AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2016WR020071

Key Points:

- Classification of the connectivity network structure of water distribution systems
- Novel neighbor nodal degree concept for the classification of spatial networks
- Vulnerability and robustness of water distribution systems

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

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

Received 4 NOV 2016
Accepted 21 MAR 2017
Accepted article online 28 MAR 2017
Published online 26 APR 2017

Network structure classification and features of water distribution systems

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

[1]Politecnico di Bari, Bari, Italy, [2]Politecnico di Torino, Torino, Italy

Abstract The network connectivity structure of water distribution systems (WDSs) represents the domain where hydraulic processes occur, driving the emerging behavior of such systems, for example with respect to robustness and vulnerability. In complex network theory (CNT), a common way of classifying the network structure and connectivity is the association of the nodal degree distribution to specific probability distribution models, and during the last decades, researchers classified many real networks using the Poisson or Pareto distributions. In spite of the fact that degree-based network classification could play a crucial role to assess WDS vulnerability, this task is not easy because the network structure of WDSs is strongly constrained by spatial characteristics of the environment where they are constructed. The consequence of these spatial constraints is that the nodal degree spans very small ranges in WDSs hindering a reliable classification by the standard approach based on the nodal degree distribution. This work investigates the classification of the network structure of 22 real WDSs, built in different environments, demonstrating that the Poisson distribution generally models the degree distributions very well. In order to overcome the problem of the reliable classification based on the standard nodal degree, we define the “neighborhood” degree, equal to the sum of the nodal degrees of the nearest topological neighbors (i.e., the adjacent nodes). This definition of “neighborhood” degree is consistent with the fact that the degree of a single node is not significant for analysis of WDSs.

Plain Language Summary Water distribution systems are networked infrastructure, which can be studied using the complex network theory. A classic study is the network classification allowing to characterize the real system feature with respect e.g. to vulnerability. The work classifies a relevant number of water distribution networks concluding that they are random-like.

1. Introduction

The increasing size of water distribution systems (WDSs) associated with current growing urbanization keeps the analysis and management of these systems a complex task. Moreover, the installation of nonlinear devices for technical tasks (e.g., the introduction of pressure control valves for the purpose of leakage management) makes the analysis of the WDN hydraulic behavior more difficult with respect to the past. Finally, planning and management actions increasingly involve the concepts of vulnerability, reliability, and safety and interact with socioeconomic and environmental aspects. It follows that the study of the network structure and connectivity of a WDS, which primarily determines the hydraulic behavior of the system, is becoming an emerging issue. From this perspective, recent complex network theory (CNT) can be very useful for a modern WDS analysis, planning, and management.

CNT is becoming one of the most powerful and versatile tool to investigate, describe, and understand the world [*Barabási*, 2012]. Networks allow the study and the interpretation of a huge number of physical, biological, and social processes. Examples range from social relationships [*Scott*, 1988] to neural connections [*Papo et al.*, 2014], from multispecies ecological interactions [*Poisot and Gravel*, 2014] to financial and economic exchanges [*Schweitzer et al.*, 2009]. Although each network exhibits its own topological and structural peculiarities, very different networks can share amazing similar features [*Albert and Barabási*, 2002; *Buchanan*, 2003].

In the last decade, CNT had unrestrained development, and researchers proposed novel approaches, metrics, and theories to explore and disentangle network features [e.g., *Boccaletti et al.*, 2006; *Newman*, 2010]. In

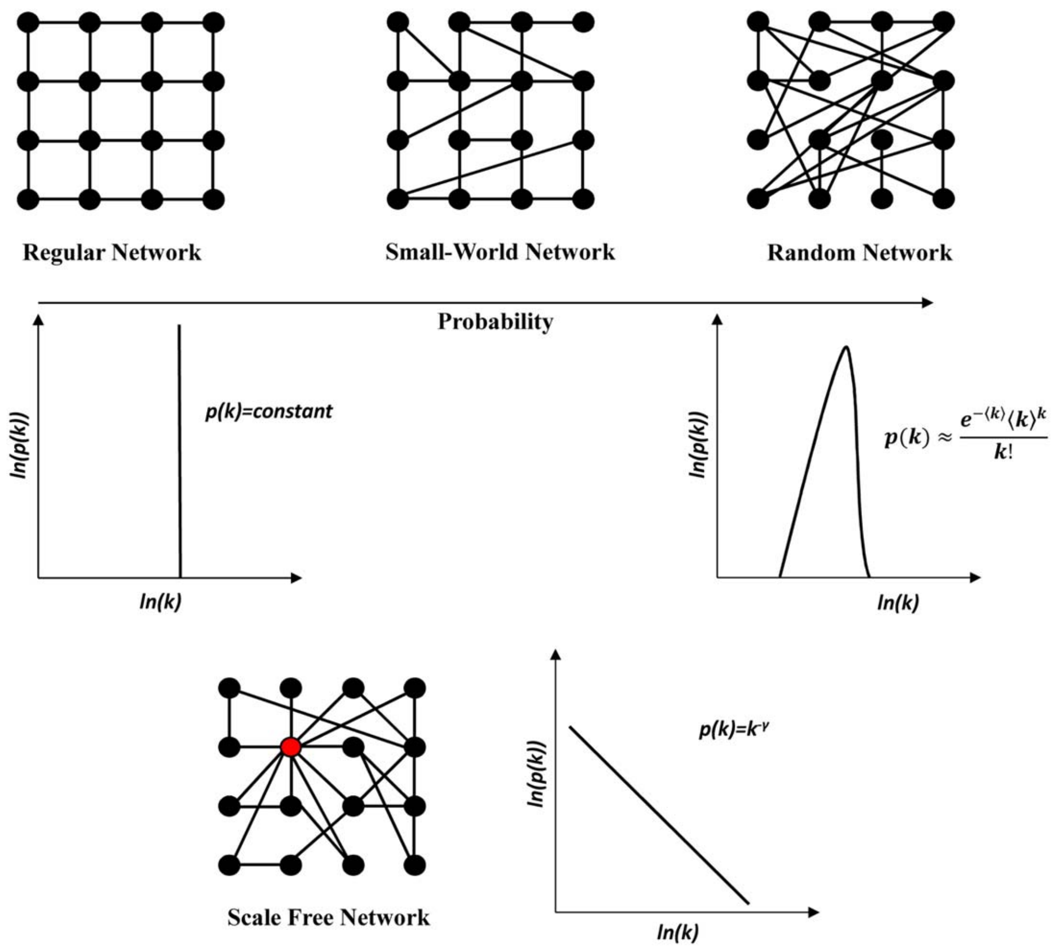


Figure 1. Degree distribution for *regular*, *small world*, and *random networks* (top) and degree distribution for *scale-free networks* (bottom).

this set of mathematical tools, the number of edges/pipes connected to each node, i.e., the nodal degree, is a relevant information for classifying networks. In fact, the nodal degree distribution describes the probability distribution of the number of edges connected with each node of the network. Several network features are associated with the shape of such distributions. *Erdös and Rényi* [1959, 1960] were the first to study the nodal degree distribution of real networks introducing the *random* networks versus *regular* networks. In the former case, the nodal degree distribution is randomly distributed around an average value and the network is characterized by a high homogeneity; whereas regular networks have a constant degree of internal nodes and are characterized by an absolute homogeneity. The Poisson model is often used to describe nodal degree distribution of random networks. The random networks were introduced because they are able to capture some features of real networks better than regular networks (e.g., being highly ordered, the shortest paths between two nodes are too large), while random networks show lower and more realistic values [e.g., *Milgram*, 1967].

Later on, *Watts and Strogatz* [1998] introduced *small world* networks and demonstrated the existence of small world effect, i.e., a behavior between the regular and random networks. Using the Poisson distribution of degree and a probability, *p*, of connections between two nodes, the model by *Watts and Strogatz* [1998] covers *regular networks* to *random networks*. For *small world networks*, the probability *p* is greater than the null value of *regular networks* but has lower values than *random networks* (see Figure 1, top).

Barabási and Albert [1999] proposed the *scale-free* networks in order to describe real networks characterized by nonhomogeneous nodal degree distributions, where many nodes have a low degree and few nodes (called hubs) have a high degree. These distributions typically exhibits a Pareto (or power law) behavior and cannot be classified as *random*, *small world*, or *regular networks* (see Figure 1, bottom).

The interest in classifying networks according to connectivity structure is related to the aim to capture emerging behaviors of real systems [*Albert and Barabási*, 2002; *Newman*, 2003, 2010]. For instance, the nodal degree distribution affects the networks vulnerability to random failures and intentional threats. In fact, *regular*, *small world*, and *random networks* present a significant structural resistance to both random failures and intentional threats, while scale-free networks show a very high structural resistance to random failures but a weak resistance to intentional threats [*Albert et al*., 2000]. In this sense, the classification of networks by associating their degree distribution to the Poisson or Pareto models is useful to assess the vulnerability and robustness of real infrastructural systems.

In spite of the fact that network classification can play a crucial role in assessing WDS behavior, this task is not easy because the network structure of WDSs is strongly constrained by the spatial characteristics of the environment where they are constructed, meaning that WDSs have spatial networks similar to other urban distribution infrastructures (e.g., gas and electricity) and transportation (e.g., railways and roads) systems. The spatial networks are constrained by (i) two-dimensional space and (ii) spatial impediments [*Barthélemy*, 2011]. This implies that the maximum nodal degree is generally very low (it is typically lower than seven); consequently, the nodal degree distribution spans a very limited range and the resulting statistical inference is unreliable. It follows a quite paradoxical situation. On the one hand, many theoretical and applicative works are built on the premise that key information is embedded in the degree distribution, while on the other hand, WDSs have characteristics that make *standard* nodal degree distribution evaluation very elusive and difficult to interpret, and do not allow for a reliable classification.

In fact, WDSs are generally planar urban infrastructure networks, strongly constrained by external geometrical/environmental factors—like the landscape topography, the structure of the cities (e.g., street network, building size, etc.)—decreasing the connectivity probability along with the distance between two nodes. It follows that WDS classification by the *standard* nodal degree distribution is very difficult and uncertain [*Yazdani and Jeffrey*, 2011] because of the very low values of the maximum *standard* nodal degree characterizing those networks.

