AGU100 ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR023966

Key Points:
- Centrality metrics are used for domain analysis of water distribution networks
- The topology of water distribution network is tailored for the calculation of centrality metrics
- WDN-tailored edge-betweenness is used for domain analysis

Correspondence to:
O. Giustolisi,
orazio.giustolisi@poliba.it

Citation:
Giustolisi, O., Ridolfi, L., & Simone, A. (2019). Tailoring centrality metrics for water distribution networks. *Water Resources Research*, *55*. https://doi.org/10.1029/2018WR023966

Received 24 AUG 2018
Accepted 3 FEB 2019
Accepted article online 12 FEB 2019

Tailoring Centrality Metrics for Water Distribution Networks

Orazio Giustolisi[1], Luca Ridolfi[2], and Antonietta Simone[1]

[1]Dipartimento di Scienze dell'Ingegneria Civile e dell'Architettura (DICAR), Politecnico di Bari, Bari, Italy, [2]Dipartimento di Ingegneria dell'Ambiente, del Territorio e delle Infrastrutture (DIATI), Politecnico di Torino, Turin, Italy

Abstract Complex network theory (CNT) is an emerging topic based on the paradigm that quite all the natural and man-made physical systems work as networks, namely, their features derive from the internal connectivity among vertices exchanging information through edges. Water distribution infrastructures are networked systems connecting the vertices named nodes, by edges named pipes, and transferring water to customers. Therefore, water distribution networks (WDNs) fall into CNT and belongs to the class of spatial networks due to their urban constraints. CNT proposed several centrality metrics for quantifying the importance of vertex and, sometime, edges. Those metrics can be potentially very useful for analyzing the key features of the physical domain (i.e., the network) where WDN hydraulics occurs, but they need to be tailored to consider that (i) pipes/edges are the relevant physical components of the WDNs, (ii) some nodes/vertices (reservoirs and tanks) play a completely different hydraulic role from the majority of nodes (demand and connection nodes), and (iii) pipes/edges have different characteristics (length, diameter, hydraulic resistance, etc.). Accordingly, this work presents and discusses the need of tailoring the most suitable centrality metrics for spatial networks: betweenness, closeness, and degree. Then the capacity of the WDN-tailored edge betweenness is demonstrated and discussed using two real WDNs, showing that it can extract useful information from the domain, that is, the *emerging* hydraulic behavior due to the network connectivity structure. Therefore, the WDN-tailored edge betweenness can assist analysis, planning, and management actions before and after the hydraulic analysis.

Plain Language Summary The aim of the work is to develop a water distribution network (WDN)-tailored centrality metric from complex network theory for the analysis of the domain of the WDN hydraulics. The WDN-tailored edge betweenness is, then, developed. Tailoring is performed considering three specific WDN features: (1) The pipes are the relevant components of such infrastructure systems. This is in contrast with standard complex network theory centrality metrics that generally focus on nodes because they represent the objects exchanging information. (2) Pipes are material components characterized by asset features such as length, diameter, and hydraulic resistance. (3) Each node can represent a different system component. It can be a source of water, a connection among pipes, or a water outflow. Accordingly, the work defines source of water, connection, and demand nodes, respectively.

1. Introduction

Comprehensive asset management of water distribution networks (WDNs) is becoming a relevant issue for the scientific community because water companies urge for solutions to novel tasks. During the last century, the main task of water companies was the construction of WDNs to deliver water extending the percentage of customers reached by the service. Starting the new millennium, aging of the hydraulic systems, demographic pressure in urban areas, increased sensibility of the customers to the service quality versus the tariffs, etc., are moving the technical needs toward an effective analysis, planning, and management of such networked hydraulic systems.

The studies carried out in the last decades about the complex network theory (CNT) can make a crucial contribution to addressing and solving these new technical needs. WDNs belong to a specific class of complex networks, the so-called spatial networks (Barthelemy, 2014; Barthélemy & Flammini, 2008) whose characteristics are conditioned by spatial constraints (Barthélemy, 2011; Giustolisi et al., 2017); that is, the geodesic distance among nodes is relevant information. The spatiality of WDNs is mainly related to the urban texture, for example, streets and buildings, constraining the pipe installation and influencing the layout. Giustolisi et al. (2017) noted that this fact limits the maximum nodal degree for WDNs, narrowing the range of possible values for nodal degree. For this reason, they introduced the neighborhood degree, which is an extension of the nodal degree concept to the adjacent nodes. Also, spatial constrains affect the temporal evolution of spatial networks (Barthelemy, 2014; Barthélemy & Flammini, 2008), as recently highlighted by Giustolisi et al. (2017) for the WDN classification.

During the last decade, several researchers used CNT for WDN analysis. Yazdani and Jeffrey (2010, 2011, 2012a, 2012b) were the first to apply CNT metrics to study WDN features such connectivity, robustness, redundancy, fault tolerance, and vulnerability in order to assess structural performance. Their conclusion was that the CNT metrics are useful but insufficient for WDN analysis because of spatial constraints. Giustolisi et al. (2017) demonstrated that spatial constraints explain why the model network structure of WDNs is Poisson-like. This fact means that the domain has random features increasing the size (i.e., the number of nodes), that is, a good structural resistance. The work did not consider reservoirs, tanks, and directional devices; therefore, the structural analysis captured the network domain features on average.

Furthermore, Yazdani and Jeffrey (2012a, 2012b) proposed the use of weighted and directed graph representations for WDN analysis, offering an illustrative novel metric based on the nature of the specific network in terms of both connectivity and physical features.

The contributions of Perelman and Ostfeld (2011), Scibetta et al. (2013), Diao et al. (2013), Giustolisi and Ridolfi (2014a, 2014b), and Laucelli et al. (2017) aimed at using the CNT studies for WDN domain segmentation. The Newman's modularity index (Newman, 2004) was generally used, and the technical task referred to district metering area (DMA) design. Perelman and Ostfeld (2011) developed a graph theory connectivity-based algorithm for WDN segmentation based on connectivity analysis; Scibetta et al. (2013) and Diao et al. (2013) were the first to apply modularity index for the WDN segmentation, and Giustolisi and Ridolfi (2014a, 2014b) tailored the modularity index obtaining a WDN-oriented modularity metric to consider that devices (flowmeters and closed gates) are installed closed to ending nodes of pipes. Finally, Laucelli et al. (2017) integrated the segmentation with the hydraulic analysis in a two-step strategy for DMA planning.

Instead, Simone et al. (2016) extended the modularity index to sampling design, defining the *pressure* districts/segments and using the overlapping community concept and the connectivity matrix of edges, that is, the line graph. Nazempour et al. (2016) also referred to CNT concepts for system monitoring and control, for example, metering water consumption and early contaminant detection.

Finally, CNT centrality metrics were adopted to assess vulnerability analysis (Agathokleous et al., 2017; Diao et al., 2014; Gutiérrez-Pérez et al., 2013; Hawick, 2012; Shuang et al., 2014), reliability assessment (Ostfeld, 2012; Trifunovic, 2012; Yannopoulos & Spiliotis, 2013), and as indicators of WDNs resilience (Pandit & Crittenden, 2016; Yazdani & Jeffrey, 2011; Yazdani et al., 2011; Zhao et al., 2015).

The novel WDN-tailored topology presented in this work can be useful to improve the previous efforts, which were based on the *standard* network connectivity structure.

In the reported scientific literature review of efforts to adopt tools of CNT, the present work focuses on the centrality metrics, aiming at tailoring for analyzing the domain structure of WDNs. Network centrality metrics represent a specific field of CNT devoted to study indicators to quantify the importance of vertices (sometime edges) of a network from different standpoints, classically for assessing system reliability and vulnerability. Several centrality metrics have been proposed in the scientific literature (Newman, 2010a), as eigenvector centrality (Bonacich, 1972), Katz centrality (Katz, 1953), PageRank (Page et al., 1998), and Hub and Authorities (Kleinberg, 1999). Despite this variety of metrics, Borgatti (2005) indicated three centralities that are more suited for infrastructure networks: betweenness, closeness, and degree. The betweenness (Freeman, 1977) quantifies the importance of a vertex for the communication inside the network, that is, the number of shortest paths passing through vertices. The closeness (Freeman, 1977) measures the centrality as the importance of a vertex in spreading information to the other vertices through minimum distance paths. Finally, the degree (Freeman, 1977; Nieminen, 1974) is the most intuitive centrality metric and it is based on the number of edges connecting to each vertex (i.e., the number of adjacent vertices): the vertex with the highest degree is the most important. The neighborhood degree proposed by Giustolisi et al. (2017) corresponds to tailoring the degree centrality concept for WDNs by involving adjacent nodes belonging to neighborhoods at different topological or geodesic distance.

Centrality metrics are generally used to rank the importance of network vertices regardless of the specific values of centrality (Bonacich, 1987). To this purpose, Freeman (1979) introduced the concept of centralization for measuring the different level of importance of vertices in the network, that is, how much a vertex is central in relation to how central all the other vertices are. From this point of view, several authors (Cadini et al., 2008; Ding et al., 2004; Furlan Ronqui & Travieso, 2015; Lee, 2006; Mihail & Papadimitriou, 2002) noted that different centrality metrics provide generally highly correlated rankings, although it is impossible to know in advance if two or more metrics provide similar rankings for a given network (Benzi & Klymco, 2015).

Note that centrality metrics are application-oriented, and the specific choice is driven by the network features to analyze, meaning that the metric selection is not a trivial task (Benzi & Klymco, 2015). For example, centrality metrics based on the concept of shortest paths are relevant to characterize the behavior of spatial networks with respect to the flux of information. This is the reason why this class of centralities is suitable for WDN hydraulic domain characterization, where water flowing into pipes correspond to the information flux on the network.

Therefore, not all metrics are suitable for WDN domain characterization and, more importantly, they have not to be applied uncritically. In this sense, the aims of this work are (i) to present and discuss the standard centrality metrics of CNT that are more suited for WDNs and (ii) to develop a WDN-tailored centrality metric. Tailoring will be performed considering three specific features of WDNs:

1. The edges, that is, pipes, are the relevant components of such infrastructure systems. This is in contrast with standard CNT centrality metrics that generally focus on nodes (vertices) because they represent the objects exchanging information.
2. Pipes are material components characterized by asset features such as length, diameter, and hydraulic resistance.
3. Each node can represent a different system component. It can be a source of water, a connection among pipes, or a water outflow, as discussed in Giustolisi et al. (2017). Accordingly, we will define nodes as source nodes, connection nodes, and demand nodes, respectively.

Considering these points, we will propose modifications of the standard centrality metrics based on the following:

1. computing the edge relevance instead of vertex one;
2. adding weights to edges to account for the specific asset characteristics of pipes, that is, the hydraulic characteristics of the domain (e.g., dissipations); and
3. modifying the connectivity structure of the network to distinguish source, connection, and demand nodes. To this aim, we will introduce fictitious nodes (whose number depends on the number of demand nodes) connected to the source nodes. In this way, the topological relevance of the source nodes will be emphasized by generating fictitious water paths that embed the different hydraulic role played by source nodes. In other words, fictitious nodes allow weighting topologically source nodes with respect to the demand nodes.

The paper is organized as follows. The next section recalls the standard topological centrality metrics proposed in the CNT that are more suitable for the spatial networks. In the third section, we begin to tailor them by focusing on the pipe component relevance in hydraulic systems. Moreover, a specific subsection shows that the hydraulic status variables can be interpreted as *hydraulic* centrality metrics, which measure the pipes relevance with respect to the flow, velocity, leakage, etc. In this way, we set a common framework for topology- and hydraulic-based centrality metrics. In the fourth section, we continue the WDN-tailoring and propose a novel connectivity structure of the network to face up the different hydraulic meaning of nodes in WDNs. Finally, in the fifth section, two case studies allow us to demonstrate and discuss the novel WDN-tailored edge betweenness centrality metric. Concluding remarks are drawn in the last section, where the perspective of using the novel WDN-tailored metric as complementary tool for WDN analysis, planning, and management tasks is indicated.

2. Complex Network Approach for WDN Domain Characterization

The hydraulic behavior of WDNs is mathematically described by three components: (i) the one-dimensional mass and momentum balance equations; (ii) the domain where equations are solved, that is the pipe network with its topological and hydraulic characteristics; and (iii) the boundary conditions, that is, demands, water tank levels, devices state, etc. The latter are usually assumed slowly time-varying in order to neglect fast transient effects, and demands are modeled as point sinks spatially distributed as Poisson pulses. The WDN hydraulics is then simulated as a sequence of steady-state snapshots representing the movie of the hydraulic system behavior over time, where the frame rate of several minutes is intended to capture the slow (with respect to the response time of the hydraulic behavior of the network) change of the network boundary conditions (Giustolisi & Walski, 2012).

In this picture, we aim at highlighting the role played by the domain features in determining the network hydraulics. Obviously, as in any mathematical problem, all the previously listed three components contribute, but the topology of the network and the most important pipes characteristics (like length and conductance or resistance) play a relevant role and we argue that the domain features drive the emerging/typical behavior of the WDN hydraulics. In other words, although different boundary conditions can alter the single values of discharges and heads, the internal connectivity structure of the network, the topological position of the reservoirs, the pipes resistance/conductance, etc.—that is, the domain of the problem—give the main traits of the hydraulic behavior of a WDN and greatly determine the system state over time.

In section 1, we recalled that in the recent years CNT is providing several refined tools to study the connectivity structure of networks and to capture its relevant characteristics. Therefore, it appears very reasonable, even if challenging, to use CNT studies and tools to investigate how the domain features affect the WDN hydraulics. The bridge is not trivial because CNT tools need to be tailored in order to capture the specificity of the hydraulic system domain. For instance, the position of reservoirs and tanks, the hydraulic features of pipes, and the presence of unidirectional devices (that correspond to pipes having null conductance in one direction) play a relevant role in determining the hydraulic state and they must be carefully considered. However, once tools have been suitably adapted, we think that CNT is a powerful way to extract and understand the role of the domain in the *emerging* hydraulic behavior of WDNs. The aim of the present work is then to propose this idea, to explore it, and to show some examples that demonstrate the importance of the domain in the hydraulic problem.

