Alma Mater Studiorum · Università di Bologna

SCUOLA DI SCIENZE

Corso di Laurea Magistrale in Informatica

Complex networks theory for water distribution networks modelling and resilience assessment. An explorative analysis

Relatore:
Chiar.mo Prof.
Stefano Ferretti

Correlatore:
Dr.
Andrea Cominola

Presentata da:
Irene Basaglia

III Sessione
Anno Accademico 2019/2020

A mia sorella Chiara.

Sommario

Il tema della sostenibilità e dell' utilizzo di nuove tecniche per la gestione delle risorse idriche è da sempre molto complesso e dibattuto. La digitalizzazione e l'impiego di nuove tecnologie in campo idrico ha ridotto sia i costi che le tempistiche di molteplici processi. Un valido esempio è costituito dalla modellazione surrogata, una tecnica ingegneristica che costruisce un modello di approssimazione imitando il comportamento del modello originale il più fedelmente possibile. I modelli surrogati hanno un basso costo computazionale e diminuiscono le tempistiche di realizzazione, per questa ragione vengono utilizzati in disparati settori in cui sono presenti esperimenti e simulazioni di difficile attuazione.
Il tema centrale della tesi ha come scopo la costruzione di un modello surrogato di rete di distribuzione idrica utilizzando i principi di Complex Network Theory. In particolare, per condurre questo tipo di studio è stato scelto di analizzare da un punto di vista interdisciplinare i concetti di robustezza e di resilienza ai malfunzionamenti. Per la loro naturale complessità e ramificazione, le reti di distribuzione idrica possono essere facilmente assimilabili alle reti complesse presenti nella teoria dei grafi. Gli elementi strutturali delle reti idriche (come valvole, pompe, serbatoi, giunzioni e tubature) sono convertiti nell'equivalente grafo complesso costituito da archi e nodi.
Al fine di indagare una possibile correlazione tra le reti di Water Distribution Network Theory e Complex Network Theory, sono state eseguite simulazioni idriche e valutazioni strutturali calibrando i modelli in base alle euristiche proposte in letteratura. Gli esperimenti eseguiti valutano l'effetto "small world" (la teoria del mondo piccolo) e altre misure per l'analisi di reti presenti nella teoria dei grafi in relazione all'intensità e alla durata dei malfunzionamenti, per esempio, deficit nella domanda idrica soddisfatta, a causa di insufficiente pressione in rete. Dai risultati è stata riscontrata una

correlazione inversa tra la proprietà di "small world" e l'indice di durata dei malfunzionamenti. In aggiunta, i risultati ottenuti evidenziano una tendenza generale alla correlazione tra metriche di entrambe le discipline, sia diretta che indiretta. Un'ulteriore evidenza emersa dall'analisi è la doppia correlazione presente tra altezza dei nodi, durata dei malfunzionamenti e indice di centralità per vicinanza. In particolare, l'importanza del ruolo che l'altezza dei nodi assume nella conversione da modello idrico a modello complesso rappresenta una novità nella modellizzazione poiché non vi sono molti studi a riguardo.
Questo lavoro può dunque aiutare nell'individuazione di parametri più puntuali e adeguati suggerendo nuove vie e tecniche per una ancora più corretta creazione di modelli surrogati tramite tecniche di Complex Network.

Abstract

Sustainable theme and usage of novel techniques for water resources management have always been debated. Digitalization and the use of new technologies in the water field has reduced both costs and time for multiple projects. A good example is surrogate modeling, an engineering technique that builds an approximation model imitating the behaviour of the original model. Surrogate models have a low computational cost and reduce implementation time, for this reason they are used in disparate areas where experiments and simulations are difficult to conduct.

The main goal of the thesis is the construction of a surrogate model of water distribution network applying Complex Network Theory principles. In particular, for this research it was analyze the concepts of robustness and resilience to failures from an interdisciplinary point of view. For their natural complexity and ramification, water distribution networks can be easily assimilated to the complex networks present in graph theory. The structural elements of water networks (e.g. valves, pumps, reservoirs, junctions and pipes) are converted into the complex equivalent graph consisting of links and nodes.

In order to investigate a possible correlation between the networks of Water Distribution Network Theory and Complex Network Theory, water simulations and structural evaluations were carried out by calibrating the models according to the heuristics proposed in the literature. The experiments performed evaluate the effect "small world" and other measures for the analysis of networks present in graph theory in relation to the intensity and duration of failures, for example, deficit in the fulfilled water demand, due to insufficient pressure in network. The results showed an inverse correlation between the "small world" property and the failure duration index. In addition, the results obtained show a general tendency to correlation between metrics in both disciplines, both direct and indirect. A further evidence emerged from the analysis is

the double correlation between nodes elevation, failures duration and closeness centrality index. In particular, the importance of the role of nodes elevation is a novelty in the conversion from WDN model to complex model.
This work can therefore help in the identification of more precise parameters suggesting new ways and techniques for an even more reliable building of surrogate models through Complex Network techniques.

Contents

Chapter 1

Introduction

Population increase, urbanization, and changing climate conditions are posing a number of challenges to urban infrastructure systems, including water networks. Water distribution networks design is a fundamental aspect of optimization improvement and strategies management. Non-trivial resources are often required and desired outcomes cannot be easily obtained. Therefore suitable methods to assess the reliability, vulnerability and overall resilience of urban water distribution systems to failure are needed.

Novel techniques based on computer analysis like surrogate modeling and simulation systems have gained great relevance for cost-effective management, sustainable operations and infrastructure maintenance. Simulation-based resilience assessment usually requires a calibrated hydraulic model, which in turn require data available for calibration and computational efforts for simulation run. Theorical data utilization offers a cheaper computationally solution and is becoming a reliable alternative to common analysis strategies. Formulating alternative methods for resilience methods that consider a reduced amount of data and can be used as computationally efficient surrogate models is still an open research challenge.

In the context of modelling water distribution systems, this thesis work integrates new emerging technologies and interdisciplinary modeling approaches. Water distribution networks characterized by different topological structures are evaluated using indices from a relatively young discipline of Complex Network Theory (CNT), a research area based on graph theory and social science. For their non-trivial properties and topologies, water distribution networks can be considered complex infrastructure systems and will be evaluated as networks with numerous interconnected elements. The main focus aims to establish the reliability of adopting CNT indices to assess the resilience of water distribution networks and build surrogate models that are computationally more efficient than traditional hydraulic models. Specifically, this research will try to address the following research questions:

- What are the most suitable criteria to model and analyze a water distribution network as a complex graph?
- Under which conditions could CNT metrics represent a hydraulic network?
- Could CNT reduce the computational effort required for water network resilience assessment?

In the attempt to pursue the above objectives, structural correlation between real networks and benchmark networks have been conducted. This research employs CNT to study one of the central topic in the state-of-the-art literature: building robust and reliable urban water infrastructure systems. In particular, the key aspect of networks resilience against failures was explored. A set of benchmark networks is utilized as support to critical node identification and evaluation of networks vulnerability. Improvements and limitations of described methods will be then pointed out.

The content of this thesis

The thesis is organized according to the following structure:

- In **Chapter 2** background theoretical knowledge of complex network theory and water distribution networks are introduced and an overview of the state-of-the-art literature is provided. Focus is set on the rationale behind evaluations and design choices.
- In **Chapter 3** software implementation and complex and hydraulic analysis techniques are explained.
- In **Chapter 4** the results obtained from methodologies used in the previous chapters are presented and discussed. Input parameters, simulations typologies and final outputs are described.
- In **Chapter 5** the main simulation results pointing out the main contributions of this work and the possible limitations are analyzed and presented. After that, a short conclusion suggesting future research and improvement is provided.

Chapter 2

Background

In this chapter, both complex network theory and water distribution theory will be explored as background for this research. State of the art of literature will be described to better understand the purposes behind the applied methods and techniques and how this interdisciplinary study has been developed. In Figure 2.1 a representative flowchart of step by step research development is given. It synthesizes the time progression of achieved milestones, starting from literature review ending to final results.

Implementation flowchart

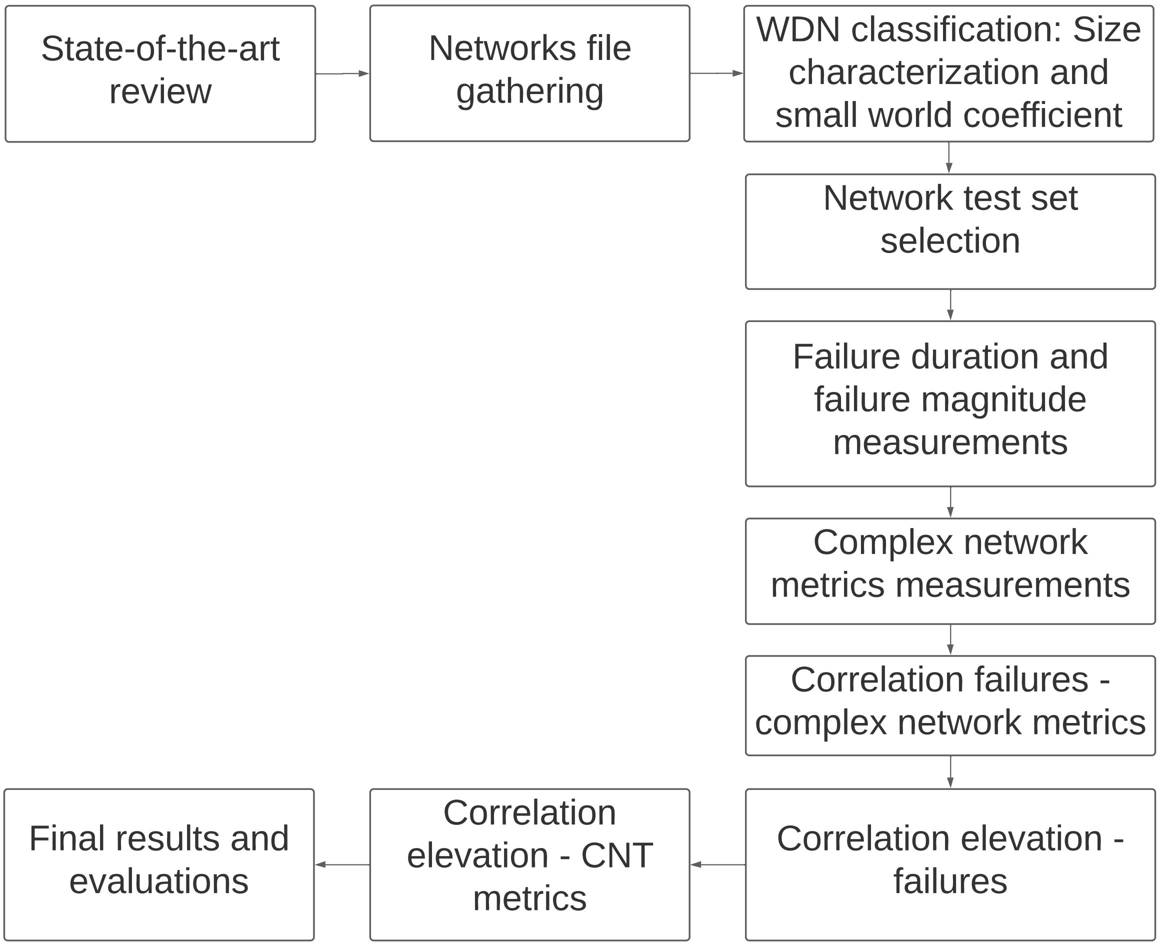


Figure 2.1: *Development flowchart*

2.1 Complex Network Theory

Historically, the study of networks was born as a branch of discrete mathematics known as *graph theory*. The first paper of graph theory was published by Euler (Euler 1736), he proposed the solution to the Seven Bridges of Königsberg problem, where each bridge of the city of Königbsberg should have been traversed *exactly one* time in a round trip. In particular, the term graph was first used by James Sylvester (Sylvester 1878) in a scientific paper: *"Every invariant and co-variant thus becomes expressible by a* *graph precisely identical with a Kekuléan diagram or chemicograph. [...] I give a rule for the geometrical multiplication of graphs, i.e. for constructing a graph to the product of in- or co-variants whose separate graphs are given".* Several years later, the study of graphs found fertile ground in social sciences. This led to the birth of Social Network Analysis (SNA), a branch discipline that explores relationships among social entities in different scenarios (political, institutional, biological, epidemiological, ...).

Unlike traditional methods, SNA focuses on interactions rather than behaviours of single individuals. It examines how the network configuration affects the functioning of individuals, groups, organisations or systems. Moving from SNA, in the last decade a new movement of research called Complex Network has grown up. Between 1998 and 1999, two important papers were published and the *"small-world network"* concept was introduced (Watts & Strogatz 1998; Barabási & Albert 1999). This networks property is closely related to structures resilience and topics covered in this thesis, therefore will be better defined in the following chapter.

Guided also by the increasing computing power and digitalization, these two papers brought the scientific community to invest in research and apply CNT in different fields, including computer networks, biological networks, brain networks and genetic. CNT is usually used for analysing and synthesizing behaviour of real-world infrastructure from a new perspective: human relationships, Internet, urban infrastructure, neural networks and biological system are only a small part of what this discipline aims to study.

The research on complex network started with the identification of new methods and concepts to define networks with common properties, e.g. similar topologies. Starting from this definition, each Water Distribution Network can be seen and analysed as a complex network graph $G = G(N, E)$ where N is the set of the graph nodes and E is the set of graph edges. As the scientific literature suggests (Sitzenfrei, Oberascher, *et al.* 2019), tanks, valves, junctions, pumps, reservoirs and demand nodes can be represented by nodes, and pipes by graph links. The evaluations set like structural patterns and resilience, are based on geometry properties and graph theory applications. In this work the studied networks have been treated both as undirected and as direct weighted graph (Yazdani & Jeffrey 2012). Evaluated metrics for research and modeling purposes are the following:

- **Node degree k**: is the number of edges that are incident to the node and is defined with regard to the adjacency matrix A, a $N \times N$ square matrix whose entry $a_{i,j}$ (i,j =1...N)

$$k_i = \sum_{j \in N} a_{i,j} \tag{2.1}$$

- **Density q**: expresses the ratio of the number of edges and the number of possible edges, thus providing a measure of network redundancy. The density is 0 for a graph without edges, 1 for a complete graph and can be higher than 1 for multigraph. In general, self-loops are counted in the total number of edges, therefore graphs with self-loops can have density higher than 1.

$$q = \frac{2 \cdot e}{n \cdot (n-1)} \tag{2.2}$$

- **Shortest path length (SPL)**: is the length of shortest geodesic path between two nodes, which is the smallest number of edges traversed to reach one node from another.

- **Average shortest path length (APL)**: is the average -shortest- distance between all pairs of nodes n in a network: σ(s,t) is the number of edges along the shortest path connecting node s to node t.

$$APL = \frac{\sum_{\forall s \neq t} \sigma(s,t)}{\frac{1}{2} n \cdot (n-1)} \tag{2.3}$$

- **Edge betweenness centrality (EBC)**: is defined by the number of shortest paths going through an edge, in other words it estimates how often an edge is part of the shortest path between all node pairs. V is the set of nodes, σ(s,t) is the number of the shortest (s,t) path and σ(s,t | e) is the number of those paths passing through edge e.

$$c_{\mathcal{B}}(e) = \sum_{s,t \in V} \frac{\sigma(s,t|e)}{\sigma(s,t)} \tag{2.4}$$

- **Eigenvector centrality (EIG)**: measures the influence of a node in a network: it assigns relative scores to all nodes in the network based on the principle that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes, e.g. for each node takes into account both the number and the quality of its connections.

- **Algebraic connectivity**: of a graph is a spectral metric that measures the network robustness: it is a nonnegative number whose magnitude represents structural robustness against effort to decouple part of the network. The algebraic connectivity may be used as a strong indicator of connectivity at network level which enables the comparison of structural robustness of different network layouts.

- **Clustering coefficient**: measures the probability that the adjacent nodes of a node are connected, in other words is the tendency for link formation between neighboring vertices.

- **Average clustering coefficient**: in a network measures the overall degree of clustering in a network as the average of the local clustering coefficients C_i of all nodes n.
$$C = \frac{1}{n} \sum_{i=n}^{n} C_i \tag{2.5}$$

- **Closeness centrality**: measures how many steps are required to access every other vertex from a given vertex and it is calculated as the reciprocal of the sum of the length ℓ of the shortest paths between the node and all other nodes in the graph.
$$C(i) = \frac{1}{\sum_{i \neq j} l_{ij}} \tag{2.6}$$

2.1.1 Small World phenomenon

An intriguing area in the field of CNT is offered by the famous *"six degrees of separation"* experiment conducted by the social psychologist Stanley Milgram in the 1960s and modelled by authors Watts and Strogatz (Watts & Strogatz 1998). The idea behind this theory is that every single person is just “six degree”, e.g. six other people, away from any other person living on Earth. In Milgram experiment, a group of people from the cities of Omaha in Nebraska and Wichita in Kansas must deliver a letter to a target/receiver, a complete stranger for them, in Boston (Massachusetts) knowing only her/his name and the final city destination. They could only send these letters to persons, used at intermediaries, they know personally. When the letters finally reached the destinations, Milgram count the number of times the letters had been forwarded. The average path length he calculated was in the range of 5 or 6 intermediaries. This hypothesis introduced for the first time the concept of a small and well-connected world called “**small-world phenomenon**”.
Starting from Milgram’s discovery, Watts and Strogatz (Watts & Strogatz 1998) applied a rewiring link process to a ring lattice with the same number of edges, increasing the level of randomness from 0 (regular lattice) to 1 (random graph), as Figure 2.2 shows. Afterwards, they calculated the characteristic path length and clustering coefficient and concluded that a small world is as a "*highly clustered network with a small average path length*" where most nodes are not neighbours of one another but can be reach from every other node by six steps.

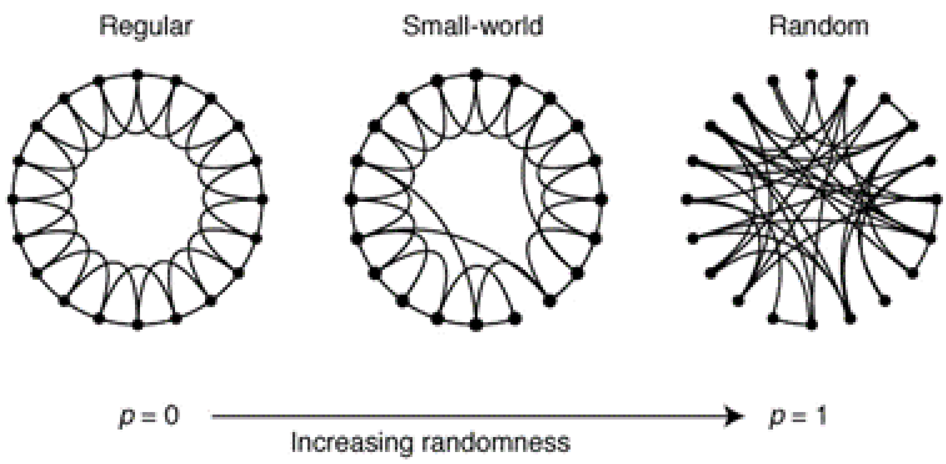


Figure 2.2: *Rewiring process of Watts and Strogatz (Watts & Strogatz 1998)*

This phenomenon was taken in consideration in this work due to its possible implications with system dynamic. Small world peculiar structure spreads the information (or generally, everything concerning transportation or connection level) along the network faster than regular or random networks and since small-world network usually have almost equal degree for all the nodes, it is robust to targeted attack (Bharali & Doley 2017).

To evaluate if a network can be considered as a small world, a mathematical comparison with a graph of the same size and random generated links among nodes should be conduct. Small world graphs, compared with their related random graphs, tend to have smaller APL and higher clustering coefficient. The following equation 2.7 expresses the small world property introducing *the small world coefficient* σ. It is measured with **cc** – clustering coefficient of considered network, **L** – average shortest path length, $\mathbf{cc_{RC}}$ – clustering coefficient of related random graph and $\mathbf{L_{RG}}$ – average path length of related random graph.

$$\sigma = \frac{cc/cc_{RG}}{L/L_{RG}} \tag{2.7}$$

A network can be classified as a small world when σ is significantly higher than 1 (Ferretti 2018).

Small world property could also be assessed through a visual representation of topol-

ogy and nodes degree distribution. The topology of a small world network, is characterized by a large number of nodes clusters where nodes are highly interconnected in subsets. Particular configuration of the degree distribution makes small world network distinguishable from equivalent random ones (o random network?).
In the next sections a description of Water Distribution Network Theory will be given, in order to better understand how both disciplines can be bound together.

2.2 Water Distribution Network basic theory

Water Distribution Networks are large and complex infrastructure system. Their structure is characterized by interconnected physical components (pipes, tanks, reservoirs, junctions, pumps) providing water supply to meet the required demands with high-quality water. The distribution network branches out towards each connection node (Loucks & van Beek 2017). Its non-trivial spatial organization and topology involve deep investigation and management in different fields: from infrastructural solution to water demand optimization.
In recent decades the problem of management and maintenance of WDNs has become more and more investigated, due to the interactions among several variables, including water resources, water demands, urbanization and climate factors. WDNs should ensure a constant water supply and, in large urban areas, water demand could vary in different periods, seasons or day. The challenge of a reliable distribution maintaining the level of failures as low as possible, is a challenge task that requires an effective and strong urban water resources planning strategy. The need of developing systems that are more adaptive and secure led researchers to the identification of the main causes for failures, including water leakage, blockages and general malfunctioning. For this reason, this research is focused on all kinds of unforeseen but frequently failures and how these failures impact on systems.
Particular attention was given to modelling and quantification of the impact of network failures, in terms of time to recovery from damage (resilience) and deficit magnitude (vulnerability). From an efficiency point of view, it is more desirable to limit the

severity of failures instead of pursuing the unrealistic goal of a complete deficit elimination (Fortunato 2012). This could lead to less episodes of failures but with higher intensity. Together, failure duration and failure magnitude indices provide support in water sizing, network rehabilitation, asset management and general planning and management strategies.

