Edge betweenness for water distribution networks domain analysis

Antonietta Simone, Francesco G. Ciliberti, Daniele B. Laucelli IWA, Luigi Berardi IWA and Orazio Giustolisi IWA

ABSTRACT

Complex network theory (CNT) studies the relevance of elements in networks using centrality metrics. From the CNT standpoint, water distribution networks (WDNs) are infrastructure networks composed by vertices, named nodes, connected to each other by edges, named pipes, that transfer water to customers following a transfer process based on shortest paths. The present paper proposes the domain analysis of several real WDNs using the edge betweenness in order to capture the hydraulic behaviour based on network structure, i.e., for understanding the role of topological features in the emergent hydraulic behaviour. The strategy is obtained by tailoring CNT studies and tools in order to (i) embed the different hydraulic roles of sources and demand nodes, (ii) move the classic concept of centrality from the nodes to the pipes, i.e., the technically relevant components for WDNs and (iii) include information related to the directional devices, because they constrain flow directions. Results show the usefulness of the novel WDN-tailored edge betweenness for the WDN domain analysis. Therefore, the metric can represent a useful tool for supporting WDNs analysis, design and management tasks.

Key words | centrality metrics, complex network theory, domain analysis, edge betweenness, water distribution networks

Antonietta Simone (corresponding author)
Francesco G. Ciliberti
Daniele B. Laucelli IWA
Orazio Giustolisi IWA
Politecnico di Bari,
Via E. Orabona 4, 70125 Bari,
Italy
E-mail: antonietta.simone@poliba.it

Luigi Berardi IWA
University 'G. d'Annunzio' of Chieti Pescara,
Viale Pindaro 42, 65127 Pescara,
Italy

NOTATION

CNT	Complex network theory
WDN	water distribution network
i	generic vertex (node) in the network
V	set of vertices in the network
l	generic edge (link) in the network
E	set of edges in the network
σ_{st}	number of shortest paths from node s to node t
$\sigma_{st}(l)$	number of shortest paths from node s to node t passing along the edge l
C_l^B	edge betweenness of link l
hub	component of a network with a high-degree node
N_f	number of connected fictitious nodes in the network
M	number of reservoirs in the network
N_d	number of demand nodes in the network
V_T	maximum water volume supplied during the emptying process
V_D	average water volume supplied during the operating cycle to the demand nodes

INTRODUCTION

Many centrality metrics have been proposed over the years in order to clarify the concept of importance, in terms of centrality, detailing components and interrelationships among various elements of networks. Freeman (1979) was among the first to propose a definition of centrality by referring to the graph theory, even if the concept of centrality was

introduced by Bavelas (1950) for studying communication between small groups. Several centrality metrics have been proposed since by many researchers in order to evaluate the most central element in the network with respect to different physical phenomena: (1) Katz centrality (Katz 1953) computes the relative influence of a node within a network by measuring the number of the immediate and secondary neighbours, i.e., the centrality is based on the number of walks deriving from a node i. (2) Eigenvector (Bonacich 1972) measures the influence of a node in a network, based on the idea that the influence of node i is related to the influence of its neighbours, i.e., a node with high eigenvector is connected to many nodes also with high eigenvector. (3) Degree (Freeman 1977) is based on the number of connections of each node where the node with the highest degree is the most central. (4) Closeness (Freeman 1977) measures the centrality of a node considering how it is relatively close to all other nodes. (5) Betweenness (Freeman 1977) measures the importance of an element quantifying how many times the element is traversed by shortest paths between two vertices. (6) PageRank (Page *et al.* 1998) is an adjustment of Katz centrality that smooths the centrality of a node when it is only related to the centrality of a connected node. (7) Cross-clique centrality (Everett & Borgatti 1998; Faghani 2013) determines the connectivity of a node to different cliques, i.e., the number of cliques to which it belongs. A node with high cross-clique connectivity facilitates the propagation of information in a graph. (8) Hub and Authorities (Kleinberg 1999) quantify the centrality of nodes as a function of the centrality of the connection nodes. The metric refers to the creation of Web pages, especially hubs. These Web pages, even if not authoritative, represent large directories that lead users directly to the authoritative pages, i.e., a good hub represents a page that points to many other pages while good authority represents a page linked by many different hubs. (9) Decay centrality (Jackson 2008) measures the proximity between a chosen vertex and every other vertex weighted by the decay. (10) Percolation (Piraveenan *et al.* 2013) quantifies the relative impact of nodes based on their percolation states. For a given node, at a given time, it represents the proportion of 'percolated paths' that go through that node, i.e., shortest path between a pair of nodes, where the source node is percolated (e.g.,

infected). (11) Neighbourhood degree (Giustolisi *et al.* 2017) corresponds to tailoring the degree centrality concept for infrastructure networks by involving adjacent nodes belonging to neighbourhoods at different topological distance.

These centrality metrics aim at ranking the nodal importance in a network, i.e., how much a node is relevant with respect to other nodes in the network. However, the ranking only orders nodes by importance, without quantifying the difference in importance between different levels of the ranking. To overcome this problem, Freeman (1979) proposed the concept of centralization, i.e., the centralization of any network is a measure of how central its most central node is in relation to how central all the other nodes are.

It is important to note that each metric identifies different elements as the most relevant (central) with respect to different applications/physical phenomena (Benzi & Klymko 2015). To this purpose, Borgatti (2005) associated specific metrics to different types of network according to the flow process that characterizes the network. He stated that the use of a metric must be appropriate to the kind of flow process that characterizes the network, i.e., not all metrics can be unconditionally applied to all types of networks. Then, he defined three typologies of flow processes and, for each, suggested the more suitable centrality metrics. Accordingly, the flow process for spatial networks, such as water distribution networks (WDNs), was defined as 'package delivery' like, i.e., a process based on the concept of shortest paths with a fixed target. Finally, Borgatti (2005) stated that for this type of process the more suitable centrality metrics to apply are betweenness and closeness centrality. Later, Demšar *et al.* (2007) proposed an approach towards computing three centrality metrics, i.e., degree, closeness and betweenness, for identifying critical locations in spatial networks by dual graph modelling (Evans & Lambiotte 2009). They claimed that the betweenness is the most suitable for identification of critical locations, especially for structural analysis useful for modelling vulnerability and survivability of spatial networks. From this general picture it emerges that specific Complex network theory (CNT) centrality metrics can be potentially useful for WDNs analysis, planning and management, e.g., in terms of reliability and vulnerability.

