Tailoring Complex Network Theory for Water Distribution Networks Analysis

Department of Civil Engineering and Architecture

SSD: ICAR02

Ph.D. in

"Conoscenza e Innovazione nel Progetto per il Patrimonio"

Ciclo XXXIII, 01/11/2017- 31/12/2020

Tailoring Complex Network Theory for Water Distribution Networks Analysis

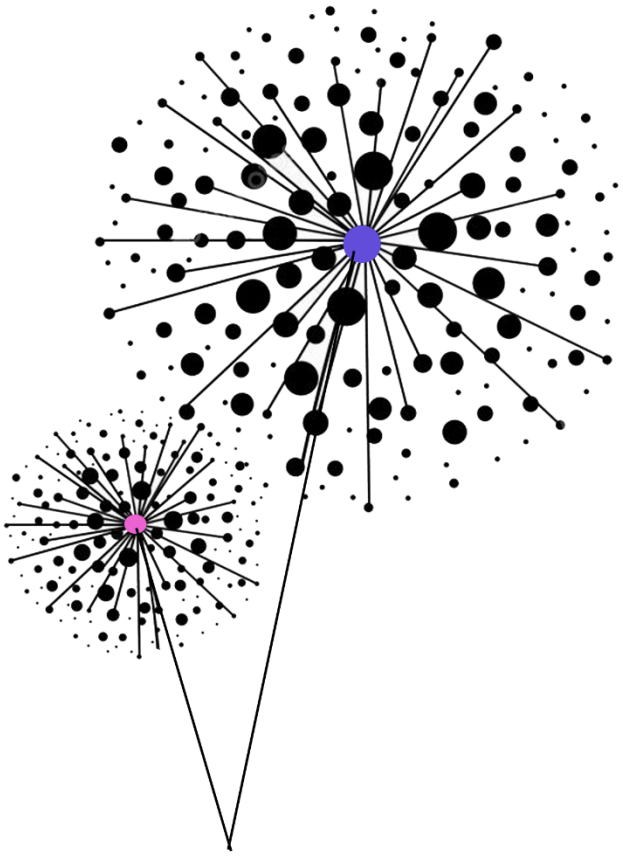
Supervisor: Prof. Orazio Giustolisi

Co-Supervisors: Prof. Luca Ridolfi

Coordinator of Ph.D. Program: Prof. Carlo Moccia

PhD Candidate: Antonietta Simone

Ai miei figli!



Lo studio e la ricerca della verità e della bellezza rappresentano una sfera di attività in cui è permesso di rimanere bambini per tutta la vita.

Albert Einstein

Agisci come se quel che fai, facesse la differenza. La fa.

William James

Acknowledgments

This thesis work is the result of a journey that led me to explore wonderful places and concepts, to overcome many challenges and to work with very special people.

I sincerely thank my Supervisor, Professor Orazio Giustolisi, first for introducing me to the real world of research. His passion in this regard was a true model to follow. I thank him for conveying me a lot of his knowledge and experience in the study and management of aqueducts as well as for always driven me in my research path with valuable advice. Thanks also for sharing with me the passion for research as a lifestyle, at the office, in a park in Seoul or on the *Promenade des Anglais* in Nice.

Special thanks go to my co-supervisor, Professor Luca Ridolfi, for his support and wise guidance, as well as for always reminding me the importance, role, and beauty of research in our conversations, making my work more and more enjoyable. I also thank him for the time spent reading and discussing the drafts of the thesis work with me and for the other projects carried out together during the PhD.

Thanks to Professors Daniele B. Laucelli and Luigi Berardi, who with their scientific rigor have been an example to follow. I also thank them for the pleasant talks and for their constant availability.

I could not have had better mentors.

Thanks to my colleague Francesco G. Ciliberti, for always showing interest in my work, greatly enriching my professional daily life. I also thank him for the availability that has always characterized him.

Un ringraziamento di cuore va a te cara Rosa, per così tanti motivi che mi è impossibile ricordarli ed elencarli tutti. Grazie a voi, Fortu e Vale, perché ovunque siamo, Aversa, Bari, Firenze, Parigi, Adelaide, Londra, o in qualsiasi altra parte del mondo, siete delle amiche straordinarie. Un grazie a te Giova, perché io ti sento sempre tanto vicina a me.

Un grande ringraziamento va alla mia famiglia, per esserci e per sostenermi sempre. Se ho raggiunto questo traguardo lo devo anche e soprattutto al vostro continuo supporto.

Abstract

Everything around us is connected: individuals in society, infrastructures in cities, cells in the brains, interacting species in ecosystems, etc... These connections allow us to represent all complex real systems as networks and help us to understand how the world really works.

Novel approaches and metrics, that often go beyond traditional analysis tools, are made available to model and interpret an impressive number of natural and anthropogenic phenomena, i.e., to investigate and understand the networks behavior. We are talking about the Complex Networks Theory (CNT), a new challenge that fascinates and involves the scientific community.

In this thesis, we aim to contribute to the study of Water Distribution Networks (WDNs), particularly to the relationship between connective structure and dynamics occurring in these complex networks, tailoring classic CNT tools. The study starts from the basic concepts proposed by the CNT, shows the limits they present when applied to infrastructural systems such as WDNs, strongly constrained spatially, and proposes their tailoring.

Then, we suggest a new complex network-based framework to model and simulate WDNs accounting for their intrinsic peculiarities and spatial limits enhancing the WDNs analysis and providing more reliable tools to identify the actual behaviour of these hydraulic systems.

Accordingly, we propose new metrics and models, supported by efficient calculation tools.

Keywords: Complex Networks; CNT; Water Distribution Networks; WDN analysis and classification; centrality metrics; intrinsic relevance of nodes; modularity index.

Résumé

Tout autour de nous est connecté : les individus dans la société, les infrastructures dans les villes, les cellules du cerveau, les espèces en interaction dans les écosystèmes, etc. Ces connexions nous permettent de représenter tous les systèmes réels complexes comme des réseaux nous aidant à comprendre comment le monde réel fonctionne vraiment.

Nouvelles approches et métriques, qui vont souvent au-delà des outils d'analyse traditionnels, sont mises à disposition pour modéliser et interpréter un nombre impressionnant de phénomènes naturels et anthropiques, c'est-à-dire explorer et comprendre les caractéristiques des réseaux. Nous parlons de la Théorie des Réseaux Complexes (CNT), un nouveau défi qui fascine et engage la communauté scientifique.

Dans cette thèse, nous entendons contribuer à l'étude des réseaux de distribution d'eau (WDN), en particulier à la relation entre la structure connective et la dynamique se produisant dans ces réseaux complexes, en adaptant les classiques outils de la CNT. L'étude part des concepts de base proposés par la CNT et montre les limites qu'ils présentent lorsqu'ils sont appliqués à des systèmes infrastructurelles tels que les WDN, fortement contraints spatialement, et propose leur adaptation.

Ainsi, nous proposons un nouveau cadre pour modéliser et simuler les WDN en tenant compte de leurs particularités intrinsèques et de leurs limites spatiales en améliorant leur analyse et en fournissant des outils plus fiables pour identifier le comportement réel de ces systèmes hydrauliques.

À cet égard, nous proposons des nouvelles métriques et modèles, soutenus par des outils de calcul efficaces.

Mots clés : Réseaux complexes ; CNT ; réseaux de distribution d’eau ; analyse et classification de réseaux de distribution d’eau ; métriques de centralité ; pertinence intrinsèque des nœuds ; indice de modularité.

Riassunto

Tutto ciò che ci circonda è connesso: individui nella società, infrastrutture nelle città, cellule nel cervello, specie interagenti negli ecosistemi, ecc. Queste connessioni ci consentono di rappresentare tutti i sistemi reali complessi come reti e ci aiutano a capire come il mondo reale funziona davvero.

Nuovi approcci e metriche, che spesso superano i tradizionali strumenti di analisi, sono resi disponibili per modellare e interpretare un numero impressionante di fenomeni naturali e antropici, ovvero per esplorare e comprendere le caratteristiche delle reti. Stiamo parlando della Teoria delle Reti Complesse (CNT), una nuova sfida che affascina e coinvolge la comunità scientifica.

In questa tesi, ci proponiamo di contribuire allo studio delle reti di distribuzione idrica (WDNs), in particolare al rapporto tra struttura connettiva e dinamiche che si verificano in queste reti complesse, adattando i classici strumenti della CNT. Lo studio parte dai concetti base proposti dalla CNT, ne mostra i limiti che presentano quando applicati a sistemi infrastrutturali come le WDNs, fortemente vincolati spazialmente, e ne propone un adattamento.

Quindi, proponiamo un nuovo framework per modellare e analizzare le WDNs tenendo conto delle loro peculiarità intrinseche e dei loro limiti spaziali migliorando la loro analisi e fornendo strumenti più affidabili per identificare il comportamento effettivo di questi sistemi idraulici.

A tal proposito, proponiamo nuove metriche e modelli, coadiuvati da strumenti di calcolo efficienti.

Parole chiave: reti complesse; CNT; Reti di distribuzione idrica; analisi e classificazione delle reti di distribuzione idrica; metriche di centralità; rilevanza intrinseca dei nodi; indice di modularità.

Contents

1

Introduction

Contents

In the last century, the advent of Complex Network Theory (CNT) allowed us to analyze the nature, complexity, and behavior of many real complex systems. The integration of both mathematical (e.g., probability, graph theory, matrix analysis, etc.) and technical tools (e.g., computer sciences, etc.) as well as the increasing availability of data, allowed one the study and interpretation of a huge number of systems, from physical to social, biological, infrastructural, etc... Such systems present relationships between connectivity structure and dynamics, and the main aspect in these interactions is the topology, i.e., how the elements are connected and interact with each other. *Euler* (1736), more than two centuries ago, for the first time faced a topology problem with graph theory, conceiving bridges (urban infrastructures) as a graph, i.e., a series of vertices connected by edges. Later, several network models and metrics have been introduced in order to associate real systems to predefined patterns and to facilitate their analysis and management. Most of the studies conducted to date using CNT tools focus on immaterial networks, such as hyperlinks for Internet networks, neurons for brain networks, air routes for air transport networks, etc. (see Figure 1- a), b), c)), for which, CNT metrics and models shown to be very useful and of immediate application.

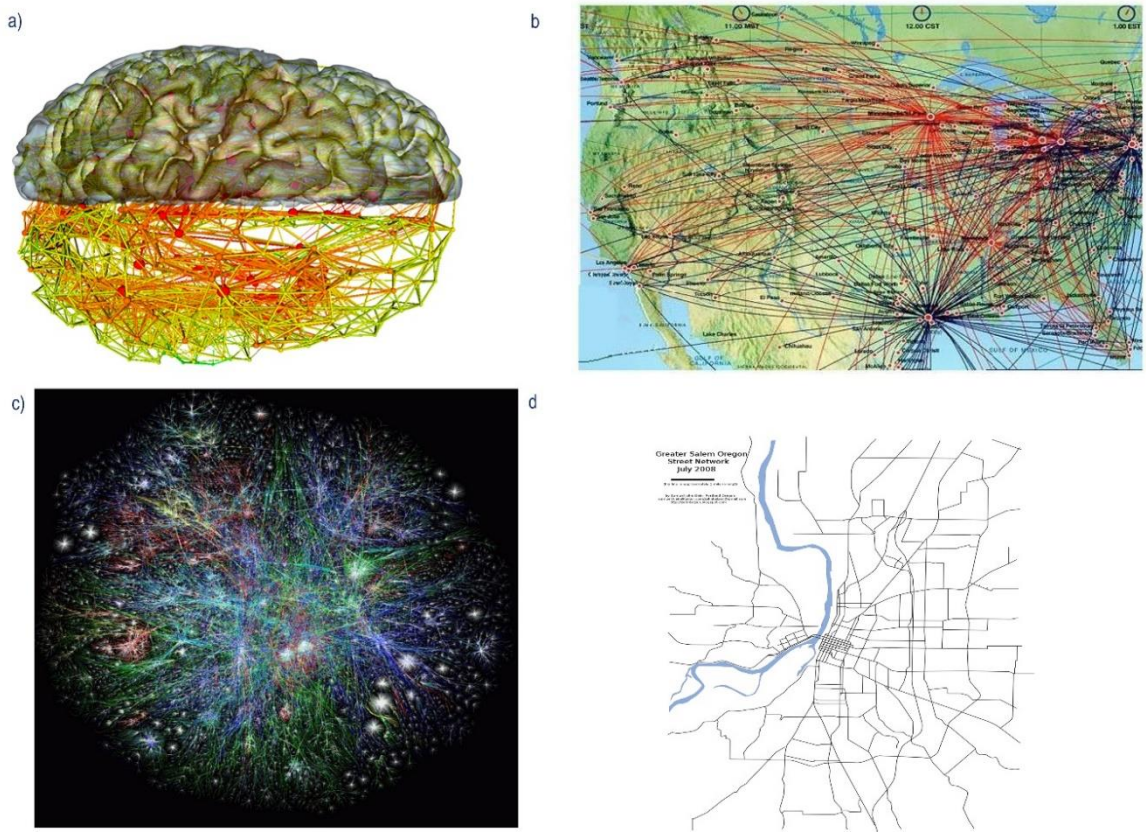


Figure 1. **a**) Brain network. Source: *Van den Heuvel and Sporns*, 2011; **b**) US air route network. Source: *Zhou*, 2004; **c**) Internet network. Source: *Pelino*, 2013; **d**) street network map of Salem, Oregon (USA).

The results in the application of these tools are so satisfactory that the theory soon assumed a general character: valid for any real system, regardless of its characteristics. When these studies also involved infrastructure systems (Figure 1-d), characterized by a tangible and often very spatially limited connective structure (e.g., Water distribution network- WDNs), the results were often unreliable, differing from the real behaviour of the systems analysed. The idea that everything was good for every system failed, i.e., classic CNT tools and framework are not suitable for the study of infrastructure systems as WDNs. The idea of proposing a new framework, tailoring that proposed by the CNT, appeared the only viable one.

Accordingly, the aim of the present thesis is to deepen the study of WDNs, by proposing a new framework, tailoring CNT tools, for WDNs classification and domain analysis. Starting from these considerations, we (i) realized a brief review on the main CNT tools and metrics proposed until today for the study of complex real networks, (ii) evaluated their actual application and usefulness/ unusefulness for WDNs, (iii) tailored the CNT tools in order to capture the specificity of the hydraulic system domain (e.g. the hydraulic features of pipes, the presence of unidirectional devices, etc.), accounting for the hydraulic and topologic features of these spatial networked systems and, finally, (iv) embedded the information about the intrinsic relevance of vertices into CNT tools (*Giustolisi et al.,* 2020), i.e., considering the different role of the vertices in the network (e.g. source nodes, demand nodes, strategic structure, etc.).

The proposal is not alternative to the usual (and necessary) multiple hydraulic simulations but represent a complementary tool to the classic ones, assisting, for example, the engineering judgment for WDN analysis, design, and management tasks (*Giustolisi et al.*, 2019).

This thesis work is the result of a journey that starts 5 years ago, when I approached the study of water distribution networks (WDNs). My attention was mainly focused on studying the domain of such infrastructural systems using approaches based on Complex Network Theory (CNT). The choice to study the WDNs using the CNT tools was linked to the need to find solutions to novel tasks. In fact, we realized that the studies carried out in the last decades by CNT, if properly tailored to the specific peculiarities of WDNs, could make a crucial contribution to addressing and solving several technical needs. The study of the domain, for example, has relevance both from a scientific and technical-practical point of view, and represents a valid support in facing increasingly complex problems in WDNs management.

We produced various contributions, analysing both the hydraulic and topological components of WDNs, fuelling more and more the idea that information on the hydraulic model

and domain should coexist and be complementary to each other to improve analysis and management of these systems.

It was interesting to study the basics of CNT, both from a historical and scientific point of view, and it was very stimulating to try to go further, in search of new paradigms and strategies that would allow us to tailor CNT tools to study WDNs.

1.1 Tailoring the classic CNT framework

Complex Network Theory (CNT) models in an efficient way the most part of the real world as networks. Networks are composed of vertices (nodes) connected by edges (links), that indicate the interactions among them. Several classes of networks exist (*Newman*, 2010), and water distribution networks (WDNs) belong to the infrastructure ones. Such systems are special infrastructure networks belonging to the subset of spatial networks because their connectivity structure is constrained by environmental factors (*Barthélemy*, 2011; *Giustolisi et al.*, 2017). WDNs differ from classic immaterial networks, for which standard CNT tools are useful and effective. In fact, their classification and analysis using the standard CNT tools is very difficult and uncertain. For example, (i) the network structure of WDSs is strongly constrained by the spatial characteristics of the environment where they are constructed; (ii) pipes and not nodes, with their asset features, are the most important network components for hydraulic systems; (iii) WDNs are equipped by directional devices that influence the flow direction; (iv) nodes have different roles in the network (demand nodes, source nodes, connection nodes) for which it is necessary to embed the information about their intrinsic relevance in the analysis. The topology- based study proposed until today is not enough.

According, in this thesis we propose a new framework for WDNs analysis tailoring the classic CNT tools.

1.1.1 Pipes as key components for WDNs analysis

CNT tools are based on a nodal perspective. In fact, generally, standard centrality metrics looks at ranking the node importance: it is customary to identify nodes with higher degree values, for classification purposes, or betweenness values, for system vulnerability purposes, etc… Preserving the importance of the nodal perspective, with the useful information that can result from it, this study focuses on the need to shift the perspective to edges. For WDNs, in fact, the relevant physical components are the edges (pipes), material components that with weights corresponding to their asset features (e.g., length, diameter, hydraulic resistance, etc.) influence the hydraulic behavior. The hydraulic features of pipes play a relevant role in

determining the hydraulic state and need to be carefully considered in the analysis. Nodes become components responsible to transfer the water (i.e., information) among pipes. There is an exception: water source nodes represent a sort of hydraulic hubs and need to be studied with a specific analysis. Therefore, pipes and source nodes are the most important component of WDNs.

Accordingly, the first step of our tailoring focus on the centrality metrics more suitable for WDNs studies from an edge perspective. To this aim, we use the edge betweenness (*Girvan and Newman*, 2002) and tailor the closeness, harmonic closeness and the extended neighbour degree centrality metrics considering edges (*Giustolisi et al*., 2020).

The next step is attributing weights to the pipes when computing tailored CNT tools as, for example, the pipes hydraulic resistance. These weights allow of moving the network domain analysis closer to the expected hydraulic behaviour of the system. For example, pipe hydraulic resistances drive the water fluxes.

1.1.2 Directional devices as prior information

The network connectivity structure (i.e., the topology) of a network is described as a graph, that can be undirected (i.e., the adjacency matrix is symmetric) when all flow path directions are unknow, or directed, when the flow path directions are known, e.g., unidirectional devices are installed. Generally, the WDNs topology is analyzed considering indirect networks, i.e., without defining preferential directions for the flow. The presence of unidirectional devices (that correspond to pipes having null conductance in one direction) plays a relevant role in determining the hydraulic state and need to be carefully considered in the analysis. Unidirectional devices introduce a prior information about the flow directions because the hydraulic domain of WDNs can be modified by the presence of devices allowing water to flow only in one direction. This prior is normally considered for pipes close to the reservoir. For directional devices installed into the network, instead, it happens that the shortest paths between two pipes, one upstream and the other downstream of a specific device, can only be traversed in one direction, i.e., that of the device, reducing the number of shortest paths allowed in the direction contrary to that of the device. This information enhances the domain analysis and can be useful to quantify for example, the feasibility of management interventions, in advance with respect to hydraulic modelling.

1.1.3 The intrinsic relevance of nodes as hydraulic information

In WDNs nodes play a different hydraulic role: source nodes, demand nodes and connection nodes. Source nodes are very few nodes in the network and can represent reservoirs, tanks, pumps and any WDN component from which the water is delivered to the network from the mass and/or energy perspective. They represent special *hubs* because all the information starts from them to be transferred to the portion of the network supplied by each of them (*Giustolisi et al*., 2019). Demand nodes represent hydraulic outflows and can assume a different importance depending on the property they serve (e.g., private properties, hospital, school, administrative establishments, etc..). Connection nodes are connections among pipes, for example in crossroads. CNT tools do not account for this difference because the classification findings are based only on topological features of the network, i.e., they refer only to the network connectivity structure, neglecting the intrinsic characteristics of elements (i.e., the hydraulic component). In fact, when computing CNT metrics, the assumption is that all nodes are equal. Such assumption can hide a part of the network information because topology does not contain all the information for defining, for example, the network centrality of elements or the network classification. For WDNs this assumption would involve a twofold mistake: (i) demand nodes have the same relevance of a source node and (ii) source nodes, that should be conceptually hydraulic hubs in WDNs, being generally connected to a single pipe (i.e., feeding the network from the source), have a very low degree of connection, thus assuming less topological importance than demand nodes. Furthermore, the presence of strategic nodes (hospitals, schools, administrative buildings, etc.) is neglected. Nodes have different intrinsic relevance, independent from their topological position and connectivity that needs to be considered and embedded in the analysis.

To enhance the WDNs analysis and their classification, the information about the intrinsic relevance of nodes, i.e., their role in the network, needs to be embedded into CNT tools. This means to couple the hydraulic information with the topological ones to make CNT tools useful for such complex systems.

1.2 Outline

This thesis is organized as follows. In Chapter 2, we recall the basic concepts of Complex Network Theory. In Chapters 3 we focus on water distribution networks: from hydraulic modelling (mass and balance equations) to hydraulic domain (material and spatial peculiarities) showing how classification and analysis of these system using classic CNT tools is not reliable and highlighting the need of a tailoring.

In Chapters 4 we propose a tailoring of the CNT tools. We start with two strategies of tailoring of the centrality metrics. The first tailoring process (*Giustolisi et al.*, 2019; *Simone et al.*, 2020) consists of three conceptual steps: (i) to move from a node-based perspective (typical of usual centrality problems) to an edge-based perspective where pipes are the key components of the network; (ii) to create a WDN-tailored network topology aimed at considering the different hydraulic roles of nodes: source nodes, tanks, and demand nodes, through a strategy based on the presence of fictious nodes connected to each source node. The number of such fictitious nodes depends on the number of demand nodes fed by the source node. In this way, even if the source node is physically connected with one pipe only to the network, its importance increases; (iii) to evaluate edge centralities in the tailored topology for betweenness, closeness, and degree, the more suitable centrality metrics for WDNs, also considering appropriate weights on pipes that contain domain information (e.g., pipe resistance); and (iv) to include the prior information about the flow direction in the pipes where directional devices are installed or for which the hydraulic conditions constrain the flow direction as, for example, next to the reservoirs.

The second tailoring process (*Giustolisi et al.*, 2020) embeds the information about the intrinsic relevance of nodes into the analysis. Such improvement allows to assign a different relevance during the analysis to each node and aims at avoiding the fictitious nodes in WDN original topology in correspondence of source nodes, decreasing the calculation times. The intrinsic relevance for each node, defined as R_n ($n=1$, …, N), is assigned equal to the demand for each demand node and to the sum of demand for the source node. Several experiments conducted on random and regular networks are reported. Finally, the paradigm of the intrinsic relevance of nodes is applied to a social network (the historical Florence family's marriage network), to demonstrate its general validity. It follows the application of the tailoring for WDNs analysis. Firstly, the paradigm of the intrinsic relevance of nodes is applied to real WDNs and results are then compared with those obtained using the previous tailoring (Tailored WDN topology).

Secondly, we report the tailoring process adopted for the modularity index (*Giustolisi and Ridolfi*, 2014 a, b) and presents the development of strategies for DMA design (*Laucelli et al.*, 2017) and pressure sampling design (*Simone et al.*, 2016).

2

Complex Network Theory (CNT)

Contents

2.1 Historical Introduction

CNT is born with Euler more than two century ago with the work untitled "Seven Bridges of Königsberg" (1736). Euler lived in Königsberg, in East Prussia, today Kaliningrad in Russia. The river Pregel with its branches divided the city into four regions, which were connected by seven bridges (see Figure 2).

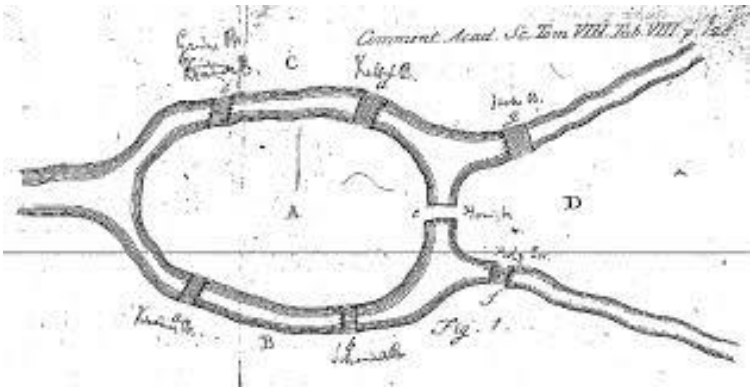


Figure 2. Scheme of Königsberg bridges at the time of Euler. Source *Commentarii academiae scientiarum Petropolitanae*, Volume 6, pp. 216-231.

Curiosity: *The inhabitants of the city were used to walk around the city searching for a path that allowed them to cross all bridges only once.*

Euler addresses the problem as an exercise in *Geometria situs*, a branch of geometry today known as topology. Euler realized that the possibility of finding a path-solution is independent on human ability and from the distance between the points, but rather dependent from the geometric characteristics of the path. He schematized the Königsberg bridge problem abstractly, only using a graph to represent lands and bridges, where each land region was a node and each bridge a connection (link) (see Figure 3).

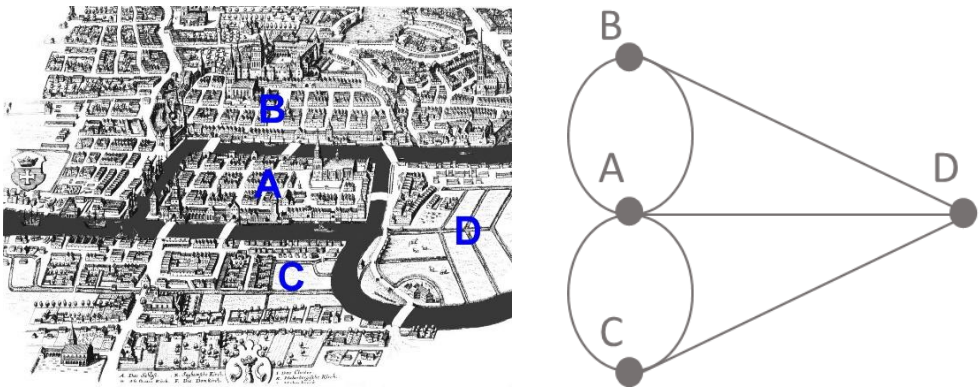
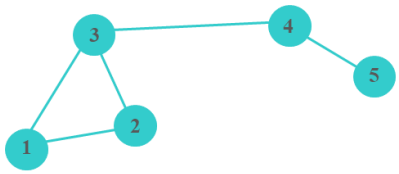
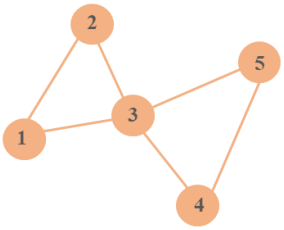


Figure 3. Map of Königsberg with four land areas (A, B, C and D) separated by the river (left). Euler's schematization of Königsberg bridges using graph theory (right).

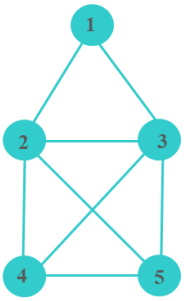
Euler used four nodes, named A, B, C, and D, to represent the land areas separated by the river, connecting among them all land areas that share a bridge. In terms of graph the scheme is reported in Figure 3 (right). The problem was to find a path, called *Eulerian path* (Figure 4-left), between two nodes in the graph where every link appears exactly once. In case the Eulerian path starts and ends at the same graph node, always traversing each graph link exactly once, it would have been a *Eulerian cycle* (Figure 4-right).



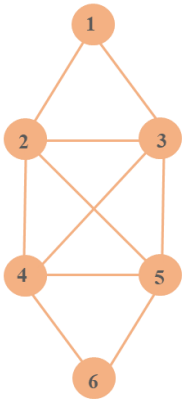
Eulerain Path: 5- 4- 3- 1- 2- 3.



Eulerian Cycle: 1- 3- 5- 4- 3- 2- 1.



Eulerain Path: 4- 5- 2- 3- 1- 2- 4- 3- 5.



Eulerian Cycle: 4- 5- 2- 3- 1- 2- 4- 3- 5- 6- 4.

Figure 4. Left: Example of *Eulerian Path* but no *Eulerian Cycle*. **Right**: Examples of *Eulerian Cycle*.

Euler solved the problem by mathematically demonstrating the impossibility of the solution through the first theorem of graph theory: in a graph with more than two nodes and an odd number of links a path that crosses all nodes only once cannot exist. The solution, in fact, required nodes with an odd number of connections only at the beginning or at the end of the path, admitting at most two.

In order to understand Euler's procedure, it is important to introduce some new concepts and notations: (i) two vertices are adjacent if they are connected by an edge/link and (ii) the number of neighbors of a node i, denoted by k_i, represents its nodal degree. Euler demonstrated that any graph with n nodes/vertices and e edges/links satisfies the relation

$$\sum_{i=1}^{n} k_i = 2e \qquad (1)$$

Hence, for the Königsberg bridges case, the sum of degrees is even (2·7=14) and, consequently, the number of nodes with odd degree is even (4), since in each of the four nodes converges an odd number of links. Euler noticed that a *Eulerian path* was possible only if the graph had all nodes of even degree or if only two of them have odd degree. In the latter case it was necessary to start the path in one of these odd nodes and finish it in the other. Resuming, is n the number of nodes of a graph G with odd degree, if $n > 2$ no *Eulerian path* exists, if $n= 0$ there are *Eulerian paths* starting from any node and if $n= 2$ only one *Eulerian path* exists that starts and arrives in the two odd nodes, respectively. Since the Königsberg Bridge graph had four nodes with odd degree, there was no solution to the problem, that is, no *Eulerian path* existed.

Euler, with this first demonstration, established the novel mathematic field named graph theory, which represents today a branch of mathematics, which guides the basis of our thinking about networks. *Cauchy* (1813) and *L'Huillier* (1861) generalized Euler's concept establishing the basis for the topological studies.

2.2 Basic concepts

A network (or graph) is a mathematical object $G= (V, E)$ where $V = \{1, . . ., V\}$ is the set of vertices (or nodes) of G, and $E = \{e_1, . . ., e_m\}$ is the set of edges (or links) of G (see Figure 5). The size of the graph G, which corresponds to the size of the set E, is indicated by $L= |E|$ and the size of the set V is indicated by $N= |V|$. The most part of the complex systems of the real world have hundreds of thousands or even millions of vertices, which makes their visualization and study not easy. To facilitate their visualization and analysis, real networks are generally represented by sparse graphs, i.e., $|E| = O(|V|)$.

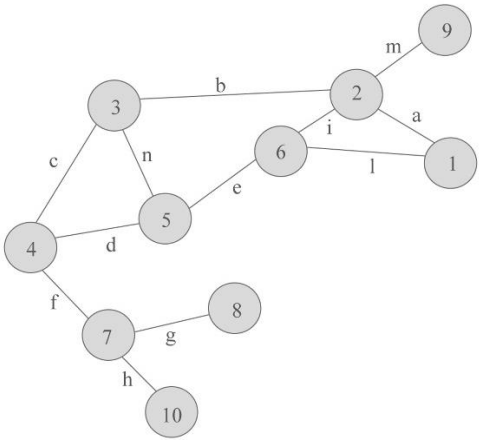


Figure 5. Example of a graph G with ten vertices represented by grey circles ($v= 1, …, 10$) and 12 edges represented by grey lines ($e= a, …, n$).

Each vertex *v* can represent several components in different real systems (nodes, crossroads, actors, people, neurons, etc.) and each edge *e* represents the connection between two components (links, pipes, roads, ties, synapses, etc.). Two vertices (*i*, *j*) are adjacent if there is an edge between them, i.e., if $e = (i, j) \in E$.

A graph $G = (V, E)$ is indirect if *E* is a set of unordered couples, i.e., edges (links) have not a direction (Figure 6-left) or is direct if *E* is a set of ordered couples, i.e., edges (links) have a direction (Figure 6-right). For indirect graphs, the relationship between nodes is symmetric, i.e., given two nodes $(i, j) \in G$, $e_{ji} = e_{ij}$, while for direct graphs $e_{ji} \neq e_{ij}$.

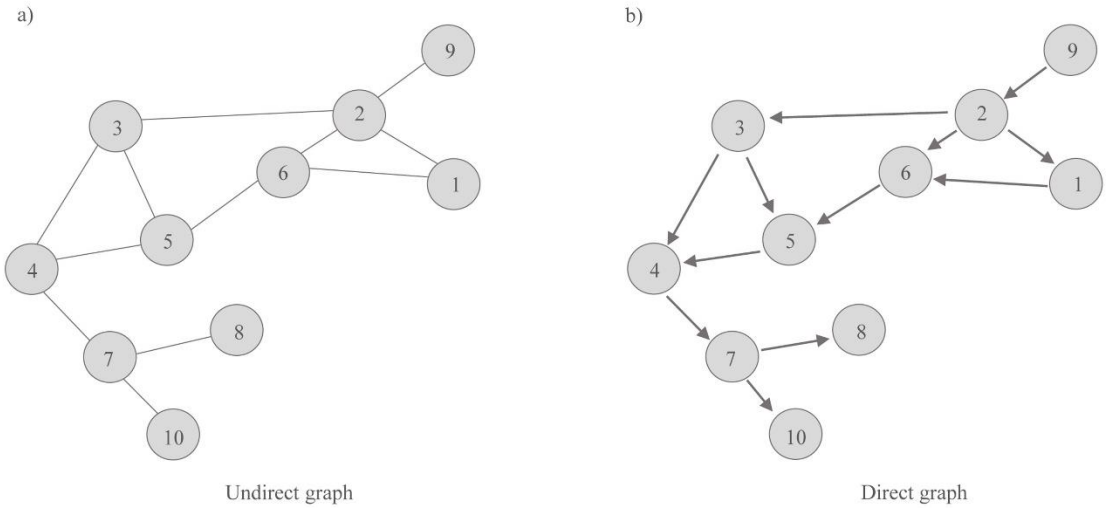


Figure 6. Left. Undirect graph, $e_{ji} = e_{ij}$ for each edge. For example, vertex 3 is connected to vertex 4 and vice versa, i.e., $a_{34} = a_{43}$. **Right.** Direct graph, $e_{ji} \neq e_{ij}$ for each edge. For example, vertex 3 is connected to vertex 4 but the vertex 4 is not connected to the vertex 3, i.e., $a_{34} \neq a_{43}$.

A *path* in a graph *G* is a sequence of vertices $\{v_1, v_2, \ldots, v_n\}$ connected by edges. Each path contains *v*+1 vertices and *v* edges. The *length* of a path is the number of traversed edges along the path. The path with the minimum number of edges between two vertices (*i*, *j*) represent the shortest path, i.e., a *shortest path* is a path between two vertices such that no shorter path exists. When *i* and *j*, i.e., starting and ending nodes of a path, coincide, their distance is 0, i.e., $d(i, j) = 0$, while when *i* and *j* are unreachable, their distance is defined to be infinity, i.e., $d(i, j) = \infty$.

Figure 7 shows, for the graph of Figure 5, all possible paths between vertices 1 and 5. One shortest path exists, and it belongs the vertices {*1*, *6*, *5*} traversing the edges *h* and *e*.

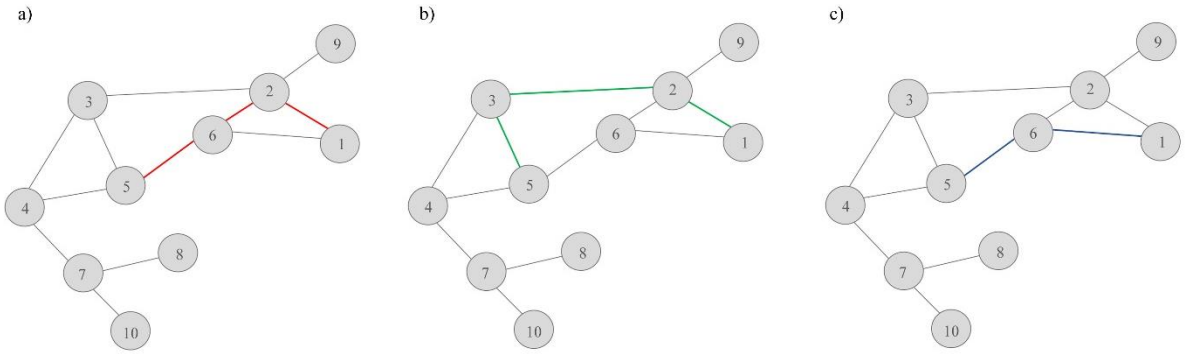


Figure 7. a) Path in red between vertices 1 and 5 crossing vertices 2 and 6. The length of the path is equal to 3 {*a*, *i*, *e*}. **b)** Path in green between vertices 1 and 5 crossing vertices 2 and 3. The length of the path is equal to 3 {*a*, *b*, *c*}. **c)** Path in blue between vertices 1 and 5 crossing the vertex 6. The length of the path is equal to 2 {*h*, *e*}. The latter is the shortest path between the vertices 1 and 5.

The average of the shortest paths between all the vertices in the network is defined as *Average path length L*,

$$L = \langle d_{ij} \rangle = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij} \qquad (2)$$

where N is the number of vertices in G and d_{ij} is the shortest network distance between all the vertices. The greatest shortest path refers to the *diameter* (D_{max}) of the graph G.

A *complete graph* $G = (V, E)$, or a *Clique*, is a graph where each pair of vertices is connected by an edge. The complete graph on n vertices is denoted by K_n. Figure 8 reports a complete graph K_6.

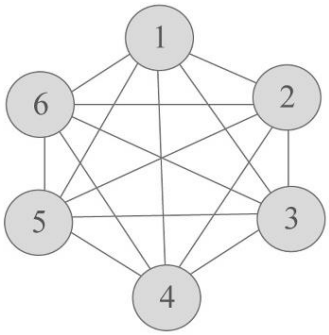


Figure 8. A complete graph G with 6 vertices (K_6). Each vertex is connected to all the other 5 vertices.

The *clustering coefficient C* (*Watts and Strogatz*, 1998) measures the relationship between the number of edges E connected to their adjacent vertices and the possible number of edges between these latter. It is based on the number of triangles present in the graph:

$$C_i = \frac{2t_i}{k_i(k_i - 1)} \qquad (3)$$

where t_i denotes the number of triangles around node i. The value of C_i varies between 0 (neighbors of a vertex are not connected to each other) and 1 (neighbors of a vertex are all connected to each other). Figure 9 shows the *clustering coefficient* C_i for the vertices 6 and 7 of the graph of Figure 5. The global clustering coefficient $C(G)$ of the graph is the average of the local values C_i:

$$C(G) = \frac{1}{n}\sum_{i=1}^{n} C_i \qquad (4)$$

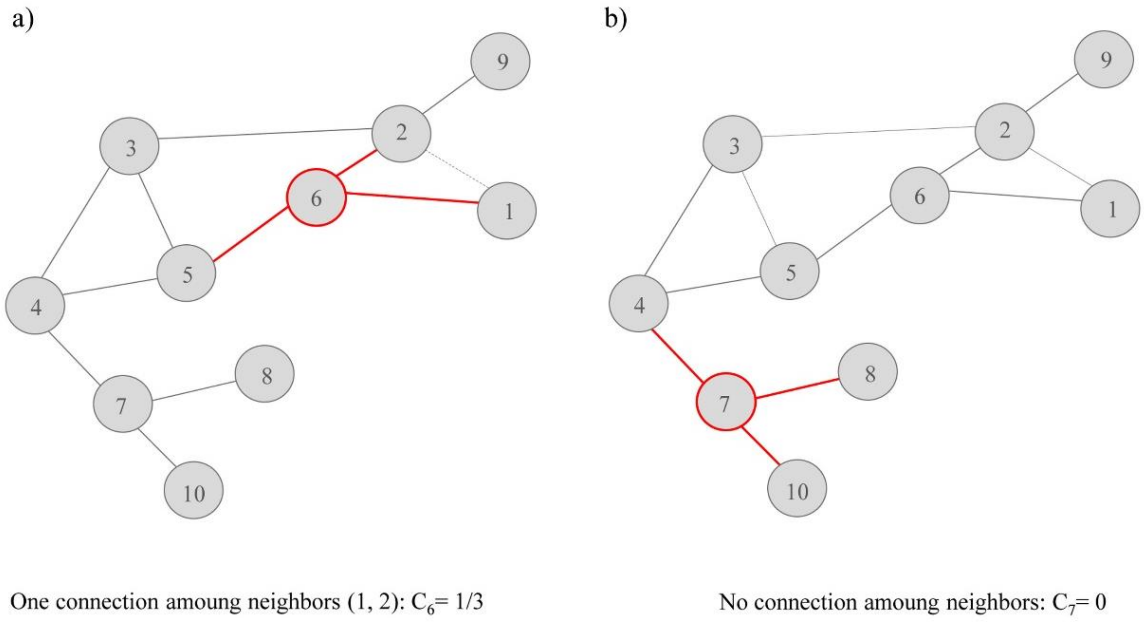


Figure 9. **a)** Clustering Coefficient C_i for the vertex 6. Three connections are possible between the vertices connected to the vertex 6 (2-5; 5-1 and 1-2) but only one exists (see dashed link between vertices 1 and 2). Then C_6= 1/3. **b)** Clustering Coefficient C_i for the vertex 7. Three connection are possible between the vertices connected to the vertex 7 (4-8; 4-10 and 10-8) but they do not exist. Then C_7= 0.

A *weighted graph G* associates a label (weight) to every edge in the graph. A *weighted graph* is represented by a tuple (V, E, w) where V and E are sets of vertices and edges of the graph G, respectively, and w: $E \rightarrow \mathbf{R}$ is a function that associates a weight $w(e)$ to each edge $e \in E$. Several networks are represented by edges having a weight, usually a real number, that can represent a cost, a distance, a capacity, etc… Figure 10 reports an example of weighted graph.

It is important noting that *shortest paths* can also be measured in weighted graphs. The shortest path between two vertices (i, j) corresponds for weighted graph with the lowest weight and not with the least number of edges (see Figure 11- left).

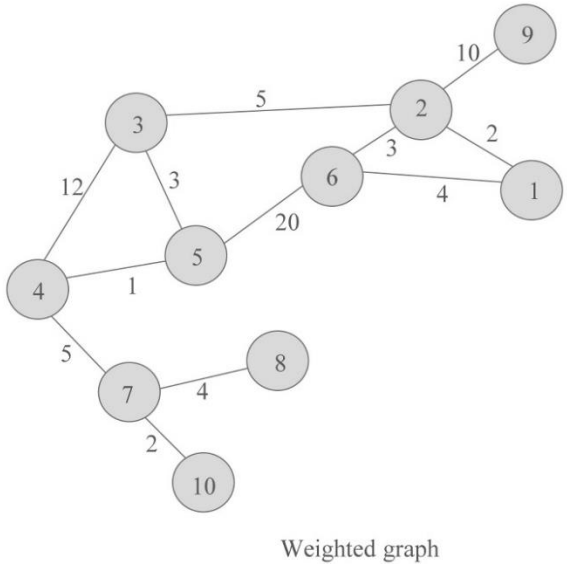


Figure 10. Weighted graph. Each edge has a weight that can correspond to several features (length, cost, etc.).

A *tree* is a graph $G= (V, E)$ without any cycles (undirected acyclic graph) and where each couple of vertices is connected by only one path. A *spanning tree* $T= (V, E')$ of a graph $G= (V, E)$ is a tree that contains the set of vertices V and a subset of edges E' of the graph G. If v is the number of vertices in the graph, each spanning tree has $v-1$ edges. A *minimum spanning tree* (*MST*) or *minimum weight spanning tree* is the spanning tree with the minimum weight in the graph, i.e., the spanning tree whose sum of edge weights is as small as possible (see Figure 11-right).

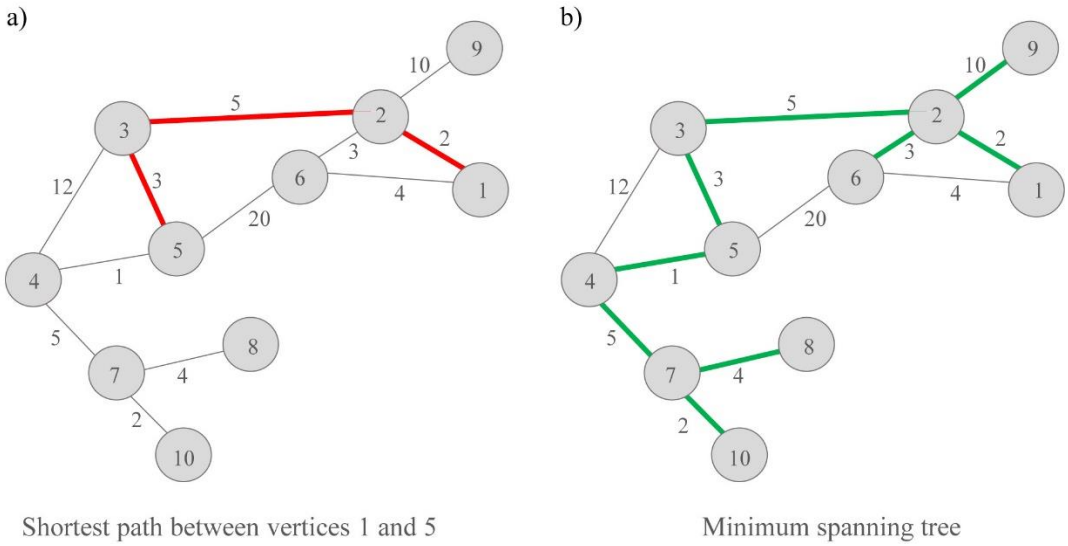


Figure 11. **a)** The weighted path between vertices 1 and 5 is equal to 10 (crossing vertices 2 and 3), 25 (crossing vertices 2 and 6) and 24 (crossing the vertex 6). The shortest path is the first one {1, 2, 3, 5}, differently from what obtained for the non-weighted graph (see Figure 5), where the shortest path was the last one {1, 6, 5}, with a minimum number of edges. **b)** A minimum spanning tree starting on node 5.

A *clique* is a subset of vertices $C \subseteq V$ of an undirected graph $G = (V, E)$ such that every two distinct vertices in the subgraph are adjacent, i.e., the subgraph is complete. The complete graph of Figure 8 is a clique with 6 vertices.

