A flexible methodology to sectorize water supply networks based on social network theory concepts and multi-objective optimization

Enrique Campbell, Joaquín Izquierdo, Idel Montalvo, Amilkar Ilaya-Ayza, Rafael Pérez-García and Mario Tavera

ABSTRACT

A novel methodology to sectorize water supply networks (WSNs) depending on a main transmission line is presented in this paper. The methodology is based on concepts derived from the social network theory and graph theory (namely, community detection and shortest path respectively); and also on a multi-objective optimization process by means of agent swarm optimization (ASO). A series of energy, operative, and economic criteria are optimized in this process. The core idea is to form sectors over the distribution network based on communities found using a community detection algorithm (Walktrap). The methodology is flexible and enables the technical staff in water utilities to make decisions at different stages. It has been tested by generating four feasible solutions over a portion of a real WSN.

Key words | community detection, optimization, reliability, sectorization, sectors

Enrique Campbell (corresponding author)
Joaquín Izquierdo
Amilkar Ilaya-Ayza
Rafael Pérez-García
Fluing-Instituto de Matemática Multidisciplinar
    (IMM)-Universitat Politècnica de València,
Camino de Vera SN,
46022, Valencia,
Spain
E-mail: encamgo1@upv.es

Idel Montalvo
3SConsult GmbH,
Albtalstraße 13,
76137, Karlsruhe,
Germany

Mario Tavera
MIYA-WATER,
21st Century Road, Nassau,
Bahamas

ABBREVIATIONS AND NOTATION

ASPV	accumulated shortest path value
ASO	agent swarm optimization
BFS	breadth first search
DN	distribution network
E	elevation
Q	flow/demand
H	head
h_L	head loss
P_{min}	minimum pressure target
P_n	nodal power
P_o	operational power
P_C	penalty cost
P	pressure
I_r	resilience index
SP	shortest path
γ	specific weight of water
σ_P	standard deviation of pressure
P_{total}	total available power

TN	trunk network
SPV	shortest path value
WSN	water supply network

INTRODUCTION

Water supply network (WSN) segmentation and/or sectorization can be described as a technique for enhancing the operative control of WSNs, including non-revenue water management, pressure control, and chemical quality monitoring. The technique implies the definition of areas that can be partially isolated from the rest of the network. Such isolation is carried out by closing some installed valves (depending on their current conditions), installing new valves, or cutting some pipes. Each of the resulting sectors is then fed by a single (or several) entrances equipped with a flowmeter, which enables the inflow to be

permanently monitored. Since the 1980s, when the first sectorization implementation was reported in British cities (Farley 2010), the technique has become very popular, especially among Latin American and European water utilities.

Despite the benefits associated with the deployment of a sectorization layout, there are also some drawbacks. From the outset, a pressure drop can be indicated as the most relevant drawback. When some pipes are closed, the general available flow area is reduced and this increases the head loss due to pipe friction, which translates into an overall pressure drop. Although a pressure reduction can be positive (given its relation with reductions in background leakage flow, frequency of appearance of new bursts and breaks, and domestic consumption (Thornton & Lambert 2005, 2006)), an excessive reduction in pressure can also be translated into supply incapacity. Moreover, a sectorization project implementation also implies the investment of economic resources which may sometimes be scarce.

Based upon the above aspects, it can be stated that a good sectorization layout has to find a good compromise among a series of positive and negative aspects. The number of possible layouts may be large, depending on the WSN extension and its looping configuration. In some cities with small and/or highly branched WSNs, it is feasible to define sectors empirically (based only on network maps) and verify them through hydraulic simulation in any hydraulic simulation package; however, in more complex contexts (the case of most large cities) with a very large number of options, the use of mathematical and computational tools becomes necessary.

Since the publication of the *District Metering Areas: Guidance Notes* by members of the Water Loss Task Force of the International Water Association (Morrison *et al.* 2007), several methodologies have been proposed to tackle the problem of WSN sectorization. Most of these studies have addressed the problem using graph theory and energy criteria (Tzatchkov *et al.* 2008; Di Nardo & Di Natale 2011; Di Nardo *et al.* 2014; Alvisi & Franchini 2014) in combination with multi-agent systems (Izquierdo *et al.* 2011; Herrera *et al.* 2012; Hajebi *et al.* 2013), optimization approaches (Gomes *et al.* 2012; Di Nardo *et al.* 2013; Hajebi *et al.* 2014; De Paola *et al.* 2014), and, very recently,

with concepts derived from social network theory (Diao *et al.* 2013; Campbell *et al.* 2014a; Giustolisi & Ridolfi 2014).

Sectorization can be addressed in two different manners depending on the topology of the WSNs. When WSNs rely on several sources within the network, sectors can be set around the sources. In this case, sectors may count on one (or several) exclusive sources. However, in other contexts, sources are far from users, and supply depends on a main transmission line (trunk network).

This work describes a flexible methodology tailored to the latter type of network. The methodology encompasses a disaggregation of the WSN into two components, namely, a trunk network (TN) and a distribution network (DN). Disaggregation is carried out in order to separate the TN pipes from the DN pipes, so the TN pipes are not included within the sectors, reducing any likely negative impact of the sectorization implementation over the DN, and enhancing leakage control (as leakage/flow monitoring is focused on the pipes of the DN where the frequency of leakage appearance is higher). The sectors that meet the established size constraints are defined over the DN. Finally, the entrance (or entrances) to each sector is determined by means of an optimization process with agent swarm optimization (ASO) (Montalvo *et al.* 2014). The water utility staff must take decisions based on their experience in three steps of the process: (1) definition of the trunk network; (2) selection of the sector size; and (3) selection of a feasible solution. These decisions define the final outcome of the whole process.

This paper is an extension of the contribution presented at the WDSA 2014 (Campbell *et al.* 2014b).

GRAPH THEORY APPLIED TO WATER SUPPLY NETWORKS' SECTORIZATION

Figure 1 depicts the proposed methodology. Its components are detailed in the following sections.