2.3 Exploring synergies between complex networks theory and water distribution networks resilience assessment

Before starting to make assumptions about a possible correlation between CNT and WDN, related scientific literature has been explored. At first glance, the focus of the research aimed to establish the validity of adopting CNT as surrogate models of water distribution networks for hydraulic resilience investigations. The main questions research are:

- Is there any direct correlation between these two disciplines?
- Is it possible to model and analyse a water transportation network as a complex graph?
- Under which conditions could a CN metric represent a hydraulic element?
- Could CNT reduce the computational work behind WDN assessment?

The theory behind this work is still explorative and the initial hypothesis are based on a reduced number of studies. Moreover, the research conducted in the examined papers generally referred to a specific application or use case since an overall shared point of view is missing.

In the majority of papers, authors try to investigate one of the main challenges of nowadays water systems: creating scalable but efficient infrastructure for water distribution (Yazdani & Jeffrey 2012; Giustolisi *et al.* 2017; Klise *et al.* 2018; Sitzenfrei, Oberascher, *et al.* 2019; Sitzenfrei, Wang, *et al.* 2020). A large number of studies on the potential of graph theory has been carried out, however additional research is needed in this direction, since effectiveness of surrogate complex models may vary significantly according to operating conditions and networks structure. Therefore, complex network adaptive hypothesis were evaluated in limited and specific scenarios or networks, since a unifying principle for defying a general behaviour of all range of WDNs has not been discovered yet. Various approaches (Yazdani & Jeffrey 2012; Di Nardo, Di Natale, *et al.* 2018; Zhan *et al.* 2020) have been put forward to find the optimal functioning of water distribution system and resiliency to vulnerability. For this reason, many studies tried to give a mathematical definition of efficiency through the concept of resilience. Since reliability and robustness against failures are the main responsible for general resilience, they have been examined in different scenario and from diverse perspectives.

In order to support resilience assessment from a CNT point of view, limits and potential of these methodologies were tested (Pagano *et al.* 2019). Different definitions of resilience can be found in literature, Laprie (Laprie 2005) and Strigini (Strigini 2012) affirm that *"Resilience is a typical feature of network systems where the redundancy in links, paths and loops allows systems to maintain and adapt their operational performance when some of their components fail or other unplanned and adverse conditions arise"*.

In WDN, resilience is linked to the network geometry and topology based on the premise that a densely looped layout offers system redundancies which are capable of overcoming operational perturbations (Mays 2000). However, a high topological redundancy in water infrastructure does not necessarily imply a better resilience. A trade-off between efficiency and redundancy has always to be taken in consideration. For example, redundancy paths or larger pipes diameter than those strictly needed, imply more water to convey (Di Nardo, Di Natale, *et al.* 2018). As consequence, a network should be redundant but efficient at the same time (nodes reachability and fulfilled demand at 75/80%). Concurrently with the identification of the best solutions for strong and lasting resilience, main causes of failures (e.g. critical components) are investigated. Studies also suggest that under certain condition, a direct correlation between hydraulic and complex metrics could be observed. For example, in a previous study of Di Nardo (Di Nardo, Di Natale, *et al.* 2018), a similar behaviour between complex metric "average shortest path length" and Todini resilience index (Todini 2000) was found. In (Sitzenfrei, Oberascher, *et al.* 2019) a direct correlation between water age and "shortest path length" was discovered. In other words, although graph techniques analysis provided only a limited overview, existing analogy between graph nodes and edges and water distribution network elements have been found (Yazdani & Jeffrey 2012).

2.3.1 Indices of resilience in water distribution networks

Resilience and system performance are generally identified by resilience index proposed by Todini (Todini 2000) or by Prasad and Park (Prasad & Park 2004). According to Todini, *"the capability of the designed system to react and to overcome stress conditions"* and it measures the available surplus of energy, Prasad and Park (Prasad & Park 2004) version of resilience index incorporate also reliable loops (nodal uniformity: majority of pipes linked to a node have the same diameter).
Besides resilience index, other two metrics, taken from Zhan *et al.* 2020, are assessed for describing system performance: **failure duration** and **failure magnitude**. They are widely used to identify nodes which are more susceptible to failure or weakness. Usually, the definition of failure rates how long a node actual demand is below the expected demand (failure duration) and how intense this shortage is (failure magnitude). These two failure metrics are often correlated.

Failure duration

Failure duration indicates how long a node is under failure and how long it takes to recover (Zhan *et al.* 2020), it determines if the state of a node is satisfactory (the pressure at that node is equal or higher than the minimum requested one and water demand is fully supplied), or unsatisfactory, e.g. the node pressure is under the minimum pressure threshold and the water demand cannot be fulfilled. Failure duration is a node related measure and indicates the average duration time of failure occurred to a specific node during simulation time:

$$T_{F,i} = \frac{\sum_{j=1}^{M_i} T_{i,j}}{M_i} \tag{2.8}$$

$\mathbf{T_{F,i}}$ = average duration of failure occurred at the i-th node during the simulation period

$\mathbf{T_{i,j}}$ = duration of the j-th failure event at the ith node

$\mathbf{M_i}$ = total number of failure episodes at the i-th node during the simulation.

Failure magnitude

Failure magnitude at nodes evaluates how bad the water shortage is, it is expressed as the percentage of unmet water demand. In reviewed literature (Zhan *et al.* 2020), failure magnitude was calculated at every time-step and for every node (equation 2.9) but in this research, in order to obtain a complex network comparable metric system, results were aggregate over all time-steps (equation 2.10).

From:

$$v_{i,t} = \frac{D_{i,t} - SW_{i,t}}{\sum_{i=1}^{n} D_{i,t}} \tag{2.9}$$

To:

$$v_i = \frac{\sum_{i=0}^{t} D_i - \sum_{i=0}^{t} SW_i}{\sum_{i=0}^{n} \sum_{i=0}^{t} D_{i,t}} \tag{2.10}$$

$\mathbf{D_i}$ = nodal water supply at node i-th

$\mathbf{SW_i}$ = water supply to node i-th

2.3.2 On the choice of network nodes weights

A key problem with much of the literature in relation to resilience studies is how to find the best corresponding complex network metric. Yazdani and Jeffrey (Yazdani

& Jeffrey 2012) demonstrated that the best solution for modelling a WDN is using a *weighted* and *direct* graph, since unweighted graph cannot guarantee enough accuracy due to the big heterogeneity of the possible relationship between nodes. In particular, link weights indicate the intensity of the relationship between two linked nodes, which is usually defined with hydraulic or geometric features (e.g diameter of pipes, length of pipes, water flow, node elevation) and the graph direction is defined by flow direction.

Di Nardo, Di Natale, *et al.* 2018 presented a use case where the behaviour of complex metric "shortest path length" with the inverse of pipes diameter as weight could be reconducted to the resilience index proposed by Todini. This theory was evaluated in similar scientific research, great importance was given to shortest water path criteria for best design solution (Di Nardo, Di Natale, *et al.* 2018; Sitzenfrei, Oberascher, *et al.* 2019).

Related observations carried out by Sitzenfrei et al. (Sitzenfrei, Oberascher, *et al.* 2019) seem to suggest a linear correlation between edge between centrality and low resilience network. After investigating different weighting function and finding the more appropriate one, this correlation could also be increased. Moreover, the highest level of correlation was reached with only geometrical-based properties as weight. This discover introduced another (branch) key point of scientific research: establish if a graph built without the need of hydraulic simulated metrics, in other words, based only on the topological elements of water network structure, could provide a reliable surrogate model. A surrogate model is a computationally efficient simplified model (often data-driven) of the complete model (often physically-based) of the system (Fiedler *et al.* 2020).

Chapter 3

Materials and Methods

A comparison between complex theory and water network distribution modelling has been conducted analyzing detailed topological features of a group of 30 benchmark networks, open-source and free distributed. Most of them are prototypes, developed for research or academical purpose. Since application of CNT in water fields is novel and related studies are still ongoing, the aim of this study has been outlined during its development and depended by the output results of each single investigation step. A challenging problem which arises in this domain is the management of sensible data. This crucial point limits the availability of WDNs due to privacy and security reasons, therefore the majority of networks data in peer-reviewed papers were generally not completely disclosed or only the final results were shown.
In the first part of this study, the networks were classified through visual inspection and size characterization. WDNs were then categorized both with complex network and hydraulic indices and on that base a 5 representative networks were selected as test set for the last part of experiments. Different simulation models were used. Final comparisons among results were evaluated in order to find a possible correlation. Numerical evaluation of this section will be reported in Chapter 4.

3.1 WDNs classification

Network classification also played a crucial role to assess WDNs vulnerability. From previously scientific paper, it is known that network structure is strongly constrained by spatial characteristics and shows a large heterogeneity (Giustolisi *et al.* 2017). Consequently, before considering WDNs as complex system, an overall hydraulic classification was carried out. Benchmark networks were classified in terms of: n° of nodes, n° edges, n° reservoirs, n° tanks, n° pumps, n° valves, n° junctions, n° pipes and size of the network (total pipes length).

This primary classification helped understand the main characteristics of the networks group and will be helpful in forthcoming correlation research. Other observations in terms of special elements type (tanks, reservoirs, valves, etc) were useful for determining network complexity and better understand the kind of vulnerability it faced during simulations. In the literature, networks with different level of complexity are threated and studied with different methods.

3.1.1 Visual inspection of water network topology

In addition to numerical classification in schematic table, a graphic plot of network topology was useful for a broader perspective. Networks visualization always reveals aspects that usually tend to be hidden or can be only visually discover, e.g. graphics generated considerable interest in nodes elevation, a key aspect in WDNs that was not taken in consideration before.

For the best evaluation method, complexity level and topological structure must be considered together. Unfortunately, plotting tool used in this thesis (Water Network Tool for Resilience - WNTR) only allows users to plot networks with the whole set of nodes or links names together, partial or single nomenclatures are not possible.

On one hand, for medium/large networks (bigger than 100 nodes) this implied that a standard desktop image resolution was not wide enough to provide a clear vision of their basic element representation (nodes and links name) together with network structure in one image. On the other hand, since nodes or links name may vary depending on specific network .INP file and there is not a general principle for naming, they are generally considered not relevant for evaluation purpose. Not surprisingly, scientific literature is more focused on node/link type rather than its name. In many networks, pumps, valves, etc. are labeled according to numbers or a mix of numbers and letters. For this reason, randomness nature of adopted nomenclature and the structure of plotting function make elements name unsuitable for a type classification. Other solution should be adopted.

To overcome the previous graphic problem, an overriden version of the WNTR plot function was implemented. In the original form "node_labels" and "link_labels" can only be set as true or false value; it is not permitted to show a specific property of a node but only its name. The new function overrode the structure of labels attribute, changing its input type from boolean to a more complex structure. After this change it was possible to specify which node and which node properties should be displayed on the graph and not only whether or not display the elements name. Two lists of entity type were created, in order to locate the exact position of tanks, reservoirs, valves and pumps in the network. Tanks and reservoirs type constituted the new node input list and were highlighted respectively by the letter **T** and a blue node color, and by the letter **R** and a yellow node color. Valves and pumps type constituted the new edge input list. In this operation junctions were not considered and the correspondence nodes were left blank.

Another adjustment was applied to name position. Since in WNTR labels position corresponds with elements position, if displayed all at the same time, it generated an overlapping problem, as it is shown in Figure 3.1. For this reason, a coordinate resize was needed and elements names have been shifted upwards (from the original location). Another problem is that position unit of measurement and length scale are not standard, each network behaviour is different, therefore the label re-positioning implied an operation for every network calibrated on its specific structure.

In the following images, a comparison between C-town network plotted with original WNTR plot function (Figure 3.1) and the new overridden version (Figure 3.2) is given.

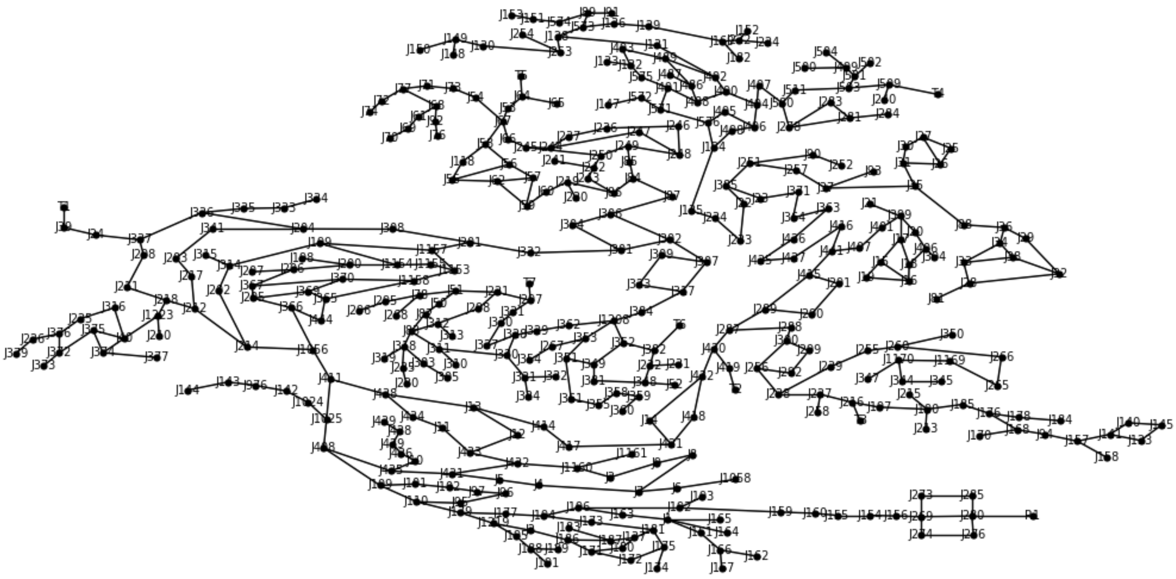


***Figure 3.1:** C-Town plotted with original plot function.*

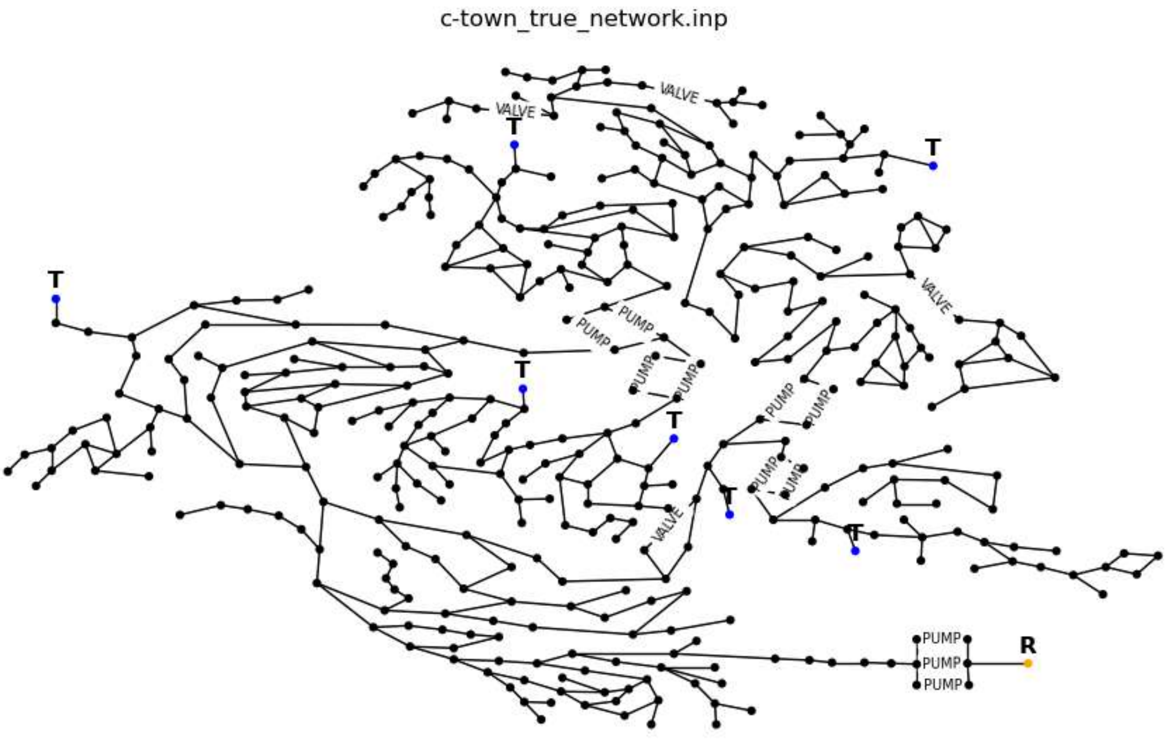


***Figure 3.2:** C-town plotted with overridden function.*

The main advantage introduced by this new function is a clearer vision of key elements (tanks, valves, pumps) positions.

3.1.2 WDNs characterisation

Number of nodes, reported as size, and total pipes length, reported as network extension, are two key aspects of WDNs. These two metrics are often linked together or considered as equivalent, but it is not necessarily true that networks with the highest number of nodes are also the most extended ones.
Before moving to further investigations, a correlation between size and extension was sought after. A common technique to examine this last observation, involves the scatter plot method. Using matplotlib tools, a scatter plot with total network length on the x-axes and number of nodes on the y-axes was created. Each point on the reslting graph corresponds to a benchmark network.
Another strategy for assessing similitude between two entities is provided by the Empirical Cumulative Distribution Function (ECDF). The ECDF orders a particular feature from least to greatest value distributed across the data set and allows a percentiles division.
In this research, nodes number ECDF and pipes length ECDF were used for splitting the 30 test networks into four subgroups, respectively small, medium, large and extra large, based on the value of 25-th/50-th/75-th percentile. Both distribution outlines were then analyzed and (the presence of) a possible similitude was investigated.

3.2 WDNs analysis with complex network indices

After the assessment of the hydraulic properties, complex network metrics were investigated. Each network was individually imported and elaborated through WNTR (Klise *et al.* 2018) library functions based on NetworkX (Hagberg *et al.* 2008). The interoperability between these two software guarantee an efficient transformation of water network models into a complex network graph.
Firstly, a **WaterNetworkModel - wn** was created from the .INP file, subsequently

this model has been transformed into a complex multidigraph with library function `get_graph()`. This function creates a graph based on attributes input (*node_weight, link_weight* and *modify_direction*) and returns a network **MultiDiGraph**. A multidigraph is a direct graph that can store multiple edges between two nodes.
In this work great importance was given to node and link weight since there are key element in this research, as previously explained in previous section 2.3.2. Complex graph was created with **nodes elevation** as node weight and **pipes length** as link weight. This decision was taken after reviewing related papers of important authors, but as previously stated, a general definition of best weight to assign is still *lacking*. Among all the proposed metrics for weight system, flowrate and pipes length were the most common. Since flowrate required a preliminary simulation, pipe length was chosen because is provided by design with the network file and does not have hydraulic dependence.
Investigated complex network metrics has been chosen among all the metrics used in scientific literature. Each one of them captures a specific property of a WDN and is related to networks tolerance to failure. They also help evaluating different networks design.
Not all the complex metrics are compatible with the previously mentioned MultiDiGraph generated by WNTR. Therefore a cast from MultiDiGraph to simple graph and undirect graph with NetworkX `Graph()` and `to_undirect()` functions was needed. Evaluated metrics with conversion from multigraph to simple graph are: betweenness centrality, eigenvector centrality and clustering coefficient. Metrics evaluated with multidigraph are: node degree, closeness centrality, density and bridge ratio. Metrics with undirect multigraph are: average shortest path length and algebraic connectivity. Interpretation of obtained measurement value largely depends on the network layout and design considerations. All of the metrics were calculated using NetworkX library and interpreted from an efficiency and robustness point of view. In particular the adopted functions were:

- **Density**: The link density is measured as ratio between the effective number of edges and the maximum possible number of links that could be present in a graph. It states the dense-connectivity and redundancy level of the network.

- **Clustering coefficient**: useful indicator of alternative path (path redundancy) in case of failure. It states the density of triangle in a network. Unfortunately, as stated authors Yazdani and Jeffrey (Yazdani & Jeffrey 2012) explained, in wide network cluster have not a triangular form. For this reason, in such networks, clustering coefficient could not provide a reliable indicator.

- **Eigenvector centrality**: measurement of the influence of a node in a network. It is assessed calculating the number of connections a node has with other highly connected node. Eigenvector centrality is slightly different from betweenness centrality because it is evaluated considering all the node path and not only the shortest one.

- **Closeness centrality**: reveals those nodes which are faster in spreading the information throughout the network. Nodes with high closeness centrality are the closest to all other nodes, in other words, they have short "shortest path length".

- **Average clustering coefficient**: The average clustering coefficient is the average of the overall clustering coefficient of nodes. This metrics was calculated in support of small-world coefficient.

- **Algebraic Connectivity**: second smallest eigenvalue of the normalized Laplacian matrix of network. In this contest is used for assessing robustness and connectedness of the network. High value of algebraic connectivity represents high robustness to network decoupling (Yazdani & Jeffrey 2012).

- **Betweenness centrality**: Larger values of betweenness centrality indicate that a node is located on many short paths.

- **Node degree**: Node degree represents the number of connections of a single node. It has been included in this research for a wider perspective even if authors Yazdany and Jeffrey (Yazdani & Jeffrey 2012) discovered that it may not be sufficiently detailed for describing network characteristics.

- **Bridge ratio**: The bridge ratio or density of bridge is a robustness metrics. It expresses the fraction between links which removal leads to graph disconnection and the total number of links. A small value of bridge ratio indicates strong robustness (Yazdani & Jeffrey 2012; Di Nardo, Di Natale, *et al.* 2018).

