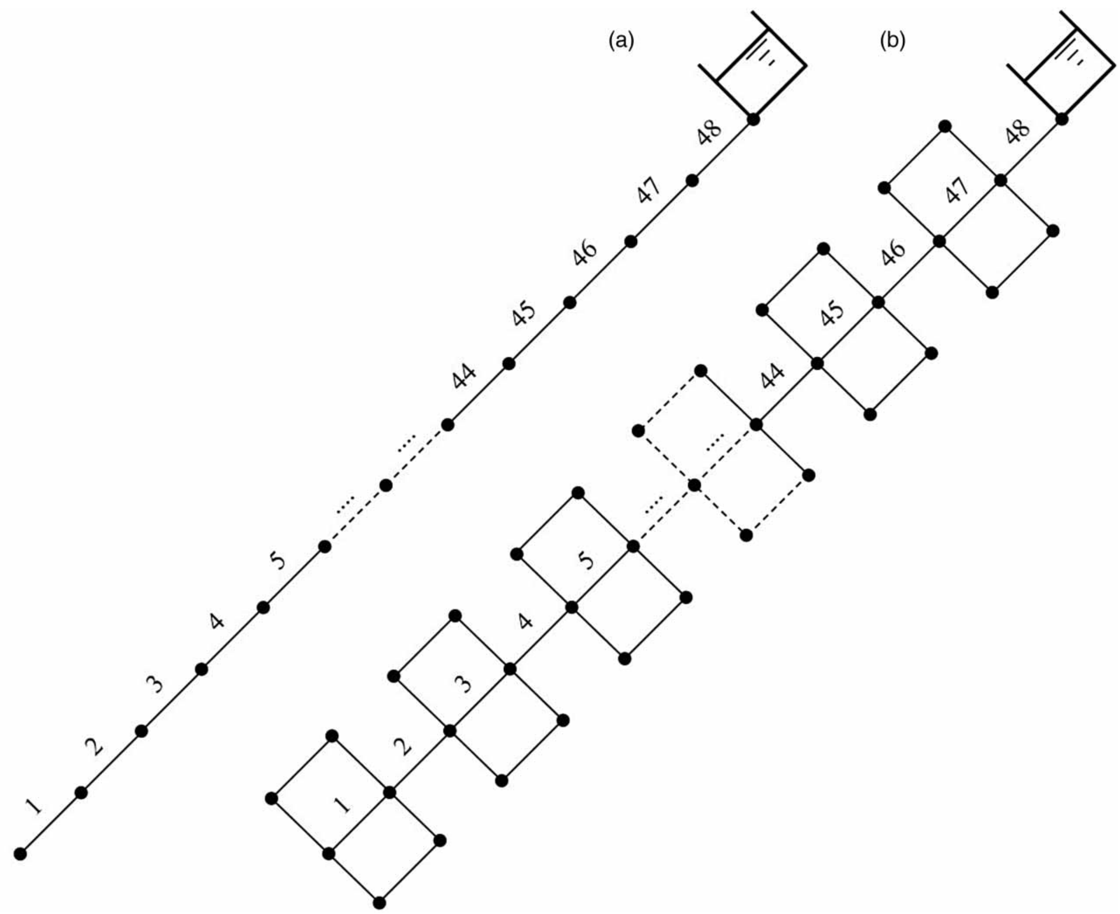

Figure 1 | (a) Linear network, i.e. fully branched system; (b) sequence of loops connected by one pipe, i.e., looped system.