Graphs can be described as structures topologically formed by nodes interconnected by links (also known as edges – in this paper, the two terms are used interchangeably). Graph theory is a branch of mathematics dedicated to the study of such structures. The topology of the WSNs coincides with the topology of their associated graphs.
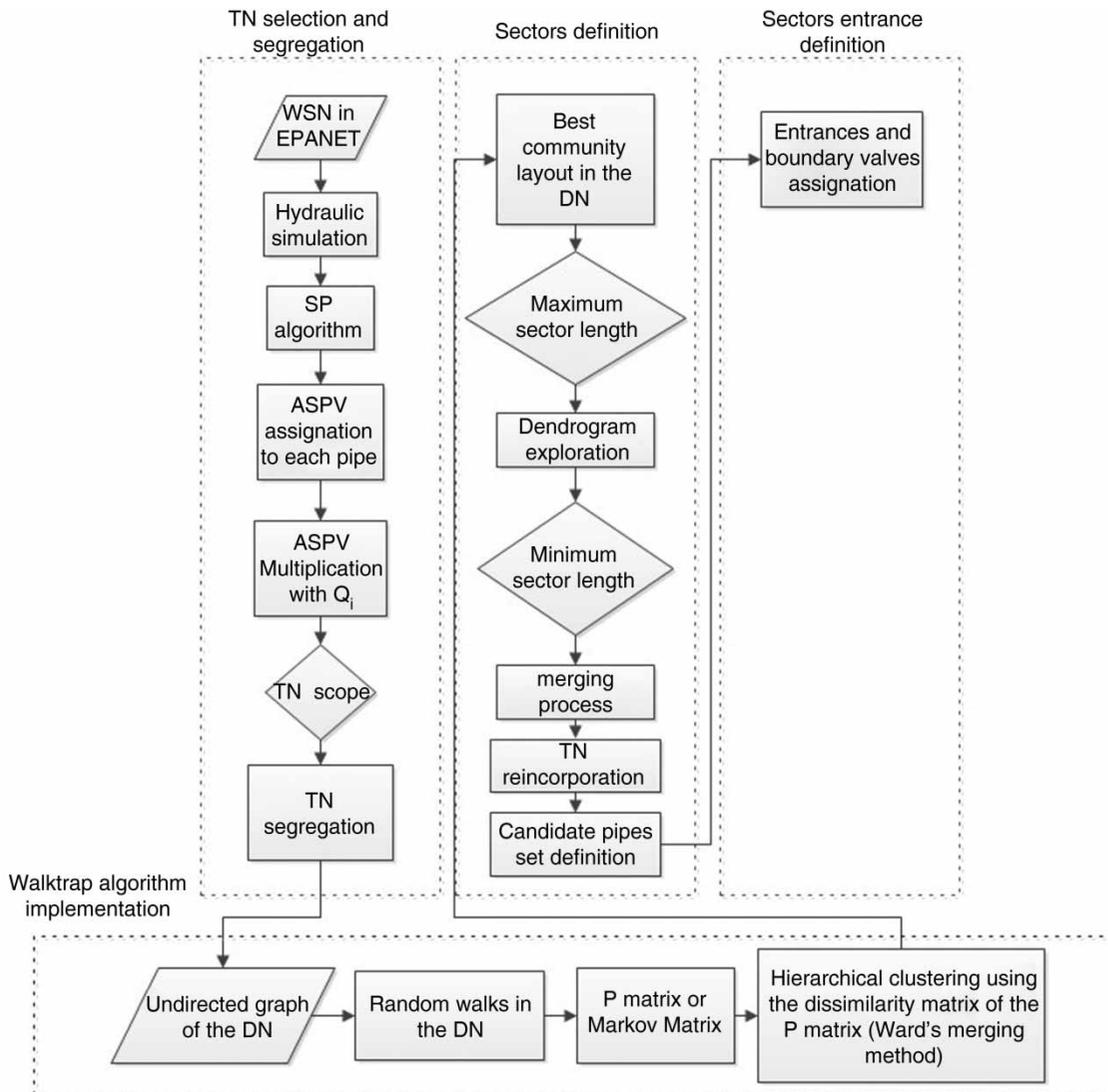
Figure 1 | Flowchart of the proposed methodology.

Therefore, WSNs can be abstracted as graphs, and the algorithms and concepts belonging to the domain of the graph theory can be implemented over them. In this work, the shortest path (SP) concept is implemented to define the TN, and a community detection algorithm is implemented to define structural communities which are used to define sectors.

The SP between nodes in a given network, or the path with minimum number of vertices, together with community detection, are two of the concepts most broadly explored within the domain of the graph and social network theory. By applying the SP concept over a digraph

(a graph with the direction of links initially set) it is possible to determine the frequency of all the possible paths in a graph. The information contained about the pipes of a given WSN can be incorporated in the edges of its graph, including the direction of flow in the pipes, in order to define the directions of the edges on the graph. Evidently, to know the flow direction, a hydraulic simulation is required. When such a hydraulic simulation (e.g., with EPANET) is carried out in the scenario of maximal demand (the point where the demand pattern curve is higher), the flow direction on each pipe corresponds to the one of the worst scenarios.

Once the directions of the edges on the graph are defined, it is possible to calculate a so-called shortest path value (SPV) between each pair of nodes. This is carried out through the widely known breadth first search (BFS) algorithm (Moore 1959) which performs a recursive search of every reachable node from a given starting node. Given a source node (any node and not only the water sources), BFS systematically explores all the vertices of the graph to find the subset of reachable nodes from that source. The distance (in terms of number of vertices) from the source to the entire set of reachable nodes is then computed. Finally, a tree is formed with the source node as the root and followed by all reachable nodes.

The SPV for each pair of nodes can be stored in a square matrix, where the nodes in the first column are the starting nodes and those on the top row are the ending nodes. The matrix is fulfilled according to the next set of rules:

if $SN = EN \rightarrow 0$

if SN cannot reach $EN \rightarrow \infty$

For any other case $\rightarrow$ number of nodes in the SP

where SN stands for starting node and EN stands for ending node.

It is important to highlight the fact that in the case of unweighted graphs, all the edges are assumed to have the same weight (1), and thus, the SP between any pair of nodes is assessed solely on the number of nodes.

The SP matrix enables ranking the nodes on their importance in the graph. A given node is considered more important the greater the number of nodes it can reach; this corresponds to the sum of every row in the former matrix. In this work, such a sum is called accumulated shortest path value (ASPV). This value can also be transferred to the links by assigning to each link a value that is equal to the ASPV on its corresponding downstream node + 1. Note that 1 is added in order to include the final node of every path, which always has 0 SPV. The greater the ASPV on a given link (or pipe), the greater its importance within the graph, and therefore, in order to maximize the reliability of the WSN, the given pipe should never be included between any sector. As the ASPV of some of the pipes might be considerably greater than the ASPV in other pipes, it is rescaled

by dividing the corresponding value of each pipe by the greatest value among all the pipes. In this way, the ASPV of the pipe connected to the main source would have a value equal to 1, and the pipes connected to the extreme nodes should have values near 0.

It is important to note that if the ASPV is used alone, some pipes with a small diameter and flow can be included among the TN if they are able to deliver water to the largest pipes. Nevertheless, if the ASPV is used as a factor that multiplies the flow that is conducted through each pipe, instead of an individual criterion, only pipes with large diameter and flow values are included in the TN. We note such a multiplication by ASPV*.

As a final outcome, a hierarchy of pipes is formed with the sources as roots. If there is more than one source, the pipes connected to each source are expected to be more important in the hierarchy depending upon the number of nodes that each is able to reach.

WSNs in most large cities are formed by blocks of looped networks connected to other blocks and to a TN (using just a few pipes). The pipes of the TN are also expected to be less frequent and to have the highest ASPV* values, whereas the pipes of the DN are expected to be more frequent and to have lower ASPV* (near 0 in the case of pipes that are connected to a node that only receives water). Based on this idea, in this paper, we propose the use of analysis of frequency to determine which lines are part of the TN and which are part of the DN. To do so, a frequency table is constructed with a very large number of ASPV* breaks, in order to obtain the smallest possible intervals. The relative frequency and percentage of cumulative frequency are then calculated. When the relative frequency and the cumulative frequency are plotted in a Pareto diagram it shows how there is a large percentage of pipes with low ASPV* (approximately 80% of the pipes). It is also evident that among the pipes with less frequent ASPV* (the remaining 20%), there are noticeable changes in the tendency of their intervals. These change points can be used as a segregation criterion and will produce several alternative scopes of the TN, from which the technical staff of the water utility can select the most convenient option. For example, the scope can be limited to the set of pipes that, if removed, cause supply problems. In a more general way, the scope may include all the pipes with values that are significantly less frequent.