Small world coefficient

Another way to couple with structural efficiency in a complex graph is to investigate the **small world effect**, previously evaluated in section 2.1.1. Networks that show a small world structure are considered locally and globally efficient (Watts & Strogatz 1998), therefore small world coefficient sigma have been measured for all the benchmark networks in relation with their other vulnerability values. Despite its great implication in networks connectivity structure, only a few studies investigated this particular graph layout in water distribution networks.
Small world coefficients were calculated using the sigma function from the small world library of NetworkX. This function returns a float type sigma coefficient of the given graph and as optional attributes takes the number of rewiring per edge for the equivalent random graph - *niter*, number of equivalent random graphs - *nrand* and random number generation state – *seed*. Since an accurate definition of small world attribute function is still missing and no practical example can be found, different combination of niter, nrand and seed were given as input (Table 3.1).

Niter - nrand - seed					
1 – 10 – 1	1 – 10 – 2	1 – 10 – 3	4 – 10 – 1	4 – 10 – 1	1 – 20 – 100
1 – 30 – 1	1 – 20 – 2	1 – 20 – 3	5 – 10 – 1	5 – 10 – 1	1 – 10 – 80
1 – 50 – 1	1 – 30 – 2	1 – 10 – 5	50 – 10 – 1	50 – 10 – 1	1 – 10 – 100
1 – 70 – 1	1 – 50 – 2	1 – 10 – 6	3 – 10 – 1	3 – 10 – 1	1 – 10 – 70
1 – 100 – 1	1 – 70 – 2	1 – 10 - 30	2 – 10 – 1	2 – 10 – 1	1 – 10 – 50

Table 3.1: *Niter, nrand and seed combination values for σ coefficient*

Sigma coefficients were calculated under the initial assumption that higher value of niter/nrand implies better accuracy in value estimation. High value of parameters implied long simulations, with numbers greater than 20 it took up to 9 hours to compute the final result on a machines with 8 GB RAM machine and Intel(R) Core(TM) i3-6006U processor. At the end of this operation, all the sigma values were stored in tables and subsequently analyzed. Collected data highlight that there is no direct implication between high value of parameters and precision.
Final small world coefficient of each network was obtained as the median of all the sigma. Median function was chosen upon mean function as a matter of accuracy. Sigma value also played a key role in final network choice. A group of five network was chosen from the original benchmark group, based on the hydraulic and complex evaluations that have been done. Selected networks were different in topology, complexity and size. The reason of this heterogeneity is to be found in coming failures investigation; different network layouts could provide results from different points of view and a more accurate answer to the main hypothesis.
As previously stated, not all the .INP file of the benchark networks interoperated with NetworkX sigma function, therefore only sigma of working networks were calculated.

3.3 Simulation-based WDN resilience assessment

Previous study of WDNs in a resilience context using metrics borrowed from complex complex network theory, already discovered important topologic similitudes. As the total water demand might not be completely satisfied when a failure in the WDN occurs, failures are better represented by a pressure-demand dependence and a pressure driven analysis (PDD). Therefore in this work a PDD simulation model has been used according to the following pressure-demand relationship (3.1):

$$d = \begin{cases} 0 & p \leq P_0 \\ D_f(\frac{p-P_0}{P_f-P_0})^{1/e} & P_0 \leq p \leq P_f \\ D_f & p \geq P_f \end{cases} \tag{3.1}$$

Figure 3.3: *Pressure-demand relationship, **d** is the actual demand (m^3/s), $\mathbf{D_f}$ is the desired demand (m^3/s), **p** is the pressure (m), P_f is the required pressure (m), P_0 is the minimum pressure and 1/e is the pressure exponent, set equal to 0.5*

Pressure driven analysis requires a **minimum pressure** P_0, the pressure below which the consumer cannot receive any water and a **requested pressure** P_f, the pressure above which the consumer should receive the desired demand. Minimum pressure was set to default value 0 (under this threshold there is no water flow in that node and also the water demand is considered 0). On the other hand, the requested pressures are not fixed values, they change depending on the examined network. Therefore, the experiment proceeded with a DD pressure simulation, this ensure that obtained value represent the maximum pressure each node could ever undergo. In WDNs, different nodes pressures can be found. Since this parameter varied from one another, an ECDF plot of general pressure was computed and the 10-th/50-th/90-th percentiles were taken as input value for requested pressure in the following pressure driven analysis.

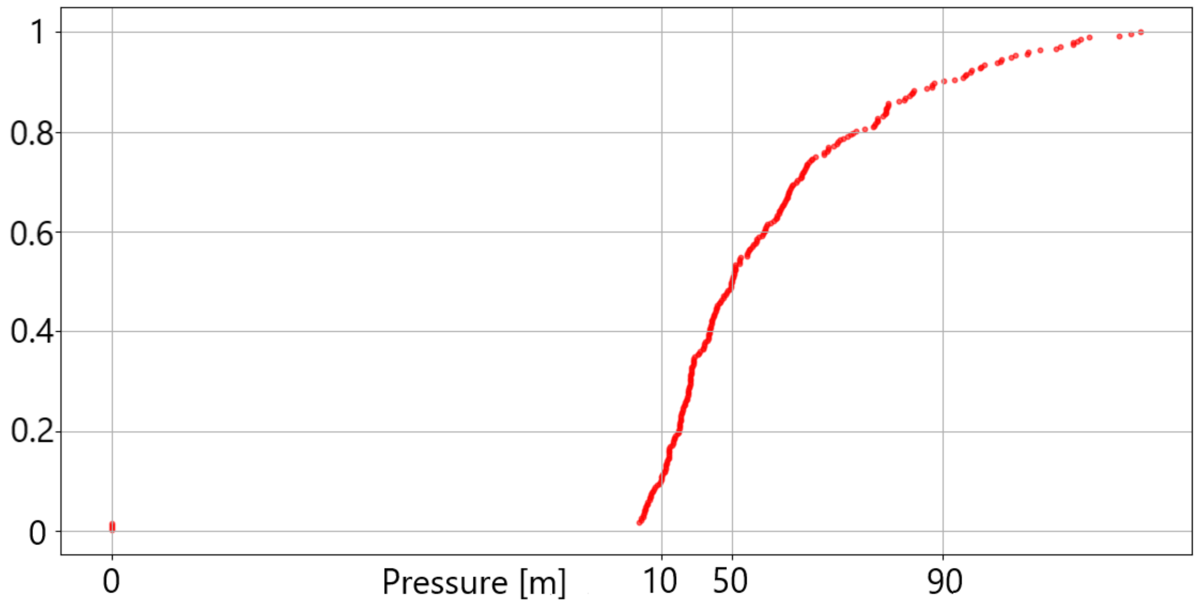


Figure 3.4: *Example: ECDF pressure plot*

At 10-th percentile only a small percentage of nodes is under requested pressure level, on the contrary at 90-th percentile the network is undergoing a severe pressure shortage. Results will be provided in Chapter 4.

3.3.1 Pressure-driven and demand-driven simulations

Simulations were executed through WNTRSimulator function. This method takes a water network model created from .INP file and the demand model (PDD or DD) as attribute input. If no model is directly specified, DD model is used as default. Other simulation parameters that WNTR allows the user to modify are time, hydraulics option, solver, results, graphic and energy. Simulation duration in this work is set to *24*7*3600 seconds* (1 week), the hydraulic timestep of simulation evolution and the report timestep are set to *3600 seconds* (1 hour).
As previously reported in Section 3.3, PDD simulations were carried out setting junctions minimum and requested pressure. For each network three simulations per percentile were conducted. Simulations results are stored in a variable, normally labelled "*results*"; WNTR provides a dedicated primitive type named "*SimulationResults*". This

variable contains simulated parameters like flowrate, pressure, demand, flowrate or more generally node and link attributes.
They then can be stored as Pandas DataFrame in single variables, the structure of DataFrame is similar to a table organized in rows and columns. Each column is given a node or link name and indicates the evolution over time of the selected parameter. Rows labels contain the timesteps and represent the time unit of performing operation, in this case each row correspond to 1 hour of simulation.

3.3.2 Failure magnitude and failure duration indices

This part of the research introduces a novelty, proposing a yet unexplored method of network resilience assessment. In previous studies evaluation of networks failures were conducted quite only from a hydraulic perspective, only a few works in literature applied a mixed approach, based both on hydraulic and complex theory. In any case, failure duration and failure magnitude were rarely analyzed.
Both failures metrics were calculated following the reinterpreted formula explained in previously theory Section 2.3.1. It is a reinterpretation of the original version proposed by authors Zhan, Meng, Liu and Fu in their paper (Zhan *et al.* 2020). Both metrics were evaluated removing time dependency: each node was characterized by a single failure value.

3.3.3 Overall resilience assessment

Concerning the code side, the WNTR function "*expected_demand*" was used jointly with "*demand*" parameter, considered as "*actual_demand*". All the average duration of failure and percentage of magnitude, occurred during the simulation period, were collected and stored into a Pandas Series variable. Although these metrics are normally referred to a single node, an overall mean value was given for an experimental comparison with other globally referred network metrics (e.g. density, algebraic connectivity, etc. . .).
Both reported images of failures are a simplified version of used networks:

```
Failure duration:
10     0.000000
11     3.714286
12     1.666667
13     1.000000
21     2.666667
22     1.000000
23     0.000000
31     6.375000
32    13.125000
```

(a) *Example failure duration structure: Node label and duration*

```
Failure magnitude:
10    0.000000e+00
11    2.427116e-04
12    1.777434e-05
13    7.046825e-07
21    7.109417e-05
22    1.446947e-05
23    0.000000e+00
31    3.155391e-04
32    1.466705e-03
```

(b) *Example failure magnitude structure: Node label and magnitude percentage*

3.4 Hydraulic resilience vs complex network indices: comparison

Last experiment of correlation investigation involved both metrics from WDNs theory and complex network theory. Before verifying similitude or direct correlation, all the metrics were cast to compatible types. This operation was carried out using Pandas data type like DataFrame or Series.

Mean values of failures were compared with complex metrics like average path length, small world coefficient sigma, network density, algebraic connectivity, average clustering coefficient and bridge ratio. These are all metrics related to the whole network. Secondly, single failure metrics together with nodes related metrics were compared with between centrality, degree, closeness centrality, clustering coefficient, eigenvector centrality.

Since failures measurements were evaluated only on junctions, different structural element like tanks, valves, etc, have been removed from complex network metrics. Direct correlation was carried out both visually with a scatter plot and numerically using correlation function based on Pearson method (Benesty *et al.* 2009). In addition to scatter plot, network structure showing different nodes feature was further provided.

3.5 Software implementation

In this section all the software applications and programs used to analyze data and conduct the research explained in previous sections will be reported. Technology implementation involved both complex and hydraulic techniques.

Water Network Tool for Resilience (WNTR)

In this thesis work, **Water Network Tool for Resilience** (WNTR, pronounced *winter*) (Klise *et al.* 2018) was used for evaluating Water Distribution Network properties in correlation with complex network theory. WNTR is a relatively new open-source software, Python-based and provided by United States Environmental Protection Agency (US EPA).
WNTR is based upon EPANET, a software application with graphical user interface for visualizing and modelling water distribution network without the need of programming skills. WNTR represent an improvement for WDNs analyses, it extends EPANET features and builds a bridge between water management and computer-based methodologies. In addition to WNTR ability to transform water distribution network in complex network, the included NetworkX extra feature makes it an indispensable tool for this research. The software can be installed from GitHub repository webpage and a 64-bit Python version is required along with several Python package dependencies including Pandas, Numpy, Scipy and Matplotlib. These packages allow the user to perform Python analysis and analyze results with scientific tool or high-quality graphics. Citing from the description on the website: "The Water Network Tool for Resilience (WNTR, pronounced *winter*) is a Python package designed to simulate and analyse resilience of water distribution networks. Here, a network refers to the collection of pipes, pumps, valves, junctions, tanks, and reservoirs that make up a water distribution system. WNTR has an application programming interface (API) that is flexible and allows for changes to the network structure and operations, along with simulation of disruptive incidents and recovery actions. An integrated development environment (IDE) is recommended for users and developers."

WNTR allows user to simulate disaster scenario as earthquake, pipe breaks or leaks, power outage, fires, environmental change, contamination, flood, damage to infrastructure and more. The WNTRSimulator does not support water quality simulation, however, through the integration of EPANET Simulator, it is possible to investigate not only technical but also water quality aspects such as water age or chemical contamination, propagating toxic chemicals through the network. In Figure 3.6, a flowchart of WNTR and EPANET functioning is presented.

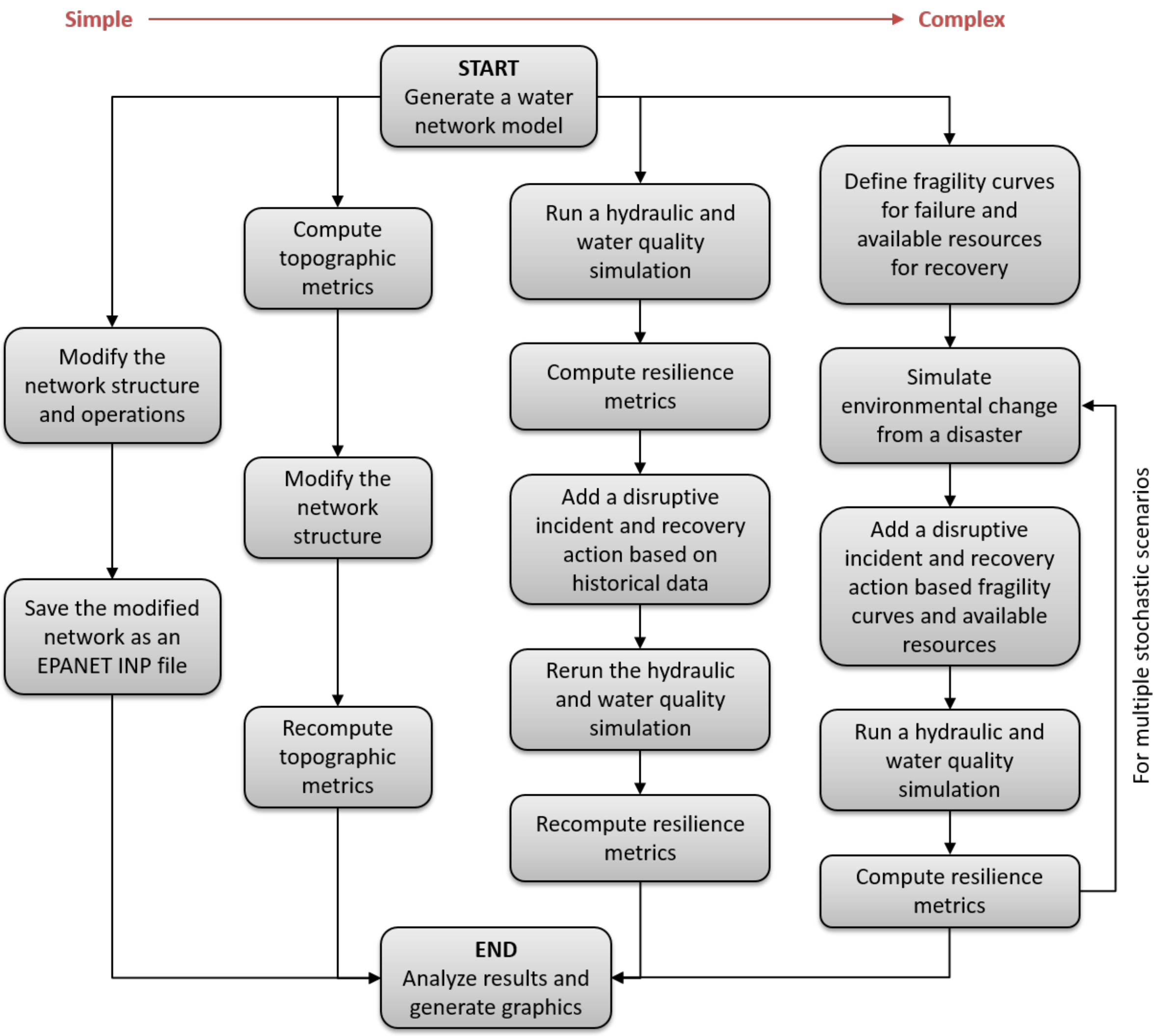


Figure 3.6: Water Network Simulator (Klise et al. *2018)*

WNTR gives the possibility both to create a new water network or to import an existing one. In this work, input benchmark networks have been used as test networks.

They were provided as an EPANET Input File Format: **.INP file**, presented in Figure 3.7

```
[TITLE]
EPANET TUTORIAL

[JUNCTIONS]
;ID   Elev   Demand
;------------------
2     0      0
3     710    650
4     700    150
5     695    200
6     700    150

[RESERVOIRS]
;ID   Head
;---------
1     700

[TANKS]
;ID  Elev  InitLvl  MinLvl  MaxLvl  Diam  Volume
;-----------------------------------------------
7    850   5        0       15      70    0

[PIPES]
;ID  Node1  Node2  Length  Diam  Roughness
;-----------------------------------------
1    2      3      3000    12    100
2    3      6      5000    12    100
3    3      4      5000    8     100
4    4      5      5000    8     100
5    5      6      5000    8     100
6    6      7      7000    10    100
```

Figure 3.7: *Structure of INP File (Klise* et al. *2018)*

An .INP file defines each single entity of network structure. It contains different sections, for example:

- Network components: *[TITLE], [JUNCTIONS], [RESERVOIRS], [TANKS], [PIPES], [PUMPS], [VALVES]* or *[EMITTERS]*
- System Operation: *[CURVES], [PATTERNS], [ENERGY], [STATUS], [CONTROLS], [RULES], [DEMANDS]*
- Water Quality: *[QUALITY], [REACTIONS], [SOURCES], [MIXING]*

- Options: *[OPTIONS], [TIMES], [REPORT]*
- Network map/tags: *[COORDINATES], [VERTICES], [LABELS], [BACKDROP], [TAGS]*

Depending on the element type, every physical component is defined by structural properties like elevation, demand, diameter, head, roughness, level and more. These parameters will then be used in water simulations and transformed in CNT metrics for correlation analysis.

Demand driven and pressure driven analysis in WNTR

Another important feature introduced in WNTR with the v0.3.0 release (November 2, 2020) is the option to switch from EPANET 2.00.12 engine which runs only **demand driven analysis** (DD or DDA) to 2.2.0 version implementing **pressure driven analysis** (PDD or PDA).
As suggested in Figure 3.8, in a demand driven simulation, the water demand does not depend on system pressure and an always-fulfilled expected demand is assumed. This hypothesis also implies that nodal piezometric elevations always satisfy the requested water demand. Those models produce reliable results only when this last assumption is always verified and the water network is operating under normal condition. The need of more realistic simulations, based on effective water demand calculations, led to the development of a pressure-driven model.
In pressure driven model the water demand is satisfied only if pressure at nodes reach the requested pressure threshold. PDD model is based on the same demand driven mathematical model plus an extra equation linking flow rate and water loss to nodes hydraulic weight. DD is not suitable for a failure analysis of a water network distribution, therefore is more useful to adopt a pressure driven methodology.
With the latest WNTR release introducing the new PDD feature, it has been possible to run more realistic and accurate simulations, recreating low pressure and failures demand scenarios. In this work both methodologies were used for evaluating the initial hypothesis.

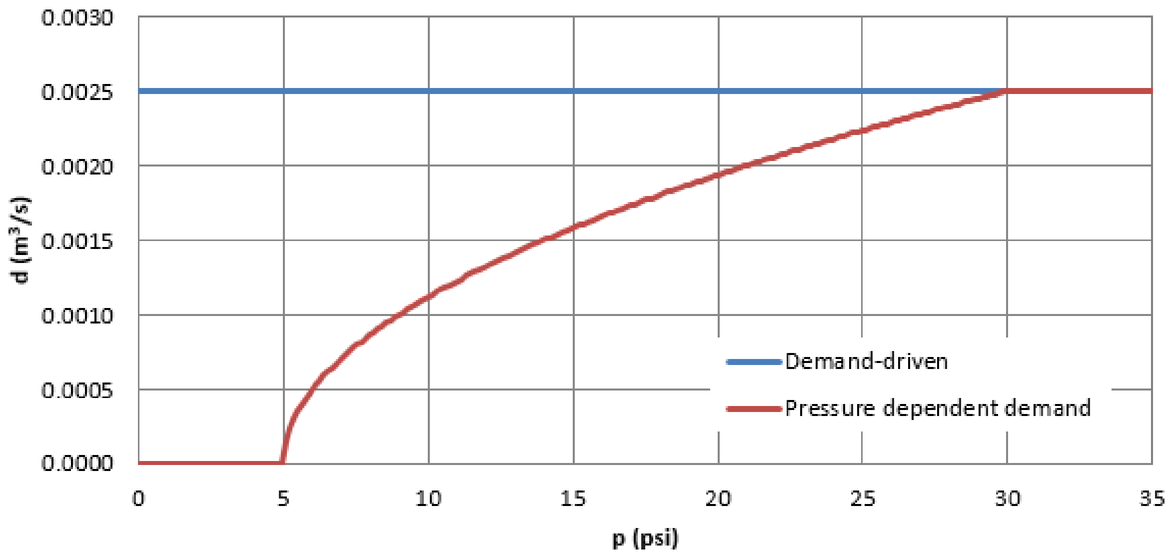


Figure 3.8: *Pressure-driven vs demand-driven model (Klise* et al. *2018)*

Python - NetworkX

In this work **NetworkX** (Hagberg *et al.* 2008), a free software for Complex Networks Analysis, was used for conducting network analysis. NetworkX is a **Python** (Van Rossum & Drake Jr 1995) library for analysing graphs and networks. As reported in the official website: *"NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. With NetworkX you can load and store networks in standard and nonstandard data formats, generate many types of random and classic networks, analyse network structure, build network models, design new network algorithms, draw networks, and much more."*

The software program used to analyze data and create plots is **Matplotlib** (Hunter 2007), a free and open-source plotting library for Python programming language and its numerical mathematics extensions **NumPy** (Harris *et al.* 2020). Networks and graph Data were stored in Dataframe and Series. Data manipulation was carried out using **Pandas** (pandas development team 2020), a data analysis library of Python. All of the source-code was implemented using **Eclipse** (Eclipse Foundation), an integrated development environment (IDE) for computer programming.