It is important to clarify that the analysis of the WDN hydraulics domain cannot substitute the hydraulic simulations, because the information about the continuity and momentum equations is not contained in the domain. In fact, domain characteristics can reveal the main features of the hydraulic system helping, for example, the engineering judgment for WDN analysis, design, and management tasks (Giustolisi & Walski, 2012).

To the purpose of a formalization of the above discussion, we report the mathematics of WDN hydraulics. A hydraulic network of n_p pipes with unknown flow rates, n_n nodes with unknown heads (internal nodes), and n_0 nodes with known heads (e.g., reservoir levels) can be analyzed by solving the following nonlinear mathematical system based on momentum and mass balance equations:

$$\begin{aligned} \mathbf{A}_{pp}\mathbf{Q}_p + \mathbf{A}_{pn}\mathbf{H}_n &= -\mathbf{A}_{p0}\mathbf{H}_0 \\ \mathbf{A}_{np}\mathbf{Q}_p - \mathbf{d}_n(\mathbf{H}_n) &= \mathbf{0}_n \end{aligned} \tag{1}$$

where $\mathbf{Q}_p = [n_p,1]$ column vector of unknown pipe flow rates, $\mathbf{H}_n = [n_n,1]$ column vector of unknown nodal heads, $\mathbf{H}_0 = [n_0,1]$ column vector of known nodal heads, $\mathbf{d}_n = [n_n,1]$ column vector of nodal demands, $\mathbf{0}_n = [n_n,1]$ column vector of null values, and $\mathbf{A}_{pp}$ = diagonal matrix of size $[n_p,n_p]$, whose elements are based on pipes resistance and flow.

The domain of the equations, that is, the network, is represented by $\mathbf{A}_{pn} = \mathbf{A}^T{}_{np}$ and $\mathbf{A}_{p0}$ = topological incidence submatrices of size $[n_p,n_n]$ and $[n_p,n_0]$, respectively, which are derived from the general topological matrix $\bar{\mathbf{A}}_{pn} = [\mathbf{A}_{pn} \mid \mathbf{A}_{p0}]$, that is, excluding the signs the structure of the edge-vertex (incidence) matrix from the graph theory standpoint, of size $[n_p,n_n + n_0]$. In other words, the general topological matrix is an incidence matrix, which includes the information for the conventional positive directions of pipes using negative or positive signs to unit values.

The solution of the system (1), that is, the computation of the unknowns ($\mathbf{Q}_p$; $\mathbf{H}_n$), requires the always present boundary conditions ($\mathbf{d}_n$; $\mathbf{H}_0$) and of the pumps curve or devices status over time if present.

$\mathbf{R}_p$ or $\mathbf{C}_p = [n_p,1]$ column vector of pipe resistances or conductance complete the characteristics of domain, ($\bar{\mathbf{A}}_{pn}$). $\mathbf{R}_p$ or $\mathbf{C}_p$ plays a similar role of terrain slopes of the shallow water equations domain, meaning that the water searches for the lower resistance or higher conductance paths. The difference resides in the fact that in pressurized systems the flow direction of pipes in not known in advance and the demands acts as sinks or water wells that sources fill at varying levels over time.

It is worth noting that the matrix multiplication of the transpose of general topological matrix and itself ($\bar{\mathbf{A}}_{np}$ x $\bar{\mathbf{A}}_{pn}$) provides the Laplace matrix, which is the difference between the degree matrix and the adjacency matrix, of the graph referred to the WDN domain.

Furthermore, the global gradient algorithm for WDN hydraulics is (Giustolisi & Walski, 2012)

$$\begin{aligned}
\mathbf{F}_n^{\text{iter}} &= -\mathbf{A}_{np}\left(\mathbf{Q}_p^{\text{iter}} - \left(\mathbf{D}_{pp}^{\text{iter}}\right)^{-1}\mathbf{A}_{pp}^{\text{iter}}\mathbf{Q}_p^{\text{iter}}\right) + \mathbf{A}_{np}\left(\mathbf{D}_{pp}^{\text{iter}}\right)^{-1}\mathbf{A}_{p0}\mathbf{H}_0 + \mathbf{d}_n^{\text{iter}}\left(\mathbf{H}_n^{\text{iter}}\right) + \mathbf{D}_{nn}^{\text{iter}}\mathbf{H}_n^{\text{iter}} \\
\mathbf{H}_n^{\text{iter}+1} &= -\left(\mathbf{A}_{np}\left(\mathbf{D}_{pp}^{\text{iter}}\right)^{-1}\mathbf{A}_{pn} - \mathbf{D}_{nn}^{\text{iter}}\right)^{-1}\mathbf{F}_n^{\text{iter}} \\
\mathbf{Q}_p^{\text{iter}+1} &= \left(\mathbf{Q}_p^{\text{iter}} - \left(\mathbf{D}_{pp}^{\text{iter}}\right)^{-1}\mathbf{A}_{pp}^{\text{iter}}\mathbf{Q}_p^{\text{iter}}\right) - \left(\mathbf{D}_{pp}^{\text{iter}}\right)^{-1}\left(\mathbf{A}_{p0}\mathbf{H}_0 + \mathbf{A}_{pn}\mathbf{H}_n^{\text{iter}+1}\right)
\end{aligned} \tag{2}$$

where iter is a counter of the iterative solving algorithm, $\mathbf{D}_{pp}$ and $\mathbf{D}_{nn}$ are diagonal matrix derivatives head loss function with respect to $\mathbf{Q}_p$ and $\mathbf{d}_n{}^{\text{iter}}$ with respect to $\mathbf{H}_n$.

The global gradient algorithm involves the iterative solution of a linear *nodal* system of equations to compute $\mathbf{H}^{\text{iter}\ +\ 1}$ and a nodal mass balance correction obtained by n_p independent equations to compute $\mathbf{Q}^{\text{iter}\ +\ 1}$ based on $\mathbf{H}^{\text{iter}\ +\ 1}$. It is to note that the matrix $\mathbf{A}_{np}$ $(\mathbf{D}_{pp})^{-1}\mathbf{A}_{pn}$ plays a relevant role for the nodal solutions, and it is a Laplace *like* matrix integrated with the information of pipes resistance $\mathbf{R}_p$, being the diagonal elements, referring to the kth pipe, $D_{kk} = D_{kk}$ (R_{kk},|Q|) of $\mathbf{D}_{pp}$.

This fact together with the presence of the term $\mathbf{A}_{np}$ $(\mathbf{D}_{pp})^{-1}$ in the first of equation (2) better shows and demonstrates the mathematical fundamentals of studying the WDN hydraulic domain.

3. Centrality Metrics Suitable for WDNs

Previous researchers (e.g., Barthélemy, 2018; Borgatti, 2005) stated that betweenness, closeness, and degree are the most suitable metrics for studying the centrality problem in spatial networks. In this section we start from this state and we recall and discuss these three metrics to the purpose of tailoring them for WDNs. To show some numerical examples, we will use the Apulian WDN whose layout is reported in Figure 1 (see Giustolisi et al., 2008, for hydraulic details). This network has 23 nodes, 34 pipes, and one reservoir.

3.1. Betweenness

The Betweenness centrality was introduced by Freeman (1977, 1979), although the first formal definition was proposed by Anthonisse (1971), but his work was never published. Given a node i, and two nodes s and t, m topological shortest paths exist between s and t (i.e., more than one path connecting s and t and traversing the minimum number of nodes), and a fraction of those shortest paths traverse the node i. The sum of those fractions for all the couples of nodes (s, t) in the network is the betweenness centrality of node i. In other words, the metric is equal to the summation of the fraction of shortest paths traversing the node i. The node with the highest betweenness centrality is the most relevant. The formulation of the betweenness is

$$C_i^B = \sum\nolimits_{s \neq i \neq t \in V} \frac{\sigma_{s,t}(i)}{\sigma_{s,t}} \tag{3}$$

where $C_i{}^B$ is the betweenness centrality of node i, $\sigma_{s,t}(i)$ is the number of shortest paths from node s to node t passing through the node i, $\sigma_{s,t}$ is the number of all shortest paths from node s to node t, and V is the set of nodes belonging to the network.

We here report the steps for the calculation of the betweenness centrality for the Apulian network to exemplify the features of the metric. The starting point is the identification of the shortest paths between each couple of nodes in the network. For example, let us consider the node couple (s,t) = (1.14). There are four shortest paths that link node 1 to node 14: {1, 19, 23, 21, 20, 14}, {1, 5, 6, 16, 15, 14}, {1, 5, 6, 7, 15, 14}, and

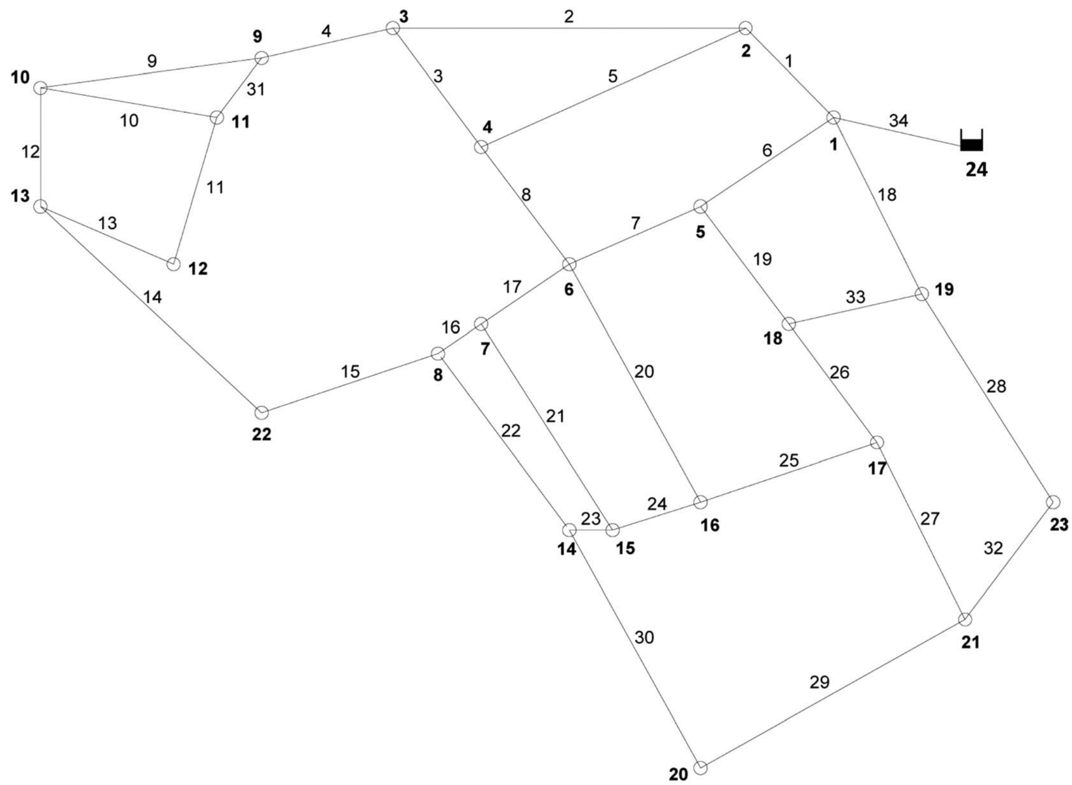

Figure 1. Apulian layout. The bold and plain numbers refer to node and pipe numbering, respectively. The node 24 is the reservoir.

{1, 5, 6, 7, 8, 14}. The length of these four shortest paths (i.e., the minimum number of topological walks) is equal to 5, and the number of nodes contained in the path equal to 6. For example, {1, 5, 6, 7, 8, 14} means that a shortest path starts from node 1 and transverses the nodes 5–8 to reach the node 14 (Table A1). Only three of these shortest paths pass through node 6, as shown in Figure 2. Therefore, considering that the total number of shortest paths between nodes 1 and 14 is 4 and that three of these traverse the node 6, equation (3) states that the contribution of node couple (1.14) to the betweenness centrality of node 6 is

$$C_6^{B(1-14)} = \frac{\sigma_{1-14}(6)}{\sigma_{1-14}} = \frac{3}{4} = 0.75 \quad (4)$$

It is important noting that a path is a walk not passing two times from a vertex initial and final excluded in the case of a loop path, see Diestel (2005) for the definition. The initial identification of the shortest paths requires a computational runtime increasing as $O(N^3)$, where N is the number of vertices/nodes. Several approximated algorithms have been proposed to reduce the computational burden when the network size increases; the Dijkstra (1959) is the most successful when the entries of the adjacency matrix are positive.