The vulnerability analysis of WDNs using centrality metrics was first addressed by Yazdani & Jeffrey (2010, 2011, 2012a, 2012b). They proposed a variety of strategies for understanding the structure, efficiency, robustness and vulnerability of WDNs using centrality metrics and concluded that CNT metrics are useful and necessary but perhaps insufficient for analysing the structural reliability or vulnerability of spatial networks because of spatial constraints. Furthermore, they proposed a novel metric based on the connectivity features for the analysis of weighted and directed WDNs. From here on, CNT centrality metrics were adopted to assess vulnerability analysis (Hawick 2012; Diao *et al.* 2014; Shuang *et al.* 2014), reliability assessment (Ostfeld 2012) and water distribution network (WDN) resilience (Pandit & Crittenden 2016). Later on, Giustolisi *et al.* (2017) proposed the neighbourhood degree, i.e., an extension of the classic nodal degree, in order to demonstrate that the WDN connectivity structure follows a Poisson-like distribution because of spatial constraints, meaning that the domain of these spatial systems, on average (i.e., not considering reservoirs, tanks and directional devices), has a good structural resistance to random failures and intentional threats (Giustolisi *et al.* 2017).

It has to be noted that all proposed works on centrality metrics focus on node relevance, while, in the case of WDNs, pipes are the most important components (Simone *et al.* 2018). This means the focus should move from nodes to pipes, because pipes, with their asset features, are the most important network components for hydraulic systems. To this end, Giustolisi *et al.* (2019) proposed the use of the edge betweenness (Girvan & Newman 2002) for analysing the domain of WDNs. In fact, they proposed to tailor the network topology to embed the different hydraulic roles of nodes (source nodes, tanks and demand nodes) including information related to the directional devices and different weights for pipes, in order to consider pipe properties (e.g., hydraulic resistance). This way, the analysis of the domain can be useful to quantify, for example, the feasibility of management interventions in advance with respect to hydraulic modelling.

The aim of the present work is to show the effectiveness of domain analysis for several real WDNs using the WDN-tailored edge betweenness proposed by Giustolisi *et al.* (2019), in order to demonstrate how it can represent a useful tool to support the hydraulic analysis of a WDN. To this end, this study takes WDNs with different size, asset features and hydraulic schemes in real decision support contexts.

The paper is organized as follows. The next section presents the main concepts of the strategy. This is followed by a section that reports several case studies, a brief discussion of the strategy and an interpretation of the main results. Conclusions are given in the final section.

EDGE BETWEENNESS AND TOPOLOGY TAILORING

This study aims at assuming the edge betweenness proposed by Girvan & Newman (2002) as centrality metric suitable for WDN domain analysis in order to consider the relevance of pipes and not of nodes (Giustolisi *et al.* 2019). In the case of WDNs, nodes are generally elements to transfer water (information) and water is delivered to customers at pipe level. In fact, in WDN models, nodes where water demand is assumed to be delivered are not significant for the connectivity structure of the WDNs because the elements that can fail are pipes while connection nodes generally do not fail.

The edge betweenness is similar to the betweenness centrality, but it refers to the generic edge l. Assume that $\sigma_{st}(l)$ is the number of shortest paths from node s to node t passing along the edge l and σ_{st} is the number of all shortest paths from node s to node t.

The edge betweenness C_l^B of edge l is the sum of the fractions $\sigma_{st}(l)/\sigma_{st}$ for all the couples of nodes s, t in the network:

$$C_l^B = \sum_{\substack{s \neq t \in V \\ l \in E}} \frac{\sigma_{st}(l)}{\sigma_{st}} \tag{1}$$

where V and E are the set of vertices and edges belonging to the network. The edge with the highest edge betweenness centrality is the most relevant.

Therefore, even if the edge and the nodal metrics are performed with the same procedure and algorithm (Dijkstra 1959), the edge betweenness is more meaningful for WDNs, accounting for flow path disruption.

The water source nodes (i.e., reservoirs and tanks) represent the only exception, because they play a different

hydraulic role with respect to other nodes and, therefore, they need a specific study in order to define a hydraulics-based topology. In fact, the strategy proposed by Giustolisi *et al.* (2019) considers the source node as a kind of hub with respect to the hydraulic system behaviour. In absence of source nodes, the hydraulics does not exist, and the connectivity structure just embeds the topology. For the water source analysis, let us assume N_f connected fictitious nodes, M reservoirs and N_d demand nodes. This assumption arises from the fact that the reservoirs are generally linked to the networks with only one (suburban) pipe. This topological characteristic gives it a low degree of connection even if, from a hydraulic point of view, it represents a hub from which all water paths towards demand nodes start. This way, the relevance of the position of original source nodes is preserved with respect to the entire network structure although amplified by N_f. Accordingly, the relevance of the pipes close to the reservoir increases, while the other pipes' relevance does not change considerably. For each reservoir there exists a star of N_f connected fictitious nodes whose number is equal to N_d/M. The repartition of the demand nodes is related to the real reservoir influence area, i.e., a different value of N_f for each reservoir exists depending on its area of influence.

The next step is inserting the domain information about tanks, i.e., nodal sources characterized by an emptying/filling process. The information is related to the maximum water volume supplied during the emptying process (V_T) by the tank. Let us assume N_f connected fictitious nodes already attributed to reservoirs and N_d demand nodes. The number of fictious nodes (usually definable as 'star of fictitious nodes') attributed to each tank equals the minimum integer value between $N_d \cdot V_T/V_D$ and N_f, where V_D is the average water volume supplied during the operating cycle to the demand nodes. It is important to note that increasing the volume V_T with respect to V_D, the tank tends to the reservoir and decreasing the volume V_T the tank node tends to a demand or connectivity node. However, tanks are always less relevant hubs than reservoirs.