Small world property in NetworkX

NetworkX provides a library function for evaluating small world property of a network. It measures the small world coefficient sigma of a given graph. Parameters takes as input are : `G` [NetworkX graph] – An undirected graph, `niter` [integer] – Approximate number of rewiring per edge to compute the equivalent random graph, `nrand` [integer] – Number of random graphs generated to compute the average clustering coefficient (Cr) and average shortest path length (Lr), `seed` [integer] – Indicator of random number generation state. This function returns a float stating the small-world coefficient sigma.

In the following chapter, the numerical results of theory and methodologies described in this chapter will be presented and discussed.

Chapter 4

Numerical results

In this section the numerical results of analysis and methods explained in Chapter 3 will be reported and commented. Numerical evaluations were conducted with technology cited in previous *"Software implementations"* section 3.5. In each section a brief explanation of errors and software problems that could possibly arise during the experiments is also included.

4.1 Network structure analysis

Since available data on WDNs are very limited, the largest number of available open-source files has been collected and examined for this research. On the one hand, this approach aimed to offer the best protection and prevention against upcoming errors or software malfunctions caused by several factors (incomplete files, technology incompatibility, etc.). On the other hand, a wider set of benchmark network improves the robustness of experiment results.

Networks that have been investigated are listed in Table 4.1:

Networks					
Anytown	C-town	HAN	MOD	Net3	Net6
NYT	Richmond	TLN	L-TOWN	Hanoi	ZJ
Jilin	KL	ky1	ky2	ky3	ky4
ky5	ky6	ky7	ky8	ky9	ky10
ky11	ky12	ky13	ky14	ky15	RuralNetwork

Table 4.1: *Benchmark networks*

Most of them are artificial network created for water research purpose. Some are built from existing water network system. Network repository was taken from KIOS Research and Innovation Center of Excellence.
For example, Anytown network is one of the most famous and used system for investigating water treatment and project demand improvements. Net3 and Net6 are two of the three sample networks provided by WNTR. C-town is a benchmark network used in Battle of Water Calibration Networks, an academic competition for calibration methodology (Ostfeld *et al.* 2012). KL is a modified form of the network first presented by Kang and Lansey (Kang & Lansey 2012). Ky-series networks are collection of grid, loop and branch system originally created by Matthew Jolly and Steven Hoagland in 2014 (Hoagland *et al.* 2015). MOD is a real size water distribution network from Modena city (Bragalli *et al.* 2008).

4.1.1 WDNs visual inspection

Water network were categorized first according to their structure and topology in the following Table 4.2.
This group of networks presented heterogeneous properties and from the results it is also clear that there is a strong division based on network size, number of special elements (tanks, pumps, valves) and network length. In particular, some of them contain several structural elements like tanks, pumps and valves (*C-town, Net6, ky9, ky10, ky11, ky15*), others are based only on reservoirs, juction and pipes instead (*HAN,*

MOD, NYT, TLN, Hanoi, Jilin, KL, RuralNetwork, ZJ).

After this primary categorization, some networks like **Ky12**, **Net6**, **TLN**, **HAN** and **NYT** were individuated and considered as not suitable for further complex investigations due to their characteristics. TLN, HAN and NYT had respectively 7, 32 and 20 nodes, 8, 34 and 42 edges, not enough to constitute a reliable water network model. On the contrary, Ky12 and Net6 were not analyzed because their amount of structural elements (2355 and 3356 nodes – 2463 and 3892 links) required too high computational effort and simulation time.

Network name	Nodes	Edges	Reservoirs	Tanks	Pumps	Valves	Junctions	Pipes	Length (m)
Anytown	22	41	3	0	1	0	19	40	35173.92
C-town	396	444	1	7	11	4	388	429	56738.77
HAN	32	34	1	0	0	0	31	34	39420.0
MOD	272	317	4	0	0	0	268	317	71806.11
Net3	97	119	2	3	2	0	92	117	65750.96
Net6	3356	3892	1	32	61	2	3323	3829	638768.34
NYT	20	42	1	0	0	0	19	42	222991.68
Richmond	872	957	1	6	7	1	865	949	75613.99
TLN	7	8	1	0	0	0	6	8	8000.0
L-TOWN	785	909	2	1	1	3	782	905	43214.75
Hanoi	32	34	1	0	0	0	31	34	39420.0
Jilin	28	34	1	0	0	0	27	34	28991.0
KL	936	1274	1	0	0	0	935	1274	252497.77
ky1	859	985	1	2	1	0	856	984	65616.31
ky2	815	1125	1	3	1	0	811	1124	152248.45
ky3	275	371	3	3	5	0	269	366	91286.96
ky4	964	1158	1	4	2	0	959	1156	260241.03
ky5	427	505	4	3	9	0	420	496	96581.4
ky6	548	647	2	3	2	1	543	644	123202.96
ky7	485	604	1	3	1	0	481	603	137048.6
ky8	1332	1618	2	5	4	0	1325	1614	247343.31
ky9	1261	1343	4	15	17	56	1242	1270	961908.03
ky10	935	1061	2	13	13	5	920	1043	430025.77
ky11	831	882	1	28	21	15	802	846	459239.53
ky12	2355	2463	1	7	15	22	2347	2426	648627.17
ky13	785	944	2	5	4	0	778	940	153297.58
ky14	384	553	4	3	5	0	377	548	103758.5
ky15	669	703	2	8	13	28	659	662	482035.72
RuralNetwork	381	476	2	0	0	0	379	476	1288420.13
ZJ	114	164	1	0	0	0	113	164	126436.0

Table 4.2: *Benchmark networks*

4.1.2 Network size

Starting from results obtained in Figure 4.2, it was clear that number of nodes (size) and total pipes length (network extension) had to be considered as key aspects in WDNs resilience investigations. Therefore, similitude between these two metrics has been investigated both graphically using the Matplotlib scatter plot and numerically with Pearson correlation function. This evaluation constituted a starting point for further hydraulic tests and networks assessments. Due to the heterogeneous length of analyzed network, in addition to the main graph, a secondary zoom plot of small networks is given (Figure 4.2).

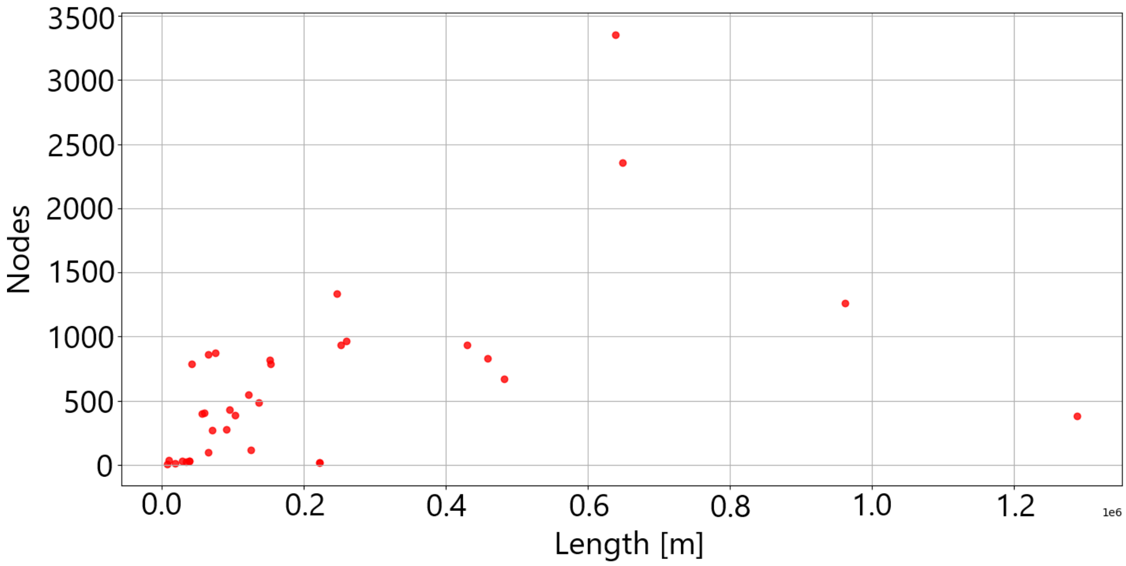


Figure 4.1: *Scatter plot of size and length shows a positive correlation*

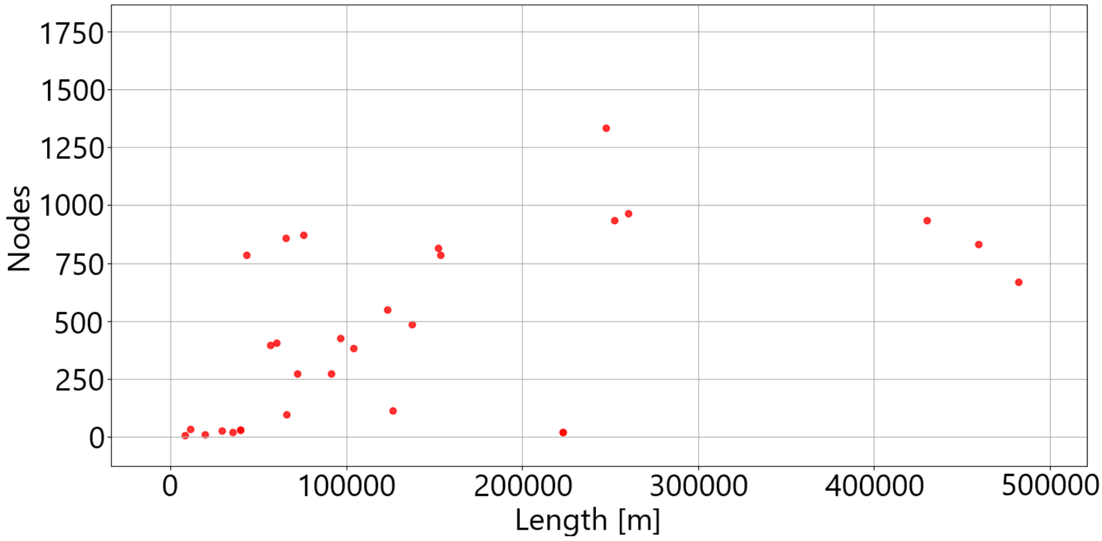


Figure 4.2: *Zoom of previous node/length scatter plot*

From the scatter plots of Figures 4.1 and 4.2 emerges a positive correlation between number of nodes and length of the network. Further proof of positive correlation could be provided by Empirical Cumulative Distribution Function of nodes number and networks length, plotted in Figures 4.3 and 4.4.

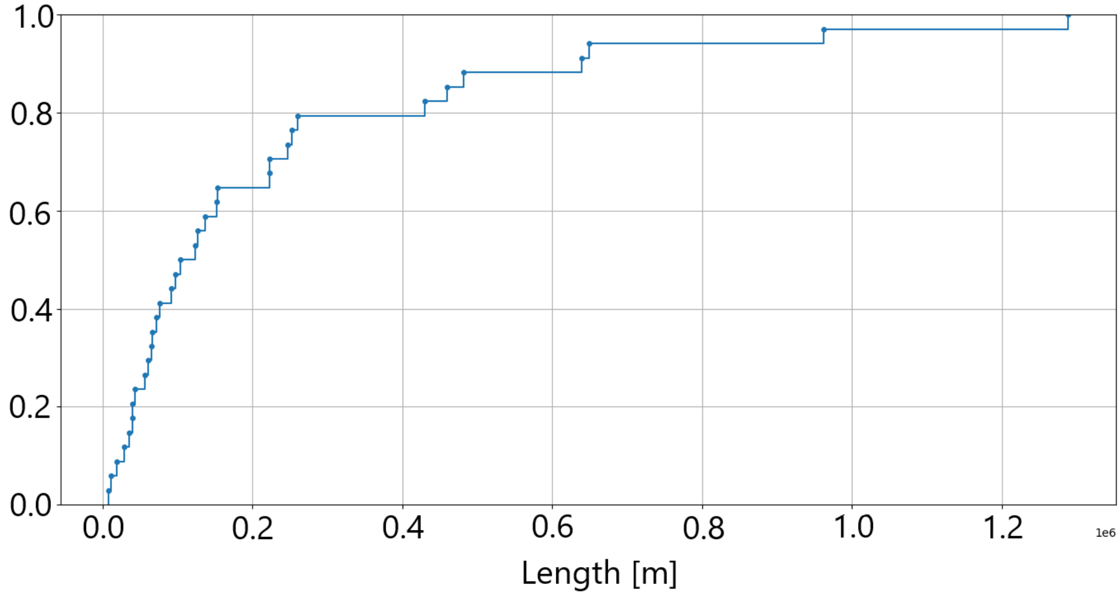


Figure 4.3: *ECDF length*

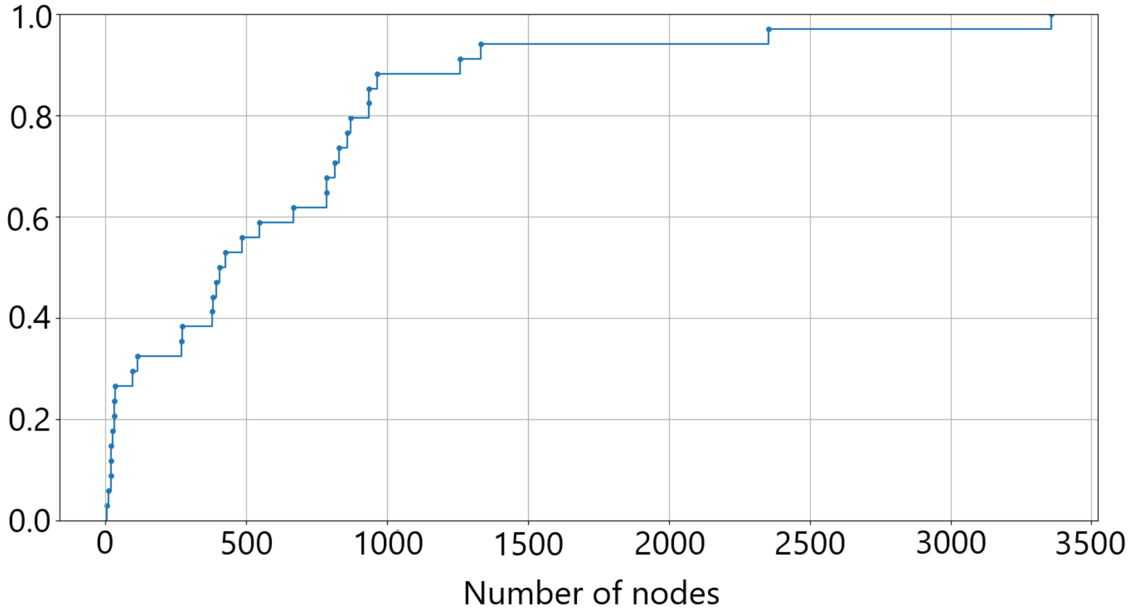


Figure 4.4: *ECDF node*

As it can be seen from Figures 4.3 and 4.4, the pattern of the corresponding curve distribution is similar for both for length and node metrics.
From these ECDF graphs, four networks subgroups (*small – medium – large - extra large*) were created according to 25-th/50-th/75-th percentiles of both metrics. This division was applied in degree distribution calculation for a better comparison between node and length quantities. In both Figures 4.5 and 4.6 the degree distributions are reported with number of nodes as y-value: in Figure 4.6 the networks are assigned to the four size-groups using nodes number as division parameter while in Figure 4.5 the division is based on networks length.
As expected this size visualization brings a clearer vision of difference among the four groups: small networks are located in the lower part of the graph, medium-large in the center and wide network in the upper part.

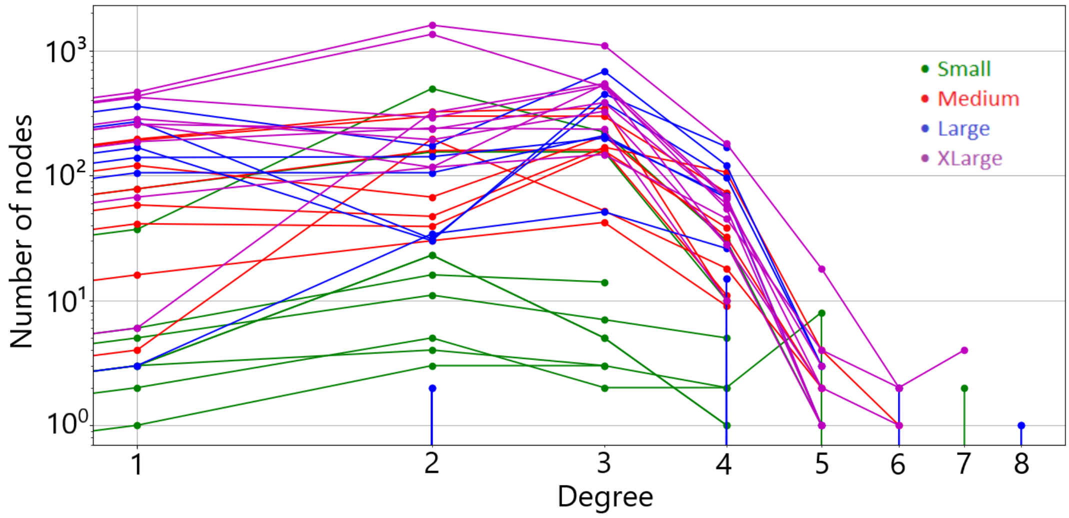


Figure 4.5: *Degree distribution: Length based division*

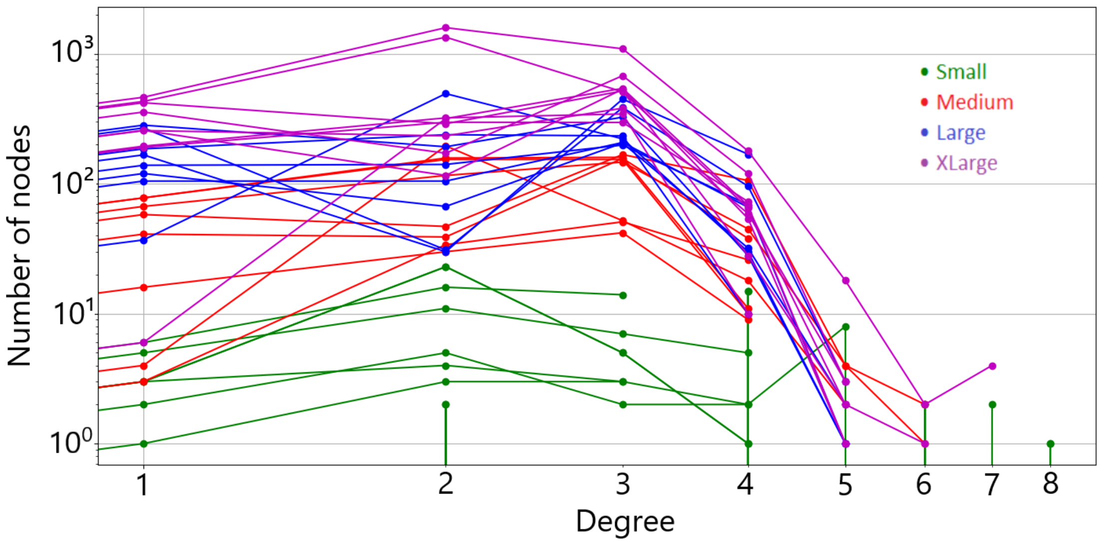


Figure 4.6: *Degree distribution: Node based division*

Degree distribution plot re-confirmed the previous thesis: no great difference in both networks segmentation can be noticed (only few networks switch to another group) and an overall metrics similitude emerges. From Figures 4.5 and 4.6, no completely separated cluster can be seen, therefore, networks categorization was made according to percentile division.

4.2 Complex networks indices

As soon as the topology classification was completed, research proceeded with small world coefficient calculation. This metric was one of the attribute used in final test set creation, networks were classified with their σ value. Subsequently, remaining complex metrics were calculated for chosen networks. Later in the text a visual example using betweenness centrality as nodes attribute will be presented. Metrics like eigenvector centrality, closeness centrality, degree and clustering coefficient are not visually reported but will be directly used in the last part of this section for correlation evaluation.

Small world coefficient σ was evaluated as the median of all the value calculated with different combinations of approximate number of rewiring per edge (NITER - NRAND). A small-world trend ($\sigma > 1$), as stated in Table 4.3, could be observed in overall networks. This phenomenon is particularly marked in some networks like KL, ky2, ky4, ky8, ky9, ky11, ky13, characterized by high σ values.

Network name	σ median
Anytown	1.7
C-town	5,05
MOD	2.9
KL	20.37
ky1	2.42
ky2	10.52
ky3	3.58
ky4	17.05
ky5	4,91
ky6	5,76

Network name	σ median
ky7	4,89
ky8	12,29
ky9	14,93
ky10	5,26
ky11	10,87
ky13	17,53
ky14	4,34
ky15	4,5
RuralNetwork	3.04
ZJ	0,81

Table 4.3: *Small world coefficient*

4.2.1 Test set selection

After WDNs visual inspection and small world classifications, **five network** were selected from the 30 original benchmark networks as test set.
Primary hypothesis was tested on networks with different values of metrics (different small world coefficient) and properties (structural differences like size-lentgh-number of tanks/pumps/valves), in order to guarantee the widest (and most complete) perspective in final results possible. Therefore, the 5-networks test set is composed by heterogeneous networks chosen according to their complex and hydraulic classifications investigated in previous chapters. The final selection was also based on the subgroup identified with node number percentiles (Figure 4.6), one network was taken from each percentile group defined in Section 4.1.2 . The selected networks are: **C-town**, **MOD**, **KL**, **Net3** and **ky6**, in Table 4.4 related features are reported.