In order to overcome this challenge, we introduce the *neighborhood* nodal degree which spans a range of values greater than the *standard* nodal degree, resulting in a statistically more reliable classification of infrastructure networks (hereinafter, we will specify *standard* as the usual nodal degree when a possible misunderstanding with *neighborhood* nodal degree can occur).

For each node in the network, the neighborhood node degree is defined as the sum of the degrees of its nearest neighbors (i.e., only nodes immediately adjacent). We demonstrate that when using this neighborhood degree approach, the Poisson distribution more reliably represents the connectivity of an infrastructure network.

Somewhat surprisingly, CNT literature reports very few investigations on WDSs despite their ubiquity, topological variety, and importance in everyday life. On the other hand, the availability of real WDSs is quite rare due to the sensible data associated with water supply service. Thus, the unusual availability of detailed topological data of 22 real WDSs built in different urban environments, allows a robust database for showing proof-of-concept of using *neighborhood* nodal degree to classify real infrastructure networks.

In the following, we analyze the detailed topological data of 22 real WDSs built in different urban environments. We demonstrate that the network connectivity structure of WDSs generally follows the Poisson distribution. The result is consistent with the fact that the network connectivity structure of WDSs is generally designed using a redundancy criterion that exceeds hydraulic capacity requirements. This provides low vulnerability of the connectivity structure with respect to any kind of failure event.

2. Degree Distribution

2.1. Degree Distribution and System Features

Robustness, vulnerability, resilience, and efficiency of real systems are increasingly investigated topics [*Albert et al*., 2000; *Cohen et al*., 2001; *Callaway et al*., 2000; *Holme et al*., 2002; *Albert et al*., 2004; *Solé et al*., 2007; *Berche et al*., 2009; *Iyer et al*., 2013; *Peng and Wu*, 2016], especially with respect to the CNT and its metrics. Among the various CNT metrics, the nodal degree distribution, i.e., the classification of a network using the Pareto or Poisson distributions, is a primary characteristic.

Vulnerability assessment of infrastructure networks is becoming a relevant issue because of socioeconomic relevance. The degree distribution, and in particular the degree of heterogeneity, greatly influences the vulnerability of real networks. Despite their low vulnerability to random attacks, networks with high heterogeneity, which usually follow a Pareto distribution, are extremely vulnerable to intentional attacks because of the presence of hubs. Instead, networks with a very low degree of heterogeneity (i.e., nodal degree distribution is random around an average value and the network is characterized by a high homogeneity), usually follow a Poisson distribution and have a low level of vulnerability to random and intentional attacks. The classification relates to the average length of the shortest paths among nodes in the network and is not dependent on size, i.e., number of nodes.

The vulnerability assessment of a network [*Latora and Marchioni*, 2005; *Berardi et al*., 2014] can be simulated by removing nodes, i.e., the connected links, at random or by targeting those corresponding to an intentional attack; node removal changes the preferential paths between nodes, increasing the shortest paths. The behavior of Pareto or Poisson networks to nodal removal is different. In fact, for networks with a high homogeneity (Poisson), i.e., where all nodes have approximately the same connectivity—a targeted or random removal of nodes causes approximately the same minimal increasing in the shortest paths, i.e., a low damage due to the redundant feature. For network with a low homogeneity (Pareto)—i.e., where few nodes have high connectivity (named hubs) with respect to the others having similar connectivity—a random removal of nodes causes approximately the same minimal increasing in the shortest paths, because the probability of removing hubs is low. On the contrary, a targeted removal of hubs causes a dramatic increase of the shortest paths, with a significant probability of network division in components.

2.2. Degree Distribution Models

The Empirical nodal degree distribution $P(k)$ is defined as the fraction of nodes in the network having degree k. Hence

$$P(k)=\frac{n_k}{n}, \tag{1}$$

where n_k is the number of nodes having degree k and n is the total number of nodes in the network. The formulation of the Poisson distribution, approximating the binomial distribution, for nodal degrees [*Watts and Strogatz*, 1998] is

$$P(k)=\binom{n-1}{p}p^k(1-p)^{n-1-\langle k\rangle}\approx\frac{e^{-\langle k\rangle}\langle k\rangle^k}{k!}, \tag{2}$$

where p is the probability of connection between two nodes and $<k>$ is the average nodal degree of the network. The Poisson distribution models random networks and it exhibits a peak value at $<k>$ and has a high probability of nodal degree around $<k>$, i.e., a high homogeneity of the degree distribution. The nodal degree distribution of small world networks has essentially the same features as the random networks [*Barthélemy*, 2011]. Therefore, the Poisson distribution can apply to small world networks arguing a narrow distribution of degrees, i.e., a very high homogeneity of the nodal degree distribution, which degenerates to a single degree (considering only internal nodes) for the case of regular networks, characterized by an absolute homogeneity. In fact, *Watts and Strogatz* [1998] introduced the small world networks using a ring lattice regular network, increasing the probability of connection, p.

In the present study, we show that the *neighborhood* nodal degree distribution of WDSs models a Poisson distribution very well and argue that it also models the *standard* nodal degree. It should be noted that *Watts and Strogatz* [1998] assumed the probability p, determining the average degree of the network to be equal to

$$\langle k\rangle=p(n-1). \tag{3}$$

In the case of the infrastructure networks, the average nodal degree strongly depends on spatial constraints and, generally, is very low (e.g., in the case of WDSs, $<k>$ ranges from 2 to 3).

Table 1. Distributions $P(k)$ and the Corresponding Networks

	Regular	Small World	Random
Poisson-binomial distribution (equation (2))		Probability →	
Pareto-power law distribution (equation (5))		Scale free	

Hence, we can write

$$p = \frac{\langle k \rangle}{n-1} \approx \frac{2.5}{n-1} \quad \Rightarrow \quad p \propto n^{-1}. \tag{4}$$

Namely, the probability of connection is inversely proportional to the size of the network because the spatial constraints strongly limit the range of variability for $<k>$. Equation (4) shows that p is generally very low for the connectivity structure typical of infrastructure networks and it decreases with the network size. As a result, the *Watts and Strogatz* [1998] scheme must be used with caution.

Finally, the formulation of the Pareto (or power law) model for the nodal degree distribution is

$$P(k) \approx k^{-\gamma}, \tag{5}$$

where γ is a constant generally ranging from 1.5 to 3 [*Newman*, 2010; *Barthélemy*, 2011]. The model of equation (5) applies to the scale-free networks, which are nonhomogeneous with respect to the nodal degree distribution (i.e., few nodes with a large degree (hubs). Table 1 shows degree distribution models.

3. Neighborhood Nodal Degree Distribution

Classification of spatial networks by the structure of the *standard* degree distribution is generally hampered by the very low maximum nodal degree. This fact is a consequence of the spatial limits constraining the network topology at the local (nodal) scale [*Barthélemy*, 2011; *Lämmer et al*., 2006], making very difficult to model the degree distribution in a statistically meaningful way. For example, the topographic structure of a city typically involves the crossing of three-four roads and rarely does the nodal degree of a node-square to exceed six-eight.

In order to classify spatial networks using the degree distribution concept, we introduce a more suitable idea: the *neighborhood* nodal degree. That is, we define a degree distribution involving the nearest neighbors or adjacent nodes. The *neighborhood* degree distribution of each node is the sum of the *standard* degrees of the topologically nearest (i.e., adjacent) nodes. The approach implies a significant increase in the maximum nodal degree and the greater number of points supports the robust identification of a specific statistical distribution.

The formulation of the neighborhood degree is

$$k_n(i) = \sum_{j \in N(i)} A_{ij} k(j), \tag{6}$$

where $k_n(i)$ is the "neighborhood" degree (involving adjacent nodes) of the i-th node, A_{ij} are the elements of the adjacency matrix, $k(j)$ is the *standard* degree of the j-th node, and $N(i)$ is the topological neighborhood of the i-th node, i.e., the set of adjacent nodes. Therefore, $k_n(j)$ is the product between the *standard* nodal degree and the i-th row of the adjacency matrix providing a nonnull value for the nearest/adjacent nodes only.

Figures 2a and 2b show illustrative examples of the neighborhood degree evaluation. Consider Figure 2a and focus on the node 1, 2. Its neighborhood is constituted by nodes labeled {2,3,4}, whose nodal degrees are equal to $k(2) = 4$, $k(3) = 3$, and $k(4) = 3$, respectively. In this case, the *neighborhood* degree of node 1 is the sum of the degree of the three adjacent nodes $k_n(1) = 4 + 3 + 3 = 10$. Similarly, this strategy assigns a degree $k_n(2) = 3 + 3 + 3 + 4 = 13$ to node 2 (see Figure 2b).

The *neighborhood* nodal degree can be interpreted as the node strength [*Boccaletti et al*., 2006] of a suitable oriented weighted network, which has the structure of the infrastructure network adjacency matrix. Each nonnull cell i,j has a value corresponding to the link weight between i and j, which is assumed equal to the degree of the starting node. Figures 2c and 2d report the weights to be assigned to the links connected to node 1 and 2. In Figure 2c the link from node 1 to node 2 has a weight, w_{12}, equal to 3 because $k(1) = 3$,