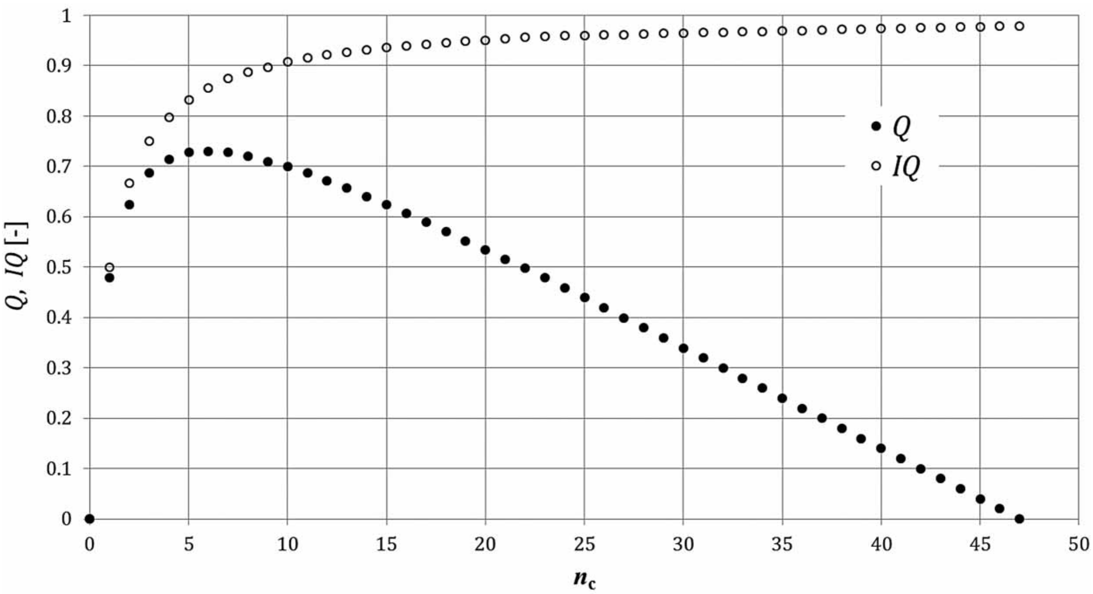

Figure 2 | Optimal segmentation max(*IQ*) vs. min(n_c) of the linear network. *Q* is the lowest curve.

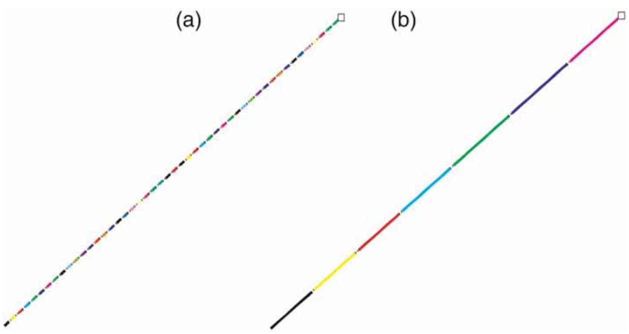

Figure 3 | Maximum number of modules: (a) metric *IQ*; (b) metric *Q*. The colors identify the modules. The full color version of this figure is available online at http://www.iwaponline.com/jh/toc.htm.

Q allows the division of the network into seven modules with six cuts (see the maximum value of *Q* in Figures 2 and 3(b)), because the resolution limit generates a mathematical barrier for the identification of smaller modules.

In summary, the exercise shows that the *IQ* does not have the resolution limit to allow the identification of any module generated by one further cut in the network, while *Q* has a strong resolution limit increasing with the number of pipes.

It is worth noting that the segmentation solution with seven modules and six cuts, corresponding to the maximum value of metric *Q* (Figure 3(b)), is the same as the solution achievable with six cuts using the metric *IQ*.

Case study II: *IQ* vs. *Q* using the looped network

This test is similar to the previous one but applied to the network of Figure 1(b). Figure 4 shows that the maximum of the curve *Q* corresponds to 12 cuts while *IQ* corresponds to 24 cuts.

Figure 5(a) and 5(b) show the maximum division into 25 modules with 24 cuts considering *IQ* and into 13 modules with 12 cuts considering *Q*. *IQ* allows each couple of loops to be divided while *Q* does not. This fact is a different confirmation of the effectiveness of *IQ* in identifying modules overcoming the resolution limit also in the presence of loops.

In summary, the case study demonstrates the effectiveness of the modularity of Equation (10(b)) to separate modules overcoming the modularity of Equation (10(a)) also in looped networks.