The idea of marking, distinguishing, and/or decoupling the TN from the DN has been previously proposed in works related to the sectorization of WSNs (e.g., Di Nardo & Di Natale 2011; Diao *et al.* 2013; Campbell *et al.* 2014b; Hajebi *et al.* 2014). Some of these studies have suggested the use of parameters associated with the hydraulics of the network (flow, head loss, diameters), or the use of a measure known as *edge betweenness*, which characterizes the centrality of the pipes throughout the network (Fortunato 2010). Using only the edge betweenness parameter has a drawback: the pipes connected to the tanks would have lower values for this parameter, making the process more complex. Using only flow as the criterion to distinguish the TN from the DN may include among the TN several pipes that are delimited for a reduced area (not so relevant for the hydraulics of the system), while the use of diameters may work well on a logical design of WSNs. However, in several real situations, it is possible to find over-dimensioned or sub-dimensioned pipes. Based on all this, the ASPV proposed here can be considered a more robust and accurate parameter to distinguish the TN from the DN.

To exemplify the concept described above, it has been applied to a small section of the WSN of Managua City, the capital of Nicaragua. As can be seen in Figure 2(a), the interval values change drastically in the area where the percentage of cumulative frequency does not vary too much.

Figure 2(b)–2(d) show three alternative scopes of the TN. Figure 2(b) includes all the pipes with values of cumulative frequency below 85%. This corresponds to the less strict scenario. Figure 2(c) corresponds to the first change in the interval values. The removal of any pipe in this subset of pipes would cut the water supply for at least a portion of the network. Finally, Figure 2(d) represents the second change of interval values, this being the stricter scenario. The removal of any of the pipes in this subset would cut the water supply in the whole network.

COMMUNITY DETECTION ALGORITHM

As mentioned above, the sectorization of WSNs depending on a TN must be approached in a different manner than the sectorization of WSNs where the sectors are established
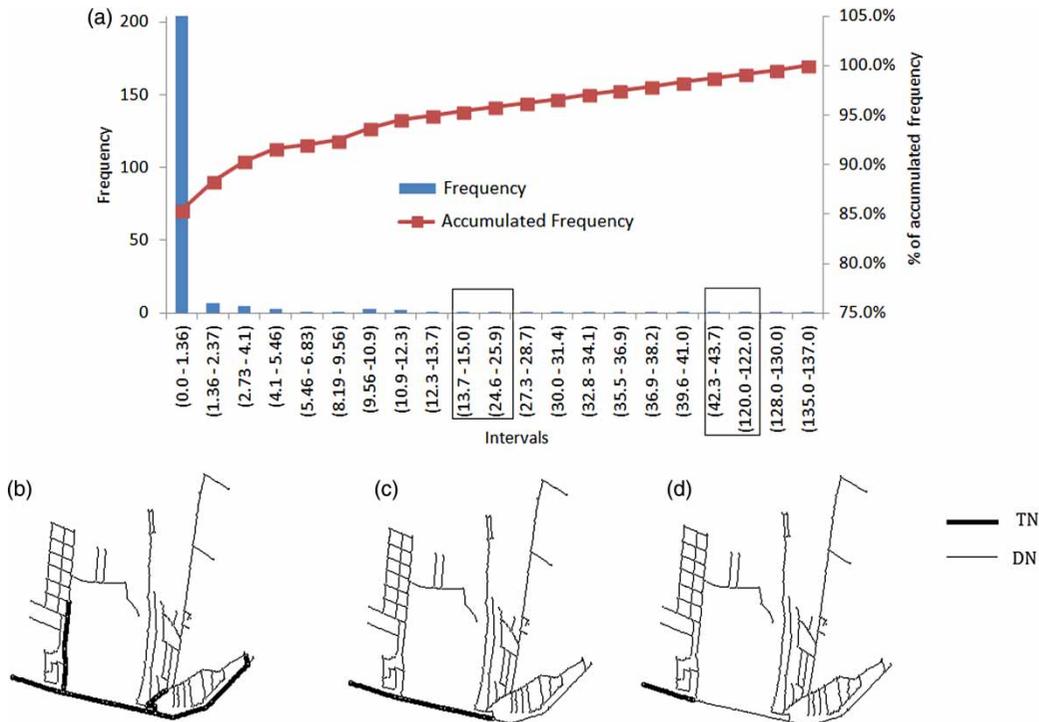

Figure 2 │ ASPV* frequency chart (Pareto diagram) and TN alternatives for the example network.

around individual water sources. In the first case, it is convenient to sectorize based on the topology of the network, which, as described below, is related to the concept of community detection in social networks.

Within graph theory, clustering techniques (or subdivision of sets of data into highly similar groups separated by few connections) are of major interest. These techniques can be subdivided into two subgroups: flat techniques (where the number of subdivisions is a priori defined (*k-means*, *clara*, spectral clustering) and an unrelated set of clusters is produced) and those techniques where there is no initial number of divisions (producing a hierarchy of related clusters – see Manning *et al.* (2008)). The main difference between community detection and clustering is that, while communities in graphs are related, explicitly or implicitly, to the concept of edge density (number of edges inside a cluster or community versus the number of edges that connect different communities), in data clustering, clusters are sets of multidimensional vectors that are 'close' to each other in distance or similarity (Newman & Girvan 2004; Newman 2006a; Fortunato 2010).

One of the most widely known algorithms to detect communities in graphs is known as *hierarchical clustering*. This algorithm works based on the construction of clusters following a hierarchical structure. It can be classified among the unsupervised machine learning methods, and is widely used for classification and pattern recognition purposes. The idea behind this method is to build pairwise groups that successively merge with other groups until all the nodes of the graph are part of the same group (when a bottom-up strategy is followed); or groups successively subdivide until all the clusters are formed by one cluster or singleton (when a top-down strategy is followed). All merges or separations may be depicted through a dendrogram, which is a diagram in an X-Y plan, where the X axis contains all the observations or vectors, and the Y axis contains a scale that establishes a numerical value for each partition. One of the advantages of this technique, with respect to other clustering techniques, is the fact that it is not mandatory to initially set a number of groups on which the data must be subdivided. In this sense, there are several indicators that can be used to determine the 'best partition'. The most widely known is the silhouette width index proposed by Rousseeuw (1987).

Despite the advantage that is represented by the fact that it is not necessary to initially predefine a number of clusters, the methods to find the most appropriate partitioning might not be suited to all situations. Most recent community detection algorithms tackle this problem by setting some quality partitioning indexes, the modularity index being the most broadly known (Newman 2006b).

In the fields of mathematics and probability theory, random walks are defined as stochastic processes where the position of a given particle at a certain instant relies solely on its position at a previous instant and on a random variable, which determines its subsequent direction and step length.

The Walktrap algorithm proposed by Pons & Lapaty (2006) is based on the idea that random walks in a graph tend to get 'trapped' into densely connected parts corresponding to communities. If random walks of a given length are preformed over a graph, the information in the P matrix (probability matrix, transition matrix, substitution matrix, or Markov matrix) about the P_{ij}^{k} probabilities to go from node i to node j in k steps is sufficiently long to gather enough information about the topology of the network.