3.2. Closeness

The Closeness centrality was introduced by Freeman (1979) to measure the node efficiency in spreading the information. Given a node i, it depends on the distances from the other nodes, j, of the network, computed as shortest path, d_{ij}. The formulation is

$$C_i^C = \frac{1}{\sum_j d_{ij}} \quad (5)$$

where C_i^C is the closeness centrality of node i, $\sum d_{ij}$ is the sum of (i, j) distances. As well as for the betweenness, the starting point of the closeness calculation is the identification of the shortest paths between each couple of nodes in the network. Let us consider the Apulian network. Table A2 reported in Appendix A

10.1029/2018WR023966

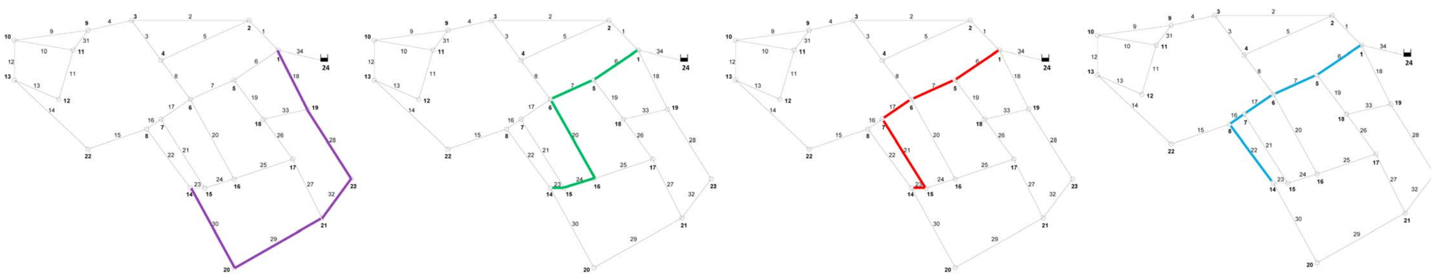

Figure 2. The four shortest paths from nodes 1 to 14: {1, 19, 23, 21, 20, 14}, {1, 5, 6, 16, 15, 14}, {1, 5, 6, 7, 15, 14}, and {1, 5, 6, 7, 8, 14}.

shows the (symmetric) 24×24 matrix of the distances for this network ($N = 24$), and the sum of the elements in the ith row (or column) gives the sum of distances from node i to all other nodes in the network. For example, if node 6 is chosen, the sum of distances from node 6 to all other nodes is equal to 61 (see the bold number in the right of the matrix) and, then, equation (5) gives

$$C_6^C = \frac{1}{\sum_{j=1-24} d_{ij}} = \frac{1}{61} = 0.0164 \tag{6}$$

As an example, Figure 3 just shows the paths (number of steps) from node 6 to the first eight nodes in the network. Such paths have from 1 to 2 steps (see bold lines), and it is trivial to observe that the distance of node 6 from itself is null.

As well as for the betweenness, the starting point of the closeness calculation is the identification of the shortest paths between each couple of nodes in the network to store them for nodal distances (i.e., the matrix in Table A2). As for the betweenness, Dijkstra (1959) is an efficient, although approximated, algorithm.

It is important to note that the closeness works for fully connected networks; otherwise, the distance between two nodes belonging to different components is infinite and, then, the metric of equation (6) is equal to the null value for each node. In order to avoid this problem, it is possible to define the metric as the summation of the inverse of each distance, which is the formulation of the harmonic centrality (Rochat, 2009).

3.3. Degree and Neighbor Degree

The degree centrality was introduced by Nieminen (1974). It is the number of edges connected to a node and describes the local connectivity of the network structure. Node degree is usually indicated with k, but here we define it with C^D to highlight its role as centrality metric and coherently to the other metrics previously recalled. The degree centrality is very fast to be calculated being equal to the sum on the rows (or on the

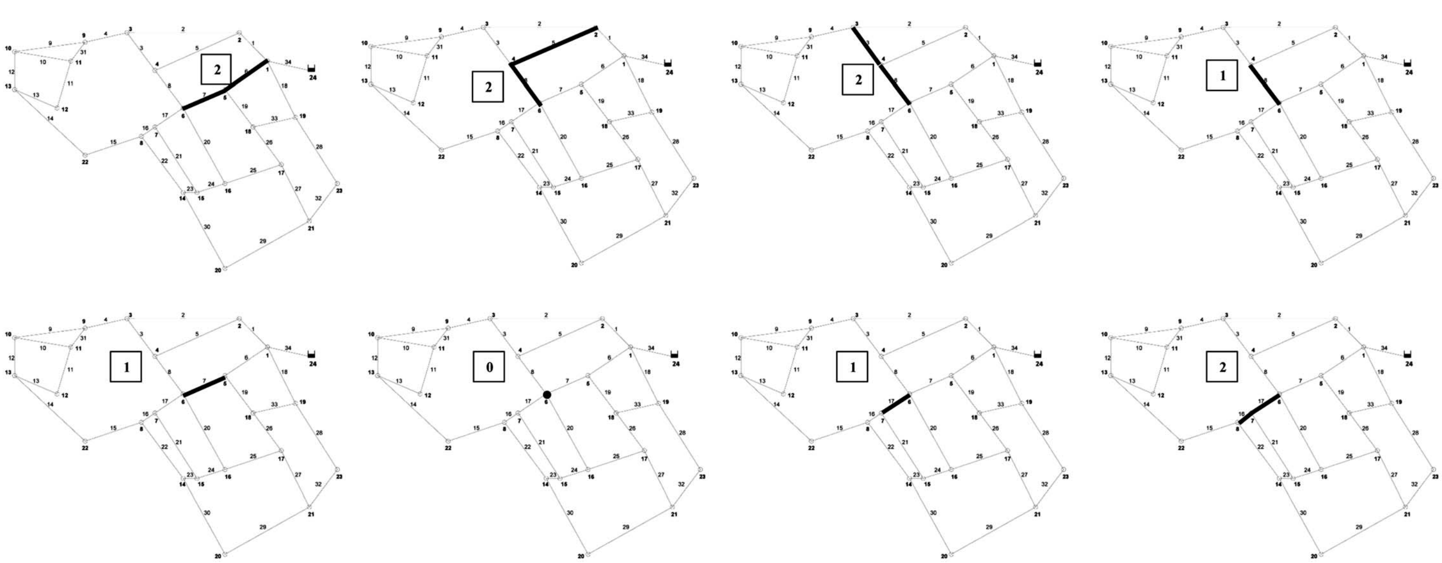

Figure 3. Paths from node 6 to the first eight nodes in the network.

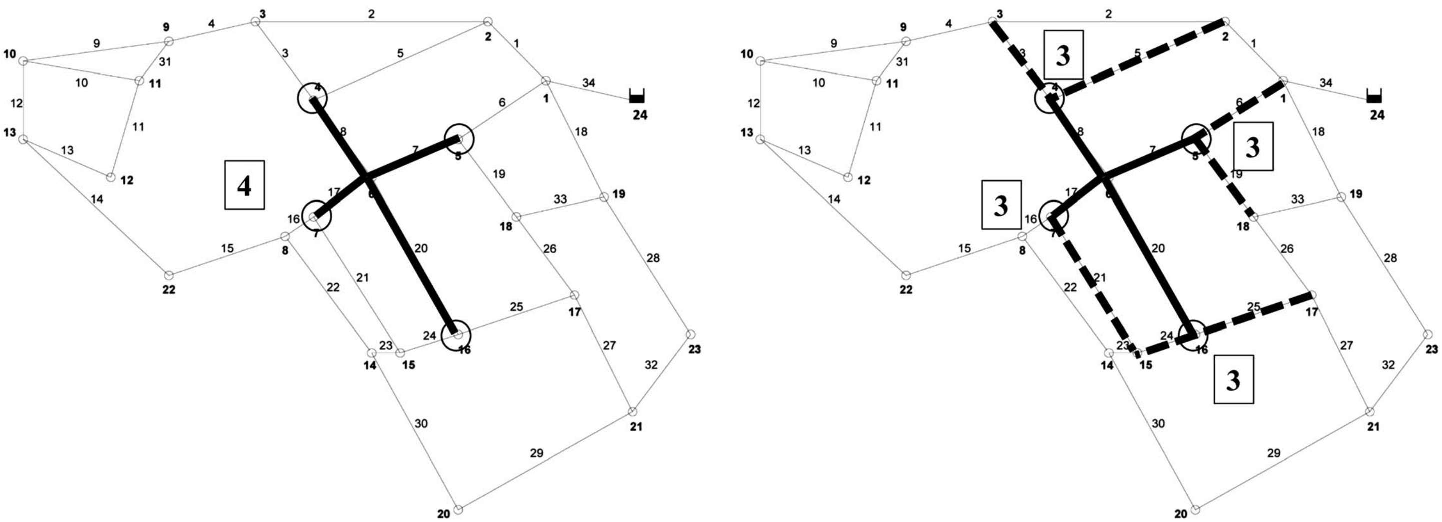

Figure 4. Standard degree (left) and 1-neighborhood degree (right) for the node 6.

columns) of the elements of the adjacency matrix. For example, C^D is equal to 4 in the case of node 6. An extension of the standard degree is the neighborhood nodal degree (C^N) introduced by Giustolisi et al. (2017) to classify WDNs. The neighborhood degree corresponds to the sum of the degrees of the adjacent nodes and its formulation is

$$C_i^N = k_n(i) = \sum_{j \in N(i)} A_{ij} C_j^D \tag{7}$$

where C_i^N is the neighborhood degree centrality of node i, $k_n(i)$ is the neighborhood degree of node i, A_{ij} are the elements of the adjacency matrix, and C_j^D is the standard degree of the node j belonging to the topological neighborhood $N(i)$ (adjacent nodes).

Therefore, the n-neighborhood nodal degree centrality measures the connectivity of the network structure grouping nodes according to the assumed level n. In Appendix A, the vectors of degree and neighborhood degree centrality for all the nodes of Apulian network are reported in Table A3.

Figure 4 shows the standard degree (equal to 4) and the 1-neighborhood degree (equal to 12) for node 6, where the edges connected to the single node (left) and to any of the adjacent nodes (right) are in bold.

Note that the extended neighbor degree represents a *local* centrality metric, while betweenness and closeness centralities are *global* metrics being based on shortest paths involving the entire network connectivity, that is, each couple of nodes.

In order to compare different metrics, Figure 5 reports betweenness and closeness centrality for the Apulian network. The centralities are normalized in the range [0, 100]. Figure 5 shows that node 6 exhibits both the maximum betweenness centrality and the maximum closeness centrality. The importance of the other nodes is lower and depends on the metric. Therefore, node 6 is the most efficient in spreading the information (closeness) because it is the closest to the others and, furthermore, for the specific network, it is also the most traversed by the shortest paths (betweenness). However, it is crucial to note that Figure 5 refers to the relevance of nodes in the network, while we are interested in that of pipes, which are the material network components characterized by asset features such as length, diameter, and hydraulic resistance. This different point of view explains why we need to tailor the previously recalled centrality metrics to adapt them to the analysis of WDN hydraulic domain features.

4. Tailoring Centrality Metrics for WDNs: Node Versus. Edge Centrality

In social, biological, and technological networks, the vertices are typically the objects sharing information through material or immaterial edges. Consequently, standard centrality metrics looks at ranking the node relevance. In the case of WDNs, pipes with their asset features are instead the most important network components: they influence the hydraulic behavior, while connection nodes are network elements that transfer

10.1029/2018WR023966

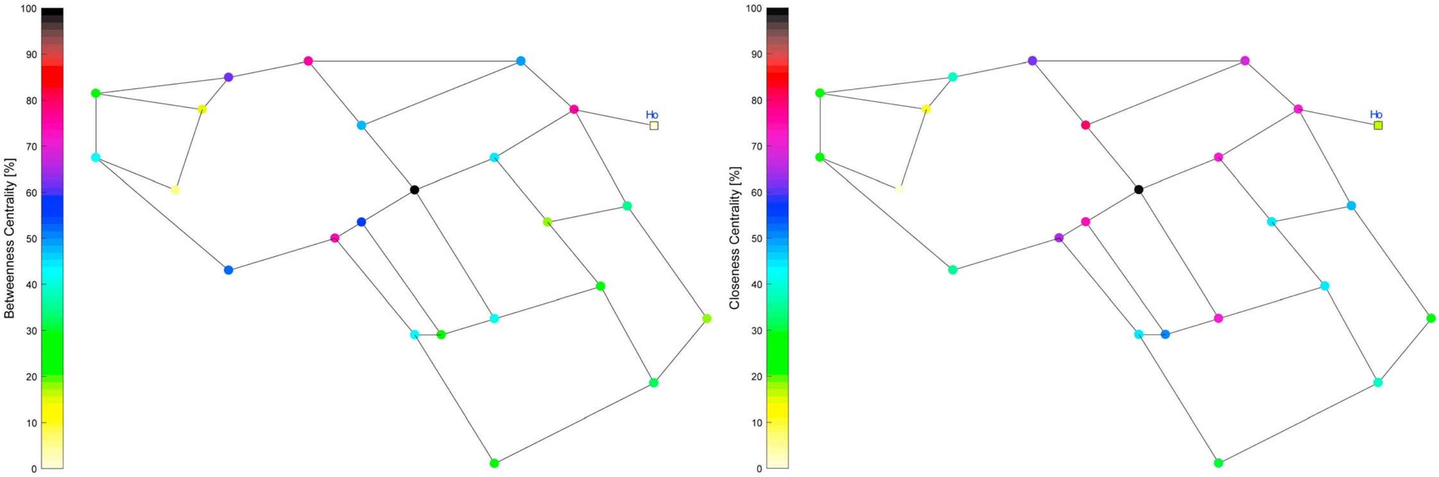

Figure 5. (left) Betweenness and (right) closeness centrality for the Apulian network. Centralities are normalized in the range [0.100].

the water (i.e., information) among pipes. The water is generally delivered at pipe level, although pipe demands are often lumped as outflows to their ending nodes for the hydraulic modeling simplification, generally without impairing the simulation accuracy (Giustolisi, 2010). Therefore, pipes are the most important asset component of WDNs together with the source nodes. Accordingly, the first step for the WDN-tailoring of centrality metrics is to move our focus from nodes to pipes. To this aim, the edge betweenness proposed by Girvan and Newman (2002) is much more consistent for WDN needs than the standard nodal metric. Similarly, the closeness and extended neighbor degree centrality metrics must be tailored considering edges.

4.1. Edge Betweenness, Closeness, and Neighbor Degree

The edge betweenness is conceptually like the *vertex* betweenness centrality described in the previous section. Given an edge *l* and two nodes *s* and *t*, *m* shortest paths exist between *s* and *t*, and a fraction of those shortest paths traverses the edge *l*. The sum of those fractions for all the couples of nodes (*s*,*t*) in the network is the betweenness of edge *l*. In other words, the metric is the summation of the fraction of shortest paths traversing the edge *l* and the formulation is

$$C_l^B = \sum_{\substack{s \neq i \neq t \in V \\ l \in E}} \frac{\sigma_{s,t(l)}}{\sigma_{s,t}} \tag{8}$$

where C_l^B is the edge betweenness of edge *l*, $\sigma_{s,t}(l)$ is the number of shortest paths from node *s* to node *t* passing along the edge *l*, $\sigma_{s,t}$ is the number of all shortest paths from node *s* to node *t*, and *V* and *E* are the set of node and edges belonging to the network. The calculation of the edge and classical *vertex* betweenness is performed with the same procedure and algorithm. The difference for the edge version is that the identification of the shortest paths among each couple of nodes is useful for storing the edge crossings ($\sigma_{s,t}(l)$), and not the node crossings, to compute equation (8).