Directional devices (e.g., pumps, pressure reduction valves, check valves, etc.) and flow direction for pipes next to reservoirs are also considered to enhance the domain analysis. Indeed, the presence of directional devices allows water to flow only in one direction. In this way, for example,

the shortest paths between two pipes, one upstream and the other downstream of a specific device, can only be travelled in one direction, i.e., that of the device. The consequence is that the edge betweenness of these pipes is automatically reduced compared to that computed on the network without devices.

The last step is to further refine the WDN-tailored edge betweenness metric attributing weights to the pipes as, for example, the pipes' hydraulic resistance. These weights allow moving the network domain analysis closer to the expected hydraulic behaviour of the system; pipe hydraulic resistances, for example, drive the water fluxes.

CASE STUDIES

The WDN-tailored edge betweenness is here applied and discussed using BBLAWN and other six Apulian real WDNs, in order to show the effectiveness of the analysis for different typology of WDNs. The relevant characteristics of analysed networks are reported in Table 1. Table 1 also reports the Spearman correlation index (Spearman 1904) between two metrics considering several weights, i.e., connectivity, length and resistance, for the edge betweenness. This way the analysis includes the domain characteristics of the networks.

Figure 1 (left panel) reports the pipe with the highest 5% of flow rate values averaged on 1-day hydraulic simulation; (right panel) the pipes with highest 5% values of WDN-tailored edge betweenness for the BBLAWN network. The most relevant pipes correspond for the two metrics. Therefore, the WDN-tailored edge betweenness is able to identify almost all the most important pipes, as ranked by pipe flow rates, only on the basis of the domain structure of the system. The correlation index for BBLAWN in Table 1, i.e., weighting the WDN-tailored edge betweenness with pipe resistance, is equal to 0.74. This result confirms the usefulness of the edge betweenness for WDN domain analysis. It can be argued that information about reservoirs, tanks and directional devices enhances the domain analysis. Furthermore, Figure 1 shows that the analysis defines a main path into the network that starts from the only reservoir and reaches the tank on the left. From this main conduct, two other paths delineate the remaining part of the network

Table 1 │ Topological characteristics and correlation index for the seven analysed WDNs

WDN Name	Node #	Pipe #	Reservoirs #	Tanks #	Demand nodes #	Inhabitants [×1000] #	Corr. index resistance	Corr. index connectivity	Corr. index length
BBLAWN	390	439	1	7	334	–	0.74	0.75	0.75
Apulia 1	986	1,105	1	0	663	12	0.64	0.71	0.73
Apulia 2	7,716	8,496	3	0	7,011	12	0.60	0.63	0.65
Apulia 3	1,270	1,472	3	0	992	199	0.54	0.50	0.53
Apulia 4	3,166	3,483	2	0	1,582	26	0.66	0.75	0.79
Apulia 5	7,164	7,895	1	0	2,918	70	0.62	0.70	0.75
Apulia 6	1,111	1,307	1	0	770	13	0.54	0.59	0.60

'skeleton', i.e., the main structure of the network. Both two metrics indicate the most relevant pipe as those close to the only reservoir, i.e., the most relevant hub.

Figure 2 reports the flow (left panel) and the WDN-tailored edge betweenness (right panel) for Apulia 1 network. The network is composed of a single reservoir and a pressure control valve, for controlling pressure and reducing water losses. The presence of a single reservoir uniquely defines the main path of the pipes, which starts from the water source, i.e., the only hub, and reaches the centre of the network (see the darker circles). Most of the pipes with slightly lower metrics values (see the lightest circles) act as links between the main path and the rest of the network. The correlation index between two metrics, considering the pipe resistance as weight for the edge

betweenness, is equal to 0.64, confirming that the information contained in the domain is relevant for the WDN hydraulics. Table 1 shows that for Apulia 1 network the correlation assumes higher values considering the length as weight for the pipes (i.e., 0.73). This occurs for six of the seven analysed networks, although graphically the pipe hydraulic resistance better corresponds to the flow distribution as expected from an engineering perspective. For this reason, pipe hydraulic resistance was chosen as weight to discuss the metric. However, it is important noting that, regardless of weight, the correlation index is always greater than 0.50, i.e., the edge betweenness well identifies the hydraulic behaviour of the networks.

Figure 3 reports the flow rates (left panel) and the WDN-tailored edge betweenness (right panel) for the Apulia 2
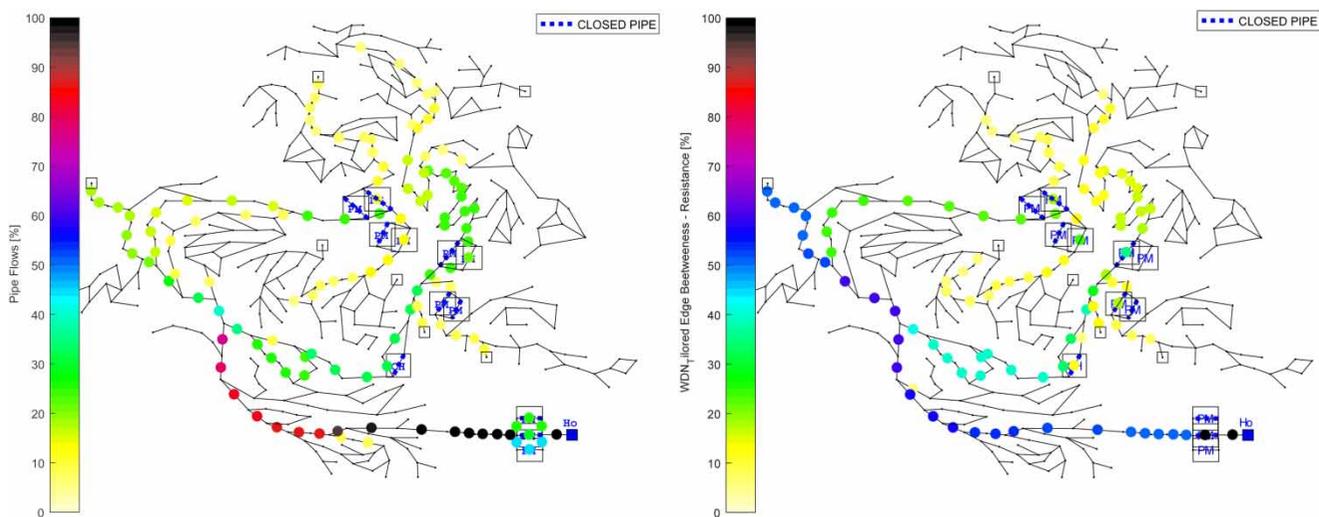