In this algorithm, a distance measure among nodes and communities based on P is computed (Equation (1) and Equation (2), respectively):

$$r_{ij}(t) = \sqrt{\frac{\sum_{k=1}^{n}\left(P_{ik}{}^{t} - P_{jk}{}^{t}\right)^2}{d(k)}} \tag{1}$$

$$r_{c_1 c_2} = \sqrt{\sum_{k=1}^{n}\frac{\left(P_{c_1 k}{}^{t} - P_{c_2 k}{}^{t}\right)^2}{d(k)}} \tag{2}$$

where $d(k)$ denotes the degree of the k nodes.

This distance is used to obtain the dissimilarity matrix of the entire graph. Based on the pairwise comparison over this dissimilarity matrix, the first sector is built. To determine which sectors to merge, the widely known Ward's method (Ward 1963) is then used. In this method, the sum of square distances of each node in a given sector of the sector mean is established as an

objective function, which has to be minimized at each merging step:

$$\sigma_k = \frac{1}{n} \sum_{C \in \mathcal{P}_k} \sum_{i \in C} r_{iC}^2 \qquad (3)$$

where $\mathcal{P}_k$ is a k partitioning and r_i corresponds to the distance described above.

It is important to point out two aspects: first, this function depends only on each cluster and its minimization does not require information about other sectors; second, the method follows a greedy strategy; thus, each pair of adjacent sectors is merged and the variation $\Delta\sigma$ is computed. The merging that induces the lowest $\Delta\sigma$ is selected as the new partition.

At each step, the algorithm forms a new partition that minimizes $\Delta\sigma$, and so the final result is not one single partition, but a sequence of partitions at different levels (the levels of the dendrogram). According to Pons & Lapaty (2005), from this set of partitions, the one that is best adjusted to requirements can be selected (for example, length of the sectors). For this purpose, the same authors propose a so-called increase ratio (η_c), which calculates the fraction that represents the $\Delta\sigma_k$ of a given partition regarding the previous partition $\Delta\sigma_{k-1}$ (Equation (4)). This is evaluated for all levels in the dendrogram. The level of the highest η_c is defined as the partition that 'best' gathers

the community structure:

$$\eta_c = \frac{\Delta\sigma_k}{\Delta\sigma_{k-1}} \qquad (4)$$

The Walktrap algorithm has several advantages over other community detection algorithms, the first being related to the distance measure described above, which is efficiently computable and which can be used in a hierarchical partitioning algorithm. The community structure can be computed in time $\mathcal{O}(mnH)$ where m is the number of edges, n is the number of nodes, and H is the height of the dendrogram. In terms of comparison with other community detection algorithms, the tests performed by Orman *et al.* (2011) rank it as the second best community detection algorithm, just after the Infomap algorithm (Rosvall & Bergstrom 2008). The results obtained by Savić *et al.* (2012) when comparing it with three community detection algorithms – namely, Newman-Girvan (Newman & Girvan 2004), label propagation (Raghavan *et al.* 2007), and greedy modularity optimization (Clauset *et al.* 2004) – show the advantages of the algorithm in terms of community quality and evolutionary stability.

To show the application of this algorithm, it is implemented over the example network shown above. First, the TN is segregated, leaving only the DN. For this example, Figure 2(c) was selected as the scope of the TN. Figure 3(a) depicts the best partition (19 communities).
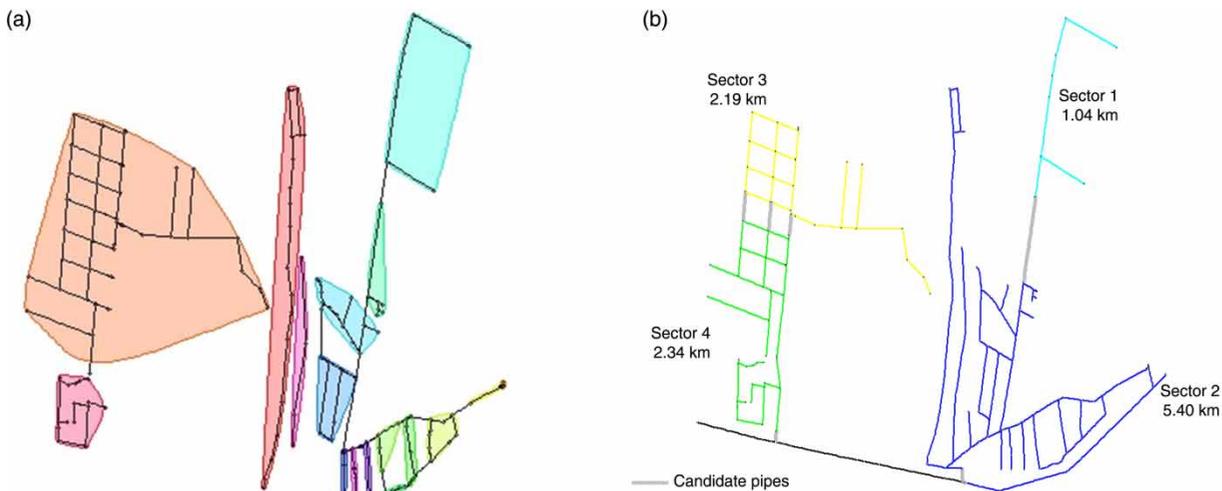

Figure 3 │ Sectors defined in the example WSN: (a) community of the best partition; (b) sectors defined.

It is important to note that the number of partitions in the resulting dendrogram can be significantly large (equal to the number of nodes), and so a manual selection is not feasible. To deal with this, a maximum threshold of pipe length per sector is defined. The threshold should be defined by the water utility staff based upon the experience in dealing with the leakage events in the network and the background leakage rate. For example, if the bursts frequency and the background leakage flow are high, then the pipe length per sector must be smaller than in the case where these parameters are lower. In this regard, GIZ *et al.* (2011) recommend 4–30 km of pipe length as an appropriate sector size range.

Once the threshold is defined, a loop is applied to compare all the partitions, starting from the best partition of the dendrogram and going either up or down, as necessary. The first partition where the length of all the sectors is under the maximum threshold value is selected. Partitions close to the best partition are stronger in terms of communitarian structure, this being the reason why the first partition that satisfies the established subdivision criterion must be selected.

In the previously described searching process, some small communities (communities that are not worth setting as sectors) can be further merged with larger sectors. Taking this into consideration, the maximum threshold can be slightly varied (lowered) so that, if a further merger affects a large sector, the final pipe length does not violate the maximum length constraint. For the above example, the maximal size constraint was set at 6 km.

Merging small communities

The process of merging small communities is carried out through the following set of rules:

$$\text{Given } m \text{ if } SN_m \in C_i \text{ and } EN_m \in C_j \rightarrow m \in CP$$

where m stands for pipe i; SN_m stands for starting node of pipe m; EN_m stands for ending node of pipe m; C stands for cluster; and CP stands for candidate pipe set.

A candidate pipe, as explained later, is a pipe that can be assigned as a boundary pipe between two given sectors, or assigned as a sector entrance.