Case study III: IQ_a vs. Q_a using both the networks

Here, we perform two tests about the linear and the looped networks of Figure 1, but the metric IQ_a of Equation (10(d)) is applied instead of that in Equation (10(b)). To this purpose the selected attribute in $\mathbf{a}_p$ are the pipe diameters. Dummy diameters in the range [1, 12] are assumed. They are

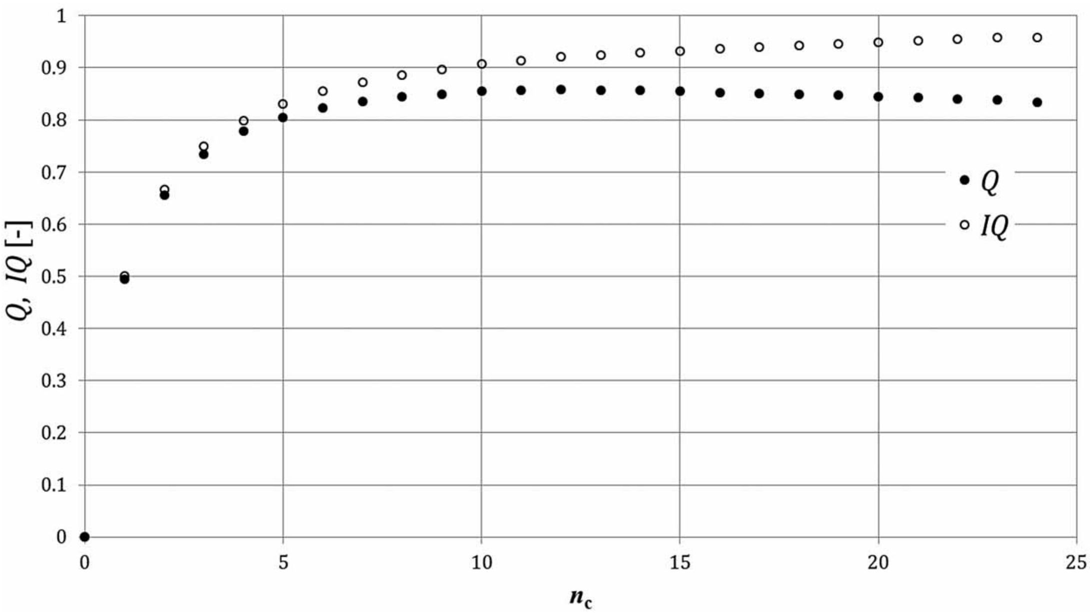

Figure 4 | Optimal segmentation max(*IQ*) vs. min(n_c) of the looped network. *Q* is the lowest curve.

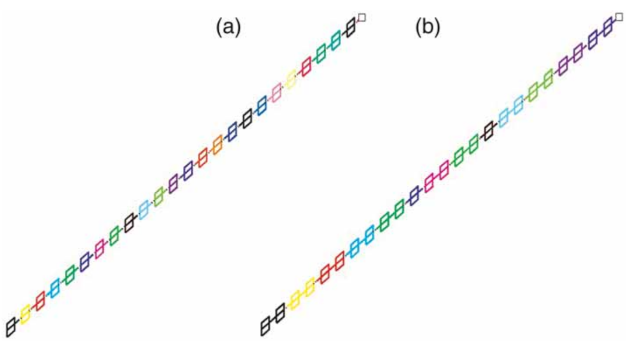

Figure 5 | Maximum number of modules: (a) metric *IQ*; (b) metric *Q*. The colors identify the modules. The full color version of this figure is available online at http://www.iwaponline.com/jh/toc.htm.

constant for each of four contiguous pipes starting from one ending node of the linear network. The looped network has the diameters of the loops constant and equal to that of the basis linear network.

Figure 6(a) and 6(b) show that the maximum number of modules of IQ_a is much higher than Q_a demonstrating the effectiveness of adding the term $(n_m - 1)/n_p$ also to Q_a. In addition, the maximum number of cuts of IQ_a is 11 which corresponds to 12 modules in both cases, as better reported in Figures 7(a) and 8(a). Therefore, IQ_a is effective to identify each group of identical pipes both in linear and in looped networks while IQ would identify the same 24 modules described in cases I and II since the topological condition prevails.

Figures 7(b) and 8(b) shows that Q_a is unable to identify the 12 groups of pipes because of the resolution limit occurring also for the attribute-based modularity of Equation (10(c)). It follows the effectiveness of the modularity of Equation (10(d)) to separate modules with the same diameters, overcoming the modularity of Equation (10(c)).

In conclusion, the three case studies demonstrate that both the metrics – namely, the pipe *weight*- and pipe *attribute*-based indices – are enhanced by adding the term $(n_m - 1)/n_p$. Furthermore, case study III demonstrates the effectiveness of the *attribute*-based metric to separate modules based on the assumed pipe characteristics.