2.3 Graph representation

The *adjacency matrix* **A** is the most used representation of a graph *G* and indicates whether couples of vertices are connected or not in the graph. The *adjacency matrix* for a graph *G* of order *n* is represented by a square matrix of order *n* where each row (column) corresponds to a vertex of the graph. For a simple graph, the *adjacency matrix* corresponds to a logical matrix where elements are all either 0 or 1. The *adjacency matrix* $\mathbf{A}(G) = (a_{ij})$ of a graph $G = (V, E)$ is defined by the conditions,

$$a_{ij} = \begin{cases} 1 \text{ if } \{i, j\} \in E \\ 0 \text{ if } \{i, j\} \notin E \end{cases} \quad (5)$$

and for $\forall i \in \{1, \ldots, n\}$ we get that $a_{ii} = 0$ (i.e., no self-loops exist).

For undirected graphs, the *adjacency matrix* is symmetric with respect to the main diagonal, while for direct graph it is not necessarily. The *adjacency matrix* of a weighted graph reports the weights of the edges between vertices in the cells.

The *incidence matrix* **B** of a graph indicates the relationship between vertices and edges, i.e., it shows whether couples of vertices-edges are connected or not. Each row corresponds to a vertex and each column corresponds to an edge of the graph. For a graph *G* the *incidence matrix* $\mathbf{B}(G) = (b_{i,e})$ corresponds to a logical matrix where elements are all either 1, if the vertex *i* and edge *e* are incident, or 0 otherwise. The sum of each column is equal to 2 because each edge has a vertex connected to each end.

The *incidence matrix* of a graph *G* is related to the *adjacency matrix* of its line graph *L*(G),

$$\mathbf{A}(L(G)) = \mathbf{B}(G)^T \mathbf{B}(G) - 2\mathbf{I}_m \quad (6)$$

where $\mathbf{A}(L(G))$ is the adjacency matrix of the line graph of *G*, $\mathbf{B}(G)$ is the incidence matrix, $\mathbf{I}_m$ is the identity matrix of dimension *m* and the apex *T* indicates the transpose of the matrix. The discrete Laplacian can be obtained from the oriented incidence matrix **B** (*G*) as **B** (*G*) **B** $(G)^T$.

The *degree* of a vertex is the number of edges connected to the vertex. The *degree* can be easily calculated by the expression:

$$\sum_{j=1}^{n} a_{ij} \quad (7)$$

The *degree* for indirect networks in terms of the adjacency matrix is

$$k_i = \sum_{j=1}^{n} \mathbf{A}_{ij} \tag{8}$$

where the sum is over all nodes in the network.

A node with a high degree, i.e., more adjacent nodes, can be considered as being well connected and a node with a relatively low degree, i.e., few adjacent nodes, can be considered weakly connected. The minimum, the maximum, and the average degree of the vertices of G represent standard characterization metrics in graph theory. The most studied vertex characteristic is the average degree,

$$\langle k \rangle = \frac{1}{n} \sum_{i=1}^{n} k_i \tag{9}$$

The *degree matrix* is a diagonal matrix that reports the information about the degree of each vertex, i.e., its number of connections. The *degree matrix* $\mathbf{D}(G) = (a_{ij})$ of a graph G is defined by the conditions,

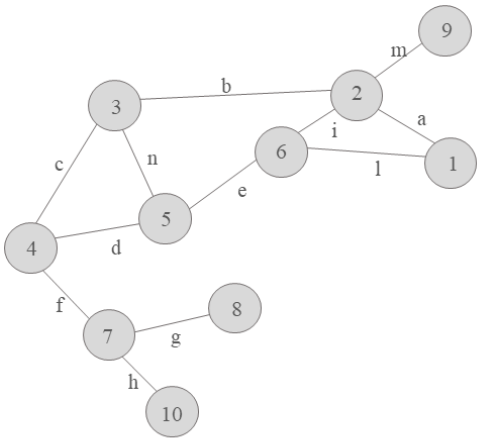
$$d_{i,j} = \begin{cases} \deg(v_i) & \textit{if} \quad i = j \\ 0 & \textit{otherwise} \end{cases} \tag{10}$$

The *Laplacian matrix* of a graph G can be used to find many useful properties of a graph, as for example its number of spanning trees. The Laplacian matrix $\mathbf{L}_{nxn}$ of a graph $G = (V, E)$, where $|V| = n$, is defined as:

$$\mathbf{L} = \mathbf{D} - \mathbf{A} \tag{11}$$

Where **A** is the adjacency matrix and **D** the diagonal matrix of degrees of the graph G. The elements of $\mathbf{L}_{nxn}$ are given by:

$$\mathbf{L}_{i,j} = \begin{cases} \deg(v_i) & \textit{if} \quad i = j \\ -1 & \textit{if} \quad i \neq j \ \textit{ and } \ v_i \ \textit{ is adjacent to } v_i \\ 0 & \textit{otherwise} \end{cases} \tag{12}$$



Simple graph *G*

	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	0	1	0	0	0	0
2	1	0	1	0	0	1	0	0	9	0
3	0	1	0	1	1	0	0	0	0	0
4	0	0	1	0	1	0	1	0	0	0
5	0	0	1	1	0	1	0	0	0	0
6	1	1	0	0	1	0	0	0	0	0
7	0	0	0	1	0	0	0	1	0	1
8	0	0	0	0	0	0	1	0	0	0
9	0	1	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	1	0	0	0

a) Adjacency matrix ***A***

	a	b	c	d	e	f	g	h	i	l	m	n
1	1	0	0	0	0	0	0	0	0	1	0	0
2	1	1	0	0	0	0	0	0	1	0	1	0
3	0	1	1	0	0	0	0	0	0	0	0	1
4	0	0	1	1	0	1	0	0	0	0	0	0
5	0	0	0	1	1	0	0	0	0	0	0	1
6	0	0	0	0	1	0	0	0	1	1	0	0
7	0	0	0	0	0	1	1	1	0	0	0	0
8	0	0	0	0	0	0	1	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	1	0
10	0	0	0	0	0	0	0	1	0	0	0	0

b) Incidence matrix ***B***

	1	2	3	4	5	6	7	8	9	10
1	2	0	0	0	0	0	0	0	0	0
2	0	4	0	0	0	0	0	0	9	0
3	0	0	3	0	0	0	0	0	0	0
4	0	0	0	3	0	0	0	0	0	0
5	0	0	0	0	3	0	0	0	0	0
6	0	0	0	0	0	3	0	0	0	0
7	0	0	0	0	0	0	3	0	0	0
8	0	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	1	0
10	0	0	0	0	0	0	0	0	0	1

c) Degree matrix ***D***

	1	2	3	4	5	6	7	8	9	10
1	2	-1	0	0	0	-1	0	0	0	0
2	-1	4	-1	0	0	-1	0	0	9	0
3	0	-1	3	-1	-1	0	0	0	0	0
4	0	0	-1	3	-1	0	-1	0	0	0
5	0	0	-1	-1	3	-1	0	0	0	0
6	-1	-1	0	0	-1	3	0	0	0	0
7	0	0	0	-1	0	0	3	-1	0	-1
8	0	0	0	0	0	0	-1	1	0	0
9	0	-1	0	0	0	0	0	0	1	0
10	0	0	0	0	0	0	-1	0	0	1

d) Laplacian matrix ***L***

Figure 12. Simple graph *G* with 10 vertices and 12 edges. **a)** Adjacency matrix **A** of *G* contains 10 rows (columns) as the number of vertices in the graph and it is symmetric with respect to the main diagonal, being the graph undirected. **b)** Incidence matrix **B** of *G* contains 10 rows as the number of vertices and 12 columns as the number of edges. **c)** Degree matrix **D** of *G* contains the number of connections of each vertex (degree) on the main diagonal. **d)** The Laplacian matrix **L** of *G* contains the degree of each vertex on the main diagonal and information about connections between all vertices of *G* in the remaining part.

2.4 Network models and common properties

The degree of a vertex represents a relevant information for classifying networks, referring to the probability of distribution of the number of edges incident each vertex of the network. The interest in classifying real systems according to its connectivity structures is related to capturing its emerging behaviour. Several models have been developed in order to simplify the study of complex systems (*Newman et al.*, 2006; *Newman*, 2010), whose connectivity structures is defined by using the nodal degree, the average path length, the clustering coefficient and the probability of connection.

The first study about the nodal degree distribution of real networks (*Erdos and Rény*, 1959; 1960) refers to *random* networks (see Figure 13), i.e., networks with a nodal degree that is randomly distributed around an average value and characterized by a high homogeneity. Random networks allow to capture real networks features better than regular networks, which are characterized by an absolute homogeneity (constant degree of internal nodes) with shortest paths between couple of nodes too large, while random networks are characterized by lower and more realistic values of the shortest paths with respect to the regular ones (*Milgram*, 1967).

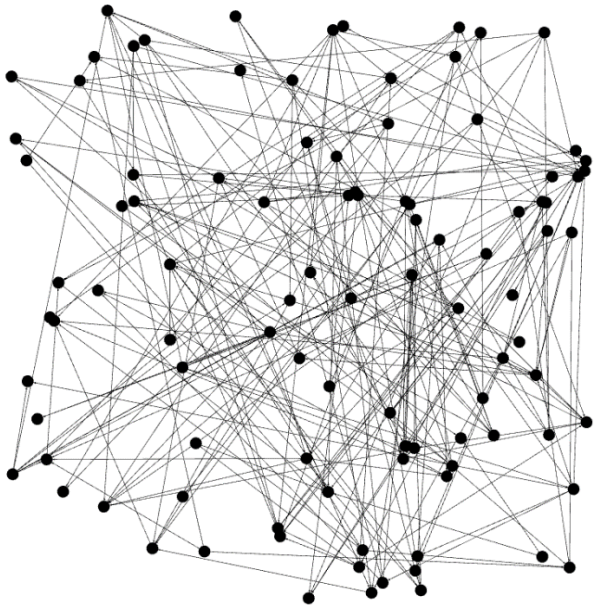


Figure 13. Random network. Source: http://corelab.mae.ufl.edu/random.html.

A mixed behaviour between *regular* and *random* networks characterizes the *small world* networks (see Figure 14), proposed by *Watts and Strogatz* (1998), based on the so-called *small world effect* (*Milgram*, 1967; *Travers and Milgram*; 1069). The most part of the vertices are not close to each other, but the neighbours of a given vertex are probably close to each other and most vertices can be reached by any other vertex with a small number of steps.

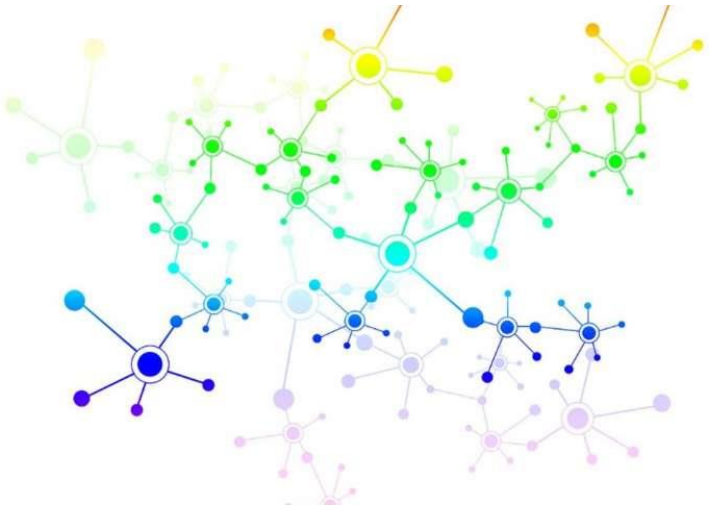


Figure 14. Small world network. Color encoding represents nodes belonging to the same cluster. Source: *Norcini Pala,* 2020.

Later on, *Barabasi and Albert* (1999) proposed the *scale-free* networks (see Figure 15) in order to describe real networks characterized by non-homogeneous nodal degree distributions, i.e., networks where many nodes have a low degree and few nodes (called *hubs*) have a high degree.

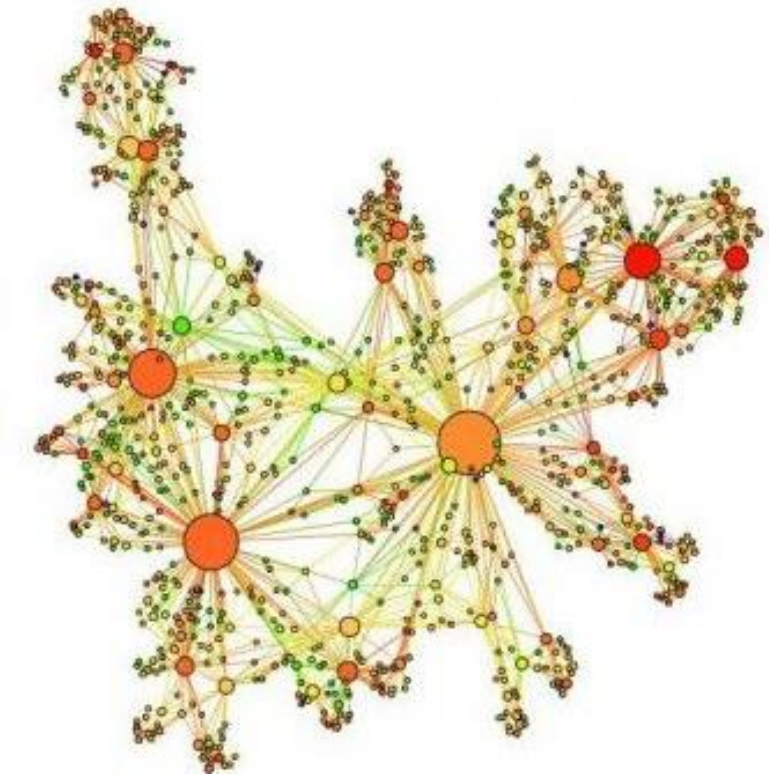


Figure 15. Small world network. Researchers from Queen Mary University of London and Karlsruhe Institute of Technology.

Networks with a homogeneous degree distribution (*regular*, *small world* and *random* networks) present a significant structural resistance to both random failures and intentional threats, while networks with a heterogeneous degree distribution (*scale free* networks) show a very high structural resistance to random failures but a weak resistance to intentional threats (*Albert et al.*, 2000). In this sense, the classification of real networks by associating their degree distribution to the Poisson (homogeneous degree distribution) or Pareto (heterogeneous degree distribution) models is useful to assess network vulnerability with respect to random failures and intentional threats.

2.4.1 Random networks

Erdös and Rényi (1959) introduced the *random* networks, i.e., networks where many nodes have a similar number of connections and the degree distribution is randomly distributed around an average value. For random networks, the existence of an edge between a pair of nodes has probability *p* and the average degree is:

$$\langle k \rangle = p \cdot (n-1) \propto p \cdot n \tag{13}$$

the number of random connections between them is:

$$\langle L \rangle = \frac{1}{2} p \cdot n \cdot (n-1) = \frac{1}{2} \langle k \rangle \cdot n \tag{14}$$

and the probability of a node having degree *k* is:

$$P(k) = \binom{n}{p} p^k (1-p)^{n-\langle k \rangle} \cong \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!} \tag{15}$$

The equation (14) highlights how in *random* networks all nodes have principally the same number of links, i.e., a great number of nodes with a similar degree. The high homogeneity of these networks results in a Poisson distribution of degrees (see Figure 17).

Random networks are proposed when looking for models more responsive to real systems than regular networks, i.e., network having a regular topology. Moreover, since randomness (i.e., elements arranged randomly) does not represent, generally, the main feature of most real systems, new models to capture the real behavior of complex systems are proposed: *small world* and *scale free* networks.

2.4.2 Small World networks

Small world networks (*Watts and Strogatz*, 1998) based on Milgram's experiment (*Milgram*, 1967; *Travers and Milgram*, 1969) on six degrees of separation of social networks, are

characterized by the presence of nodes that can reach other nodes through a small number of edges. Milgram Stanley tested the existence of *shortest paths* in social networks, persuaded that the typical distance between actors (i.e., the number of vertices in a social network) was small. He attributed this phenomenon to the growing globalization that led to an increasingly interconnected world. Therefore, a small-world network refers to an ensemble of networks in which shortest paths, i.e., the distance between vertices, increases sufficiently slowly as a function of the number of vertices in the network.

Curiosity: *Milgram's most famous experiment recalls the famous six degrees of separation. He sent 96 packages to randomly selecting people residing in Omaha, NE, (USA) from a telephone directory. Each package contained a leaflet from Harvard University (where Milgram worked) and instructions for recipients. The goal was for each recipient to attempt to deliver the package to a target actor, a Milgram's friend who lived in Boston, MA, USA. The only information about the target actor were the name, address, and job (agent of exchange). Each recipient had to send the package to someone who he believed could meet Milgram's friend or know someone who lived in Boston or that for work purposes had more chances to find him. The instructions were repeated equally for each recipient. Milgram asked the recipients to record each step of the path in the package. The average number of steps for the paths that reached the goal (18 of the 96) was 5.9. From this experiment it is customary to state that there are no more than 6 steps between each pair of people in the world, i.e., 6 degrees of separation.*

Modern versions report much lower success rates (Dodds, 2003).

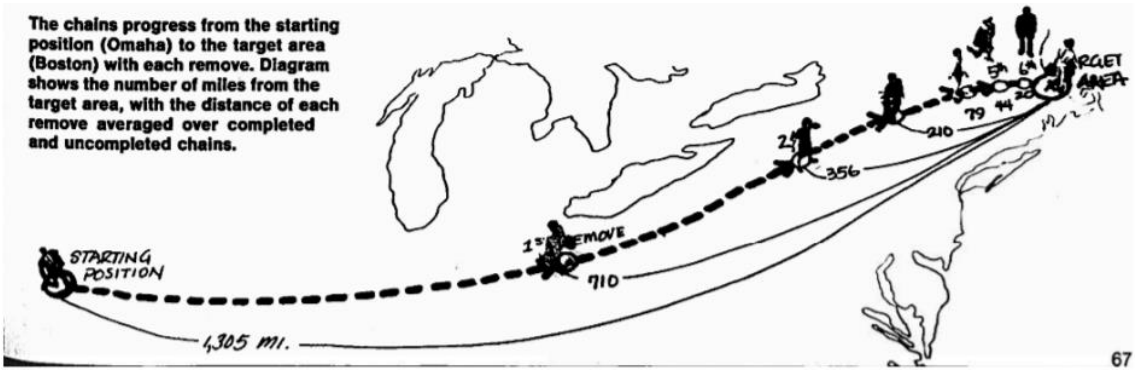


Figure 16. Image from Milgram's original article in Psychology Today, showing a "composite" of the successful paths converging on the target person. Each intermediate step is positioned at the average distance of all chains that completed that number of steps. Source: *Easley and Kleinberg*, 2010.

The most famous *small world model* was proposed by *Duncan Watts and Steve Strogatz* (1998), with the aim to describe networks with both small mean geodesic path lengths and

significant local clustering coefficient. In fact, despite the large size of real system, *small world* networks have the average distance L between two randomly chosen vertices, i.e., the number of edges along the shortest path connecting them, that grows proportionally to the logarithm of the number of nodes N in the network,

$$L \propto \log N \tag{16}$$

The clustering coefficient of *small world* networks, instead, is generally high, differently than for *random* networks, where it is considerably small since the edges are randomly distributed. The *Watts and Strogatz*'s model (1998) conciliate the existence of a large clustering coefficient with a characteristic path length. They started from a ring lattice with N nodes in which every node was connected to its neighbours and then each edge of the lattice was randomly rewired with probability p. The procedure covered from regular to random networks, with a Poisson distribution of degree and a probability p of connections between two nodes (see Figure 17). In fact, although *small world* networks are not completely *random* networks, a Poisson distribution of degree can model them, being characterized by a certain level of homogeneity, which is lower than regular and higher than random networks. For *small world* networks, the probability p is greater than the null value of regular networks but rather lower than values of random networks. The small world effect in the network is favoured by the new connections.

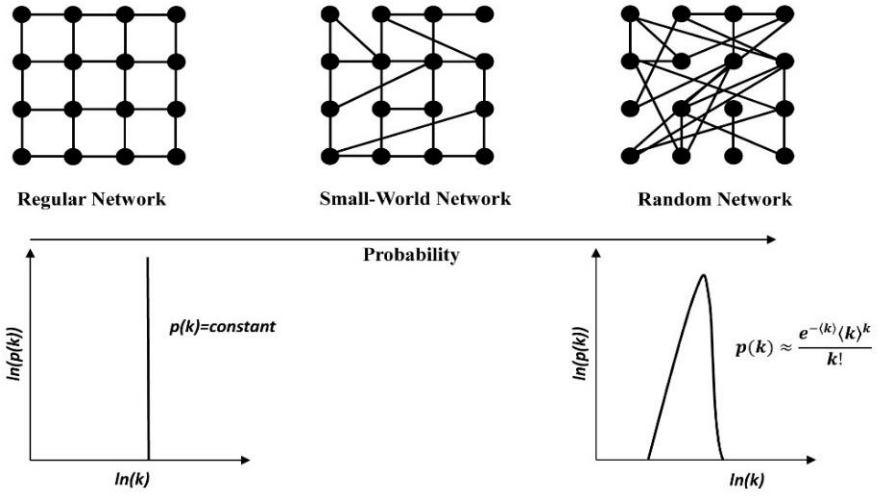


Figure 17. **a)** A regular network in which each internal vertex has the same number of connection e=4. The absolute homogeneity of internal vertices produces a constant degree distribution for the internal vertices ($p(k)$= constant). **b)** A small world network is created by removing each edge with uniform probability p and rewiring it to produce an edge between a couple of vertices that are chosen uniformly at random. **c)** Random network. The high homogeneity of vertices produces a degree distribution randomly distributed around an average value (Poisson distribution). Source: *Giustolisi et al*., 2017.

The Poisson distribution models *random* networks and exhibits a peak value at <*k*> and a high probability of nodal degree around <*k*>, i.e., a high homogeneity of the degree distribution, and the nodal degree distribution of *small world* networks has essentially the same features as the *random* ones. The Poisson distribution degenerates to a single degree (considering only internal nodes) for the case of *regular* networks, characterized by an absolute homogeneity. Finally, the *small world* networks introduced by *Watts-Strogatz* was obtained starting from a ring lattice regular network (Figure 18), increasing the probability of connection, *p*.

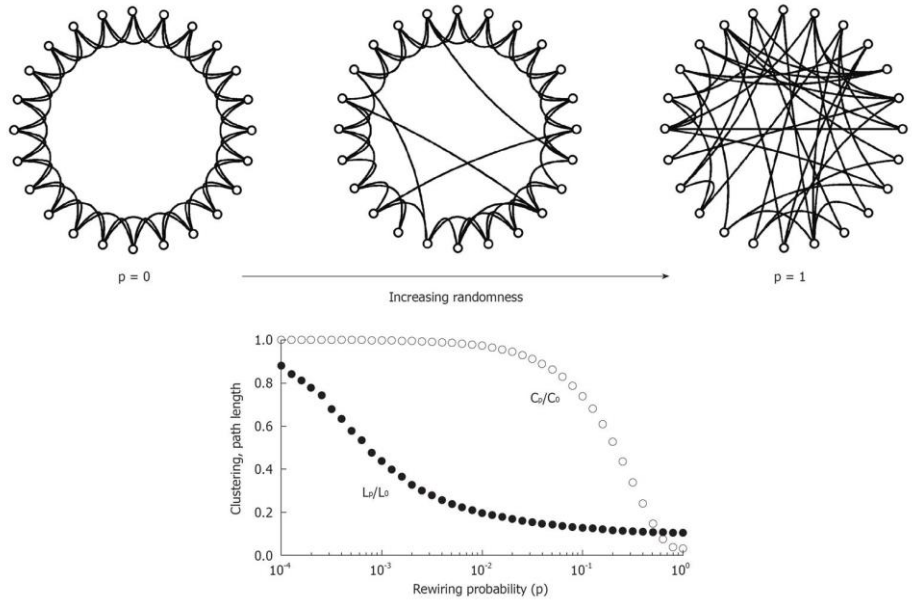


Figure 18. From Watts-Strogatz (1998), the explanation of *small world networks* with respect to *regular* and *random networks*. **a)** Ring lattice. Starting from this configuration, connections are randomly rewired with a given rewiring probability *p*. For $p = 0$ the network does not change its regular lattice topology while for $p = 1$ the network become random. Intermediate values of *p* result in small world networks, i.e., networks with characteristics mixed between regular and random. **b)** Clustering coefficient C_p and the Average path length L_p, both normalized by clustering C_0 and average path length L_0 of the ring lattice regular network. When the rewiring probability *p* gives networks with clustering coefficient like that of regular networks and average path length like that of the random networks, we obtain small-world networks. Source *Giustolisi et al., 2017*.

2.4.3 Scale-free networks

Barabasi and Albert (1999) proposed a new network model (Figure 19), named *scale free*, according to which the degree distribution follows a power law distribution. The *scale free* networks are characterized by a high clustering coefficient and many short paths between all couple of vertices, i.e., small world features. In other words, few nodes having a much higher degree than many nodes, named *hubs*, characterize *scale free* networks.

The substantial difference with respect to the previous network models resides in three main factors: growth, preferential attachment, and fitness. Differently from *random* and *small world* models, which assume a fixed and constant number *N* of nodes randomly connected or rewired in the network, *Albert and Barabasi* proposed a model for real-world networks that allowed of increasing the number of nodes over time, presenting ever-larger systems (e.g., WWW grows in time by the addition of new web pages). Whit regard to the preferential attachments, instead, *small world* and *random* networks assume that the probability that two nodes are connected is random, while in most real networks, new nodes prefer to connect to the highly connected nodes in the network, i.e., prefer to be connected to a node rather than another in the network depending on the nodal degree. Finally, the *fitness model* (*Bianconi and Barabasi*, 2001) refers to the evolution of a network according to which the *fitness* of a nodes influences how the links between nodes change over time. Fitter nodes attract more edges than less fit nodes. For example, a web page whose network structure is realized using the fitness model, will more likely include hyperlinks to popular documents with already high degrees (preferential attachment), because such highly connected documents are easy to find and thus well known (*Chapela et al*., 2015). Therefore, the model algorithm of scale free networks (i) starts with a small number of nodes, (ii) the number of nodes in the network increases, (iii) the probability of the new node to be connected to an existing node depends on the degree and (iv) nodes with high fitness can attract more than nodes with high degree, despite the high ability of hubs to take edges added to the network. It is to report that also *scale free* networks can have a low degree of separation of nodes, although they are not *random* networks, similarly to *small world* (*Xiao and Guanrong*, 2003) because the hubs reduce the path lengths among nodes. The formulation of the Pareto (or power law) model for degree distribution is expressed as a polynomial.

$$P(k) \approx k^{-\gamma} \tag{17}$$

where γ represents the power-law exponent and it is a constant (ranging from 1.5 to 3) (*Newman*, 2010; *Barthelemy*, 2011). A power law distribution is a relatively slow decreasing function without peak at its average value. Furthermore, power law model is scale invariance, i.e., it does not change if scales (length, weight, or other variables) are multiplied by a common factor.

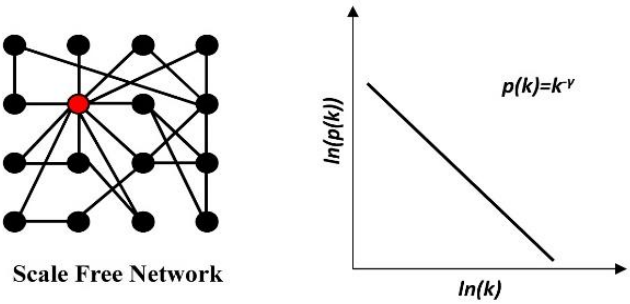


Figure 19. Left. A scale-free network. **Right.** The degree distribution follows a power law, at least asymptotically, indicating that many vertices have a low degree, and few vertices have a high degree (hubs). Source *Giustolisi et al.*, 2017.

A typical value for the degree power-law exponent in the majority of real networks is $2 \leq \gamma \leq 3$. Resuming, the Barabási-Albert model produces a degree power-law distribution with exponent $\gamma = 3$, meanwhile the Watts-Strogatz and the Erdös-Rénji follow a Poisson distribution.

2.5 Assortativity

Assortativity is the preference, for a node in a network, to attach to other nodes that are similar in terms of degree (*Newman*, 2002). It represents a degree-degree correlations measure between nodes that tend to attach to similar degree nodes (assortativity) or different degree nodes (disassortativity). I.e., a network is said to be assortative when high degree nodes are connected to other high degree nodes and *vice versa*, while a network is said to be disassortative when high degree nodes are connected to low degree nodes and *vice versa* (see Figure 20).

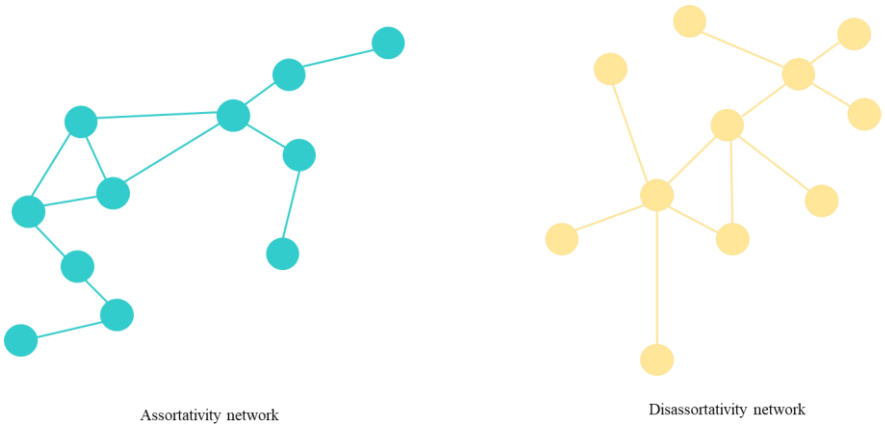


Figure 20. Example of assortativity. **Left)** Assortativity network where high-degree nodes attach to other high-degree nodes and *viceversa*. **Right)** Disassortativity network where hubs connect to low degree nodes.

In terms of network models, assortative networks follow a Poisson distribution, considering most edges connected to nodes that exhibit similar degrees and disassortative networks follow a Pareto distribution, considering most edges connected to nodes that exhibit different degrees.

For instance, social networks (e.g., friendship network) can be highly assortative (see Figure 21) because nodes tend to be connected with other nodes with similar degree values while biological networks can be highly disassortative (see Figure 22) because high degree nodes tend to attach to low degree nodes (*Newman*, 2002).

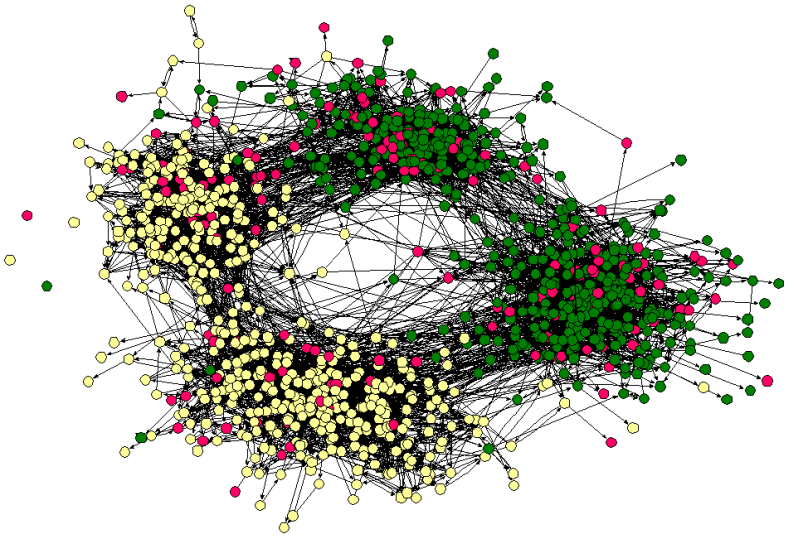


Figure 21. Assortativity network: well-connected nodes tend to join other well-connected nodes. Source: *Moody*, 2001.

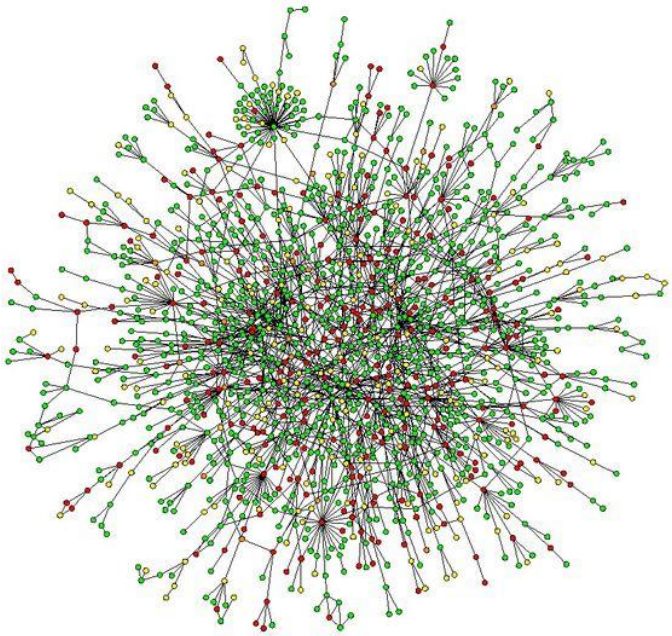


Figure 22. Disassortativity network: well-connected nodes join to a much larger number of less-well-connected nodes. Source: *Jeong, et al*., 2001.

Assortativity is defined with ρ and ranges between -1 and 1. The degree assortativity, identified as ρ_D, provides information about structure and behavior of a network (when subject to random or targeted attack, virus spread, etc.). The original definition of assortativity proposed by *Newman* (2002) for non-weighted and non-directed networks, based on the correlation

between random variables, has been successively quantified by the Pearson correlation coefficient (*Van Mieghem*, 2014) and extensively analyzed by *Piraveenan* (2010), whose proposed studies of networks perfectly assortative or perfectly disassortative. A structural metric that characterizes the degree similarity of adjacent nodes is the degree-degree correlation. This correlation is characterized by the assortativity *r* and defined as the Pearson correlation coefficient:

$$r = \frac{(\langle ij \rangle - \langle i \rangle \langle j \rangle)}{(\langle i^2 \rangle - \langle i \rangle^2)} \tag{18}$$

where *i* and *j* are the degrees at the two ending nodes of an edge and the ⟨ ⟩ notation represents the average over all edges. If the assortativity coefficient of a network is negative, a hub tends to be connected to non-hubs, and *vice versa*. Therefore, for $r > 0$ the network exhibits an assortative mixing pattern, for $r < 0$ a disassortative mixing pattern and for $r=0$ the network a neutral degree-mixing pattern (*Xu et al.*, 2009).

2.6 Centrality metrics

Complex Network Theory (CNT) studies the relevance of elements in networks using centrality metrics. The CNT approach for studying network's centrality is based on the idea that critical components (nodes and edges) stand between others, playing the role of an intermediary in the interactions or in the communications. Greater is the number of connections and paths in which a node or edge participates, higher is its importance in the network.

Many centrality metrics are proposed over the years to clarify the concept of importance, in terms of centrality, detailing components and interrelationships among the various elements of networks. *Freeman* (1977-1979) is among the first to propose a comprehensive study of the vertex topological centrality, by identifying three main metrics: (i) the degree as capability to spread information because of the number of local connections (*Freeman*, 1977), (ii) the position in the network as importance of the vertices for the global exchange of information into the entire system (*Freeman*, 1977), and (iii) the connections which identify preferential shortest paths into the network (*Freeman*, 1979; *Giustolisi et al*., 2020). Several centrality metrics are proposed later by many researchers in order to evaluate the most central element in the network with respect to different physical phenomena, (e.g., see *Newman*, 2018); examples are the Katz centrality (*Katz*, 1953), the eigenvector centrality (*Bonacich*, 1987), the PageRank (*Page et al*., 1998), and the so-called Hub and Authorities (*Kleinberg*, 1999), Cross-clique centrality (*Everett and Borgatti*, 1998).

Degree Centrality (*Freeman*, 1977) is a local centrality metric that represents the number of edges incident upon a node. It describes the local structural importance of the nodes thought the number of its neighbours, i.e., the node with highest number of first neighbours (adjacent nodes) is the most important (central). The *degree centrality* of a node *i* can be defined as (*Nieminen*, 1974; *Freeman*, 1979):

$$C_i^D = k_i = \sum_j a_{ij} \qquad i = 1, \ldots, V \tag{19}$$

where C^D_i is the *degree centrality*, k_i is the degree of the node *i* in the network *G*. Since a node *i* can at most be adjacent to (N−1) other nodes, it is possible normalize the metric dividing for the factor (N−1):

$$C_i^D = k_i \cdot \frac{1}{(N-1)} = \sum_j a_{ij} \cdot \frac{1}{(N-1)} \tag{20}$$

The value of C^D_i is now independent from the size of the network and range between 0 and 1.

Closeness Centrality (*Freeman*, 1977) is a centrality metric that measures the geodesic distance (shortest path) from a node to all other nodes in a network, denoted by the reciprocal of the sum of all distances of a node from all other nodes. It measures the centrality of a node considering how it is relatively close to all other nodes. *Closeness centrality* can be regarded as a measure of how long it will take to spread information from *i* to all other nodes sequentially.

The *closeness centrality* of a node *i* can be defined as (*Sabidussi*, 1966; *Freeman*, 1979):

$$C_i^C = \frac{1}{\sum_j d_{ij}} \tag{21}$$

where C^C_i is the *closeness centrality* and $\sum d_{ij}$ is the average distance from node *i* to all the other nodes in the network. The normalized form of the *closeness centrality* considers that a node *i* can at most be adjacent to (N−1) other nodes and represents the average length of the shortest paths instead of their sum, given by the previous formula multiplied by *N*-1:

$$C_i^C = \frac{N-1}{\sum_j d_{ij}} \tag{22}$$

where *N* is the number of nodes in the graph. The value of C^C_i range between 0 and 1 for the normalized form.

The *closeness centrality* only works for fully connected networks. In fact, when the distance between two nodes belonging to different components is infinite, the metric is equal to the null value for each node. It is possible to define the metric as the summation of the inverse of each distance, which is the formulation of the *Harmonic Centrality* (*Rochat*, 2009).

$$H_i^C = \sum_t \frac{1}{dist_{i,j}} \tag{23}$$

Betweenness Centrality (*Freeman*, 1977) is a global centrality metric that measures the importance of an element counting how many times it is traversed by shortest paths between two vertices, quantifying how a node intervenes in the transfer of information through the network. It seems clear that the node with the most strategic location is the most important. The *betweenness centrality* of a node *i* can be defined as (*Freeman*, 1977; 1979):

$$C_i^B = \sum_{s \neq i \neq t \in G} \frac{\sigma_{st}(i)}{\sigma_{st}} \tag{24}$$

The normalized form of the *betweenness centrality* is obtained by dividing for the number of couples of nodes, not including the node *i*, which for directed graphs is (N−1) (N−2) and for undirected graphs is (N−1) (N−2)/2. This way we obtain:

$$C_i^B = \sum_{s \neq i \neq t \in G} \frac{\sigma_{st}(i)}{\sigma_{st}} \cdot \frac{1}{(N-1)(N-2)} \tag{25}$$

Table 1 and Figure 23 report degree, closeness, harmonic and betweenness values for all nodes of the graph of figure 5.

Table 1. Degree, closeness, harmonic closeness and betweenness values for all vertices/nodes of the network in Figure 23.

Node	Degree	Closeness	Harmonic	Betweenne
1	2	0,040	4,48	0.00
2	**4**	0,053	**5,83**	11.00
3	3	**0,059**	5,67	11.00
4	3	**0,059**	5,67	**16.50**
5	3	0,056	5,50	4.00
6	3	0,050	5,33	1.00
7	3	0,048	5.17	14.00
8	1	0,035	3,57	0
9	1	0,037	3,82	0
10	1	0.035	3.57	0

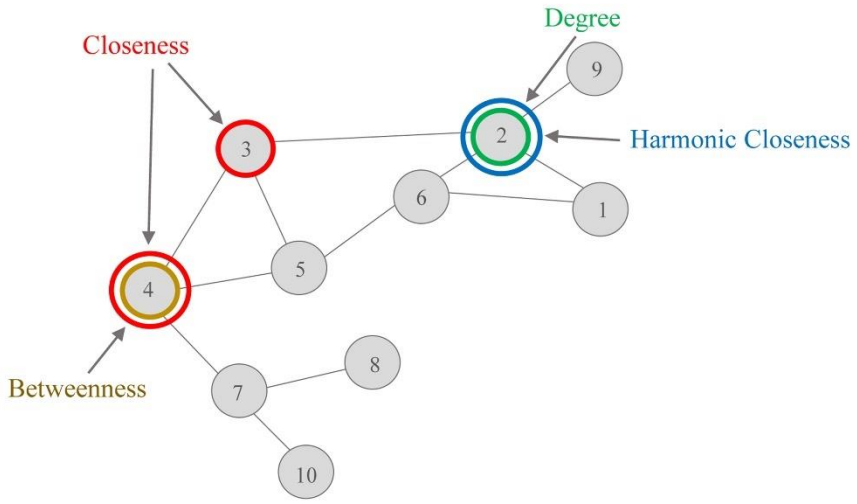


Figure 23. The figure highlights the vertices/nodes into the network with higher values of degree (2), closeness (3 and 4), harmonic closeness (2) and betweenness (4).

2.7 Network partitioning and modularity index

Graphs help us to understand, in a visual way, the global organization of complex real systems. When the graph is small this type of representation works very well but, as soon as the system become of hundreds or thousands of vertices, the graph seems to disappear to make way for a meaningless cloud of nodes. When *N* is large, to find a division of the network into communities, i.e., uncovering communities/modules in the network, could be helpful to understand the structure of the network and make it more readable. In fact, complex systems can be naturally divided into several modules that represent groups of densely connected nodes with sparse connections to the nodes of other groups.

Graph partitioning and *community detection* are the two main methods that refer to the division of a network into modules. In the first one number and size of modules are fixed *apriori* by the user, i.e., the method divides the system into a required number of modules, regardless of the organization of the system. In the second one number and size of modules are not fixed *apriori* but determined by the organization of the system. This method should be able, independently, to identify various modules of different size in the same system and to divide the latter only when a good subdivision exists, leave it undivided otherwise.

In order to detect groups/communities, several methods have been proposed (*Fortunato*, 2010) and the modularity index was introduced as the most widely accepted and used metric to measure the propensity of the system division into modules/communities (*Newman and Girvan*, 2004). The modularity index is a CNT topological index and relies strictly on the network structure, depending only on the identification of network adjacency matrix (*Steinhaeuser and*

Chawla, 2010). Furthermore, its assessment does not require high computational costs. For a given network, a higher value of the modularity index indicates a better identification of modules, i.e., a network with nodes densely connected within modules but lightly connected between different modules. The maximum value of the modularity index corresponds to the maximum degree of division of the network.

The high successful application of the modularity index in a wide range of fields is due (i) to the flexibility of the metric in embedding information about properties of the network links, (ii) to the range of the maximum value between zero and unit, (iii) to the availability of efficient maximization algorithms for very large size problems and (iv) to the applicability to different and varied types of networks (*Giustolisi and Ridolfi*, 2014a).

The modularity index has been proposed as a comparison between the real network and a random network which share some general topological characteristics (e.g., the total number of links into the network and the nodal degree distribution) with the real one. Modularity is defined as (*Newman and Girvan*, 2004),

$$Q = \frac{1}{2n_p}\sum_{ij}\left(A_{ij} - P_{ij}\right)\delta\left(\mathrm{M}_i, \mathrm{M}_j\right) = \frac{1}{2n_p}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2n_p}\right)\delta\left(\mathrm{M}_i, \mathrm{M}_j\right) \tag{26}$$

where A_{ij} are the elements of the adjacency matrix $\bar{\mathbf{A}}$ of the network, P_{ij} is the expected fraction of edges between nodes i and j in the null/random network (i.e., the expected number of edges in the network if they are randomly distributed), M_i is the identifier of network modules, δ is the function to apply the summation to the elements of the same module (i.e., δ = 1 if $\mathrm{M}_j = \mathrm{M}_i$ and $\delta = 0$ otherwise), and summation runs on all the possible node pairs (i,j), with $i \neq j$. In Eq. (25), the expected fraction P_{ij} is computed using nodal degree: i.e., k_i (k_j) is the degree of the i-th (j-th) node (i.e., the number of edges incident that node). The two following terms of the modularity in Eq. (26),

$$\frac{\sum_{ij} A_{ij}\delta\left(\mathrm{M}_i, \mathrm{M}_j\right)}{2n_p}; \quad \frac{\sum_{ij} k_i k_j \delta\left(\mathrm{M}_i, \mathrm{M}_j\right)}{\left(2n_p\right)^2} \tag{27}$$

are, respectively, the fraction of edges connecting nodes which are in the same module and the expected fraction of edges connecting nodes in the same module in a random graph having the same degree distribution of the original graph.

The modularity can be formulated also as,

$$Q = \sum_{m=1}^{n_m} e_{mm} - \sum_{m=1}^{n_m} a_m^2$$
$$a_m = \sum_i \frac{k_i}{2n_p} \quad \forall i \in \mathrm{M}_m \tag{28}$$

where e_{mm} is the fraction of the edges with both the ending nodes belonging to the m-th module, n_m is the number of network modules, and a_m is the fraction of edges having at least one end node in the module m. The pipes dividing modules are counted ½ when computing a_m. In fact, a_m is ½ of the summation of nodal degrees of the module m, divided by n_p. It is possible to write

$$\sum_{m=1}^{n_m} e_{mm} = 1 - \frac{n_c}{n_p} \tag{29}$$

where n_c is the number of edges separating modules, namely the number of "cuts" in the network. Those cuts are virtual and relate to the division into blocks (modules), i.e., the metric Q can be explained as a measure of the block decomposition of the adjacency matrix of the network graph.

Figure 24 reports the network of Figure 5 (left) and its division in three modules (right) obtained with three virtual cuts.

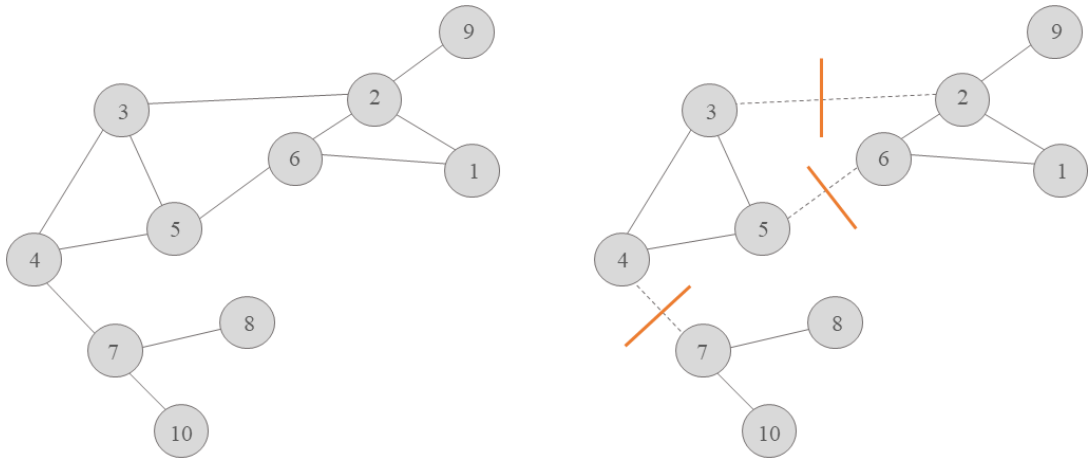


Figure 24. Entire network (**left**) and its division in three modules (**right**) with three virtual cuts. Modularity index Q=0.34.