Figure 1 │ Pipe with flow rates in the highest 5% values (left) and pipes with the highest 5% values of the WDN-tailored edge betweenness (right) for BBLAWN network. Spearman correlation index is equal to 0.74.
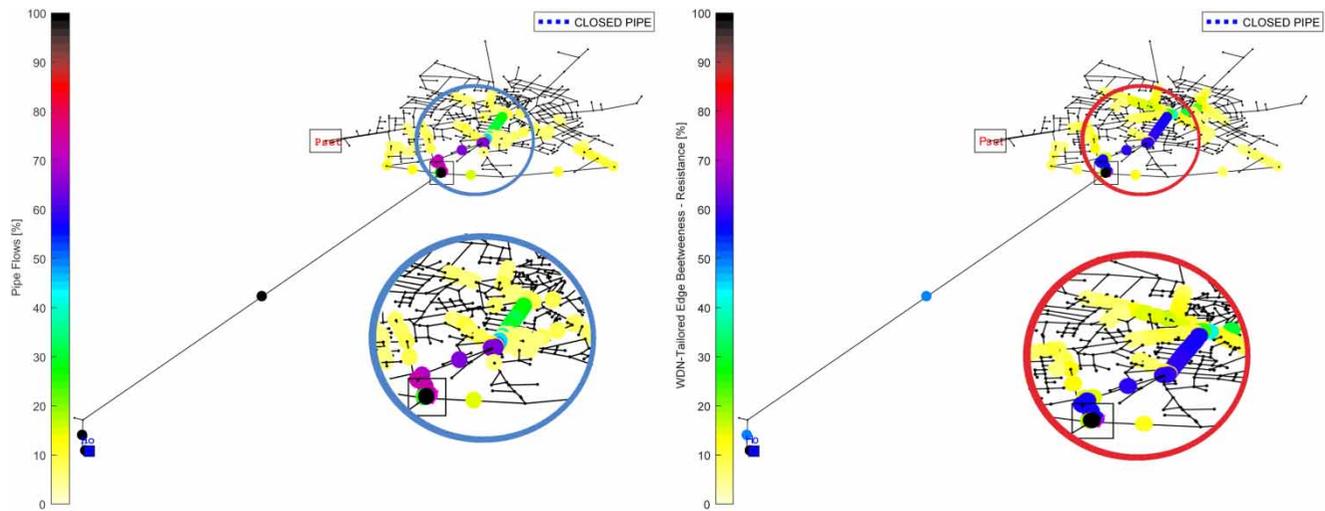
Figure 2 | Pipe with flow rates in the highest 5% values (left) and pipes with the highest 5% values of the WDN-tailored edge betweenness (right) for Apulia 1 network. Spearman correlation index is equal to 0.64.

network. Apulia 2 is a large looped network with three reservoirs, five controlled entries and a wide range of diameters. Comparing the two panels of Figure 3 and looking at the correlation indices, it can be stated that also for such a large network, the proposed tailored-metric identifies quite well all the most important pipes in terms of pipe flow rates. This is also confirmed by correlation indices equal to 0.60, 0.63 and 0.65 in Table 1. It is important to note that, in addition to the five suburban pipes, some of the most important paths are positioned into the looped part of the network, corresponding to the city centre. The

complexity of the network, also due to the presence of many reservoirs and pressure control valves, influences the correlation index values. This means that, for Apulia 2, the values obtained considering devices are lower than those without devices. The high correspondence between the two panels of Figure 3 confirms this consideration.

Figure 4 reports pipe flow rates (left panel) and the WDN-tailored edge betweenness (right panel) for the Apulia 3 network. The presence of three reservoirs leads to identifying a higher number of main paths in the network. Nevertheless, since each reservoir has its own area of influence in terms of
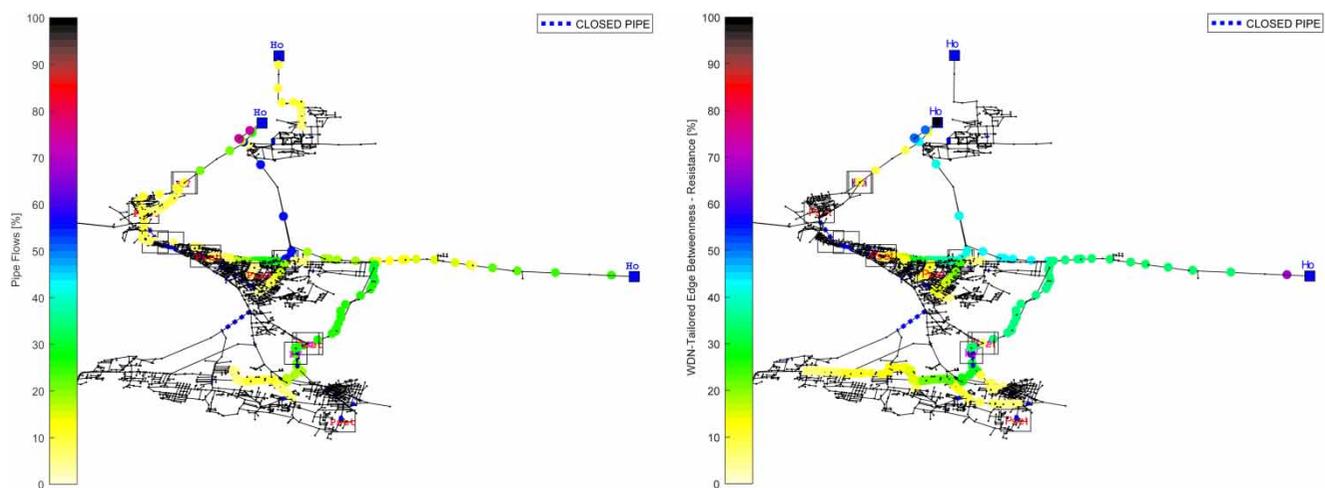