If a given sector is connected to another sector, and the length of one of those sectors is smaller than the minimum size constraint, then the candidate pipes between those two sectors are eliminated from the list of candidate pipes. Conversely, if a given community is connected to another community with lengths greater than the minimum size restriction, the corresponding intra sector pipes remain in the list of candidate pipes:

$$\text{Given } CP_i \text{ if Size } C_i \text{ or Size } C_j \geq MSL \rightarrow CP_i \notin CP$$

where *MSL* stands for minimum size length and CP_i stand for candidate pipe *i*.

Once the sectors are defined in the DN, the next step implies the reincorporation of the TN. When the TN is reincorporated, the pipes connecting the TN and the DN have to be added to the candidate pipe set:

$$\text{Given } m \text{ if } SN_m \text{ or } EN_m \in C_0 \text{ and } SN_m \text{ or } EN_m \in C_i \rightarrow m_i \in CP$$

where C_0 represents the group of nodes of the TN.

There may be some sectors connected only to the TN (and not to other sectors) with size length lower than the minimum size constraint and, thus, they are not worth being established as sectors. Therefore, the pipes connecting those sectors to the TN are removed from the set of candidate pipes:

$$\text{Given } m \text{ if } SN_m \text{ or } EN_m \in C_0 \text{ and } SN_m \text{ or } EN_m \in C_i \text{ and size } C_i \leq MSL \Rightarrow m_i \notin CP$$

Figure 3(b) shows the sectors detected in the DN. For this example, the minimum size constraint was set to 1 km. In this case, four sectors (seven candidate pipes) were finally defined.

MULTI-OBJECTIVE OPTIMIZATION

Once the list of candidate pipes is obtained, the next step corresponds to the selection of the entrance of each sector. As explained above, the pressure drop caused by closing some pipes is proportional to the number of those pipes. Nevertheless, when a candidate pipe remains open, a flowmeter must be positioned on it to control the flow. Setting more candidate pipes as closed pipes than sector entrances is preferred for two reasons: first, the smaller

the number of entrances to each sector, the greater the accuracy in controlling inlet flow and, therefore, the greater the probability of detecting a new leak; on the other hand, the cost of installing a flowmeter in economic and operative terms is expected to be higher than the cost of installing a boundary valve (when it is necessary to install a new valve).

The problem is addressed as a multi-objective optimization problem with five objectives: (objective 1) the minimization of the deviation ($\triangle I_r$), of the resilience index (I_r) proposed by Todini (2000); (objective 2) the minimization of the increase of operational power ($\triangle P_O$); (objective 3) the minimization of the variation of standard deviation of the pressure (σ_P) between the network nodes; (objective 4) a penalty cost (P_c), for the nodes that are unable to meet a minimum pressure target (P_{def}); and finally (objective 5) the minimization of the cost associated with the purchase of new valves (C_{cv}) and flow meters (C_{fm}). The first two objectives can be considered as energy criteria, the next two as operative criteria, and the fifth objective as an economic criterion. The optimization problem is solved using the ASO algorithm proposed by Montalvo *et al.* (2014). Agent swarm optimization is a novel paradigm that exploits swarm intelligence and borrows some ideas from multi-agent based systems. It is aimed at supporting decision-making processes by solving multi-objective optimization problems. ASO offers robustness through a framework where various population-based algorithms co-exist. In addition, agents are endowed with sets of problem-dependent specific rules with two clear objectives, namely fine-tuning the agent behavior to specialize in the optimization problem being tackled, while, at the same time, reducing the search space, thus enabling decisions to be produced with increased reliability and within a reasonable time frame.

Each of the former functions was included as it was considered that they fairly represent the most important aspects that need to be considered for the sectorization of a middling or large WSN. However, as explained below, there are still some other important aspects that should be taken into consideration.

The objectives may be formulated as follows:

Objective 1: $f(x) = \text{Min}(\triangle I_r)$

Objective 2: $f(x) = \text{Min}(\triangle P_O)$

Objective 3: $f(x) = \text{Min}(\triangle \sigma_P)$

Objective 4: $f(x) = \text{Min}(P_{def})$

Objective 5: $f(x) = \text{Min}(C_{cv} + C_{FM})$

No constraint is defined, as the ASO algorithm works with penalties. A penalty cost is added to objective 4 (nodes with pressure lower than 15 m), so the greater the number of nodes with pressure below that value, the greater the cost of objective 4. It was also assumed that there were no pre-existing valves so any candidate pipe resulting closed in the optimization process is associated with a symbolic cost (100 monetary units/closing valve), lower than the cost associated with pipes that should remain open (1,000 monetary units/flow meter). A symbolic cost was assumed as there was no information about real costs.

In the following sections the objectives are described.

Objective 1: variation of the resilience index

The I_r measures the capacity of a network to overcome pipe failures and still be able to satisfy a minimum pressure restriction.

The total available power (P_{total}) in the network can be calculated by adding the power supplied by all tanks and pumps (Equation (5)):

$$P_{total} = \frac{\gamma \times \left\{ \left[\sum_{t=1}^{T} (Q_t \times H_t) \right] + \left[\sum_{p=1}^{P} (Q_p \times H_p) \right] \right\}}{1000} \quad (5)$$

where P_{total}: total available power (kW); γ: specific weight of water ($9800\ N/m^3$); Q_t: exit flow from tank (m^3/s); H_t: head of a tank (m); Q_p: flow of a pump (m^3/s); H_p: head of a pump (m).

The power in the form of pressure that is provided to all nodes or nodal power (P_n) can be estimated adding the output of multiplying the total demand at each node and the total head in the same node (Equation (6)):

$$P_n = \frac{\gamma \times \left(\sum_{i=1}^{n} Q_i \times H_i \right)}{1000} \quad (6)$$

where P_n: nodal power (kW); Q_i: demand (consumptions and leaks) at node $i (m^3/s)$; H_i: head at node $i (m)$.

The value of P_n may be expressed in real terms, when a real head value is used in its calculation. It may also be expressed in theoretical terms, when a required head value is used. As shown in Equation (7), the required head is calculated by adding the node elevation and the required pressure value:

$$P_n^{\text{required}} = \frac{\gamma \times \left[\sum_{i=1}^{n} Q_i \times (E_i + P_i^{\text{required}}) \right]}{1000} \quad (7)$$

where P_n^{required}: nodal power required to satisfy a pressure target (kW); E_i: elevation of node $i (m)$; P_i^{required}: lower limit of pressure requirement at node $i (m)$.

Considering that P_{total} may be expressed as the sum of P_n and the dissipated power (operational power) due to friction, this may be calculated by subtracting P_n from P_{total} (Equation (8)):

$$P_o = P_{\text{total}} - P_n \quad (8)$$

where P_o: operational power (kW).

Depending on the value of P_n used (real or required), in the evaluation of Equation (8), its output can be either real or required. Both are then used to evaluate the I_r (Equation (9)):

$$I_r = 1 - \frac{P_o^{\text{real}}}{P_o^{\text{required}}} \quad (9)$$

Finally, the values of I_r in the original network and in the sectorized network are used to calculate the $\triangle I_r$ (%) proposed by Di Nardo *et al.* (2014) (Equation (10)):

$$\triangle I_r = \left(1 - \frac{I_r{}'}{I_r{}^{\text{original}}} \right) \times 100 \quad (10)$$

where $I_r{}'$: I_r after sectorization; $I_r{}^{\text{original}}$: I_r before sectorization.