TOWN-C CASE STUDY

This section compares the results of segment design achievable by using the *attribute*-based index (Q_a) and the *attribute*-based infrastructure index (IQ_a) metrics on the TOWN-C water distribution network. This network is

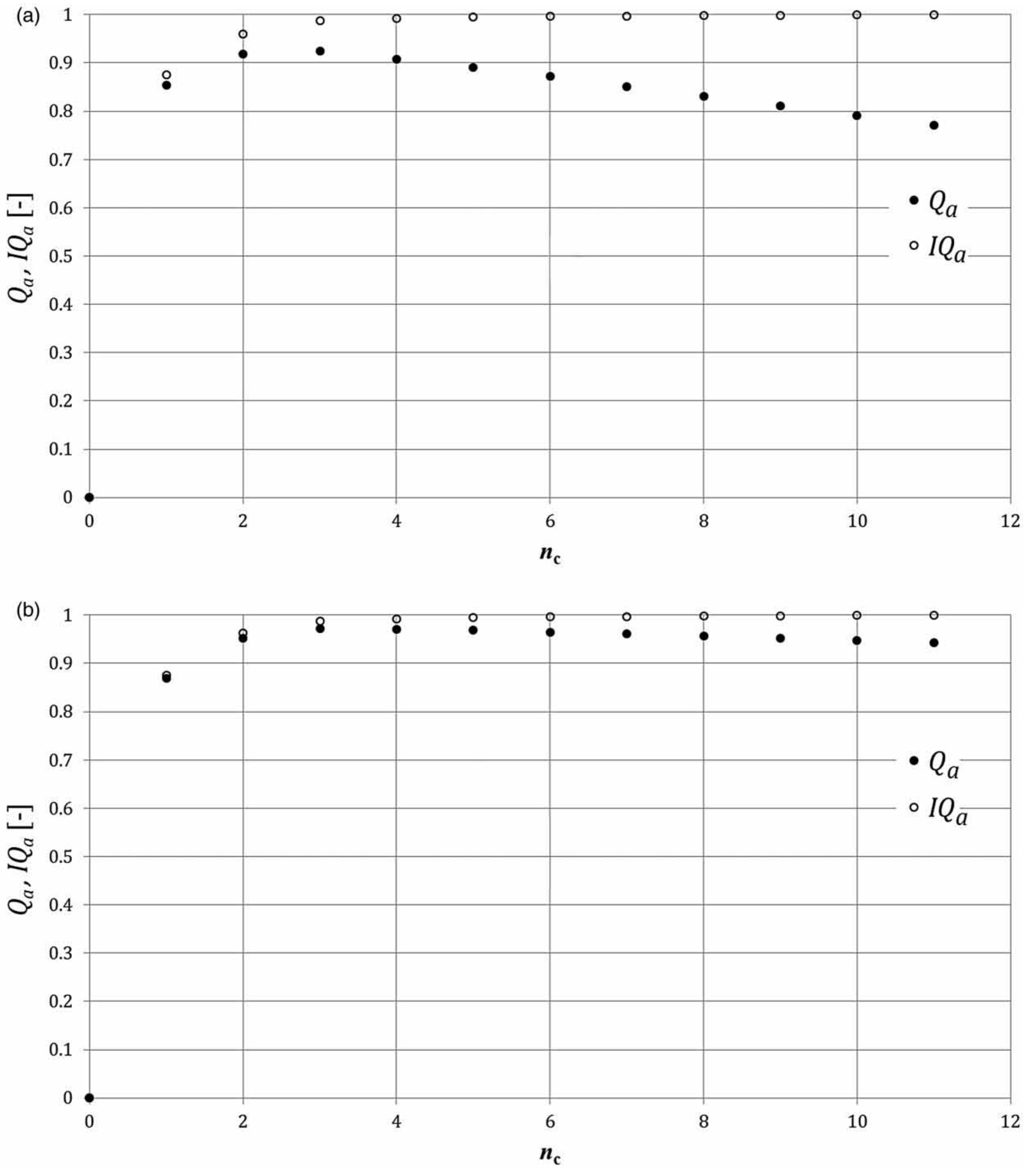

Figure 6 | Optimal segmentation max(IQ_a) vs. min(n_c): (a) linear network; (b) looped network.

chosen because it is well-known in the technical literature (data are available as supplementary material of Ostfeld *et al.* (2012)) and its layout allows segments and relevant cuts to be clearly visualized. In addition, the presence of hydraulic devices (i.e., pumping stations and tanks) already installed in the network results in a scenario that is closer to the real context of designing segments into an existing system.