Figure 3 | Pipe with flow rates in the highest 5% values (left) and pipes with the highest 5% values of the WDN-tailored edge betweenness (right) for the Apulia 2 network. Spearman correlation index is equal to 0.60.
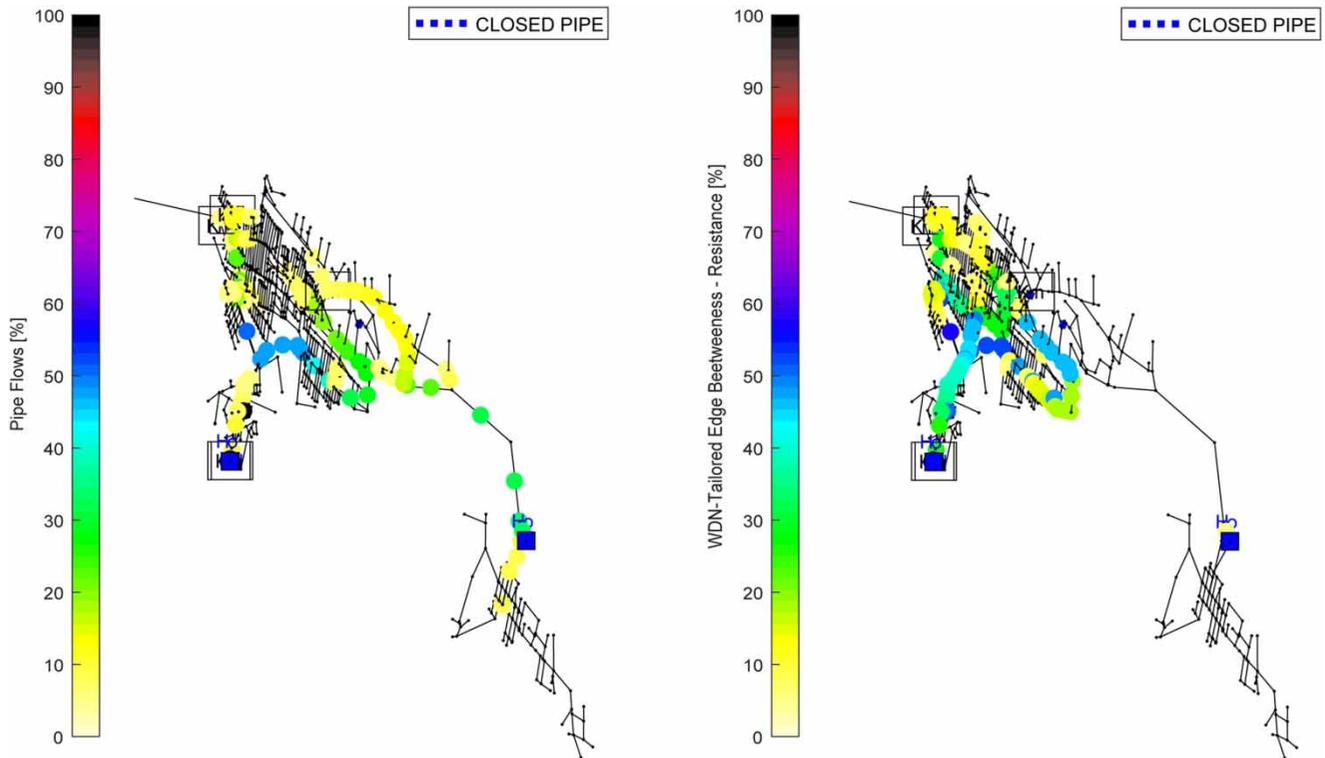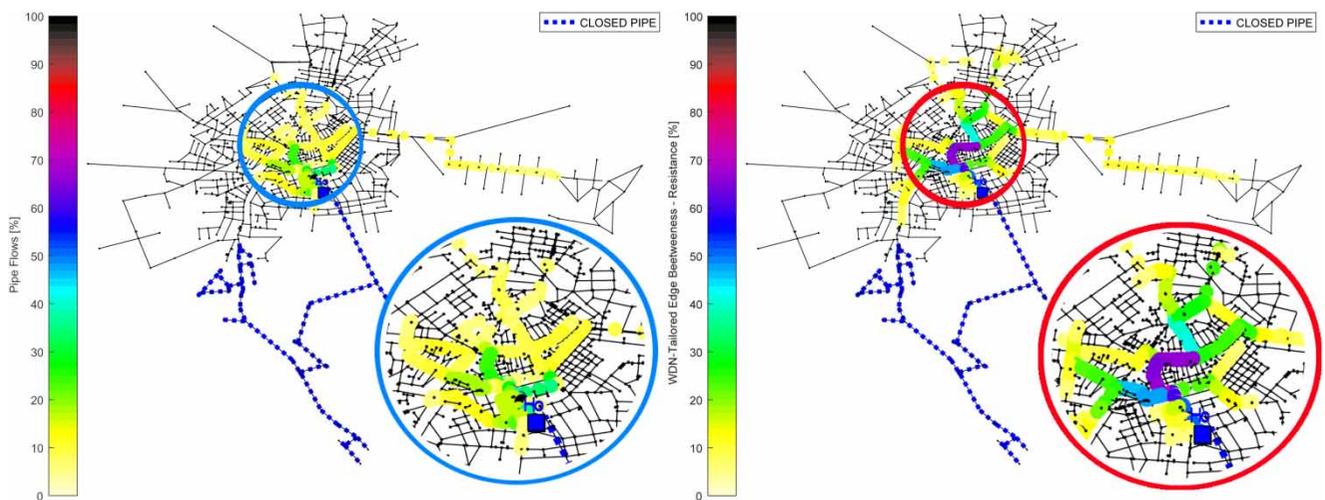
Figure 4 │ Pipe with flow rates in the highest 5% values (left) and pipes with the highest 5% values of the WDN-tailored edge betweenness (right) for the Apulia 3 network. Spearman correlation index is equal to 0.54.

demand nodes, each main path has a different relevance. The number of reservoirs influences the correlation index, obtaining a value equal to 0.54 considering the pipe resistance as weight, while for connectivity and length the value decreases

to 0.53 and 0.50, respectively, i.e., values lower than those evaluated for less complex networks.