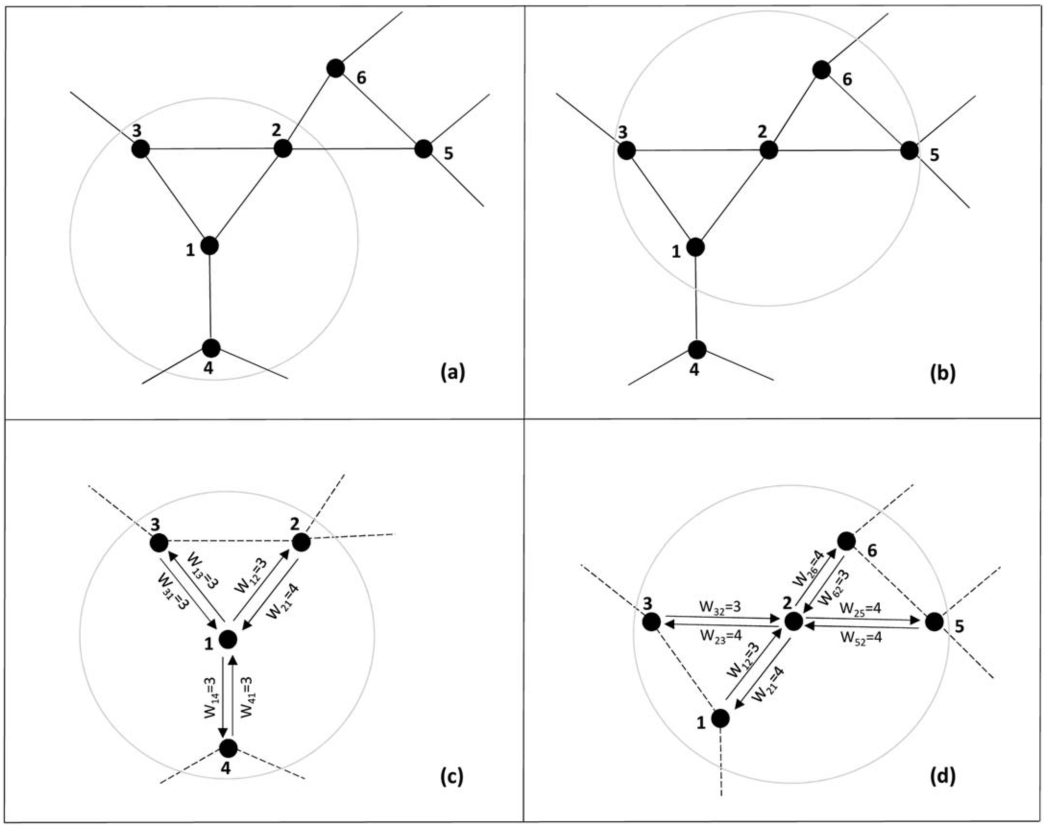


Figure 2. Example of a portion of a network. Figures 2a and 2b show the *neighborhood* of node 1 and node 2, and Figures 2c and 2d show the weights assigned to links incident nodes 1 and 2, respectively.

while the link from node 2 to node 1 has a weight $w_{21}=4$ because $k(2)=4$. The strategy is similarly applied to all nodes in the network.

The neighborhood degree of this oriented weighted network may be seen as the node strength [*Boccaletti et al*., 2006], namely

$$k_n(i)=\sum_{j\in N(i)} w_{ji}, \quad (7)$$

where w_{ij} are the previously defined topological weights. In this way, the weights can be interpreted as a sort of flow of topological information, which travels on each link, and the *neighborhood* degree is the overall information that flows into each node from the nearest neighbors.

In the previous sections, we discussed the importance of considering the nodal degree distribution of networks as tool for vulnerability analysis. In this section, we introduced the concept of *neighborhood* degree for the classification of infrastructure networks. While the *standard* degree measures the nodal connectivity with adjacent nodes in terms of number of links, the *neighborhood* degree measures the nodal connectivity at the neighborhood scale/level.

This extension represents a benefit in the analysis of WDS vulnerability, where the single node may not be relevant from a technical standpoint because it represents, for example, a connection to a private property. However, it may exist for reaching numerous customers. Furthermore, the existence of isolation valves allows one to argue that the failure of a single node actually involves the change of paths not only related to that node, but also related to its neighbors. In this sense, it is possible to extend the neighborhood degree concept to further levels of neighbors, which should depend on the purpose of the isolation valve system.

Therefore, the advantage of using the neighborhood degree for infrastructure networks is twofold. On the one hand, it guarantees a more robust identification of a specific statistical distribution; on the other hand, it allows accounting for the relevance of a single node in terms of its connectivity in spatial networks.

4. Classifying the Network Structure: Neighborhood Versus Standard Nodal Degree Distribution

In order to test the effectiveness of *neighborhood* nodal degree to classify infrastructure networks, we start by showing some results using a small but realistic infrastructure network. Figure 3 shows the network layout of the battle of background leakage assessment water network (BBLAWN) described in greater detail in *Giustolisi et al*. [2016]. The network is composed of 390 nodes and 439 links, and it is frequently used as a benchmark network in studies about WDSs [*Ostfeld et al*., 2008]. Figure 4 reports the Empirical density distributions, $P(k)$ and $P(k_n)$, and the corresponding cumulative distributions, $P_{cum}(k)$ and $P_{cum}(k_n)$, of the BBLAWN related both to the *standard* nodal degree, k (Figures 4a and 4c), and the *neighborhood* nodal degree, k_n (Figures 4b and 4d). Actually, $P_{cum}(k_n)$ is not the cumulative of $P_{cum}(k)$, but one minus its cumulative in order to be consistent with the common use in CNT literature.

The same figure reports the corresponding theoretical Poisson and Pareto distributions, for comparison to the Empirical distribution. To allow for easier comparison of the Pareto distribution, we also report the cumulative distributions also in logarithmic scale (Figures 4e and 4f).

Figures 4a, 4b and 4c illustrate that identifying a specific probabilistic model is very difficult when using the *standard* nodal degree. The maximum degree is five and, consequently, only five points are available to fit the distribution. It follows that discerning between the Poisson and Pareto distributions is difficult and statistically unreliable. The only reasonable conclusion from Figures 4a, 4c, and 4e is that the Empirical nodal degree distribution seems to be qualitatively more similar to the Poisson model than the Pareto model.

On the contrary, Figures 4b, 4d, and 4f relate to the *neighborhood* nodal degree allow one to discern between the Poisson and Pareto distributions because the maximum degree is equal to 12, and therefore distributions have a wider interval of degrees. In fact, the 11 points (degrees) make the statistical analysis reported in Table 2 more reliable. The Empirical distributions of Figure 4 follow the Poisson distribution very well especially considering the cumulative distribution. The differences between the Empirical data and the Pareto distribution are more evident in the logarithmic scale of the cumulative distribution.

We systematically compared *standard* and *neighborhood* nodal degree distributions for 22 networks corresponding to existing WDSs. Table 2 reports their relevant characteristics. In particular, it is worth noting that network size spans 2 orders of magnitude; in fact, the number of nodes ranges from 390 to 26,761, while the number of links varies from 439 to 32,096. In spite of the size variability, the average nodal degree remains consistently very low—ranging from 2.08 to 2.59—as well as the maximum nodal degree, ranging from 5 to 11. These values confirm the situation reported in the previous paragraphs about the BBLAWN. The small number of points of the Empirical distribution frustrates any attempt to infer a statistical model. In contrast, the *neighborhood* nodal degree depicts a very different picture: the average and maximum values range from 4.68 to 7.98 and from 12 to 33, respectively. The minimum *neighborhood* nodal degree is equal to 2 as it corresponds to the extreme nodes of network branches comprised of serial pipes/links. Furthermore, on a logarithmic scale, Empirical curves do not report a final value because the log of zero is not defined.

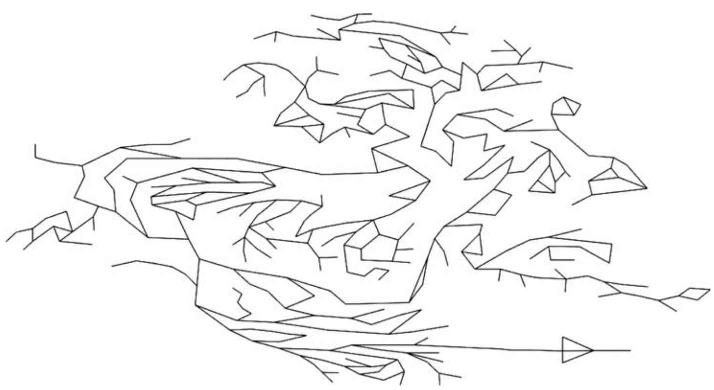


Figure 3. BBLAWN layout.

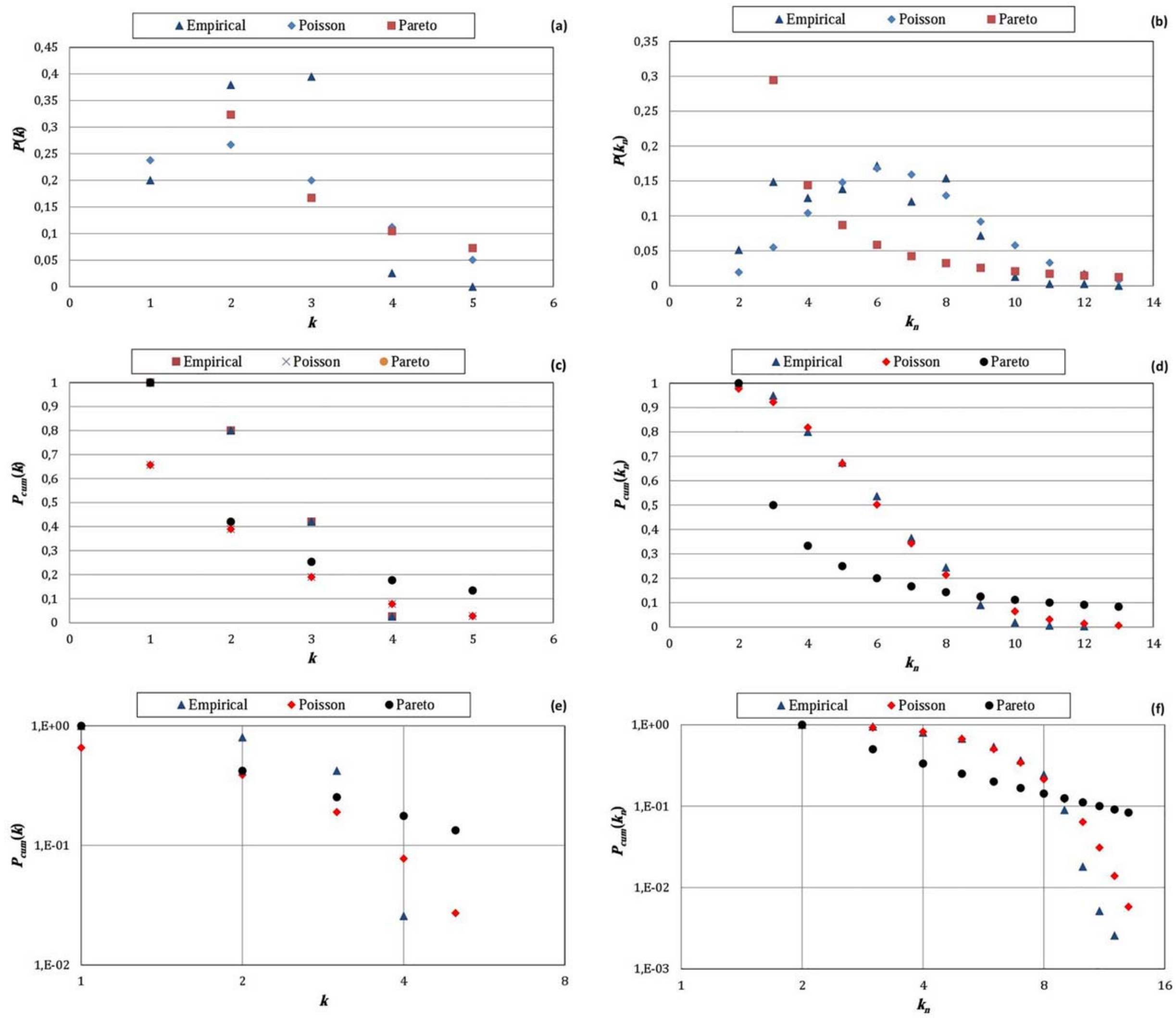


Figure 4. Empirical, *standard*, and *neighborhood*, nodal degree versus Poisson and Pareto distributions.