Figure 6 reports the edge betweenness for the Apulian network. It is evident that the edge betweenness is more meaningful for the analysis of centrality of the WDN because it suggests that the most relevant (traversed) pipes are 4 and 15, which is a different (and hydraulically more revealing) information with respect to the classic betweenness showing that the node 6 is the most relevant (Figure 5, left). The nodal information assesses that the node 6 is strategic for network structure to deliver water. Nevertheless, the network is a real hydraulic system and failing of node 6 corresponds to failing of pipes 6, 7, 8, and 17 because node 6 is not a real technical component of the WDN. In other words, a nodal failure is generally unrealistic, while an interruption of a pipe is a common technical occurrence. Consequently, the edge betweenness is more informative for flow path disruption. Only a very specific class of node, namely, the water source nodes, represents the exception to this statement and needs a node-specific analysis, as will be discussed later in the section referring to the WDN-tailoring of the centrality metrics.

10.1029/2018WR023966

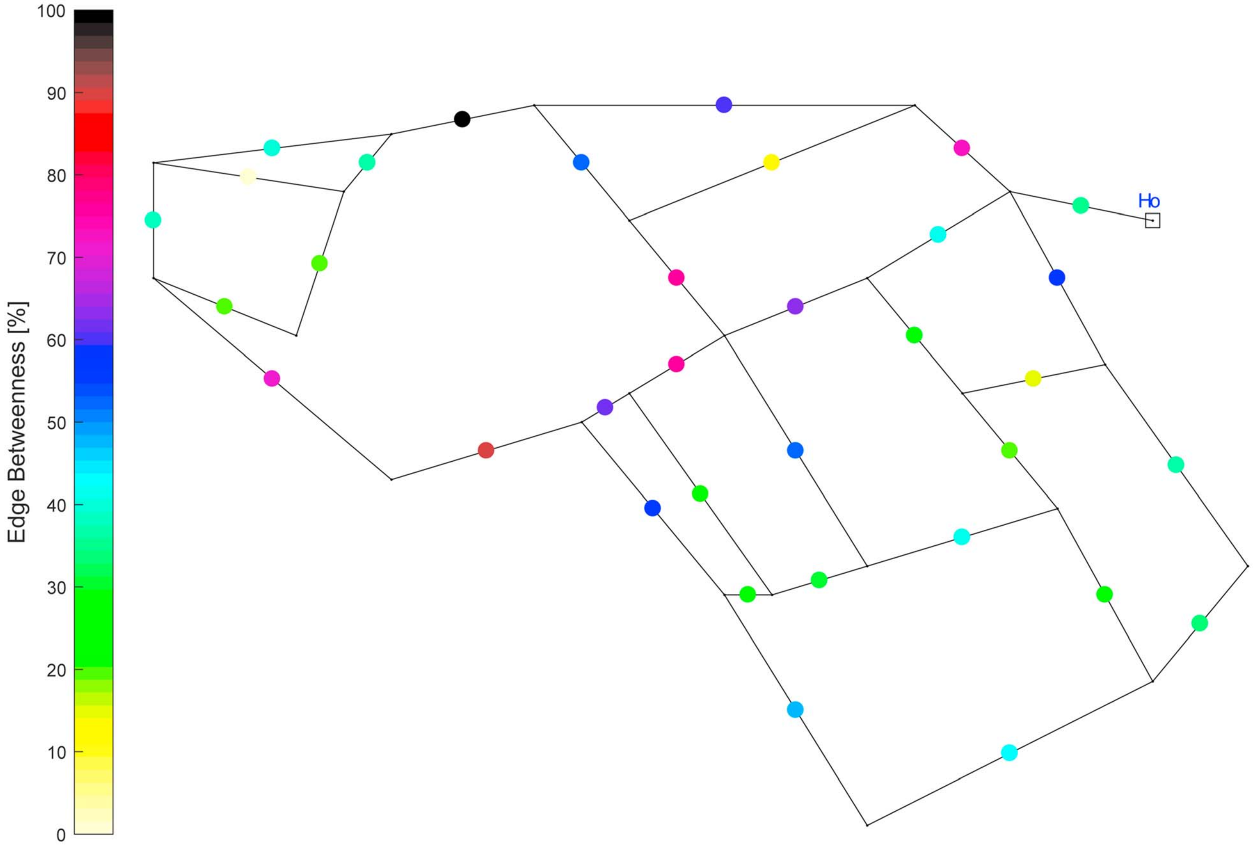

Figure 6. Normalized edge betweenness for the Apulian network.

The edge closeness is a measure of the edge efficiency in spreading the information, and it consists in considering the distance between pipes rather than between nodes. To this aim, one can refer to the line graph or the edge adjacency matrix, which represents the adjacencies between edges of the network (Simone et al., 2016). Therefore, the formulation is

$$C_l^C = \frac{1}{\sum_e d_{l,e}} \tag{9}$$

where $C_l{}^C$ is the edge closeness of the edge l and $\sum d_{l,e}$ is the sum of the steps (path) from edge l to edges e of the network. To compute $C_l{}^C$, a matrix similar to the Table A3 but referred to edge distances needs to be computed for the specific network. In the case of the Apulian network, this matrix has dimension 34 × 34 (being the number of pipes equal to 34) and each (l,e) cell reports the path between edge couple (l,e). The matrix is symmetric for undirected graph, and Dijkstra's algorithm allows one to compute efficiently the shortest paths for each couple of pipes in the network using the line graph matrix, that is, the adjacency matrix of pipes/edges.

Figure 7 reports the edge closeness for the Apulian network and, similarly to the case of Figure 6, it shows that the most important pipes for spreading information (i.e., for distributing water) are 6, 7, 8, and 17. In this case, they have in common the most important node selected by the standard closeness (see the right panel of Figure 5), although the edge closeness clarifies that the four pipes are characterized by a similar relevance.

Finally, the edge degree centrality is the number of edges connected to one of the two ending nodes of that edge, that is, the degree computed using the line graph matrix. The extension of the edge degree (n-EN) is similar to the n-neighborhood degree involving edges at level n. In this case, the rationale beyond n-neighborhood edge degree is to group pipe features with respect to the neighbor connectivity as discussed for nodes in Giustolisi et al. (2017).

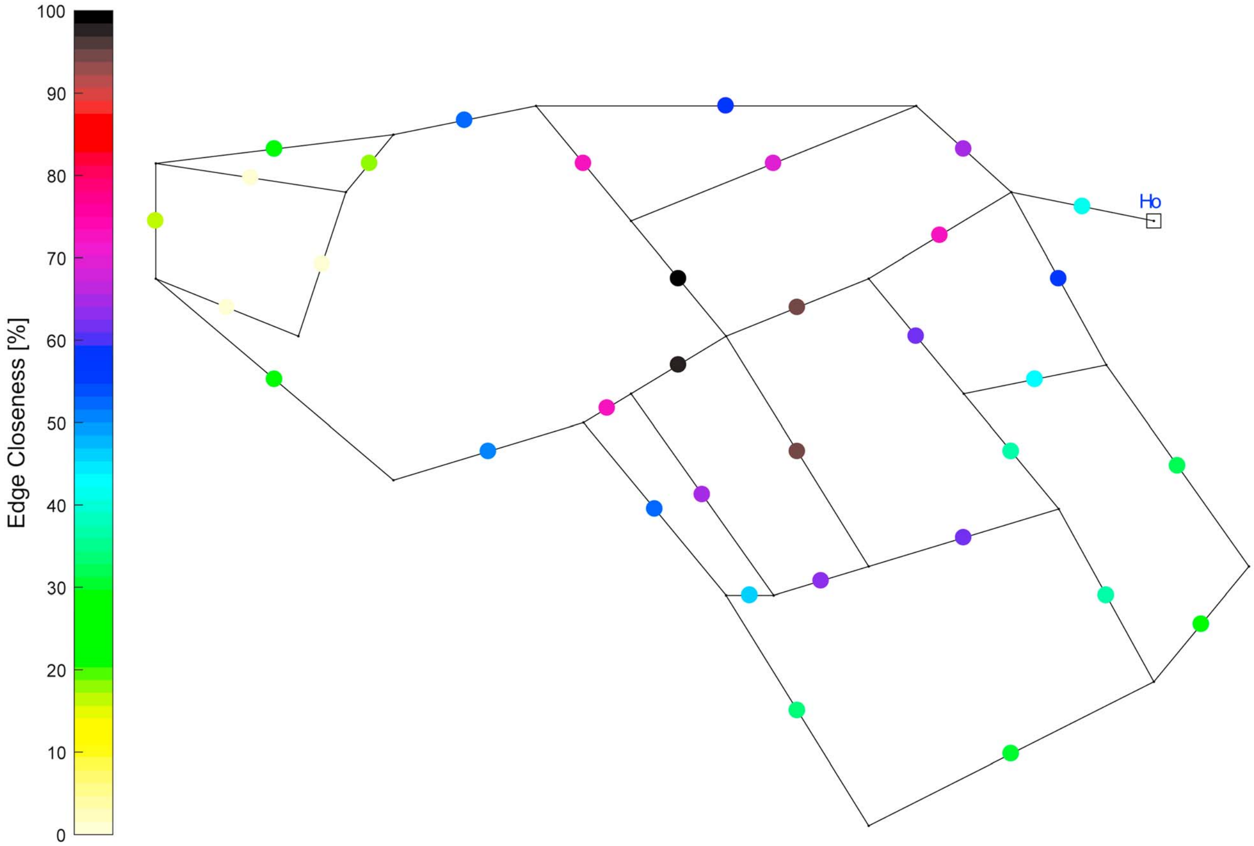

Figure 7. Normalized edge closeness for the Apulian network.

4.2. Hydraulic Metrics for WDNs

CNT defines centrality metrics and allows one to assess the pipe relevance based only on the network connectivity structure (i.e., the topology) described as a graph, that is undirected (i.e., the adjacency matrix is symmetrical) when all the flow path directions are unknown. Otherwise, the graph can be directed for the pipes, for example, where unidirectional devices are installed, to introduce the prior information about the known directions. On the other hand, WDN hydraulic modeling allows one to calculate the hydraulic status of the system, for example, the flow values and paths over time, the nodal pressure, the pipe leakages, etc. Therefore, the graph used in WDN modeling (see section 2) is undirected because the information is not a prior. Considering section 2, we can conceive the hydraulic variables as a sort of centrality metrics generated by coupling WDN domain and the momentum and mass balance equations. Therefore, we can argue that *topological* (e.g., edge betweenness) and *hydraulic* centralities (e.g., pipe discharges) are correlated to each other; namely, the hydraulic network behavior inherits the structural features of the domain coupled with the physically based De Saint Venant equations, although simplified. From this perspective, the information gained from the topological centralities can be helpful (i) for inferring in advance the *emerging* hydraulic behavior without the need of hydraulic simulations independently on the specific boundary conditions and (ii) to compare the real hydraulic behavior with the domain analysis to achieve further information about the effect of the boundary conditions driving the specific hydraulic status.

In addition, domain analysis through CNT WDN-tailored metrics can allow understanding in advance the role of pipes in determining the system hydraulics independently on the specific set of boundary conditions, which can be, for example, helpful for districtualization tasks (Apulian-T).

The previous parallel between topology-based centralities and hydraulics-based centrality can be very useful for the WDN analysis. However, some key differences occur. For instance, the pipe flow rate centrality informs about the most relevant pipes transferring water similarly to the edge betweenness, but this latter topologic-driven metric lacks information about the source node. Moreover, despite topological edge

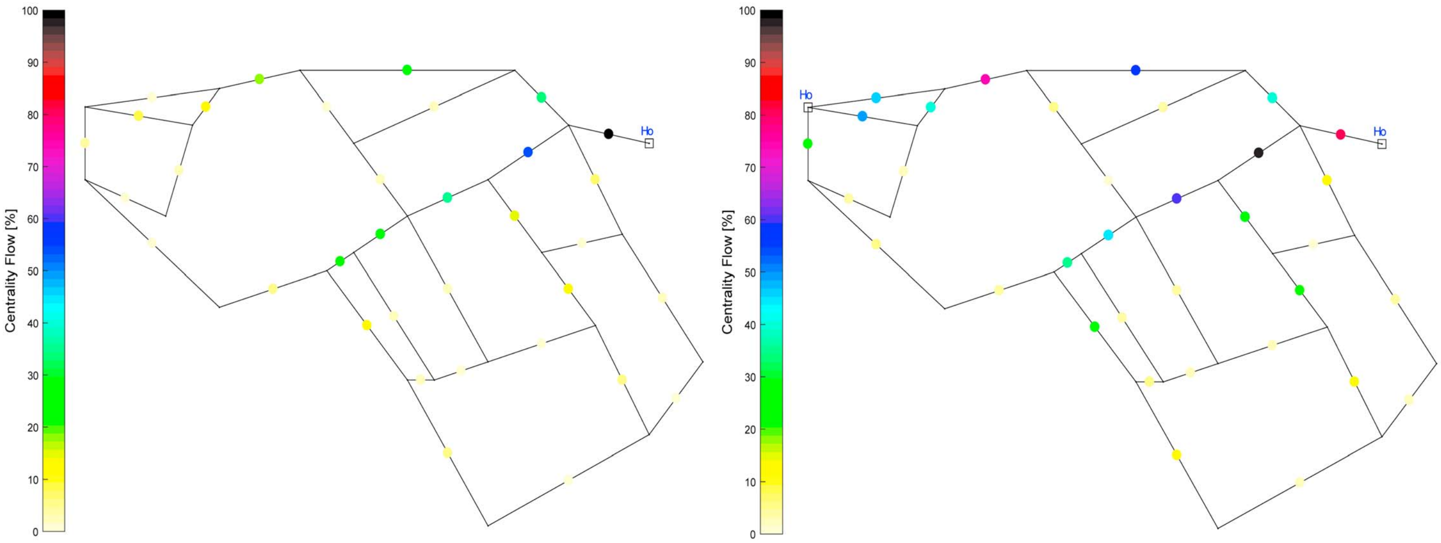

Figure 8. Normalized flow-based centrality considering (left) one and (right) two reservoirs.

betweenness is a global measure as well as the hydraulic metric *flow rate*, it does not consider pipe features (e.g., length and hydraulic resistances) that instead are embedded in the hydraulic metric.