Figure 5 reports pipe flow rates (left panel) and the WDN-tailored edge betweenness (right panel) for the

Figure 5 │ Pipe with flow rates in the highest 5% values (left) and pipes with the highest 5% values of the WDN-tailored edge betweenness (right) for the Apulia 4 network. Spearman correlation index is equal to 0.66.

Apulia 4 network. The dotted lines indicate a portion of the network no longer in use. The network is strongly looped, i.e., it has a very high redundancy, and this aspect favours the presence of several pipes/paths with similar relevance instead of a few pipes/paths with much higher relevance than the others. It is important to note that the reservoir always represents a hub regardless of the network structure. The most relevant pipes coincide for both the metrics, even if the edge betweenness identifies a greater number of paths as relevant, probably because the topologic metric weighted with pipe resistance considers the diameters regardless of the flow that actually crosses them, while the hydraulic metric defines them uniquely. The very regular nature of the network and the presence of a single reservoir provides a good correlation index between the two metrics, i.e., 0.66 considering the pipe resistance as weight, 0.75 and 0.79 assuming the connectivity and the length, respectively. Once again, the correlation index assumes a higher value considering the length as weight. However, this result makes the analysis very close to the hydraulics of the system.

Figure 6 reports pipe flow rates (left panel) and the WDN-tailored edge betweenness (right panel) for the Apulia 5 network. Apulia 5 is a large network composed of a single reservoir and a pressure control valve downstream of the water source in order to control and manage the pressure into the network. The size of the network means that the hydraulic metric (i.e., flow rates) identifies several main paths, not all directly connected to the reservoir. The edge betweenness identified the same main paths as the most important (see the darker circles). However, Figure 6 (right) shows that in the highest 5% values of the topologic metric there are other pipes/paths that are less relevant (see the lightest circles) from the hydraulic perspective, i.e., the flow has no direct correspondence with the resistance/diameters for all the pipes. The correlation index shows once again that the analysis is close to the hydraulic of the system, obtaining values equal to 0.62, 0.70 and 0.75 in Table 1 weighting the edge betweenness with pipe resistance, connectivity and length, respectively.

Figure 7 reports pipe flow rates (left panel) and the WDN-tailored edge betweenness (right panel) for the
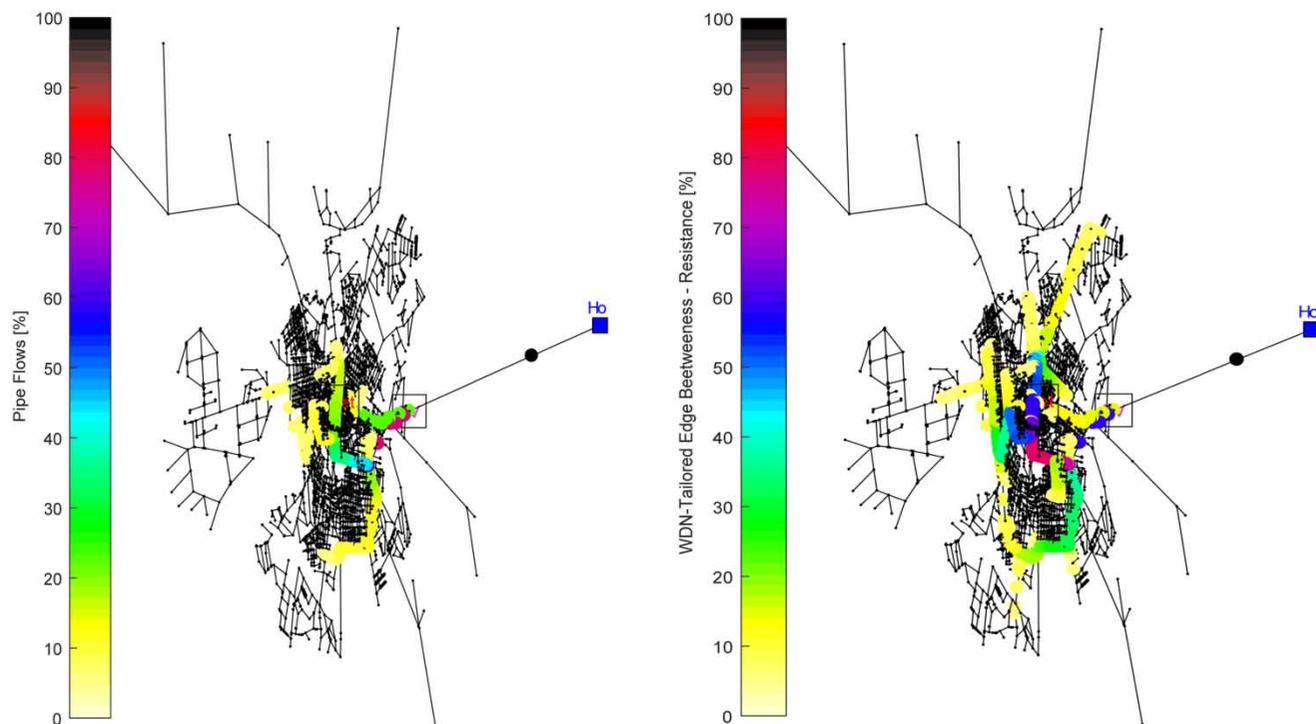