Figure 5 refers to the case of a large urban network. Figure 5a shows the network structure, which retraces clearly the urban infrastructure, while the other panels compare the Empirical density and cumulative distributions of *neighborhood* nodal degree to the theoretical Poisson and Pareto distributions. These latter are fit using the average value of the *neighborhood* nodal degree and a calibrated value of γ of equation (5), respectively. As in Figure 4, we report the cumulative distributions using both arithmetical and logarithmic scales. The comparisons clearly show that the Empirical distribution of *neighborhood* nodal degree is very similar to the Poisson distribution, while it substantially differs from the Pareto distribution.

In Appendix, we report the layouts and the cumulative distributions in logarithmic scale for all the WDSs listed in Table 2, excluding BBLAWN, Big Town, and Exnet [*Laucelli and Giustolisi*, 2014], with the exception of the Exnet network, all of the comparisons demonstrate that the Poisson distribution models very well the Empirical distribution using the *neighborhood* nodal degree. The visual inspection is confirmed by the two-sample Kolmogorov-Smirnov (KS) goodness-of-fit hypothesis test [*Massey*, 1951], which discerns whether if the distributions of the values in the Empirical and Poisson samples are drawn from the same underlying population. The test results are reported in the third last column of Table 2 and clearly show a high level of significance. It is worth noting that the Exnet network is the only one having a low significance level (about 26%).

Figure 6 shows the Exnet results. The Empirical cumulative distribution (in log scale) partially follows the Poisson model up to value of k_n equal to about 13 and then it becomes linear, similarly to the Pareto model. This is the only case, among the analyzed network connectivity structures that show features of both

Table 2. Relevant Data of the 22 Networks (for Few Benchmark Real WDSs the Inhabitants Are Not Available)

WDN Name	Node #	Pipe #	Mean Standard	Min Standard	Max Standard	Mean Nearest Neighbor	Min Nearest Neighbor	Max Nearest Neighbor	KS Test [%]	Length [km]	Inhabitants [×1000]
BBLAWN	390	439	2.25	1	5	5.69	2	13	78.6	57	
Apulia 15	1,263	1,406	2.23	1	5	5.56	2	15	99.5	30	16
Apulia 14	1,263	1,428	2.26	1	5	5.78	2	15	99.7	46	17
Apulia 13	1,762	2,098	2.38	1	5	6.38	2	17	100.0	70	25
Exnet	1,776	2,300	2.59	1	11	7.98	2	33	25.8	594	
Apulia 12	1,918	2,153	2.25	1	6	5.74	2	16	100.0	108	56
Apulia 11	2,403	2,820	2.35	1	5	6.11	2	15	99.7	145	94
Norway 2	2,520	2,651	2.10	1	7	4.87	2	18	99.8	129	24
Piedmont 1	2,784	2,894	2.08	1	5	4.89	2	12	98.5	171	49
Apulia 10	2,810	3,307	2.35	1	5	6.26	2	16	99.8	175	398
Apulia 9	2,895	3,333	2.30	1	5	5.86	2	16	99.8	85	30
Apulia 8	2,968	3,400	2.29	1	5	5.89	2	16	100.0	130	60
Apulia 7	3,000	3,189	2.13	1	5	5.12	2	13	99.1	91	49
Apulia 6	3,547	3,881	2.19	1	5	5.37	2	17	100.0	113	33
Apulia 5	4,188	4,727	2.26	1	5	5.82	2	14	99.5	261	94
Apulia 4	4,242	4,940	2.33	1	6	6.06	2	17	100.0	161	70
Norway 1	5,035	5,292	2.10	1	6	4.68	2	13	78.6	239	
Apulia 3	5,036	5,848	2.32	1	6	5.98	2	16	99.8	199	107
Apulia 2	5,288	6,116	2.31	1	6	6.01	2	17	99.9	277	151
BWSN	12,518	14,314	2.29	1	6	5.80	2	17	99.9	1,844	
Apulia 1	18,718	19,990	2.14	1	7	5.01	2	16	99.8	678	326
Big Town	26,761	32,096	2.40	1	8	6.18	2	17	91.2	2,054	1,347

Poisson and Pareto models consistent with *Boccaletti et al*. [2006]. It is worth noting that (Table 2) the Exnet network has the greatest maximum *standard* and *neighborhood* nodal degree, respectively, equal to 11 and 33, revealing (Figure 6) the presence of a few hubs.

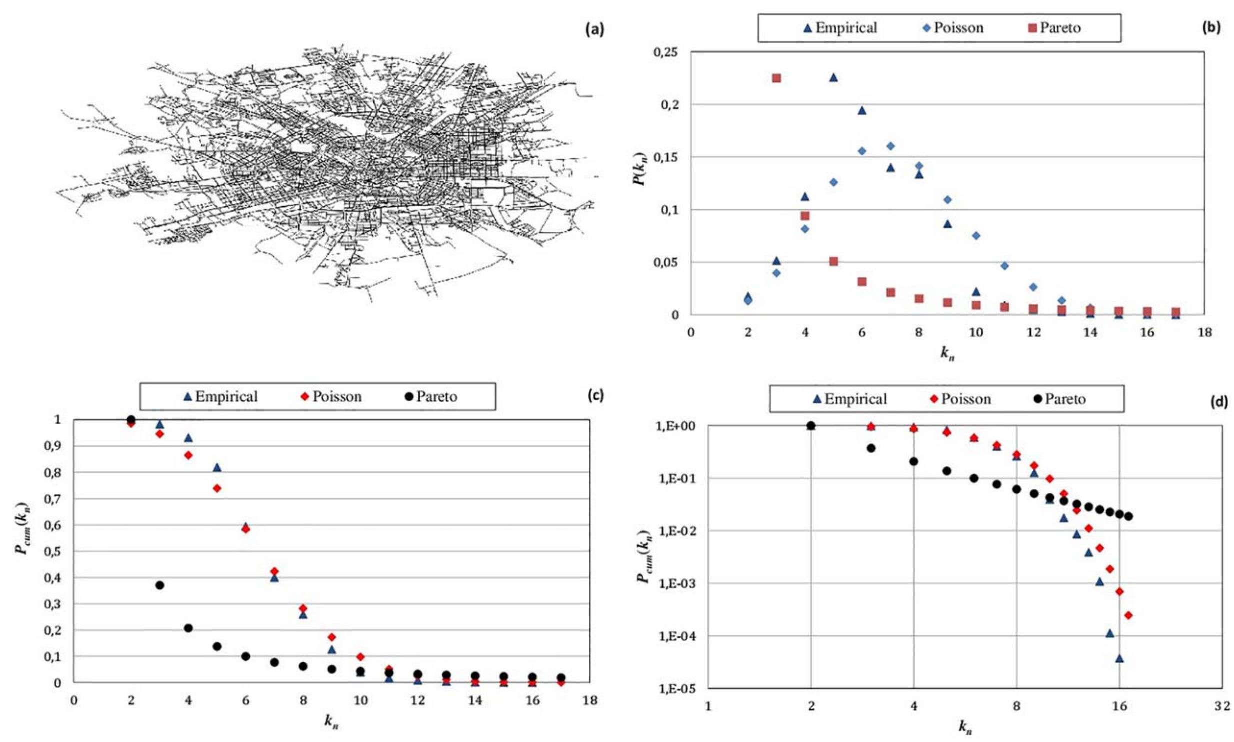


Figure 5. Layout of the Big Town network and Empirical versus Poisson and Pareto distributions.

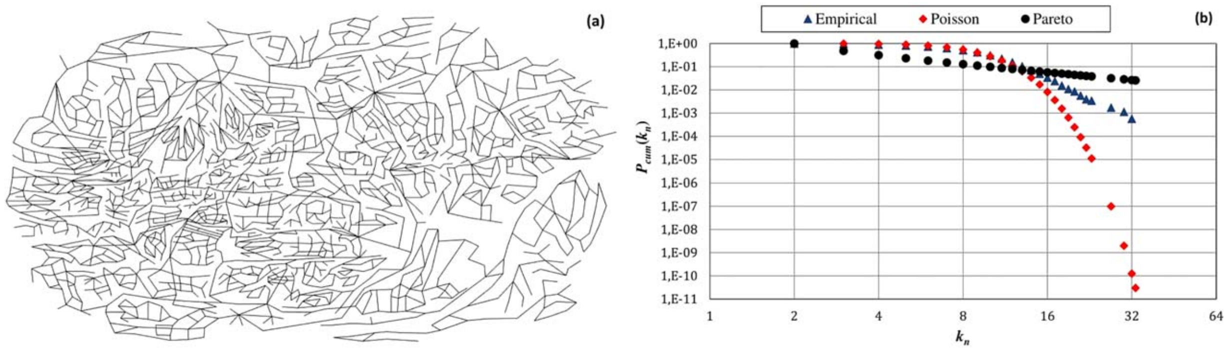


Figure 6. Layout of the Exnet network and Empirical versus Poisson and Pareto cumulative distributions.