Therefore, dividing networks into modules by optimizing the modularity index is essential to discover the connections between subnets and to make various practical applications simpler in many real systems. At the same time, however, the optimization of modularity can have limits, failing to identify modules smaller than a scale that depends on the total size of the network and the degree of interconnection of the modules. Modularity compares the number of edges in a module with the expected number of edges in the same module if the network was

random, i.e., with edges randomly attached assuming the same number of nodes, that keep their original degree. This assumption implicitly assumes that each node can get attached to any other node of the network, that can result unreasonable if the network is very large, since, physically speaking, a node cannot be directly connected to any other node within the network. Moreover, as the size of the network increases, the number of edges between two modules decreases, and, for very large networks, it may even become less than one. In this case, the optimization of the modularity could lead to the merger of the two modules, which would be viewed as a single module, thus limiting the formation of small modules (*Fortunato and Barthélemy*, 2007).

3

Water distribution networks (WDNs) and CNT

Contents

Water distribution networks (WDNs) represent complex systems of water drawing, treatment, transport, and distribution whose purpose is supplying city centers. They are composed of several interconnected components, structured in non-trivial configurations, whose behavior is largely influenced by their connective structure, spatial limits, and interactions between components (*Yazdani and Jeffrey*, 2011; *Giustolisi et al*., 2017). Such complex systems must be reliable both in relation to random failures and intentional threats, in the identification of contaminants, in the control of pressures and in the leakages management, as well as for purposes related to monitoring, calibration of hydraulic models and system control, for example by optimally dividing the network into modules, in order to improve its design and management (*Simone et al*., 2016; *Berardi et al*. 2017; *Mazzolani et al*., 2017; *Giustolisi et al*. 2017b; *Laucelli et al*., 2017). WDNs planning and management focused, for a long time, on empirical assessments based on two mainly concepts: increase the level of energy and the topological redundancy of the networks (*Mays*, 2000), increasingly, consequently, number of loops and hydraulic diameters. Furthermore, most studies on the vulnerability and the robustness of such systems neglects the topological aspects, losing a lot of useful information during the analysis.

Such incomplete knowledge of the actual behavior of WDNs led, often, to the choice of inadequate management and planning solutions, and sometimes to disadvantageous interventions, because based on incomplete data and analysis. The need to develop alternatives for WDNs studies is increasingly pressing and in the last years CNT attracted attention by making possible a new, interesting, and structured design approach from the point of view of their analysis, management and optimization, both in their functioning and evolution (*Giustolisi et al.*, 2017).

Numerous studies of WDNs based on CNT methods and metrics have been proposed. *Kesavan and Chandrashekar* (1972) began using graph theory for nonlinear pipe networks. *Jacobs and Goulter* (1988, 1989) and *Kessler et al*. (1990) proposed few works about vulnerability and robustness of WDN based on topological metrics, showing how regular networks, i.e., networks with equal number of links incident to each node, are less vulnerable to failures. The concept of vulnerability and system failure is also faced by *Ostfeld and Shamir* (1996) mainly referred to the WDNs redundancy when failure occur. *Gupta and Prasad* (2000) proposed a steady-state analysis of flow and pressure based on linear graph theory. *Deuerlein* (2008) proposed a network simplification based on the connectivity structure of WDNs, in order to improve the understanding of the single network components and their interaction. *Yazdani and Jeffrey* (2010; 2012a) represented WDNs as large sparse planar graphs showing

relationship between topological features and network resilience, also introducing two new metrics, meshedness and algebraic connectivity, in order to quantifying WDNs redundancy and robustness, respectively. Successively, they proposed several strategies for understanding WDNs structure (*Yazdani and Jeffrey*, 2012b), studying their connectivity and the relationship with system robustness and susceptibility to damage, modeling weighted and directed networks. Finally, in 2011, they proposed the link-node representation of WDNs and a wide range of metrics in order to establish potential relationships between structural features and performance of WDNs. *Hawick* (2012) used betweenness centrality metric and rank network to study WDNs robustness and fragmentation properties. *Gutiérrez-Pérez et al*. (2013) proposed a methodology based on spectral measurements using PageRank and HITS algorithms in order to achieve an efficient vulnerability analysis and to establish the relative importance of smaller areas in WDNs. *Sheng et al*. (2013) explored the formation of isolated communities due to pipeline failures using a graph-algebraic model. *Shuang et al*. (2014) proposed the study of nodal vulnerability of WDNs under cascading failures. *Nazempour et al*., (2016) proposed a new approach for the optimal placement of sensors for contamination detection in WDNs. Successively, *Giustolisi et al*., (2017) proposed the *neighborhood degree* strategy for augmenting the information with respect to the original nodal degree distribution increasing the maximum nodal degree with respect to the original one, also reducing the dependency of the distribution by the spatial constraints. They demonstrated that, when using the neighborhood degree, the Poisson distribution more reliably represents the connectivity of WDNs. Various authors proposed strategies to understand the actual role of the topological structure in the functioning of WDNs (*Giudicianni et al*., 2018; *Giustolisi et al*., 2019; *Simone et al*., 2020) and of spatial networks in general (*Barthelemy*, 2018).

The graph theory concepts are used also for monitoring and control purposes and optimal network division with respect to topology and asset characteristics (*Laucelli et al*., 2017; *Torres et al*., 2016; *Perelman and Ostfeld*, 2011). A paradigm from CNT largely used for WDNs division/segmentation bases on the modularity index (*Newman and Girvan*, 2004). Starting from the original Newman's formulation, suggested for immaterial networks (*Barthélemy*, 2011), *Scibetta et al*. (2013) and *Diao et al*. (2013) proposed the first segmentations of WDNs. Later on, *Giustolisi and Ridolfi* (2014a) proposed (i) the WDN-oriented modularity index, tailoring the original one, in order to accounting for the features of WDNs and (ii) the infrastructure modularity index (2014b), modifying the WDN-oriented modularity index, in order to overcome the resolution limit of the original one (*Fortunato and Barthélemy*, 2007), which causes the non-identifiability of smaller modules depending on the size of the network.

Simone et al. (2016) extended the concepts of network segmentation to pressure sampling design, introducing the sampling-oriented modularity index and the concept of "pressure" DMAs.

However, as well as network analyzes only based on energy parameters are not exhaustive, the use of purely topological studies for WDNs analysis has a limited scope (*Yazdani and Jeffrey,* 2010; *Giustolisi et al*., 2019).

3.1 Hydraulic modelling and hydraulic domain

In the early twentieth century the aim of the hydraulic analysis was to calculate the operating pressures at the nodes of the network, to ensure the satisfaction of water demands to users and the fire protection. Progressive urban development produced ever larger, more complex and dated WDNs, implying increasingly greater management needs with respect to leakages management, reliability, water quality, energy optimization, rehabilitation, etc. With the advent of automatic calculation, *Todini and Pilati* (1988) conceived the Global Gradient Algorithm (GGA) based on the assumption of fixed demands at the nodes of the WDNs. The hydraulic solver, created for hydraulic verification, assumed demand nodes known at any time step independently on network head/pressure status, i.e., it allowed a demand-driven analysis (DDA). No relationship between pressures and demand was assumed and the leakages component was commonly kept as percentage of the fixed nodal demands. Successively, *Todini* (2003) introduced the need for a hydraulic simulation capable of evaluating the effective demand that could be supplied to users in conditions of pressures below the minimum pressures for correct service, i.e., allowing a pressure-driven analyses (PDA). He stated that a relationship between pressure and demand was necessary because demands were not fixed but generally depend on the pressure status. In fact, the pressure-driven simulation models a more realistic prediction of leakages, where the leakage component is evaluated as a function of the network pressure (*Giustolisi et al*., 2008).

Bhave (1981) was the first to model WDNs considering PDA. He proposed to model the networks by defining a minimum pressure required for normal working conditions. *Germanopoulos* (1985) combined the concept of leakages to that of pressure-dependent demand. *Wagner et al*. (1988) proposed a pressure-demand model for controlled outlets based on a function that is hydraulically significant but not differentiable (*Ackley et al*., 2001), particularly close to zero value of the flow rate. Successively, Wagner' approach, which reflects well the real hydraulic operation of WDNs, is considered the most appropriate for pressure dependent demands (*Gupta and Bhave*, 1996). In 2008 *Giustolisi et al*. proposed the

introduction of an adaptive overrelaxation parameter to pressure-driven analysis within GGA and using *Wagner et al*. (1988) and *Germanopoulos* (1985) models within GGA and EGGA, *Giustolisi et al*. (2008), accounted for demand and leakage dependencies on network head/ pressure status. The leakage component was evaluated as a function of the network pressure at the level of each individual pipe preserving the accuracy of the hydraulic calculation.

Giustolisi and Walski (2012), studying Wagner's model, shown how it was consistent with the real hydraulic operation of the network where users statistically check the flow rates to the taps as long as the pressures are sufficient (Torricelli's law for completely open devices): equations (29) expresses an effective demand $d(i, t)$ at the i-th node at hour t which, only for pressures greater than P_{ser} (i) minimum pressure for a correct service, is equal to the request statistical water of the user d_{req} (i, t), while it is zero for pressures lower than the height of the tap P_{min} (i) and responds to Torricelli's law for intermediate pressures, or in conditions of insufficient pressure (pressure-deficient condition). Figure 25 reports the correspondent Wagner's model diagram.

$$d(i,t)=\begin{cases} d^{req}(i,t) & P(i,t)\geq P^{ser}(i) \\ \dfrac{d^{req}(i,t)}{\sqrt{P^{ser}(i)-P^{min}(i)}}\sqrt{P(i,t)-P^{min}(i)} & P^{min}(i)<P(i,t)<P^{ser}(i) \\ 0 & P(i,t)\leq P^{min}(i) \end{cases} \tag{30}$$

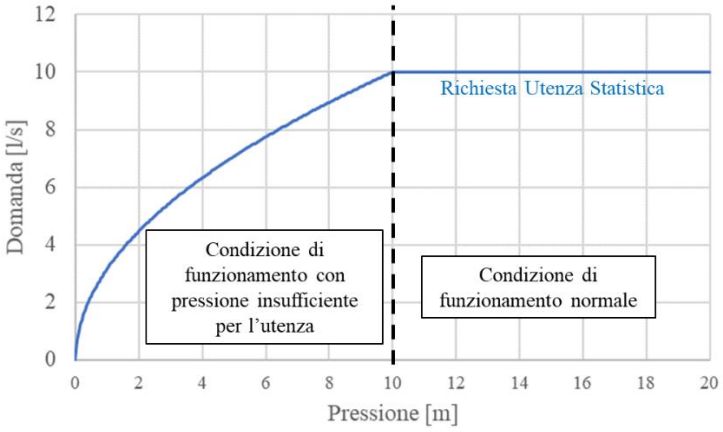


Figure 25. Pressure-demand diagram for users.

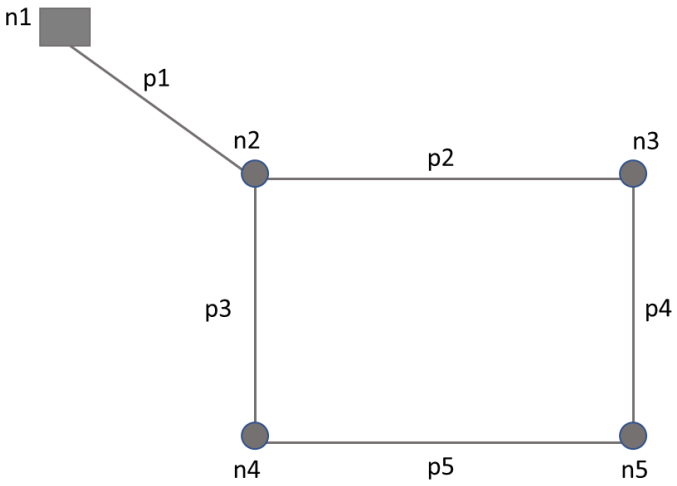
Generally, the hydraulic behavior of WDNs can be described by three components:

- mass and momentum balance equations
- domain (topological and and the most important pipes characteristics (e.g., length, pipe resistance or conductance, etc.) where equations are solved

- boundary conditions (e.g., demands, water tank levels, devices state, etc.) usually assumed slowly time-varying to neglect fast transient effects.

The hydraulics behavior is then simulated as a sequence of steady-state snapshots over time, which captures the slow change of the network boundary conditions (*Giustolisi and Walski*, 2012). Although these different boundary conditions alter the single values of flows and pressure, the domain of the problem (i.e., the internal connectivity structure of the network, the topological position of the reservoirs, the pipes resistance/conductance, etc.) determines the main traits of the hydraulic behaviour of a WDN and greatly determine the system state over time. Therefore, the analysis of the WDN hydraulics domain cannot substitute the hydraulic simulations, because the information about the continuity and momentum equations is not contained in the domain, but it can represent a useful tool to identify the main features of the hydraulic system (*Giustolisi et al.*, 2019). Therefore, it is essential to analyse the topology of WDNs and make it complementary to hydraulic analyses to provide increasingly reliable results and to drive technicians in the most appropriate management choices.

Considering a WDNs as a graph, we obtain that edges and vertices of the graph represent the number of pipes (n_p) and nodes (n_n) of the hydraulic system, respectively. The nodal degree, *k*, of the specific node *n* is the number of pipes *p* incident the node *n*. The network connectivity can be described by the topological incidence matrix, which is generally used also for hydraulic system representation (*Giustolisi et al.*, 2008b). Figure 26 (left) reports a simple water system composed of five nodes, n_n=5 (four demand nodes and one reservoir) and five pipes, n_p=5. Figure 26 (right) reports the corresponding incidence matrix composed by n_p rows, each corresponding to one pipe of the hydraulic system, and n_n columns. The sum of the elements of each column gives the number of pipes connected to the corresponding node, i.e., its degree.



	n1	n2	n3	n4	n5
p1	1	1	0	0	0
p2	0	1	1	0	0
p3	0	1	0	1	0
p4	0	0	1	0	1
p5	0	0	0	1	1

Figure 26. Simple hydraulic network (**left**) and corresponding topological incidence matrix (**right**).

If we also consider the hydraulic side of the system, we can state that a network of n_p pipes with unknown flow rates, n_n nodes with unknown heads (internal nodes) and n_0 nodes with known heads (e.g., reservoir) can be analysed by solving the following non-linear mathematical system based on momentum and mass balance equations:

$$\begin{aligned} \mathbf{A}_{pp}\mathbf{Q}_p + \mathbf{A}_{pn}\mathbf{H}_n &= -\mathbf{A}_{p0}\mathbf{H}_0 \\ \mathbf{A}_{np}\mathbf{Q}_p - \mathbf{d}_n\left(\mathbf{H}_n\right) &= \mathbf{0}_n \end{aligned} \tag{31}$$

where $\mathbf{Q}_p$ = [n_p,1] column vector of unknown pipe flow rates; $\mathbf{H}_n$ = [n_n,1] column vector of unknown nodal heads; $\mathbf{H}_0$ = [n_0,1] column vector of known nodal heads; $\mathbf{d}_n$ = [n_n,1] column vector of nodal demands; $\mathbf{0}_n$ = [n_n,1] column vector of null values; $\mathbf{A}_{pp}$ = diagonal matrix of size [n_p,n_p], whose elements are based on pipes resistance and flow; $\mathbf{A}_{pn}$=$\mathbf{A}^T_{np}$ and $\mathbf{A}_{p0}$ = topological incidence sub-matrices of size [n_p,n_n] and [n_p,n_0] represent the domain of the equations, i.e., the network. The general topological matrix $\bar{\mathbf{A}}_{pn}$=[$\mathbf{A}_{pn}$ ¦ $\mathbf{A}_{p0}$] of size [n_p,n_n+n_0], is the incidence matrix (*Todini and Pilati*, 1988), which includes the information about the direction of pipes, negative or positive. The solution of the system (30) requires the identification of the unknowns ($\mathbf{Q}_p$; $\mathbf{H}_n$) considering known the boundary conditions ($\mathbf{d}_n$; $\mathbf{H}_0$) and information about pumps curve or devices status over time if present. $\mathbf{R}_p$ or $\mathbf{C}_p$ = [n_p,1] column vector of pipe resistances or conductance complete the characteristics of domain, ($\bar{\mathbf{A}}_{pn}$) (*Giustolisi et al*., 2019).

It is worth noting that the matrix multiplication of the transpose of general topological matrix and itself ($\bar{\mathbf{A}}_{np}$ *x* $\bar{\mathbf{A}}_{pn}$) provides the Laplace matrix, which is the difference between the degree matrix and the adjacency matrix, of the graph referred to the WDN domain. Furthermore, the global gradient algorithm (GGA) for WDN hydraulics is (*Giustolisi and Walski*, 2012),

$$\begin{aligned} \mathbf{F}_n^{iter} &= -\mathbf{A}_{np}\left(\mathbf{Q}_p^{iter} - \left(\mathbf{D}_{pp}^{iter}\right)^{-1}\mathbf{A}_{pp}^{iter}\mathbf{Q}_p^{iter}\right) + \mathbf{A}_{np}\left(\mathbf{D}_{pp}^{iter}\right)^{-1}\mathbf{A}_{p0}\mathbf{H}_0 + \mathbf{d}_n^{iter}\left(\mathbf{H}_n^{iter}\right) + \mathbf{D}_{nn}^{iter}\mathbf{H}_n^{iter} \\ \mathbf{H}_n^{iter+1} &= -\left(\mathbf{A}_{np}\left(\mathbf{D}_{pp}^{iter}\right)^{-1}\mathbf{A}_{pn} - \mathbf{D}_{nn}^{iter}\right)^{-1}\mathbf{F}_n^{iter} \\ \mathbf{Q}_p^{iter+1} &= \left(\mathbf{Q}_p^{iter} - \left(\mathbf{D}_{pp}^{iter}\right)^{-1}\mathbf{A}_{pp}^{iter}\mathbf{Q}_p^{iter}\right) - \left(\mathbf{D}_{pp}^{iter}\right)^{-1}\left(\mathbf{A}_{p0}\mathbf{H}_0 + \mathbf{A}_{pn}\mathbf{H}_n^{iter+1}\right) \end{aligned} \tag{32}$$

where *iter* is a counter of the iterative solving algorithm, $\mathbf{D}_{pp}$ is a diagonal matrix whose elements are the derivatives of the head loss function with respect to $\mathbf{Q}_p$ and $\mathbf{D}_{nn}$ is a diagonal matrix whose elements are the derivatives of $\mathbf{d}_n^{iter}$ with respect to network pressure status as $\mathbf{d}_n$= $\mathbf{d}_n$ ($\mathbf{H}_n$). The GGA involves the iterative solution of a linear nodal system of equations, to compute $\mathbf{H}^{iter+1}$, and a nodal mass balance correction obtained by n_p independent equations to compute $\mathbf{Q}^{iter+1}$ based on $\mathbf{H}^{iter+1}$. The matrix $\mathbf{A}_{np}(\mathbf{D}_{pp})^{-1}\mathbf{A}_{pn}$ plays a relevant role for the nodal

solutions and it is a Laplace like matrix integrated with the information of pipes resistance $\mathbf{R}_p$, being the diagonal elements, referring to the kth pipe, $D_{kk} = D_{kk}(R_{kk},|Q|)$ of $\mathbf{D}_{pp}$ (*Giustolisi et al*., 2019).

This fact shows and demonstrates the mathematical fundamentals of studying the WDN hydraulics domain.

3.2 Characteristics of the network: material and spatial peculiarities

Studies about structure and evolution of distribution/transportation systems started with quantitative geographers in the 60-70s. *Hagget and Chorley* (1969) were the first to address the problem of the spatial peculiarities studying topology and geometry of transportation networks. Many other authors proposed metrics and models to characterize such spatial systems and study their temporal and spatial evolution (*Kanksy*, 1963; *Taafe et al*., 1963). With the advent of the first computers and an ever-increasing availability of data relating to networks, it has been possible to represent real systems in more detail even if the purely spatial component of some of them, linked to geometry as well as to their topology, was often neglected. WDNs represent one of the cases of much studied infrastructural networks for which, however, the spatial component has often been ignored. Actually, such systems are limited in the space and governed by a metric: the Euclidian distance (*Barthelemy*, 2018). Furthermore, pipes, which represent the key elements for WDNs, are material components characterized by asset features such as length, diameter, and hydraulic resistance. This means that to characterize this type of networks we need of the topological information (stored in the adjacency matrix), of the spatial information (coordinates of nodes) and of the asset characteristics about pipes. The spatial constraints affect the structure and properties of these networks, whose performance depend on distance, i.e., the probability of finding a link between two nodes decreases with the distance. In fact, WDNs are manmade distribution systems that progressively grow filling the space based on connection costs and nodal distances, but also constrained by the impracticality of some connections (*Buhl et al*., 2006). Furthermore, the growth of these systems affects the temporal component as well as the spatial one. They constantly change over time for multiple factors (e.g., human behavior, climatic conditions, political and social dispositions, etc.). Each variation adds new data to the system, which should be updated. One of the most obvious examples of infrastructure network subject to changes over time is the street network. Figure 27 shows the evolution of the Paris street network over time (*Barthelemy*, 2018). The network in 1789 only considers the main level of transportation of the city and presents a different scale representation (low) with respect to those in 1999 (high), which reports the secondary level of

the transportation with quite all the streets: it is evident that the system is much more constrained. The classification of the network from the first scale representation to the last scale representation could vary at least with respect to the categories of *small world* and *random networks*. This means that considering only the mains of WDNs (primary level of the distribution), i.e., neglecting a secondary level of the distribution and/or connections to properties, means to analyse a different system with a different behaviour (e.g., vulnerability). In this perspective, the temporal component, linked to the availability of data, becomes increasingly important in the classification of infrastructural networks.

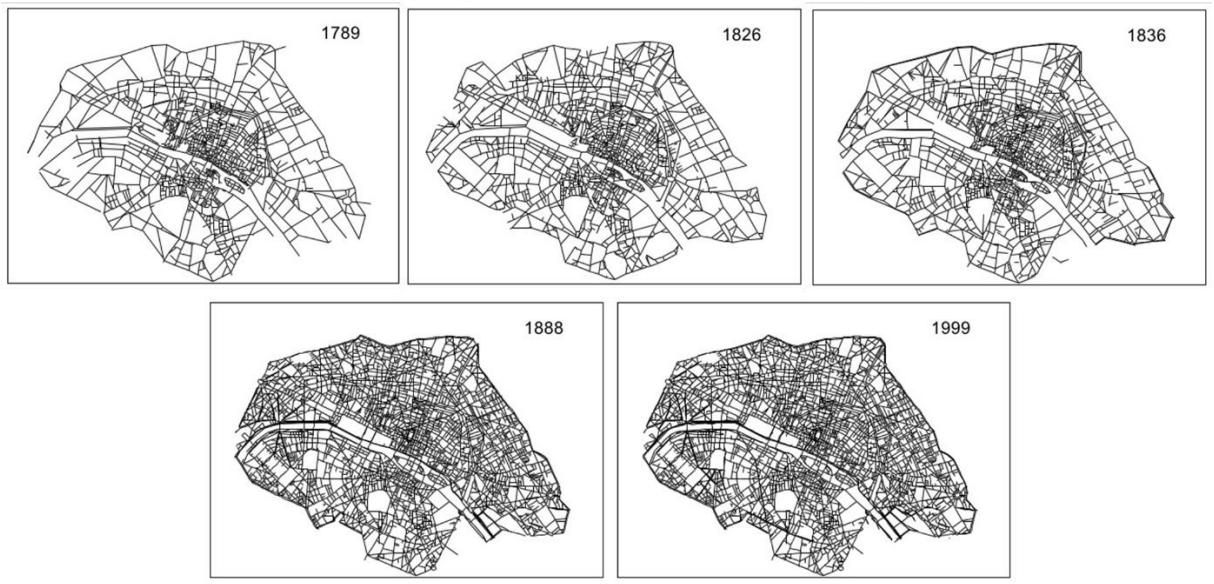


Figure 27. Evolution of the Paris street network. Source: *Barthelemy*, 2018.

3.3 WDNs classification and assessment using the classic CNT framework

Most of the studies on CNT concern the characterization and classification of complex real systems and the identification of the most suitable metrics for understanding and describing their behaviour. Several CNT metrics represent irrelevant parameters for WNDs study, or, in any case, they need to be adapted to be efficiently applied to such systems.

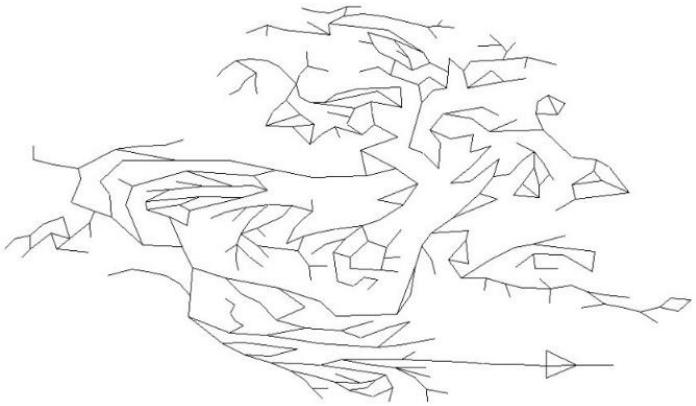
Degree

CNT generally classifies the connectivity structures of real systems using the nodal degree, the average path length, the clustering coefficient, and the probability of connection. For WDNs most of these metrics are irrelevant.

The spatiality of WDNs is mainly related to the urban texture, e.g., streets and buildings, constraining the pipe installation and influencing the layout. In fact, nodes represent

intersections between pipes for transferring information (flow), and since they are generally located at road junctions, the number of pipes incident the node corresponds to the number of streets that converge at the intersection. This fact limits the maximum nodal degree for WDNs, reducing the range of possible values for nodal degree. It results that WDS classification by the CNT standard nodal degree distribution is very difficult and uncertain (*Yazdani and Jeffrey*, 2011) because of the very low values of the maximum standard nodal degree and the resulting unreliable statistical for those networks. A first step to overcome this limit is represented by the *neighborhood degree* (*Giustolisi et al*., 2017), defined as the sum of the degrees of the nearest neighbours (i.e., only nodes immediately adjacent). The method augments the information with respect to the original nodal degree distribution increasing the maximum nodal degree with respect to the original one and reducing the dependency of the distribution by the spatial constraints. When using the *neighborhood degree*, the connectivity structure of WDNs follows a Poisson distribution, i.e., WDNs network connectivity presents a low vulnerability with respect to failure events, consistent with the fact that technicians generally design WDNs using a criterion of redundancy beyond the hydraulic capacity requirements.

Figure 28 report the BBLAWN WDN layout and Figure 29 reports its classification using both classic degree and *neighborhood degree*. Discerning between the Poisson and Pareto distributions when using the classic nodal degree is difficult, also because the maximum degree is five and, when only five points are available to fit the distribution, it can result statistically unreliable. Differently, discerning between the Poisson and Pareto distributions when using the *neighborhood degree* is possible and quite immediate. In this last case, in fact, twelve points are available, and it is evident that the pattern follows well the cumulative Poisson distribution.



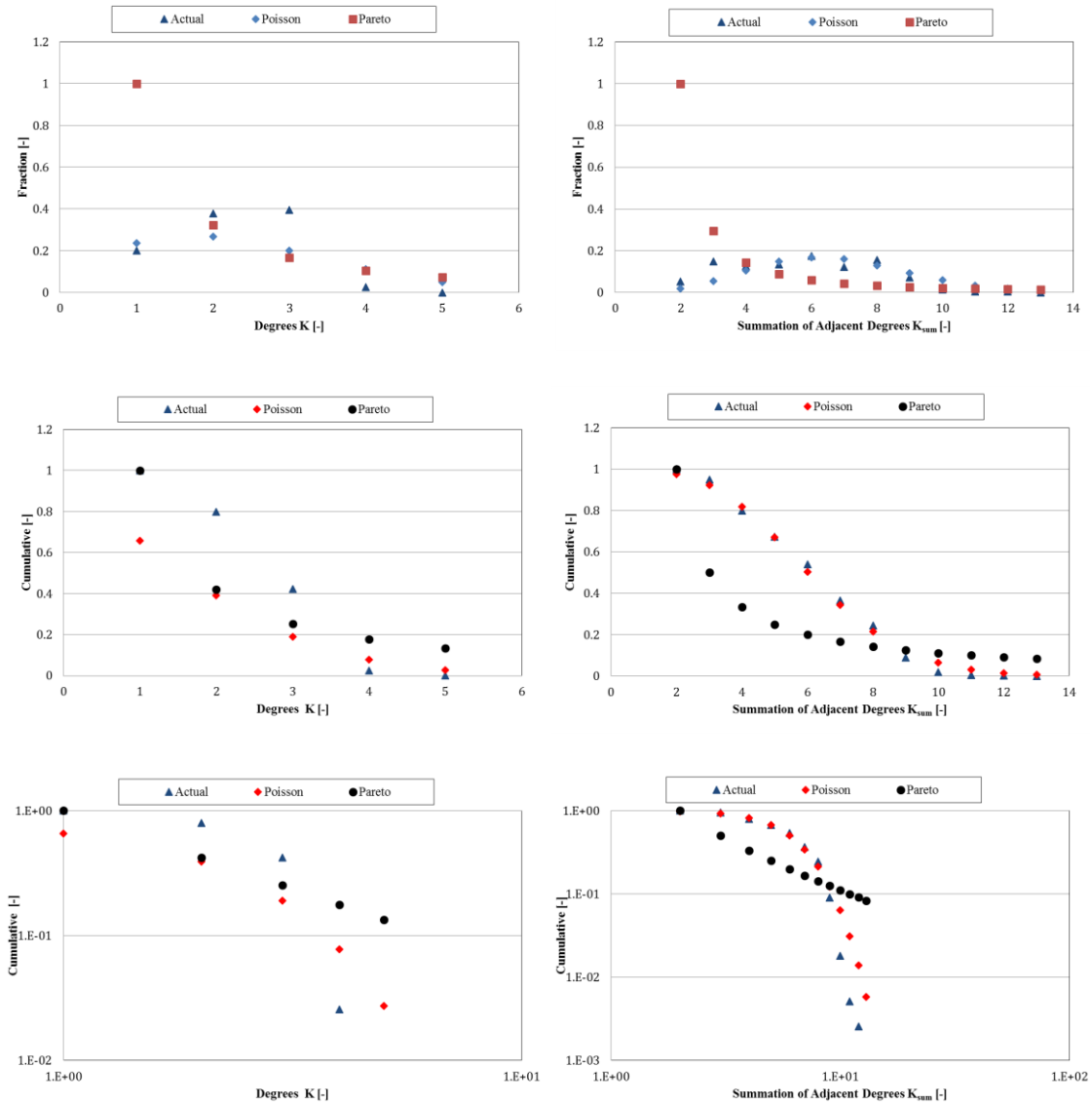


Figure 28. Top. WDN layout. **Bottom**. Actual distribution of classic degrees and neighbourhood nodal degree versus Poisson and Pareto distributions for BBLAWN WDN. Source: *Giustolisi et al.*, 2017.

Clustering coefficient and probability of connection

The Poisson distribution models the fact that each pair of nodes is connected randomly with a probability *p*, which generates a network having a great number of nodes with a similar degree. For regular networks, the probability *p* is zero and tends to increase with increasing random connections in the network, and for small world networks, the probability *p* is greater than the null value of regular networks but rather lower than values of random networks. Regular networks are highly ordered with shortest paths linking two nodes very high, differently from what happens to real systems. The shortest paths for random network, instead, is very low. Therefore, a real system, generally, has an average path length typical of small world networks,

i.e., lower than regular network and greater than random networks. It is worth to note that also scale free networks can have a low average path length, similarly to small world, because the presence of hubs reduces the degree of separation among nodes, but WDNs connectivity structures are not scale free, i.e., characterized by nodes working as hubs.

Random and small world networks, modelled by the Poisson distribution, can be characterized by the clustering coefficient C_p and the average path length L_p, normalized by clustering C_0 and average path length L_0 of the ring lattice regular network (see Figure 18). Regular networks, with null p, have high L_p/L_0 and C_p/C_0, *small world networks*, with intermediate p, have a low L_p/L_0 and a high C_p/C_0 and *random networks*, with p close to unit value, have low both L_p/L_0 and C_p/C_0. Generally, the probability p determines the average degree of the network,

$$\langle k \rangle = p(n-1) \tag{33}$$

The situation changes for WDNs, whose average degree is determined by spatial constraints. It is possible to state that the average degree for WDNs ranges from 2 to 3, hence, we can write,

$$p = \frac{\langle k \rangle \approx 2.5}{n-1} \quad \Rightarrow \quad p \propto n^{-1} \tag{34}$$

where the probability of connection is very low for the network connectivity structure of WDNs because of spatial constraints, in addition to being inversely proportional to the size of the network (*Simone et al*., 2018). As a result, the *Watts-Strogatz* (1998) scheme must be used with caution for this type of system.

Figures 29 and 30 report the diagrams of L_n/L_0, C_n/C_0 versus the decreasing network size, and not versus the probability of connection p, for 21 analysed WDNs. The diagram in Figure 29 clearly demonstrates that L_n/L_0 increases with decreasing of the network size as also shown by the trend line there reported, meaning that decreasing the WDN size, the network structure exhibits a higher regularity. Therefore, Figure 29 indicates that n should substitute p in *Watts-Strogatz* diagram (1998).

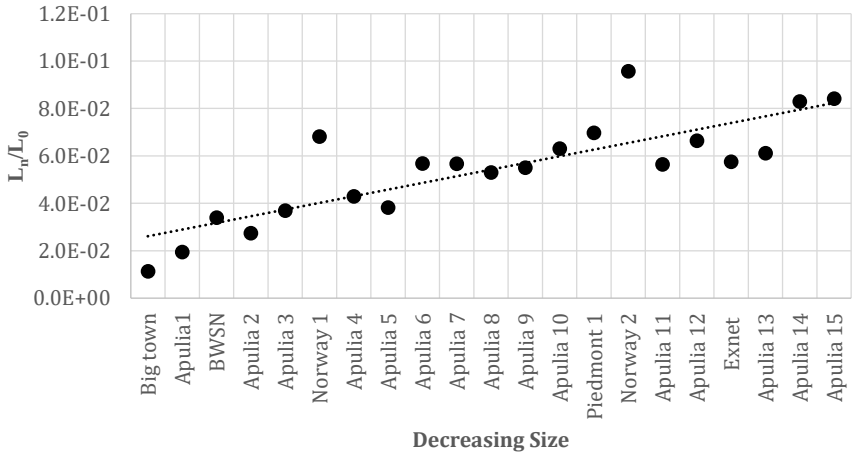


Figure 29. Normalized average path length L_n/L_0 versus decreasing network size. Source: *Simone et al*., 2018.

The diagrams of Figure 30 reports normalized clustering coefficient C_n/C_0 versus decreasing network size. It shows that the parameter has very low values and it does not depend on network size. This because the spatial constraints of analysed networks limit the number of elementary topologies (triangles). In fact, this metric will be of significant interest when applied to networks with a large range of degree variations (*Barthelemy*, 2018) that is not the case in WDNs.

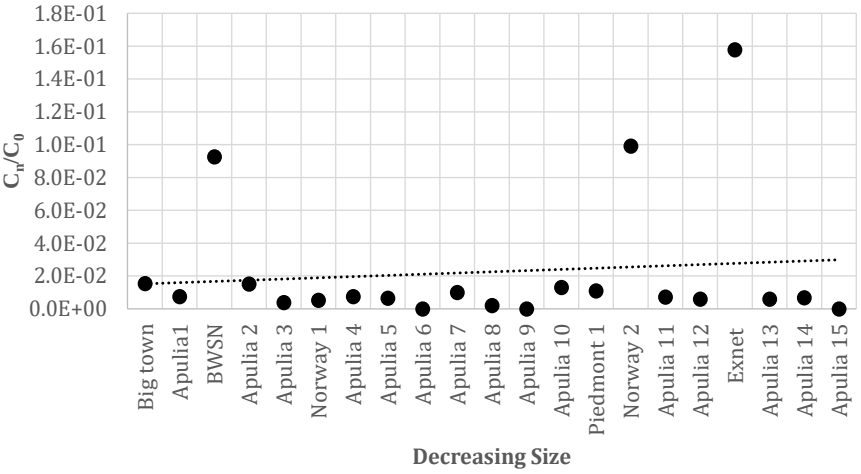


Figure 30. Normalized clustering coefficient C_n/C_0 versus decreasing network size. Source: *Simone et al*., 2018.

Betwenness and Closeness

Several CNT centrality metrics have been proposed for quantifying the importance of elements in networks, analysing the key features of the physical domain (i.e., the network) where WDN hydraulics occurs. Betweenness and closeness, together with degree, are identified as the most suitable centrality metrics for studying infrastructure networks (e.g., *Borgatti*, 2005; *Barthélemy*, 2018; *Giustolisi et al.*, 2019). As example, here we report their application for a benchmark WDN, named Apulian network (*Giustolisi et al*.; 2008), composed of 23 nodes, 34

pipes and one reservoir (Figure 31), in order to demonstrate the usefulness of the metrics for WDNs analysis.

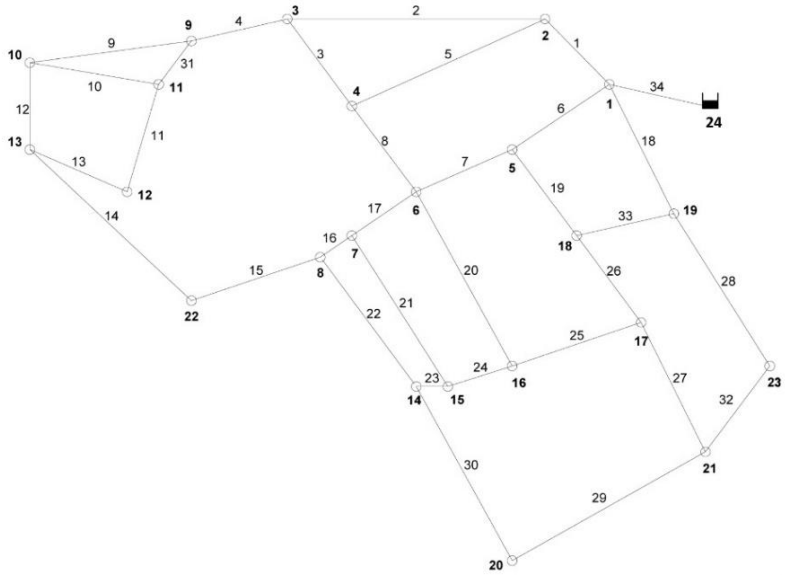


Figure 31. Apulian layout.

We start computing the shortest paths between each couple of nodes in the network using the *Dijkstra's algorithm* (1959). Betweenness and closeness centrality for the Apulian network are normalized in the range [0, 100] and reported in Figure 32. The figure refers to the centrality values for nodes and not pipes. In this framework, node 6 exhibits both the maximum betweenness centrality and the maximum closeness centrality, i.e., it results the most efficient in spreading the information (closeness) because it is the closest to the others and, furthermore, it is also the most traversed by the shortest paths (betweenness) (*Giustolisi et al*., 2019). However, since a nodal failure is generally unrealistic for WDNs, failing of node 6 corresponds to failing of pipes connected to it, being an interruption of a pipe a common technical occurrence. Furthermore, this topological analysis does not highlight the importance of source waters within the network.

These facts clarify the need of tailoring the classing centrality metrics to adapt them to the analysis of WDN hydraulic domain features.

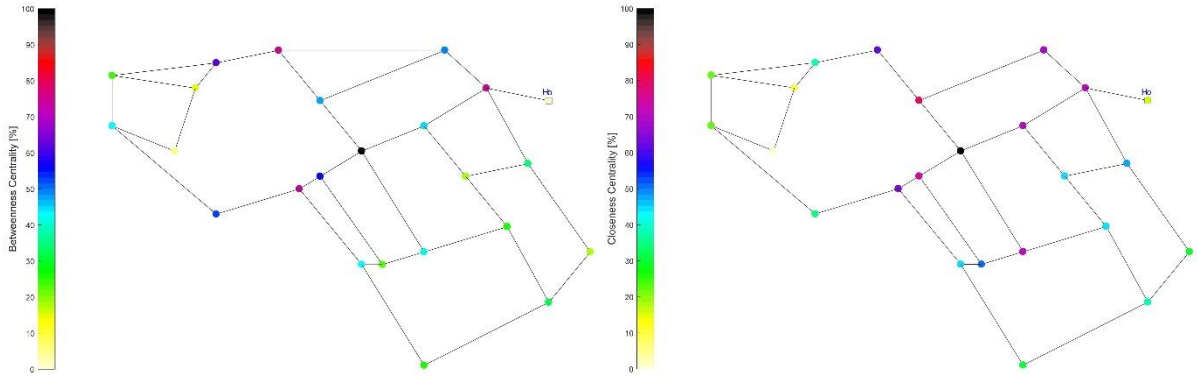


Figure 32. Betweenness centrality (**left**) and closeness centrality (**right**) for Apulian network.

3.4 Need for CNT tailoring

During the last decade, several researchers started using classic CNT tools for WDNs analysis. The usefulness of these tools is recognized in all works as well as the limits they present in the analysis of infrastructural systems. The analysis fully reflects the purely topological character of the CNT tools: results depend on the connectivity structure. Various attempts have been made to assign/associate specific physical features to the connectivity structure to make much more reliable the WDNs analysis, for example introducing weights and using direct graphs (*Yazdani and Jeffrey*, 2012). However, the results have not always been the expected ones.

In order to correctly perform a technical transfer of the CNT tools to the specific infrastructure network, a tailoring process began in 2014 with *Giustolisi and Ridolfi*. They proposed of using the modularity index (*Newman and Girvan*, 2004) for the optimal segmentation design of WDNs. Since the original formulation of the modularity index was not advisable for WDNs (the random network assumed as benchmark for community detection has not correspondence for infrastructure networks), they proposed a tailoring in order to account for WDNs peculiarities. Their tailoring affects (i) the general topological incidence matrix, which is commonly adopted to represent the network topology of the hydraulic systems; (ii) pipe characteristics (weights), in order to account for the specific technical task of segmentations; (iii) the developing of a cut position-sensitive metric that allows to install devices (cuts) close to the end nodes of pipes and not in the middle of pipes, as expected with the classic modularity. The novel WDN-oriented modularity index allows multi-objective (MO) optimization strategy that maximizing the modularity index versus the minimization of the number of observations (e.g., pressure and flow devices), result in a Pareto set of optimal

segmentations (selected in order to optimally segment according to a specific management criteria) versus the solution cost. In a second work, *Giustolisi and Ridolfi* (2014b) focused on the resolution limit issue (*Fortunato and Barthélemy*, 2007), that increases with network size and that limits the identifiability of small modules with respect to the size. They overcome the problem by proposing a novel infrastructure modularity index starting from the WDN-oriented.

We continue the tailoring process (*Giustolisi et al*., 2017) proposing the *neighborhood nodal degree*, instead of the classic degree, for the WDNs classification. The novel metric allows to overcome the problem of spatial limits, typical of WDNs, increasing the degree of each node and finding a greater number of points for the identification of a distribution shape. We showed that the Poisson distribution well characterizes the WDNs connectivity structure. This means that the domain has random features increasing the size (i.e., the number of nodes), i.e., a good structural resistance. However, we leaved out all the information related to the hydraulic behaviour of these systems, e.g., the presence of strategic structures like reservoirs and public buildings, whose intrinsic relevance is independent from their topological position and connectivity, and directional hydraulic devices, which can influence the flow inside the network. I.e., the structural analysis captured the *network domain features* on average.

Novel studies and further tailoring are necessary to adapt CNT methods to the actual behaviour of WDNs in order to obtain a reliable classification and assessment of these systems.

4

Tailoring CNT tools for WDNs: System analysis and management

Contents

This chapter addresses and concretizes the need to adapt CNT metrics for WDNs. The first part presents and discusses two strategies of tailoring referred to the most suitable centrality metrics for spatial network (betweenness, closeness, harmonic closeness, and degree) accounting for WDNs peculiarities. The topic has been the subject of three publications in international journals (Giustolisi et al., 2019; Simone et al., 2019; Giustolisi et al., 2020) that will be recalled along the text. The second part briefly recall the tailoring process adopted for the modularity index and presents two of the works produced on the subject, about the development of strategies for DMA design (Laucelli et al., 2017) and pressure sampling design (Simone et al., 2016).

The need to adapt CNT metrics for WDNs has been addressed for several years and referred to numerous tasks. In particular, the use of centrality metrics has always been addressed to the network classification and vulnerability, while the use of the modularity index to the network planning and management through their optimal division in segments.

Several centrality metrics have been proposed and each one identifies different elements as the most important (central) with respect to different applications/ physical phenomena (Benzi et al., 2015), i.e., the use of a metric must be appropriate to the kind of flow process that characterizes the network. The flow process for spatial networks, such as WDNs, is based on the concept of shortest paths (Borgatti, 2005).

Now, once the most suitable centrality metrics for studying WDNs have been identified, they need to be tailored in order to considering the specific peculiarities of these hydraulic systems. We start presenting a first tailoring approach, enhanced in the second phase of my PhD by a second one.

- The first approach aimed to develop a WDN-tailored centrality metric to identify the WDN domain, i.e., a tailored edge betweenness able to identify the role of topological domain features in the emergent hydraulic behaviour. Particularly, the tailoring is performed considering three specific features of WDNs: (i) edges, i.e., pipes, are the relevant components, in contrast with standard CNT centrality metrics, that generally focus on nodes; (ii) pipes are material components characterized by asset features (e.g., length, diameter, hydraulic resistance, etc.); (iii) each node can represent a different system component (e.g., source of water, connection node and demand node) (Giustolisi et al., 2017). Consequently, we computed the edge relevance instead of nodal one, we ed weights to edges to account for the specific asset characteristics of pipes and we modified the connectivity structure of the network to distinguish source, connection and

demand nodes. To this aim, we introduced fictitious nodes (whose number depends on the number of demand nodes) connected to the source nodes. In this way, the topological relevance of the source nodes is emphasised by generating fictitious water paths that embed the different hydraulic role played by source nodes. In other words, fictitious nodes allow weighting topologically source nodes with respect to the demand nodes. The tailored metric represents a useful tool to support WDNs analysis, design, and management tasks, although adding a fictitious node for each demand node increases the network size and the number the shortest paths to be calculated.