Objective 2: minimization of the variation of the operational power

$\triangle I_r$ can be used to assess the suitability of a sectorization layout relative to a level of reliability based on a pressure

target. Nevertheless, a more general assessment can be carried out by evaluating the $\triangle P_O$ produced after the deployment of a given sectorization layout:

$$\triangle P_O = \frac{\gamma \times \left[P_{\text{total}} - \sum_{i=1}^{n} Q_i^o \times H_i^o \right] - \left[P'_{\text{total}} - \sum_{i=1}^{n} Q_i^x \times H_i^x \right]}{1000} \quad (11)$$

where $\triangle P_O$: variation of the operational power (kW); Q_i^O: demand (consumption and leaks) at node i before sectorization (m^3/s); H_i^O: head at node i before sectorization (m); Q_i^x: demand (consumption and leaks) at a node i after sectorization (m^3/s); H_i^x: head at a node i after sectorization (m); P'_{total}: total available power after sectorization (kW).

Objective 3: minimization of variation of the standard deviation of pressure between the network nodes

Pressure uniformity in WSNs plays an important role in the reduction of h_L. More pressure-uniform WSNs are expected to be more efficient in energy terms, and to some extent, in terms of leakage management as well. The second aspect is related to the control of excessive pressures. When pressure is reduced, a reduction in the background leakage flow is to be expected. However, leakage control by means of pressure reduction has to be carried out with extreme precaution as it might also be translated into supply incapacity (Morrison *et al.* 2007).

Moreover, as can be logically devised, it also has social implications, as it guarantees customer satisfaction. For this reason, it is considered as an individual objective in this work. The way to consider it in the optimization process is by means of the σ_P (Equation (12)). The idea is to minimize the variation of the pressure values between the nodes of the network:

$$\sigma_P = \sqrt{\sum_{i=1}^{n} \frac{(P_i - \bar{P})^2}{n-1}} \quad (12)$$

where $\bar{P}$: average pressure of all the nodes of the WSN (m).

Then, $\triangle \sigma_P$ is calculated with Equation (13):

$$\triangle \sigma_P = \sigma_P^{\text{original}} - \sigma_P' \quad (13)$$

where $\sigma_P{}^{\text{original}}$: σ_P before sectorization; $\sigma_P{}'$: σ_P after sectorization.

Objective 4: penalty cost for deficit of nodal pressure

A penalty cost is defined to lead the solution search process. The penalty is calculated as the sum of the differences between the real pressure and the target pressure in the subset of nodes that are unable to satisfy the minimum pressure requirement, multiplied by a penalty cost value:

$$P_{\text{def}} = P_{c*} \sum_{i=1}^{n} \left| (P^{\min} - P_i) \right| \tag{14}$$

where P_c: penalty cost (monetary units/m); $P_{\min}$: minimun pressure target (m); P_i: pressure in node i (m).

In the equation above, n represents the number of nodes where the pressure constraint is not satisfied.

A symbolic value of 0.8 monetary units/m was assumed for P_c.

Objective 5: economic criteria

As described in the Introduction, when performing sectorizations, water utilities require some investment for the purchase and installation of boundary valves (if the existing valves are in poor condition or there are no valves) and flowmeters. Normally, valves are cheaper than flowmeters. In some countries (especially developing countries), they are completely buried, with a small empty cylinder giving access. Conversely, flowmeters are more expensive and must be located in special chambers. Such chambers need enough room to install the flowmeter as well as a bypass, pressure gauge, and at least three control valves (two for flowmeter maintenance and one for the bypass). The chambers also require enough room for a person to be able to perform maintenance. In a real situation, the construction of a flowmeter chamber disturbs normal city life as it interrupts traffic, especially when chambers have to be located over the TN pipes. This type of impact can be considered as an externality and is not normally taken into consideration. As this methodology isolates the TN from the DN, the impact described above is significantly reduced.

Limitations of the optimization model

It is very important to stress that although this optimization scheme may work to conveniently address the problem of sectorization, the results cannot be considered as completely optimized. A reduction of the network pressure not only reduces the background leakage flow, but also the frequency of new pipe breaks and bursts (Pearson *et al.* 2005; Fantozzi & Lambert 2007). This is very important, as a 28–80% reduction in break-bursts can be translated into a reduction of annual repair costs that is usually far greater than the value of the water saved (Thornton & Lambert 2006). A pressure reduction is also expected to have an effect on domestic consumption (Lambert & Thorton 2005). Therefore, such effects need to be considered among the optimization functions.

It is also very important to note that the optimal point of leakage reduction varies according to the cost of water production and transportation; therefore, this point must be considered as a restriction of the optimization model.

Based on these two aspects, a more complete optimization model must initially define the economic level of leakage reduction, and based on that level, estimate the required pressure reduction and the set of feasible alternative sectorization layouts. We are currently working on this new model.

APPLICATION TO AN EXAMPLE NETWORK

Figure 4 shows the resulting assignation of flowmeters and boundary valves in the example network. Figure 4(a) and 4(b) depict the found feasible solutions 1 and 2, respectively. Table 1 displays a comparison between the values obtained for the two feasible solutions with the values in the original network (network with all the pipes open). Solution 1 is better in terms of flow control, as it only requires one flowmeter for the larger sector. However, it also implies a greater negative impact on network reliability (30.68% of resilience index variation whereas feasible solution 2 has only 0.65% for the same parameter), a greater increase in the operational power, and a few nodes with pressure below 15 m (but above 6 m). Therefore, at this point the technical staff of the network must decide which
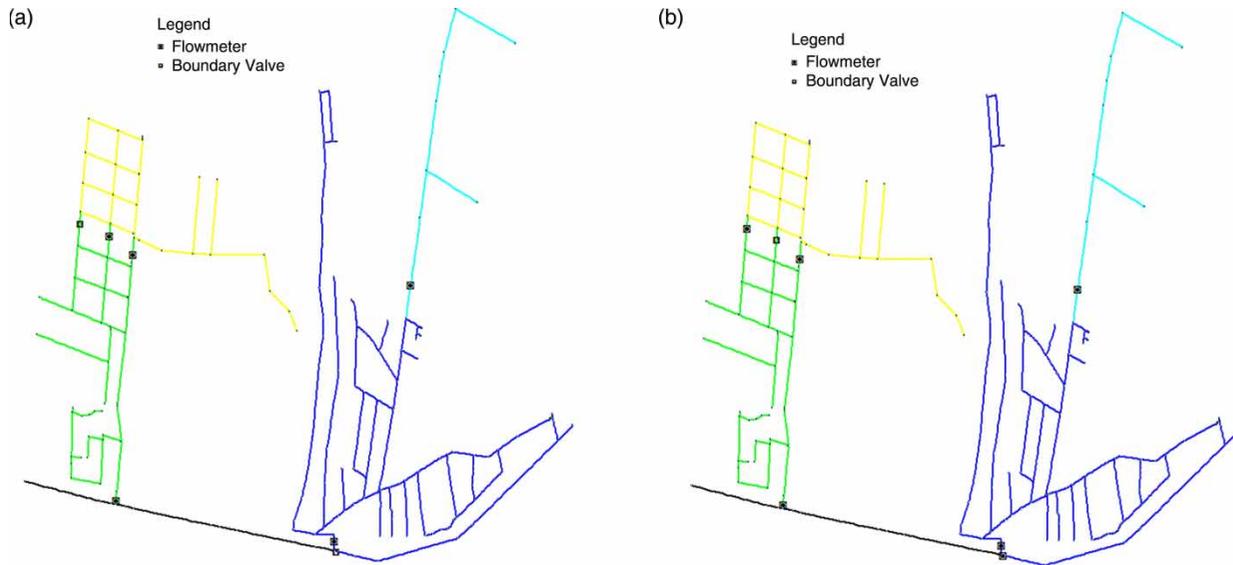
Figure 4 | Results of the optimization process applied to the small network example.