5. Network Connectivity for Classification of WDNs

From a hydraulic standpoint, network nodes can represent any WDN component from which the water is delivered to the network from the mass and/or energy perspective (e.g., reservoirs, tanks, pumps, etc.), each one with different functionality (e.g., supply, demand, and storage), while the network connectivity analysis does not account for these differences.

Very few nodes are sources in the network and their characterization is trivial for vulnerability assessment [*Berardi et al*., 2014; *Laucelli and Giustolisi*, 2014]. In fact, in a network, the water source nodes represent special hubs, because the information departs from them, to be then transferred to the areas that they subtend. This means that they do not characterize the major part of the network connectivity structure, which is the focus of the present work.

Demand nodes are connection to properties—domestic, agriculture, industrial area, etc.—and represent hydraulic outflows of WDNs. Connection nodes, instead, are connections among pipes, e.g., in crossroads. Actually, the water delivered along pipes is often concentrated as outflows to the ending nodes of those pipes for hydraulic modeling needs [*Giustolisi*, 2010].

The finding that the network connectivity structure of WDNs can be generally modeled referring to random networks essentially refers to the classification of demand and connection nodes, which are fairly the majority of the nodes in the network. Therefore, our work does not focus on the vulnerability assessment, which has to account for the specific hydraulic behavior of a WDN coupled with topologic metrics [*Yazdani and Jeffrey*, 2012]. Instead, our purpose is a basic classification of the network connectivity structure of WDNs aimed at characterizing the "emerging" behavior of the topological domain, which drives the hydraulic performance and changes over time due to the connectivity evolution, as reported in the next section.

6. Network Classification and Temporal Evolution

Infrastructure networks are one of most prevalent cases of spatial networks. The term *spatial* characterizes networks in which nodes are located in a space equipped with a metric [*Barthélemy*, 2011]. For most of the infrastructure networks, the space is two-dimensional and the metric is the usual Euclidean distance. The spatial layout/architecture of the network is typically constrained by the topology of the environment in which it is constructed. In fact, infrastructure networks are generally man-made and they progressively grow, filling the space and balancing connection costs and nodal distance, but they are also constrained by the impracticality of some connections [*Buhl et al*., 2004]. The existence of such constrains explains why spatial networks generally are not scale-free networks [*Barthélemy*, 2003, 2011], and in this line, our findings indicate that the Poisson distribution is generally the best model to describe the network connectivity (as seen by the *neighborhood* nodal degree) of WDSs.

The work by *Barthélemy and Flammini* [2008] on the temporal evolution of urban road networks (Figure 7) is helpful to understand our finding about WDSs, because these latter networks are strongly influenced by the topology of urban patterns. An initial network (see Figure 7, top left) has a low size and appears quite regular (a ring with branches is the typical initial configuration). Afterward, the network evolves and new

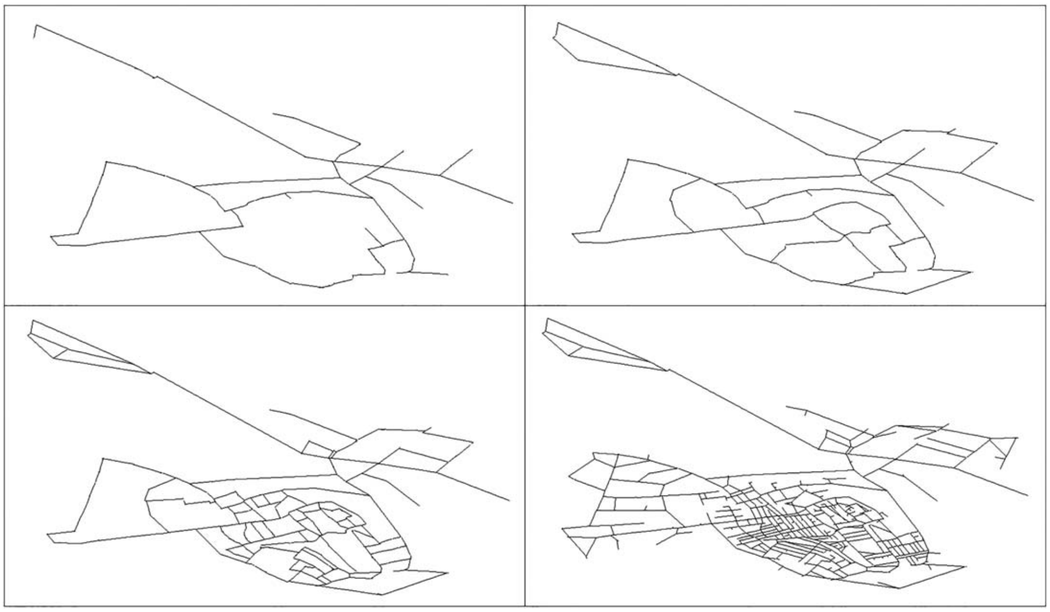


Figure 7. Example of snapshots of an urban network at different times of its evolution.

connections to customer proprieties are built. During this growth phase, the network is affected by the *spatial* constrains but a certain level of randomness emerges because of decisions to design system redundancy for WDS management. This is due to the different local shapes of such constrains (e.g., different buildings, housing blocks, districts, etc.). During its evolution and increasing size, the network can be increasingly classified into the category of random networks (and possibly small world networks). It follows that the Poisson distribution is the best suited to model the network structure of WDSs because the evolution of such systems, although constrained, introduces random characteristics increasing the network size.

7. Concluding Remarks

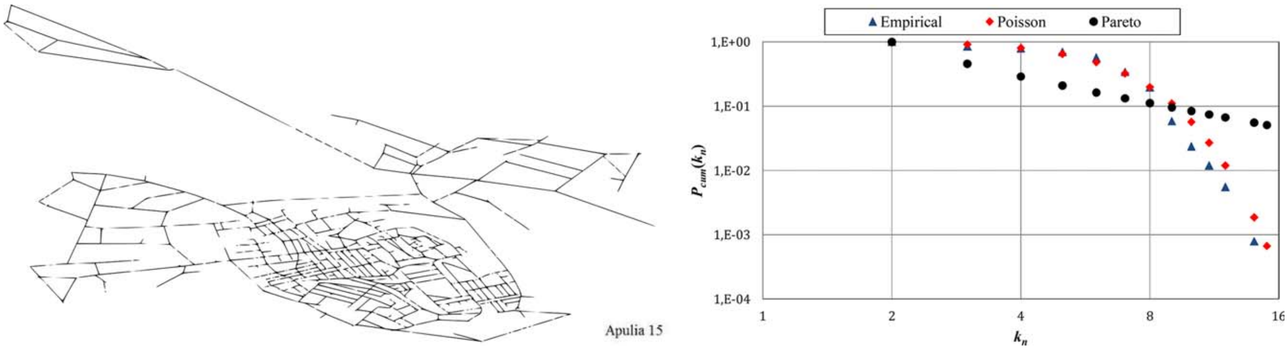
The aim of the present work was twofold. First, we proposed the *neighborhood* nodal degree. It is a novel quantity geared to describe the topology in the neighborhood of each node of a network. The second aim has been to show that *neighborhood* nodal degree distribution is suitable to classify infrastructure networks. Differently from the *standard* nodal degree distribution, which generally is uninformative due to the very limited range of nodal degree values, the *neighborhood* nodal degree exploits the topological information of the nearest neighbors and allows one to infer reliable probabilistic models. In particular, we have investigated 22 real water distribution networks having different sizes and characteristics. In almost all cases, a Poisson distribution fits the Empirical *neighborhood* nodal degree distribution very well. This result appears in agreement with the characteristics of networks whose evolution is constrained by urban patterns.

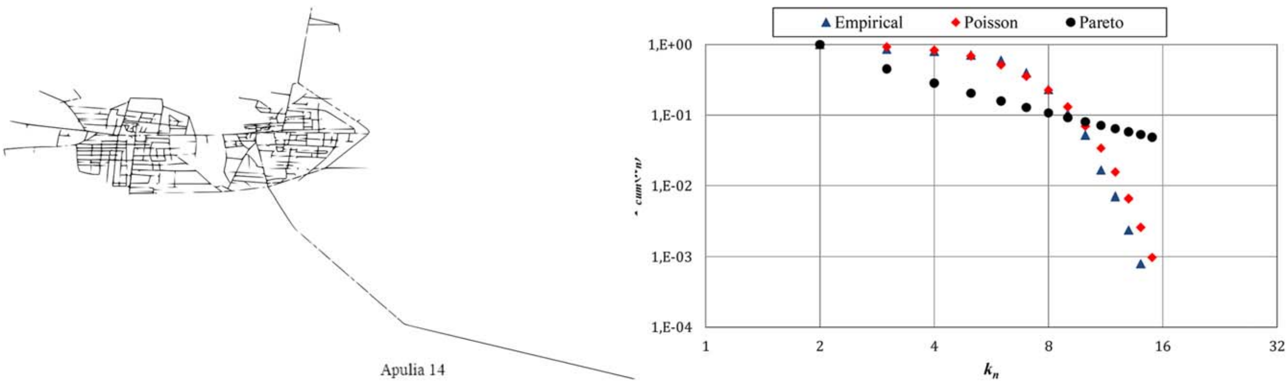
The above reported results about the classification of the network connectivity structures for WDNs are useful for analyzing their emerging behavior, for example related to the vulnerability due to random failures and intentional threats. The fact that most of the studied WDNs can be modeled using the Poisson distribution of the neighborhood nodal degree means that they present a significant structural resistance to random failures and intentional threats as connectivity structure.