- The second approach aimed to develop a novel tailoring of the centrality metrics considering the information about the intrinsic relevance of each node, exogeneous data independent from their topological position and connectivity, and to embed it into the analysis to make useful CNT tools for WDNs. The intrinsic relevance for each node, defined as R_n ($n=1$, …, N), is assigned equal to the demand for each demand node and to the sum of demand for the source node. This way, all nodes assume a different relevance during the analysis. The novel strategy is performed using several functions $f(R_s, R_t)$ that depend on the intrinsic relevance of nodes s and t.

  This tailoring approach is more flexible and general with respect to the previous one because it allows to adopt "intrinsic relevance" also different from the demand for water networks, or to adopt others relevance for networks other than WDNs.

Regarding the network division in segments, the concept of modularity was among the first to be borrowed from the CNT to be applied to the WDNs analysis and management. The tailoring procedure was proposed, through various scientific papers, by Giustolisi and Ridolfi (2014a, b) who introduced various basic formulations of the modularity index accounting for WDN peculiarities. Starting from these searches, we used the WDN-oriented modularity index to perform the optimal DMA design strategy of WDNs and extended the metric to define the Sampling-oriented modularity index and to perform the optimal Sampling design strategy (Simone et al., 2016) aimed at the optimal location of pressure meters in WDNs.

4.1 WDN-oriented Centrality metrics for WDNs assessment

We conceived and developed the tailoring strategy with the aim to produce a tool that would allow the identification of the domain analysis of WDNs. The tailoring process involved four conceptual steps. The first one was to move from a node-based perspective (typical of CNT applications) to an edge-based perspective, where pipes are the key elements of the network characterized by specific asset features (e.g., length, diameter, pipe resistance, etc.). The second

step was to create a WDN-tailored network topology to embed the different hydraulic roles of nodes: source nodes, tanks, and demand nodes. The third step was to evaluate edge centralities in the tailored topology where pipes are weighted with appropriate domain information (e.g., pipe resistance). The last step involved the prior information about the flow direction in pipes where unidirectional devices are installed or next to the reservoirs.

4.1.1 Edge centralities

The functioning of a WDN entails two main operations: (i) the water is delivered at pipe level and (ii) nodes are network elements that transfer the water (i.e., information) among pipes. This first definition clarifies why pipes, with their asset features, represent the most important network components for WDNs. Furthermore, focusing on pipes allowed us to attribute to them asset characteristics, generally neglected in topological analysis before. For example, considering the hydraulic resistance as attribute for pipes allows to move the network domain analysis closer to the expected hydraulic behaviour of the system.

We moved the focus from nodes to pipes assuming, as metric to be used, the edge betweenness (Girvan and Newman, 2002) and tailoring closeness, standard and harmonic, and degree centrality metrics with respect to edges. One way to evaluate the effectiveness of the new topological metric was to identify the hydraulic behaviour of the networks through the evaluation of a hydraulic metrics and then proceed to a comparison of them. Obviously, the greater the correspondence between the two metrics, the more the topological metric was suitable in identifying the WDNs domain. We referred to the pipe flowrate as hydraulic metric, also based on the concept of shortest hydraulic paths within the network.

Edge Betweenness: *Edge Betweenness* is a global centrality metric that measures the importance of an edge counting how many times it is traversed by shortest paths between two vertices, i.e., quantifying how an edge intervenes in the transfer of information through the network. It seems clear that the edge with the most strategic location is the most important. The *edge betweenness centrality* of an edge *l* can be defined as (*Girvan and Newman*, 2002):

$$C_l^B = \sum_{\substack{s \neq t \in G \\ l \in E}} \frac{\sigma_{st}(l)}{\sigma_{st}} \tag{35}$$

Edge Closeness Centrality: *Edge Closeness* is a centrality metric that measures the edge efficiency in spreading the information and it considers the geodesic distance (shortest path) between pipes rather than between nodes, denoted by the reciprocal of the sum of all distances

of a pipe from all other pipes. I.e., it measures the centrality of a pipe considering how it is relatively close to all other pipes. The *Closeness centrality* of an edge *l* can be defined as:

$$C_l^C = \frac{1}{\sum_e d_{l,e}} \tag{36}$$

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

Edge Degree Centrality: *edge degree* is the number of edges connected to one of the two ending nodes of that edge, i.e., the degree computed using the line graph matrix. The extension of the edge degree (n-EN) is similar to the *n*-neighborhood degree involving edges at level *n* (*Giustolisi et al*., 2017).

Hydraulic centrality metrics: *hydraulic centrality metrics* can be defined considering the hydraulic status of WDNs (e.g., pipe flowrate) and given boundary conditions (e.g., costumers demand, tanks level, etc). We assumed that the domain analysis using the WDN-tailored metrics allows understanding in advance the role of pipes in determining the system hydraulics independently on the specific set of boundary conditions. For example, the edge betweenness informs about the most relevant pipes transferring water similarly to the pipe flowrate centrality. This means that topological (e.g., edge betweenness) and hydraulic centralities (e.g., pipe flowrate) are correlated to each other, i.e., the information obtained computing the topological centralities can be helpful for identifying, in advance, the "emerging" hydraulic behaviour of WDNs without the need of hydraulic simulations and independently on the specific boundary conditions. The drawback is that the topological centrality metrics lack information about the hydraulic status, e.g., the presence of source nodes, that represent a sort of hubs for WDNs, and of directional devices, that constrains the flow direction. Furthermore, (i) although the edge betweenness is very similar to the flowrate, the topological metric does not consider pipe features (e.g., length, hydraulic resistances, etc.) that instead are embedded in the hydraulic one; (ii) hydraulic metrics are time varying, differently from the topological ones, since the WDN hydraulic behaviour changes over time; (iii) the hydraulic metrics accounts for any change concerning the network, e.g., addition of reservoirs or tanks in existing nodes, installation of unidirectional devices, addition or removal of nodes, while the topological metrics account only for the last one. In fact, the transformation of a demand node into a source node significantly affects the hydraulics of the system without, however, affecting its topology.

4.1.2 Source nodes

We highlighted several times the need to assign different roles to the nodes within the system. Accordingly, we created a hydraulic-based topology that accounts for the different hydraulic role of source nodes with respect to other nodes. The first step was inserting the domain information about reservoirs assuming: (i) the presence of N_f connected fictitious nodes, (ii) M reservoirs and (iii) N_d demand nodes in the network. For each reservoir exists a star of N_f connected fictitious nodes whose number is equal to N_d/M (See Figure 33). Increasing the number of nodes connected to a reservoir increases its topological relevance and the reservoir act as a sort of hubs with respect to the hydraulic system behaviour.

In fact, in this way, the shortest paths starting from the fictious nodes (all forced to cross the reservoir to reach the other nodes of the network) will cross at least N_f times each reservoir and the relevance of the original position of the reservoir with respect to the entire network structure is preserved although amplified according to N_f, i.e., the topological relevance of the reservoirs is multiplied by N_f. It is important noting that without reservoirs, and therefore stars of fictious nodes, the connectivity structure degenerates into the original one, which coincides with the topological one, i.e., hydraulics does not exist.

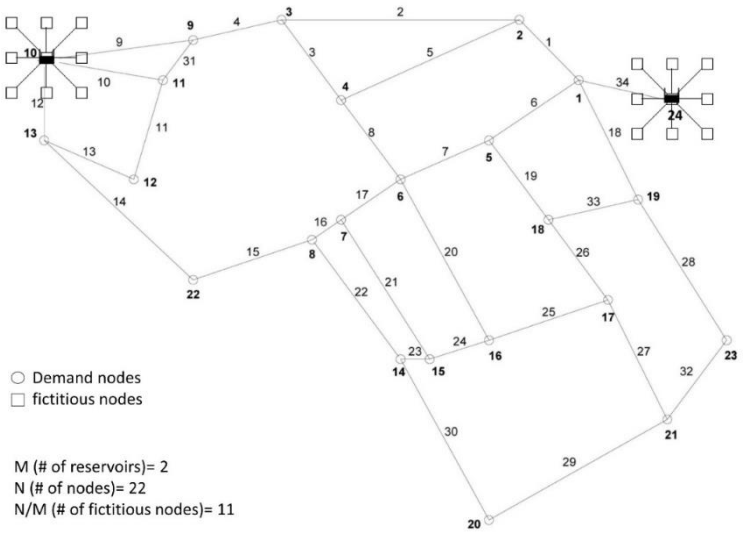


Figure 33. Hydraulic-based topology obtained inserting the domain information about reservoirs. Stars of fictious nodes N_f are inserted in correspondence of each reservoir in order to increases its topological relevance. The Figure reports Apulian network with two tanks. In this case, M= 2, N_d = 22 and, therefore, N_f = 11 for each reservoir. Source: *Giustolisi et al*., 2019.

It is important to keep in mind that if the influence area of each reservoir, i.e., how many nodes it feed, is known, a different number of demand nodes N_d and therefore of fictious nodes

N_f exists for each reservoir. Conversely, if there is no information about the influence area, each reservoir has the same number of N_d and N_f.

The next step was inserting the domain information about tanks, i.e., nodal sources characterized by an emptying/filling process that occurs during the operating cycle of the hydraulic system. Here we acted considering the maximum water volume V_T that can be supplied during the emptying process by the tank and the average water volume V_D supplied during the operating cycle to the demand nodes. The procedure is like that proposed for reservoirs, with the difference that in this case the star of fictitious nodes considers the tank volume and the number of demand nodes based on the emptying/filling process. We assumed N_f connected fictitious nodes, already attributed to reservoirs, and N_d demand nodes. The number of fictious nodes attributed to each tank equals the minimum integer value between $N_d \cdot V_T/V_D$ and N_f. Decreasing the volume V_T the tank node tends to a demand or connectivity node and increasing the volume V_T the tank node tends to a reservoir. However, even if tanks are hubs too, they always have less or equal importance of reservoirs because the number of their fictitious nodes can be equal or less to those traversing the reservoir. Furthermore, V_T also influence the relevance of pipes close to the tanks. Looking at the Figure 34, it is possible to state that if V_D is larger than V_T, the pipe between reservoir and tank (pipe 1) is much more relevant than the pipe between tank and distribution system (pipe 2), because it still represents the main source of supply for the network. Differently, when V_T and V_D are comparable, pipe 2 becomes the most relevant because is able to ensure the supply of the network for an operating cycle, i.e., for the time necessary for repairing/replacing the pipe 1.

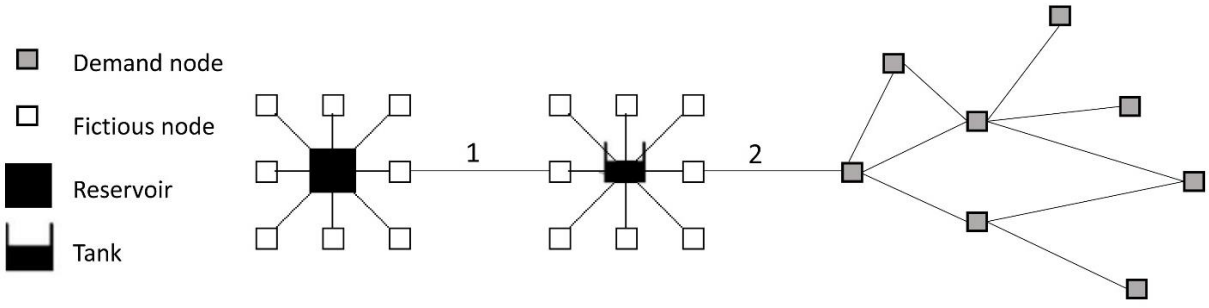


Figure 34. Scheme of the WDN-oriented topology modifications to account reservoirs and tanks. Source: *Giustolisi et al*., 2019.

4.1.3 Directional devices

The WDNs topology has been commonly analysed by considering the association of these systems to indirect networks, i.e., networks whose features do not present a flow direction. For WDNs this association translates into considering networks without unidirectional devices, or,

in any case, neglecting their possible existence. In order to perform a complete topological analysis of WDNs, the next step has been to consider unidirectional devices and their variation over time in the tailoring strategy. In fact, the boundary conditions, linked to the presence of such devices, define the flow direction, influencing the shortest paths and, therefore, the domain of the system. It can happen, for example, that the shortest paths between two pipes, one upstream and the other downstream of a specific device, can be traversed only in one direction, i.e., that of the device. The consequence is that the edge betweenness of these pipes is automatically reduced compared to that computed on the network without devices. An example of how the presence of devices affects the flow within a network is given in Figure 35 (Top), that reports a simple network fed by two reservoirs (A and B), composed of 2 connection nodes and 4 demand nodes. A check valve (CH) is installed between the reservoir A and the demand node 1. The direction of the CH is from node 1 toward reservoir A. The behaviour of the network changes if we consider direct or indirect network, i.e., whether or not we consider the presence of the CH. Not assuming the unidirectional device, the repartition of the demand nodes is equal for each reservoir, i.e., each reservoir can reach all four demand nodes. Differently, assuming the presence of the unidirectional device, the flow starting from the reservoir A is unable to reach the four demand nodes because the CH directs the flow in the opposite direction. Differently, the flow that starts from reservoir B reaches all demand nodes, going in the same direction of the CH. Figure 35 (Bottom) enriches the example with other cases.

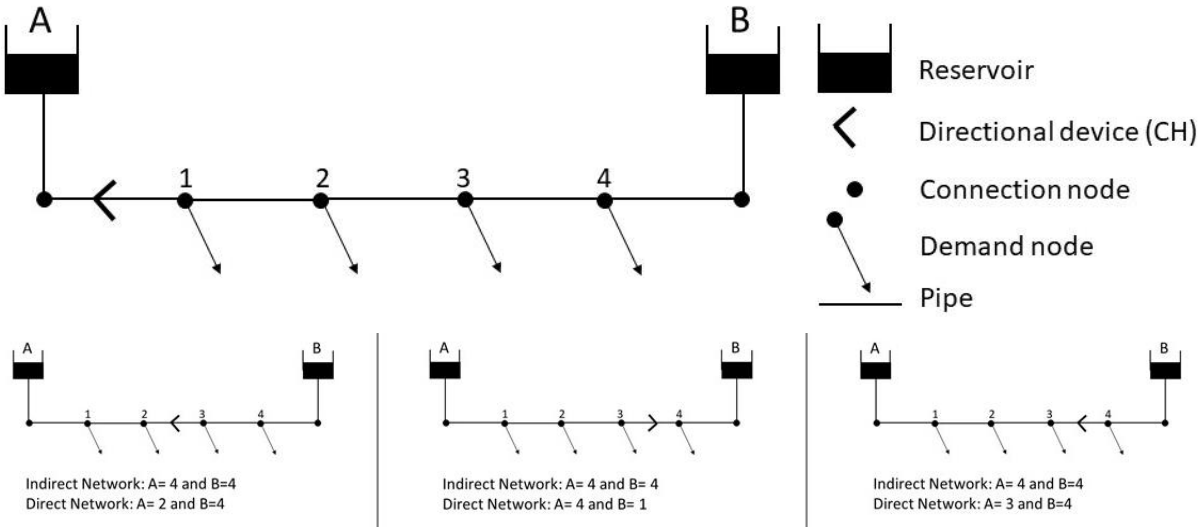


Figure 35. Top. Simple network fed by two reservoirs (A and B), composed of 2 connection nodes and 4 demand nodes and a check valve (CH). The system domain, influenced by the flow, changes considering the network as direct or indirect, i.e., considering, or not the presence of the unidirectional device (CH). **Bottom**. The same network positioning the CH in three different ways. The reachable nodes, both in the case of a direct or indirect network, are reported below each example. The results do not change for the indirect network, while they vary with the variation of the position and of the directionality of the CH. This example makes us

understand how it is important considering unidirectional devices even in the topological analysis, i.e., in defining network domain.

4.1.4 Case studies

The present chapiter show how the WDN-tailored edge betweenness, properly tailored for accounting the WDNs features, represents a useful tool for the domain analysis of real WDNs and a support for the hydraulic analysis of these systems. Here we report the application of the WDN-tailored edge betweenness to several real networks with different size, asset features and hydraulic schemes. The procedure is detailed using the BBLAWN network (*Giustolisi et al.*, 2019) while it only reports and discusses the results for the other six networks.

BBLAWN WDN is composed of 390 nodes, 439 pipes, 1 reservoir, 7 tanks, 11 pumps, 1 control valve and 1 check valve. A star of fictitious nodes (N_f=N_d as M=1) has been associated to the only reservoir and allows it of acting as a topological hub in the system. The hydraulic metric flowrate is computed for the network, and it is compared with the standard edge metric, weighted with the connectivity, in Figure 36.

The metric pipe flowrate identifies the pipes close to the reservoir as the most significant (Figure 36-left), while the standard edge betweenness fails in ranking correctly pipes close to the reservoir, not recognizing them as the most significant (Figure 36- right). It is important noting that the analysis only reports the 5% highest values, both for topological and hydraulic metrics. In this perspective, we can state that pipes with edge betweenness, weighted with connectivity, falling in the 5% highest values, does not correspond to the pipes with the highest flow rates, i.e., the highest 5% values of the average values for one-day hydraulic simulation.

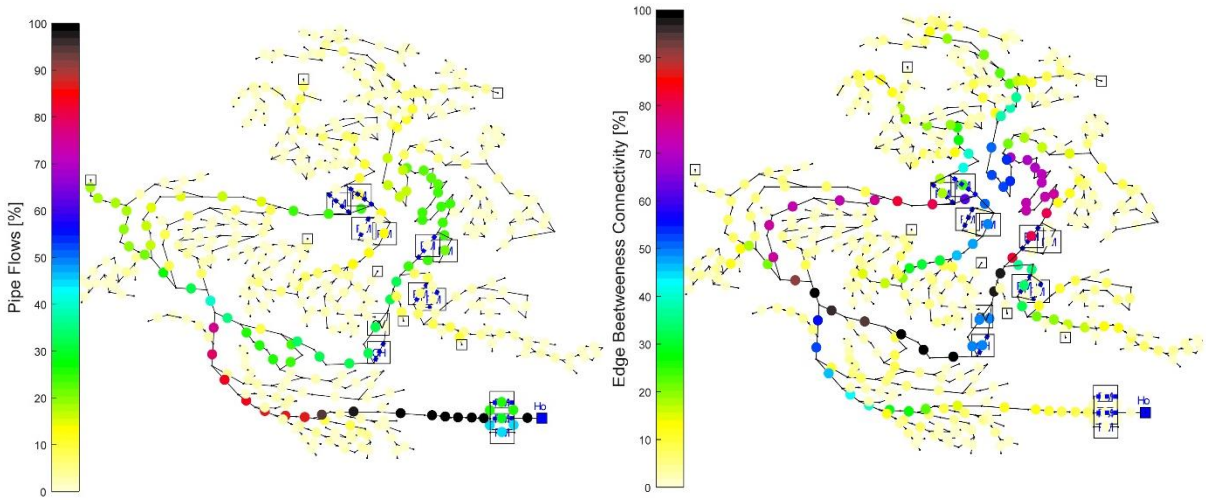


Figure 36. Hydraulic metric pipe flowrate (**left**) and normalized standard edge betweenness (**right**) for BBLAWN network. The standard topological metric does not identify the most important paths as identified by the flowrate. In particular, the metric does not recognize the pipe close to the reservoir as the most important. Source: *Giustolisi et al.*, 2019.

Applying the tailoring process, the relevance of pipes/ paths close to the reservoir increases (Figure 37- left). I.e., the tailored metric is now able to capture the relevant hydraulic features of the system with respect to the source nodes. Using the pipe hydraulic resistance as weight for pipes (see Figure 37- right), the analysis is getting closer and closer to the hydraulic behaviour of the system because the resistance drives the water fluxes.

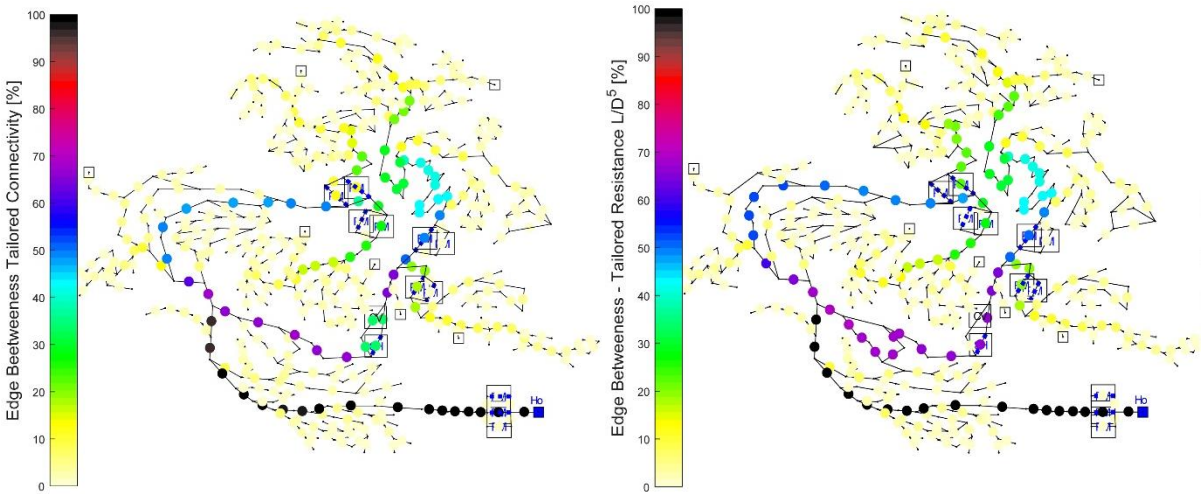


Figure 37. WDN-tailored edge betweenness for BBLAWN network using pipes connectivity (**left**) and hydraulic resistance (**right**) as weight. Source: *Giustolisi et al*., 2019.

The comparison between flowrate (Figures 36- left) and WDN-tailored edge betweenness weighted with the hydraulic resistance (Figure 37- right) demonstrates that the two metrics identify the same pipes as more important, i.e., the modification of the network topology considering source nodes and weighting edge betweenness by pipe resistance allows one to obtain an effective domain analysis. Inserting the domain information about tanks, the connectivity structure of the system changes again (Figure 38-left). This time, the metric does not identify the pipe close to the only reservoir as the most important for WDN hydraulics, attributing a low relevance to the line starting from the main source of water. Inserting the information about the directional devices the problem is solved (see Figure 38-right), i.e., the metric well captures the emerging hydraulic behaviour of the system. This last step highlights that the information on directional devices enhance the domain analysis.

Results show that the WDN-Tailored Edge betweenness, just considering the domain structure of the system, identifies quite all the most important pipes as ranked by pipe flow rates. Pipes close to the only reservoir result the most important for the WDN hydraulics, while the relevance of the other pipes does not change although it has a new parametrization referred to the most important one. The Spearman correlation index (*Spearman*, 1904) between WDN-Tailored Edge betweenness, weighted with hydraulic resistance, and pipe flow rate is equal to

0,74. i.e., there is a high correspondence between two metrics, although the topological metric focuses only on the domain structure and the hydraulic one depends also on the boundary conditions and on momentum and continuity equations.

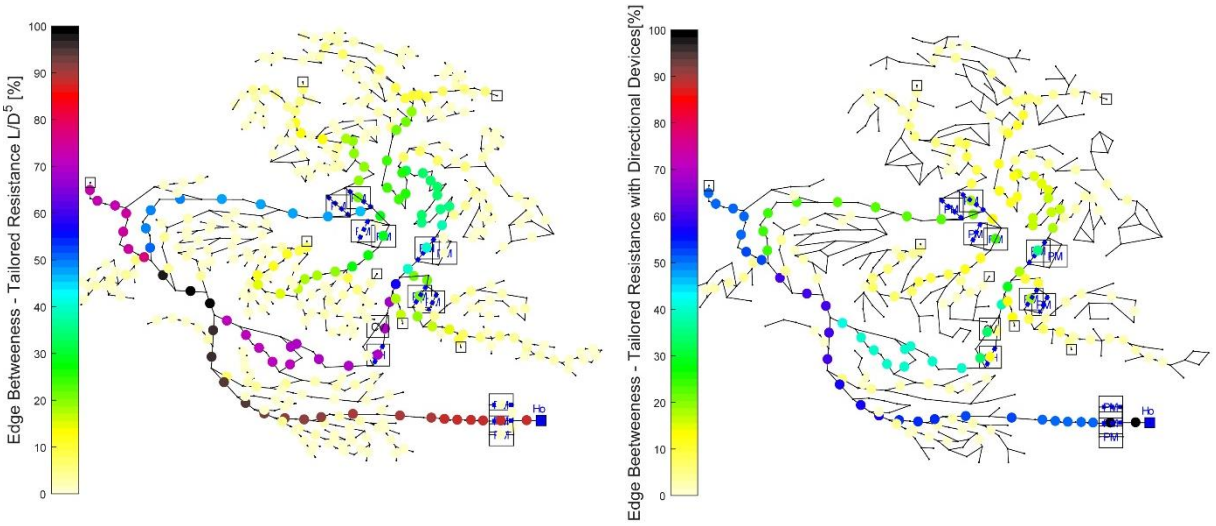


Figure 38. WDN-tailored edge betweenness using pipes resistance and tank information. The right panel also considers information about directional devices. Comparing the right panel with Figure 36-left we can state that the topological metric is able to capture the emerging hydraulic behaviour due to the connectivity structure of the network. Source: *Giustolisi et al.*, 2019.

Following (Figure 39-44), we report the flowrate (left panel) and the WDN-Tailored edge betweenness (right panel) for the remaining real networks.

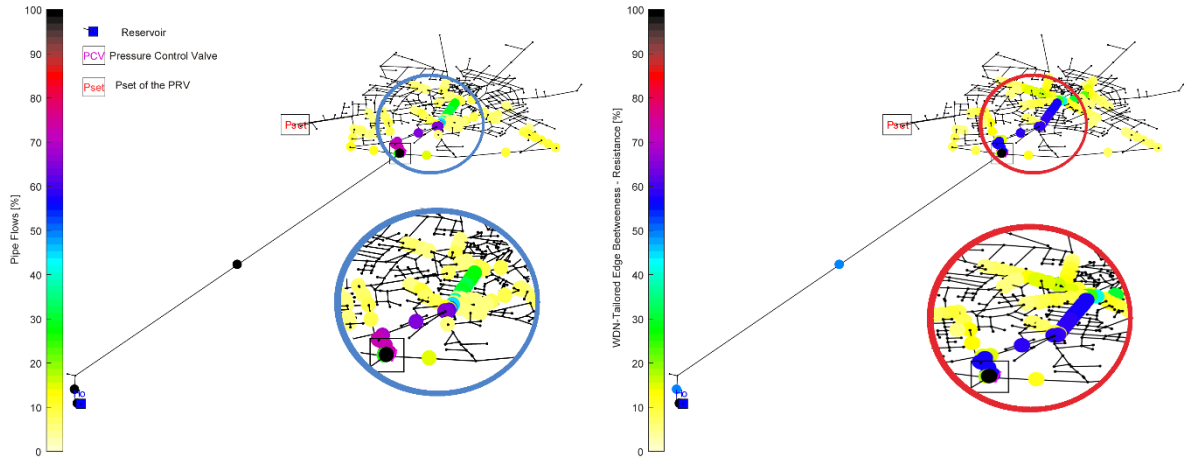


Figure 39. Pipe flow (**left**) and WDN-tailored edge betweenness weighted with pipes resistance (**right**) for Apulia 1 network. The network is composed of 986 nodes, 1,105 pipes and a single reservoir from which the main path starts to reach the centre of the network. The correlation index between two metrics is equal to 0,64, i.e., the edge betweenness well identifies the hydraulic behaviour of the networks. Source: *Simone et al.*, 2019.

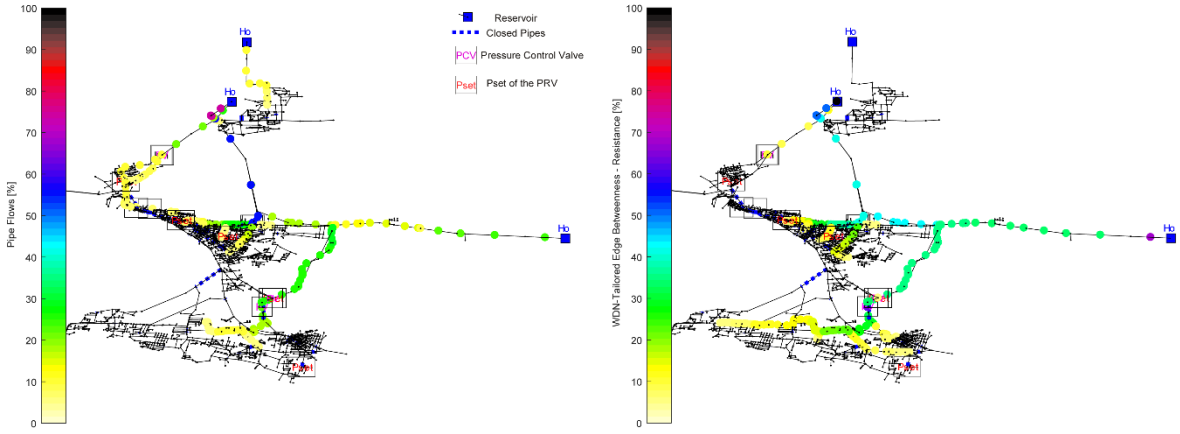


Figure 40. Pipe flow (**left**) and WDN-tailored edge betweenness weighted with pipes resistance (**right**) for Apulia 2 network. The network is composed of 7,716 nodes, 8,496 pipes, 3 reservoirs five pressure control valves and presents a wide range of diameters. The WDN-Tailored edge betweenness identifies quite well all the most important pipes in terms of pipe flow rates. In fact, the correlation index between two metrics is equal to 0,60, i.e., the edge betweenness well identifies the hydraulic behaviour of the networks. Source: *Simone et al.*, 2019.

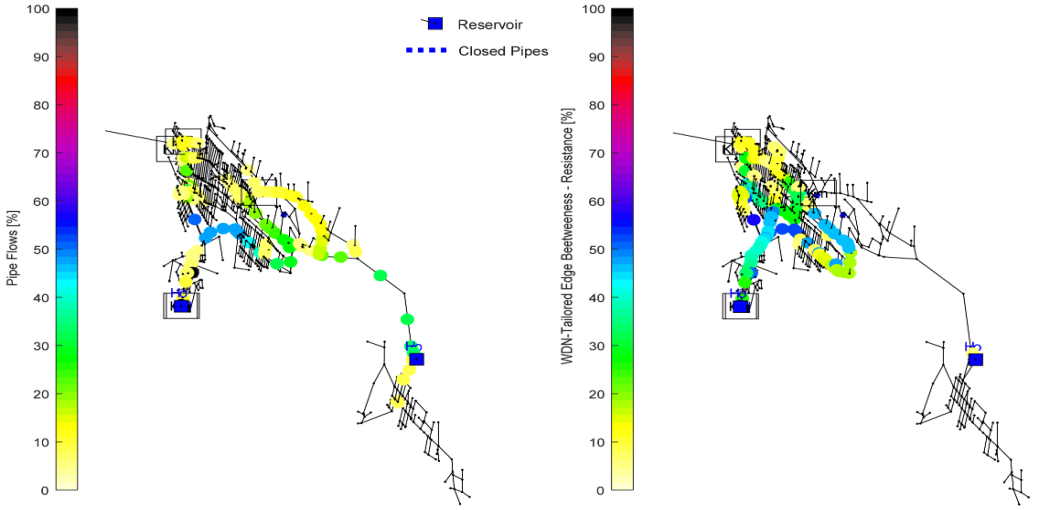


Figure 41. Pipe flow (**left**) and WDN-tailored edge betweenness weighted with pipes resistance (**right**) for Apulia 3 network. The network is composed of 1,270 nodes, 1,472 pipes and 3 reservoirs. The metric identifies various main paths within the network probably since of the presence of three reservoirs. Since each reservoir has its own area of influence in terms of demand nodes, each main path has a different relevance. The result is consistent with the information provided by the flow rate. In fact, the correlation index between two metrics is equal to 0,54, i.e., the WDN-Tailored the edge betweenness well identifies the hydraulic behaviour of the network. Source: *Simone et al.*, 2019.

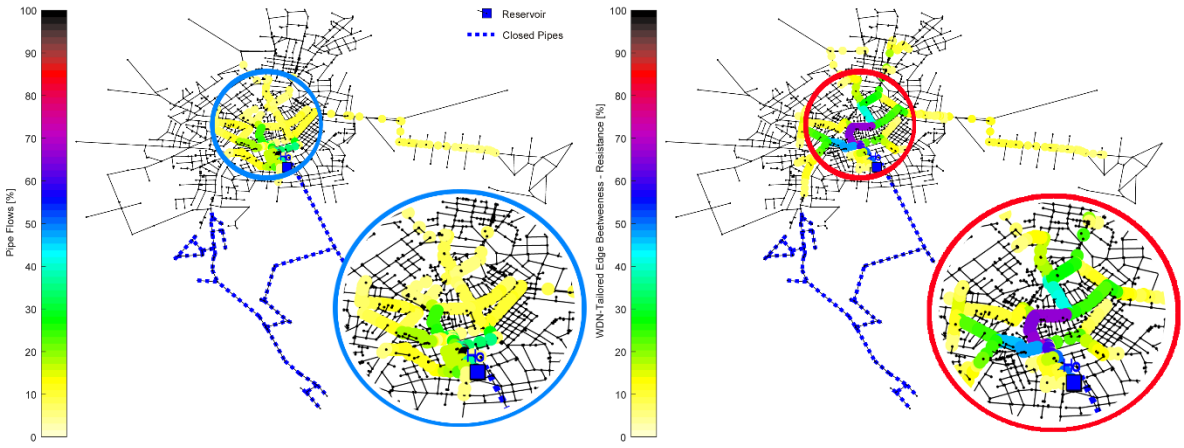


Figure 42. Pipe flow (**left**) and WDN-tailored edge betweenness weighted with pipes resistance (**right**) for Apulia 4 network. The network is a very looped network composed of 3,166 nodes, 3,483 pipes, 2 reservoirs and a portion of network no longer in use (dotted blue). The high redundancy favours the presence of several paths with similar relevance. The WDN-Tailored edge betweenness identifies a greater number of paths as relevant with respect to the pipe flow rates, probably because the network is oversized, and the topologic metric considers the diameters and not the flow that effectively crosses them. The correlation index between two metrics is equal to 0,66, i.e., the edge betweenness well identifies the hydraulic behaviour of the networks. Source: *Simone et al.*, 2019.

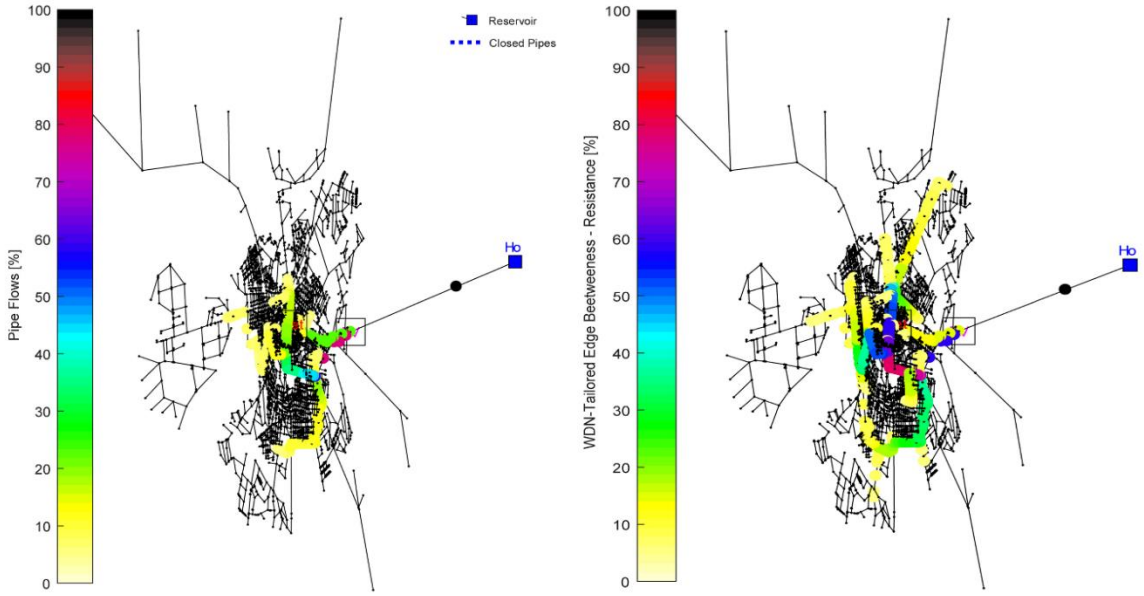


Figure 43. Pipe flow (**left**) and WDN-tailored edge betweenness weighted with pipes resistance (**right**) for Apulia 5 network. The network is composed of 7,164 nodes, 7,895 pipes, 1 reservoir and a pressure control valve downstream of the reservoir. The WDN-Tailored edge betweenness identifies a different number of paths as relevant with respect to the pipe flow rates, i.e., the flow has no direct correspondence with the resistance for all the pipes. However, the correlation index, equal to 0,62, shows once again that the analysis is close to the hydraulic of the system. Source: *Simone et al.*, 2019.

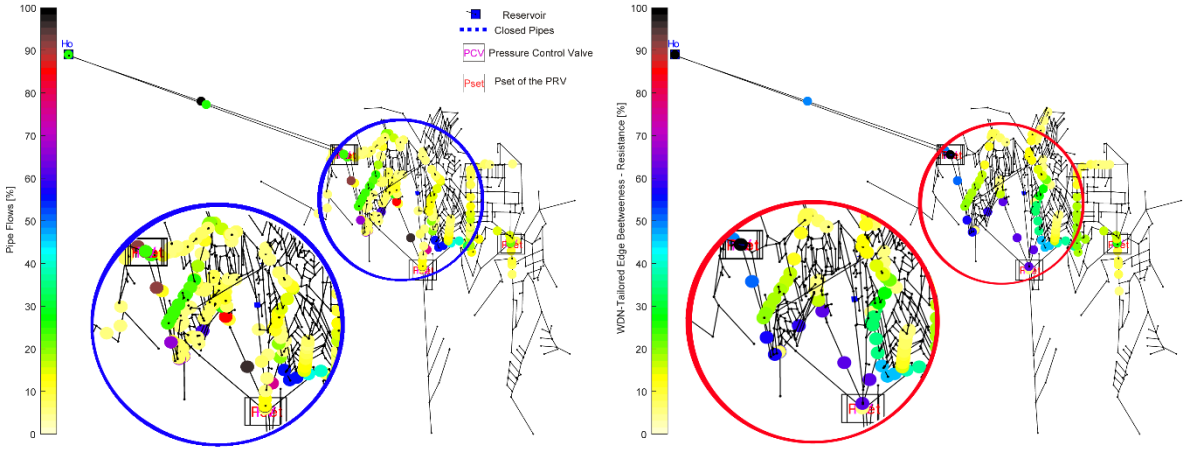


Figure 44. Pipe flow (**left**) and WDN-tailored edge betweenness weighted with pipes resistance (**right**) for Apulia 6 network. The network is a very looped network composed of 1,111nodes, 1,307 pipes, 4 pressure control valves and 1 reservoir. The hydraulic metric highlights the mains pipes without defining properly a main path. The same happens with the topological metric, confirming once again that the edge betweenness well identifies the hydraulic behaviour of the network. The correlation index between two metrics is equal to 0,54. Source: *Simone et al*., 2019.

Table 2 reports the Spearman correlation index (*Spearman*, 1904) between two metrics of the six analysed networks considering several weights for the edge betweenness, i.e., connectivity, length, and resistance.

WDN Name	Node #	Pipe #	Reservoirs #	Tanks #	Corr. Index Resistance	Corr. Index Connectivity	Corr. Index Length
Apulia 1	986	1,105	1	0	0,64	0,71	0,73
Apulia 2	7,716	8,496	3	0	0,60	0,63	0,65
Apulia 3	1,270	1,472	3	0	0,54	0,50	0,53
Apulia 4	3,166	3,483	2	0	0,66	0,75	0,79
Apulia 5	7,164	7,895	1	0	0,62	0,70	0,75
Apulia 6	1,111	1,307	1	0	0,54	0,59	0,60

Table 2. Topological characteristics and correlation index for the six analysed WDNs.

The results obtained by applying the WDN-tailored edge betweenness to real WDNs, confirmed the usefulness of the metric for the WDN domain analysis. The domain analysis provided correlation index values ranging in the interval [0,50; 0,79].

Indeed, although the WDN-Tailored edge betweenness cannot replace the hydraulic simulation, it might provide useful indications to drive calibration, maintenance works as well as relevance of pipes to plan operations.

4.2 Paradigm of the Intrinsic Relevance of Vertex/ Node

We conceived and developed a strategy with the aim to enhance the domain analysis of real systems without resorting to artifices, for example linked to the need to increase the topological relevance of specific nodes (e.g., source nodes for WDNs). The assumption was that whatever the system analysed, each vertex/node has an own intrinsic relevance, independent from the topological one, that must be embedded into the analysis.

CNT tools have always evaluated the importance of elements only considering the topological features, i.e., the connective structure and the position of the elements. Consequently, each type of analysis, assuming an identical intrinsic relevance for all nodes, neglected a part of the network information.

For example, the strategy presented in the previous paragraph, i.e., the idea to assign a star of fictitious nodes in correspondence of the source nodes in order to increase their topological relevance, arises precisely from the impossibility of considering an intrinsic relevance for this type of hubs. For the specific case, since demand nodes have an intrinsic relevance generally similar to each other and much lower than that of the source nodes, assuming identical relevance for them did not significantly affected the success of the analyses. The problem arises from the fact that a percentage of demand nodes is represented by strategic structures (e.g., hospital, schools, etc.) other than the source nodes. The assumption of the same intrinsic relevance for these nodes led to a loss of useful information on the main paths between the strategic structures within the networks. Therefore, also for these nodes it would be necessary to propose a way to increase their topological relevance. We found the solution to this problem. By embedding the intrinsic relevance of all nodes within the analysis, it will no longer be necessary to propose artifices to increase the topological relevance of specific nodes to obtain reliable results.

In WDNs, nodes with many connections, which can represent release nodes in the network, present a high topology-based centrality correctly detected by classical CNT tools, because high number of connections corresponds to high centrality. Differently, a source node generally presents a low number of connections and CNT tools cannot correctly quantify their actual centrality in the network, because a low number of connections does not correspond to low centrality. The finding is that not considering the information about the intrinsic relevance of nodes can give misleading outcomes for some networked systems. Therefore, with reference to the study of WDNs it is important to consider the intrinsic relevance of all nodes, i.e., the intrinsic relevance referred to demand, strategic and sources nodes need to be embedded in the centrality evaluation to make useful CNT for such networked systems.

The intrinsic relevance R_n ($n=1, \ldots, N$) of each node is an exogenous information depending on the type of network and analysis to be performed. For WDNs, characterized by source nodes supplying water to other nodes, the intrinsic relevance could be assigned equal to the demand for each demand node and to the sum of demand for the source node. This way, all nodes assume a different relevance during the analysis. The strategy is performed using several functions $f(R_s, R_t)$ that depend on the intrinsic relevance of the ending nodes s and t of the generic edge l. The values of the function $f(R_s,R_t)$ can be stored in a matrix having the same dimension of the adjacency matrix with zero diagonal elements, i.e., a matrix, generally asymmetric ($f(R_s,R_t) \neq f(R_t,R_s)$), is constructed each time by observing the specific system analyzed.

The function $f(R_s,R_t)$ has a problem-dependent structure, i.e., the selection of a specific function allows to embed a diverse exogenous information in the standard centrality metrics. For example, when we assume $f(R_s,R_t) = R_s$ the degree $d(s)$ and closeness $C(s)$ are simply scaled by R_s and the relevance of the node s does not influence its neighbors, meaning that we are only increasing the degree and the capacity to transfer information of the specific node. I.e., centrality metrics increase as the intrinsic relevance of the vertices increases. It follows that the intrinsic relevance of the nodes does not coincide with the pipe weights because the first scales the inverse distance (harmonic centrality) or determines the importance of the shortest paths (betweenness) while the second considers the strength of the connections and the effect in determining the shortest path and the distance between the vertices (*Giustolisi et al.*, 2020).

It is important noting that the novel centrality metrics, obtained embedding the intrinsic relevance of nodes, represent a generalization of the classic ones, i.e., $f(1,1) = 1$. Examples of functions are: (i) $f(R_s, Rt) = R_s \cdot R_t$; (ii) $f(R_s, Rt) = (R_s + R_t)/2$; (iii) $f(R_s,R_t) = R_s$ and (iv) $f(R_s,R_t) = \max[R_s, R_t]$; (v) $f(R_s, R_t) = \Sigma R_{path}$ and (vi) $f(R_s, R_t) = \Pi R_{path}$ (where R_{path} indicates the relevance of vertices crossed in the path).

Furthermore, $f(R_s, R_t)$ can be considered as a constitutive function that allows to specify the characteristics of the network to be analyzed and depends on the specific characteristics of the problem under examination. By increasing the knowledge of the constitutive function, and therefore of the network, the accuracy of the analysis increases. From this point of view, it is possible to identify the functions $f(R_s,R_t)$ most suitable for the various physical systems, also considering that it should be monotonically increasing.

The strategy draw attention to the importance of embedding the information about the intrinsic relevance of nodes into CNT tools to enhance the WDNs analysis. This choice allowed to obtain concrete results on the actual role of topological domain features in the emergent hydraulic behaviour.

4.2.1 Embedding the intrinsic relevance of vertices to the centrality metrics: degree, harmonic and betweenness centrality

We extended the standard definitions of degree, closeness, harmonic closeness and betweenness to embed the intrinsic relevance of nodes, showing that this information can increase the capability of centrality tools to assess the relevance of vertices and edges.

Degree centrality

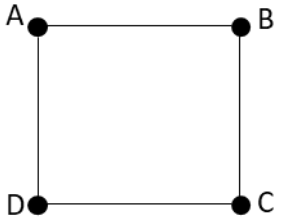
The standard degree of Eq. (19) assumes that all vertices are characterized by an identical intrinsic relevance and that the metric exclusively depends on the connectivity structure,

$$C_s^D = D(s) = \sum_t A_{st} \qquad t = 1 \ldots V \tag{37}$$

We assumed that each node has an own intrinsic relevance, defined as R_v (v=1, …, V), i.e., the metric does not exclusively depend on the connectivity structure. The function $f(R_s, Rt)$, that depends on the relevance of vertices s and t, weights the importance of the connection of the vertex s with its neighbors. Hence, the definition of the standard degree can be extended to embed the information about the intrinsic relevance of the vertices through the function f (R_s, R_t) as follows,

$$C_s^D = D(s) = \sum_t A_{st} f(R_s, R_t) \qquad t = 1 \ldots V \tag{38}$$

The use of a pure regular network, composed of four vertices and four edges (Figure 45), help us to explains how the new metric works. The function $f(R_s, R_t) = R_s \cdot R_t$ is considered. With the assumption of identical relevance of nodes, i.e., $R_A = R_B = R_C = R_D = 1$, the degree is equal to two for each node. Doubling only the relevance of the vertex A (i.e., $R_A = 2$), $D_A = 4$, $D_B = D_D = 3$ and $D_C = 2$. I.e., increasing the intrinsic relevance of the node A increases its relevance-embedding degree and that of the adjacent nodes B and D. Therefore, associating information on intrinsic relevance and connective structure generates hubs and leads to networks more and more scale free.



$$\begin{aligned} D_A &= f(R_A, R_B) + f(R_A, R_D) \\ D_B &= f(R_B, R_C) + f(R_B, R_A) \\ D_C &= f(R_C, R_D) + f(R_C, R_B) \\ D_D &= f(R_D, R_A) + f(R_D, R_C) \end{aligned}$$

Figure 45. Regular network: the degree case. Source: *Giustolisi et al*., 2020.