Network name	Nodes	Edges	Reservoirs	Tanks	Pumps	Valves	Junctions	Pipes	Length (m)
C-town	396	444	1	7	11	4	388	429	56738.77
MOD	272	317	4	0	0	0	268	317	71806.11
Net3	97	119	2	3	2	0	92	117	65750.96
KL	936	1274	1	0	0	0	935	1274	252497.77
ky6	548	647	2	3	2	1	543	644	123202.96

Table 4.4: *Selected networks test set*

C-town network was chosen for its particular structure and hydraulic properties, it is one with the most complex topology inside the benchmark group. It is also an important academic network used as sample in many scientific papers.
MOD is one with the simplest structure. It does not contain tanks, valves or pumps but has four reservoirs and has a low small world coefficient (σ = 2.9). It was develop starting from Modena city, Italy.
KL was chosen for its size and particularly high small world coefficient (σ = 20.37). It is a simple but widespread network.
Net3 is a WNTR network test, it has the smallest number of element and its structure is particularly adapted for water evaluations.

Ky6 network is a loop system and was chosen as the most representative of the ky-group. It one of the most widespread and has a particular structure characterized by long distance between reservoirs and the network center.

In following Figures 4.7, 4.8, 4.9, 4.10 and 4.11, plotted structures of network is reported, symbol **T** indicates tanks and **R** indicates reservoirs.

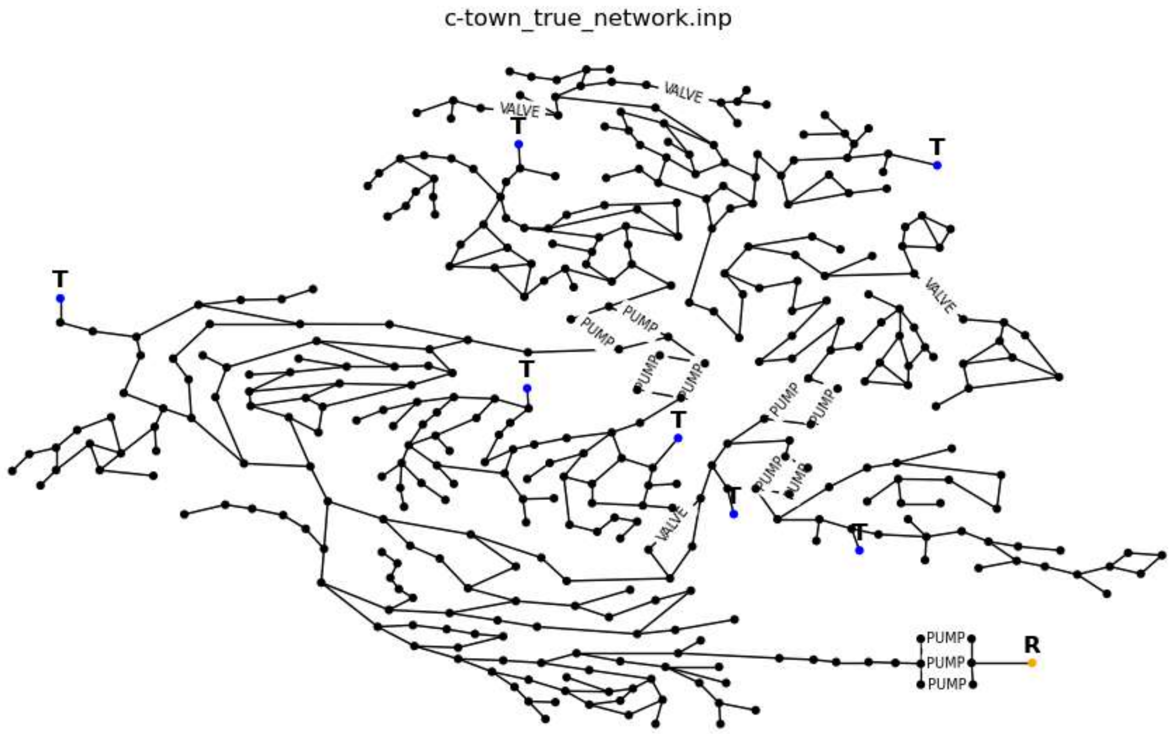


Figure 4.7: C-town structure

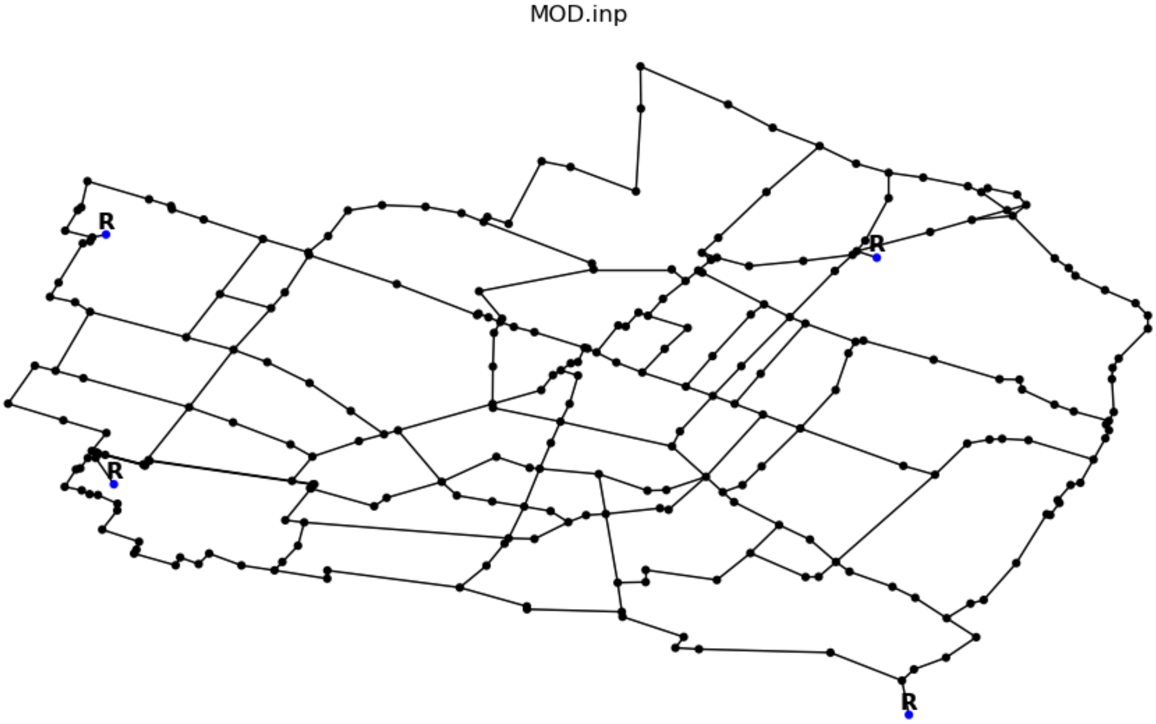


Figure 4.8: MOD structure

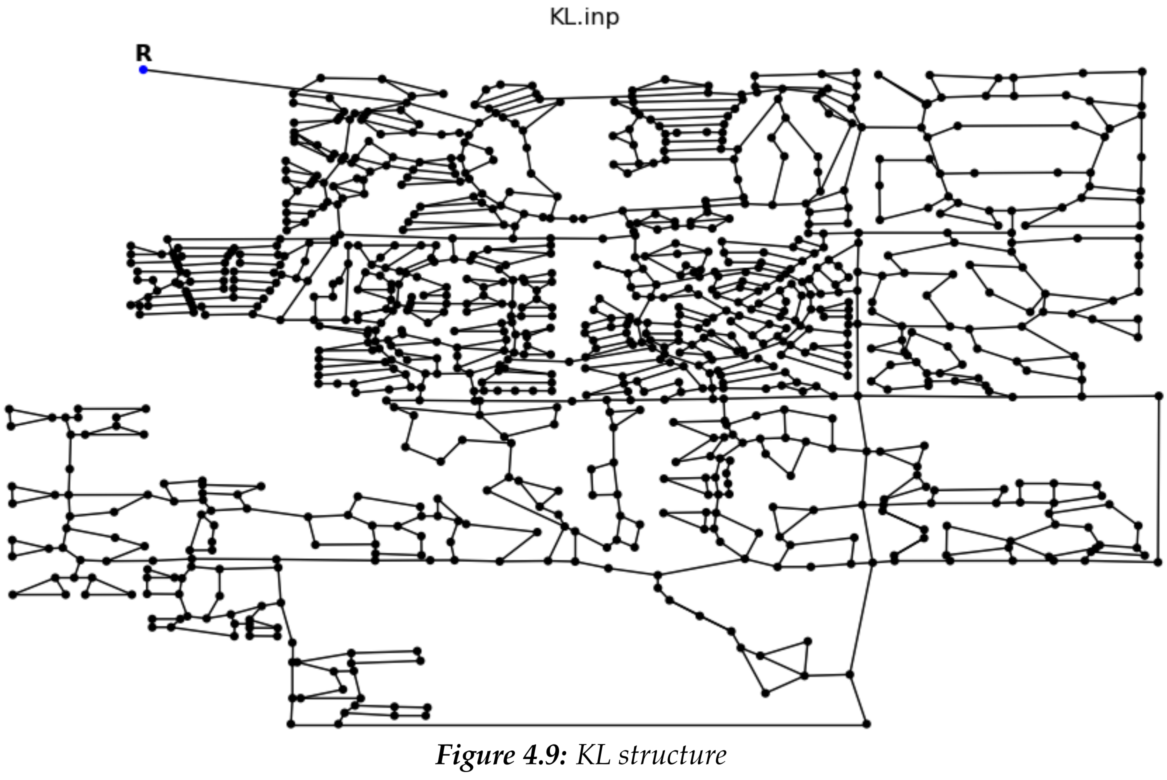


Figure 4.9: KL structure

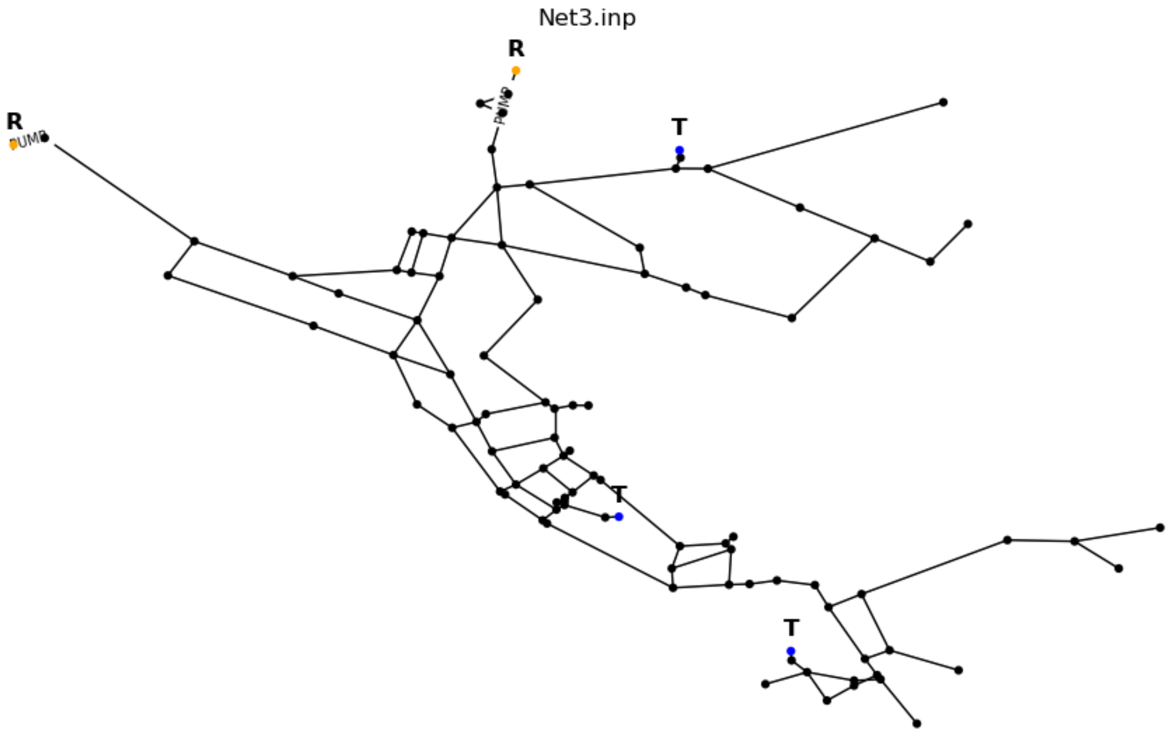


Figure 4.10: *Net3 structure*

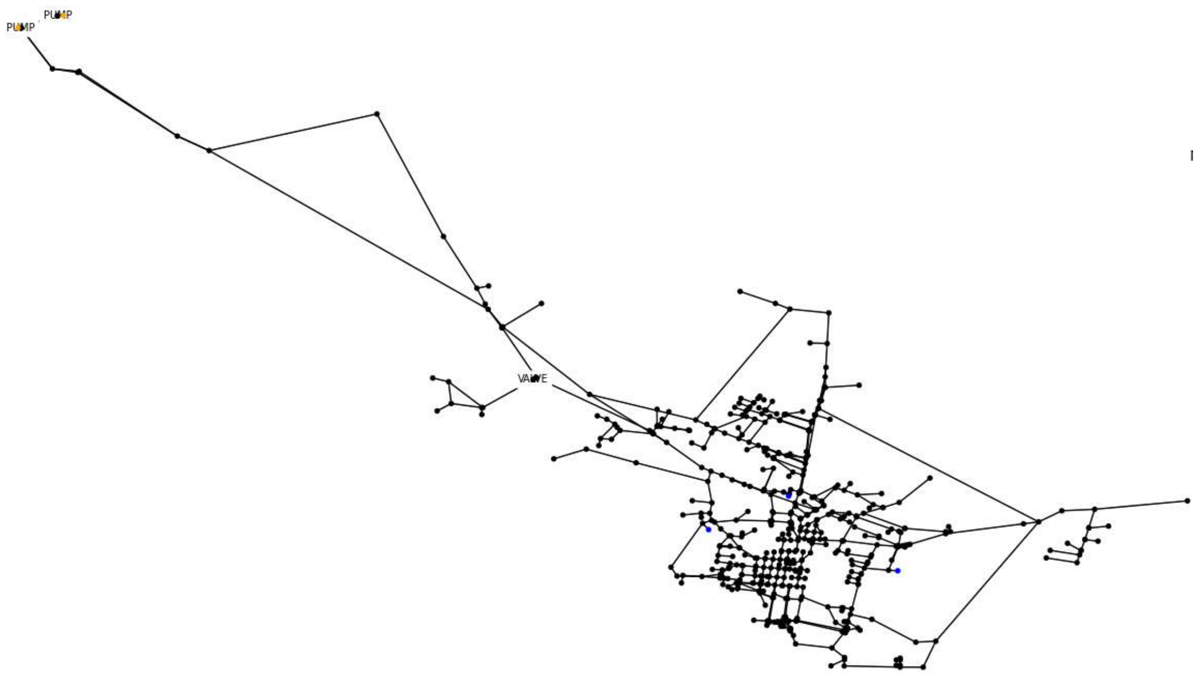


Figure 4.11: *Ky6 structure*

After the networks for the final test set had been chosen, their relative complex network metrics were singularly plotted and studied. Metrics related to a network general property were grouped in Table 4.5, metrics referring to a single node property were graphically analyzed on network structure.

Name	**APL [m]**	σ	**Density**	**Algebraic connectivity**	**Average CC**	**BR**
C-town	3487.32	5.08	0.003	0.00849	0.002	0.495
MOD	3323.78	2.9	0.004	0.91014	0.001	0.013
KL	4117.58	20.37	0.001	0.32592	0.078	0.015
Net3	5496.40	2.03	0.013	0.42076	0.0004	0.261
Ky6	3290.52	5.76	0.002	0.11243	0.0003	0.335

Table 4.5: *Complex network metrics. APL is the average shortest path length,* AVERAGE CC *is the average clustering coefficient,* BR *is the Bridge Ratio.*

APL value is similar for the majority of the networks, exceptions are represented by KL and Net3 networks. KL, despite its high small world coefficient, has the longest APL but the highest average clustering coefficient. These results are coherent with small world definition.

The values of networks **densities** are low, this means the graphs are in general sparse with only a few edges. For the nature of this work this discovery does not provide any relevant information related to the thesis questions.

The **bridge ratio** was evaluated for overall connectivity and a better understanding of network robustness, C-town, Net3 and Ky6 are the least robust, instead of KL and MOD showing a low percentage of bridge-link. This implies that the presence of bridge and cut-set point inside the networks is very limited and this results in a higher level of robustness. A possible implication of bridge ratio with other metrics were sought after, but no significant result were identified.

Algebraic connectivity evaluate the strength of the network connections and the resilience to network disconnection, highest values have been found in network with reduced number of tanks, pipes and valves, e.g. C-town has low algebraic connectivity and a high number of tanks (7), pipes (11) and valves (4).

Small world coefficient σ was primarily investigate to discover any potential hidden implication of small world properties. In support of σ value, the average clustering coefficients were calculated since small world network tend to have higher values than random network (e.g. KL clustering coefficient = 0.078). Moreover, small world networks are usually referred as "locally and globally efficient" and robust to targeted attacks, this could be confirmed by KL network analysis (σ = 20.37). Among the benchmark networks, KL values state a good network condition in terms of resilience and robustness.

4.3 Simulation results

In this section results from different simulations and experiments are reported and evaluated. Methods from Chapter 3 are used on the new test set to quantify the criticality of failures. This measurements will then be used to pursue a possibly similarity with previous complex metrics.

4.3.1 Pressure driven - demand driven analysis results

In order to establish the requested pressure ($\mathbf{P}_{req}$) for the PDD model, since a unifying principle is still missing, a simulation for every networks in **demand driven** (DD) mode was carried out. The DD pressure results were displayed as ECDF and **10-th/50-th/90-th** percentiles were taken as input pressure value ($\mathbf{P}_{req}$) for the following PDD simulation. At 10-th percentile only a small percentage of nodes is under requested pressure level, on the contrary at 90-th percentile the network is undergoing a severe pressure shortage. These values were useful for the following operational status evaluations. In Table 4.6 a pressure value is given for each percentile:

Name	10-th percentile [m]	50-th percentile [m]	90-th percentile [m]
C-town	29.7	55.8	74.3
MOD	20.9	23.6	31.7
KL	33.2	39.8	47.2
Net3	29.1	42.4	48.3
Ky6	42.4	54.4	79.2

Table 4.6: *10-th -50-th -90-th percentile of networks pressure in demand-driven mode*

4.3.2 Failure duration and failure magnitude results

The requested pressure values can now be inserted in pressure-driven formula and evaluate networks failure duration and failure magnitude coefficients. As previously done, both failure coefficients are metrics referred to a single node, therefore a median value referring to the whole network was calculated (Tables 4.7a and 4.7b). This unifying value is given as example and is an approximation calculated for assessing correlation among other network referred metrics, like bridge ratio or algebraic connectivity.

Name	10-th percentile [h]	50-th percentile [h]	90-th percentile [h]
C-town	3.91	53.2	112.75
MOD	13.16	72.09	134.15
KL	10.96	47.61	93.61
Net3	0.1	37.59	88.46
Ky6	9.09	57.54	135.67

(a) *Failure duration*

Name	10-th percentile [%]	50-th percentile [%]	90-th percentile [%]
C-town	$1.6 \cdot 10^{-5}$	0.000177	0.000397
MOD	$4 \cdot 10^{-6}$	$5.8 \cdot 10^{-5}$	0.000317
KL	$2 \cdot 10^{-6}$	$2.5 \cdot 10^{-5}$	$6.7 \cdot 10^{-5}$
ZJ	$9.8 \cdot 10^{-5}$	$9.8 \cdot 10^{-5}$	$9.8 \cdot 10^{-5}$
Net3	$1 \cdot 10^{-6}$	0.00011	0.000574
Ky6	$1.3 \cdot 10^{-5}$	$6.7 \cdot 10^{-5}$	0.000277

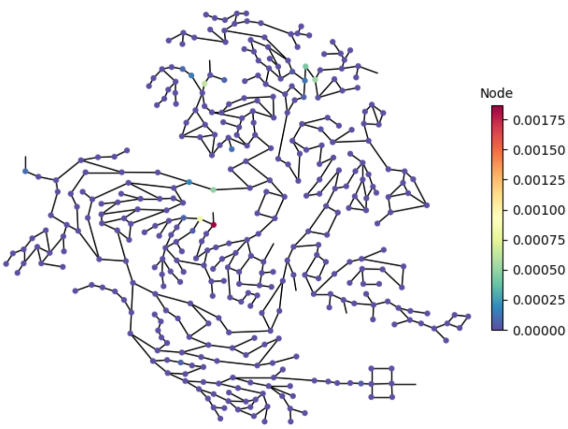
(b) *Failure magnitude*

Table 4.7: *Mean value of networks failures*

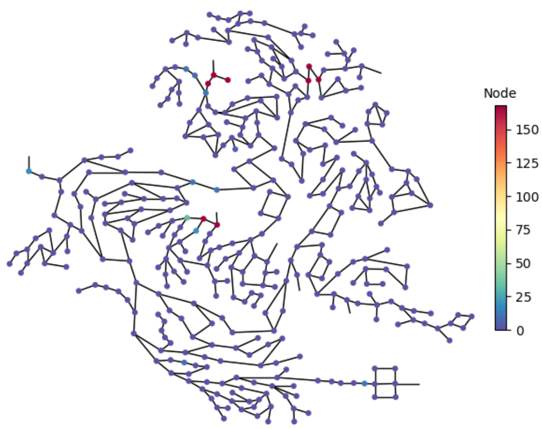
At 10-th percentile all the network show low level of failures, in particular Net3 failure duration is close to zero. Most reliable indicator can be found at 90-th percentile where two different typology of networks are highlighted. Taking into account that the maximum period of failure is 1-week simulation time (*168 hours*), KL and Net3 show great resilience to nodes failures and a quick recovery. Remaining networks

tend to be highly subjected to long duration failure. The most remarkable result to emerge from the data comparison is that lowest value of failures belong to those network with the longest average shortest path length (KL = 4117.58 and Net3 = 5496.4) and, on the contrary, the worst failure duration belongs to the network with the shortest average path length (Ky6). This interesting result may evidence how the correct functioning of a network is not only attributable to structural properties, but multiple factors have to be taken in consideration. The discovery is in line with previous findings in literature and further supports the idea that WDN surrogate model should rely upon different aspects and factors.
Next Figures 4.12, 4.13, 4.14b, 4.15 and 4.16 report previous failure metrics at node level for each percentile as node property. Failure magnitude and duration calculated at every percentile are plotted for each network in test set as **node attribute**. Each node is colored following the intensity of failure that is undergoing, according to the color map on the right side of figures. The same color scale of previous graphic was maintained: red tones implies high values and more blue/violet shades indicate low values. In this case red values stand for high-impact failures ongoing. Networks are sorted following evolution of failures level, from 10-th to 90-th percentile.