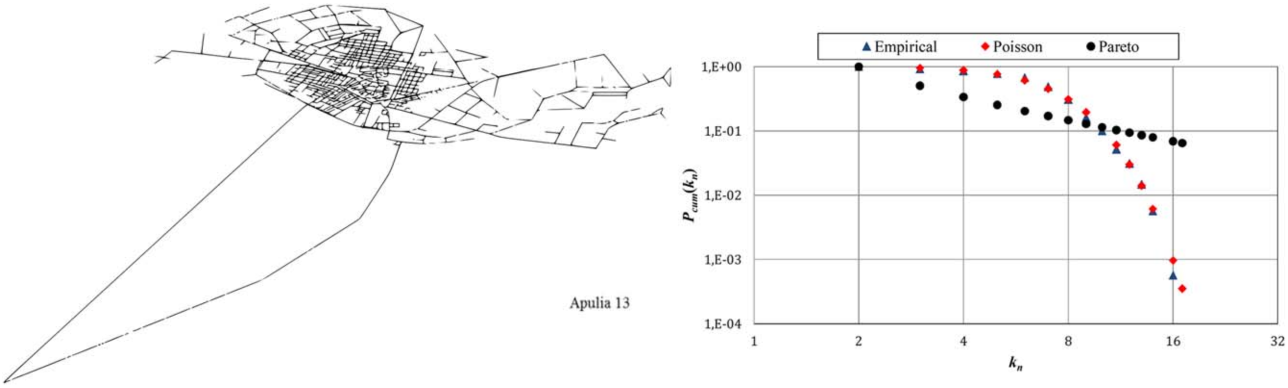
Different ways can be used to analyze the WDN behavior with respect, for example, to the vulnerability, but the classic classification of the connectivity structure nodal degree-based has wide implications for classifying them as small world or purely random considering temporal evolution or for assessing vulnerability considering for example hydraulic flows as weights while analyzing the connectivity structure.

Appendix A

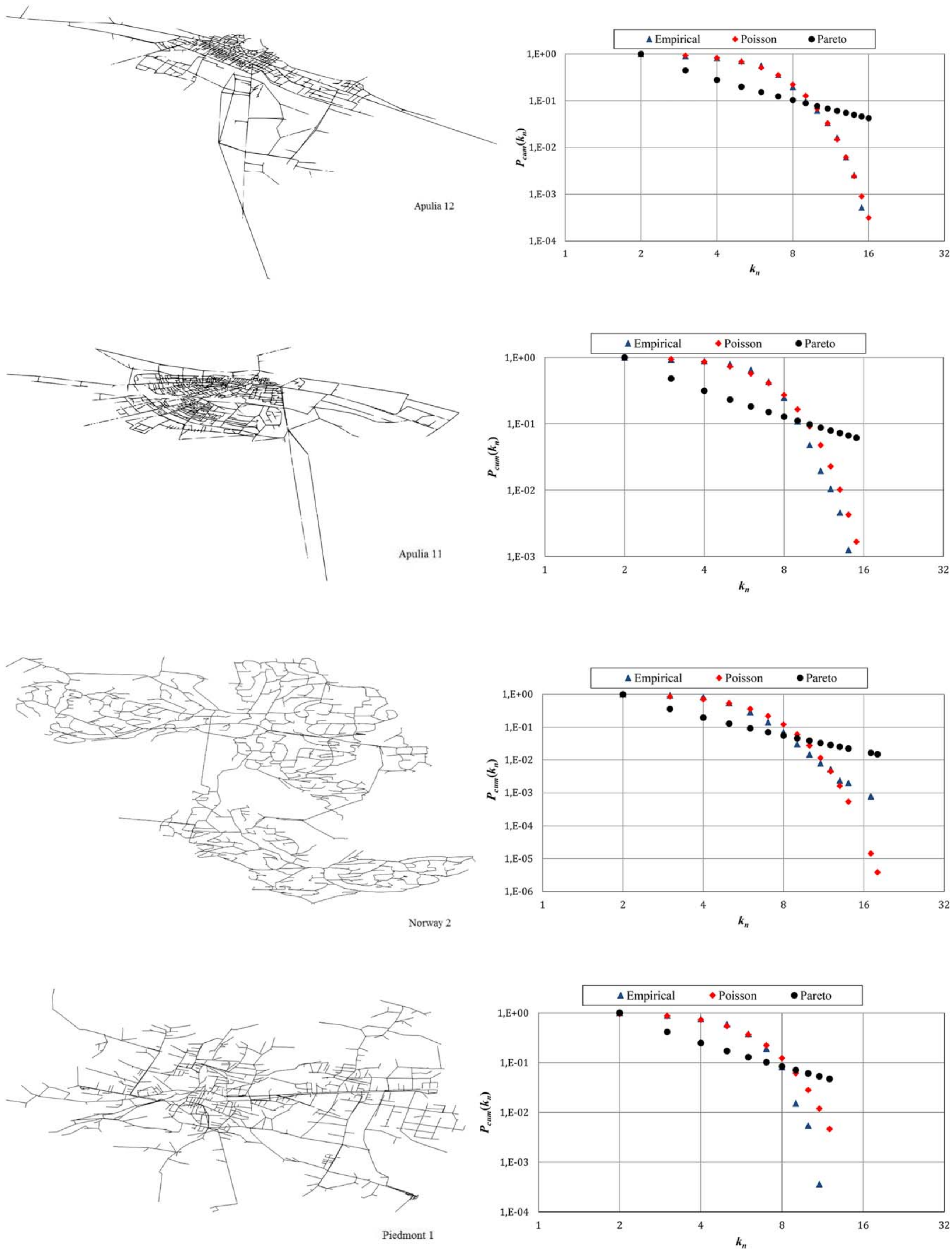
Layout and Empirical cumulative distribution of data versus the theoretical Poisson and Pareto distributions for the nineteen WDSs, which are not reported in the main text.