To show some examples of hydraulics-based centralities and compare them with the topology-based ones, let us add in node 10 a second reservoir to the Apulian network having piezometric head equal to the original one (see Figure 8, right). The internal diameter of pipes 9, 10, and 12 are increased to 0.327 m considering that the internal diameter of pipe 34 is equal to 0.368 m. Looking at the Apulian network with one reservoir of Figure 8 (left) it is trivial to note that the flow rate metric informs that pipe 34 is the most relevant.

Figure 8 (right) shows that the flow rate metric for the Apulian network with two reservoirs provides nontrivial information. In fact, Figure 8 (right) shows that pipe 6 is the most important component because the presence of the second reservoir that reduces the importance of pipe 34 with respect to the one-reservoir case. This fact means that the hydraulic metric informs that pipe 6 is relevant for both reservoirs. Another nontrivial information of the Figure 8 (right) is the fact that each single pipe (9, 10, 12) connected to the added reservoir is less important than the pipe 34 because they share the relevance. Finally, Figure 8 (right) highlights the most relevant path of pipes connecting the two reservoirs.

On the contrary, the edge betweenness shown in the Figure 6 does not change adding the second reservoir and it is not informative about the hydraulic relevance of pipes because the connectivity structure does not account for water sources. This fact does not mean that CNT centrality metrics cannot support the analysis of WDNs but that we need to consider the hydraulic differences among components represented by nodes in the specific networked system (Giustolisi et al., 2017). In fact, source nodes represent a sort of hubs for WDNs and the pipes are real network components, which are characterized by length, diameter, hydraulic resistance, supplied water, etc.

Previous examples show that edge topology-based centralities must be further tailored so that they can be fully useful for WDN applications. Therefore, in the next section we will propose a hydraulic-oriented tailoring of the topological centralities based (i) on a suitable modification of the connectivity structure of the WDNs and (ii) on weighting topological centralities, where weights can embed different pipe hydraulic features depending on the analysis being performed.

5. Tailoring Centrality Metrics for WDNs: Hydraulic-Based Topology and Edge Weighting

Giustolisi et al. (2017) proposed a basic classification of the network connectivity structure of WDNs describing different shapes of the neighborhood degree distribution (random, small world, and scale free) for 22 WDNs. They demonstrated that the Poisson distribution (random network) generally models WDNs very

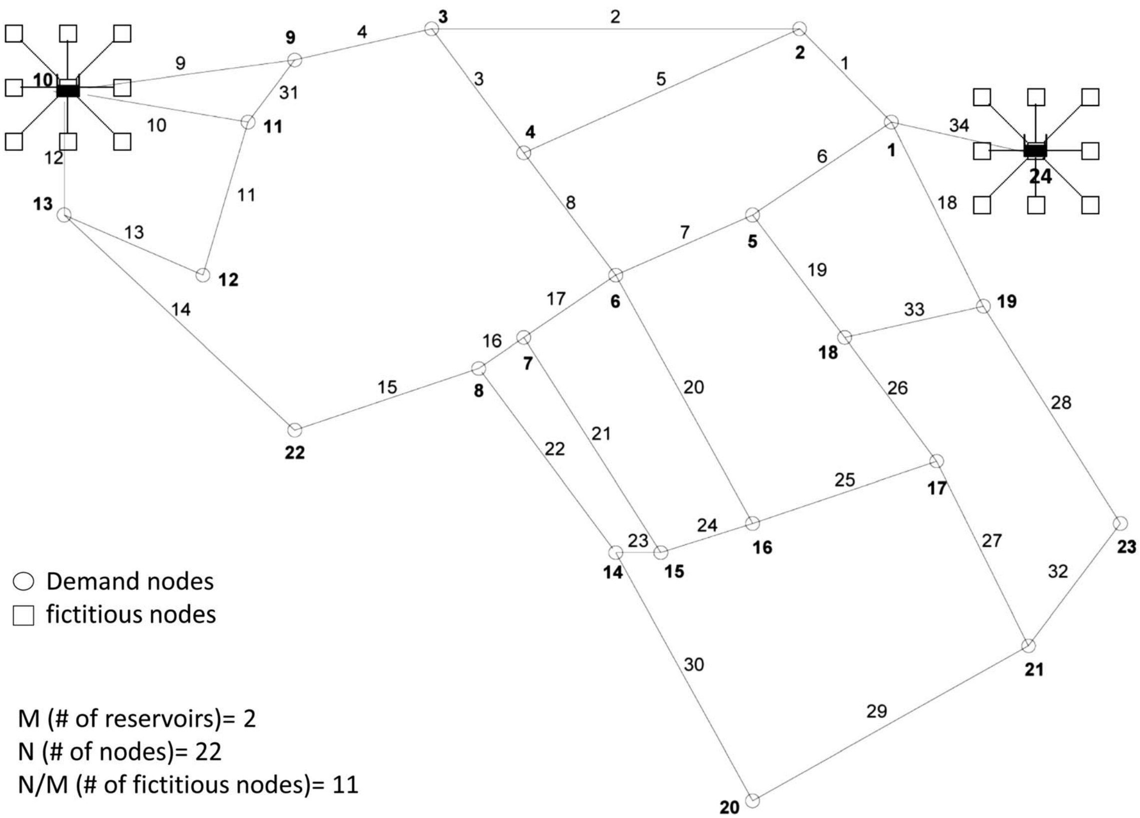

Figure 9. Apulian network modified with stars of fictitious nodes introduced where source nodes occur (nodes 10 and 24).

well and discussed the fact that beyond each node there is a different hydraulic object (water source, demand outflow, and connectivity of pipes) despite the topological classification findings. Namely, there is the need of tailoring the connectivity structure of WDNs to account for source, demand, and connection nodes (according to the node nomenclature recalled in the Introduction section) to make CNT tools really informative of the emerging hydraulic behavior; otherwise, those tools remain too generic to be useful for networked hydraulic systems. In this section, we will describe tailoring of the network topology to embed node hydraulic features. Moreover, we will discuss edge weights to consider that pipes are material components characterized by different asset features. Notice that the edge betweenness is always the number of times a pipe is traversed, while the edge weights make different the identified shortest paths.

5.1. Source Nodes

The first issue for tailoring topological edge centralities to WDNs is to represent source nodes in the connectivity structure to obtain a hydraulics-based topology. The strategy is to consider for each source node a star of N_f connected fictitious nodes (see Figure 9). The number N_f is assumed a round integer of N_d/M, where N_d is the number of demand nodes and M is the number of source nodes; notice that rounding to the lower or greater integer is irrelevant for the result. In this way, the source nodes act as a sort of hubs with respect to the hydraulic system behavior, assuming the same relevance for each one in supplying water to demand nodes despite the topological position and other unavailable hydraulic information.

As the shortest paths will pass at least N_f times across each source node, the proposed approach corresponds to multiply by N_f the topological relevance of source nodes with respect to the original connectivity structure. This means that the original source node position relevance with respect to the entire network structure is preserved although amplified according to N_f. It is worth noting that in absence of source nodes, the connectivity structure degenerates to the original one, which is consistent with the fact that the hydraulic system and the hydraulics-based topology do not exist without source nodes. Figure 9 exemplifies the described strategy using the Apulian network with two reservoirs. In this case, $N_d = 22$ and, then, $N_f = 12$ for each reservoir.

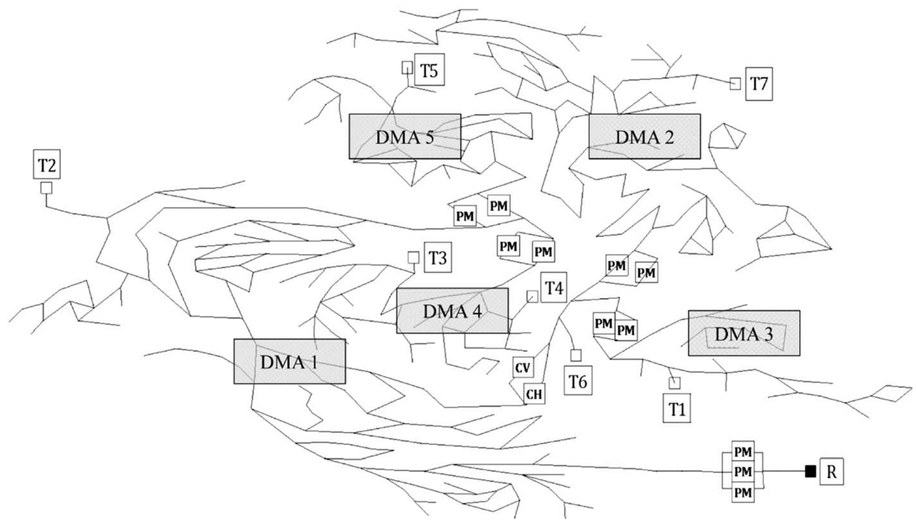

Figure 10. BBLAWN layout composed of five district metering areas (DMAs).

Notice that the repartition of the demand nodes in the same number of fictitious nodes for each source node is the basic choice if no information is available about the actual influence area of reservoirs. However, if some data are known (also approximately) about the real reservoir influence area, the strategy can be refined adopting a different value of N_f for each reservoir depending on its hydraulic relevance.

6. Case Study and Dealing With Tanks and Directional Devices

Without loss of generality, the strategy of tailoring described in the previous section is here applied and discussed using the edge betweenness centrality and the network BBLAWN (Giustolisi et al., 2016). This WDN is more complex than the Apulian one and allows us to discuss in more detail the strategy. The network layout is reported in Figure 10, and it is composed of 389 internal nodes, 444 pipes, one reservoir (R), seven tanks (Tx), 11 pumps (PM), one control valve (CV), one check valve (CH), and five DMAs.

Figure 11 shows the standard and the WDN-tailored edge betweenness, which is computed using the hydraulic-based topology without pipe weights. The difference between standard and the WDN-tailored metrics is evident and derives from the attribution of a different relevance to the single reservoir with respect to internal nodes. The introduction of a star of fictitious nodes (in this case $N_f = N_d$ as $M = 1$) allows the single source node of acting as a topological hub in the networked system. As consequence, the pipes relevance of the paths starting from the reservoir increases, while the other pipes relevance does not change considerably. Therefore, the tailored metric becomes able to capture through the domain analysis the relevant hydraulic features of the system with respect to the source nodes, while the standard metric fails in ranking correctly pipes close to the reservoir.

Attributing weights to the pipes, the WDN-tailored edge betweenness metric can be further refined considering the other domain features as reported in the section “CNT and study of the WDN hydraulics domain.” The pipe internal diameter and length are usually known after taking a bearing of the network topology and of the system components (reservoir, tanks, devices, etc.). The pipe resistance strongly depends on those asset data as the fifth power of the internal diameter. Figure 12 (left) shows the metric computed using pipes hydraulic resistance as weights, which makes the analysis closer to the hydraulic of the system because the resistances drive the water fluxes. In fact, the refining identifies an alternative path (see red circle) with respect to Figure 11 (right).

Figure 12 (right) reports the pipe centrality based on the average pipe flows of the weekly operating cycle of the BBLAWN. The comparison of Figure 12 demonstrates that the domain analysis using the WDN-tailored betweenness is useful to assess the hydraulic pipe relevance. Therefore, (i) the modification of the network topology considering source nodes and (ii) weighting edge betweenness by pipe resistance allows one to obtain an effective domain analysis, which captures the emerging hydraulic behavior of the system.

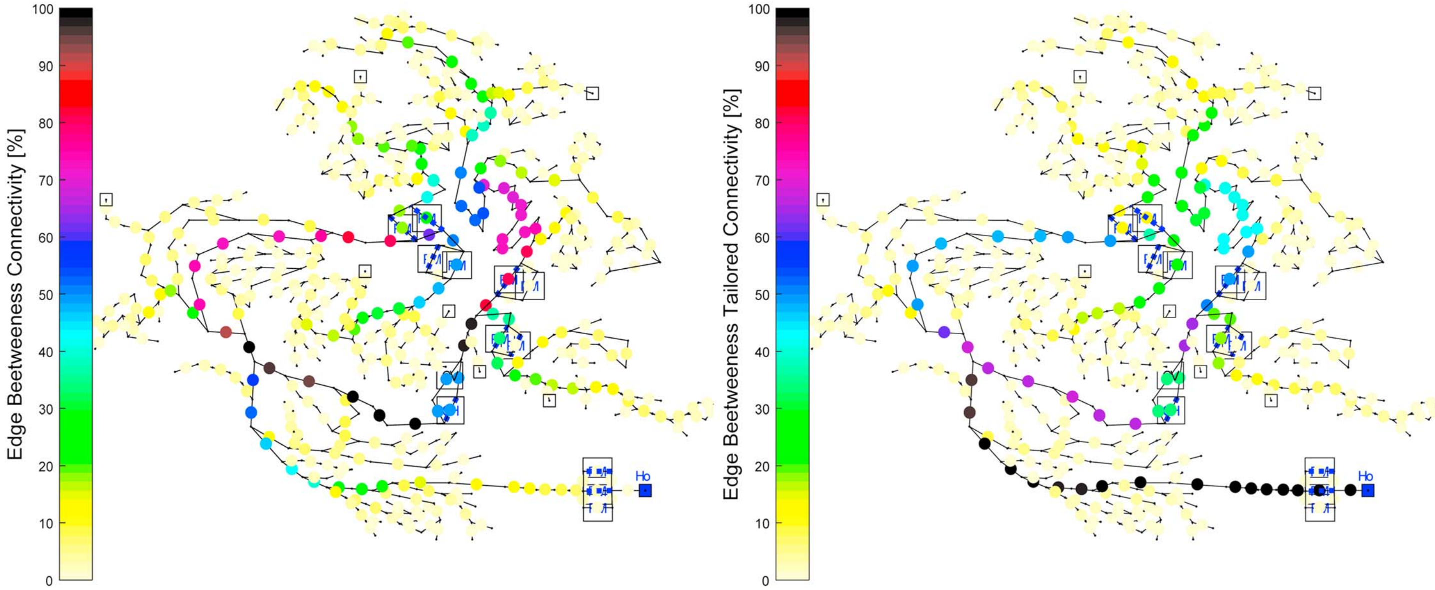

Figure 11. Normalized standard (left) and water distribution network-tailored (right) edge betweenness.