The analysis considers two attributes focusing on different technical purposes of WDN segmentation: the pipe diameter and the average elevation.

As for previous examples, the segmentation solutions descend from the two-objective optimization where the chosen WDN-oriented modularity index (IQ_a or Q_a) should be maximized with the minimum number of cuts

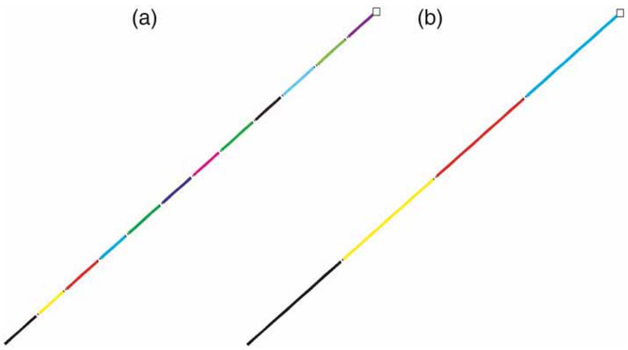

Figure 7 | Maximum number of modules: (a) metric IQ_a; (b) metric Q_a. The colors identify the modules. The full color version of this figure is available online at http://www.iwaponline.com/jh/toc.htm.

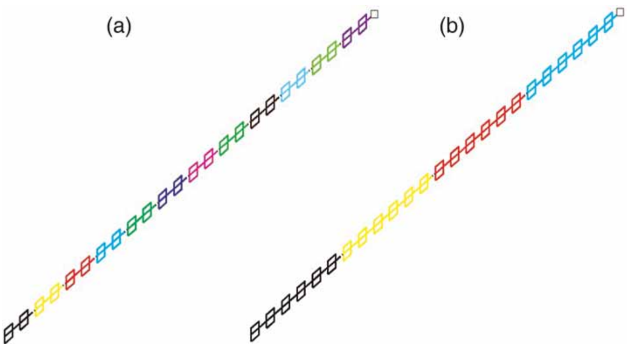

Figure 8 | Maximum number of modules: (a) metric IQ_a; (b) metric Q_a. The colors identify the modules. The full color version of this figure is available online at http://www.iwaponline.com/jh/toc.htm.

(devices) (see Giustolisi & Ridolfi (2014a) for optimization problem formulation).

TOWN-C case (a): diameter-based metrics

Identifying WDN modules composed of homogeneous pipe diameters has technical relevance from both asset management and hydraulic modeling purposes. In fact, pipe diameters generally reflect the hydraulic functioning ranging from larger trunks, mainly to transport water (e.g., from the water sources), to smaller pipes, mainly used to distribute water to users (e.g., in the peripheral WDN areas). Thus, looking for modules composed of similar pipe diameters is expected to return WDN portions with different preeminent hydraulic functioning. This information can be useful for many practical purposes including, for example, the selection of candidate location of gate valves to isolate WDN portions where water is distributed (e.g., in case of malfunctioning) without affecting the main transport lines.

In addition, internal hydraulic resistances of pipes strongly depends on pipe diameters (i.e., on power 5 based on Darcy-Weisbach head loss formulation), thus it is of direct relevance for WDN model calibration purposes. Grouping pipes that are expected to show similar values of hydraulic resistance per unit length is a pragmatic way to reduce the number of unknowns of the calibration problem by introducing technical/engineering insight. Identifying WDN modules with homogeneous pipe diameters allows preservation of information on the network topology, returning modules with similar and contiguous pipes; thus, going beyond strategies for clustering the pipe database only. From such perspectives, the cuts that separate modules represent the most effective points to allocate pressure/flow sampling devices in order to maximize the observability of pipe hydraulic resistances (Giustolisi & Berardi 2011).