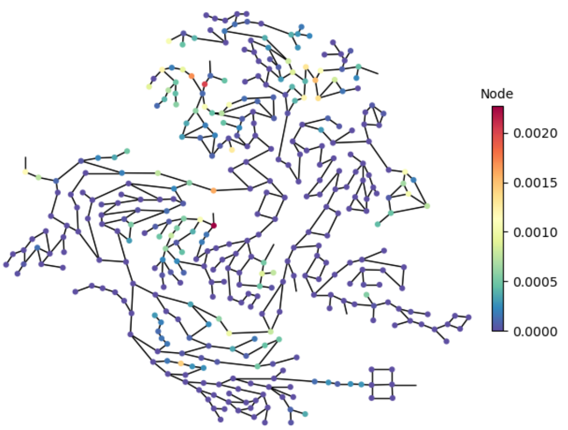
Figure 4.12: C-Town network



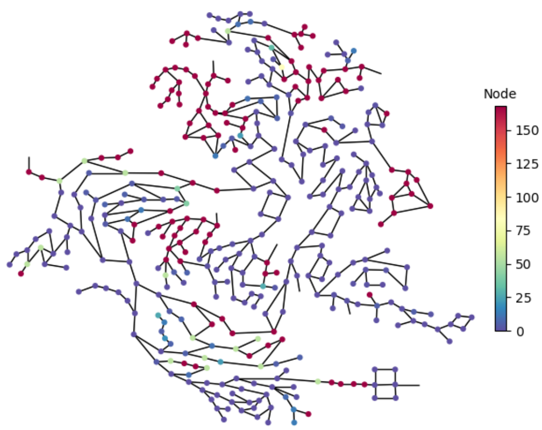
(a) *10-th percentile: Failure magnitude*



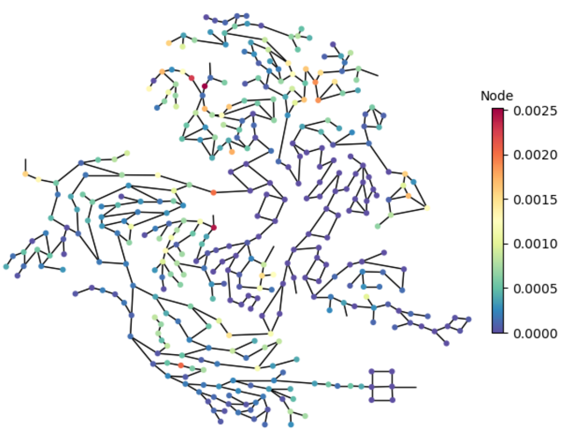
(b) *10-th percentile: Failure duration*



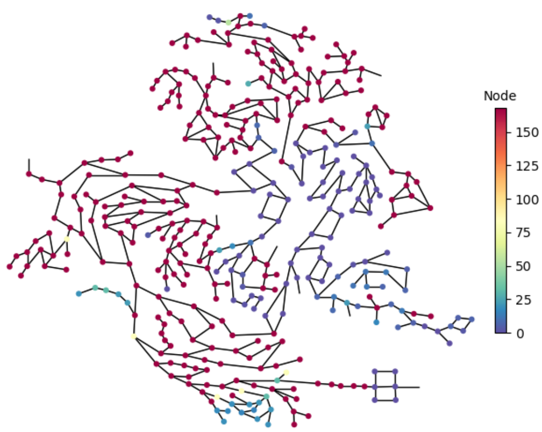
(c) *50-th percentile: Failure magnitude*



(d) *50-th percentile: Failure duration*

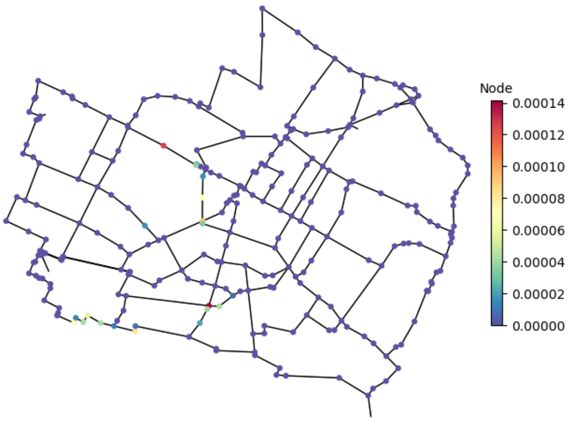


(e) *90-th percentile: Failure magnitude*

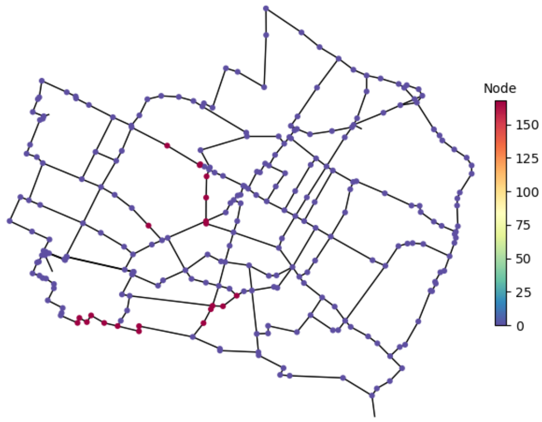


(f) *90-th percentile: Failure duration*

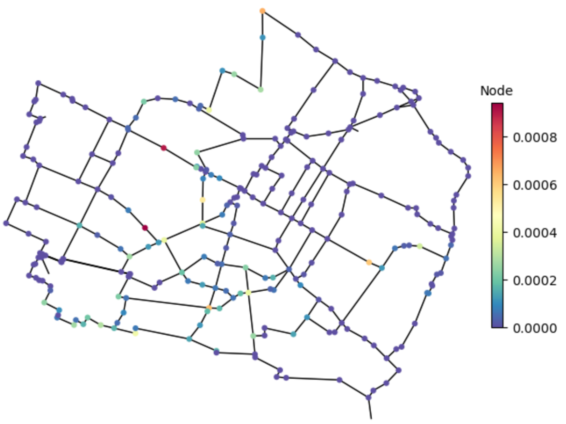
Figure 4.13: *MOD network*



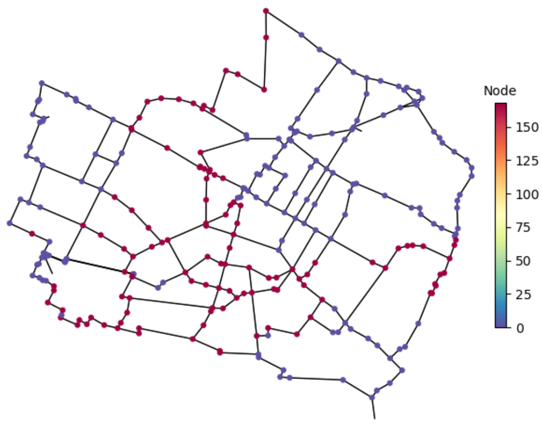
(a) *10-th percentile: Failure magnitude*



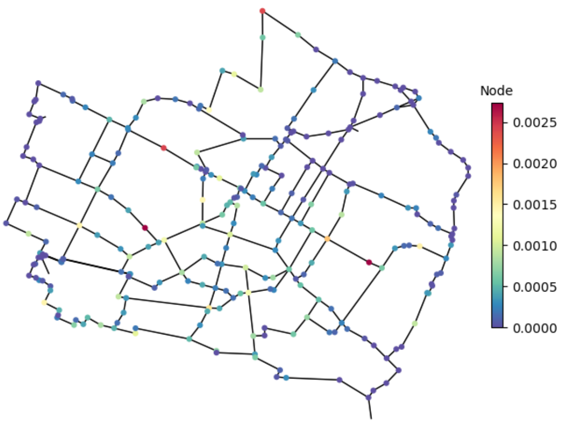
(b) *10-th percentile: Failure duration*



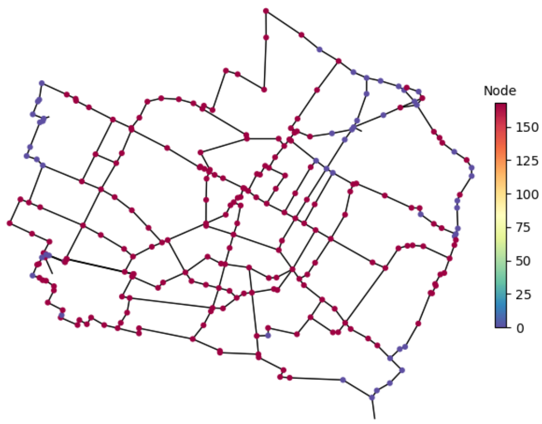
(c) *50-th percentile: Failure magnitude*



(d) *50-th percentile: Failure duration*

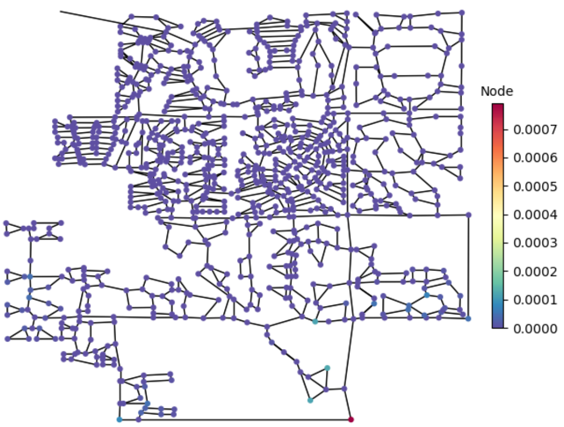


(e) *90-th percentile: Failure magnitude*

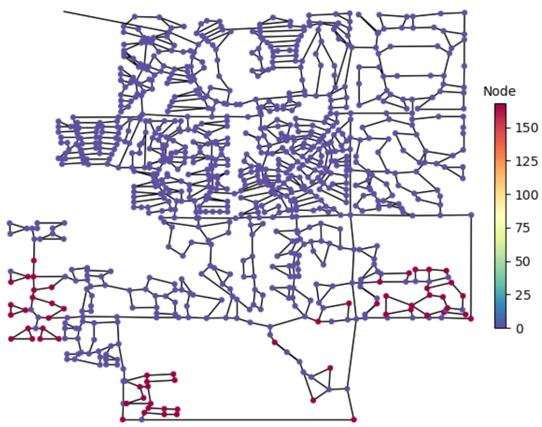


(f) *90-th percentile: Failure duration*

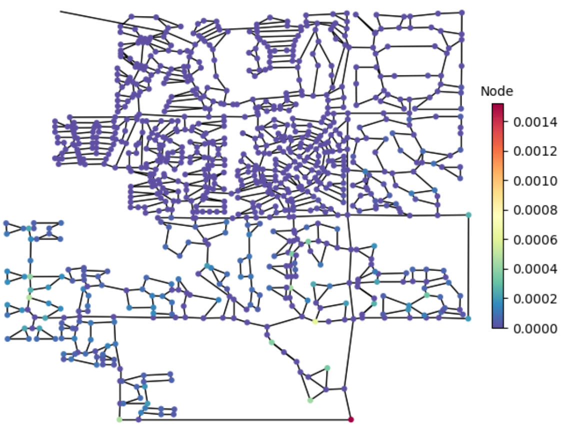
Figure 4.14: *KL network*



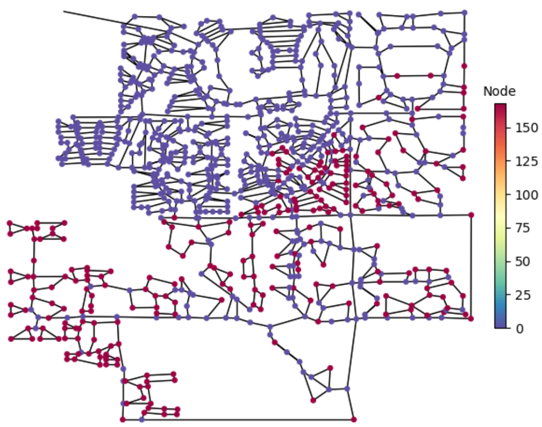
(a) *10-th percentile: Failure magnitude*



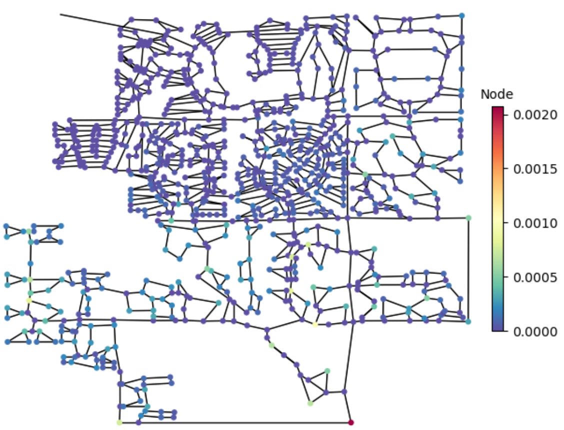
(b) *10-th percentile: Failure duration*



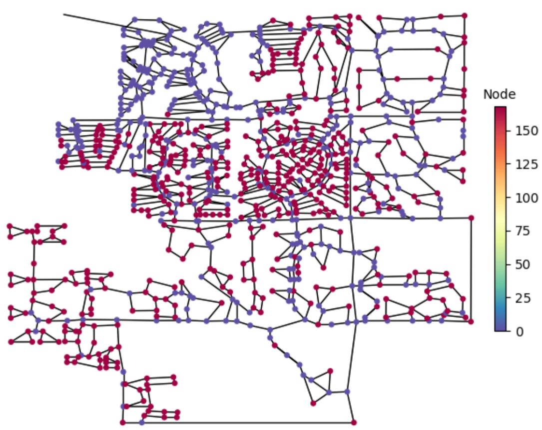
(c) *50-th percentile: Failure magnitude*



(d) *50-th percentile: Failure duration*

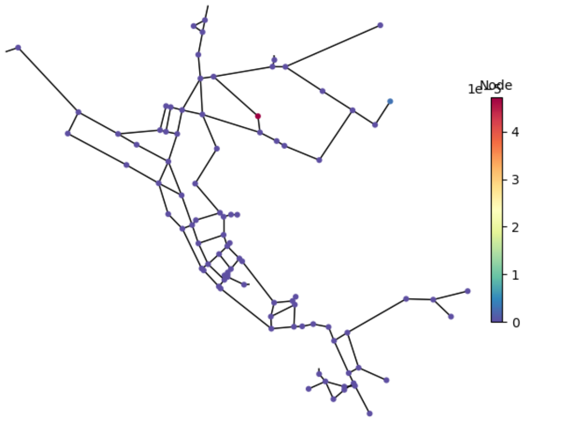


(e) *90-th percentile: Failure magnitude*

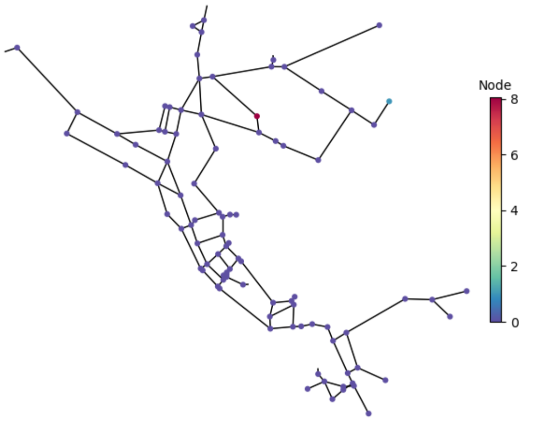


(f) *90-th percentile: Failure duration*

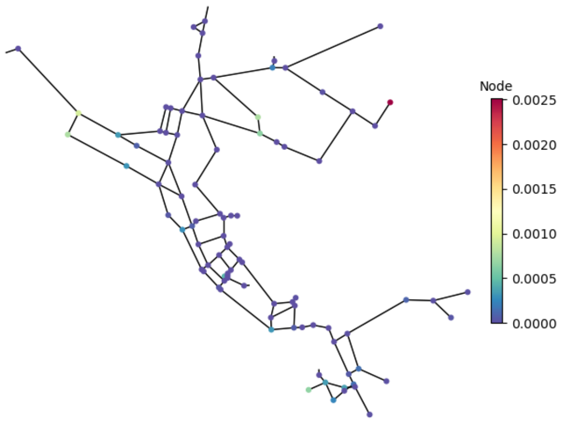
Figure 4.15: *Net3 network*



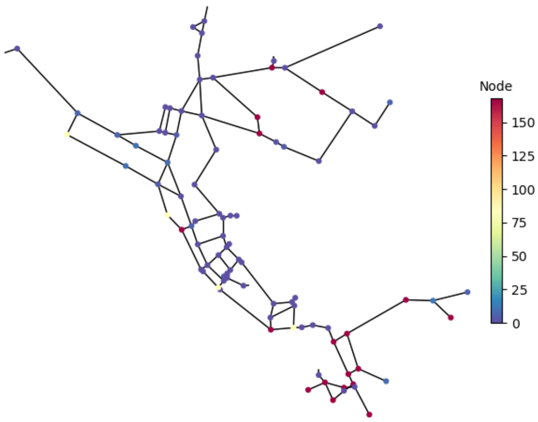
(a) *10-th percentile: Failure magnitude*



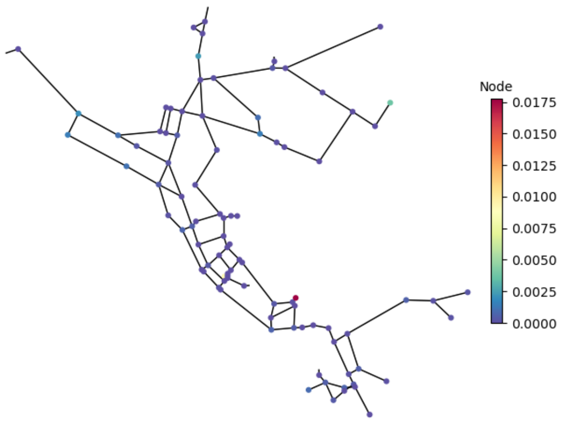
(b) *10-th percentile: Failure duration*



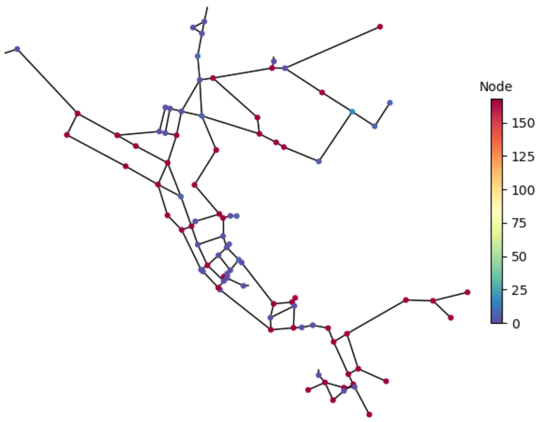
(c) *50-th percentile: Failure magnitude*



(d) *50-th percentile: Failure duration*

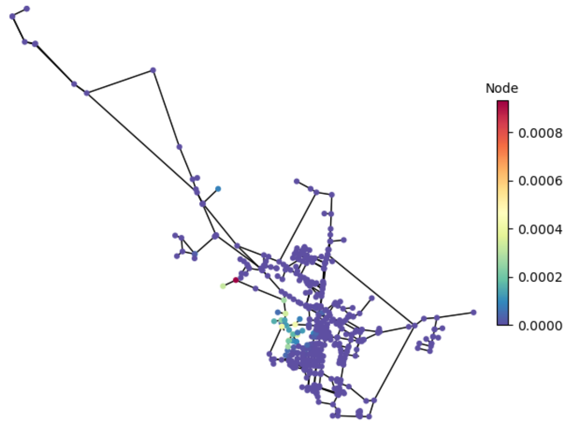


(e) *90-th percentile: Failure magnitude*

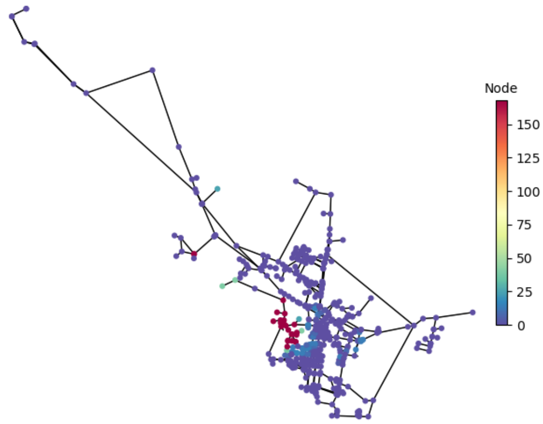


(f) *90-th percentile: Failure duration*

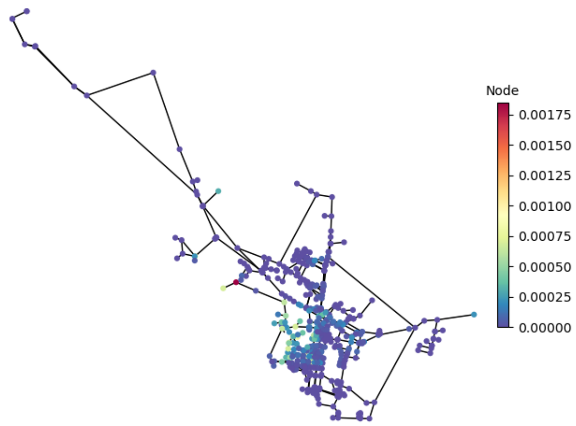
Figure 4.16: *Ky6 network*



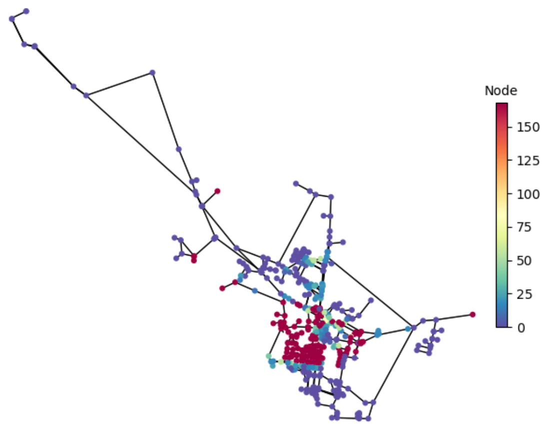
(a) *10-th percentile: Failure magnitude*