Table 1 | Objective values obtained in the application of the methodology over the small network example

Parameters	Original network		Feasible solution 1	Feasible solution 2
	Values	Functions	Values	Values
I_r	0.48	$\triangle I_r(\%)$	30.68 (current value: 0.30)	0.65 (current value: 0.47)
$P_O(kW)$	20,524.04	$\triangle P_o(kW)$	2210.97 (current value: 22735.02)	86.73 (current value: 20610.77)
$\sigma_p\ (m)$	6.98	$\triangle \sigma_p(m)$	3.30 (current value: 10.28)	0.029 (current value: 6.95)
P_{def} (monetary units)	0.00	P_{def} (monetary units)	259.59	0.00
N_{CV}	0	N_{CV}	2	2
N_{FM}	0	N_{FM}	5	5

configuration is most appropriate. It is important to highlight that pressure values below 15 m are only reported for one period of time (the worst scenario).

To test the whole methodology in a larger network, it was applied over a large section of the WSN of Managua City. This new network has 1,374 nodes and 1,325 pipes. It is supplied by one main pipe that receives water from a series of pumps and a lake. Since these sources are far from the studied network, they are represented by one reservoir. Figure 5 shows the TN definition and the division of the network into eight sectors, all between 3 and 30 km of pipe length. In total, 21 pipes were included in the candidate pipe set. Table 2 shows the value of the objectives in the original network and in the solutions found by the optimization process. In this case, three feasible solutions were found. The three of them imply the installation of six boundary valves (not the same in each case), with a maximum reduction in the resilience index of 16.52%.

CONCLUSIONS

This paper tackles the problem of WSN sectorization depending on a TN by using a novel methodology based on the concepts of SP derived from graph theory and the Walktrap community detection algorithm derived from social network theory, and a multi-objective optimization
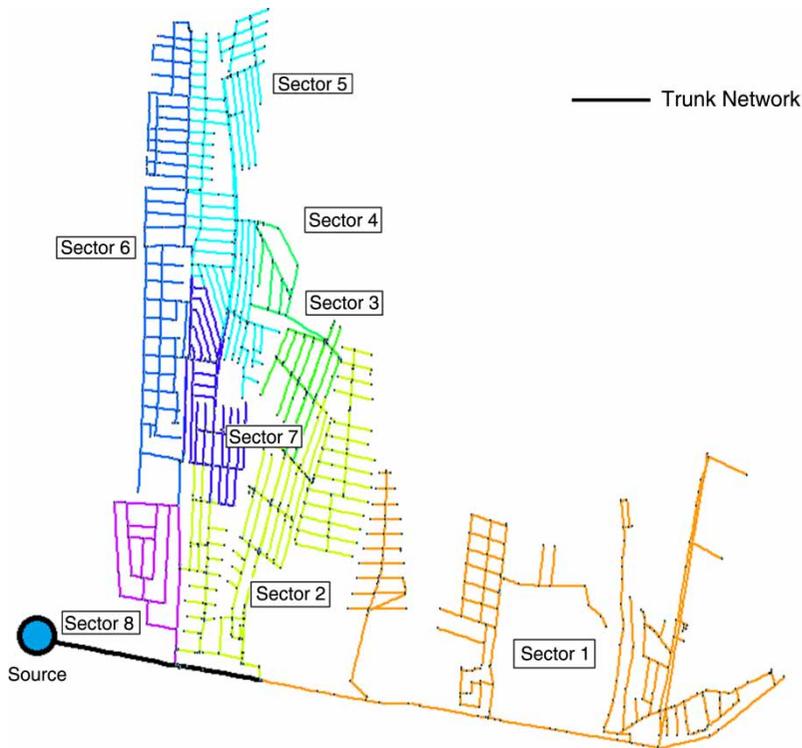
Figure 5 │ Subdivision of the large network example.

Table 2 │ Objective values obtained in the application of the methodology over the large network example

Parameters	Original network Values	Functions	Solution 1 Values	Solution 2 Values	Solution 3 Values
I_r	0.55	$\triangle I_r(\%)$	12.53	14.93	16.52
$P_O(kW)$	24788.61	$\triangle P_O(kW)$	171.58	161.77	173.76
$\sigma_p(m)$	5.47	$\triangle \sigma_p(m)$	0.511	1.20	0.51
P_{def}(monetary units)	0	P_{def}(monetary units)	0	0	0
N_{CV}	0	N_{CV}	6	6	6
N_{FM}	0	N_{FM}	15	15	15

process using ASO. The methodology provides at different steps, feasible solutions, from where the technical staff of the water utilities may select the most convenient solution. To address the definition of the TN in a WSN (where it is not yet defined) a hierarchy of pipes, according to their importance for the supply, is established based on the ASPV* coefficient proposed here. This criterion is more robust than the use of a single characteristic of the pipes because it considers the flow magnitude and the direction

of each pipe in the most critical scenario. Such a hierarchy of pipes also enables the water utilities to define the scope of the TN based on the characteristics of each context. Once the TN is defined, it is detached from the DN, in order not to include it among the sectorization, thus, ensuring the reliability of the WSN.

The Walktrap algorithm creates a structural hierarchy of communities in the DN, from where a partition can be selected out of a set of obtained partitions, depending on

the given context. The advantage behind the use of a community detection algorithm is related to the fact that it reveals the community structure that is expected to be found in real WSNs, which is normally formed by blocks of looped subnetworks separated from other blocks by a few links. It also guarantees homogeneity throughout the sectors, since it is expected that each block is also formed by similar nodes. Finally, when sectors are defined, the TN is reincorporated and a set of candidate pipes is established. This set is formed by the pipes that separate sectors, and those that separate sectors from the TN. From this set, the pipes that are established as entrances to each sector are defined. This definition is carried out by means of a multi-objective optimization process performed with ASO. In this process, reliability, operative, and economic criteria are optimized. Once again, the process does not provide a single solution, but a set of feasible solutions, from where the staff of the water utility can select the most convenient solution.

Despite the good results obtained by the application of the proposed process, the results cannot be considered completely optimized, as some aspects are still unconsidered. The first two are the effect of the pressure reduction over the frequency of new pipe break-bursts and over a reduction of domestic consumption. The third aspect is the economic level of leak reductions, which must be evaluated before any exploration of sectorization layouts.

This analysis is carried out in the most critical scenario in order to give a high level of importance to the reliability of the network. Therefore, once a specific sectorization layout (or a few options) is selected it is necessary to perform extended period simulations.

REFERENCES

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.