Figure 9(a) and 9(b) report WDN segments obtained by minimizing the number of cuts (e.g., pressure/flow meters) and maximizing the index Q_a (a) and IQ_a (b) using pipe diameters as attribute. The metric Q_a is maximized for 18 modules and 18 cuts. It is evident (see Figure 10(a)) that links representing pumps (i.e., with the same diameter) are in the same modules, as expected. In addition, the largest trunks, mainly used to transport water from tanks/reservoirs, are identified as belonging to the same modules; while some branched WDN portions, mainly used for water distribution to customers, are identified as separate segments. Nonetheless, in some cases the identified modules are visibly composed of portions with different diameters that should be separated from the rest of the network; this is the case of some branches connected with the upstream looped network portions.

Figure 9(b) shows that maximizing the *attribute*-oriented infrastructure index IQ_a allows this resolution limit to be overcome, returning up to 83 modules. It is worth noting that, consistently with the expected WDN hydraulic behavior, three types of sub-modules are identified: (i) branched portions separated by one cut only; (ii) looped inner portions, separated by multiple cuts from the rest of the WDN; (iii) transport modules, linking the WDN to tanks/reservoir or to other distribution modules.

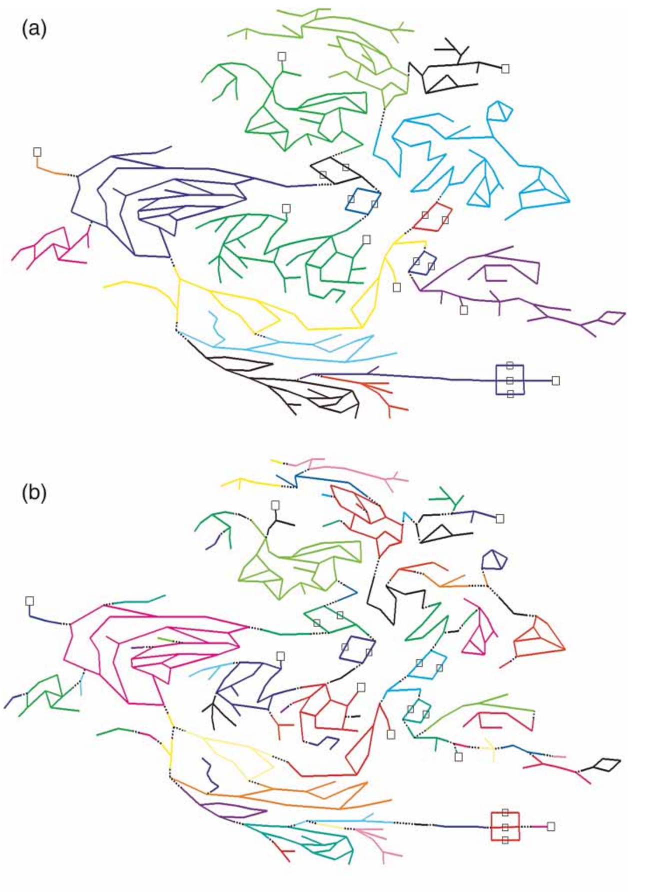

Figure 9 | TOWN-C segmentation based on diameter-based modularity metric Q_a (a) and diameter-based infrastructure metric IQ_a (b).