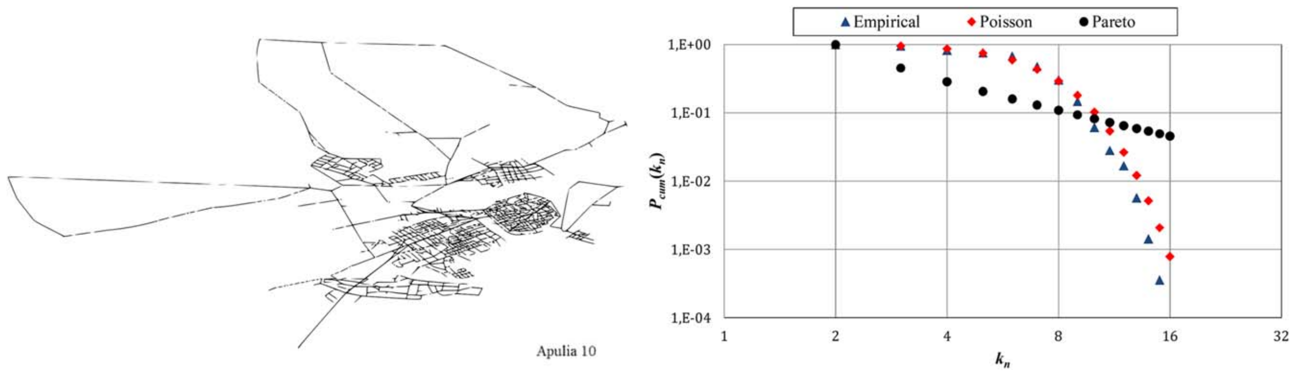


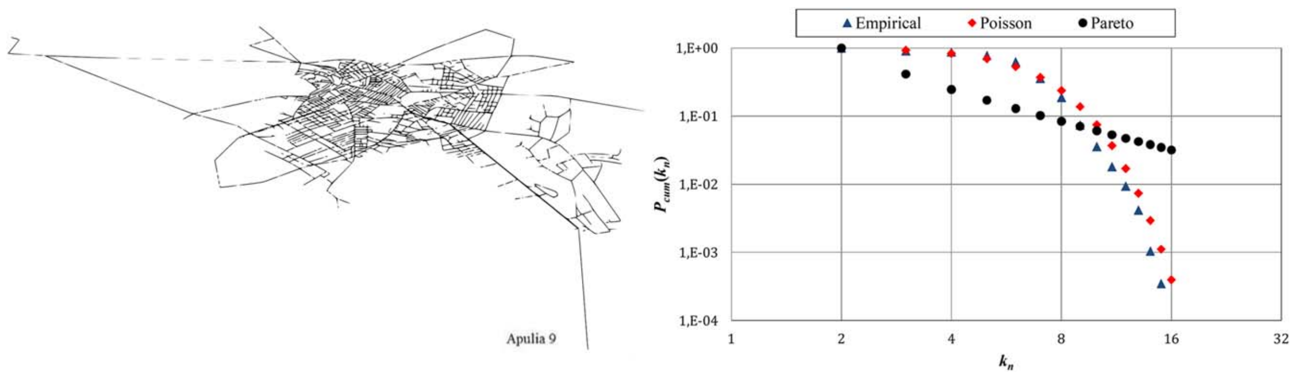


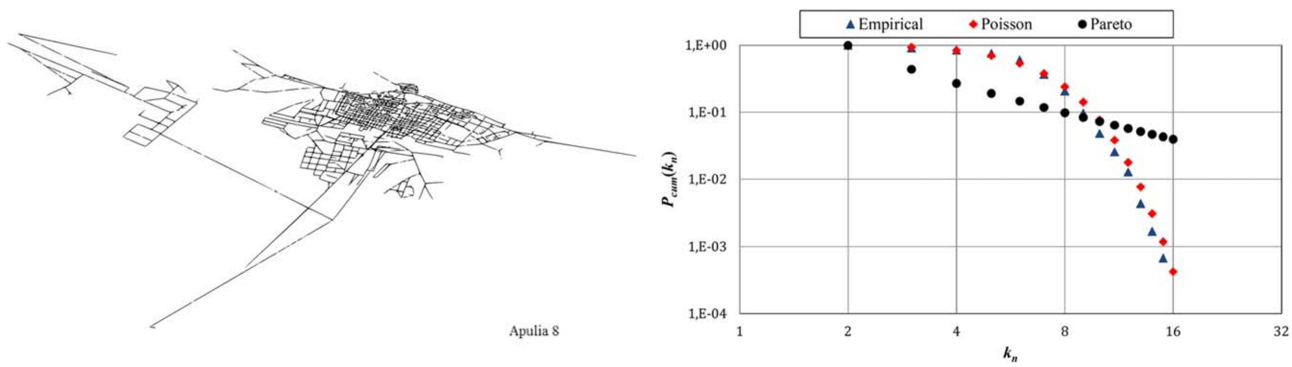


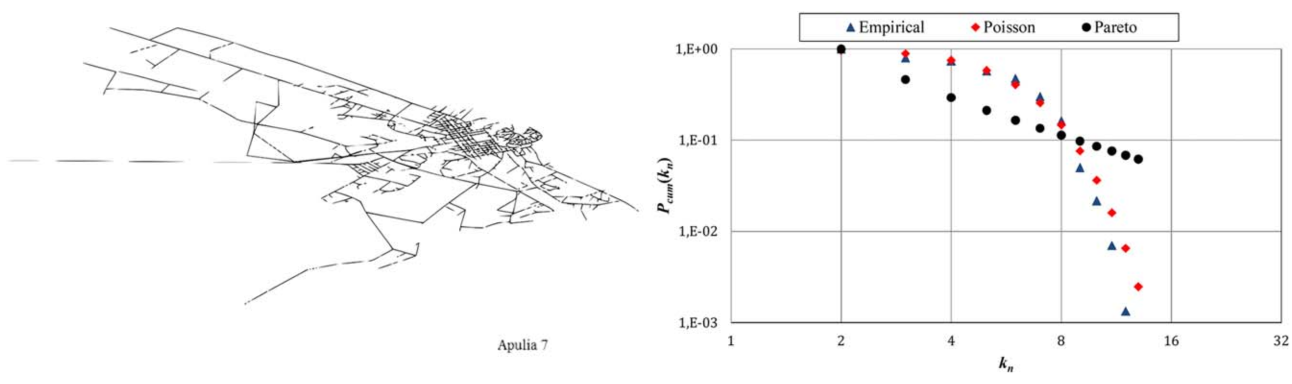
10.1002/2016WR020071

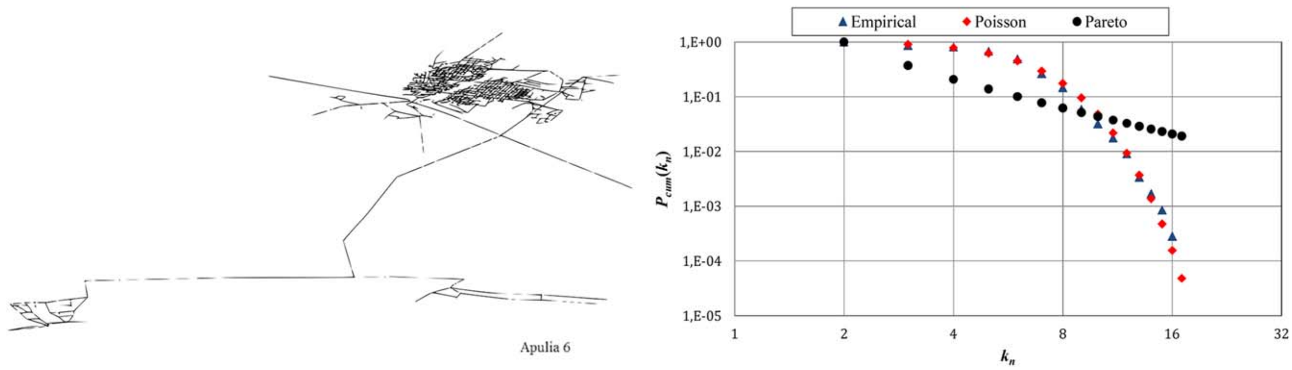


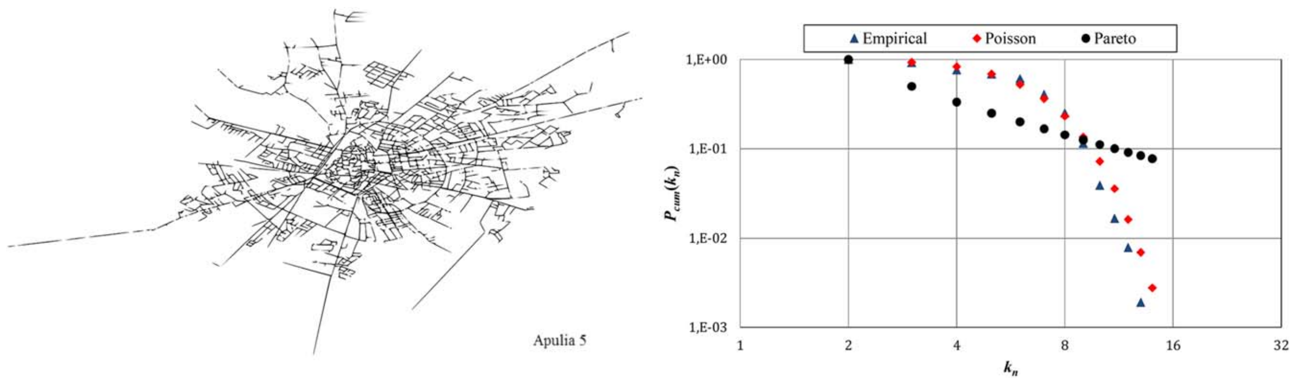

Apulia 10
Empirical
Poisson
Pareto
1,E+00
1,E-01
1,E-02
1,E-03
1,E-04
$P_{cum}(k_n)$
1
2
4
8
16
32
k_n

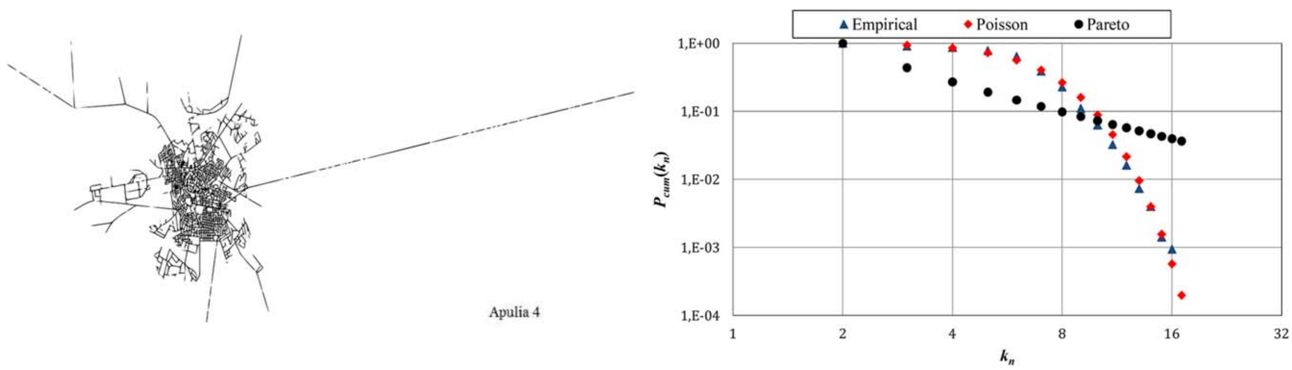

Apulia 9
Empirical
Poisson
Pareto
1,E+00
1,E-01
1,E-02
1,E-03
1,E-04
$P_{cum}(k_n)$
1
2
4
8
16
32
k_n

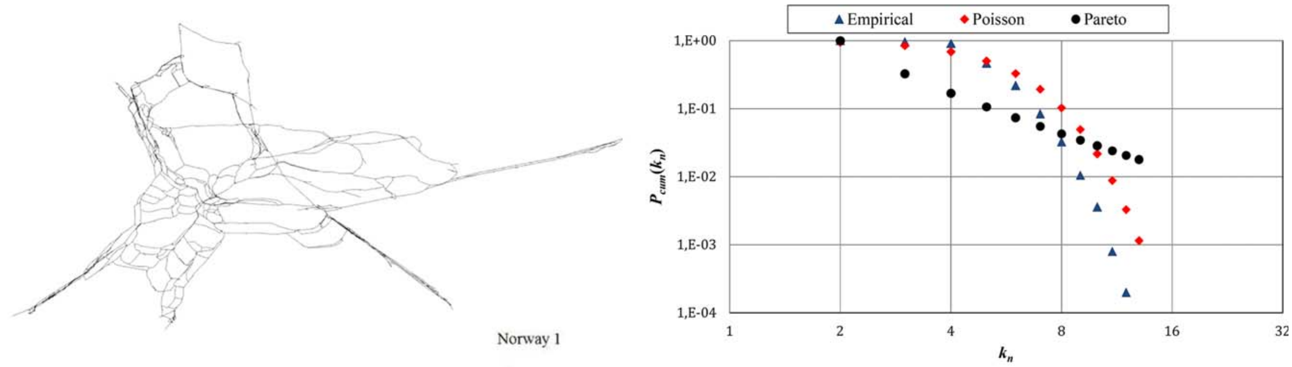

Apulia 8
Empirical
Poisson
Pareto
1,E+00
1,E-01
1,E-02
1,E-03
1,E-04
$P_{cum}(k_n)$
1
2
4
8
16
32
k_n