Figure 6 │ Pipe with flow rates in the highest 5% values (left) and pipes with the highest 5% values of the WDN-tailored edge betweenness (right) for the Apulia 5 network. Spearman correlation index is equal to 0.62.
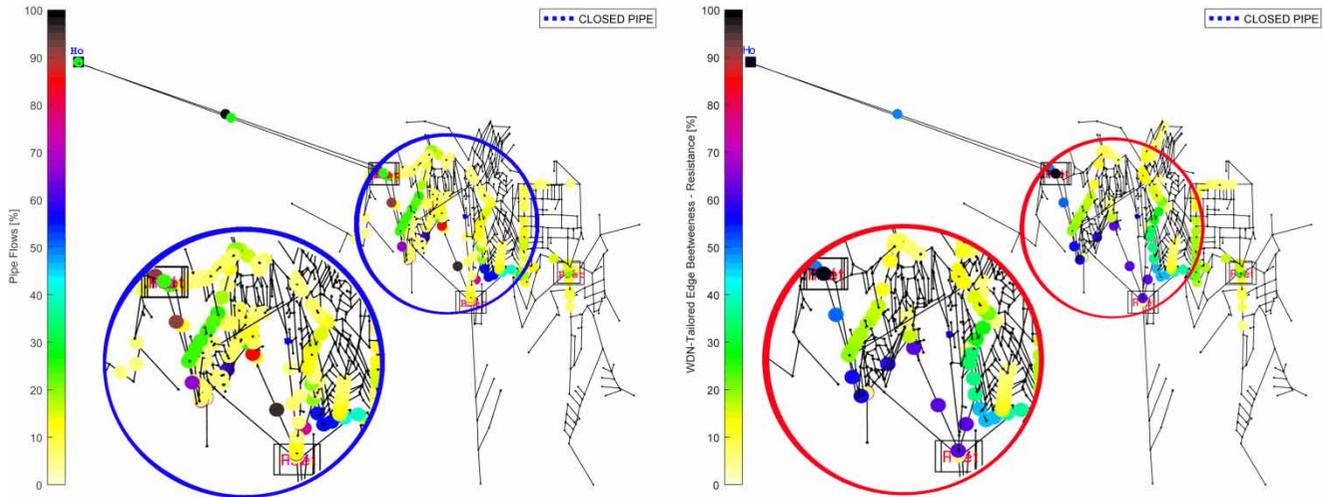
Figure 7 │ Pipe with flow rates in the highest 5% values (left) and pipes with the highest 5% values of the WDN-tailored edge betweenness (right) for the Apulia 6 network. Spearman correlation index is equal to 0.54.

Apulia 6 network. The left panel shows that the pipe flow rates do not define properly a main path, but rather highlight the most relevant pipes, i.e., where the value falls in the highest 5% flow rate value, with circles that range from dark to light. The right panel shows the same pipes as the most relevant for the topological metric, confirming that the pipes close to the reservoir are the most relevant. The correlation index between two metrics, considering the pipe resistance as weight, is equal to 0.54, and as for the previous case, Table 1 reports a higher correlation index if considering the length as weight (i.e., 0.60). The presence of four pressure control valves, i.e., four directional devices into the network, probably influenced the evaluation of the correlation indices, because of the directional devices.

The comparison between topological and hydraulic metrics is really positive, also considering that the pipe flows depend on the boundary conditions and on the momentum and continuity equations while the proposed metric focuses only on the connectivity structure. Such results confirm that the hydraulic behaviour is strongly dependent on the domain features and that the proposed metric can be useful to predict it and support further decisions on network operation and management.

In order to show that the hydraulic behaviour is strongly dependent on the domain features embedded in the edge betweenness, Figure 8 shows a diagram that compares the distribution of the pipe flow rates (sorted in descending order) with the corresponding distribution of the tailored edge betweenness for Apulia 6. The comparison confirms
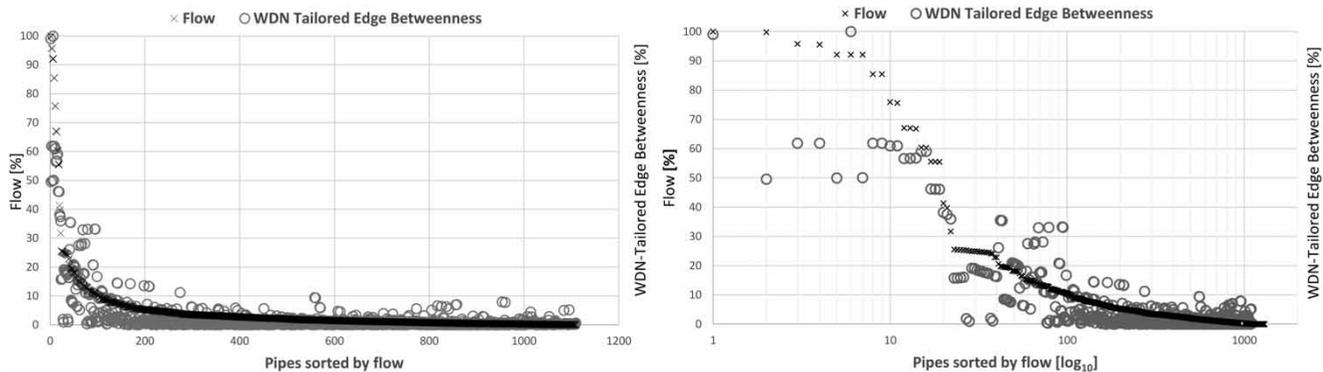

Figure 8 │ Comparison between the pipe flows and the tailored edge betweenness.

that the correspondence between the two metrics is very good. The same trend has been obtained for the other WDNs reported herein. Please note that the diagram on the right of Figure 8 has the x-axis in logarithmic scale.