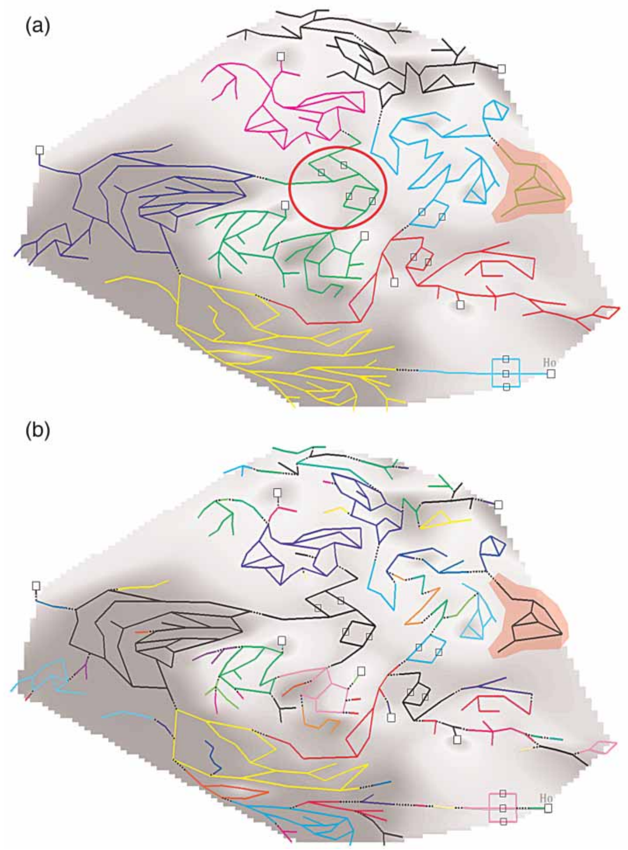

Figure 10 | TOWN-C segmentation based on *elevation*-based modularity metric Q_a (a) and *elevation*-based infrastructure metric IQ_a (b).

From a WDN model calibration perspective, the layout of modules achievable by maximizing IQ_a is consistent with the general criteria for collecting pressure/measurement at nodes where such measures are maximally informative for the calibration of variables. This is the case of many cuts separating branches, since the measurements collected at these points actually maximize the topological observability of the unit hydraulic resistance for homogeneous pipes in that module (Walski 1983; Giustolisi & Berardi 2011).

TOWN-C case (b): elevation-based metrics

Elevation is an essential driver to model and run water distribution systems fed by gravity or by pumps, since differences in elevations affect pressure regimes and, in turn, the capacity of satisfying water demand as well as the leakage outflow from pipes. Accordingly, considering average pipe elevation as the attribute for identifying modules has two technical purposes. On the one hand, this is a way to design WDN modules that are expected to experience similar pressure regime. On the other hand, due to the multi-objective strategy used to design modules, resulting solutions entail the minimum number of cuts separating WDN zones with homogeneous average elevation (and expected pressure regime) inside the modules but different elevation between contiguous modules. Thus, cuts are suited to representing the optimal location of pressure control devices (e.g., PCV) that are usually required to be the fewest possible, due to their cost of installation and maintenance. Although minimizing such devices (i.e., cuts) directly reflects economic criteria of water utilities, increasing the resolution of elevation-based modules is expected to provide additional information on the most suitable candidate nodes to be used as pressure set points of PCV (i.e., controlled by remote pressure readings).

Figure 10(a) reports the 11 modules entailing the maximum *attribute*-based index Q_a (i.e., Equation (10(c)) achievable with the minimum number of cuts (equal to 10). It is evident that modules correspond to differences in elevation and in some cases (shadowed in Figure 10(a)) cuts indicate feasible candidate location of pressure reduction valves to control branched WDN portions. Nonetheless, the resolution limits of Q_a result in some inconsistencies such as, for example, pipes joining tanks (i.e., at higher elevation) still belong to modules that have lower average elevation.

When the *attribute-based infrastructure index* IQ_a, i.e., Equation (10(d)) is used, the increased resolution identifies 88 modules. It is evident that a number of branches are identified as different modules. However, this happens only if branches serve areas with different elevation from the nearest upstream WDN module. In fact, in some cases, like the shadowed module in Figure 10(b), the differences in elevation are not statistically significant to justify the creation of a new module; thus this module is the same as that of Figure 10(a) (i.e., based on Q_a). A similar behavior occurs in the area near to tank T1 that is located on a flat area, where the increased resolution allows identification of branches as separate modules while the main looped WDN portion still belongs to the same module.

It should be mentioned that tanks and reservoir are all located in modules apart from the rest of the network, consistently with their different elevation and WDN hydraulic behavior. In addition, the long pipeline connecting the main pumping station (near to the reservoir) to the rest of the network is divided into three modules since its elevation drops from 56 m to about 12 m above sea level.

CONCLUSIONS

The modularity concept has been recently borrowed from complex network theory to infrastructure networks and has been tailored for WDN analysis and management. The work by Giustolisi & Ridolfi (2014a) introduced the base formulation for WDN-oriented modularity indexes that were aimed at matching the peculiarities of WDN infrastructure as well as the technical meaning of pipes and modules.

The *weight*-based modularity index was introduced to maximize the similarity of modules with each other. Starting from the *weight*-based modularity index, Giustolisi & Ridolfi (2014b) proposed an *infrastructure modularity index* in order to overcome the resolution limit that resulted from the original formulation of the classic modularity index upon which the WDN-oriented modularity was developed.