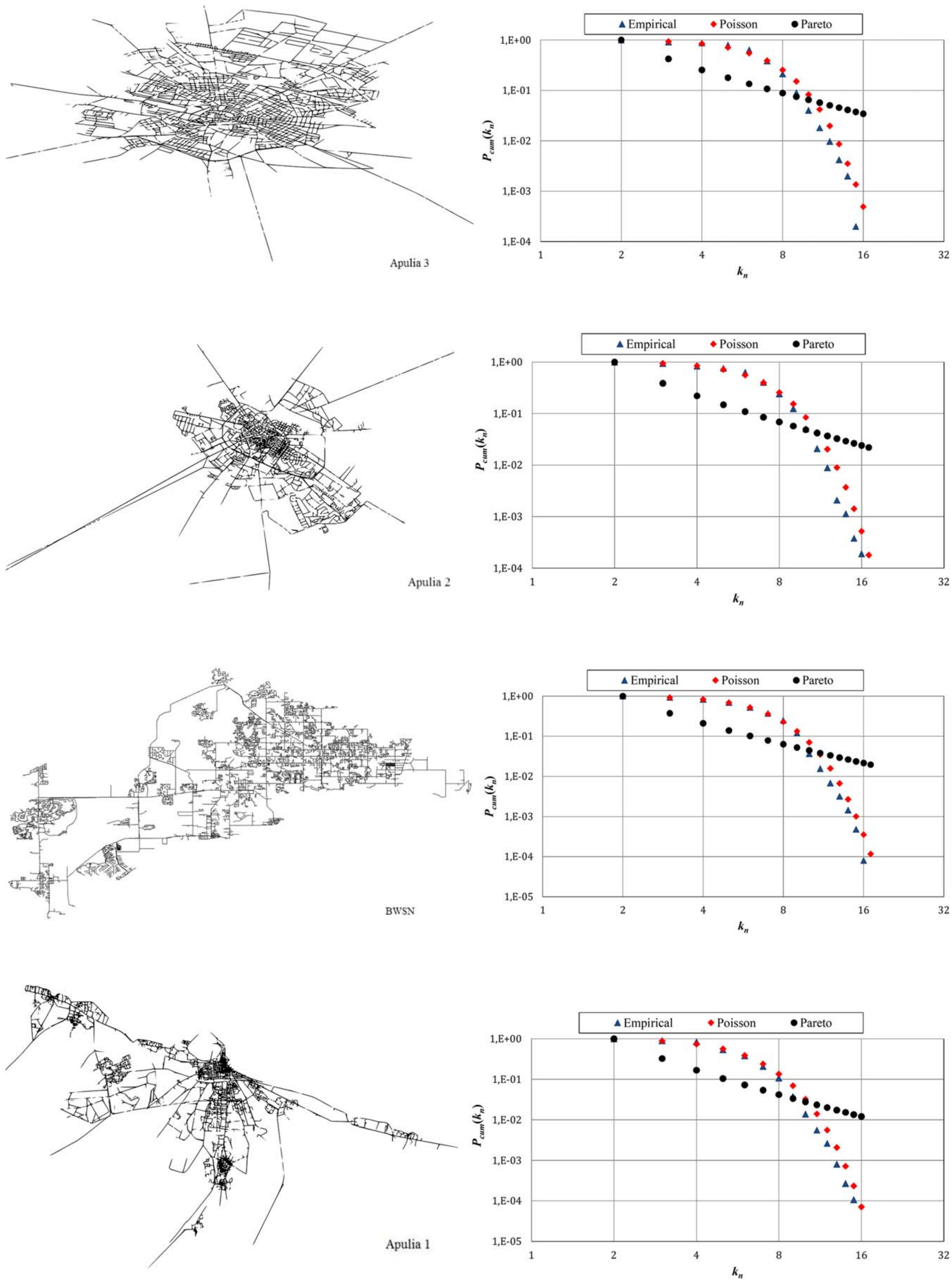

Apulia 7
Empirical
Poisson
Pareto
1,E+00
1,E-01
1,E-02
1,E-03
$P_{cum}(k_n)$
1
2
4
8
16
32
k_n


Apulia 6
Empirical
Poisson
Pareto
1,E+00
1,E-01
1,E-02
1,E-03
1,E-04
1,E-05
$P_{cum}(k_n)$
1
2
4
8
16
32
k_n


Apulia 5
Empirical
Poisson
Pareto
1,E+00
1,E-01
1,E-02
1,E-03
$P_{cum}(k_n)$
1
2
4
8
16
32
k_n


Apulia 4
Empirical
Poisson
Pareto
1,E+00
1,E-01
1,E-02
1,E-03
1,E-04
$P_{cum}(k_n)$
1
2
4
8
16
32
k_n


Norway 1
Empirical
Poisson
Pareto
1,E+00
1,E-01
1,E-02
1,E-03
1,E-04
$P_{cum}(k_n)$
1
2
4
8
16
32
k_n


Apulia 3
Empirical Poisson Pareto
$P_{cum}(k_n)$
k_n
1,E+00
1,E-01
1,E-02
1,E-03
1,E-04
1 2 4 8 16 32
Apulia 2
Empirical Poisson Pareto
$P_{cum}(k_n)$
k_n
1,E+00
1,E-01
1,E-02
1,E-03
1,E-04
1 2 4 8 16 32
BWSN
Empirical Poisson Pareto
$P_{cum}(k_n)$
k_n
1,E+00
1,E-01
1,E-02
1,E-03
1,E-04
1,E-05
1 2 4 8 16 32
Apulia 1
Empirical Poisson Pareto
$P_{cum}(k_n)$
k_n
1,E+00
1,E-01
1,E-02
1,E-03
1,E-04
1,E-05
1 2 4 8 16 32

Acknowledgments
The data used for the classification of network connectivity structures are available contacting Orazio Giustolisi (orazio.giustolisi@poliba.it). The Italian Ministry of Education, University and Research (MIUR) has supported this research, under the Projects of Relevant National Interest "Advanced analysis tools for the management of water losses in urban aqueducts" and "Tools and procedures for an advanced and sustainable management of water distribution systems."

References

Albert, R., and A.-L. Barabási (2002), Statistica mechanics of complex networks, *Rev. Mod. Phys.*, *74*, 47.
Albert, R., H. Jeong, and A.-L. Barabási (2000), Error and attack tolerance of complex networks, *Nature*, *406*(6794), 378–382.
Albert, R., I. Albert, and G. L. Nakarado (2004), Structural vulnerability of the North American power grid, *Phys. Rev. E*, *69*, 025103.
Barabási, A.-L. (2012), The network takeover, *Nat. Phys.*, *8*, 14–16.
Barabási, A.-L., and R. Albert (1999), Emergence of scaling in random networks, *Science*, *286*(5439), 509–512.
Barthélemy, M. (2003), Crossover from scale-free to spatial networks, *Europhys. Lett.*, *63*(6), 915–921.
Barthélemy, M. (2011), Spatial networks, *Phys. Rep.*, *499*(1), 1–101.
Barthélemy, M., and A. Flammini (2008), Modeling urban street patterns, *Phys. Rev. Lett.*, *100*, 138702.
Berardi, L., R. Ugarelli, J. Rostum, and O. Giustolisi (2014), Assessing mechanical vulnerability in water distribution networks under multiple failures, *Water Resour. Res.*, *50*(3), 2586–2599.
Berche, B., C. Von Ferber, and T. Holovatch (2009), Resilience of public transport networks against attacks, *Eur. phys. J. B*, *71*(1), 125–137.
Boccaletti, S., V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwa (2006), Complex networks: Structure and dynamics, *Phys. Rep.*, *424*(4–5), 175–308.
Buchanan, M. (2003), *Nexus: Small Worlds and the Groundbreaking Science of Networks*, edited by W. W. Norton and Co., Inc. New York, N. Y.
Buhl, J., J. Gautrais, R. V. Solé, P. Kuntz, S. Valverde, J. L. Deneubourg, and G. Theraulaz (2004), Efficiency and robustness in ant networks of galleries, *Eur. Phys. J. B*, *42*(1), 123–129.
Callaway, D. S., M. E. J. Newman, and S. H. Strogatz (2000), Network robustness and fragility: Percolation on random graphs, *Phys. Rev. Lett.*, *85*(25), 5468–5471.
Cohen, R., K. Erez, D. Ben-Avraham, and S. Havlin (2001), Breakdown of the internet under intentional attack, *Phys. Rev. Lett.*, *86*, 3682–3685.
Erdös, P., and A. Rényi (1959), On random graphs, *Publ. Math.*, *6*, 290–297.
Erdős, P., and P. Rényi (1960), On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci.*, *5*, 17–60.
Giustolisi O. (2010), Considering actual pipe connections in water distribution network analysis, *J. Hydraul. Eng.*, *136*(11), 889–900.
Giustolisi, O., L. Berardi, D. Laucelli, D. Savic, and Z. Kapelan (2016), Operational and tactical management of water and energy resources in pressurized systems: Competition at WDSA 2014, *J. Water Resour. Plann. Manage.*, *142*(5), 1–12.
Holme, P., B. J. Kim, and C. N. Yoon (2002), Attack vulnerability of complex networks, *Phys. Rev. E*, *65*, 056109.
Iyer, S., T. Killingback, B. Sundaram, and Z. Wang (2013), Attack robustness and centrality of complex networks, *Plos One*, *8*(4), e596163.
Lämmer, S., B. Gehlsen, and D. Helbing (2006), Scaling laws in the spatial structure of urban road networks, *Physica A*, *363*(1), 89–95.
Latora, V., and M. Marchioni (2005), Vulnerability and protection of infrastructure networks, *Phys. Rev. E*, *71*, 015103.
Laucelli, D., and O. Giustolisi (2014), Vulnerability assessment of water distribution networks under seismic actions, *J. Water Resour. Plann. Manage.*, *140*(10), 04014082-1-13.
Massey, F. J. (1951), The Kolmogorov-Smirnov test for goodness of fit, *J. Am. Stat. Assoc.*, *46*(253), 68–78.
Milgram, S. (1967), The small world problem, *Psychol. Today*, *1*, 60–67.
Newman, M. E. J. (2003), The structure and function of complex networks, *SIAM Rev.*, *45*, 167–256.
Newman, M. E. J. (2010), *Networks: An Introduction*, Oxford Univ. Press, Oxford, U. K.
Ostfeld, A., et al. (2008), The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms, *J. Water Resour. Plann. Manage.*, *134*(6), 556–568.
Papo, D., J. M. Buldú, S. Boccaletti, and E. T. Bullmore (2014), Complex network theory and the brain, *Philos. Trans. R. Soc. B*, *369*(1653), 20130520.
Peng, G. S., and J. Wu (2016), Optimal network topology for structural robustness based on natural connectivity, *Physica A*, *443*, 212–220.
Poisot, T., and D. Gravel (2014), When is an ecological network complex? Connectance drives degree distribution and emerging network properties, *PeerJ*, *2*, e251.
Schweitzer, F., D. Dornette, F. Vega-Redondo, and D. R. White (2009), Economic networks: What do we know and what do we need to know?, *Adv. Complex Syst.*, *12*, 407.
Scott, J. (1988), Social network analysis, *Sociology*, *22*(1), 109–127.
Solé, R. V., M. R. Casals, and B. C. Murtra (2007), Robustness of the European power grids under intentional attack, *Phys. Rev. E*, *77*(2), 026102.
Watts, D. J., and D. H. Strogatz (1998), Collective dynamics of small-world networks, *Nature*, *393*, 440–442.
Yazdani, A., and P. Jeffrey (2011), Complex network analysis of water distribution systems, *Chaos*, *21*, 016111.
Yazdani, A., and P. Jeffrey (2012), Water distribution system vulnerability analysis using weighted and directed network models, *Water Resour. Res.*, *48*, W06517, doi:10.1029/2012WR011897.