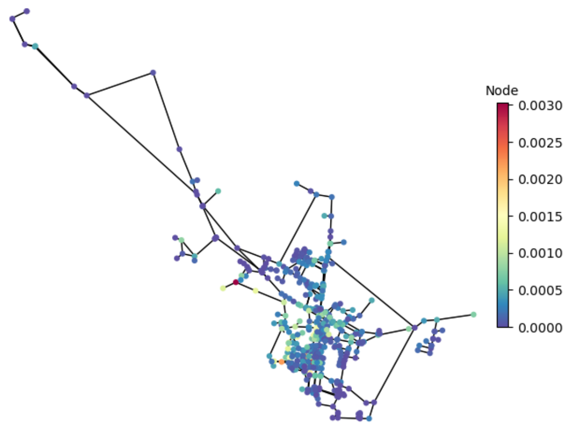
(b) *10-th percentile: Failure duration*



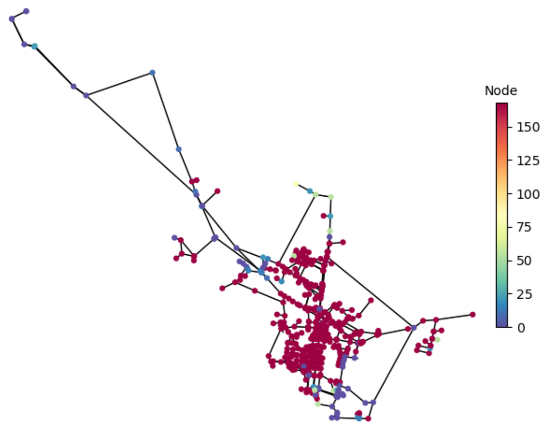
(c) *50-th percentile: Failure magnitude*



(d) *50-th percentile: Failure duration*



(e) *90-th percentile: Failure magnitude*



(f) *90-th percentile: Failure duration*

Visual comparison of results from previous section is evaluated in Figures 4.12, 4.13, 4.14b, 4.15 and 4.16. Failure levels at 10-th percentile are relatively low with brief recovery-time: networks have enough resilience to sustain unsatisfied water demand. At 90-th percentile some networks are affected from severe water shortage, the average time of failure in MOD and Ky6 network is quite long as 1-week simulation (168 hours).

Another interesting aspect that emerges from previous visual inspection of failures is the immediate identification of nodes more susceptible to failure, they are characterized by the red color. Intensity of failures is highlighted by the nodes color: from a light blue for low intense failure to red for severe failures. In the next section these results will be furtherly deepen with correlation analysis.

4.4 Metrics correlation

After all proposed metrics were gathered, correlation evaluations were carried out. Both failures values are complementary and bound together: they describe two different aspects of the water shortage phenomenon. For this reason, before deepen into cross-discipline correlation, relationship between failure values was investigated. For each percentile a **Pearson correlation** between failure duration and failure magnitude was calculated and results are reported in Table 4.8

Name	**10-th percentile**	**50-th percentile**	**90-th percentile**
C-town	0.690	0.702	0.537
MOD	0.769	0.509	0.359
KL	0.338	0.496	0.445
Net3	0.998	0.244	0.198
Ky6	0.664	0.508	0.352

Table 4.8: *Correlation coefficient between failure duration and magnitude for each percentile*

This analysis found evidence for what has been stated before. There is a direct correlation between both failure values. Among the test networks, in Net3 use case was found both the highest value with a correlation coefficient close to 1 and also the lowest with value 0.198. The results of this investigation are in agreement with previous study of Zhan et al. (Zhan *et al.* 2020), from where both failures formulas were taken. Starting from this positive evaluation, research went on investigating correlation between failures and complex network metrics and it is possible factors responsible.

Failure duration and graph metrics

In Table 4.9, 4.10 and 4.11 correlation values between failure duration and complex network metrics at each percentile are reported. Highest values of correlation are emphasised in bold and evaluated with a visual comparison.

10-th percentile					
Name	**BCE**	**Node degree**	**Closeness centrality**	**Clustering coeff**	**EIG**
C-town	-0.031	0.047	-0.039	-0.027	-0.015
MOD	0.02	-0.03	**-0.159**	-0.036	**0.154**
KL	-0.123	-0.095	**-0.223**	0.112	0.091
Net3	-0.114	-0.091	-0.021	-0.029	-0.018
Ky6	0.031	0.015	-0.006	-0.05	-0.024

Table 4.9: *Correlation failure duration and complex network metrics at 10-th percentile*

50-th percentile					
Name	**BCE**	**Node degree**	**Closeness centrality**	**Clustering coeff**	**EIG**
C-town	-0.127	0.01	**-0.138**	0.11	-0.059
MOD	0.085	0.02	**-0.258**	-0.106	**0.150**
KL	**-0.171**	**-0.208**	**-0.174**	0.088	0.023
Net3	-0.115	-0.025	-0.066	-0.134	-0.086
Ky6	**0.158**	-0.055	-0.039	-0.056	-0.07

Table 4.10: *Correlation failure duration and complex network metrics at 50-th percentile*

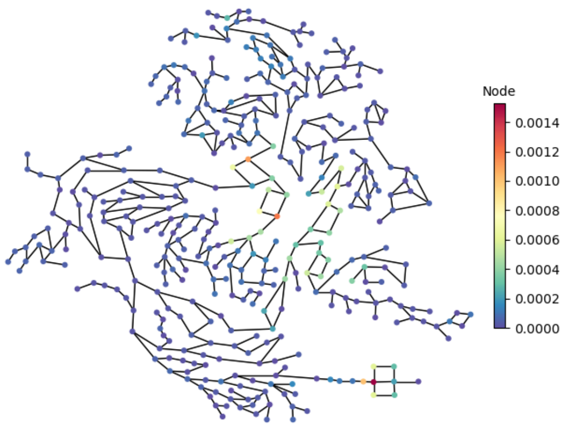
90-th percentile					
Name	**BCE**	**Node degree**	**Closeness centrality**	**Clustering coeff**	**EIG**
C-town	0.166	0.033	**-0.348**	0.054	0.032
MOD	**0.192**	0.014	**-0.232**	-0.091	0.066
KL	**-0.331**	**-0.251**	**-0.140**	0.081	-0.011
Net3	0.015	-0.07	**-0.117**	-0.108	**-0.154**
Ky6	0.085	**-0.128**	-0.039	-0.033	**-0.183**

Table 4.11: *Correlation failure duration and complex network metrics at 90-th percentile*

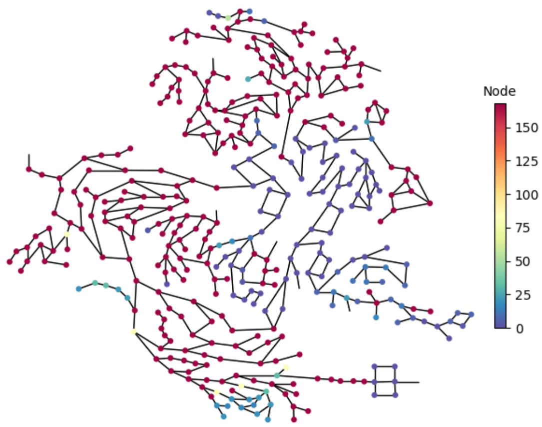
Table 4.11 underlines a partial negative correlation between failure duration and closeness centrality. This results cast a new light to nodes positions inside the network and the possible implications with failures probability. Closeness centrality evaluate closest node to other in the graph (in term of distance, previously fixed as pipe length), inverse correlation could imply that peripheral (furthest) nodes are more involved in failures. In support of this assumption, visual comparison of the percentile with the highest correlation values is provided in Figures 4.17 and 4.18. Since an indirect correlation was present between failure duration and closeness centrality, an inversion of nodes color is detected among figures. Light colored nodes tend to become darker in correlated figure and vice versa. In other words, nodes with high closeness centrality tend to have lower failure duration.

A deeper evaluation related to closeness centrality will be provided in the next section investigating the role of nodes elevations.

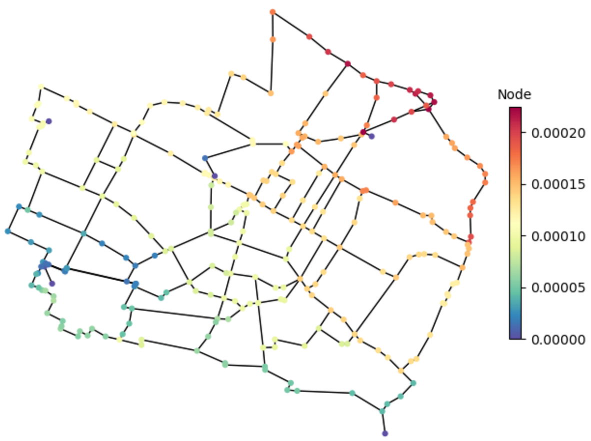
Figure 4.17: *Visual comparison closeness centrality - failure duration*



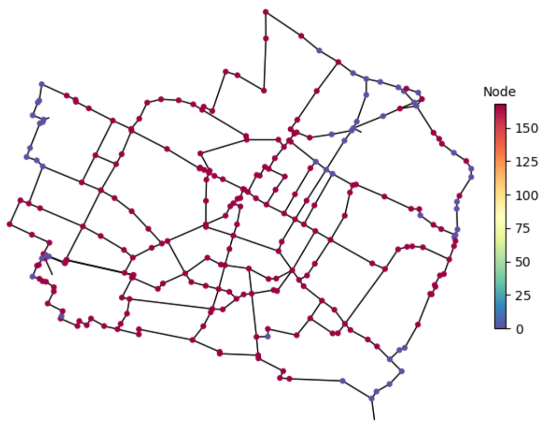
(a) *C-town: Closeness centrality*



(b) *C-town: Failure duration, 90-th percentile*

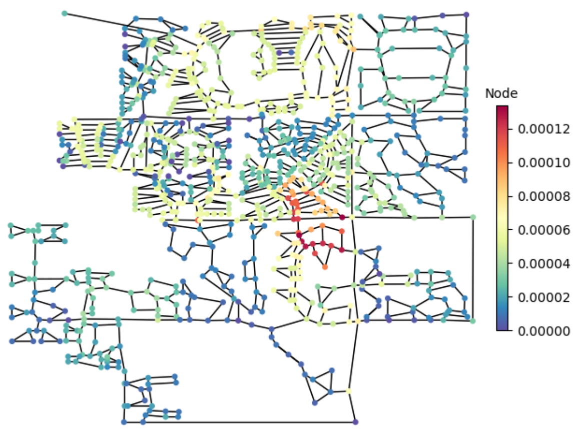


(c) *MOD: Closeness centrality*

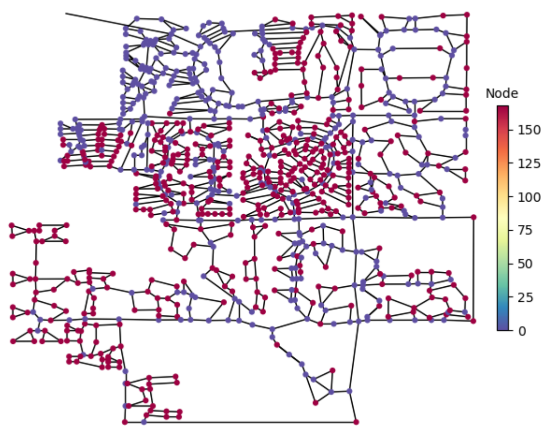


(d) *MOD: Failure duration, 90-th percentile*

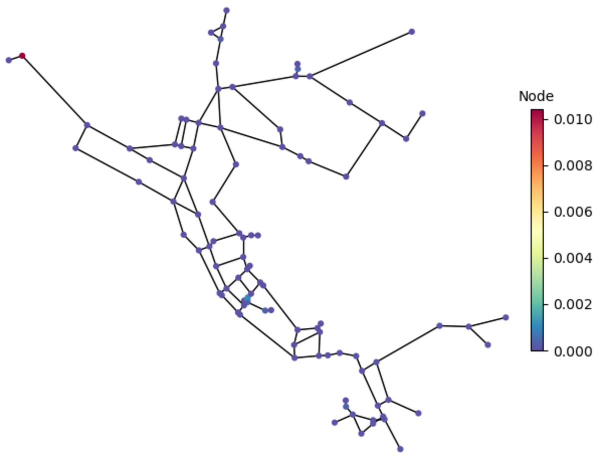
Figure 4.18: Visual comparison closeness centrality - failure duration



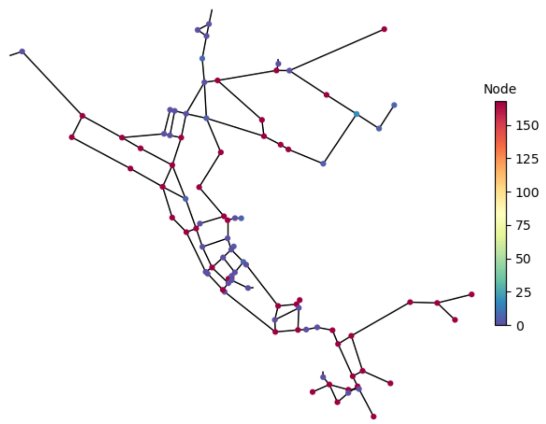
***(a)** KL: Closeness centrality*



***(b)** KL: Failure duration, 90-th percentile*



***(c)** Net3: Closeness centrality*



***(d)** Net3: Failure duration, 90-th percentile*

Failure magnitude and graph metrics

Tables 4.12, 4.13 and 4.14 report correlation values between failure magnitude and complex network metrics at each percentile. Highest values of correlation are emphasised in bold and evaluated with a visual comparison.

10-th percentile					
Name	**BCE**	**Node degree**	**Closeness centrality**	**Clustering coeff**	**EIG**
C-town	0.013	0.05	-0.045	-0.021	-0.012
MOD	0.056	0.049	-0.121	-0.027	**0.360**
KL	-0.039	-0.05	-0.092	0.019	**0.666**
Net3	-0.109	-0.08	-0.02	-0.028	-0.017
Ky6	-0.005	0.03	-0.047	-0.037	-0.018

Table 4.12: *Correlation failure magnitude and complex network metrics at 10-th percentile*

50-th percentile					
Name	**BCE**	**Node degree**	**Closeness centrality**	**Clustering coeff**	**EIG**
C-town	-0.047	0.009	-0.124	0.066	-0.042
MOD	0.066	0.029	-0.015	-0.054	**0.205**
KL	-0.062	-0.069	-0.115	0.098	**0.441**
Net3	**-0.221**	-0.19	-0.055	-0.084	-0.053
Ky6	0.029	0.107	-0.027	-0.007	-0.039

Table 4.13: *Correlation failure magnitude and complex network metrics at 50-th percentile*

90-th percentile					
Name	**BCE**	**Node degree**	**Closeness centrality**	**Clustering coeff**	**EIG**
C-town	-0.096	0.018	**-0.205**	0.082	-0.045
MOD	0.11	0.115	-0.098	-0.063	**0.132**
KL	-0.094	-0.092	**-0.156**	0.094	**0.343**
Net3	**-0.187**	**-0.287**	-0.042	-0.054	0.042
Ky6	0.003	**0.181**	-0.023	0.104	-0.061

Table 4.14: *Correlation failure magnitude and complex network metrics at 90-th percentile*

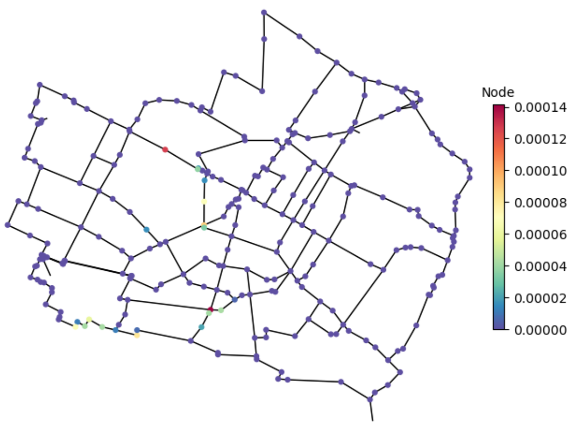
Unlike failure duration, low correlation with closeness centrality has been found. In this case, it is interesting to note the correlation between **eigenvector centrality** and failure magnitude in KL and MOD network. In both networks high values have been measured (at 10-th percentile respectively 0.666 and 0.360). As level of magnitude increase, on the contrary, the correlation tend to decrease. This behaviour could led to a primary identification of a nodes property related to those that will be more targeted by failures. Starting from eigenvector centrality definition, centrality of a node is based on the centrality of its neighbors. This could suggest that for certain networks, first nodes targeted by failures may play big roles both in network supply system and complex graph. Visual comparison of networks KL and MOD at highest correlation percentile is given in Figures 4.19.

Figure 4.19: *Visual comparison eigenvector centrality - failure magnitude*

(a) *KL: Eigenvector centrality*

(b) *KL: Failure duration, 10-th percentile*

(c) *MOD: Eigenvector centrality*



(d) *MOD: Failure magnitude, 10-th percentile*

As mentioned in Chapter 3, **nodes elevation** plays an important role in hydraulic network and water distribution. Separately from failure values, this fundamental network property was carried out as side investigation. Since node elevation, by design constrictions, was *not* taken in consideration during failure assessment and CNT measurements, a possible correlation with complex metrics was studied.

Elevation correlation

Further confirmation of nodes elevation importance is given by correlation investigations. Even if elevation was not taken into account, high correlation between elevation and failure duration was measured, as confirmed in Tables 4.15 and 4.16.

Failure duration			
Name	**10-th percentile**	**50-th percentile**	**90-th percentile**
C-town	**0.291**	**0.389**	0.063
MOD	**0.264**	**0.236**	**0.174**
KL	**0.415**	**0.543**	**0.199**
Net3	**0.188**	0.113	-0.046
Ky6	**0.263**	**0.470**	**0.524**

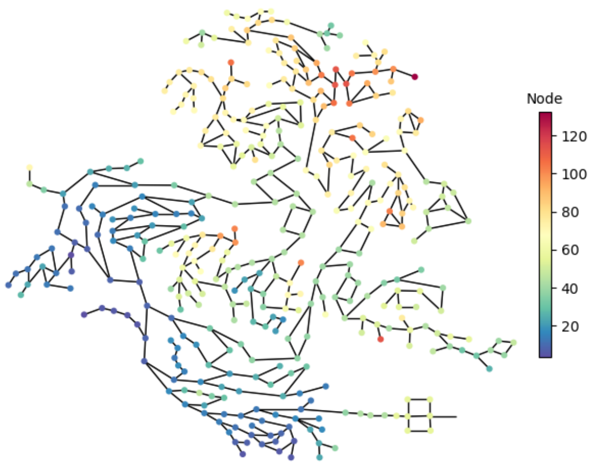
Table 4.15: *Correlation between elevation and failure duration*

Failure magnitude			
Name	**10-th percentile**	**50-th percentile**	**90-th percentile**
C-town	**0.305**	**0.384**	**0.214**
MOD	0.105	**0.140**	**0.191**
KL	**0.402**	**0.393**	**0.188**
Net3	-0.023	**0.190**	**0.186**
Ky6	**0.276**	**0.297**	**0.208**

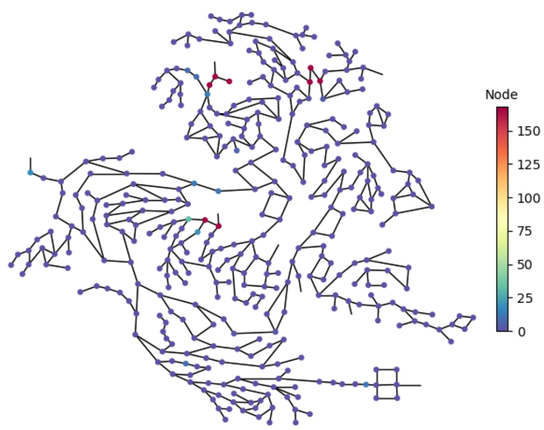
Table 4.16: *Correlation between elevation and failure magnitude*

Direct correlation between elevation and both failure duration and magnitude values has been found. Further evidence is given in Figures 4.20 and 4.21 where both elevation and failure duration at 10-th percentile are displayed beside as visual comparison, showing how nodes with a high elevation are more likely to be the first targeted by failures when requested pressure level is not completely fulfilled. For this reason elevation may play an important role in resilience investigation and surrogate modelling.