However, the topological metric fails close to tanks because, similarly to reservoirs, there is the need of embedding the domain information about those specific nodal sources in the connectivity structure. In other words, it is necessary to consider those specific nodal sources which are different from reservoirs because an emptying/filling process occurs during the operating cycle of the hydraulic system. It follows that another step is necessary to tailor topological centrality metrics for WDN domain analysis. In addition, the domain can be modified by the presence of devices allowing water to flow only in one direction: e.g., pumps, pressure reduction valves, check valves, etc., that is, directional devices.

6.1. Tank Nodes and Directional Devices

The starting point is that a tank is characterized by a volume (V_T). It is the maximum volume that can be supplied to demand nodes during the operating cycle when the emptying process occurs. In this sense, the hydraulic role played by tanks is similar to source nodes. This fact can be accounted by modifying the

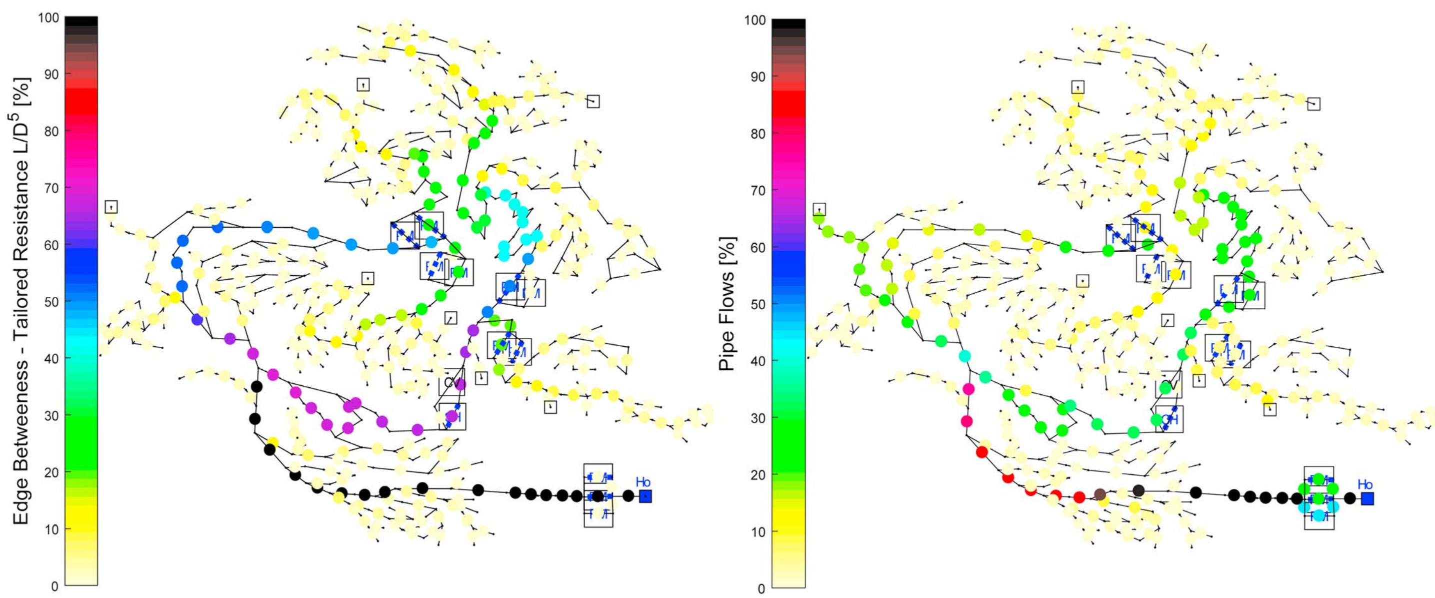

Figure 12. Water distribution network-tailored edge betweenness using pipes resistance (left) and pipe flow (right).

10.1029/2018WR023966

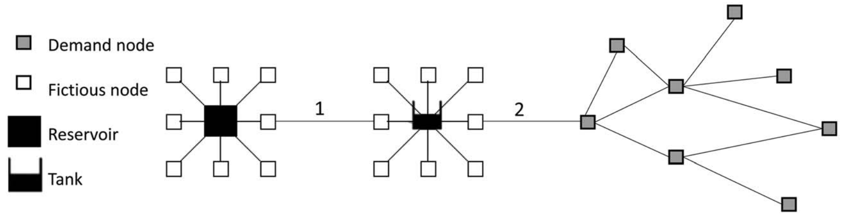

Figure 13. Scheme of the water distribution network (WDN)-oriented topology modifications to account reservoirs and tanks.

network topology by a suitable star of N_f connected fictitious nodes (see Figure 13) depending on the tank volume and the number of demand nodes because of the emptying/filling process. Therefore, to evaluate the number of these fictitious nodes we compare the volume V_T (maximum water volume supplied during the emptying process) with the average water volume supplied during the operating cycle to the demand nodes (V_D), considering the constraint of N_f fictitious nodes already attributed to reservoirs. It is then possible to attribute to each tank a star of fictious nodes equal to the minimum integer value between $N_d{\cdot}V_T/V_D$ and N_f, where N_d is the number of demand nodes. This way, the tanks relevance is linearly depending on the ratio of the volume and of the required demand volume in the operative cycle.

Note that as the volume V_T increases with respect to V_D, the tank node tends to the source node in the connectivity structure because it has a similar feature toward demand nodes and, possibly, N_f needs to be iteratively determined considering that a tank node acts as a source node. On the other hand, decreasing the volume V_T the tank node tends to a demand or connectivity node. In any case, as for the reservoirs, we account for the fact that tanks act as a sort of hubs with respect to the hydraulic system behavior and that their relevance depends on their specific volume V_T. It is important to note that, generally, tanks are less relevant hubs than reservoirs. The relevance of a tank can be less than or equal to that of a reservoir but never greater, because the number of their fictitious nodes, can be equal or less to those traversing the reservoir.

In the same way, the relevance of pipes close to the tanks depends on V_T. Figure 13 illustrates the relationship between pipe relevance and V_T. If V_D is much larger than V_T, pipe 1 is much more relevant than the pipe 2 because it is still the main source of supply for the network. Instead, if V_T is comparable with V_D, pipe 2 becomes more relevant than the pipe 1 because even in the absence of pipe 1, the pipe 2 is still able to ensure the supply of the network for an operating cycle, i.e., for the time necessary for repairing/replacing the pipe 1.

Figure 14 (left) shows the same metric of Figure 12 (left) but modifying the connectivity structure of the network in order to account the tank information as above discussed. It is worth to note that the metric accounting for tanks increases the domain information about the relevance of pipes connecting the reservoir with the tank T2, differently from Figure 12 (left) where the tank information was not used. The previous consideration is still valid for the other tanks although they are characterized by lower V_T; that is, they are less relevant hubs with respect to the other tank and the reservoir.

However, the domain analysis with the improved WDN tailored metric does not capture the fact that the most relevant pipe for WDN hydraulics is the one at next to the single reservoir. Figure 14 (left) shows that the most important pipes are those that deliver water to DMA2 and DMA3 from tank T2 and the reservoir, but this domain analysis attributes a lower relevance to the line starting from the main source of water.

This drawback is solved accounting for the directional devices as demonstrated by Figure 14 (right). In fact, the most relevant pipe from the domain study is now the one next to the single reservoir, while the relevance of the other pipes does not change although it has a new parametrization referred to the most important one. Notice that this very good accordance with hydraulic behavior is not strictly related to the presence of the pump installation close to the single reservoir, but to the fact that the direction of the flow into the pipe next to reservoir is a prior, assuming that pumps exist, or the water level is sufficient to supply water to the network. Consequently, the previous finding indicates that introducing the flow direction for pipes next to reservoirs (assuming that the hydraulic condition allows the prior) enhanced the domain analysis. Finally, it is worth noting for the specific case study that the directional devices and the prior close to the reservoir play the same role for the domain analysis.

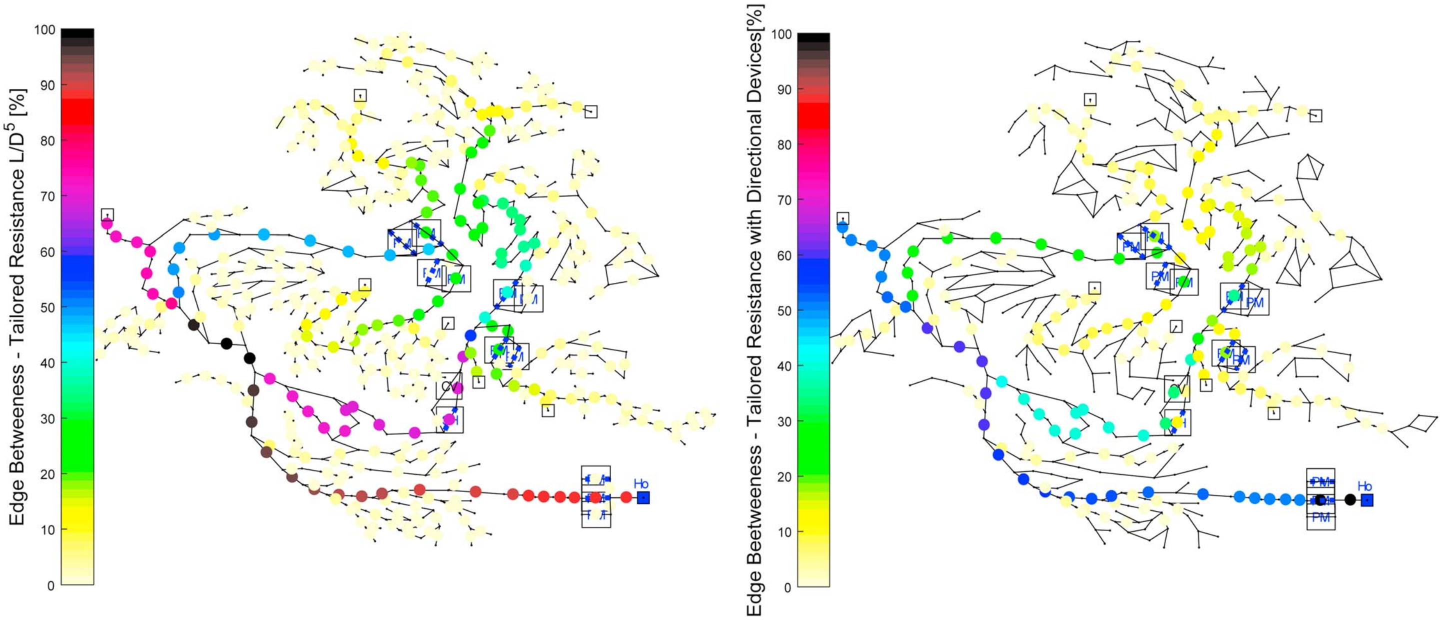

Figure 14. Water distribution network (WDN)-tailored edge betweenness using pipes resistance and tank information. The right panel considers also information about directional devices.

7. Domain Analysis of a Large Size WDN: The Apulian-T Case

The WDN-Tailored Edge Betweenness is here applied and discussed using another real hydraulic networked system. The WDN, that we will name Apulian-T for confidentiality, has a length of 440 km and it is composed of 7,217 nodes, 8,496 pipes, 3 reservoirs, and 5 main suburban pipes. The pressure inside the WDN is controlled by five pressure reduction valves (see red circles in the Figure 15) installed at the end of the five suburban pipes. The WDN has several closed gates because DMAs have been planned. In comparison with the BBLAWN, the Apulian-T is a large size hydraulic system with greater complexity than the BBLAWN due to the multisources (three reservoirs), the five controlled entries, the internal very looped network, a wide range of diameters (spanning from 1,200 to 40 mm) and lengths of pipes. Despite the large network size, the calculation of the proposed tailored metric takes about 60 s on a standard laptop equipped with a CPU Intel i7, which is very satisfactory for a domain analysis to be performed one time.

Figure 15 (left) shows the most relevant pipes corresponding to the values of edge betweenness that fall in the highest 5% of the maximum value, while, similarly, Figure 15 (right) highlights the pipes where the flow is greater than 5% with respect to the maximum value about equal to 500 L/s, that is, the flow metric below defined, reminding that we refer to the average values of the daily hydraulic simulation. Comparing the two panels of Figure 15, it is evident that the proposed tailored-metric identifies quite all the most important pipes as ranked by pipe flows corresponding to the daily hydraulic simulation. Notice that the metric does not identify the five suburban pipes only, but also the most important internal paths inside the looped part of Apulian-T.