Harmonic centrality

The harmonic version of the closeness of Eq. (23) assumes that all the vertices are characterized by an identical intrinsic relevance and that the metric exclusively depends on the connectivity structure,

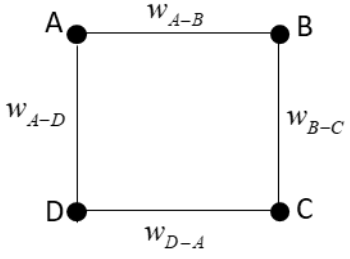
$$H_s^C = C(s) = \sum_t \frac{1}{d_{s,t}} \qquad t = 1 \ldots V \tag{39}$$

where $d_{s,t}$ is the shortest distance between vertices s and t.

We assumed that each node has an own intrinsic relevance, defined as R_v (v=1, …, V), i.e., the metric does not exclusively depend on the connectivity structure. The definition of the standard harmonic can be extended to embed the information about the intrinsic relevance of the vertices through the function $f(R_s, R_t)$ as follows,

$$H_s^C = C(s) = \sum_t \frac{f(R_s, R_t)}{d_{s,t}} \qquad t = 1 \ldots V \tag{40}$$

We use the same pure regular network of the degree case and the same function $f(R_s, R_t) = R_s \cdot R_t$. With the assumption of identical relevance of nodes (i.e., $R_A = R_B = R_C = R_D = 1$) and identical weighs for all edges, $C_A = C_B = C_C = C_D$. To explain the different role of edge weights with respect to the vertex intrinsic relevance, we assume a different weight for the edge AB, i.e., $w_{A\text{-}B} = 0.5$ and $w_{B\text{-}C} = w_{C\text{-}D} = w_{D\text{-}A} = 1$ (Figure 46), obtaining $C_A = C_B = 3.66$ and $C_C = C_D = 2.66$. Nodes A and B are characterized by a greater capacity to spread information with respect to C and D, being $w_{A\text{-}B}$ lower than the other weights.



$$w_{A-B} = 0.5$$
$$w_{B-C} = w_{C-D} = w_{D-A} = 1$$

$$C_A = \frac{f(R_A, R_B)}{w_{A-B}} + \frac{f(R_A, R_D)}{w_{A-D}} + \frac{f(R_A, R_C)}{w_{A-B} + w_{B-C}}$$
$$C_B = \frac{f(R_B, R_A)}{w_{A-B}} + \frac{f(R_B, R_C)}{w_{B-C}} + \frac{f(R_B, R_D)}{w_{A-B} + w_{D-A}}$$
$$C_C = \frac{f(R_C, R_B)}{w_{B-C}} + \frac{f(R_C, R_D)}{w_{C-D}} + \frac{f(R_C, R_A)}{w_{A-B} + w_{B-C}}$$
$$C_D = \frac{f(R_D, R_A)}{w_{D-A}} + \frac{f(R_D, R_C)}{w_{C-D}} + \frac{f(R_D, R_B)}{w_{A-B} + w_{D-A}}$$

Figure 46. Regular network: harmonic centrality case. Source: *Giustolisi et al*., 2020.

Assuming that each vertex has its own intrinsic relevance, doubling R_A, the intrinsic relevance of A increased its capacity to spread information (i.e., the value of the harmonic centrality) and that of the nearest (as distance d) nodes, obtaining $C_A = 7.33$, $C_B = 5.66$, $C_C = 3.33$ and $C_D = 3.66$. Therefore, associating information on intrinsic relevance, connective

structure, and edge weights (for weighted networks), increases the capacity to spread information of the most relevant nodes and of the nearest ones.

Betweenness centrality

The standard betweenness of Eq. (24) assumes that all the vertices are characterized by an identical intrinsic relevance and that the metric exclusively depends on the connectivity structure. The metric assigns the number of shortest paths traversing a vertex *v* for all the couples of vertices *s* and *t* of the network,

$$C_v^B = B(v) = \sum_{s \neq v \neq t \in N} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \tag{41}$$

Similarly, the edge betweenness, *EB*(*e*), assigns the number of shortest paths traversing an edge *e*, for all the couples of vertices *s* and *t* of the network,

$$C_l^B = EB(e) = \sum_{s \neq t \in N} \frac{\sigma_{s,t}(e)}{\sigma_{s,t}} \tag{42}$$

It is important noting that for weighted networks a single shortest path exists between each couple of vertices *s* and *t*,

$$\begin{aligned} B(v) &= \sum_{s \neq v \neq t \in N} \delta\left(v \in S_{s,t}\right) \\ EB(e) &= \sum_{s \neq t \in N} \delta\left(e \in S_{s,t}\right) \end{aligned} \tag{43}$$

where the Kronecker's δ function value is unitary if the vertices *v* (or the edge *e*) is member of the shortest path $S_{s,t}$ and zero otherwise.

Figure 47 indicates how betweenness is computed. Generally, given a shortest path between two nodes *s* and *t*, the standard betweenness of each traversed node (or edge) has a unitary increment (Figure 47-a), while the novel betweenness of each traversed node (edge) increases of a quantity $f(R_s,R_t)$ (Figure 47-b).

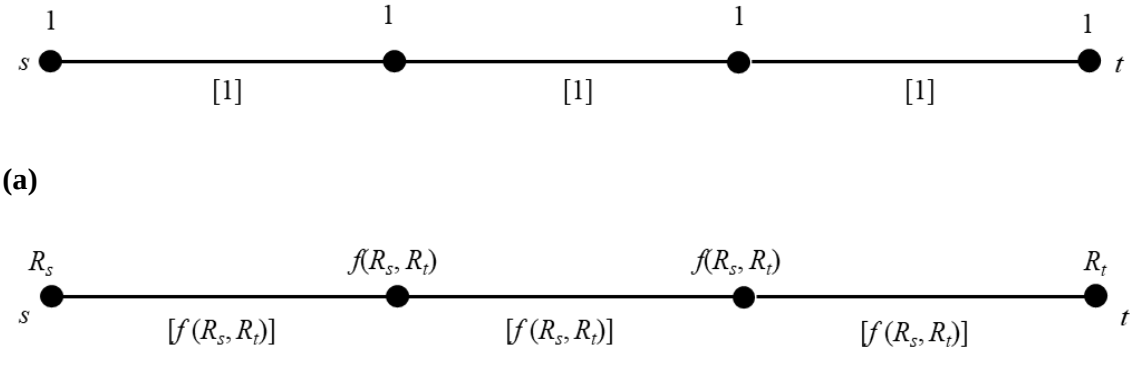


Figure 47. (**a**) Assignment of the standard betweenness to internal elements of a shortest path between vertices *s* and *t*. The values referred to the vertices are reported above the path and those

referred to the edges, shown in square brackets, are reported below the path. Notice that the betweenness of traversed vertices (or edges) has a unitary increment. After the identification of all the shortest paths between all the couples *s* and *t*, the betweenness assigns to each vertex (edge) the number of times has been traversed. (**b**) Betweenness embedding intrinsic relevance. In this case, the betweenness increment of traversed vertices (or edges) is computed from $f(R_s, R_t)$. Source: *Giustolisi et al.*, 2020.

The standard betweenness can be extended to embed the information about the intrinsic relevance of the vertices through the function $f(R_s, R_t)$ as follows,

$$\begin{aligned} B(v) &= \sum_{s\neq v\neq t\in N} f\left(R_s, R_t\right)\delta\left(v \in S_{s,t}\right) \\ EB(e) &= \sum_{s\neq t\in N} f\left(R_s, R_t\right)\delta\left(e \in S_{s,t}\right) \end{aligned} \tag{44}$$

or considering the most general case of Eqs. (40) and (41),

$$\begin{aligned} B(v) &= \sum_{s\neq v\neq t\in N} f\left(R_s, R_t\right)\frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \\ EB(e) &= \sum_{s\neq t\in N} f\left(R_s, R_t\right)\frac{\sigma_{s,t}(e)}{\sigma_{s,t}} \end{aligned} \tag{45}$$

Let us consider the same network and function $f(R_s, R_t) = R_s \cdot R_t$ of the degree and harmonic cases. Figure 48 reports the shortest paths and the formulae to compute the betweenness and the edge betweenness considering the network as unweighted. With the assumption of identical relevance of vertices, of nodes (i.e., $R_A = R_B = R_C = R_D = 1$) and identical weighs for all edges, $B_A = B_B = B_C = B_D = 1$ and $EB_{A\text{-}B} = EB_{B\text{-}C} = EB_{C\text{-}D} = EB_{D\text{-}A} = 4$. Doubling the relevance of A ($R_A = 2$), increases the betweenness of vertices and edges which need to be traversed to reach the most relevant vertices and the relevance-embedding betweenness becomes: $B_A = B_C = 1$ and $B_B = B_D = 2$ and $EB_{A\text{-}B} = EB_{D\text{-}A} = 7$ and $EB_{B\text{-}C} = EB_{C\text{-}D} = 5$. It is important noting that when two paths exist between two nodes, as for example between A and C, the factor ½ ($\sigma_{A,C} = 2$) accounts for that occurrence.

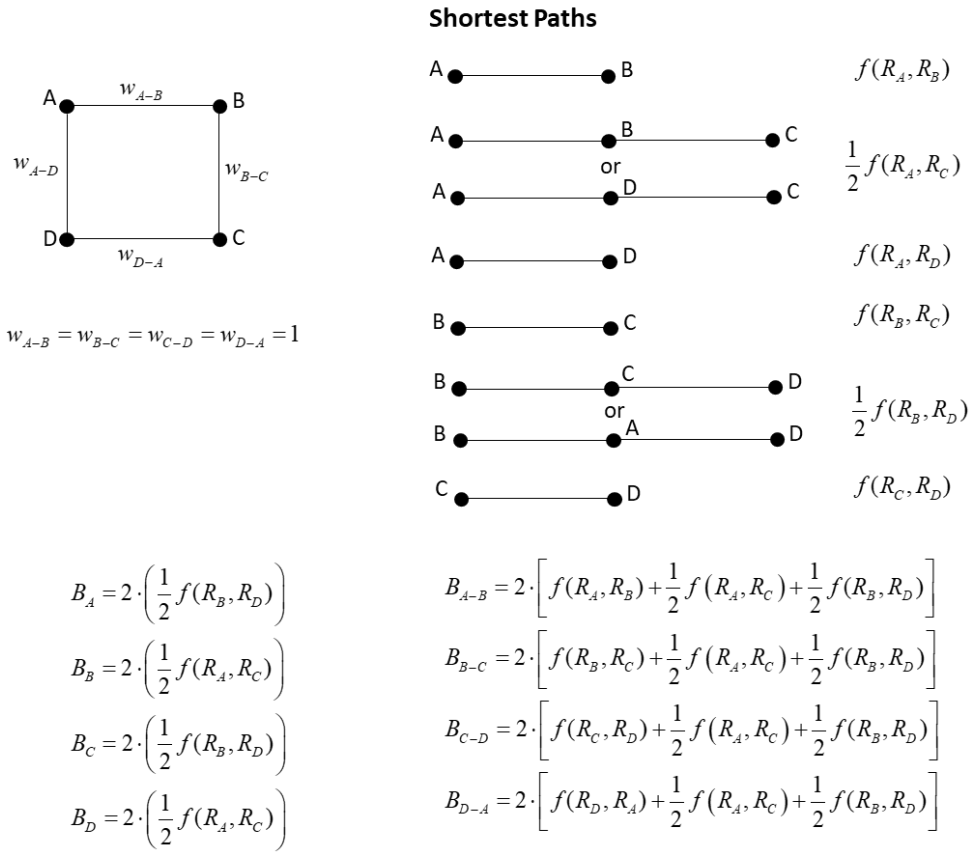


Figure 48. Regular network: betweenness case using an unweighted network. Source: Giustolisi et al., 2020.

To explain the different role of edge weights with respect to the vertex intrinsic relevance, we assume a different weight for the edge AB, i.e., $w_{A-B} = 0.5$ and $w_{B-C} = w_{C-D} = w_{D-A} = 1$. Figure 49 show the different roles of edge weights and relevance of vertices in determining the betweenness.

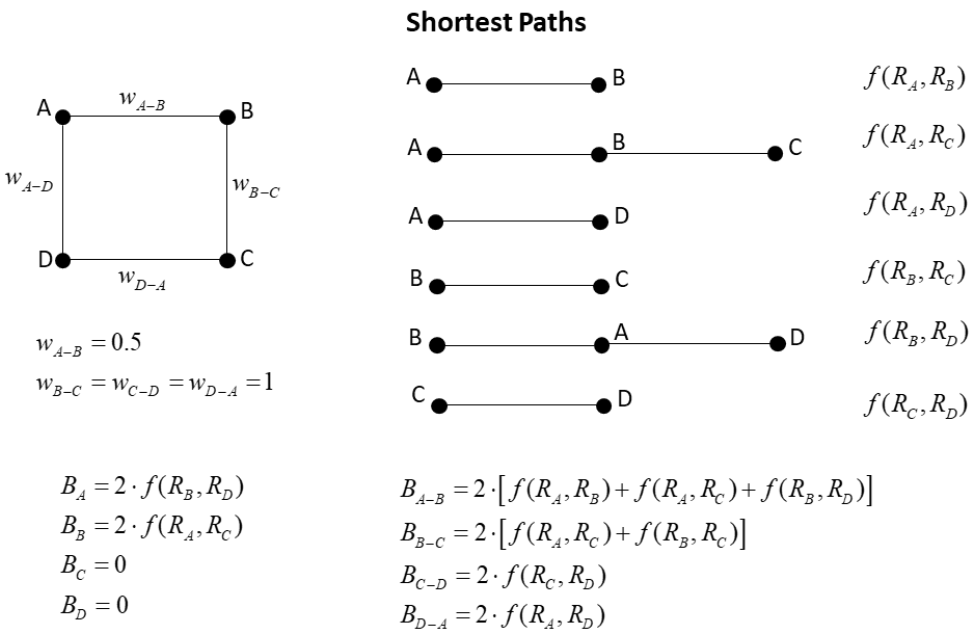


Figure 49. Regular network: betweenness case using a weighted network. Source: *Giustolisi et al*., 2020.

The assumption of identical relevance of vertices ($R_A = R_B = R_C = R_D = 1$) provides $B_A = B_B = 2$ and $B_C = B_D = 0$ and $EB_{A-B} = 6$, $EB_{B-C} = 4$ and $EB_{C-D} = EB_{D-A} = 2$. Doubling the relevance of the vertex A ($R_A = 2$), the betweenness becomes: $B_A = 2$, $B_B = 4$, $B_C = B_D = 0$ and $EB_{A-B} = 8$, $EB_{D-A} = 6$, $EB_{B-C} = 2$, $EB_{C-D} = 4$, i.e., embedding the intrinsic relevance increases the centrality of the vertices and edges which need to be traversed to reach the most relevant vertices.

The results confirm that the vertex intrinsic relevance plays an important role together with the network topology (connectivity structure and edge weights) in assessing the centrality ranking.

4.2.2 Regular and random networks

The paradigm of the intrinsic relevance is applied to two regular and random networks of 100 and 1,000 vertices, generated using the Watts and Strogatz model (Watts and Strogatz, 1999) with mean degree $d = 10$ and probability $p = \{0,1\}$. We proceeded by calculating both the standard metrics, betweenness and closeness, and the new metrics with the previously defined six functions, i.e., (i) $f(R_s, R_t) = R_s \cdot R_t$; (ii) $f(R_s,R_t) = (R_s + R_t)/2$; (iii) $f(R_s,R_t) = R_s$ and (iv) $f(R_s,R_t) = \max[R_s, R_t]$; (v) $f(R_s, R_t) = \Sigma R_{path}$ and (vi) $f(R_s, R_t) = \Pi R_{path}$. The intrinsic relevance was assigned to vertices adding to the basic unit value a random value sampled from a uniform distribution in the range $[0; d]$.

We started by considering that the generated networks were made up of nodes having the same intrinsic relevance. We proceeded by randomly selecting a fraction r= 10% of the vertices assigning them, always randomly, an intrinsic relevance. The analysis continued by randomly assigning an intrinsic relevance to all vertices (r=100%) in the same range [1; 1+d]. Therefore, we have three different cases:

- all vertices share the same relevance equal to one;
- a fraction r=10% of vertices have random relevance in the range [1; 1+d];
- all vertices (r=100%) exhibit random relevance in the range [1; 1+d].

For each network (regular and random, with 100 and 1,000 vertices), analysed with the function $f(R_s, R_t)= R_s \cdot R_t$ and pattern of intrinsic relevance (10% of randomly assigned intrinsic relevance to vertices and 100% assigned to all vertices) we report the scatter plots of:

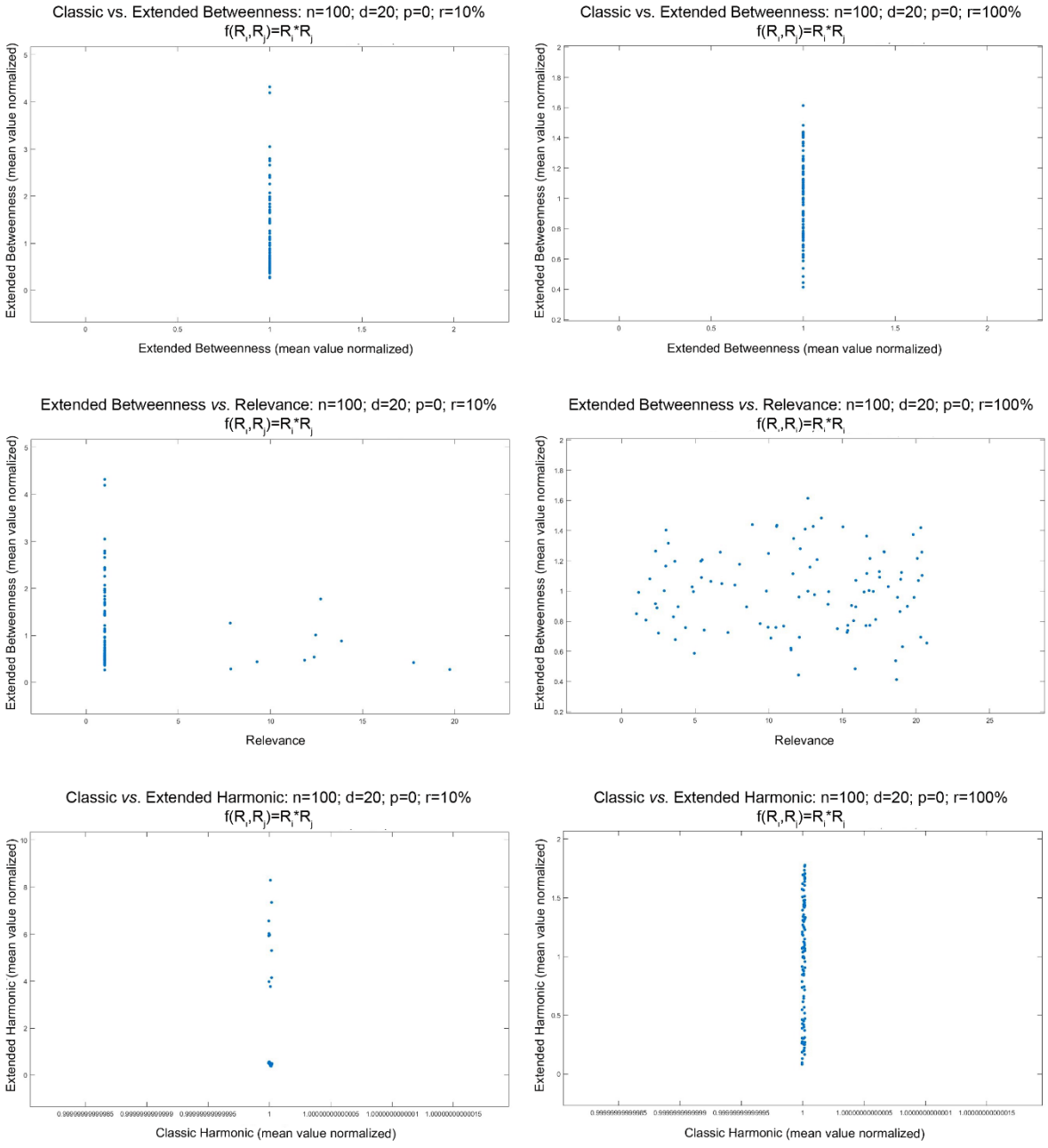
- standard versus the relevance-based betweenness;
- relevance-based betweenness versus intrinsic relevance;
- standard versus the relevance-based harmonic centrality;

- relevance-based harmonic centrality versus intrinsic relevance;

and networks with coloured vertices referring to a colour-bar indicating the values of the:

- standard betweenness;
- relevance-based betweenness;
- intrinsic relevance of vertices;
- standard harmonic centrality;
- relevance-based harmonic centrality.

The same tests were done using the remaining five functions and they are available as supplementary material.



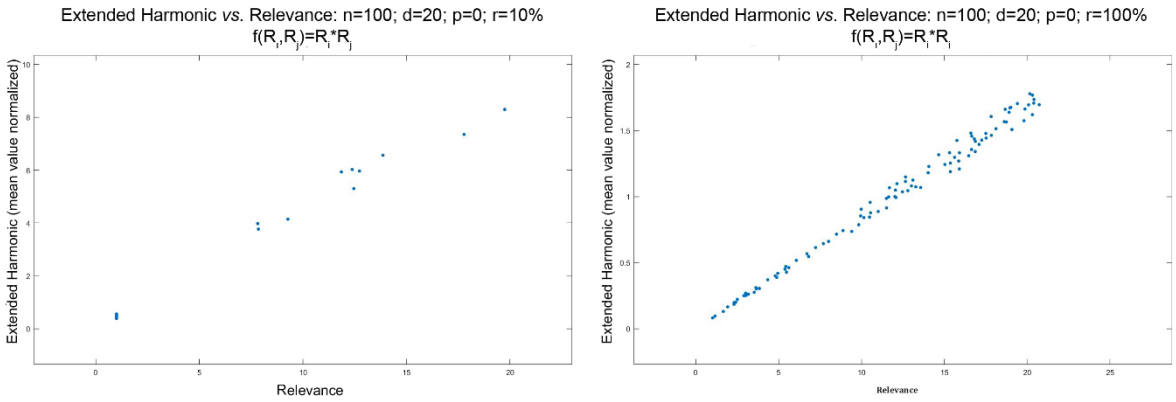
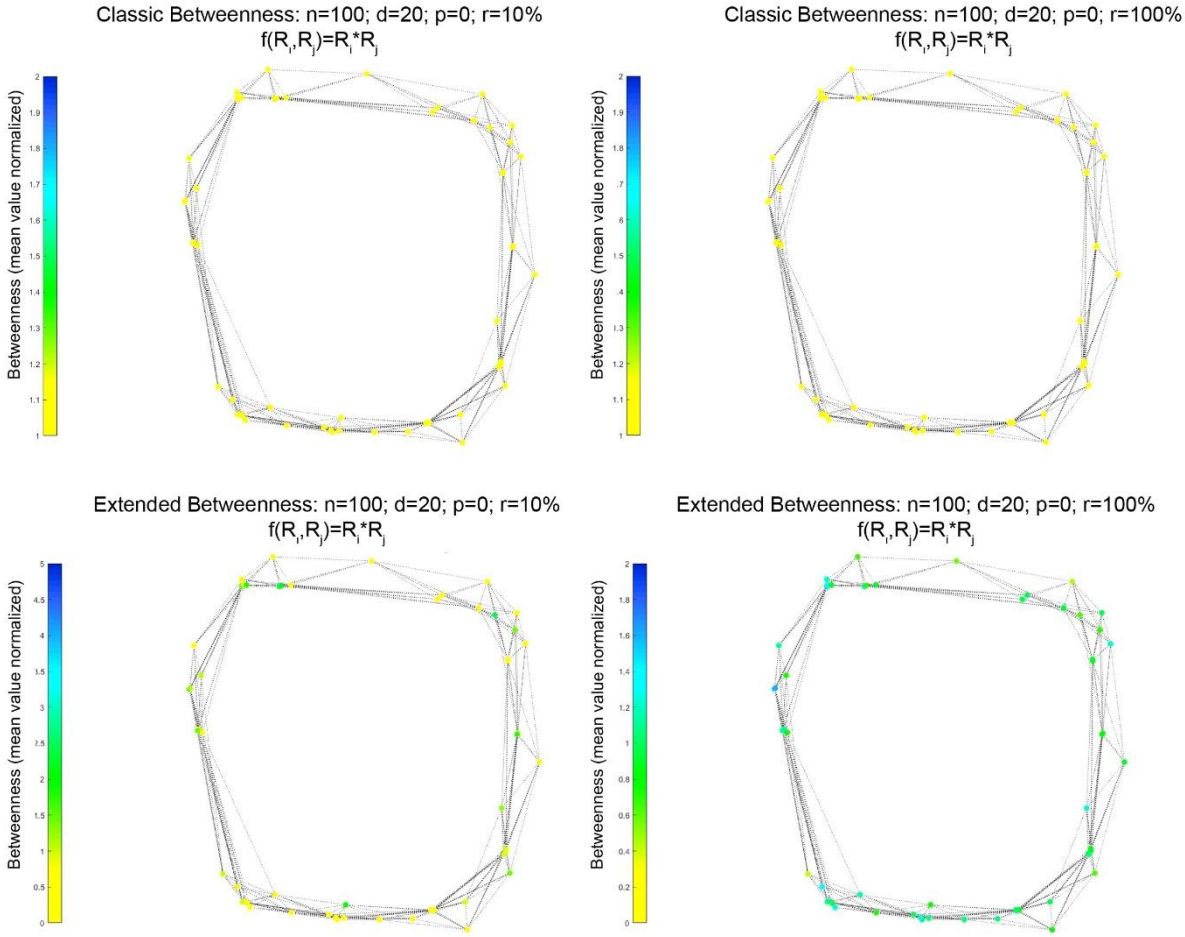


Figure 50. $f(R_i, R_j) = R_i \cdot R_j$ performed for a regular network with 100 vertices. It reports standard *vs* relevance-based betweenness with r=10% (1st panel-left) and r=100% (1st panel-right), intrinsic relevance *vs* relevance-based betweenness with r =10% (2nd panel-left) and r =100% (2nd panel-right), classic *vs* relevance-based harmonic centrality with r =10% (3rd panel-left) and r =100% (3rd panel-right), and intrinsic relevance *vs* relevance-based harmonic with r =10% (4th panel-left) and r =100% (4th panel-right).



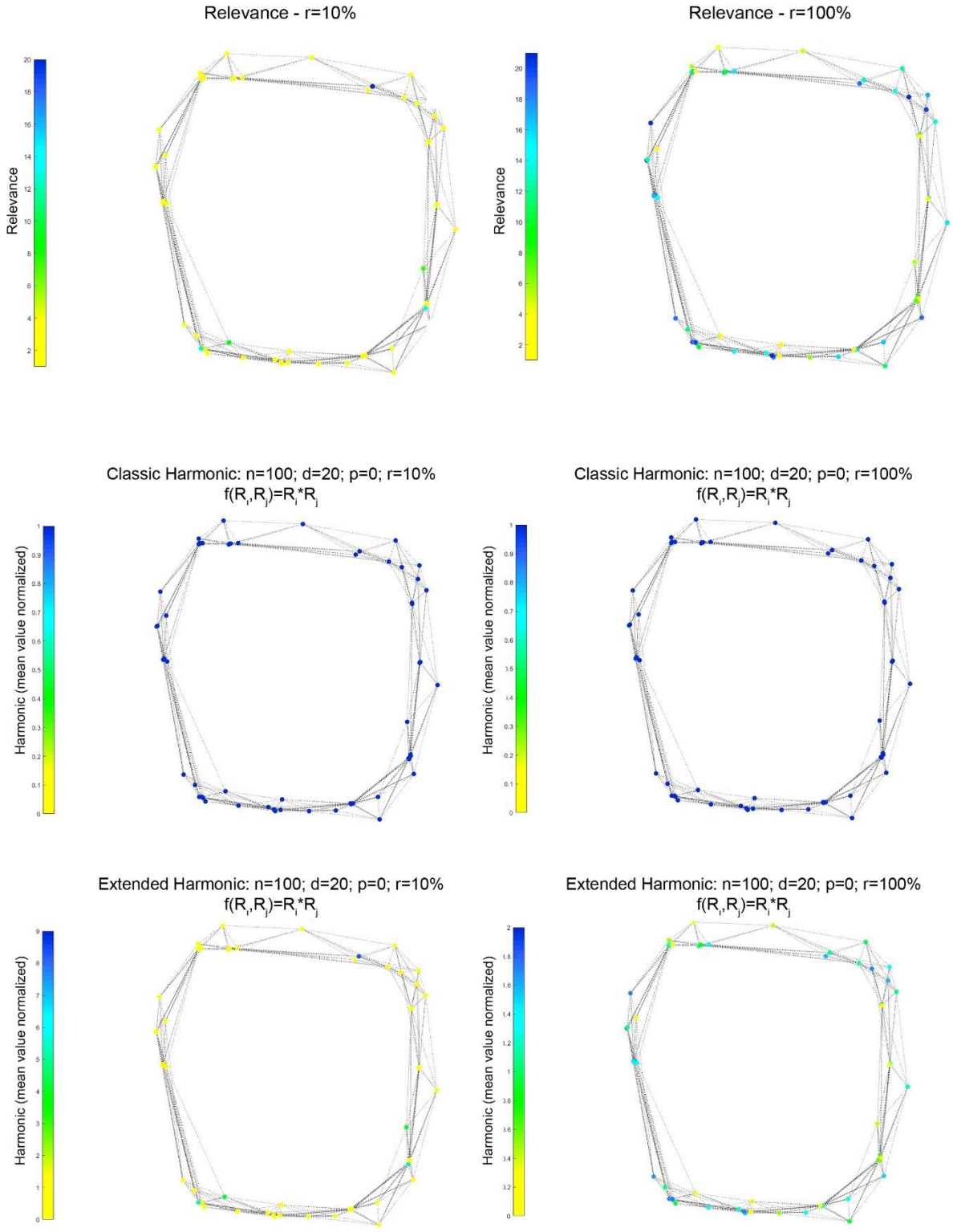


Figure 51. Regular network with 100 vertices corresponding to r=10% (left-panel) and r=100% (right-panel) of the randomly assigned intrinsic relevance. The coloured vertices refer to the colour-bar indicating the values of the: standard betweenness (1st panel), relevance-based betweenness (2nd panel), intrinsic relevance of vertices (3rd panel), standard harmonic centrality (4th panel) and relevance-based harmonic centrality (5th panel). The relevance-based metrics refers to the function $f(R_i, R_j) = R_i \cdot R_j$ as in Figure 50.

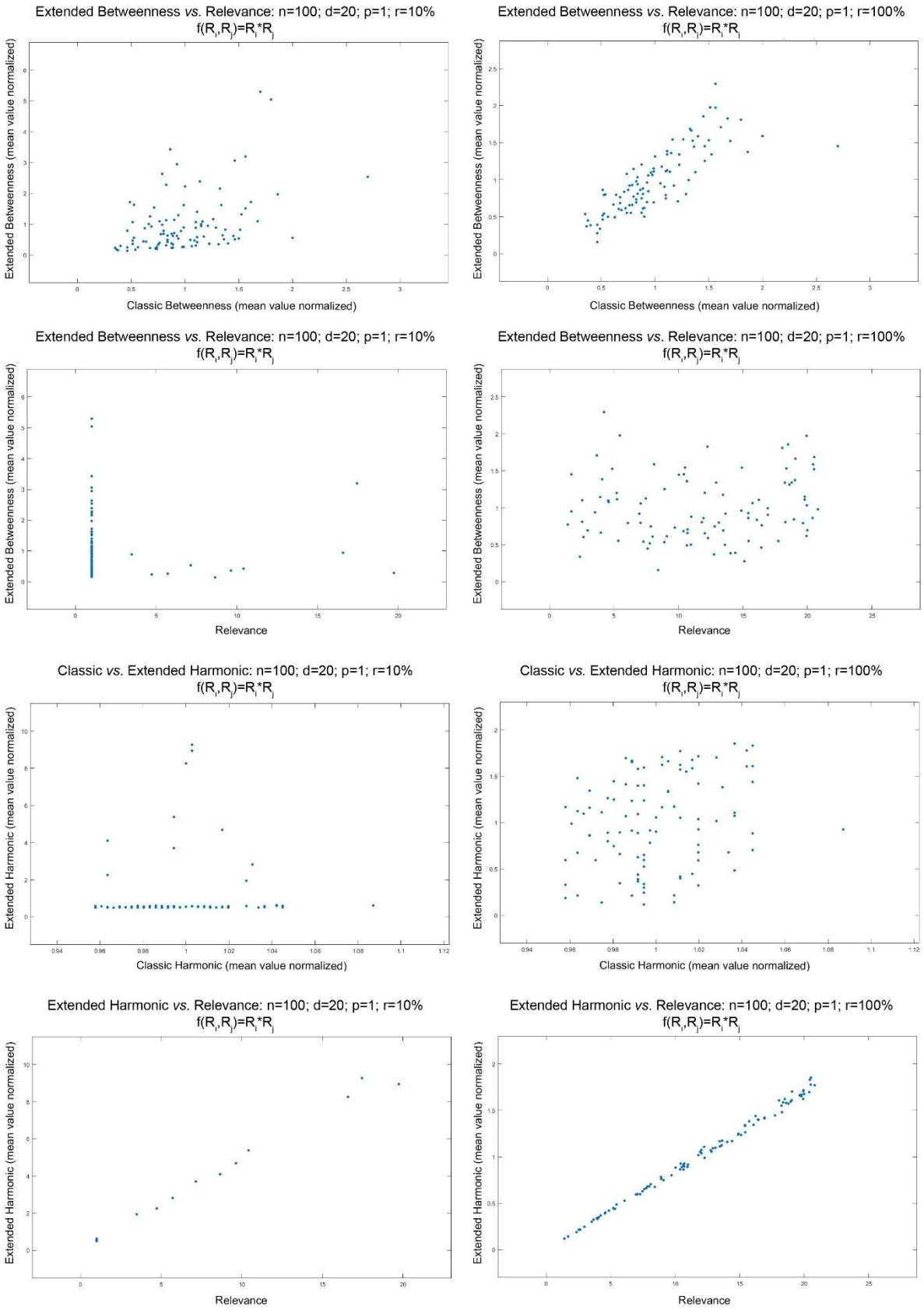
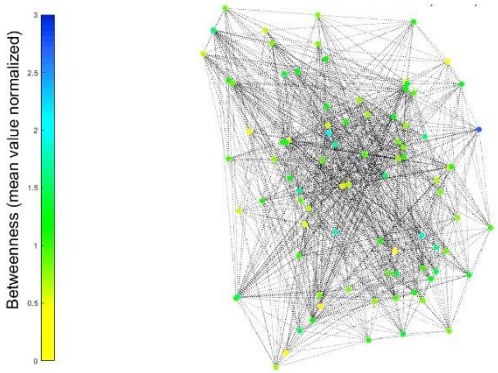
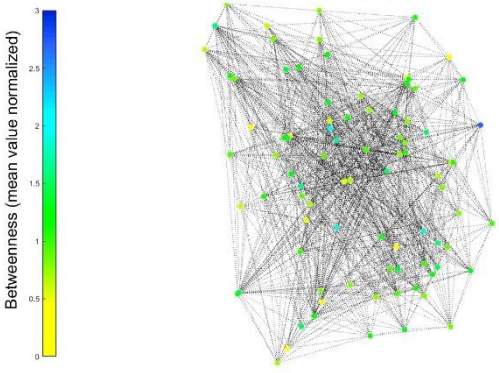


Figure 52. $f(R_i, R_j) = R_i \cdot R_j$ performed for a random network with 100 vertices. It reports standard *vs* relevance-based betweenness with r=10% (1st panel-left) and r=100% (1st panel-right), intrinsic relevance *vs* relevance-based betweenness with r =10% (2nd panel-left) and r =100% (2nd panel-right), classic *vs* relevance-based harmonic centrality with r =10% (3rd panel-left) and r =100% (3rd panel-right), and intrinsic relevance *vs* relevance-based harmonic with r =10% (4th panel-left) and r =100% (4th panel-right).

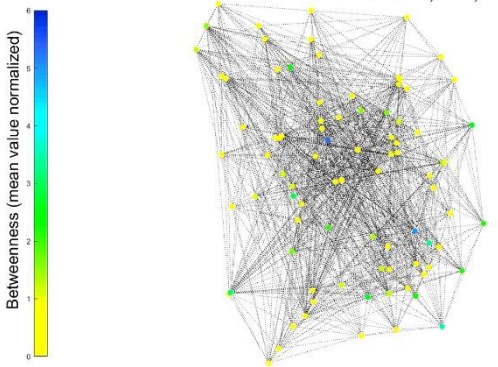
Classic Betweenness: n=100; d=20; p=1; r=10%
$f(R_i,R_j)=R_i*R_j$



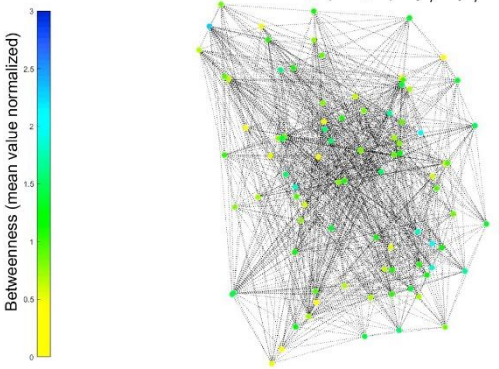
Classic Betweenness: n=100; d=20; p=1; r=100%
$f(R_i,R_j)=R_i*R_j$



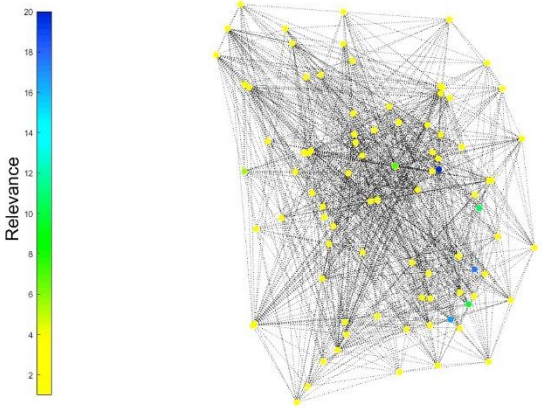
Extended Betweenness: n=100; d=20; p=1; r=10%
$f(R_i,R_j)=R_i*R_j$



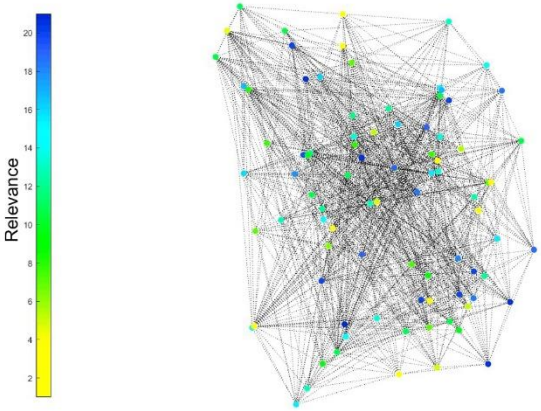
Extended Betweenness: n=100; d=20; p=1; r=100%
$f(R_i,R_j)=R_i*R_j$



Relevance - r=10%



Relevance - r=100%



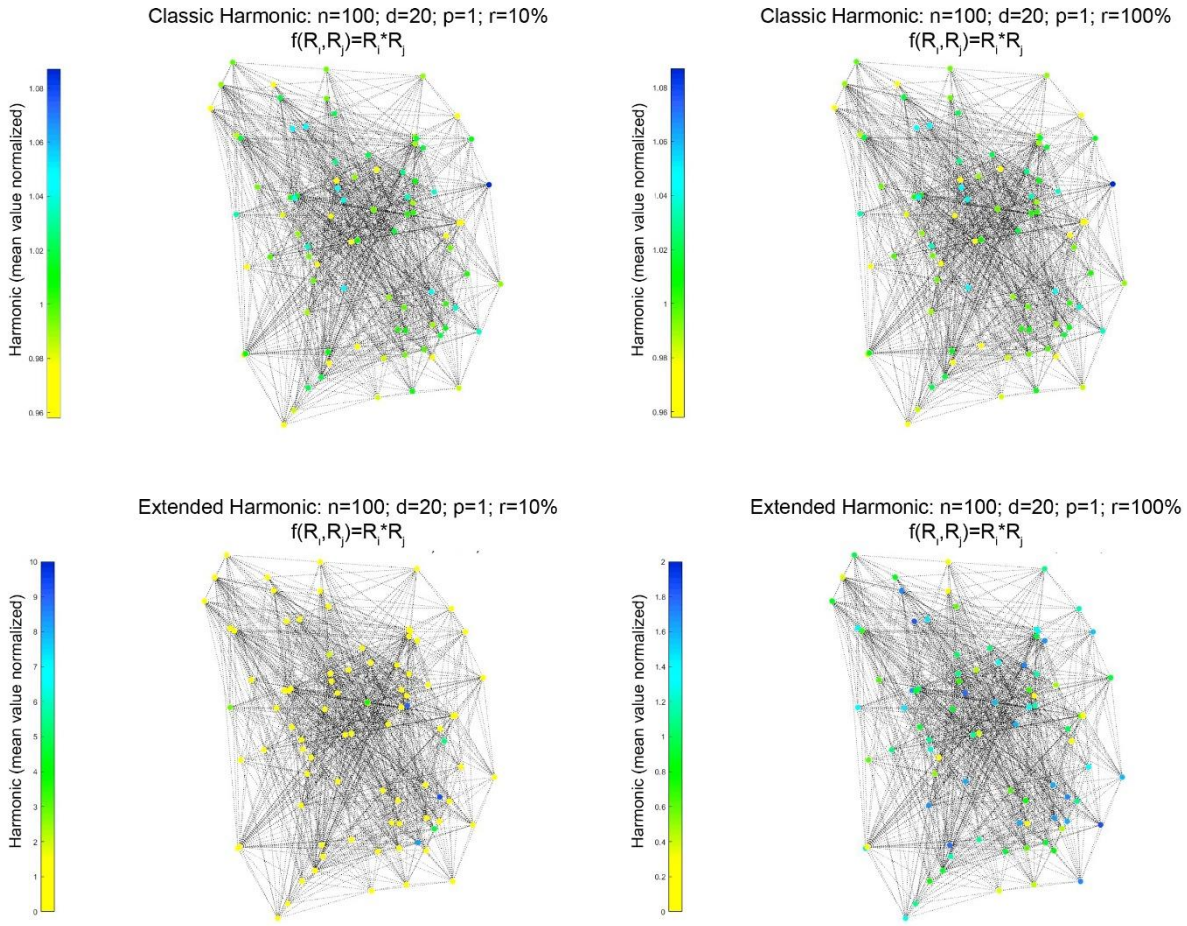


Figure 53. Random network with 100 vertices corresponding to r=10% (left-panel) and r=100% (right-panel) of the randomly assigned intrinsic relevance. The coloured vertices refer to the colour-bar indicating the values of the: standard betweenness (1st panel), relevance-based betweenness (2nd panel), intrinsic relevance of vertices (3rd panel), standard harmonic centrality (4th panel) and relevance-based harmonic centrality (5th panel). The relevance-based metrics refers to the function $f(R_i, R_j) = R_i \cdot R_j$ as in Figure 52.