Figure 4.20: *Visual comparison elevation - failure duration*

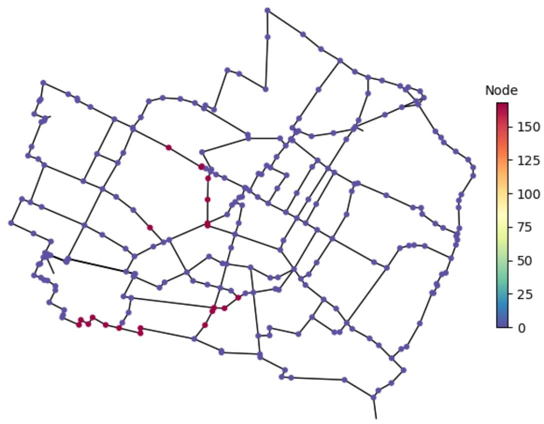


(a) *C-town: Elevation*



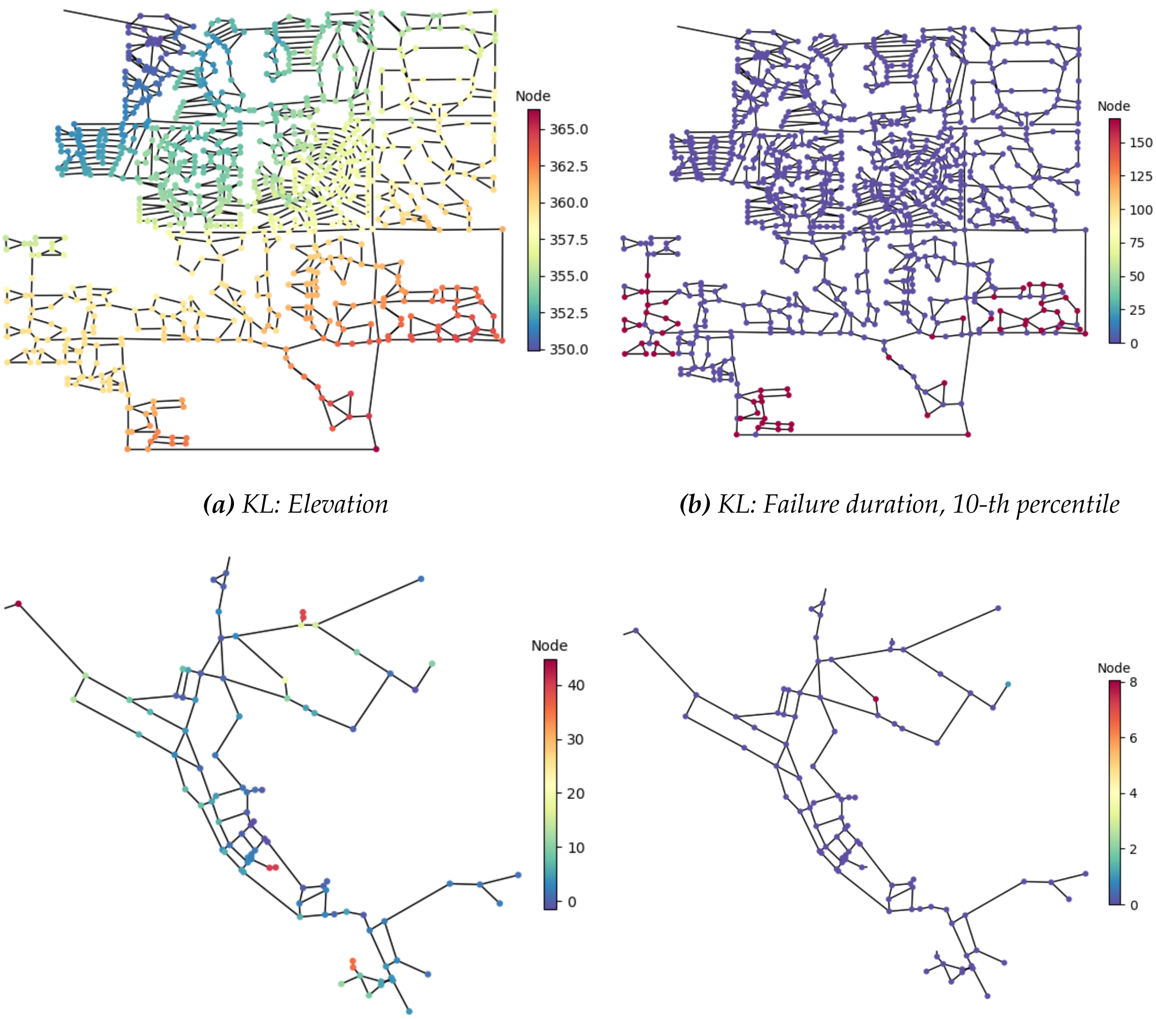
(b) *C-town: Failure duration, 10-th percentile*

(c) *MOD: Elevation*



(d) *MOD: Failure duration, 10-th percentile*

Figure 4.21: *Visual comparison elevation - failure duration*



(a) *KL: Elevation*

(b) *KL: Failure duration, 10-th percentile*

(c) *Net3: Elevation*

(d) *Net3: Failure duration, 10-th percentile*

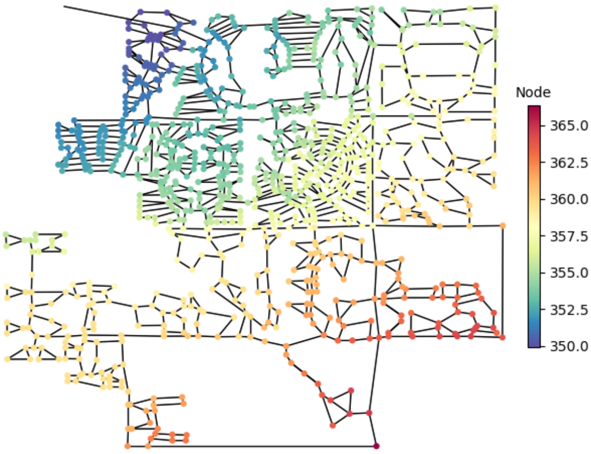
A similar pattern of results was obtained between elevation and some complex network metrics (Table 4.17): **closeness centrality** of some networks is highly indirect or direct correlated with nodes elevation. This may find an explanation in closeness centrality definition, a node is classified with high closeness centrality if its distance to other nodes is relatively small (The node is close to other nodes). Starting from previous discovery and since peripheral nodes tend to have greater elevation, they are likely to be the most isolated and the one with the lowest closeness centrality values.

Name	**BCE**	**Node degree**	**Closeness centrality**	**Clustering coeff**	**EIG**
C-town	-0.049	0.034	0.059	**0.199**	**-0.118**
MOD	-0.038	-0.029	**-0.814**	**-0.212**	0.08
KL	0.002	-0.098	**-0.137**	0.034	**0.123**
Net3	**-0.297**	**-0.153**	**0.504**	**-0.122**	-0.068
Ky6	**0.140**	-0.049	-0.012	-0.003	-0.067

Table 4.17: *Correlation between elevation and complex metrics*

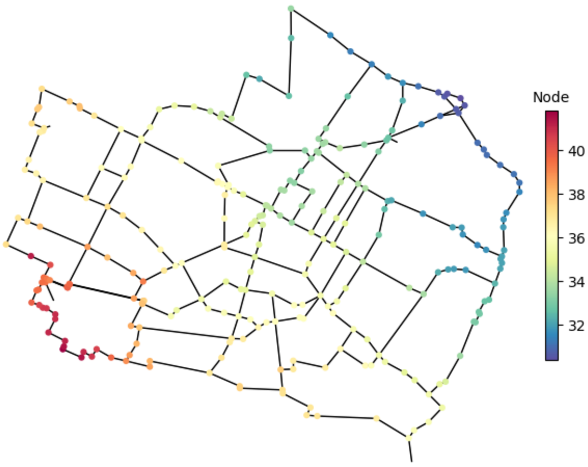
For example, networks **MOD** (correlation value -0.841) and **KL** (-0.137) provide a partial support to previous hypothesis showing indirect correlation. Key result can be found in MOD network visual comparison, both figures could be considered quite complementary since an inverse colour tendency can be noticed. As is reported in next Figure 4.22, highest closeness centrality belongs to the node with the lowest elevation. In some network, peripheral nodes tend to have the highest elevations and be the most isolated, their centrality values is low and, as expected, indirect correlation could be found. Together with evaluation made in Figure 4.20 and 4.21, this may imply that nodes with a high centrality are less subjected to failure. One concern about finding a possible similitude is represented by positive correlation in **Net3** (correlation value 0.504). This change in trend may be due to aforementioned technical errors or bad water network model conversion.

Figure 4.22: Visual comparison elevation - closeness

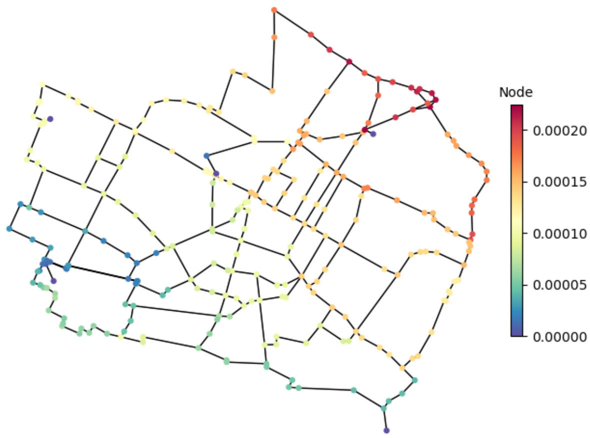


(a) KL: Elevation

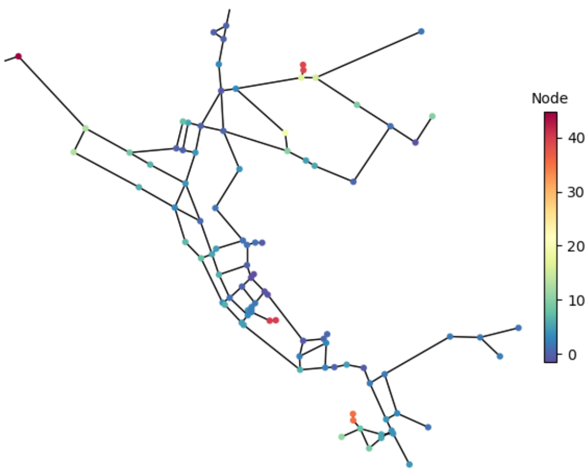
(b) KL: closeness centrality



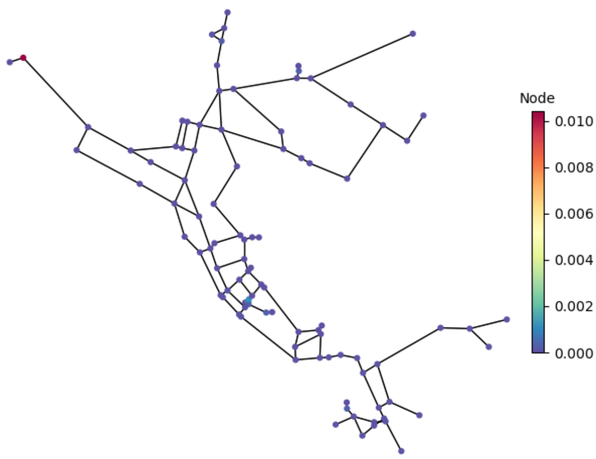
(c) MOD: elevation



(d) MOD: closeness



(e) Net3: elevation



(f) Net3: closeness

Chapter 5

Discussion

As stated in the title of this work, this research is an explorative analysis of an interdisciplinary approach for resilience assessment of water distribution systems. Initial questions and hypothesis on complex network are based on literature review. The lack of a common unifying principle due to ongoing studies still on an early stage emphasises even more the explorative nature of this case study. The obtained findings, as reported in reviewed literature, underline existing analogies between complex network application and water distribution system. Yet, the preliminary discoveries of this thesis need to be interpreted with caution since all the investigations and supposition are not based on empirical data from real networks.
In this work water distribution networks have been considered complex networks with links and nodes representing interconnected hydrualic components. Different topological and physical parameters were used to analyze a potential relationship between CNT and WDNs metrics. The main conclusion from the results highlights the non-trivial and complex nature of WDNs, proving the technical discoveries in previous research. Advanced analysis techniques are necessary to provide a more reliable and effective surrogate models, e.g. more accurate graph weight involving several types of metrics (nodes elevation, water quality, operational status, ecc) should be studied. Moreover, multiple aspects and scales (spatial, timing, efficiency, resilience, optimization, etc.) have to used when evaluating WDNs. The results confirm what was previously discovered by authors Yazdani and Jeffrey (Yazdani & Jeffrey 2012): a

sectoral analysis or an exclusive characterization relying only on few metrics cannot provide a sufficiently accurate model due to the complexity nature of WDNs.

Furthermore, this experimental set up and simulation calibration were inspired by the ones proposed in the literature (Di Nardo & Di Natale 2011; Yazdani & Jeffrey 2012; Di Nardo, Cavallo, *et al.* 2016; Giustolisi *et al.* 2017; Di Nardo, Di Natale, *et al.* 2018; Simone *et al.* 2018; Sitzenfrei, Wang, *et al.* 2020). Analysis referred to single nodes attributes are found to be a better evaluation parameter for resilience and robustness than network referred attributes, since values approximated as average at a global scale are too generic for a valid representation. The present study confirmed the complexity of WDNs, findings on metrics similitude hint that a probably correlation among them is present. Confirmation may be found in positive correlation between CNT and WDNs (closeness centrality metrics and failures values): it appears that under certain conditions (proposed in Chapter 3 and evaluated in Chapter 4), complex networks theory could represent a useful tool for preliminary structural hypothesis and this discovery is in line with previous research. On combining this result with correlation between elevation and failures values, it may be deduced that node elevation should be taken in consideration in surrogate modelling and resilience investigation. This correlation involving three different metrics not only puts in evidence the role of node elevation in WDNs conversion into complex graph, but also the direct implication it has with failures impact.

In addition to previous discovery, small world application in WDNs also find validation in this thesis: small world property may disclose hidden features like higher robustness against failures that could be useful criteria for structural and resiliency analysis. Moreover, association between low failures values and high level of small world coefficient were found. Proof of this hypothesis is provided by KL network results at Tables 4.7a and 4.7b, where, despite its large size, it is one of the most resilient network with low failures values.

Conclusion

Interdisciplinary research using CNT on WDNs was conducted. The answers to the main hypothesis questions were investigated looking for a possible structural correlation between the two disciplines. Focus was centered on the most suitable criteria to model a WDNs and reducing computational effort for water network resilience assessment. In particular, the key aspect of networks resilience against failures was explored.

Starting from reviewed literature, this study provides further evidence for existing correlation between complex network theory and water distribution network management. The effectiveness of the proposed methodology was conducted on five benchmark WDNs, selected from a heterogeneous group of thirty networks according to main complex network and topology criteria. Diversity among them was kept in order to test the same hypothesis on different network typologies. The main results from this thesis highlight that CNT can provide an useful tool to support WDNs analysis, overall this first attempt of WDNs modeling casts a new light on small world property implications: association between low failures values and high level of small world coefficient were found.

Another evidence of this study points toward the idea that nodes elevation plays an important role in WDNs modeling. Its importance was discovered from preliminary correlation investigation between complex network metrics and hydraulic parameters, highest nodes tend to be less resilient and more targeted by system failures. To this aim, a not static parameter, e.g. pipe length was used as graph weight. A general correlation tendency was evaluated in all the involved complex network metrics, for this reason more accurate solution implementing nodes elevation and simulation-dependent parameters, e.g. water flow, water age, networks status as graph weight should be developed. Satisfactory results were obtained proving what was previously stated: assessment relying only on a single or a few factors are not able to provide a sufficiently detailed model. As mentioned several times in the thesis, the explorative nature of this work also highlights several challenges that remain open for further research:

- Research discoveries should be evaluated on a different type of WDN. Particularly, small world findings should be tested on a wider test set of small world networks.
- Graph links weight should be calibrated including other structural and simulation-dependent parameters like nodes elevation, water flow, water quality, operational status, etc.
- Simulation parameters should be more specific and calibrated on each input .INP file.

Overall, the findings of this thesis can be used to assess networks resilience and robustness starting from critical nodes discovery and suitable surrogate models for water distribution networks.

Bibliography

1. Euler, L. Mechanica, volume 1 (1736).
2. Sylvester, J. J. On an application of the new atomic theory to the graphical representation of the invariants and covariants of binary quantics, with three appendices. *American Journal of Mathematics* **1,** 64–104 (1878).
3. Van Rossum, G. & Drake Jr, F. L. *Python reference manual* (Centrum voor Wiskunde en Informatica Amsterdam, 1995).
4. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world'networks. *nature* **393,** 440–442 (1998).
5. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *science* **286,** 509–512 (1999).
6. Mays, L. W. *Water distribution system handbook* (McGraw-Hill Education, 2000).
7. Todini, E. Looped water distribution networks design using a resilience index based heuristic approach. *Urban water* **2,** 115–122 (2000).
8. Prasad, T. D. & Park, N.-S. Multiobjective genetic algorithms for design of water distribution networks. *Journal of Water Resources Planning and Management* **130,** 73–82 (2004).
9. Laprie, J. *Resilience for the scalability of dependability* in *Fourth IEEE International Symposium on Network Computing and Applications* (2005), 5–6.
10. Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **9,** 90–95 (2007).

11. Bragalli, C., D'Ambrosio, C., Lee, J., Lodi, A. & Toth, P. Water network design by MINLP. *Rep. No. RC24495, IBM Research, Yorktown Heights, NY* (2008).

12. Hagberg, A., Swart, P. & S Chult, D. *Exploring network structure, dynamics, and function using NetworkX* tech. rep. (Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008).

13. Benesty, J., Chen, J., Huang, Y. & Cohen, I. in *Noise reduction in speech processing* 1–4 (Springer, 2009).

14. Di Nardo, A. & Di Natale, M. A heuristic design support methodology based on graph theory for district metering of water supply networks. *Engineering Optimization* **43,** 193–211 (2011).

15. Fortunato, A. SCELTA DEGLI INTERVENTI DI RIABILITAZIONE DELLE RETI IDRICHE CON UNA BASE INFORMATIVA LIMITATA (2012).

16. Kang, D. & Lansey, K. Revisiting optimal water-distribution system design: issues and a heuristic hierarchical approach. *Journal of Water resources planning and management* **138,** 208–217 (2012).

17. Ostfeld, A., Salomons, E., Ormsbee, L., Uber, J. G., Bros, C. M., Kalungi, P., Burd, R., Zazula-Coetzee, B., Belrain, T., Kang, D., *et al.* Battle of the water calibration networks. *Journal of Water Resources Planning and Management* **138,** 523–532 (2012).

18. Strigini, L. in *Resilience assessment and evaluation of computing systems* 3–24 (Springer, 2012).

19. Yazdani, A. & Jeffrey, P. Water distribution system vulnerability analysis using weighted and directed network models. *Water Resources Research* **48** (2012).

20. Hoagland, S., Schal, S., Ormsbee, L. & Bryson, S. *Classification of water distribution systems for research applications* in *World Environmental and Water Resources Congress 2015* (2015), 696–702.

21. Di Nardo, A., Cavallo, A., Di Natale, M., Greco, R. & Santonastaso, G. F. Dynamic control of water distribution system based on network partitioning. *Procedia engineering* **154,** 1275–1282 (2016).

22. Bharali, A. & Doley, A. On small-world and scale-free properties of complex network. *International Journal of mathematics and statistics invention* **5,** 11 (2017).

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

24. Loucks, D. P. & van Beek, E. Water resources planning and management: An overview. *Water Resource Systems Planning and Management,* 1–49 (2017).

25. Di Nardo, A., Di Natale, M., Giudicianni, C., Greco, R. & Santonastaso, G. F. Complex network and fractal theory for the assessment of water distribution network resilience to pipe failures. *Water Science and Technology: Water Supply* **18,** 767–777 (2018).

26. Ferretti, S. On the complex network structure of musical pieces: analysis of some use cases from different music genres. *Multimedia Tools and Applications* **77,** 16003–16029 (2018).

27. Klise, K. A., Murray, R. & Haxton, T. An Overview of the Water Network Tool for Resilience (WNTR). (2018).

28. Simone, A., Ridolfi, L., Berardi, L., Laucelli, D. & Giustolisi, O. Complex Network Theory for Water Distribution Networks analysis. *EPiC Series in Engineering* **3,** 1971–1978 (2018).

29. Pagano, A., Sweetapple, C., Farmani, R., Giordano, R. & Butler, D. Water distribution networks resilience analysis: A comparison between graph theory-based approaches and global resilience analysis. *Water Resources Management* **33,** 2925–2940 (2019).

30. Sitzenfrei, R., Oberascher, M. & Zischg, J. *Identification of network patterns in optimal water distribution systems based on complex network analysis* in *World Environ-*

mental and Water Resources Congress 2019: Hydraulics, Waterways, and Water Distribution Systems Analysis (2019), 473–483.

31. Fiedler, F., Cominola, A. & Lucia, S. Economic nonlinear predictive control of water distribution networks based on surrogate modeling and automatic clustering (2020).

32. Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del R'ıo, J. F., Wiebe, M., Peterson, P., G'erard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C. & Oliphant, T. E. Array programming with NumPy. *Nature* **585,** 357–362. `https://doi.org/10.1038/s41586-020-2649-2` (Sept. 2020).

33. Pandas development team, T. *pandas-dev/pandas: Pandas* version latest. Feb. 2020. `https://doi.org/10.5281/zenodo.3509134`.

34. Sitzenfrei, R., Wang, Q., Kapelan, Z. & Savić, D. Using complex network analysis for optimization of water distribution networks. *Water resources research* **56,** e2020WR027929 (2020).

35. Zhan, X., Meng, F., Liu, S. & Fu, G. Comparing performance indicators for assessing and building resilient water distribution systems. *Journal of Water Resources Planning and Management* **146,** 06020012 (2020).