The comparison of the two panels of Figure 15 is then really positive, taking into account that the proposed metric focuses only on the domain structure, while the pipe flows obviously depend also on the boundary conditions and the physical principles contained in the momentum and continuity equations. However, despite these lacking, the proposed metric, properly tailored and weighted, can capture the emerging hydraulic behavior due to the connectivity structure of the network. The good performance of the tailored edge betweenness is confirmed by the Spearman correlation index (Spearman, 1904) between the pipe ranking obtained by the proposed metric and the one corresponding to the pipe flows; this index is equal to 0,60, which confirms that the information contained in the domain is relevant for the WDN hydraulics. In order to evaluate in a different way, the importance of information contained in the domain structure, the diagram in Figure 16 compares the distribution of the pipe flows (ordered from the maximum value) with the corresponding distribution of the tailored edge betweenness. The agreement is very good (see Figure 16, left),

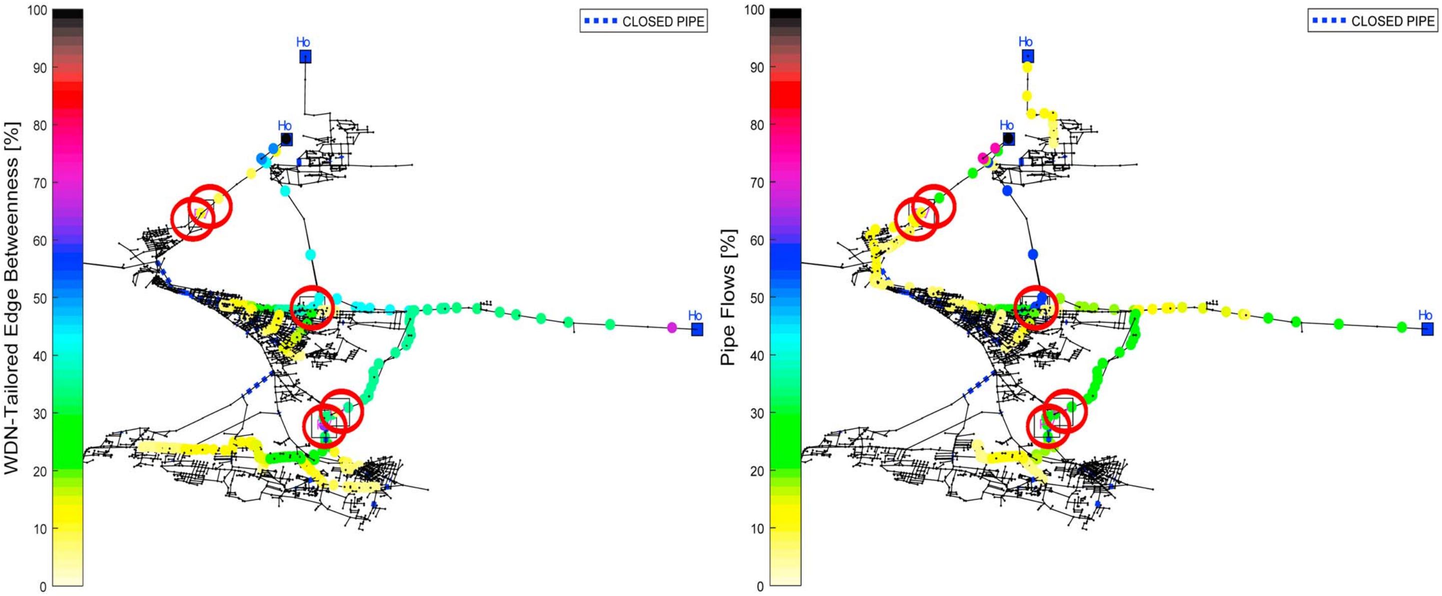

Figure 15. Water distribution network (WDN)-tailored edge Betweenness of Apulian-T. The values greater than 5% of the maximum values (of edge betweenness or flow) are reported.

demonstrating also for this large size WDN that the hydraulic behavior is strongly dependent on the domain features embedded in the proposed metric. The *x* axis in logarithmic scale (Figure 16, right) better shows the similar trend of the highest values, that is, it confirms the comparison based on the panels of Figure 15.

The Spearman correlation index was also evaluated for the previous case study—that is, the BBLAWN-obtaining a value around 0.74. This higher value is expected because it is reasonable that the correlation index is a function of the network complexity and of the number of water sources. In fact, the relevance of the suburban pipes of the water sources depends on the hydraulic influence (as portion of the supplied network) of each one, and such influence region generally varies over time and is not only related to the topological position. For example, the boundary conditions, such demand values and reservoir levels, affect the flow into the suburban pipes and they are not information contained in the domain analysis tool. In any case, we performed this correlation analysis for many real WDNs obtaining values ranging in the interval [0.55; 0.80]; this confirms that the domain analysis using the WDN-Tailored Edge Betweenness is robust, and it allows one to capture the core of the hydraulic behavior at a good extend for a wide spectrum of WDN typologies.

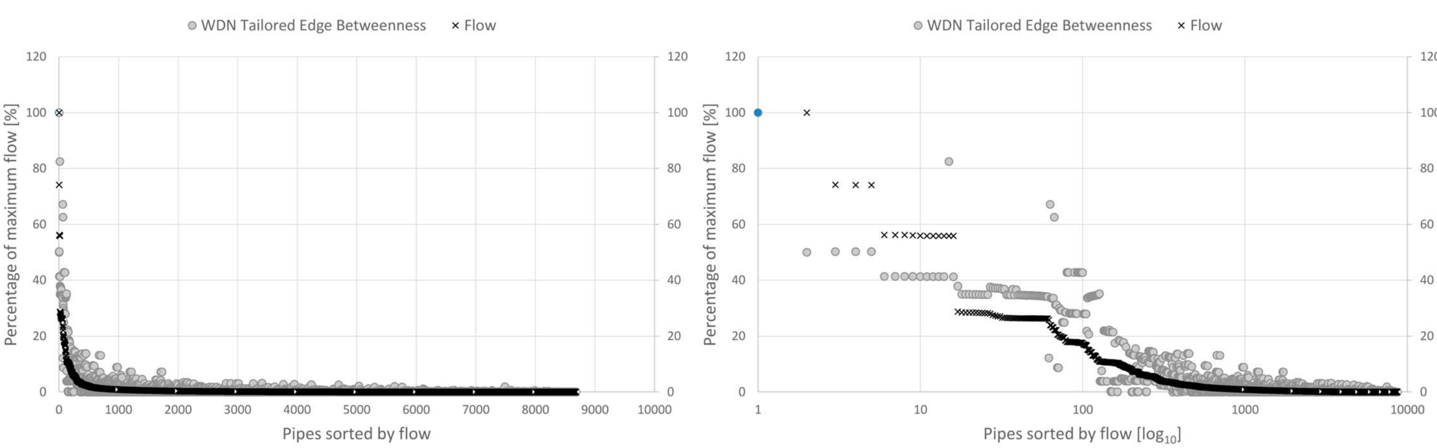

Figure 16. Comparison between the pipe flows and the tailored edge betweenness.

10.1029/2018WR023966

8. Conclusions

The Betweenness centrality can be used for analyzing the WDN domain if carefully tailored to consider the hydraulic and topologic characteristics of these spatial networked systems. This way, the use of CNT can represent a complementary and effective tool for supporting analysis, design, and management tasks and not only a fashionable (but fruitless) exercise.

The tailoring process proposed in this work consists of three conceptual steps. The first one is to move from a node-based perspective (typical of usual centrality problems) to an approach where pipes are the right key *brick* of the network and; therefore, centralities must concern the edges. The second step is to create a WDN-tailored network topology to embed the different hydraulic roles of nodes: source nodes, tanks, and demand nodes. The last step is to evaluate edge centrality in the tailored topology by considering the more suitable centrality metrics to explore hydraulic networks (betweenness, closeness, and degree) and appropriate weights that contain domain information (e.g., pipe resistance). The last step should involve also the prior about the flow direction in the pipes where directional devices are installed or for which the hydraulic condition constrain the flow direction as, for example, next to the reservoirs.

Appendix A

Table A1 shows the identification of the shortest paths between each couple of nodes and traversing node 6 for the network reported in Figure 1.

Table A1
All Shortest Paths Passing Through Node 6 and a Column (Bold Number) That Reports the Total Number of Shortest Paths for Each Couple of Nodes

Shortest paths through node 6 and number of shortest paths from node s to node t (σ_{st})

1 5 6 7	**1**	4 6 5 18	**1**		
1 5 6 7 8	**1**	4 6 7 8 14 20	**2**	8 7 6 5 1 24	**1**
1 5 6 7 8 14	**4**	4 6 16 17 21	**1**	9 3 4 6 7 15	**2**
1 5 6 7 15	**2**	4 6 7 8 22	**1**	9 3 4 6 16	**1**
1 5 6 16	**1**	5 6 7	**1**	9 3 4 6 16 17	**1**
1 5 6 7 8 22	**1**	5 6 7 8	**1**	10 9 3 4 6 16	**1**
2 4 6 7	**1**	5 6 7 8 22 13	**1**	10 9 3 4 6 16 17	**1**
2 4 6 7 8	**1**	5 6 7 8 14	**2**	11 9 3 4 6 7 15	**6**
2 4 6 7 8 14	**3**	5 6 7 15	**1**	11 9 3 4 6 16	**1**
2 4 6 7 15	**2**	5 6 16	**1**	11 9 3 4 6 16 17	**1**
2 4 6 16	**1**	5 6 7 8 22	**1**	12 11 9 3 4 6 16	**3**
3 4 6 7	**1**	7 6 4 3 9	**1**	12 11 9 3 4 6 16 17	**5**
3 4 6 7 8	**1**	7 6 4 3 9 11	**3**	13 22 8 7 6 16	**1**
3 4 6 7 8 14	**3**	7 6 16	**1**	13 22 8 7 6 16 17	**2**
3 4 6 7 15	**2**	7 6 16 17	**1**	13 22 8 7 6 5 18	**1**
3 4 6 16 0	**1**	7 6 5 18	**1**	14 8 7 6 5 1 24	**4**
3 4 6 16 17	**1**	7 6 5 1 19	**2**	15 7 6 5 1 24	**2**
4 6 5	**1**	7 6 16 17 21	**1**	16 6 7 8 22	**2**
4 6 7	**1**	7 6 5 1 19 23	**4**	16 6 5 1 24	**1**
4 6 7 8	**1**	7 6 5 1 24	**1**	17 16 6 7 8 22	**1**
4 6 7 8 14	**3**	8 7 6 16	**1**	18 5 6 7 8 22	**1**
4 6 7 15	**2**	8 7 6 16 17	**1**	19 1 5 6 7 8 22	**3**
4 6 16	**1**	8 7 6 5 18	**1**	22 8 7 6 5 1 24	**1**
4 6 16 17	**1**	8 7 6 5 1 19	**3**		

Table A2 shows the (symmetric) 24 × 24 matrix of the distances for the network reported in Figure 1 ($N = 24$). The sum of the elements in the ith row (or column) gives the sum of distances from node i to all other nodes in the network.