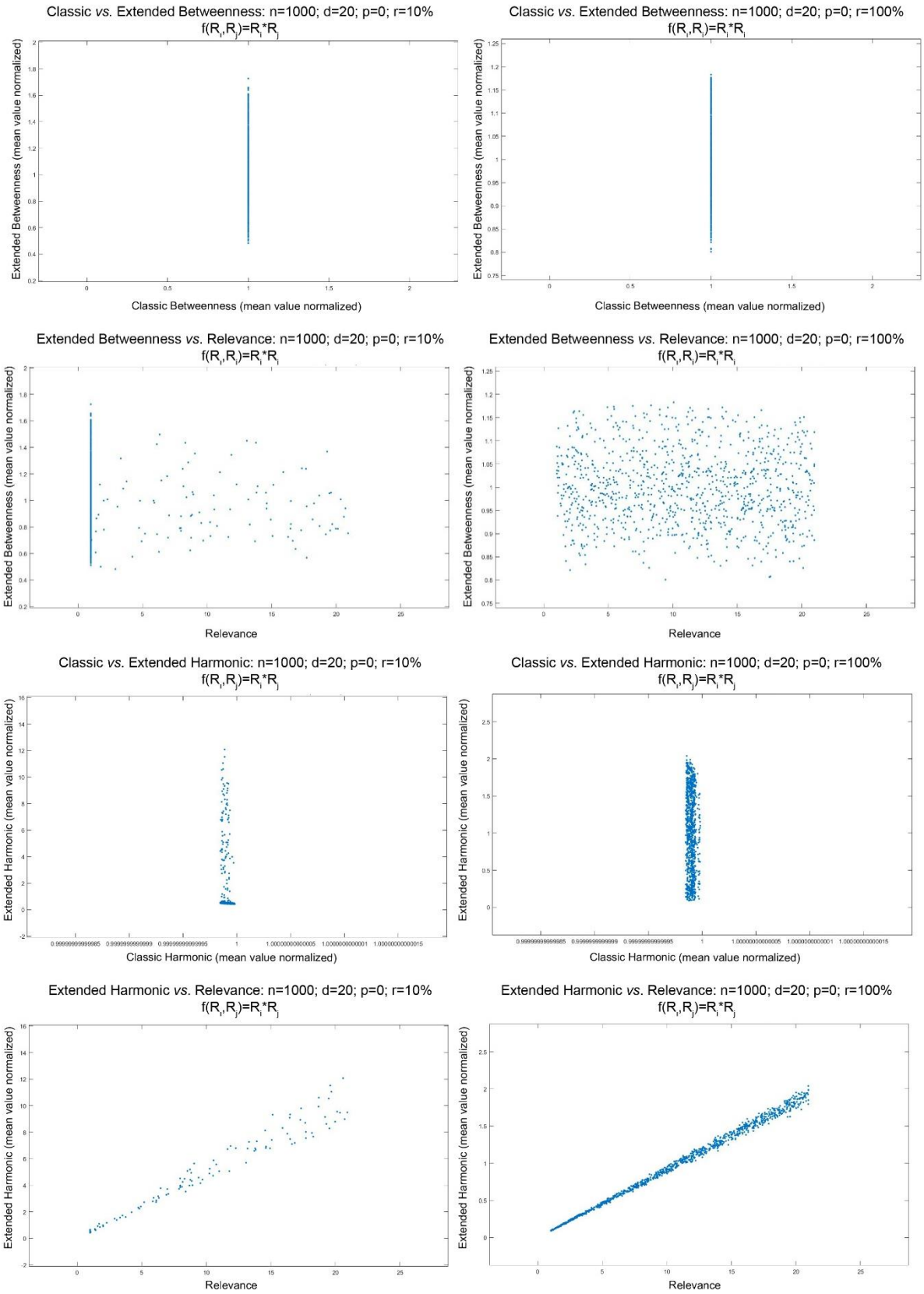
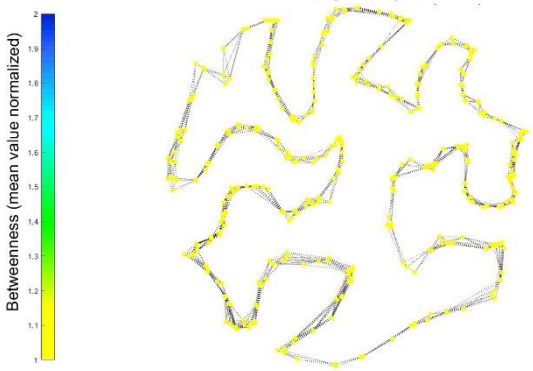
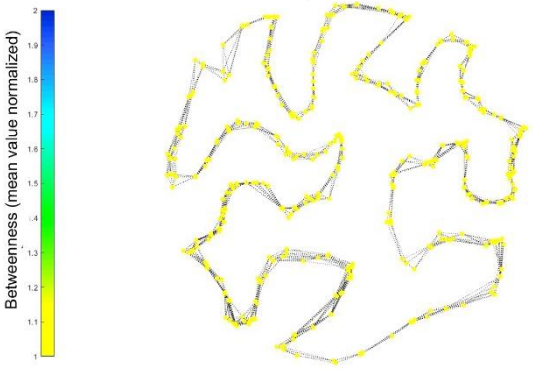


Figure 54. $f(R_i, R_j) = R_i \cdot R_j$ performed for a regular network with 1,000 vertices. It reports standard *vs* relevance-based betweenness with r=10% (1st panel-left) and r=100% (1st panel-right), intrinsic relevance *vs* relevance-based betweenness with r =10% (2nd panel-left) and r =100% (2nd panel-right), classic *vs* relevance-based harmonic centrality with r =10% (3rd panel-left) and r =100% (3rd panel-right), and intrinsic relevance *vs* relevance-based harmonic with r =10% (4th panel-left) and r =100% (4th panel-right).

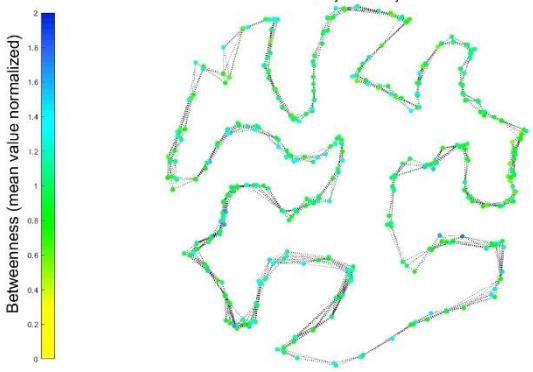
Classic Betweenness: n=1000; d=20; p=0; r=10%
$f(R_i,R_j)=R_i*R_j$



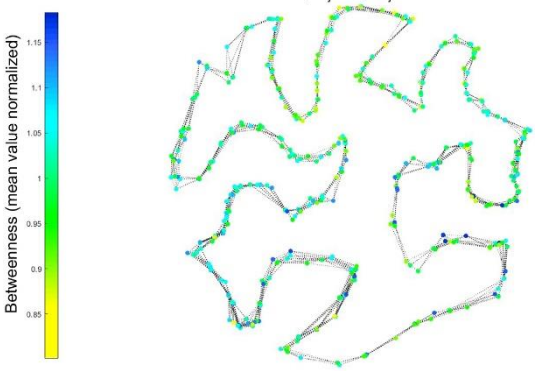
Classic Betweenness: n=1000; d=20; p=0; r=100%
$f(R_i,R_j)=R_i*R_j$



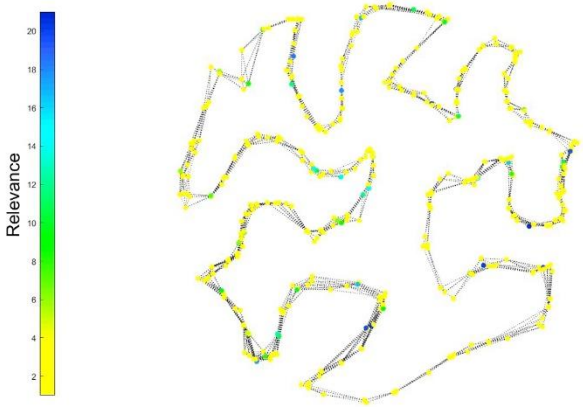
Extended Betweenness: n=1000; d=20; p=0; r=10%
$f(R_i,R_j)=R_i*R_j$



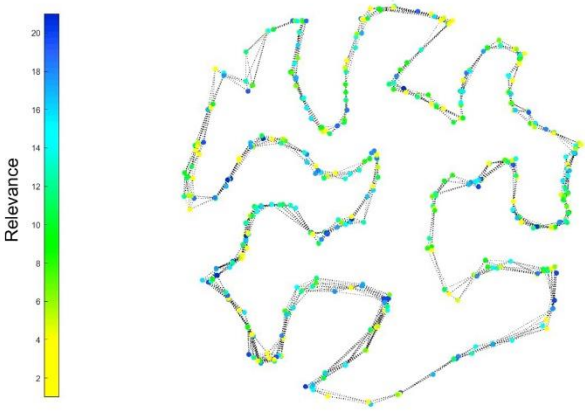
Extended Betweenness: n=1000; d=20; p=0; r=100%
$f(R_i,R_j)=R_i*R_j$



Relevance - r=10%



Relevance - r=100%



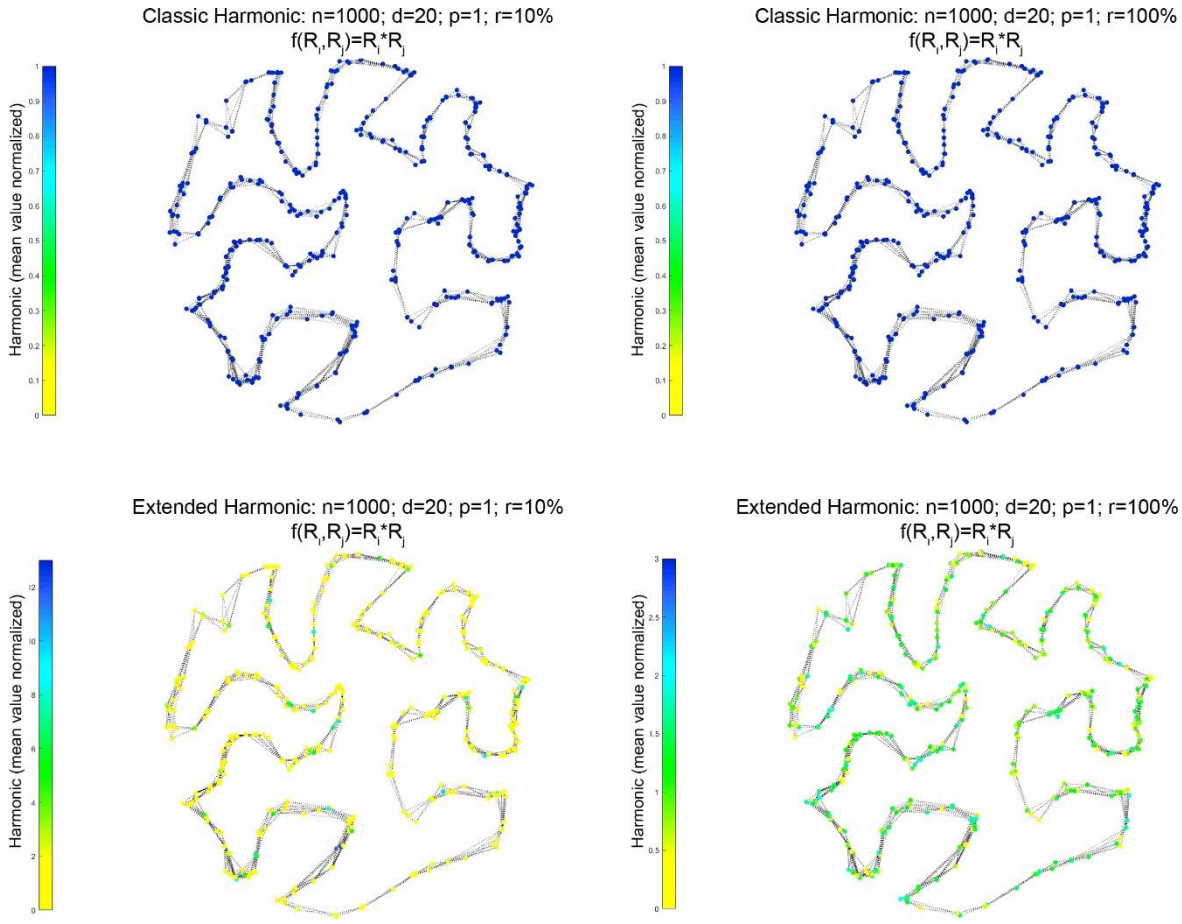


Figure 55. Regular network with 1,000 vertices corresponding to r=10% (left-panel) and r=100% (right-panel) of the randomly assigned intrinsic relevance. The coloured vertices refer to the colour-bar indicating the values of the: standard betweenness (1st panel), relevance-based betweenness (2nd panel), intrinsic relevance of vertices (3rd panel), standard harmonic centrality (4th panel) and relevance-based harmonic centrality (5th panel). The relevance-based metrics refers to the function $f(R_i, R_j) = R_i \cdot R_j$ as in Figure 54.

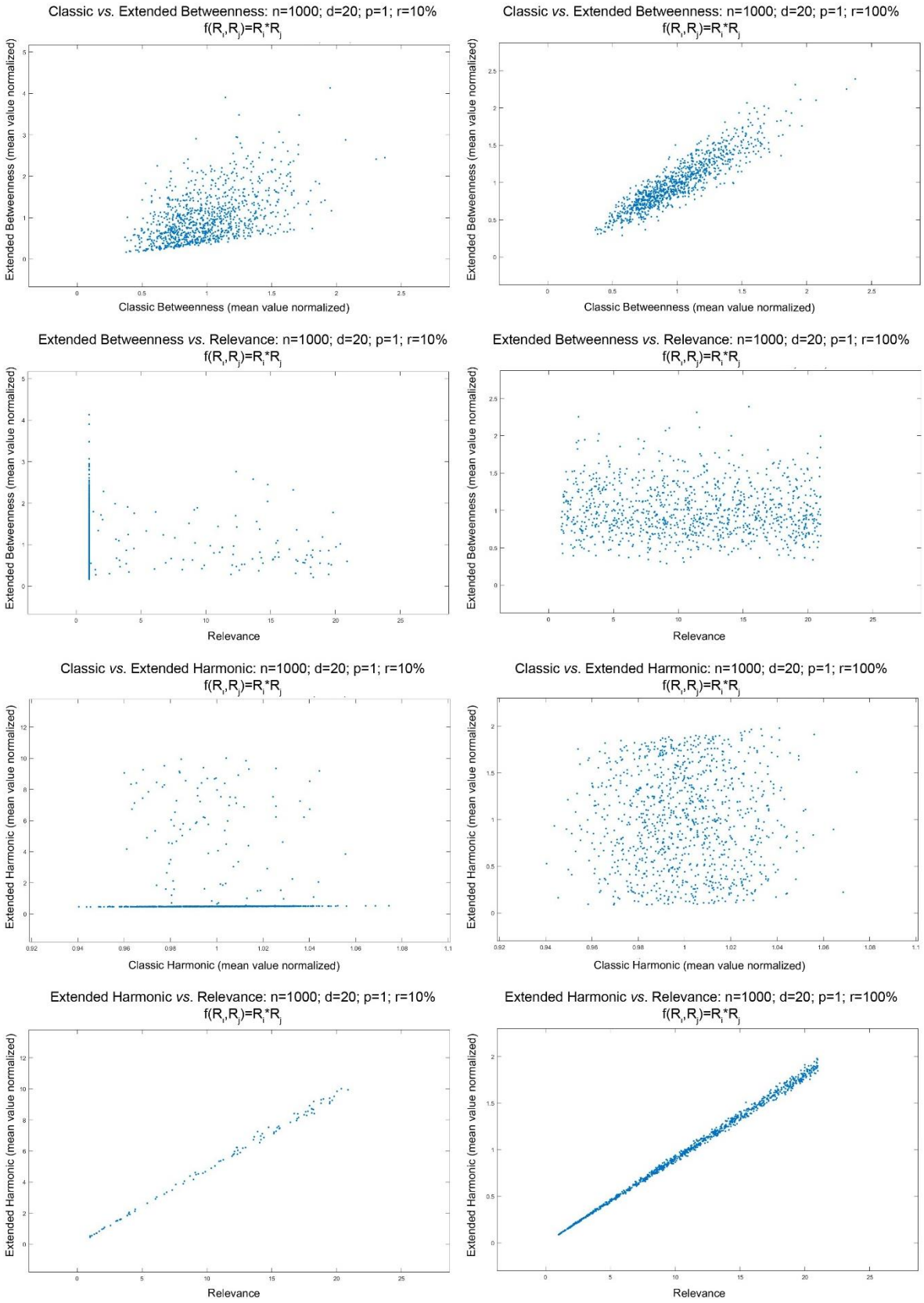
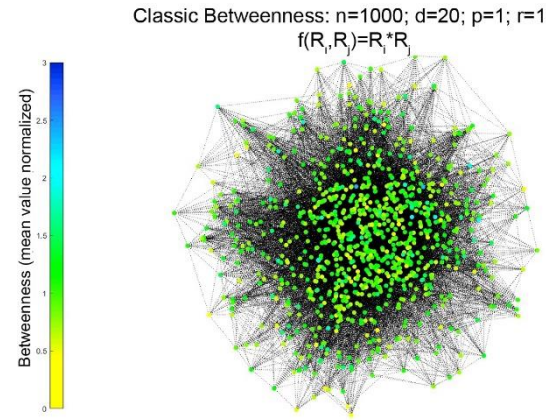
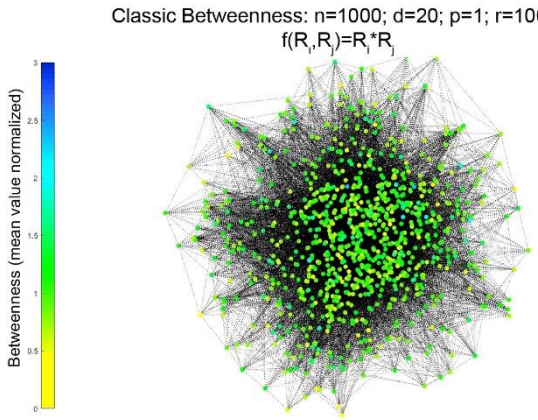
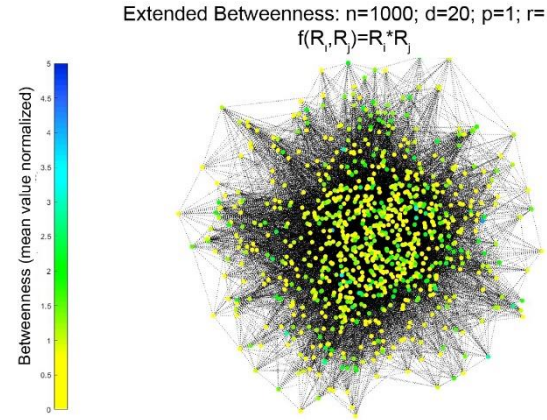
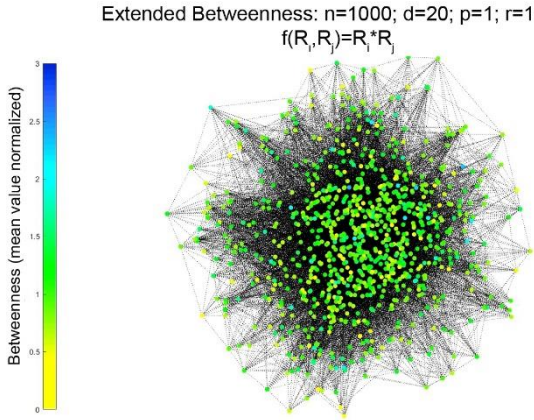


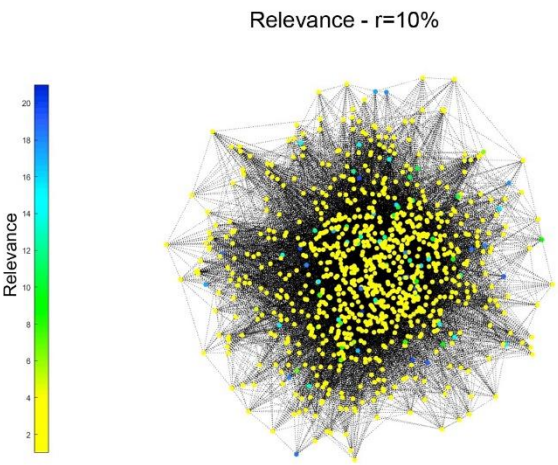
Figure 56. $f(R_i, R_j) = R_i \cdot R_j$ performed for a random network with 1,000 vertices. It reports standard *vs* relevance-based betweenness with r=10% (1st panel-left) and r=100% (1st panel-right), intrinsic relevance *vs* relevance-based betweenness with r =10% (2nd panel-left) and r =100% (2nd panel-right), classic *vs* relevance-based harmonic centrality with r =10% (3rd panel-left) and r =100% (3rd panel-right), and intrinsic relevance *vs* relevance-based harmonic with r =10% (4th panel-left) and r =100% (4th panel-right).

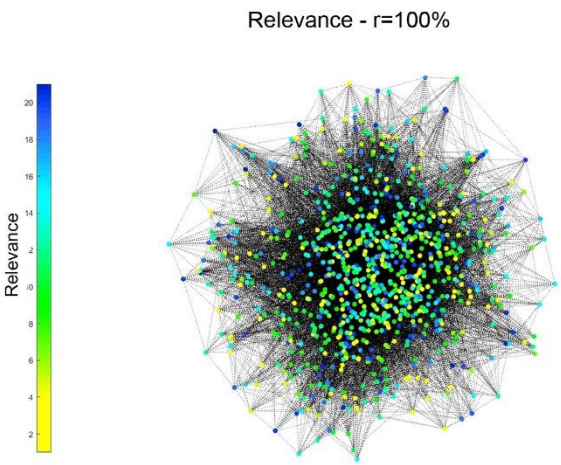

Classic Betweenness: n=1000; d=20; p=1; r=10%
f(R_i,R_j)=R_i*R_j
Betweenness (mean value normalized)


Classic Betweenness: n=1000; d=20; p=1; r=100%
f(R_i,R_j)=R_i*R_j
Betweenness (mean value normalized)


Extended Betweenness: n=1000; d=20; p=1; r=10%
f(R_i,R_j)=R_i*R_j
Betweenness (mean value normalized)


Extended Betweenness: n=1000; d=20; p=1; r=100%
f(R_i,R_j)=R_i*R_j
Betweenness (mean value normalized)


Relevance - r=10%
Relevance


Relevance - r=100%
Relevance

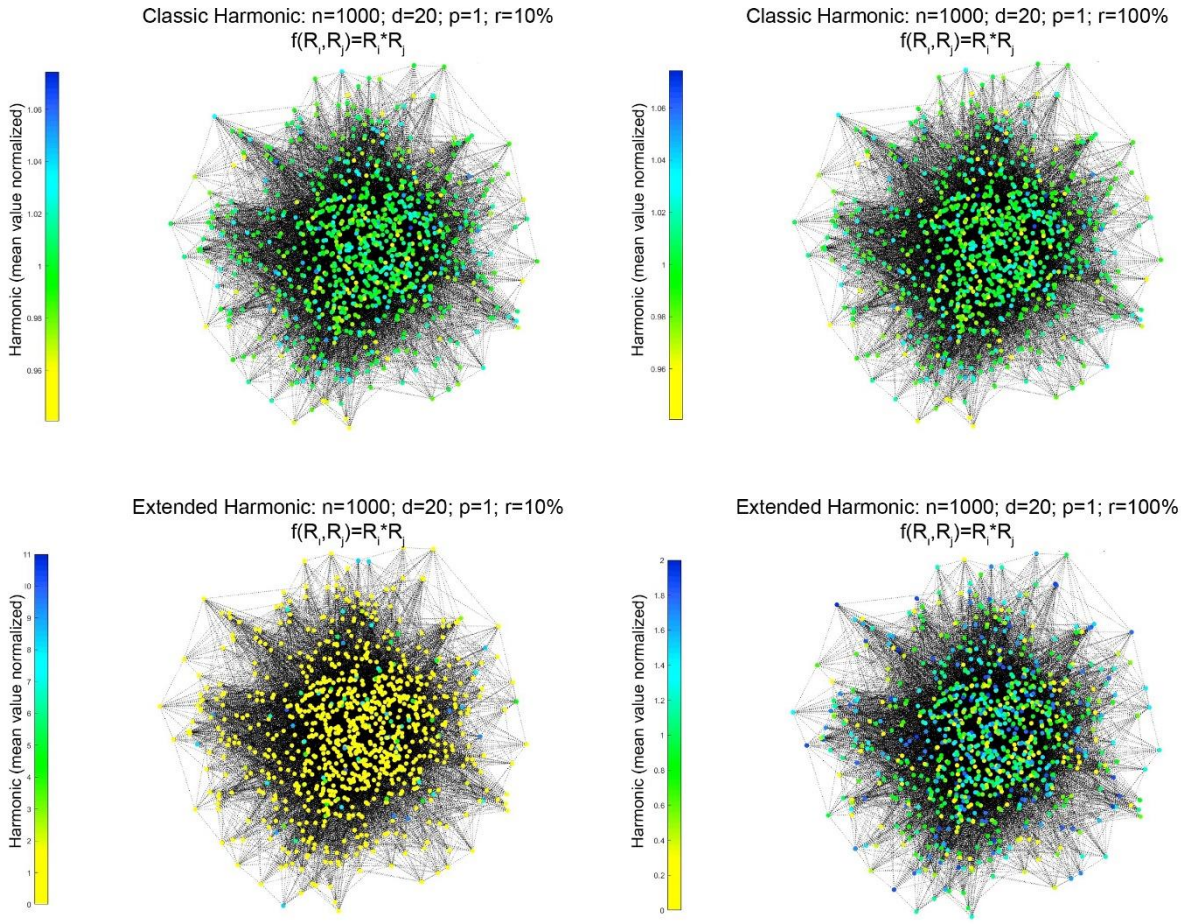


Figure 57. Random network with 1,000 vertices corresponding to *r*=10% (left-panel) and *r*=100% (right-panel) of the randomly assigned intrinsic relevance. The coloured vertices refer to the colour-bar indicating the values of the: standard betweenness (1st panel), relevance-based betweenness (2nd panel), intrinsic relevance of vertices (3rd panel), standard harmonic centrality (4th panel) and relevance-based harmonic centrality (5th panel). The relevance-based metrics refers to the function $f(R_i, R_j) = R_i \cdot R_j$ as in Figure 56.

About the betweenness centrality, the tests shown that:

- for regular networks, the correlation between standard and novel metric is low, in fact the intrinsic relevance is always an important external information to characterize the specific problem;
- for random networks, the correlation between classic and extended metric is low for r=10% and increases for r=100%, because the intrinsic relevance is an important external information to characterize the specific problem when it concerns the 10% of vertices, and it is less so increasing the number of nodes, until it vanishes for r = 100%;
- increasing the size of the random network the correlation between classic and extended metric increases, because the importance of the intrinsic relevance vanishes when randomly assigned to all the vertices (r=100%);

- functions $f(R_s,R_t)= (R_s \cdot R_t)$ and $f(R_s,R_t) = \Pi R_{path}$ emphasize the role of the intrinsic relevance;
- the intrinsic relevance is not correlated to the standard betweenness when the function $f(R_s,R_t)$ depends on the ending vertices of the paths since of the specific meaning of the metric;
- the intrinsic relevance of vertices becomes to be correlated to the standard betweenness when the function $f(R_s,R_t)$ depends on all the vertices of the paths because the most relevant vertices contribute to the metric values when traversed.

About the harmonic centrality, the tests show that:

- both for regular and random networks, the correlation between standard and novel metric is always low, in fact the intrinsic relevance is always an important external information to characterize the specific problem;
- the intrinsic relevance of vertices is quite always correlated to the harmonic centrality because the value of the metric for a vertex is at least scaled by its intrinsic relevance $f(R_s, R_t) = R_s$;
- the intrinsic relevance of vertices is not correlated to the harmonic centrality when the function $f(R_s, R_t)$ depends on all the vertices of the paths, because the intrinsic relevance of all the vertices in the shortest paths determines the metric value.

4.2.3 The social network case: the Florence Families network

Here we report the application of the extended metrics on a social network, precisely the Florence Families network (*Padgett and Ansell*, 1993; *Alvarez-Socorro et al*., 2015; *Latora et al*., 2017), composed of seventeen vertices, corresponding to the most notables Renaissance Florentine Families, and twenty-three edges, corresponding to the marriages between them. Computing the standard betweenness with the assumption of identical intrinsic relevance of vertices, i.e., R_v=1 for all nodes, we obtained the vertex and edge ranking shown in Figure 58, whose values are reported in Tables 3 and 4. The size of the family emblem indicates the node centrality, while the marriage link colour indicates the edge betweenness. Guasconi is the most central family, and the marriage between Peruzzi and Castellani is the most important link. However, it turned out strange that the Medici family, which has always been considered the most central/ influential of the Renaissance, was not the most central here too. We associated

this inconsistency with a lack of information useful for defining the metric itself, i.e., the intrinsic relevance of the vertices.

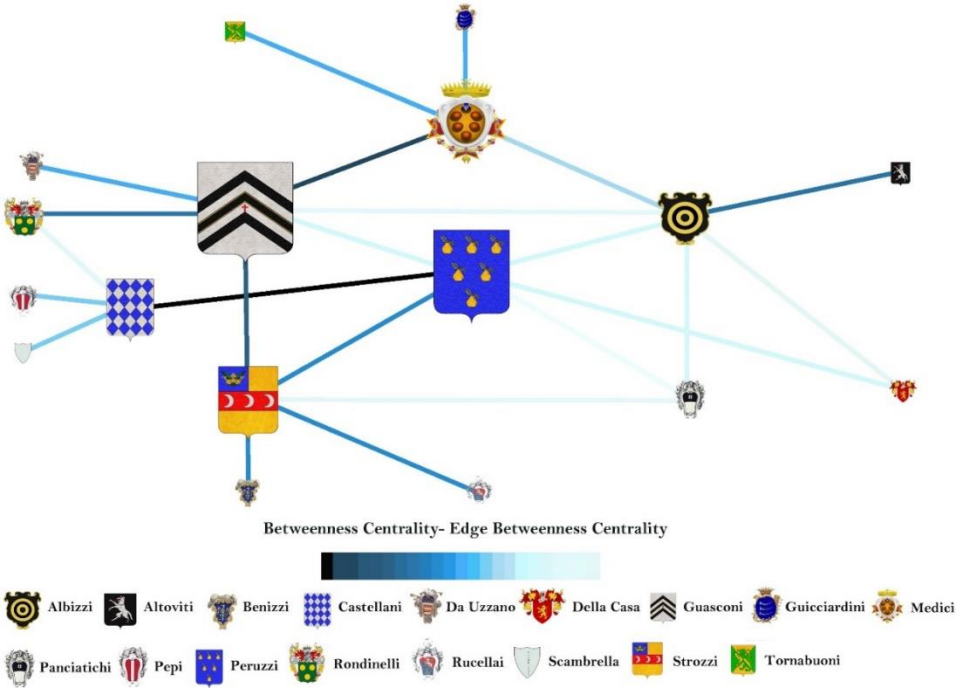


Figure 58. Standard betweenness for the Florence Families network. Vertices are characterized by identical intrinsic relevance, equal to 1. Each family is represented by its own emblem, whose size is proportional to the corresponding vertex betweenness centrality. The edge betweenness is shown by coloured edges, ranging from black (maximum value) to cyan (minimum value). The colour bar of the figure is obtained subtracting the minimum value and normalizing to 100. Source: *Giustolisi et al*., 2020.

Table 3. Standard betweenness centrality for the Florence families' network.

Family	Standard Betweenness Centrality
Guicciardini	0,00
Tornabuoni	0,00
Medici	24,00
Da Uzzano	0,00
Rondinelli	3,00
Guasconi	**34,92**
Pepi	0,00
Scambrilla	0,00
Benizzi	0,00
Castellani	23,00
Strozzi	26,33
Rucellai	0,00
Peruzzi	29,75
Albizzi	19,00
Altoviti	0,00
Della Casa	0,00
Panciatichi	2,00

Table 4. Standard edge betweenness centrality for the Florence families' network

Marriage Edge		Standard edge betweenness
Guicciardini	Medici	8,50
Tornabuoni	Medici	8,50
Medici	Guasconi	17,50
Medici	Albizzi	7,00
Da Uzzano	Guasconi	8,33
Rondinelli	Guasconi	9,08
Rondinelli	Castellani	1,00
Guasconi	Strozzi	13,00
Guasconi	Peruzzi	4,25
Guasconi	Albizzi	1,00
Pepi	Castellani	8,00
Scambrilla	castellani	8,00
Benizzi	Strozzi	9,67
Castellani	**Peruzzi**	**22,00**
Strozzi	Rucellai	9,67
Strozzi	Peruzzi	10,00
Strozzi	Panciatichi	0,00
Peruzzi	Albizzi	3,00
Peruzzi	Della Casa	3,83
Peruzzi	Panciatichi	0,42
Albizzi	Altoviti	9,00
Albizzi	Della Casa	0,00
Albizzi	Panciatichi	2,00

When we assumed that each family had an intrinsic relevance equal to the gross wealth (i.e., the sum of marital, trade, partnership, bank, and real estate relations) quantified in florins (*Padgett and Ansell*, 1993), the coupling and interaction between topology and intrinsic relevance made Medici the most central family. In fact, the function $f(R_s,R_t) = R_s \cdot R_t$ identified Medici as the most central family and the marriage between Medici and Guasconi as the most central link. The gross wealth of each family is reported close to the family name in Figure 55. These results highlight the fact that Medici represent the most central family in the socio-economic structure of the Florence families, i.e., the most traversed by the socio-economic activities and not the one that holds the greatest gross wealth (i.e., intrinsic relevance). Therefore, Medici results the most central in the network because it is always traversed by the paths between the four of five richest Florentine families (Medici excluded), not being these four families linked by marriages. This consideration supports even more the fact that the

importance of an element in the network is the result of the combination of topology and intrinsic relevance.

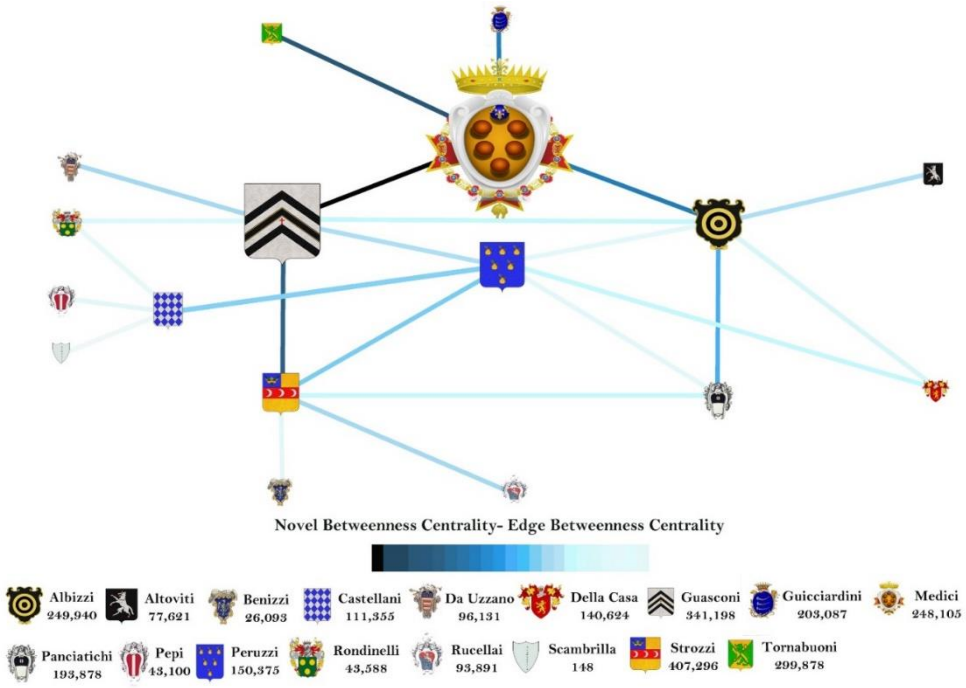


Figure 59. Proposed betweenness centrality, embedding the family intrinsic relevance of the Florence Family network. The intrinsic relevance corresponds to the family gross wealth (in florins) and its value is reported close to the name of each family. Emblem size and edge color (from black for maximum value to cyan for minimum value) highlight the vertex and edge centrality, respectively. Source: *Giustolisi et al*., 2020

Table 5. Relevance-embedding betweenness centrality for Florence family's network, assuming the vertex intrinsic relevance based on gross wealth in florins. The colour bar of the corresponding figure is obtained subtracting the minimum value and normalizing to 100.

Family	Intrinsic relevance in florins	Relevance-embedding betweenness centrality
Guicciardini	203,087	0,00E+00
Tornabuoni	299,878	0,00E+00
Medici	248,105	**9,65E+11**
Da Uzzano	96,131	0,00E+00
Rondinelli	43,588	3,87E+10
Guasconi	341,198	7,62E+11
Pepi	43,100	0,00E+00
Scambrilla	148	0,00E+00
Benizzi	26,093	0,00E+00
Castellani	111,355	8,01E+10
Strozzi	407,296	2,40E+11
Rucellai	93,891	0,00E+00
Peruzzi	150,375	3,69E+11
Albizzi	249,940	4,10E+11
Altoviti	77,621	0,00E+00
Della Casa	140,624	0,00E+00
Panciatichi	193,878	5,76E+10

Table 6. Relevance-embedding edge betweenness for Florence family's network, assuming the vertex intrinsic relevance based on gross wealth in florins. The colour bar of the corresponding figure is obtained subtracting the minimum value and normalizing to 100.

Marriage Edge		Relevance-embedding edge betweenness
Guicciardini	Medici	4,76E+11
Tornabuoni	Medici	6,74E+11
Medici	**Guasconi**	**8,52E+11**
Medici	Albizzi	4,97E+11
Da Uzzano	Guasconi	2,26E+11
Rondinelli	Guasconi	1,35E+11
Rondinelli	Castellani	4,51E+10
Guasconi	Strozzi	6,49E+11
Guasconi	Peruzzi	2,22E+11
Guasconi	Albizzi	1,57E+11
Pepi	Castellani	8,43E+10
Scambrilla	castellani	0,00E+00
Benizzi	Strozzi	6,45E+10
Castellani	Peruzzi	2,41E+11
Strozzi	Rucellai	2,26E+11
Strozzi	Peruzzi	2,35E+11
Strozzi	Panciatichi	1,60E+11
Peruzzi	Albizzi	9,96E+10
Peruzzi	Della Casa	1,63E+11
Peruzzi	Panciatichi	1,03E+11
Albizzi	Altoviti	1,78E+11
Albizzi	Della Casa	1,51E+11
Albizzi	Panciatichi	2,66E+11

We computed the standard degree and the extended one. In the first case we obtained the vertex ranking shown in the Figure 60 (left), whose values are reported in Table 7. The degree values range from 1 to 6 and Guasconi, Albizzi and Peruzzi represent the most central families, i.e., the most important hubs, being connected to six other families by marriages. For the second one, embedding the family gross wealth (in florins) as intrinsic relevance during the analysis, the degree values range between 0 to 4,08E+12 (Figure 60- right). Guasconi family remains the most important socio-economic hub, being a rich family (intrinsic relevance equal to 341,198 florins) connected with six families by marriages. The extended degree informs that Strozzi, the richest family (with an intrinsic relevance equal to 407,296 florins) and connected with five families with marriages, becomes the second socio-economic hub. Overall, the example shows that socio-economic influence of a family on its neighbours is relevant if the family is rich and well connected (by means of marriages) with other rich families.

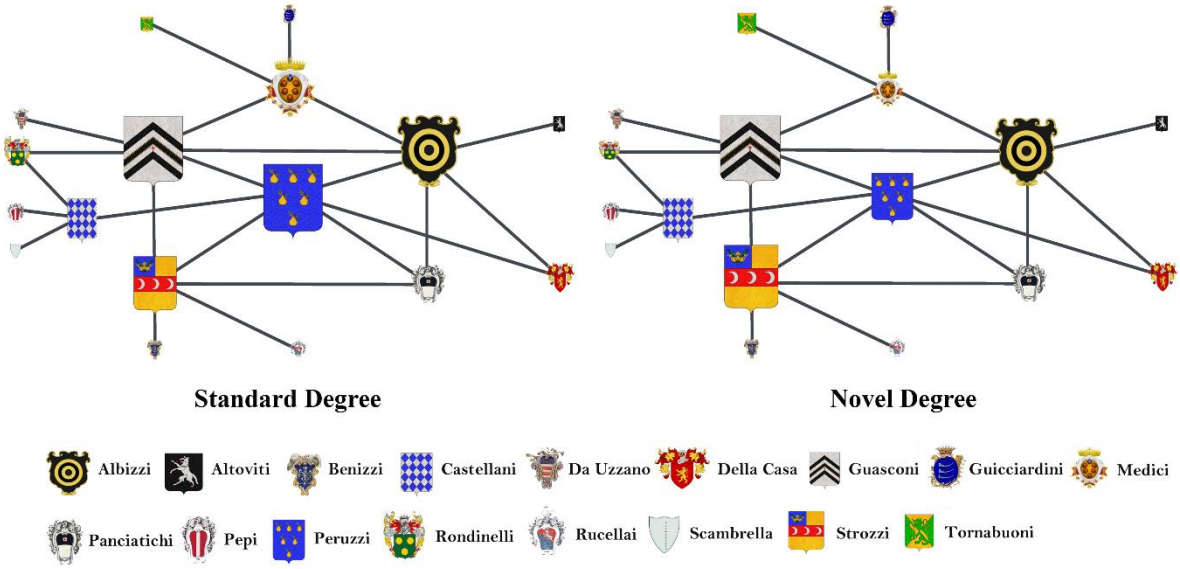


Figure 60. Standard degree centrality (**left**) and proposed degree centrality embedding the family intrinsic relevance of the Florence Family network (**right**). The novel metric shows that the most central family within the network is not only the most connected but also the one with high relevance, i.e., the metric is a combination of topology and relevance of the elements.

Table 7. Standard degree and Degree embedding vertex intrinsic relevance for Florence family's network.

Family	Standard Degree	Degree embedding vertex intrinsic relevance
Guicciardini	1	5,00E+10
Tornabuoni	1	7,40E+10
Medici	4	2,71E+11
Da Uzzano	1	3,20E+10
Rondinelli	2	1,90E+10
Guasconi	**6**	**4,08E+12**
Pepi	1	4,00E+09
Scambrilla	1	0.00E+00
Benizzi	1	1,00E+10
Castellani	4	2,60E+10
Strozzi	5	3,28E+12
Rucellai	1	3,80E+10
Peruzzi	**6**	2,17E+12
Albizzi	**6**	2,88E+12
Altoviti	1	1,90E+10
Della Casa	2	5,60E+10
Panciatichi	3	1,57E+12

We also computed the standard harmonic centrality and the extended one. In the first case we obtained the vertex ranking shown in Figure 61 (left), whose values are reported in Table 8. The harmonic centrality values range from 5,817 to 10,667. Guasconi and Peruzzi represent the most efficient families to spread socio-economic activities in the network, followed by Albizzi. For the second one, embedding the family gross wealth (in florins) as intrinsic relevance during the analysis, the harmonic values range between 1,35E+08 to 6,08E+11 (Figure 61- right) and Guasconi remains the most central family to spread socio-economic activities through the network because of its economic relevance and topological position into the network of marriages. The novel harmonic informs that Strozzi, the richest family, becomes the second one in ranking because of its economic relevance and position in the network of marriages.

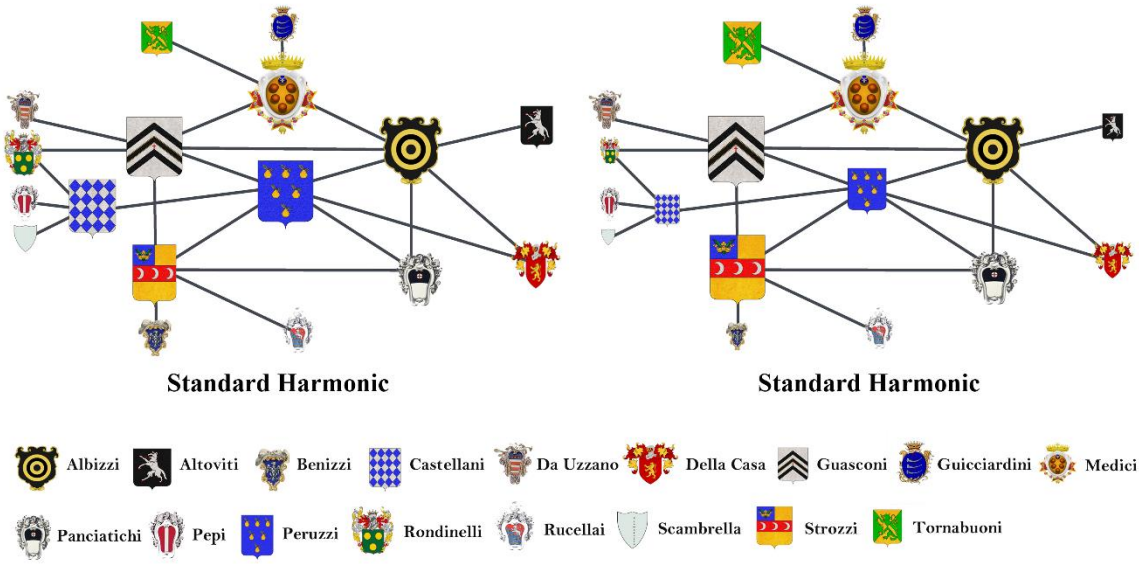


Figure 61. Standard harmonic centrality (**left**) and proposed harmonic centrality embedding the family intrinsic relevance (**right**) of the Florence Family network. The novel metric shows that the most central family within the network is both the richest one and the most connected one.

Table 8: Novel degree and harmonic centrality for Florence family's network, assuming the vertex intrinsic relevance based on gross wealth in florins. The embeds considers the intrinsic relevance of vertices in florins.

Family	Degree	Harmonic
Guicciardini	5,983	2,29E+11
Tornabuoni	5,983	3,24E+11
Medici	9,000	4,31E+11
Da Uzzano	6,667	1,23E+11
Rondinelli	7,833	6,08E+10
Guasconi	**10,667**	**6,08E+11**
Pepi	5,817	3,83E+10
Scambrilla	5,817	1,35E+08

Benizzi	6,250	3,26E+10
Castellani	8,667	1,35E+11
Strozzi	9,667	**5,94E+11**
Rucellai	6,250	1,14E+11
Peruzzi	**10,667**	2,90E+11
Albizzi	10,333	4,47E+11
Altoviti	6,500	9,43E+10
Della Casa	7,667	1,91E+11
Panciatichi	8,500	3,02E+11

The extended degree and harmonic allowed to integrate the economic and social information of the families, obtaining a different result from the standard CNT analyses.

4.2.4 Enhanced WDNs Assessment using the Paradigm of the Intrinsic Relevance of Vertices

Here we report the application of the extended metrics on Apulian network (see Figure 34), composed of twenty-four vertices, that represent twenty-three demand nodes and one source node, and thirty-four edges, corresponding to the pipes for the water distribution. The strategy is performed using the function $f(R_s, R_t) = R_s \cdot R_t$.

Computing the standard betweenness with the assumption of identical intrinsic relevance of vertices, i.e., R_v=1 for all nodes, we obtained the edge ranking shown in the figure 62 (left), whose values are reported in Table 9. The metric does not identify the pipes close to the reservoir as the most important, and this result is somewhat unrealistic from a technical point of view, since it is known that the most important edge for the distribution network corresponds to the pipe connecting the network to the source of water because all the paths from the source of water always traverse that edge. When we embedded the intrinsic relevance in the analysis, equal to the water demand for each demand node and to the sum of supplied water for the source node, the importance of pipes/ paths close to the reservoir increases, i.e., the extended metric correctly identifies the pipe close to the reservoir as the most important edge (See Figure 62-right), consistently with the hydraulic engineering knowledge. In fact, the metric identifies the edge 34 as the most important and the main path is that between vertices 24-8. Results are reported in Table 9.