Campbell, E., Ayala-Cabrera, D., Izquierdo, J. & Pérez-García, R. 2014a Graph clustering based on social network community detection algorithms. In: *Proceedings of the 7th International Congress on Environmental Modelling and Software* (D. P.

Ames, N. W. T. Quinn & A. E. Rizzoli, eds), June 15–19, San Diego, California, USA.

Campbell, E., Ayala-Cabrera, D., Izquierdo, J., Pérez-García, R. & Tavera, M. 2014b Water supply network sectorization based on social networks community detection algorithms. In: *16th Conference on Water Distribution System Analysis (WDSA'2014)*, Bari, Italy. American Society of Civil Engineers, Reston, VA, USA.

Clauset, A., Newman, M. E. J. & Moore, C. 2004 Finding community structure in very large networks. *Phys. Rev. E* **70** (6), 066111.

De Paola, F., Fontana, N., Galdiero, E., Giugni, M., Savic, D. & Sorgenti Degli Uberti, G. 2014 Automatic multi-objective sectorization of a water distribution network. In: *16th Conference on Water Distribution System Analysis (WDSA'2014)*, Bari, Italy. American Society of Civil Engineers, Reston, VA, USA.

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

Di Nardo, A. & Di Natale, M. 2011 A heuristic design support methodology based on graph theory for district metering of water supply networks. *Eng. Optimiz.* **43** (2), 193–221.

Di Nardo, A., Di Natale, M., Santonastaso, G., Tzatchkov, V. & Alcocer-Yamanaka, V. 2013 Water network sectorization based on genetic algorithm and minimum dissipated power paths. *Water Sci. Technol.* **13** (4), 951–957.

Di Nardo, A., Di Natale, M., Santonastaso, G., Tzatchkov, V. & Alcocer-Yamanaka, V. 2014 Water network sectorization based on graph theory and energy performance indices. *J. Water Res. Pl-ASCE* **140** (5), 620–629.

Fantozzi, M. & Lambert, A. 2007 Including the effect of pressure management in calculations of short run economic leakage levels. In: *Proc. IWA Special Conference 'Water Loss' (2007)*, Bucharest, Romania. IWA, London.

Farley, M. 2010 Are there alternatives to DMA? *Asian Water* **26** (10), 10–16.

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

Giustolisi, O. & RidoIfi, L. 2014 Modularity index for hydraulic system segmentation. In: *16th Conference on Water Distribution System Analysis (WDSA'2014)*, Bari, Italy. American Society of Civil Engineers, Reston, VA, USA.

GIZ (Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH), VAG (VAG Armaturen GmbH), FHNW (Fachhochschule Nordwestschweiz), KIT (Karlsruhe Institute of Technology) 2011 *Guidelines for Water Loss Reduction: A Focus on Pressure Management*. Eschborn, Bonn, Germany.

Gomes, R., Marques, A. S. & Sousa, J. 2012 Identification of the optimal entry points at district metered areas and implementation of pressure management. *Urban Water J.* **9** (6), 365–384.

Hajebi, S., Barret, S., Clarke, A. & Clarke, S. 2013 Multiagent simulation to automate water distribution network partitioning. In: *27th European Simulation and Modelling Conference (ESM'2013)*, Eurosis, Lancaster, UK.

Hajebi, S., Temate, S., Barret, S., Clarke, A. & Clarke, S. 2014 Water distribution network sectorization using structural graph partitioning and multiobjective optimization. In: *16th Conference on Water Distribution System Analysis (WDSA'2014)*, Bari, Italy. American Society of Civil Engineers, Reston, VA, USA.

Herrera, M., Izquierdo, J., Pérez-García, R. & Montalvo, I. 2012 Multi-agent adaptive boosting on semi-supervised water supply clusters. *Environ. Modell. Softw.* **50** (August 2012), 131–136.

Izquierdo, J., Herrera, M., Montalvo, I. & Pérez-García, R. 2011 Division of water supply systems into district metered areas using a multi-agent based approach. In: *Software and Data Technologies* (J. Cordeiro, A. Ranchordas & B. Shishkov, eds), CCIS 50, Springer, Berlin, Heidelberg, pp. 167–180.

Manning, C. D., Raghava, P. & Schutze, H. 2008 *Introduction to Information Retrieval*. Cambridge University Press.

Montalvo, I., Izquierdo, J., Pérez-García, R. & Herrera, M. 2014 Water distribution system computer-aided design by agent swarm optimization. *Computer-Aided Civ. Inf.* **29** (6), 433–448.

Moore, E. F. 1959 The shortest path through a maze. In: *Proc. International Symposium on the Theory of Switching*, Cambridge, Massachusetts, USA.

Morrison, J., Stephen, T. & Roger, D. 2007 *District Metered Areas: Guidance Notes*. Technical Report 1. Water Loss Task Force, International Water Association (IWA), London, UK.

Newman, M. E. J. 2006a Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E.* **74** (3), 036104.

Newman, M. E. J. 2006b Modularity and community structure in networks. *PNAS* **103** (23), 8577–8582.

Newman, M. E. J. & Girvan, M. 2004 Finding and evaluating community structure in networks. *Phys. Rev. E.* **69** (2), 026113.

Orman, G. K., Labatut, V. & Cherifi, H. 2011 On accuracy of community structure discovery algorithms. *J. Converg. Inform. Technol.* **6** (11), 283–292.

Pearson, D., Fantozzi, M., Soares, D. & Waldron, T. 2005 Searching for N₂: how does pressure reduction reduces bursts frequency? In: *Proc. IWA Conference for Leakage (2005)*, Halifax, Nova Scotia, Canada.

Pons, P. & Lapaty, M. 2006 Computing communities in large networks using random walks. *J. Graph Algorithms* **10** (2), 191–218.

Raghavan, U., Albert, R. & Kumara, S. 2007 Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E.* **76** (3), 036106.

Rosvall, M. & Bergstrom, C. 2008 Maps of information reveal community structure in complex networks. *PNAS* **105** (4), 1118–1123.

Rousseeuw, P. 1987 Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65.

Savić, M., Radovanović, M. & Ivanović, M. 2012 Community detection and analysis of community evolution in Apache Ant class collaboration networks. In: *Proc. 5th Balkan Conference in Informatics (BCI'2012)*, Novi Sad, Serbia. ACM.

Thornton, J. & Lambert, A. 2005 Progress in practical prediction of pressure: leakage, pressure: burst frequency and pressure: consumption relationships. In: *Proc. IWA Water Loss Management Conference (2005)*, Halifax, Nova Scotia, Canada.

Thornton, J. & Lambert, A. 2006 Managing pressures to reduce new break. *Water 21* **8.6**, 24–26.

Todini, E. 2000 Looped water distribution networks design using a resilience index based heuristic approach. *Urban Water J.* **2** (2), 115–122.

Tzatchkov, V., Alcocer-Yamanaka, V. & Bourguett-Ortiz, V. 2008 Graph theory based algorithms for water distribution network sectorization projects. In: *Proc. 8th Water Distribution Systems Analysis Symposium (2006)*, Cincinnati, Ohio, USA. American Society of Civil Engineers, Reston, VA, USA.

Ward Jr., J. 1963 Hierarchical grouping to optimize an objective function. *J. Am. Statist. Assoc.* **58** (301), 236–244.