Table A2
Matrix of Distance for Apulian Network

$$\begin{bmatrix}
0 & 1 & 2 & 2 & 1 & \mathbf{2} & 3 & 4 & 3 & 4 & 4 & 5 & 5 & 5 & 4 & 3 & 3 & 2 & 1 & 4 & 3 & 5 & 2 & 1 \\
1 & 0 & 1 & 1 & 2 & \mathbf{2} & 3 & 4 & 2 & 3 & 3 & 4 & 4 & 5 & 4 & 3 & 4 & 3 & 2 & 5 & 4 & 5 & 3 & 2 \\
2 & 1 & 0 & 1 & 3 & \mathbf{2} & 3 & 4 & 1 & 2 & 2 & 3 & 3 & 5 & 4 & 3 & 4 & 4 & 3 & 6 & 5 & 4 & 4 & 3 \\
2 & 1 & 1 & 0 & 2 & \mathbf{1} & 2 & 3 & 2 & 3 & 3 & 4 & 4 & 4 & 3 & 2 & 3 & 3 & 3 & 5 & 4 & 4 & 4 & 3 \\
1 & 2 & 3 & 2 & 0 & \mathbf{1} & 2 & 3 & 4 & 5 & 5 & 6 & 5 & 4 & 3 & 2 & 2 & 1 & 2 & 4 & 3 & 4 & 3 & 2 \\
\mathbf{2} & \mathbf{2} & \mathbf{2} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{4} & \mathbf{5} & \mathbf{4} & \mathbf{3} & \mathbf{2} & \mathbf{1} & \mathbf{2} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{3} & \mathbf{3} & \mathbf{4} & \mathbf{3} \\
3 & 3 & 3 & 2 & 2 & \mathbf{1} & 0 & 1 & 4 & 4 & 5 & 4 & 3 & 2 & 1 & 2 & 3 & 3 & 4 & 3 & 4 & 2 & 5 & 4 \\
4 & 4 & 4 & 3 & 3 & \mathbf{2} & 1 & 0 & 4 & 3 & 4 & 3 & 2 & 1 & 2 & 3 & 4 & 4 & 5 & 2 & 3 & 1 & 4 & 5 \\
3 & 2 & 1 & 2 & 4 & \mathbf{3} & 4 & 4 & 0 & 1 & 1 & 2 & 2 & 5 & 5 & 4 & 5 & 5 & 4 & 6 & 6 & 3 & 5 & 4 \\
4 & 3 & 2 & 3 & 5 & \mathbf{4} & 4 & 3 & 1 & 0 & 1 & 2 & 1 & 4 & 5 & 5 & 6 & 6 & 5 & 5 & 6 & 2 & 6 & 5 \\
4 & 3 & 2 & 3 & 5 & \mathbf{4} & 5 & 4 & 1 & 1 & 0 & 1 & 2 & 5 & 6 & 5 & 6 & 6 & 5 & 6 & 7 & 3 & 6 & 5 \\
5 & 4 & 3 & 4 & 6 & \mathbf{5} & 4 & 3 & 2 & 2 & 1 & 0 & 1 & 4 & 5 & 6 & 7 & 7 & 6 & 5 & 6 & 2 & 7 & 6 \\
5 & 4 & 3 & 4 & 5 & \mathbf{4} & 3 & 2 & 2 & 1 & 2 & 1 & 0 & 3 & 4 & 5 & 6 & 6 & 6 & 4 & 5 & 1 & 6 & 6 \\
5 & 5 & 5 & 4 & 4 & \mathbf{3} & 2 & 1 & 5 & 4 & 5 & 4 & 3 & 0 & 1 & 2 & 3 & 4 & 4 & 1 & 2 & 2 & 3 & 6 \\
4 & 4 & 4 & 3 & 3 & \mathbf{2} & 1 & 2 & 5 & 5 & 6 & 5 & 4 & 1 & 0 & 1 & 2 & 3 & 4 & 2 & 3 & 3 & 4 & 5 \\
3 & 3 & 3 & 2 & 2 & \mathbf{1} & 2 & 3 & 4 & 5 & 5 & 6 & 5 & 2 & 1 & 0 & 1 & 2 & 3 & 3 & 2 & 4 & 3 & 4 \\
3 & 4 & 4 & 3 & 2 & \mathbf{2} & 3 & 4 & 5 & 6 & 6 & 7 & 6 & 3 & 2 & 1 & 0 & 1 & 2 & 2 & 1 & 5 & 2 & 4 \\
2 & 3 & 4 & 3 & 1 & \mathbf{2} & 3 & 4 & 5 & 6 & 6 & 7 & 6 & 4 & 3 & 2 & 1 & 0 & 1 & 3 & 2 & 5 & 2 & 3 \\
1 & 2 & 3 & 3 & 2 & \mathbf{3} & 4 & 5 & 4 & 5 & 5 & 6 & 6 & 4 & 4 & 3 & 2 & 1 & 0 & 3 & 2 & 6 & 1 & 2 \\
4 & 5 & 6 & 5 & 4 & \mathbf{4} & 3 & 2 & 6 & 5 & 6 & 5 & 4 & 1 & 2 & 3 & 2 & 3 & 3 & 0 & 1 & 3 & 2 & 5 \\
3 & 4 & 5 & 4 & 3 & \mathbf{3} & 4 & 3 & 6 & 6 & 7 & 6 & 5 & 2 & 3 & 2 & 1 & 2 & 2 & 1 & 0 & 4 & 1 & 4 \\
5 & 5 & 4 & 4 & 4 & \mathbf{3} & 2 & 1 & 3 & 2 & 3 & 2 & 1 & 2 & 3 & 4 & 5 & 5 & 6 & 3 & 4 & 0 & 5 & 6 \\
2 & 3 & 4 & 4 & 3 & \mathbf{4} & 5 & 4 & 5 & 6 & 6 & 7 & 6 & 3 & 4 & 3 & 2 & 2 & 1 & 2 & 1 & 5 & 0 & 3 \\
1 & 2 & 3 & 3 & 2 & \mathbf{3} & 4 & 5 & 4 & 5 & 5 & 6 & 6 & 6 & 5 & 4 & 4 & 3 & 2 & 5 & 4 & 6 & 3 & 0
\end{bmatrix}
=
\begin{bmatrix}
69 \\ 70 \\ 72 \\ 66 \\ 69 \\ \mathbf{61} \\ 68 \\ 71 \\ 81 \\ 88 \\ 95 \\ 101 \\ 88 \\ 78 \\ 76 \\ 69 \\ 78 \\ 78 \\ 77 \\ 84 \\ 81 \\ 82 \\ 85 \\ 91
\end{bmatrix}$$

Note. The sixth row and column in bold refer to node 6.

Table A3 shows the vectors of degree and neighborhood degree centrality for all the nodes of the network reported in Figure 1.

Table A3
Vectors of Degree and Neighborhood Degree Centrality for All the Nodes of Apulian Network

#node	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Degree	4	3	3	3	3	4	3	3	3	3	3	2	3	3	3	3	3	3	3	2	3	2	2	1
Neighborhood degree	10	9	8	9	11	12	9	8	8	7	6	6	7	8	8	10	10	9	9	7	8	7	7	5

Acknowledgments

Authors wish to thank the Water Company Acquedotto Pugliese for the use of Apulian-T. The work is theoretical proposal of the Edge-Betweenness tailoring for WDNs. Therefore, readers can reproduce the findings computing the Edge-betweenness, by means of any code, tailoring the network topology according to the procedure reported in the manuscript. The BBLAWN is a well-known benchmark WDN used first during the international conference Water Distribution System Analysis (WDSA 2014) Giustolisi et al. (2016).

References

Agathokleous, A., Christodoulou, C., & Christodoulou, S. E. (2017). Robustness and vulnerability assessment of water networks by use of centrality metrics. *European Water*, *58*, 489–495.

Anthonisse, J. (1971). *The rush in a graph*. Amsterdam: Mathematisch Centrum.

Barthélemy, M. (2011). Spatial networks. *Physics Reports*, *499*(1–3), 1–101. https://doi.org/10.1016/j.physrep.2010.11.002

Barthelemy, M. (2014). Discussion: Social and spatial networks. *Les Nouvelles de l'Archéologie*, (135), 51–61. https://doi.org/10.4000/nda.2374

Barthélemy, M. (2018). *Morphogenesis of spatial networks*. Cham: Springer International Publishing AG, Springer. https://doi.org/10.1007/978-3-319-20565-6

Barthélemy, M., & Flammini, A. (2008). Modeling urban street patterns. *Physical Review Letters*, *100*(13), 138702. https://doi.org/10.1103/PhysRevLett.100.138702

Benzi, M., & Klymco, C. (2015). On the limiting behavior of parameter-dependent network centrality measures. *SIAM Journal on Matrix Analysis and Applications*, *36*(2), 686–706. https://doi.org/10.1137/130950550

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, *2*(1), 113–120. https://doi.org/10.1080/0022250X.1972.9989806

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, *92*(5), 1170–1182. https://doi.org/10.1086/228631

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, *27*(1), 55–71. https://doi.org/10.1016/j.socnet.2004.11.008

Cadini, F., Zio, E., & Petrescu, C. A. (2008). *Using centrality measures to rank the importance of the components of a complex network infrastructure. In: Critical Information Infrastructure Security*. Critis: Springer.

Christodoulou, S. E. (2017). Robustness and vulnerability assessment of water networks by use of centrality metrics. *European Water*, *58*, 489e495.

Diao, K., Farmani, R., Fu, G., Astaraie-Imani, M., Ward, S., & Butler, D. (2014). Clustering analysis of water distribution systems: Identifying critical components and community impacts. *Water Science & Technology*, *70*(11), 1764–1773. https://doi.org/10.2166/wst.2014.268

Diao, K., Zhou, Y., & Rauch, R. (2013). Automated creation of district metered area boundaries in water distribution systems. *Journal of Water Resources Planning and Management*, *139*(2), 184–190. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000247

Diestel, R. (2005). *Graph Theory* (3rd ed.). Heidelberg: Springer.

Dijkstra, E. (1959). A note on two problems in Connexion with graphs. *Numerische Mathematlk*, *1*(1), 269–271. https://doi.org/10.1007/BF01386390

Ding, C., Zha, H., He, X., Husbands, P., & Simon, H. (2004). Link analysis: Hubs and authorities on the world wide web. *SIAM Review*, *46*(2), 256–268. https://doi.org/10.1137/S0036144501389218

Freeman, L. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, *1*(3), 215–239.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, *40*(1), 35–41. https://doi.org/10.2307/3033543

Furlan Ronqui, J., & Travieso, G. (2015). Analyzing complex networks through correlations in centrality measurements. *Journal of Statistical Mechanics: Theory and Experiment*, (5).

Girvan, M., & Newman, M. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, *99*(12), 7821–7826. https://doi.org/10.1073/pnas.122653799

Giustolisi, O. (2010). Considering actual pipe connections in WDN analysis. *Journal of Hydraulic Engineering*, *136*(11), 889–900. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000266

Giustolisi, O., Berardi, L., Laucelli, D., Savic, D., & Kapelan, Z. (2016). Operational and tactical Management of Water and Energy Resources in pressurized systems: Competition at WDSA 2014. *Journal of Water Resources Planning and Management*, *142*(5), 1–12.

Giustolisi, O., & Ridolfi, L. (2014a). A new modularity-based approach to segmentation of water distribution network. *Journal of Hydraulic Engineering*, *140*(10), 04014049. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000916

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

Giustolisi, O., Savìc, D., & Kapelan, Z. (2008). Pressure-driven demand and leakage simulation for water distribution networks. *Journal of Hydraulic Engineering*, *134*(5), 626–635. https://doi.org/10.1061/(ASCE)0733-9429(2008)134:5(626)

Giustolisi, O., Simone, A., & Ridolfi, L. (2017). Network structure classification and features of water distribution systems. *Water Resources Research*, *53*, 3407–3423. https://doi.org/10.1002/2016WR020071

Giustolisi, O., & Walski, T. (2012). Demand components in water distribution network analysis. *Journal of Water Resources Planning and Management*, *138*(4), 356–367. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000187

Gutiérrez-Pérez, J. A., Herrera, M., Pérez-García, R., & Ramos, M. E. (2013). Application of graph-spectral methods in the vulnerability assessment of water supply networks. *Mathematical and Computer Modelling*, *57*(7-8), 1853–1859. https://doi.org/10.1016/j.mcm.2011.12.008

Hawick, K. A. (2012). Water distribution network robustness and fragmentation using graph metrics. *Proceedings- International Conference on Water Resource Management- AfricaWRM*, *2012*, 304–310.

Katz, L. (1953). A new status index derived from socio metric analysis. *Psychometrika*, *18*(1), 39–43. https://doi.org/10.1007/BF02289026

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, *46*(5), 604–632. https://doi.org/10.1145/324133.324140

Laucelli, D., Simone, A., Berardi, L., & Giustolisi, O. (2017). Optimal design of district metering areas for the reduction of leakages. *Journal of Water Resources Planning and Management*, *143*(6), 403–410.

Lee, C. (2006). Correlations among centrality measures in complex networks. arXiv, p. physics/0605220v1.

Mihail, M., & Papadimitriou, C. (2002). On the eigenvalue power law. In *Lecture Notes in Computer Science* (pp. 254–262). Berlin, Heidelber: Springer.

Nazempour, R., Monfared, M. A. S., & Zio, E. (2016). A complex network theory approach for optimizing contamination warning sensor location in water distribution networks. arXiv 1601.02155.

Newman, M. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E*, *69*(6), 066133. https://doi.org/10.1103/PhysRevE.69.066133

Newman, M. (2010a). *Networks: An introduction*. Oxford, UK: Oxford University Press.

Nieminen, J. (1974). On centrality in a graph. *Scandinavian Journal of Psychology*, *15*(1), 322–336.

Ostfeld, A. (2012). Optimal reliable design and operation of water distribution systems through decomposition. *Water Resources Research*, *48*, W10521. https://doi.org/10.1029/2011WR011651

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The PageRank Citation Ranking: Bringing Order to the Web. Brisbane, Australia, s.n.

Pandit, A., & Crittenden, J. C. (2016). Index of network resilience for urban water distribution systems. *International Journal of Critical Infrastructures*, *12*(1/2), 120–142. https://doi.org/10.1504/IJCIS.2016.075865

Perelman, L., & Ostfeld, A. (2011). Topological clustering for water distribution systems analysis. *Environ. Modell. & Software*, *26*(7), 969–972. https://doi.org/10.1016/j.envsoft.2011.01.006

Rochat, Y. (2009). *Closeness centrality extended to unconnected graphs: The harmonic centrality index.* Zürich: CDH.

Scibetta, M., Boano, F., Revelli, R., & Ridolfi, L. (2013). Community detection as a tool for complex pipe network clustering. *EPL*, *130*(4), 48001.

Shuang, Q., Zhang, M., & Yuan, Y. (2014). Node vulnerability of water distribution networks under cascading failures. *Reliability Engineering and System Safety*, *124*, 132–141. https://doi.org/10.1016/j.ress.2013.12.002

Simone, A., Giustolisi, O., & Laucelli, D. B. (2016). A proposal of optimal sampling design using a modularity strategy. *Water Resources Research*, *52*, 6171–6185. https://doi.org/10.1002/2016WR018944

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*(1), 72–101. https://doi.org/10.2307/1412159

Trifunovic, N. (2012). *Pattern recognition for reliability assessment of water distribution networks*. London: CRC Press. https://doi.org/10.1201/b11795

Yannopoulos, S., & Spiliotis, M. (2013). Water distribution system reliability based on minimum cut – Set approach and the hydraulic availability. *Water Resources Management*, *27*, 1821–1836. https://doi.org/10.1007/s11269-012-0163-5

Yazdani, A., & Jeffrey, P. (2010). A complex network approach to robustness and vulnerability of spatially organized water distribution networks. arXiv1008.1770, p. 18.

Yazdani, A., & Jeffrey, P. (2011). Complex network analysis of water distribution systems. *Chaos*, *21*(1), 016111. https://doi.org/10.1063/1.3540339

Yazdani, A., & Jeffrey, P. (2012a). Robustness and vulnerability analysis of water distribution networks using graph theoretic and complex network principles. *Proceedings of the National Academy of Sciences of the United States of America, 2000*, 933–945.

Yazdani, A., & Jeffrey, P. (2012b). Applying network theory to quantify the redundancy and structural robustness of water distribution systems. *Journal of Water Resources Planning and Management*, *138*(2), 153–161. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000159

Yazdani, A., Otoo, R., & Jeffrey, P. (2011). Resilience enhancing expansion strategies for water distribution systems: A network theory approach. *Environmental Modelling and Software*, *26*(12), 1574–1582. https://doi.org/10.1016/j.envsoft.2011.07.016

Zhao, X., Chen, Z., & Gong, H. (2015). Effects comparison of different resilience enhancing strategies for municipal water distribution network: A multidimensional approach. *Mathematical Problems in Engineering*, *2015*, 1–16.