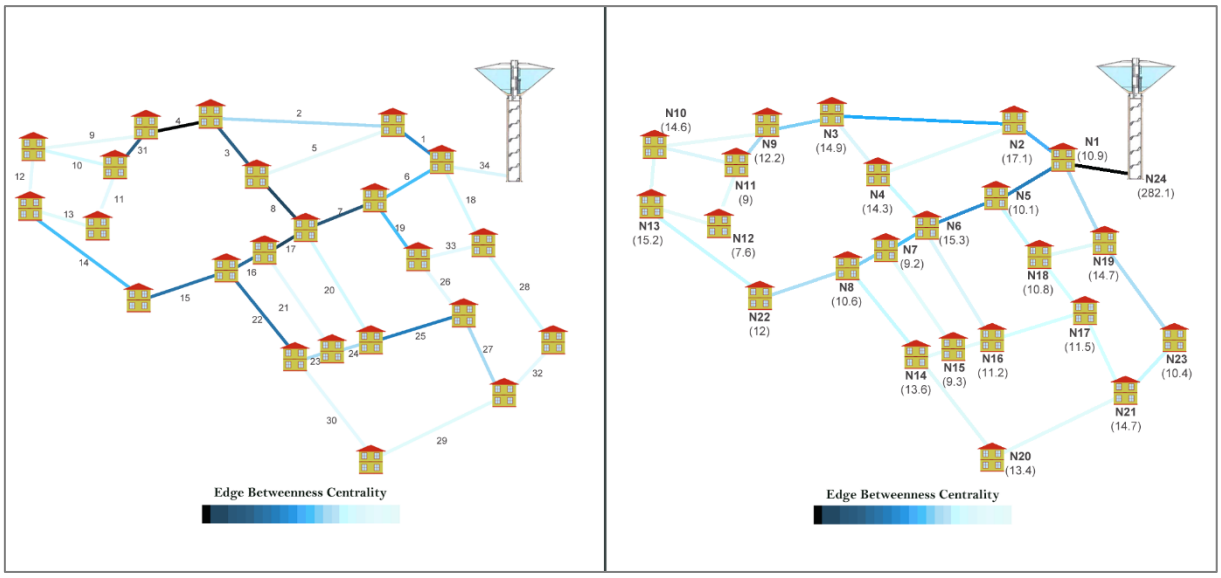


Figure 62. **Left**. Standard edge betweenness for Apulian network. Source and demand nodes have the same intrinsic relevance equal to 1. The most important pipe is the number 4. **Right.** Proposed edge betweenness embedding the intrinsic relevance in terms of demand for each node and as sum of demands for the source node (values in l/s are reported close to each node). The pipe 34 results the most important edge. The edge betweenness is identified by coloured edges ranging from black (maximum value) to cyan (minimum value). Source: *Giustolisi et al*., 2020.

Table 9. Standard edge betweenness (source and demand nodes are characterized by the identical intrinsic relevance) and relevance-embedding edge betweenness (intrinsic relevance of vertices is assumed equal to the demand, for nodes, and the sum of demands, for the source node) for all pipes of the water supply system. The colour bar of the corresponding figure is obtained subtracting the minimum value and normalizing to 100.

Pipe	Standard edge betweenness	Relevance-embedding edge betweenness	Pipe	Standard edge betweenness	Relevance-embedding edge betweenness
1	38	0,0305	18	22	0,0179
2	26	0,0198	19	31	0,0105
3	52	0,0075	20	25	0,0069
4	**65**	0,0209	21	13	0,0040
5	3	0,0044	22	44	0,0105
6	32	0,0396	23	26	0,0037
7	58	0,0344	24	28	0,0040
8	60	0,0089	25	40	0,0060
9	0	0,0000	26	15	0,0053
10	22	0,0080	27	26	0,0041
11	15	0,0035	28	22	0,0137
12	15	0,0026	29	8	0,0050
13	8	0,0007	30	15	0,0024
14	30	0,0087	31	52	0,0156
15	43	0,0140	32	17	0,0102
16	48	0,0213	33	17	0,0027
17	62	0,0279	**34**	23	**0,0796**

In order to understand how the presence of strategic structures was considered in the analysis, we added a hospital in node 10, amplifying its demand, i.e., its intrinsic relevance, by 10 times. The edge 34 results again the most important pipe, but the metric now identifies a new main path, i.e., those between vertices 24-10. This fact demonstrates that the function $f(R_s,R_t) = R_s \cdot R_t$ correctly emphasizes the most important pipes for water supplying security at the hospital.

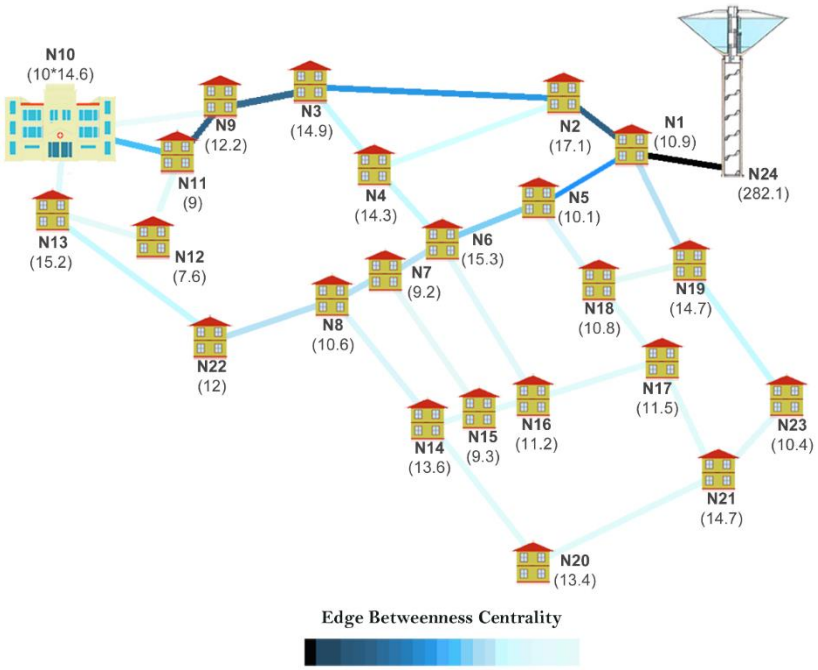


Figure 63 Intrinsic relevance-embedding edge betweenness when a hospital is present in the vertex 10. Its intrinsic relevance is ten times its demand. The proposed edge betweenness again identifies pipe 34 as the most important, but the main path in the network changes in favor of those connecting the source node and the hospital. Source: *Giustolisi et al*., 2020.

The analysis highlights even more the need of embedding the intrinsic relevance of nodes for the centrality evaluation of WDNs.

Here we report the application of the Relevance-embedding edge betweenness to several real networks with different size and asset characteristics, already analysed with the WDN-Tailored edge betweenness (*Giustolisi et al*., 2019). The metric is computed using the function $f(R_s,R_t) = R_s \cdot R_t$, because it correctly emphasizes the most important pipes/ path for WDNs. Figures 64-69 and Table 10 demonstrated the crucial role played by intrinsic relevance in assessing the correct vertex and edge ranking. The comparison between flowrate (left) and the Relevance-embedding edge betweenness weighted with the hydraulic resistance (right) demonstrated that the topological metric identifies the domain of the network, i.e., it captures the emerging hydraulic behaviour of the system.

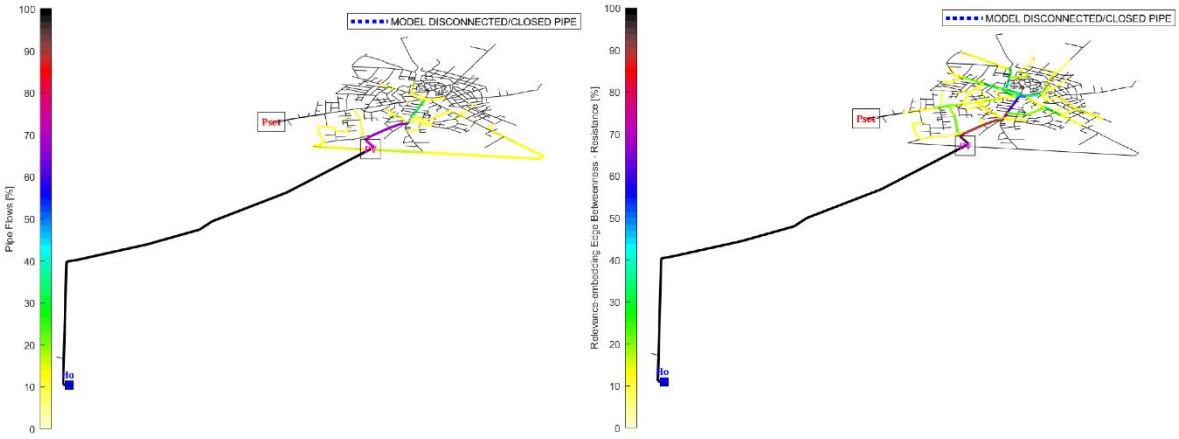


Figure 64. Hydraulic metric pipe flowrate (left) and Relevance-embedding edge betweenness for Apulian 1 network (right). The topological metric identifies the pipe close to the only reservoir as the most important for WDN hydraulics. The main path, from the reservoir to the city center, also matches the two metrics. The correlation index between two metrics is equal to 0,815, i.e., the relevance-embedding edge betweenness very well identifies the hydraulic behavior of the networks.

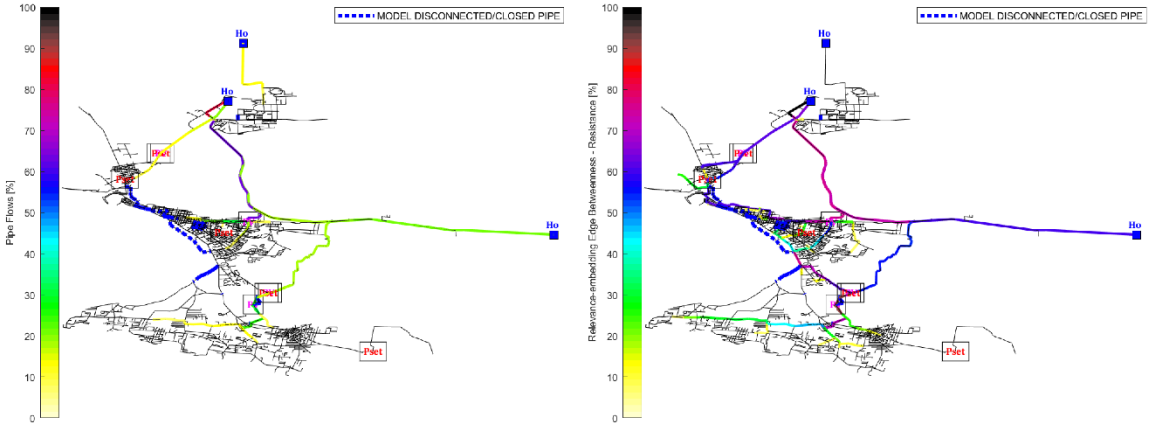


Figure 65. Hydraulic metric pipe flowrate (left) and Relevance-embedding edge betweenness for Apulian 2 network (right). The network is fed by 3 reservoirs. Even if the topological metric is not able to capture the importance of the main paths, it identifies both the range of importance of the reservoirs and of the main paths. In fact, the correlation index between two metrics is equal to 0,691, i.e., it well identifies the hydraulic behavior of the network.

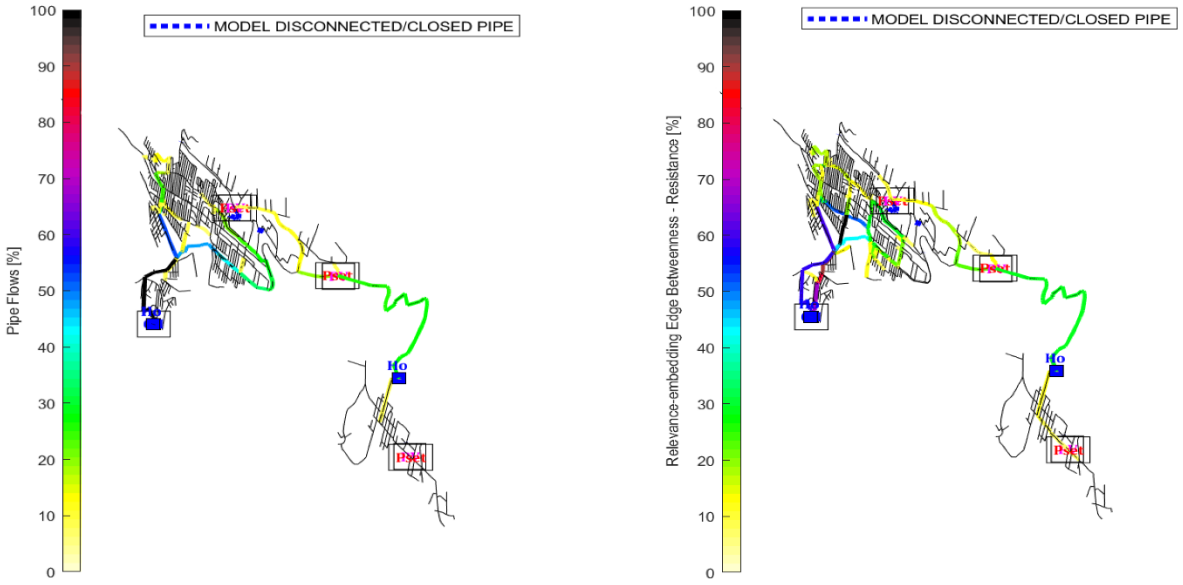


Figure 66. Hydraulic metric pipe flowrate (left) and Relevance-embedding edge betweenness for Apulian 3 network (right). The network is fed by 3 reservoirs, each with its own intrinsic relevance. The two metrics identify the same pipes as more relevant within the network. The correlation index between two metrics is equal to 0,751, i.e., the Relevance-embedding edge betweenness well identifies the hydraulic behavior of the network.

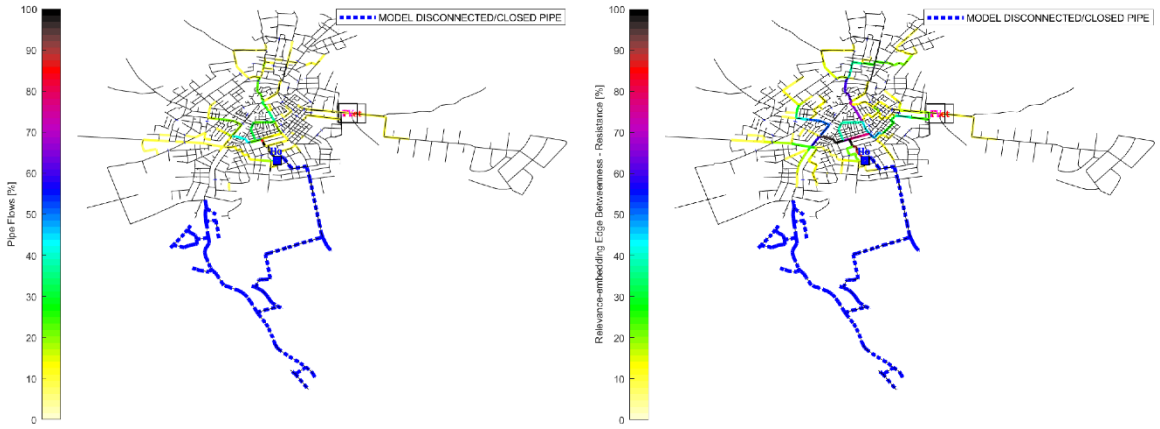


Figure 67. Hydraulic metric pipe flowrate (left) and Relevance-embedding edge betweenness for Apulian 4 network (right). Apulian 4 is a very looped network composed with a portion of network no longer in use (dotted blue). The novel metric correctly identifies the pipe close to the reservoir as the most important edge, even if, since the high redundancy of the network, it identifies other paths as relevant with respect to the pipe flow rates. This may be due to the fact that the topologic metric considers the diameters and not the flow that effectively crosses them. The correlation index between two metrics is equal to 0,843, i.e., the Relevance-embedding edge betweenness well identifies the hydraulic behavior of the networks.

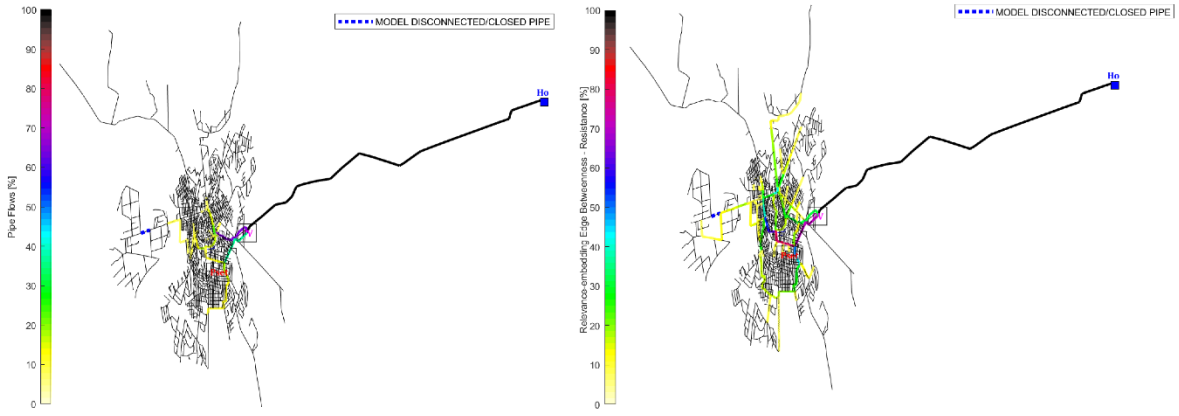


Figure 68. Hydraulic metric pipe flowrate (left) and Relevance-embedding edge betweenness for Apulian 5 network (right). Apulian 5 is a very looped network of great size. The topological metric identifies a greater number of main paths within the network than the hydraulic metric, even if the main paths coincide for both. In fact, the correlation index, equal to 0,818, shows once again that the analysis is close to the hydraulic of the system.

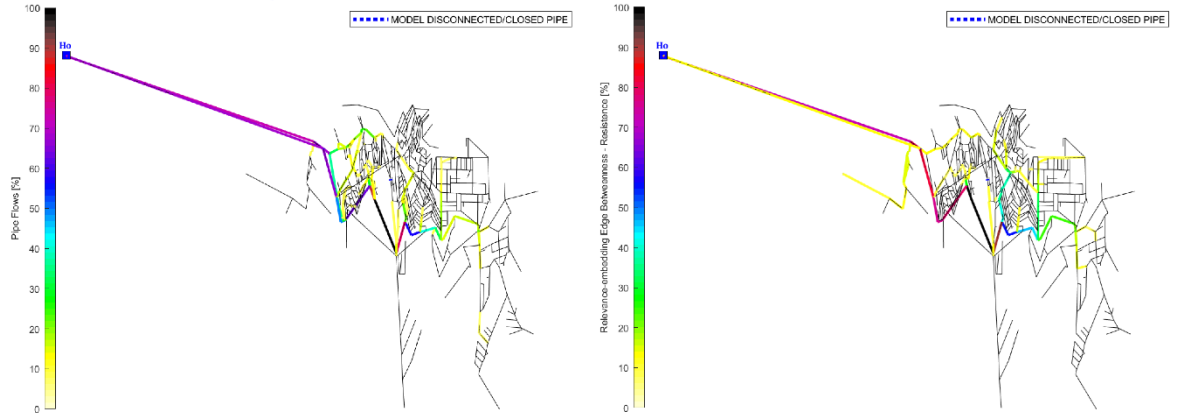


Figure 69. Hydraulic metric pipe flowrate (left) and Relevance-embedding edge betweenness for Apulian 6 network (right). The topological metric identifies the main paths as identified by the flow, even if the range of values differs. The correlation index between two metrics is equal to 0,677, i.e., the Relevance-embedding edge betweenness well identifies the hydraulic behavior of the networks.

Table 10 reports the Spearman correlation index (*Spearman*, 1904) between two metrics of the six analysed networks considering several weights for the edge betweenness, i.e., connectivity, length, and resistance, in order to includes the domain characteristics of the networks into the analysis.

The Figures demonstrated how the Relevance-embedding edge betweenness is effective in identifying the emerging behaviour of WDNs. These results are confirmed by the Spearman correlation index, which for the metric vary between 0,60 and 0,84.

Table 10. Topological characteristics and correlation index for the six analysed WDNs using flow rate and Relevance-embedding edge betweenness.

WDN Name	Node #	Pipe #	Reservoirs #	Tanks #	Corr. Index Resistance	Corr. Index Connectivity	Corr. Index Lenght
Apulia 1	986	1,105	1	0	0,82	0,75	0,79
Apulia 2	7,716	8,496	3	0	0,69	0,71	0,71
Apulia 3	1,270	1,472	3	0	0,75	0,69	0,63
Apulia 4	3,166	3,483	2	0	0,84	0,81	0,83
Apulia 5	7,164	7,895	1	0	0,82	0,70	0,81
Apulia 6	1,111	1,307	1	0	0,68	0,60	0,63

4.3 WDN-oriented Centrality metrics vs. Intrinsic Relevance of Vertices

The topology-based metrics, as they come from CNT, cannot provide reliable information about the actual importance of nodes and pipes for WDNs, i.e., they cannot be useful to discover the key features of the physical domain (i.e., the network) where hydraulics occurs. To overcome this problem, we firstly proposed the WDN-tailored edge betweenness as centrality metric able of describing the emergent hydraulic behaviour of WDNs. The tailoring is successively enhanced with the Relevance-embedding edge betweenness. The advantages deriving from the second strategy compared to the first are evident looking at the case studies analysed. In particular, the paradigm of the intrinsic relevance of vertices:

- avoid the presence of fictitious nodes in WDN original topology, thus reducing the calculation times;
- consider for the various nodes an intrinsic relevance that can refer to different situations and networks. This makes the metric more flexible and general than the previous one. In the case of WDNs we decided to assign, as intrinsic relevance, the demand for the demand nodes and the sum of the demands for the reservoirs. In this way it is not necessary to topologically increase the relevance of the source nodes, since there is already a ranking of intrinsic relevance of all the nodes;
- presents much more reliable results than those obtained with the WDN-tailored edge betweenness.

In order to demonstrate that the Relevance-embedding edge betweenness better identifies the emergent hydraulic behaviour of WDNs compared to WDN-tailored edge betweenness, Figures 70-75 report diagrams that compare the distribution of the pipe flow rates (sorted in descending order) both with the corresponding distribution of the WDN-tailored edge betweenness and Relevance-embedding edge betweenness, also in logarithmic scale (right

panel) for the six real networks analysed previously. The comparison confirms that Relevance-embedding edge betweenness works better than its predecessor in detecting main paths within WDNs, i.e., in identifying the physical domain of WDNs. The good performance of the two metrics, and the added value of Relevance-embedding edge betweenness, is confirmed by the Spearman correlation index (*Spearman*, 1904), weighted with length, connectivity and resistance (see Table 11).

Table 11. Comparison of correlation indices using flowrate and both WDN-Tailoring edge betweenness and Relevance-embedding edge betweenness. Results show that Relevance-embedding edge betweenness works better in identifying the emergent hydraulic behaviour as described by pipe flow rates.

	WDN-Tailoring Edge Betweenness			**Relevance-embedding Edge Betweenness**		
WDN Name	***Corr. Index Resistance***	***Corr. Index Connectivity***	***Corr. Index Length***	***Corr. Index Resistance***	***Corr. Index Connectivity***	***Corr. Index Length***
Apulia 1	**0,64**	0,71	0,73	**0,82**	0,75	0,79
Apulia 2	**0,60**	0,63	0,65	**0,69**	0,71	0,71
Apulia 3	**0,54**	0,50	0,53	**0,75**	0,69	0,63
Apulia 4	**0,66**	0,75	0,79	**0,84**	0,81	0,83
Apulia 5	**0,62**	0,70	0,75	**0,82**	0,70	0,81
Apulia 6	**0,54**	0,59	0,60	**0,68**	0,60	0,63

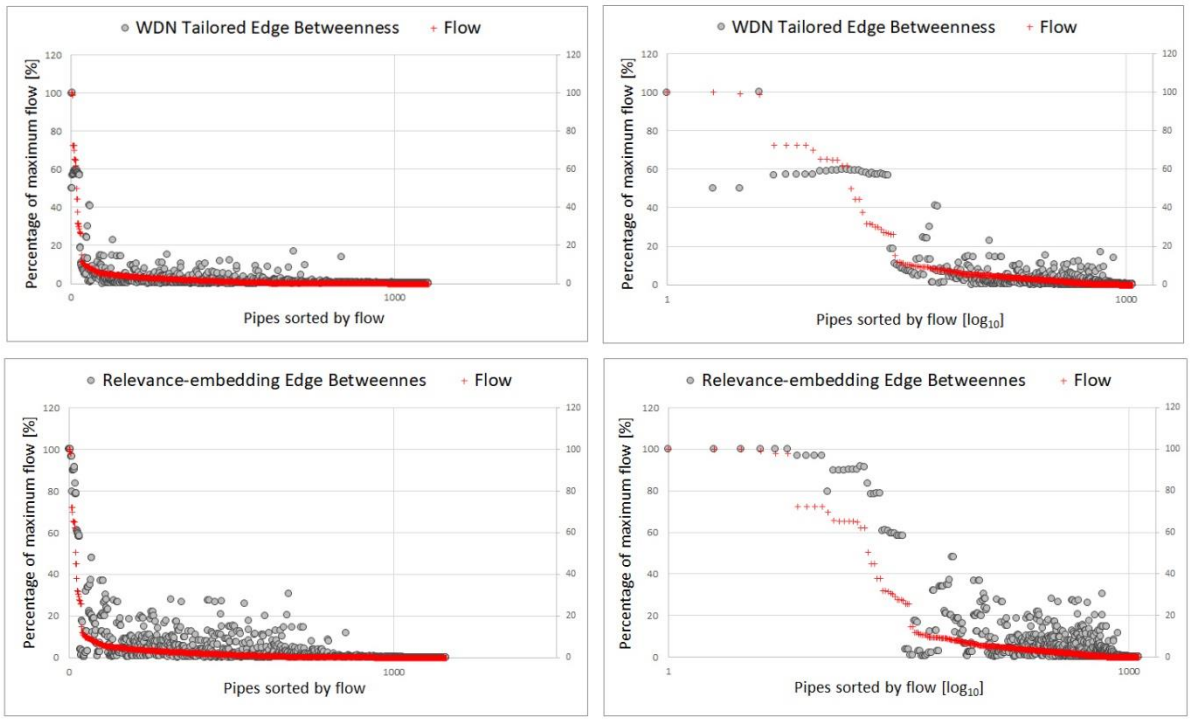


Figure 70. **Top**. Comparison between the pipe flows and the WDN-Tailored edge betweenness for Apulian 1. The correlation index is equal to 0,64. **Bottom**. Comparison between the pipe flows and the Relevance-embedding edge betweenness for Apulian 1. The correlation index is equal to 0,82. The diagrams on the right panel have the x-axis in logarithmic scale.

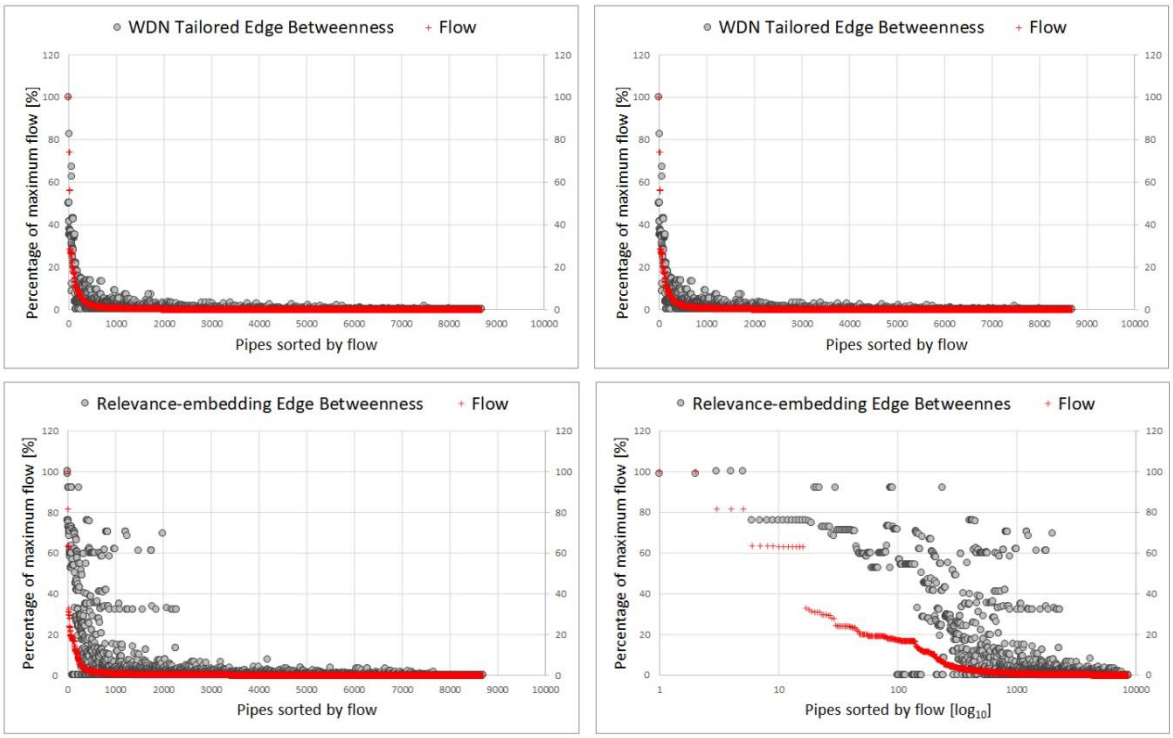


Figure 71. **Top**. Comparison between the pipe flows and the WDN-Tailored edge betweenness for Apulian 2. The correlation index is equal to 0,60. **Bottom**. Comparison between the pipe flows and the Relevance-embedding edge betweenness for Apulian 2. The correlation index is equal to 0,69. The diagrams on the right panel have the x-axis in logarithmic scale.

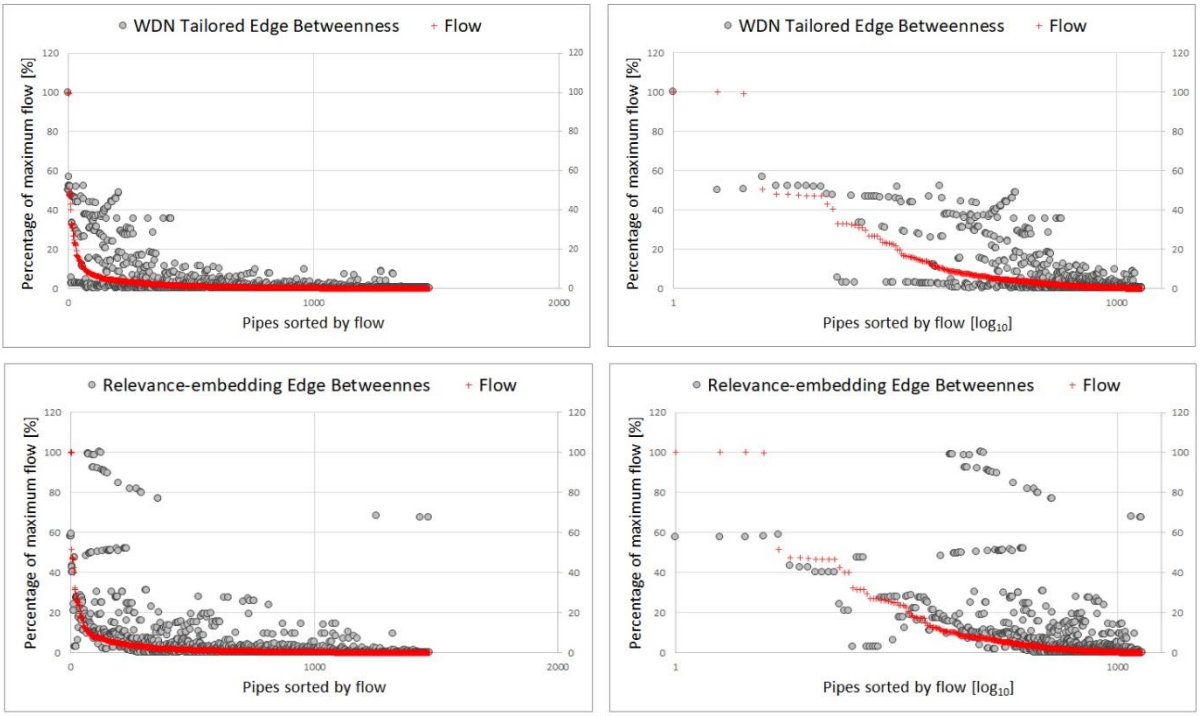


Figure 72. **Top**. Comparison between the pipe flows and the WDN-Tailored edge betweenness for Apulian 3. The correlation index is equal to 0,54. **Bottom**. Comparison between the pipe flows and the Relevance-embedding edge betweenness for Apulian 3. The correlation index is equal to 0,75. The diagrams on the right panel have the x-axis in logarithmic scale.

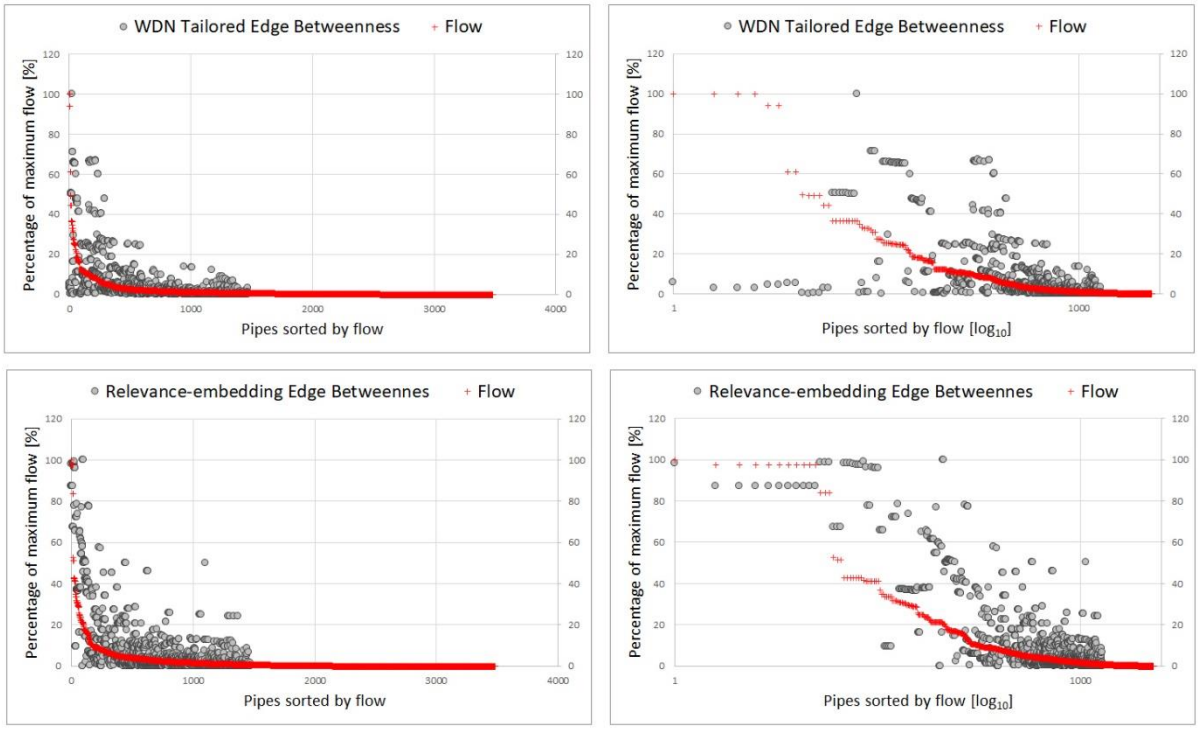


Figure 73. **Top**. Comparison between the pipe flows and the WDN-Tailored edge betweenness for Apulian 4. The correlation index is equal to 0,66. **Bottom**. Comparison between the pipe flows and the Relevance-embedding edge betweenness for Apulian 4. The correlation index is equal to 0,84. The diagrams on the right panel have the x-axis in logarithmic scale.

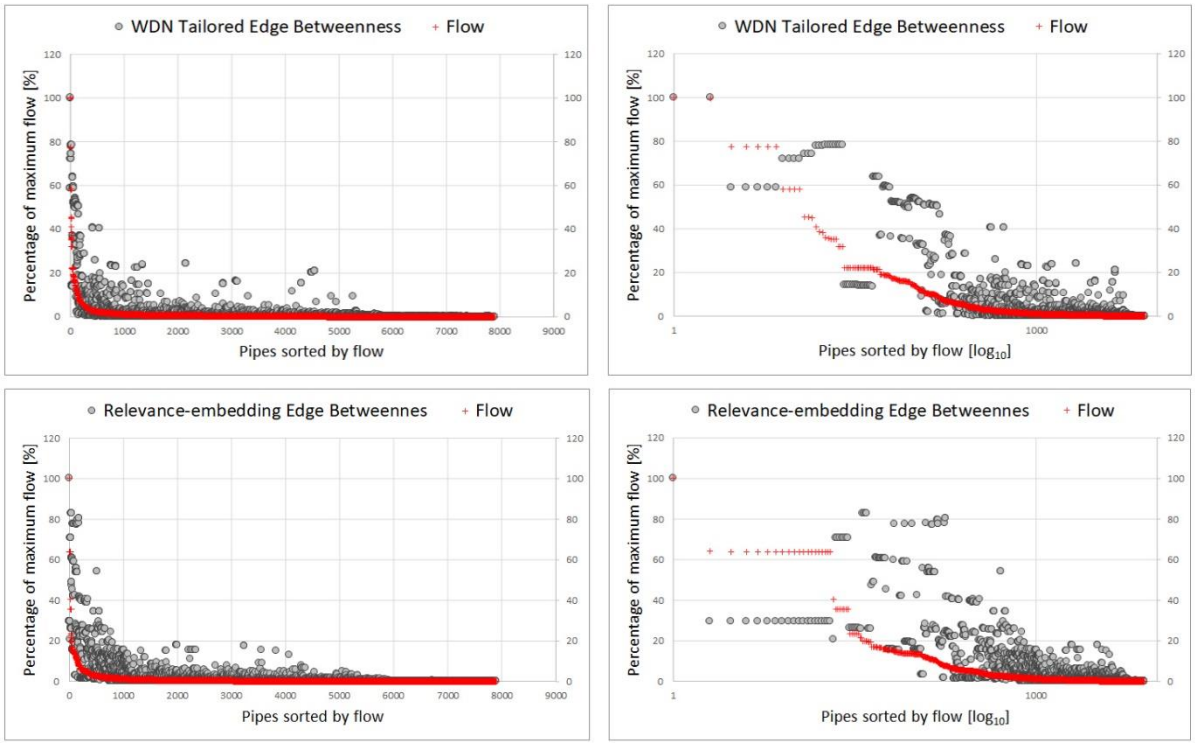


Figure 74. **Top**. Comparison between the pipe flows and the WDN-Tailored edge betweenness for Apulian 5. The correlation index is equal to 0,62. **Bottom**. Comparison between the pipe flows and the Relevance-embedding edge betweenness for Apulian 3. The correlation index is equal to 0,82. The diagrams on the right panel have the x-axis in logarithmic scale.

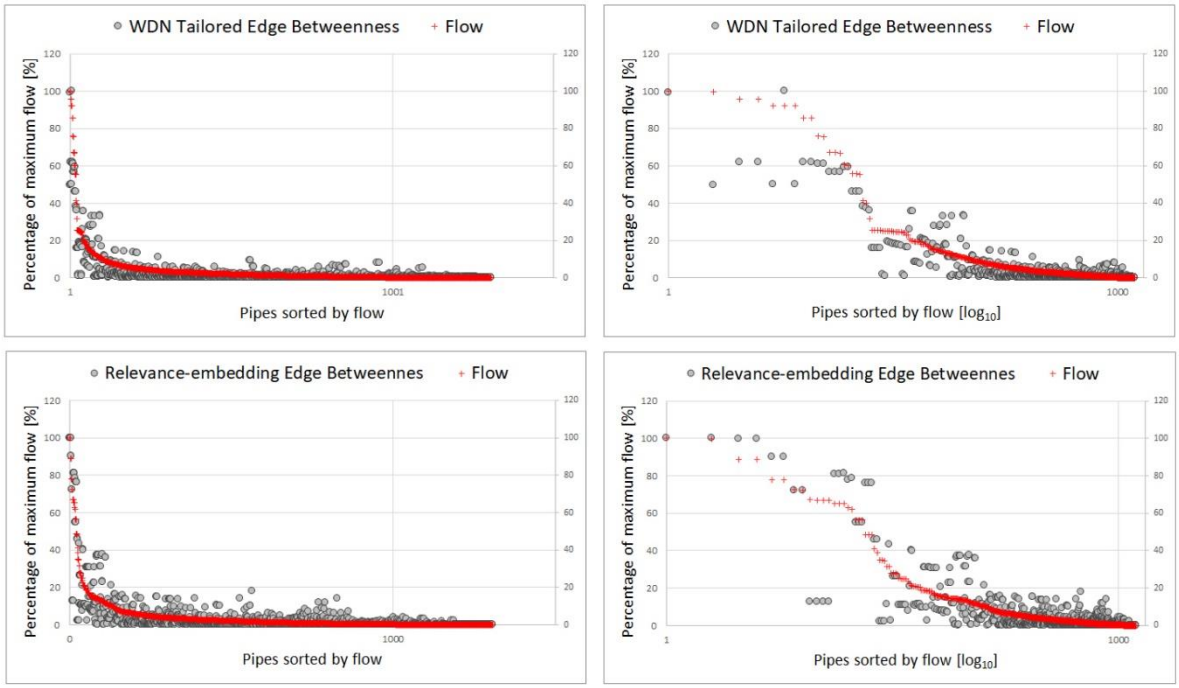


Figure 75. **Top**. Comparison between the pipe flows and the WDN-Tailored edge betweenness for Apulian 5. The correlation index is equal to 0,54. **Bottom**. Comparison between the pipe flows and the Relevance-embedding edge betweenness for Apulian 3. The correlation index is equal to 0,68. The diagrams on the right panel have the x-axis in logarithmic scale.

4.4 WDNs oriented modularity index for DMAs and sampling design

The concept of modularity was among the first to be borrowed from the CNT to be applied to infrastructural networks, adapting it in particular to the analysis and management of WDNs. The tailoring process was proposed in the course of various scientific papers, by *Giustolisi and Ridolfi* (2014a, b) who introduced various basic formulations for the modularity index aimed at combining the peculiarities of the WDN infrastructure in addition to the meaning technician of tubes and modules. This chapter briefly retraces the tailoring process that took place over the years and presents two of the works we produced on the subject. The first work proposed the development of a district strategy, i.e., DMA design, (*Laucelli et al*., 2017) based on a topological segmentation and a hydraulic distrectualization of WDNs. In the second work we extended the concept of modularity index, already tailored, to pressure sampling design (*Simone et al*., 2016), introducing a new metric to perform the optimal location of pressure meters in WDNs. Both works demonstrated the effectiveness of the metric in management applications on WDNs.

Analysis and management of WDNs has always presented many problems related with a non-homogeneous behaviour of these infrastructural systems. Several solutions have been proposed to manage this problem, and the most used and studied is the community detection/segmentation strategy, i.e., the division of the network into smaller modules. Researchers proposed several approaches to segmentation (*Jacobs and Goulter*, 1988; *Yang et al*., 1996; *Walski*, 1983; *Davidson et al*., 2005; *Deuerlein*, 2008; *Perelman and Ostfeld*, 2011; *Alvisi and Franchini*, 2014; *Yazdani and Jeffrey*, 2012; *Scibetta et al*., 2013; *Diao et al*., 2013; *Albert and Barabasi*, 2002; *Newman*, 2004) in order to identify an optimal division of the network with respect to WDN topology and characteristics. Also, CNT studies proposed several approaches/ metrics for real systems segmentation (*Steinhaeuser and Chawla*, 2009) and the modularity index (*Girvan and Newman*, 2004) resulted the most widely accepted and used metric to measure the propensity of the network division into modules. The modularity index derives from a comparison between a real network and a random network that share some common topological characteristics (e.g., the total number of connections to the network). However, the direct use of the original formulation of the modularity index for WDNs is inadequate (*Giustolisi and Ridolfi*, 2014; *Barthélemy*, 2011) because the structure of this type of network is strongly constraints by several limits, and the random network taken as a reference point for the detection of segments has not a clear correspondence with the infrastructural one. To overcome this limit, *Giustolisi and Ridolfi* (2014a) proposed a tailoring of the modularity index for WDNs, defining the WDN-oriented modularity index, accounting for the features of

WDNs using the general topological incidence matrix commonly adopted in hydraulic modeling,

$$Q\left(\mathbf{w}_p\right)=\left\{1-\frac{n_c}{n_p}\right\}+\left\{-\sum_{m=1}^{n_m}\left[\sum_{k=1}^{n_p}\frac{\left(\mathbf{w}_p\right)_k\,\delta\left(\mathrm{M}_m,\mathrm{M}_k\right)}{W}\right]^2\right\}=Q_1+Q_2 \tag{46}$$

where n_c is the number of pipes linking modules of the infrastructure, namely the number of conceptual cuts in the network (i.e., the decision variables of the WDN segmentation problem) and n_m is the number of network modules. The summation inside the square brackets is related to pipe weights stored in the vector $\mathbf{w}_p$, whose sum is W, and Kronecker's δ function makes that the sum refers only to the weights of pipes belonging to the m-th module (i.e., $\delta = 1$ if $M_m=M_k$ and $\delta = 0$ otherwise). The term Q_1 of Eq. (46) decreases with the number of cuts, while Q_2 generally increases with the number of modules. For a given number of modules, Q_2 increases with the similarity of modules to each other with respect to the assigned weights.

They developed a cut position-sensitive index in order to account for the actual position of conceptual cuts, which would later be devices, along pipes (i.e., close to ending nodes instead of the middle of pipes) generating districts and introduced in the formulation of the modularity index the pipe weights, i.e., asset and/or hydraulic information at pipe level. This way, the tailored metric allowed dividing the network into districts similar to each other with respect to an assumed weight. The tailored modularity index was successively extended proposing the attribute–based segmentation index to allow the division of the network into districts having internal similar attribute (e.g., diameter, average elevation, average pressure, etc.),

$$Q_a=1-\frac{n_c}{n_p}-\sum_{m=1}^{n_m}\left[\frac{\sum_{k=1}^{n_p}\left|\mathbf{a}_p-\bar{a}\left(\mathrm{M}_m\right)\right|_k\,\delta\left(\mathrm{M}_m,\mathrm{M}_k\right)}{\sum_{k=1}^{n_p}\left|\mathbf{a}_p-\bar{a}\left(\mathrm{N}\right)\right|_k}\right]^2 \tag{47}$$

where $\mathbf{a}_p$ is the vector of pipe attributes, $\bar{a}(\mathrm{N})$ is the mean value of the pipe attributes of the network N, i.e., of $\mathbf{a}_p$, and $\bar{a}(\mathrm{M}_m)$ is the mean value of the pipe attributes in M_m. Function δ limits the summation of the pipe attributes to the elements belonging to the same module.

Furthermore, since the tailored metric was still affected by resolution limit (*Fortunato and Barthélemy*, 2007), which causes the non-identifiability of smaller modules depending on the size of the network, *Giustolisi and Ridolfi* (2014b) proposed the infrastructure modularity index to overcome such limit,

$$IQ\left(\mathbf{w}_p\right)=1-\frac{n_c-\left(n_{act}-1\right)}{n_p}-\sum_{m=1}^{n_m}\left[\sum_{k=1}^{n_p}\frac{\left(\mathbf{w}_p\right)_k\,\delta\left(\mathrm{M}_m,\mathrm{M}_k\right)}{W}\right]^2 \tag{48}$$

where n_{act} is the actual number of modules satisfying given constraints (e.g., the minimum length of the modules, the minimum number of pipes, etc.).