In addition, the same authors introduced an *attribute*-based index that was suited to identify modules with the maximum similarity of the attributes within each module. This means that returned modules do not simply entail groups of similar pipes through the network, but implicitly preserve the information on WDN topology (i.e., modules with similar and contiguous pipes), which is essential for an infrastructure management perspective.

The present contribution demonstrates that also the *attribute*-based index can suffer from resolution limit and extend the concepts of *infrastructure index* variant also to *attribute*-based index.

The infrastructure segmentation metrics are better suited to identify segments in a multi-objective optimization paradigm, where the *weight*-based and *attribute*-based indices should be maximized while minimizing the number of cuts (i.e., costly devices). Also, it was found that the modules identified using the Q_a index are identified when the IQ_a is used; this means that the additional modules identified by maximizing IQ_a are actually nested in the previous ones.

The proposed didactical examples demonstrate the advantages in identifying modules by using the *infrastructure* metrics IQ and IQ_a instead of the *weight*- and *attribute*-based indexes Q and Q_a, respectively. Finally, the well-known TOWN-C literature network is used to discuss the practical implications of the increased resolution of segmentation by using the *attribute*-based infrastructure index IQ_a. Pipe diameters and average elevations were selected as pipe attribute, respectively. In all cases, the resolution limits that are typical of Q_a clearly prevent identifying modules that are more suited for possible final technical purposes of the segmentations (e.g., ranging from WDN model calibration to pressure control and leakage detection plans). Differently, IQ_a overcomes such limits and gives a very detailed and technically sound network segmentation.

ACKNOWLEDGEMENT

The research reported in this paper was founded by the Italian Scientific Research Program of National Interest 'Tools and procedures for an advanced and sustainable management of water distribution systems' – PRIN2012 (Prot. 20127PKJ4X) – Italian Ministry of University and Research (MIUR).

REFERENCES

Albert, R. & Barabasi, A. L. 2002 Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74** (1), 47–97.

Alvisi, S. & Franchini, M. 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.

Davidson, J., Bouchart, F., Cavill, S. & Jowitt, P. 2005 Real-time connectivity modelling of water distribution networks to predict contamination spread. *J. Comput. Civil Eng.* **19** (4), 377–386.

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

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

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

Fortunato, S. & Barthélemy, M. 2007 Resolution limit in community detection. *Proc. Nat. Acad. Sci. USA* **104**, 36–41.

Giustolisi, O. & Berardi, L. 2011 Water distribution network calibration using enhanced GGA and topological analysis. *J. Hydroinform.* **13** (4), 621–641.

Giustolisi, O. & Ridolfi, L. 2014a A new modularity-based approach to segmentation of water distribution networks. *J. Hydraul Eng.* **140** (10), 04014049.

Giustolisi, O. & Ridolfi, L. 2014b A novel infrastructure modularity index for the segmentation of water distribution networks. *Water Resour. Res.* **10**, 7648–7661.

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

Newman, M. E. J. 2010 *Networks: An introduction*. Oxford University Press, Oxford, UK.

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

Ostfeld, A., Salomons, E., Ormsbee, L., Uber, J. G., Bros, C. M., Kalungi, P., Burd, R., Zazula-Coetzee, B., Belrain, T., Kang, D., Lansey, K., Shen, H., Mcbean, E., Wu, Z. Y., Walski, T., Alvisi, S., Franchini, M., Johnson, J. P., Ghimire, S. R., Barkdoll, B. D., Koppel, T., Vassiljev, A., Kim, J. H., Chung, G., Yoo, D. G., Diao, K., Zhou, Y., Li, J., Liu, Z., Chang, K., Gao, J., Qu, S., Yuan, Y., Prasad, T. D., Laucelli, D., Vamvakeridou Lyroudia, L. S., Kapelan, Z., Savic, D., Berardi, L., Barbaro, G., Giustolisi, O., Asadzadeh, M., Tolson, B. A. & Mckillop, R. 2012 The battle of the water calibration networks (BWCN). *J. Water Res. Plann. Manage.* **138** (5), 523–532.

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

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

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

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

Yazdani, A. & Jeffrey, P. 2012 Water distribution system vulnerability analysis using weighted and directed network models. *Water Resour. Res.* **48** (6), W06517.

First received 10 October 2014; accepted in revised form 4 February 2015. Available online 28 February 2015