Therefore, looking at the results obtained with the domain analysis, it is possible to state that the edge betweenness, tailored and weighted, can capture the emergent hydraulic behaviour as described by pipe flow rates, due to the connectivity structure of the network. In fact, the domain analysis for seven real WDNs provided correlation index values ranging in the interval [0.50; 0.79], confirming the usefulness of the topological metric for WDN analysis. Furthermore, a further advantage is the possibility to get effective information on the network without the need of many hydraulic simulations and with a low computational cost. For instance, with respect to the largest network herein (Apulia 2), the calculation takes about 60 seconds on a standard laptop equipped with a CPU Intel i7, 2.6 GHz with 16 GB of RAM. Although the time taken is larger than a single hydraulic simulation, it pertains to the main topological domain of the system and needs to be performed only once, unless major connectivity changes happen.

CONCLUSIONS

This contribution proves the effectiveness of the domain analysis of WDNs using the edge betweenness, tailored for accounting the WDNs constraints, as proposed by Giustolisi *et al.* (2019), on seven real networks. The strategy embeds the different hydraulic roles of nodes, moves the concept of centrality from the nodes to the pipes and includes information related to the directional devices. Furthermore, the tailored edge betweenness is weighted with domain information (e.g., pipe resistance). The Spearman correlation index between hydraulic and topological metrics has been evaluated for the seven WDNs considering as weights the connectivity, the length and the resistance. The results confirm the usefulness of the edge betweenness for the WDN domain analysis, i.e., the WDN-tailored edge betweenness can represent a useful tool for supporting WDN analysis, design and management tasks. Indeed, although the WDN-tailored edge betweenness cannot replace hydraulic

simulation, it might provide useful indications to drive model validation, calibration, maintenance works as well as relevance of pipes to plan operations.

REFERENCES

Bavelas, A. 1950 Communication patterns in task-oriented groups. *Journal of the Acoustical Society of America* **22**, 725–730. https://doi.org/10.1121/1.1906679.

Benzi, M. & Klymko, C. 2015 On the limiting behavior of parameter-dependent network centrality measures. *SIAM Journal on Matrix Analysis and Applications* **36** (2), 686–706.

Bonacich, P. 1972 Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology* **2** (1), 113–120.

Borgatti, S. P. 2005 Centrality and network flow. *Social Networks* **27** (1), 55–71.

Demsar, U., Spatenkova, O. & Virrantaus, K. 2007 Centrality measures and vulnerability of spatial networks. In: *4th International Conference on Information Systems for Crisis Response and Management, ISCRAM 2007*, Delft, the Netherlands.

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.

Dijkstra, E. 1959 A note on two problems in connexion with graphs. *Numerische Mathematik* **1** (1), 269–271.

Evans, T. S. & Lambiotte, R. 2009 Line graphs, link partitions, and overlapping communities. *Physical Review E* **80**, 016105. doi: 10.1103/PhysRevE.80.016105.

Everett, M. G. & Borgatti, S. P. 1998 Analyzing clique overlap. *Connections* **21** (1), 49–61.

Faghani, M. R. 2013 A study of XSS worm propagation and detection mechanisms in online social networks. *IEEE Transactions on Information Forensics and Security* **8** (11), 1815–1826. doi: 10.1109/TIFS.2013.2280884.

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

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

Girvan, M. & Newman, M. 2002 Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* **99** (12), 7821–7826.

Giustolisi, O., Simone, A. & Ridolfi, L. 2017 Network structure classification and features of water distribution systems. *Water Resources Research* **53** (4), 3407–3423.

Giustolisi, O., Ridolfi, L. & Simone, A. 2019 Tailoring centrality metrics for water distribution networks. *Water Resources Research* **55** (3), 2348–2369.

Hawick, K. A. 2012 Water distribution network robustness and fragmentation using graph metrics. In: *Proceedings of the International Conference on Water Resources Management AfricaWRM 2012*, Gaborone, Botswana, pp. 304–310. doi: 10.2316/P.2012.762-037.

Jackson, M. O. 2008 *Social and Economic Networks*. Princeton University Press, Princeton, NJ, USA.

Katz, L. 1953 A new status index derived from socio metric analysis. *Psychometrika* **18** (1), 39–43.

Kleinberg, J. M. 1999 Authoritative sources in a hyperlinked environment. *Journal of the ACM* **46** (5), 604–632.

Ostfeld, A. 2012 Optimal reliable design and operation of water distribution systems through decomposition. *Water Resource Research* **48** (10), W10521. doi:10.1029/2011WR011651.

Page, L., Brin, S., Motwani, R. & Winograd, T. 1998 The PageRank Citation Ranking: Bringing Order to the Web. In: *Proceedings of the 7th International World Wide Web Conference*, Brisbane, Australia, pp. 161–172.

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.

Piraveenan, M., Prokopenko, M. & Hossain, L. 2013 Percolation centrality: quantifying graph. Theoretic impact of nodes during percolation in networks. *PLoS ONE* **8** (1), e53095. https://doi.org/10.1371/journal. pone.0053095.

Shuang, Q., Zhang, M. & Yuan, Y. 2014 Node vulnerability of water distribution networks under cascading failures. *Reliability Engineering and System Safety* **124**, 132–141.

Simone, A., Ridolfi, L., Laucelli, D. B., Berardi, L. & Giustolisi, O. 2018 Centrality Metrics for Water Distribution Networks. In: *HIC 2018-13th International Conference on Hydroinformatics*, Palermo, Sicily.

Spearman, C. 1904 The proof and measurement of association between two things. *American Journal of Psychology* **15**, 72–101.

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

Yazdani, A. & Jeffrey, P. 2011 Complex network analysis of water distribution systems. *Chaos* **21** (1), 016111.

Yazdani, A. & Jeffrey, P. 2012a Robustness and vulnerability analysis of water distribution networks using graph theoretic and complex network principles. In: *12th Annual Conference on Water Distribution Systems Analysis (WDSA)*, Tucson, Arizona, USA, pp. 933–945. https://doi.org/10.1061/41203(425)85

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). https://doi.org/10.1061/(ASCE) WR.1943-5452.0000159