Starting from these findings, we proposed a comprehensive application of the infrastructural modularity, i.e., a distrectualization (DMA design) strategy (*Laucelli et al*., 2017) based on two phases: topological segmentation and hydraulic distrectualization. The first phase of the strategy involves a virtual segmentation of the topological structure of the network with conceptual cuts in order to obtain virtual segments. This phase is based on a two-objective optimization that, on the one hand, maximizes the modularity index which measures the efficiency of the topological division into segments and, on the other, minimizes the number of conceptual cuts that separate the segments. The second phase involves the installation of devices, flow meters or closed valves, in the location of the conceptual cuts. This phase is based on a three-objective optimization that, on the one hand, minimizes the number of flow meters and the total unsupplied customer demand and, on the other, maximize the reduction of background leakages. Finally, the application of the segmentation strategy divides the system in a number of segments bounded by installed devices allowing to define district metering areas (DMAs), which are useful for different technical purposes, e.g., demand and background leakages management, burst detections, rehabilitation works, model calibration, etc.

We further extended the concept of network segmentation to pressure sampling design (*Simone et al*., 2016), introducing the sampling-oriented modularity index to perform the optimal location of pressure meters, i.e., to optimize the pressure monitoring system mainly based on network topology and weights assigned to pipes according to the specific technical tasks.

Starting from Eq.s (46-48) we defined, respectively, the sampling-oriented modularity Q_s, sampling-oriented infrastructure modularity IQ_s, sampling-oriented attribute–based modularity $Q_{a\text{-}s}$ and sampling-oriented infrastructure attribute–based modularity $IQ_{a\text{-}s}$,

$$
\begin{aligned}
Q_s\left(\mathbf{w}_p\right) &= 1-\frac{n_{obs}}{n_n}-\sum_{m=1}^{n_m}\left[\sum_{k=1}^{n_p}\frac{\left(\mathbf{w}_p\right)_k\,\delta\left(\mathrm{M}_m,\mathrm{M}_k\right)}{W}\right]^2 \\
IQ_s\left(\mathbf{w}_p\right) &= 1-\frac{n_{obs}-\left(n_{act}-1\right)}{n_n}-\sum_{m=1}^{n_m}\left[\sum_{k=1}^{n_p}\frac{\left(\mathbf{w}_p\right)_k\,\delta\left(\mathrm{M}_m,\mathrm{M}_k\right)}{W}\right]^2 \\
Q_{a-s}\left(\mathbf{a}_p\right) &= 1-\frac{n_{obs}}{n_n}-\sum_{m=1}^{n_m}\left[\frac{\sum_{k=1}^{n_p}\left|\mathbf{a}_p-\bar{a}\left(\mathrm{M}_m\right)\right|_k\,\delta\left(\mathrm{M}_m,\mathrm{M}_k\right)}{\sum_{k=1}^{n_p}\left|\mathbf{a}_p-\bar{a}\left(\mathrm{N}\right)\right|_k}\right]^2 \\
IQ_{a-s}\left(\mathbf{a}_p\right) &= 1-\frac{n_{obs}-\left(n_{act}-1\right)}{n_n}-\sum_{m=1}^{n_m}\left[\frac{\sum_{k=1}^{n_p}\left|\mathbf{a}_p-\bar{a}\left(\mathrm{M}_m\right)\right|_k\,\delta\left(\mathrm{M}_m,\mathrm{M}_k\right)}{\sum_{k=1}^{n_p}\left|\mathbf{a}_p-\bar{a}\left(\mathrm{N}\right)\right|_k}\right]^2
\end{aligned}
\tag{49}
$$

where n_{obs} is the number of removed nodes, i.e., where pressure meters will be installed, n_n is the number of nodes, n_m is the number of modules n_m identified by removing nodes and n_{act} is the actual number of modules matching the technical constraints to avoid excessive resolution of the segmentation. The term Q_2 of Eqs. (49) refers to the characteristics of modules, which are now created by removed nodes.

The strategy is based on a two-objective optimization that, on the one hand, maximize the sampling-oriented modularity index and, on the other, minimizes the cost of pressure meters. The strategy has several analogies with the optimal segmentation design for district monitoring areas (DMAs) and integrates the need of dividing a WDN considering pressure meters in order to know the pressure status. Therefore, with respect to the previous metrics the sampling-oriented modularity index segments the network using conceptual removal of nodes and not conceptual cuts positioned on pipes and for segment identification it uses the connectivity matrix of the edges/pipes $\mathbf{L}= \mathbf{A}_{pn}$ x $\mathbf{A}_{np}$, and not the classic adjacency matrix **A**.

Furthermore, pressure DMAs do not involve the modification of the network topology, which also affects the hydraulic status of the WDN, differently from the classic DMAs, that modifies both the network topology and hydraulic status installing closed valves. I.e., for the sampling design strategy, the network division is conceptual.

4.4.1 Optimal DMA Design

The division of WDNs into districts is a common practice for multiple technical purposes, generally aiming at planning and management problems. With the aim of finding effective solutions to relevant management problems, we proposed the Optimal Design of District

Metering Areas (Optimal DMA Design) strategy. The strategy includes a first topological phase that focuses on the network structure, in order to segment the hydraulic domain, and a second phase that focuses on the hydraulic behaviour, in order to define the location of flow measurement and closed valves to create operative districts. Therefore, once again, the network connectivity structure represents the main driver of the hydraulic behaviour, i.e., the network connectivity structure of the hydraulic system dominates the district design with respect to the hydraulic behavior.

Optimal network segmentation. The two-objective optimization of the first phase minimizes the number of conceptual cuts and maximizes the WDN-oriented modularity index,

$$\begin{cases} [M, n_c, n_{act}] = \text{Connectivity}\left(I_c, \left|\mathbf{A}_{np}\right|\right) \\ f_1 = \max\left\{Q\left(\mathbf{L}_p\right)\right\} = \max\left\{\left(1 - \dfrac{n_c}{n_p}\right) - \sum_{m=1}^{n_{act}}\left[\sum_{k=1}^{n_p}\dfrac{\left(\mathbf{L}_p\right)_k \delta\left(\mathrm{M}_m, \mathrm{M}_k\right)}{L}\right]^2\right\} \\ f_2 = \min\left\{n_c\right\} \end{cases} \tag{50}$$

where $\mathbf{A}_{np}$ is the incidence matrix, I_c is the set of n_c cuts in the network, decision variables of the optimization, the Connectivity (I_c,|$\mathbf{A}_{np}$|) stands for component analysis of the undirected graph for the given cuts, L is the sum of pipes lengths, stored in the vector $\mathbf{L}_p$, and n_{act} indicate the cuts that are *actually* used to separate modules. The modularity index measures the similarity of the segments with respect to selected pipe weights (e.g., in the case of lengths the similarity of segment lengths) for a given number of conceptual cuts dividing the network in n_m segments. Therefore, its maximization provides the maximum number of modules, which are similar with respect to pipe weights, obtained with the minimum number of cuts. The already existing devices (e.g., control valves, pumps, closed valves, and reservoir) represent priors and can be considered as already existing conceptual cuts. The optimization problem in Eq. (50) returns a Pareto front of optimal segmentation solutions generating modules that are similar to each other in terms of total length for a given number of cuts, i.e., each segmentation solution consists of a set of conceptual cuts dividing the network structure of the WDN into modules.

The optimal segmentation is constrained to search for nested solutions, i.e., solutions with higher resolution (with more conceptual cuts) are nested in the more parsimonious segmentations (with less conceptual cuts). This fact provides flexibility to the segmentation design because it is possible to start from any optimal scenario, depending on the preferred resolution of the districts, which is generally driven by the budget uncertainty and the growing knowledge of the system (*Laucelli et*

al., 2017). One solution is selected among the optimal segmentation solutions as the basis for the actual DMA design.

Optimal design of the network DMAs. The three-objectives optimization of the second phase minimizes the number of flow observations and the unsupplied customer demand and maximizes the reduction of leakages,

$$\begin{cases} f_1 = \min\left\{n_{fm}\right\} \\ f_2 = \max\left\{1 - \dfrac{V_T^{leak}(\mathbf{v})}{V_T^{leak}}\right\} \\ f_3 = \min\left\{1 - \dfrac{V_T^{cust}(\mathbf{v})}{V_T^{cust}}\right\} \\ \\ \text{subject to:} \\ \begin{bmatrix} \mathbf{A}_{pp}\mathbf{Q}_p(t) + \mathbf{A}_{pn}\mathbf{H}_n(t) & = & -\mathbf{A}_{p0}\mathbf{H}_0(t) + \mathbf{H}_p^{pump}(t) \\ \mathbf{A}_{np}\mathbf{Q}_p(t) - \left[\mathbf{d}_n^{leaks}(\mathbf{H}_n,t) + \mathbf{d}_n(\mathbf{H}_n,t)\right] & = & \mathbf{0}_n \end{bmatrix} \quad \forall t=1,...,T \\ \\ \text{and} \quad \textit{Technical Constraints} \end{cases} \tag{51}$$

where n_{fm} is the number of flow meters; V_T^{cust} and V_T^{leak} are namely the total customer demand and leakages over T-steps extended period simulation (EPS), in the original WDN configuration (i.e., without accounting for devices); $V_T^{cust}(\mathbf{v})$ and $V_T^{leak}(\mathbf{v})$ are the same figures but referred to the set $\mathbf{v}$ of closed gate valves located at "conceptual cuts" (i.e., representing the implementation of the DMAs). $\mathbf{d}_n$ and $\mathbf{d}_n^{leaks}$ are the vectors of water demand and background leakages (lumped at nodes) which depend on time and current pressure status. In order to account for the emptying/filling process of tanks during optimal DMA design the following constraints have been added to the formulation (50),

$$\begin{cases} H_s^{ini} \le H_s^{final} \\ H_s(t) \ge H_s^{\min} \end{cases} \tag{52}$$

where s is the subscript of the sth tank node; $H_s^{\min}$ is the minimum head level at the sth tank node; H_s^{ini} is the initial head level at the sth tank node varying over time t and H_s^{final} is the final level at the sth tank node. The constraints in (52) guarantee system reliability because the first equation states that the initial level (i.e., volume) in each tank should be lower than the final of the EPS, while the second states that the tank level cannot be lower than the minimum during

the extended period simulation (EPS). It is important noting that the strategy imposes a pressure-driven analysis (*Giustolisi et al*., 2008; *Giustolisi and Walski*, 2012).

This phase is performed using as basis one of the optimal segmentation solutions. The procedure aims at determining the position of flow meters and closed valves in the conceptual cuts dividing the network structure of the WDN in segments. The installation of closed valves allows the design of actual districts in the hydraulic systems and the reduction of leakages thanks to the reconfiguration of flows and to the consequent reduction of the pressure status within the system.

A connectivity analysis between the first and second phase is performed in order to verify that the closure of any candidate closed valve in the conceptual cuts does not cause the disconnection of portions of the network. In this case the valve is automatically removed from the conceptual cuts by replacing it with a flow meter.

Case study: Exnet WDN

Exnet network has been used to show the effectiveness of the Optimal DMA strategy (*Laucelli et al*., 2017). The network has a real medium size and is composed of 1,894 nodes, 2,471 pipes and 7 reservoirs, whose layout is reported in Figure 76. It is important noting that in both phases, the presence of existing devices is considered as a prior information in the analysis.

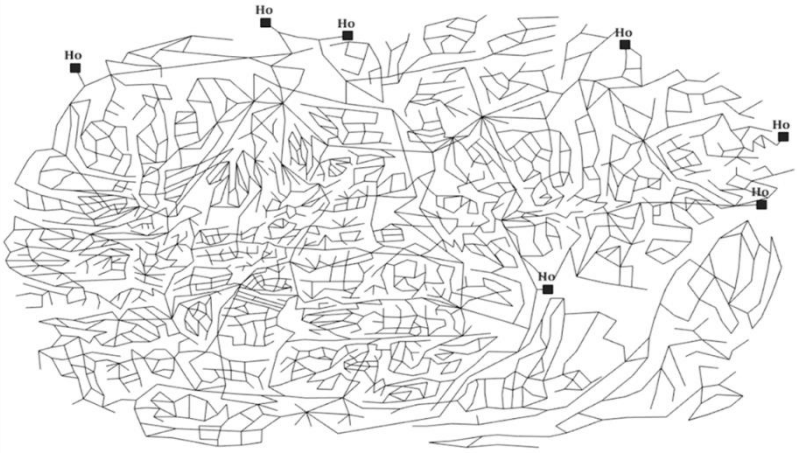


Figure 76. Layout of Exnet network.

The first optimization procedure returned a Pareto front of optimal segmentations solutions by solving the problem of Eqs. (50). The optimal solutions are reported in Figure 77. In fact, the first optimal solution contains the minimum number of conceptual cuts, i.e., those constrained by the pipes delivering water from the seven reservoirs, while the last optimal solution contains ninety-nine conceptual cuts that divide the network in 32 modules.

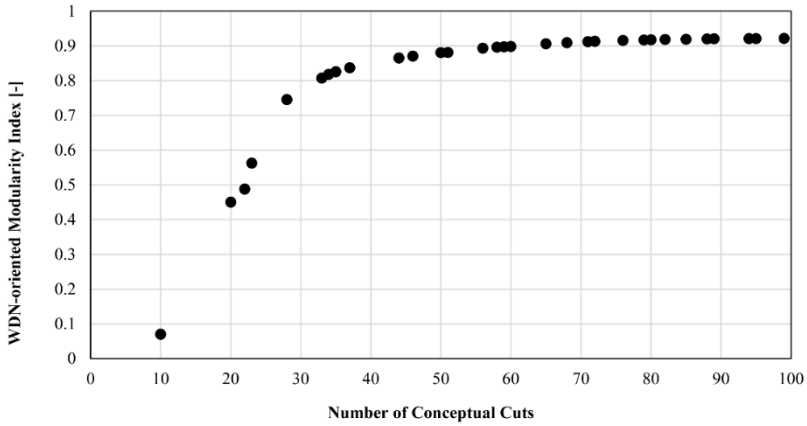


Figure 77. Pareto set of optimal segmentation solution for Exnet network. Source: *Laucelli et al*., 2017.

This last solution, with the maximum number of cuts is reported in Figure 78 and corresponds to the maximum value of the WDN-oriented modularity index. It is important noting that two constraints have been assigned to the analysis: (i) flow meters are mandatory close to reservoirs and branched portions of the network; (ii) high diameter (>500 mm) are not considered as potential position for closed valves, because their closure could result into massive reduction of system hydraulic capacity for which the WDN was originally designed. The technical constraints result into 31 mandatory flow meters and, therefore, the problem concerns the decision about installing closed valves or flow metering for 68 out of 99 conceptual cuts.

Figure 78. Optimal segmentation for Exnet with 32 modules and 99 cuts, corresponding to maximum value of modularity index. Source: *Laucelli et al*., 2017.

The optimal segmentation solution, with the maximum value of the WDN-oriented modularity index, is selected as the basis for the second phase. The second optimization procedure returned a Pareto front of 12 optimal DMA solutions (Figure 79), all characterized by null unsupplied demand.

The solutions present a number of flow meters ranging from 16 to 27, in addition to those already defined above, corresponding to closed valves ranging between 52 and 41, respectively.

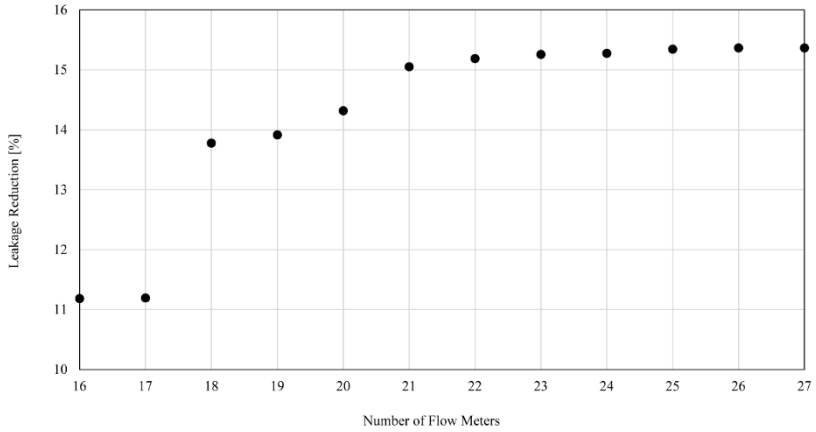


Figure 79. Twelve DMA design solutions with null unsupplied water. Source: *Laucelli et al*., 2017.

In terms of leakage, the solutions guarantee a reduction between about 11% and 15 % with respect to the original configuration. Each solution generates a different distribution of flow paths with respect to the original ones of the hydraulic system (i.e., without DMAs definition) since number and position of closed valves changes between the various configurations. This fact also affects the pressure conditions for each of them. Figures 80 and 81 report the different pressure conditions for solutions 1 and 12, which are obtained with a different number of flow meters, 27 and 16 respectively (corresponding to 41 and 52 closed valves).

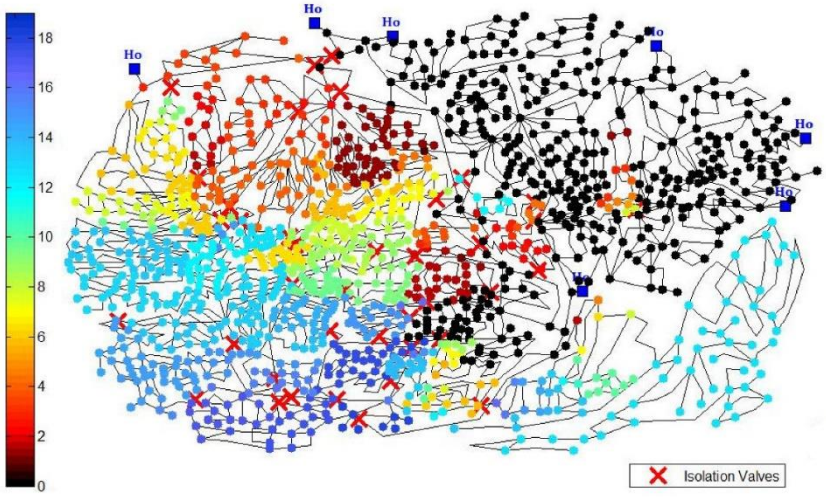


Figure 80. Pressure drop of Solution 1 causing a leakage reduction of 15.4% using 27 m and 41 valves. Source: *Laucelli et al*., 2017.

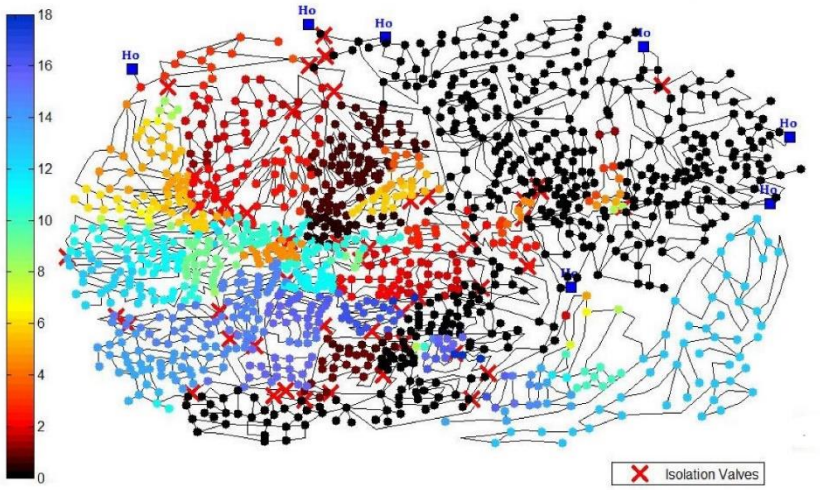


Figure 81. Pressure drop of Solution 12 causing a leakage reduction of 11.2% using 16 m and 52 valves. Source: *Laucelli et al*., 2017.

The importance of certain pipes in the network is highlighted by their presence in almost all optimal configurations. I.e., installing a closed valve on such pipes almost always corresponds to prompting a reconfiguration of the flows which brings advantages both in terms of pressure and leakages reduction. Obviously, this analysis is local and must be inserted in the wider context of the water network. Figure 82 reports the percentage of times that each candidate conceptual cuts is selected for installing a closed valve, over the twelve optimal DMA design solutions.

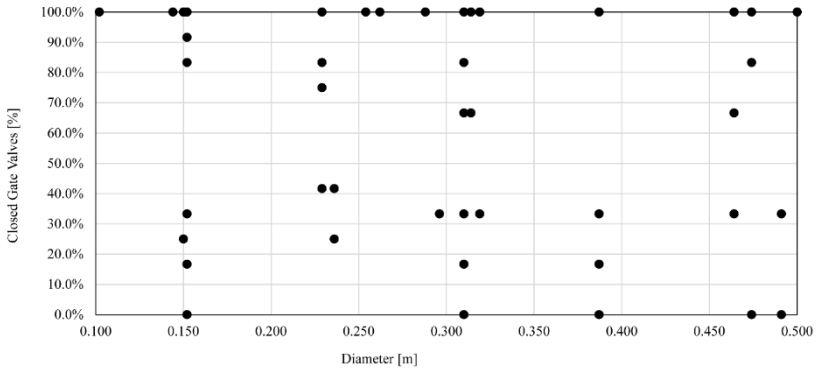


Figure 82. Average use of closed valves in the 68 candidate positions versus diameter. Source: *Laucelli et al*., 2017.

A comparison between solution 1 and 12, in terms of the number of valves installed in the network, highlights the fact that installing a few closed valves on large diameters (solution 1) is more effective than installing many valves on small diameters (solution 12). Therefore, the trend indicates a greater leakage reduction for solutions with a lower number of closed valves because they are installed in strategic topologic positions (i.e., more important flow paths) with respect to the original flow paths of the hydraulic system, which is not the case when many valves are installed aiming only at the need of topological division of the network structure without significantly changing the original system configuration.

4.4.2 Optimal Sampling Design

Sampling design is of key importance for various WDN analysis and management tasks by means of analysis and monitoring of pressure status. The present optimal sampling design strategy extends the concept of WDN-oriented modularity metrics (*Giustolisi and Ridolfi*, 2014) defining the sampling-oriented modularity indexes, i.e., the metrics are modified in order to consider the positioning of nodal pressure meters and introducing the "conceptual removal of nodes" that divides the networks into "pressure DMAs".

The strategy includes a multi-objective optimization based on a topological analysis that minimizes the cost of pressure meters and maximizes the sampling-oriented modularity index.

The two-objective optimization of the problem, that minimizes the number of conceptual removal of nodes and maximizes the sampling-oriented infrastructure modularity index, are

$$\begin{cases} [M, n_{obs}, n_{act}] = connectivity\left(I_c, \mathbf{L}\right) \\ f_1 = \max\left\{IQ\left(\mathbf{w}_p\right)\right\} = \max\left\{1 - \dfrac{n_{obs} - \left(n_{act} - 1\right)}{n_n} - \sum_{m=1}^{n_m}\left[\sum_{k=1}^{n_p}\dfrac{\left(\mathbf{w}_p\right)_k \delta\left(\mathrm{M}_m, \mathrm{M}_k\right)}{W}\right]^2\right\} \\ f_2 = \min_{I_c}\left\{n_{obs}\right\} \end{cases}$$

considering

$$\Delta\mathbf{H}_n = \left(\mathbf{A}_{nn}\right)^{-1}\Delta\mathbf{d}_n$$
$$\Delta\mathbf{Q}_p = \left(\mathbf{D}_{pp}\right)^{-1}\mathbf{A}_{pn}\left(\mathbf{A}_{nn}\right)^{-1}\Delta\mathbf{d}_n \qquad (53)$$

and

$$C_1 = \frac{\sum_{k=1}^{n_p}\left(\mathbf{L}_p\right)_k \delta\left(\mathrm{M}_m, \mathrm{M}_k\right)}{L} \geq x\%$$

where **L** is the edge/pipe adjacency matrix, I_c is the set of nobs conceptually removed nodes in the network, the decision variables, and connectivity (**L**, I_c) stands for component analysis of the graph with respect to edge, x% is a threshold percentage to be assumed. C_1 is a sort of "pressure" for the optimization to search for solutions characterized by modules having length larger than x% of the total network length L. The already existing pressure meters are considered as constraints.

Case study: Exnet WDN

The optimization procedure returns a Pareto front of optimal solutions by solving the problem of Eqs. (53). Figure 83 reports the two Pareto sets corresponding to the solution of the two-optimization procedures using the classical and the infrastructural version of the metric. The solutions present a number of flow meters ranging from 2 to 41 for the first version of the metric and from 2 to 101 for the second one.

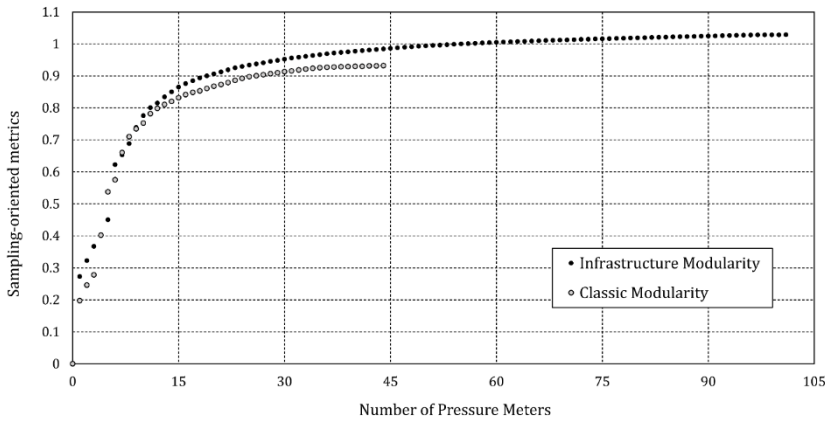


Figure 83. Pareto set of optimal solutions, i.e., optimal sampling configurations, corresponding to the use of the classic and infrastructural sampling-oriented modularity. Source: *Simone et al*., 2016.

The last solutions corresponding to the maximum value of the sampling-oriented metrics are reported in Figure 84 and 85, where the 44 pressure meters determine 45 "pressure DMAs" and 101 pressure meters determine 205 "pressure DMAs".

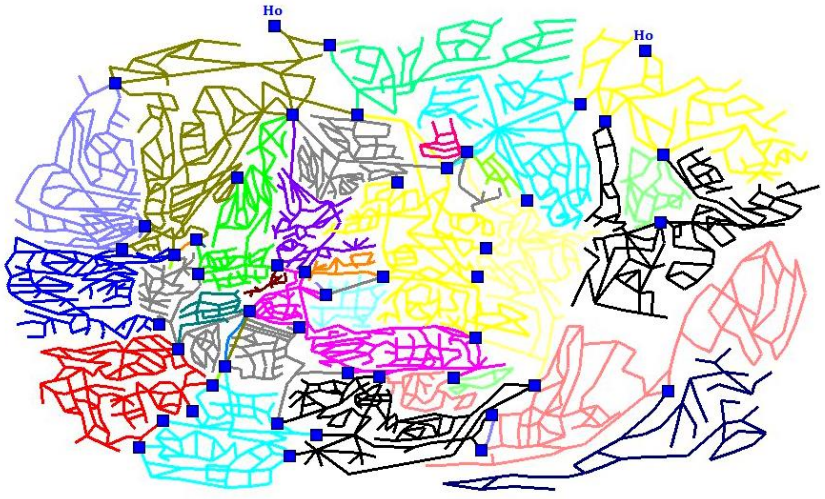


Figure 84. Sampling design configuration corresponding to the maximum value of the sampling-oriented modularity. The bleu squares represent the locations of pressure meters. Source: *Simone et al*., 2016.

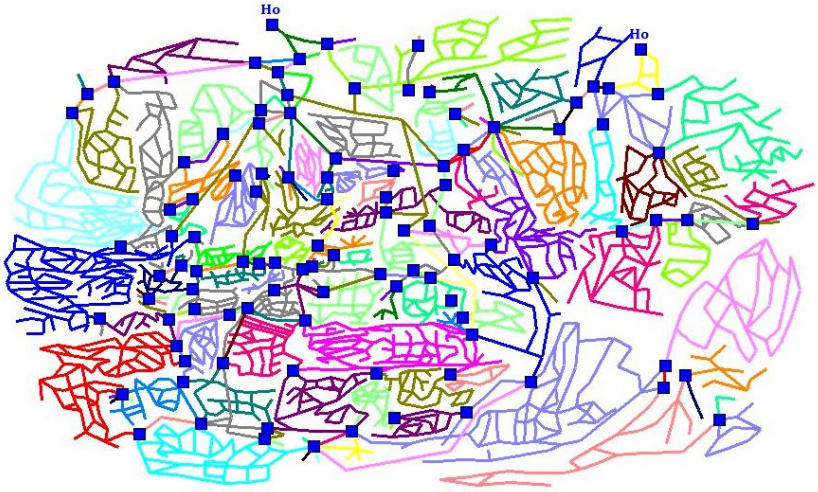


Figure 85. Sampling design configuration corresponding to the maximum value of the sampling-oriented infrastructure modularity. The bleu squares represent the locations of pressure meters. Source: *Simone et al*., 2016.

The sampling design strategy is refined implementing further meters internal to the pressure DMAs, in order to increase the information collected by the pressure monitoring system (see Figure 86). The betweenness centrality is used as topological metric to identify the position of the most relevant nodes within each pressure DMA. The node with the maximum value of betweenness centrality represent the most central within each pressure DMA.

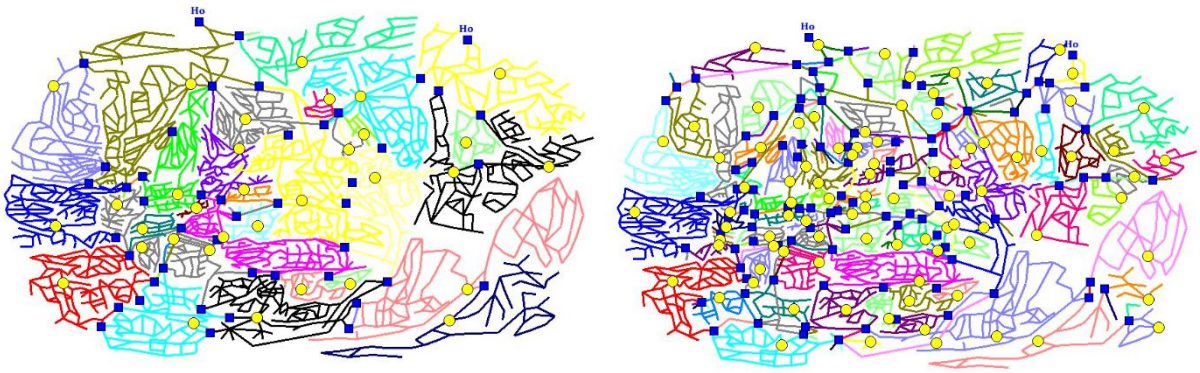


Figure 86. Sampling design configuration corresponding to the maximum value of the classic (left) and infrastructure (right) sampling-oriented modularity. The bleu squares represent the locations of pressure meters and the yellow circles represent the locations of pressure meters positioned by means of betweenness centrality. Source: *Simone et al*., 2016.

A greater number of meters, and therefore of pressure DMAs, is recommended for management intervention that focuses on leakages detection.

5

Future Perspectives

Complex Network Theory (CNT) centrality metrics represent one of the most used tools for network classification, whose reliability increases as more data becomes available. This is valid for most real complex systems, but not for infrastructural networks (e.g., WDNs). In fact, for WDNs the situation is different, because even if many topological data about position of nodes (coordinates) and connectivity between pipes are available, the classification of these systems using the classic CNT tools cannot, however, be reliable. This because, due to the spatial limits that affect them, WDNs have quite a uniform degree distribution, and regardless of the number of nodes that characterize the system, its distribution presents a range of low nodal degree values, which prevents the identification of a reliable curve to classify such networks (*Giustolisi et al.*, 2017; *Yadzany and Jeffrey*, 2011).

We took a first step towards solving this problem by introducing the neighborhood nodal degree (*Giustolisi et al*., 2017). We demonstrated that WDNs network connectivity structure generally follows a Poisson-like distribution, typical of random networks, with low vulnerability with respect to any kind of failure event. However, the classification findings only referred to the network topology, neglecting the hydraulic component, i.e., the different role of nodes in the analysis. Each node, including the sources, had the same intrinsic relevance.

To go one step further in the classification of WDNs, our future perspectives concern the possibility of using the Relevance-embedding degree, instead of the standard degree and of the neighborhood nodal degree, as metric for the WDNs classification. In fact, information about the intrinsic relevance of nodes can enhance the WDNs classification because while from a topological point of view they are very homogeneous systems, from a hydraulic point of view there is a wide heterogeneity that arises from the fact that each node has a specific role in the network (i.e., source nodes, nodes feeding strategic buildings, demand nodes, etc.), and, therefore, a specific intrinsic relevance, as we deepened in this thesis.

Again, using the Apulian network (Figure 34), we report a simple example to show how the values obtained by the extended degree (Figure 87 -right) are very different and at the same time much more realistic than those obtained using the standard metric (Figure 87-left). In particular:

- the final values of the metric reflect the high hydraulic heterogeneity. This means having a range of values large enough to guarantee the definition of a reliable curve to classify such networks;
- increasing the intrinsic relevance of the reservoir (source nodes in general), its Relevance-embedding degree increases. This way, the reservoir becomes a hub, and the importance of the nodes close to it increases;

- associating information on intrinsic relevance and connective structure generates hubs and leads to networks more and more scale free.

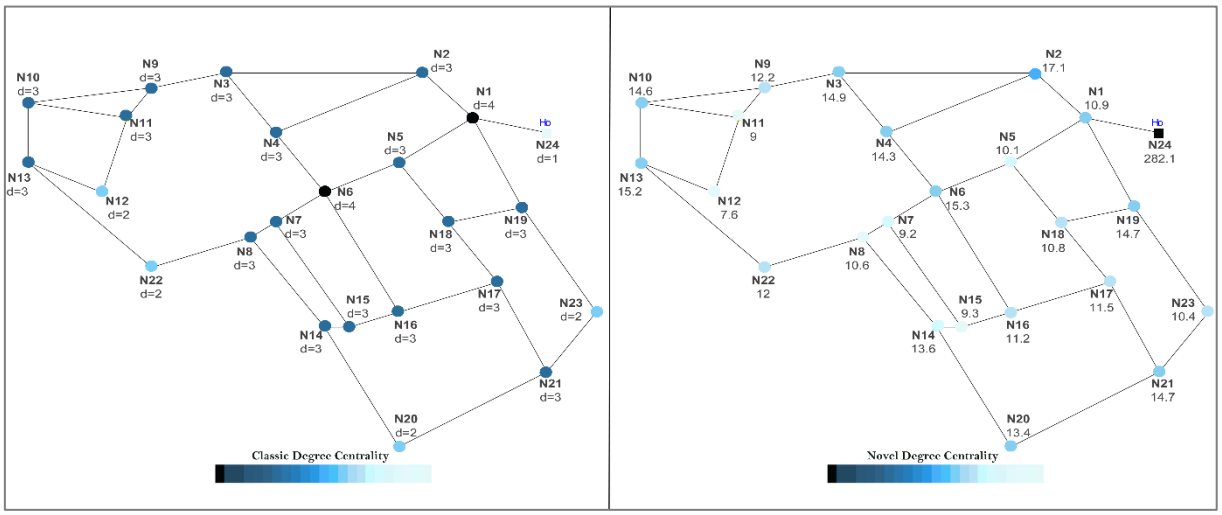


Figure 87. Classic (left) and extended degree (right) for Apulian network. In the left panel the most part of the nodes (17/24) has the same degree value (d=3), five nodes have degree 2, two nodes have degree 4 and only the reservoir has a degree equal to 1. The network is highly homogeneous with a scarcity of values (4) for the identification of a reliable classification curve. Furthermore, the most important node is not the one corresponding to the reservoir (N24) but nodes N6 and N1, due to their greater number of connections. In the right panel we used the Relevance-embedding degree, using as intrinsic relevance the demand node for each demand nodes and the sum of the demands for the reservoir. Demands are reported close to each node. The network is less homogeneous in the degree of connection and each node assumes a different degree value. Furthermore, the most important node is now the reservoir, i.e., the hub, consistently with the hydraulic engineering knowledge.

6

Conclusions

Complex Network Theory (CNT) is a very large and attractive world both from a historical and scientific point of view. Many events over time characterized and embellished this theory laying the foundations for the development and use of tools that today may seem to be immediately applicable. Each event represented an important step forward in discovering complex systems that surround us. In fact, the analyses that such theory allows are often exhaustive in determining the behaviour of a large percentage of complex systems.

The choice of resorting to the CNT for the WDNs analysis was linked to the need to find increasingly sophisticated tools to solve new and critical tasks. However, after various studies, we realized that classic CNT tools do not represent a usable tool for WDNs. For several years, many researchers employed CNT tools unconditionally and passively to WDNs, as simple applications, with too high expectations for such a new sector. Generally, they did not consider the peculiarities of these infrastructural systems. The analyses were unrealistic and the results, which evidently differed from their actual behaviour, not very encouraging.

There were beliefs that confronted us with a twofold problem:

- it was common to think that CNT was characterized by tools unconditionally applicable to all real complex systems to know their behaviour and obtain reliable analyses;
- it was common to think that WDNs were well-known systems, having been the subject of study for decades, and that nothing else about their characteristics and their limitations could be useful before introducing them into the world of the CNT.

In reality, (i) the CNT tools had always given excellent results as they were applied to systems (e.g., immaterial networks) that did not have the limits of infrastructural ones, such as WDNs, and (ii) knowledge of WDNs should not be taken for granted, since they represent extremely current systems, with a dynamic behaviour that vary their characteristics, performances and needs, both in space and in time. Hence the importance of the spatial as well as the temporal component in the analysis of WDNs.

Accordingly, in this thesis, with the aim to contribute to the study of Water Distribution Networks (WDNs) using CNT tools, we proposed a tailoring process that involved the most suitable CNT metrics for WDNs also considering the peculiarities of these hydraulic systems.

Starting from previous research that privileged the node-based approach, we proposed a change of perspective: pipes play a fundamental role in the study of such systems. Later, we constructed a tailored topology, embedded the intrinsic relevance of nodes in the analyses, captured the emergent hydraulic domain of these systems and proposed the optimal division in segments for planning and management purposes.

The first three chapters of the thesis are dedicated to the presentation of the basic tools of the CNT, to their inapplicability to WDNs and to the need to create a new framework that tailors them accordingly to the WDNs peculiarities. We initially clarified the usefulness/ relevance that the CNT tools have in the study and analysis of complex systems, highlighting, at the same time, their limits when applied to infrastructural networks, such as WDNs. In the second chapter we recalled the birth of the CNT and its basic concepts. In the third chapter we (i) proposed a review of the previous applications of the classic CNT tools to WDNs, (ii) explained the concepts of hydraulic model and hydraulic domain and (iii) highlighted the importance of spatial and temporal components for these systems. Finally, we highlighted the need of a tailoring after applying the classic CNT metrics to WDNs.

In the fourth chapter we concretized the need to adapt CNT metrics for WDNs presenting two strategies of tailoring referred to the most suitable centrality metrics for spatial network (betweenness, closeness, harmonic closeness, and degree). Both strategies arose from the need to develop a tool capable of identifying the emerging hydraulic behavior of the WDNs, that is, the domain of the system. The first strategy based on the WDN-tailored edge betweenness, when developed, showed to respond very well to this need, managing to detect the main paths/ pipes in the system, as identified by the flow, with high correlation. The second strategy based on Relevance-embedding centralities, considering the information about the intrinsic relevance of nodes into the analysis represents a real added value to the study of all complex real systems. The identification of the main paths/ pipes in the network is much more reliable, the behavior of the new topological metric almost mirrors that of the hydraulic metric, with a very high correlation. The comparison between the two metrics by analyzing six real WDNs showed that the Relevance-embedding edge betweenness better identifies the emergent hydraulic behaviour of WDNs compared to WDN-tailored edge betweenness.

Finally, we briefly recalled the tailoring process adopted for the modularity index. We used the tailored metric to perform the optimal DMA design strategy of WDNs and extended it to define the Sampling-oriented modularity index in order to perform the optimal Sampling design strategy aimed at the optimal location of pressure meters in WDNs.

We think that this thesis represents only the first step towards the study of WDNs using tools borrowed from the CNT, as the sector is still very young and with a long and prosperous future ahead. At the same time, precisely for what has been said, carrying out this work has been extremely useful, training and interesting because (i) it was possible to clarify some points that have remained obscure until now on the application of the CNT to the WDNs; (ii) it was

possible to know in an ever more in-depth way the WDNs behavior; (iii) it was possible to put the foundations for a new framework and to develop new ad-hoc strategies and metrics for WDNs analysis starting from a framework full of possibilities and tools. Furthermore, we highlighted the importance of topology in the study of WDNs and how the real behavior of such systems is the result of the interplay between network topological structure and intrinsic relevance of nodes.

Publications on International Journals

1. GIUSTOLISI, O., RIDOLFI, L., **SIMONE, A**. (2020). Embedding the intrinsic relevance of vertices in network analysis: the case of centrality metrics. *Scientific Reports* 10, 3297. (https://doi.org/10.1038/s41598-020-60151-x).

2. **SIMONE A.**, CILIBERTI F. G., LAUCELLI D. B., BERARDI L., GIUSTOLISI O., (2020). Edge Betweenness for Water Distribution Networks domain analysis, *Journal of Hydroinformatics*, 22 (1), 121–131. (https://doi.org/10.2166/hydro.2019.030).

3. GIUSTOLISI, O., RIDOLFI, L., **SIMONE, A.**, (2019). Tailoring Centrality Metrics for Water Distribution Networks, *Water Resources research*, 55(3), 2348–2369. (https://doi.org/10.1029/2018WR023966).

4. LAUCELLI D.B., BERARDI L., **SIMONE A.**, GIUSTOLISI O., (2018). Towards serious gaming for water distribution networks sizing: A teaching experiment, *Journal of Hydroinformatics*, 21(2), 207-222. (https://doi.org/10.2166/hydro.2018.038).

5. BERARDI L., **SIMONE A**., LAUCELLI D.B., UGARELLI R.M., GIUSTOLISI O., (2017). Relevance of Hydraulic Modeling in Planning and Operating Real-Time Pressure Control: Case of Oppegård Municipality, *Journal of Hydroinformatics*, 20(3),535–550. (https://doi.org/10.2166/hydro.2017.052).

6. GIUSTOLISI, O., **SIMONE, A.**, RIDOLFI, L., (2017). Network structure classification and features of water distribution systems, *Water Resources Research*, 53(4), 3407-3423. (https://doi.org/10.1002/2016WR020071).

7. GIUSTOLISI O., UGARELLI R. M., BERARDI L., LAUCELLI D.B., **SIMONE A.**, (2017b), Strategies for the Electric Regulation of Pressure Control Valves, *Journal of Hydroinformatics*, 19(4). (https://doi.org/10.2166/hydro.2017.101).

8. MAZZOLANI G., BERARDI L., LAUCELLI, D., **SIMONE A.**, MARTINO R. GIUSTOLISI O., (2017). Estimating Leakages in Water Distribution Networks Based Only on Inlet Flow Data, *Journal of Water Resources Planning and Management*, 143(6). (https://doi.org/10.1061/(ASCE)WR.1943-5452.0000758).

9. LAUCELLI, D., **SIMONE A.**, BERARDI L., GIUSTOLISI O., (2017). Optimal Design of District Metering Areas for the Reduction of Leakages, *Journal of Water Resources Planning and Management*, 143(6). (https://doi.org/10.1061/(ASCE)WR.1943-5452.0000768).

10. **SIMONE A.**, GIUSTOLISI O., LAUCELLI D.B., (2016). A proposal of optimal sampling design using a modularity strategy, *Water Resources Research*, 52(8), 6171–6185. (https://doi.org/10.1002/2016WR018944).

International Conferences

1. LAUCELLI D. B., DI SPIRIDIONE S., BERARDI L., **SIMONE A.**, CILIBERTI F., GIUSTOLISI O., (2019) "Supporting Design of Combined Energy Recovery and Pressure Control in a Water Distribution System," 2019 International Conference on ENERGY and ENVIRONMENT (CIEM), Timisoara, Romania, 2019, 279-283.

2. LAUCELLI D. B., BERARDI L., **SIMONE A.**, CILIBERTI F. G., BALACCO G., GIUSTOLISI O., (2019) "Advanced Hydraulic Analysis for Energy Assessment in a Real Water Distribution Network," 2019 International Conference on ENERGY and ENVIRONMENT (CIEM), Timisoara, Romania, 2019, 284-288.

3. **SIMONE A**., CILIBERTI F.G., BERARDI L., LAUCELLI D.B., GIUSTOLISI O., (2019) WDN-Tailored Edge Betweenness for Analysis of District Metering Areas Planning. E-proceedings of the 38th IAHR World Congress September 1-6, 2019, Panama City, Panama. 6249-6255.

4. **SIMONE A.**, RIDOLFI L., LAUCELLI D.B., BERARDI L., GIUSTOLISI O., (2018) "Centrality metrics for Water Distribution Networks" in 13° International Conference HIC 2018 (Hydroinformatics Conference) - Palermo, Luglio 2018 - EPiC Series in Engineering, 3, 1979–1988.

5. **SIMONE A.**, RIDOLFI L., LAUCELLI D.B., BERARDI L., GIUSTOLISI O., (2018) "Complex Network Theory for Water Distribution Networks analysis" in 13° International Conference HIC 2018 (Hydroinformatics Conference) - Palermo, Luglio 2018 - EPiC Series in Engineering, 3, 1971–1978.

6. LAUCELLI D.B., BERARDI L., **SIMONE A.**, GIUSTOLISI O., (2018) "A teaching experiment using a serious game for WDNs sizing" in 13° International Conference HIC 2018 (Hydroinformatics Conference) - Palermo, Luglio 2018, EPiC Series in Engineering, 3, 1104–1112.

7. SPAGNUOLO S., PERRONE G., BERARDI L., LAUCELLI D. B., **SIMONE, A.**, GIUSTOLISI O., (2018) "A practical application of WDNETXL system to DMA Design of Apulian networks, 1st International WDSA/CCWI 2018 Joint Conference, Kingston, Ontario, Canada.

8. PERRONE G., PALMA F., LAUCELLI D. B., BERARDI L., **SIMONE, A.**, GIUSTOLISI O., (2018) "The relevance of enhanced hydraulic modelling for asset management and related performance indicators" - 1st International WDSA/CCWI 2018 Joint Conference, Kingston, Ontario, Canada.

9. BERARDI L., LAUCELLI D.B., **SIMONE A.**, PERRONE G.C., GIUSTOLISI O., (2018) "Integrated pressure control strategies for sustainable management of water distribution networks" - EENVIRO Conference, Cluj Napoca, Romania.

10. LAUCELLI, D. B., BERARDI, L., **SIMONE, A.**, RASPATI, G., UGARELLI, R. M., GIUSTOLISI, O., (2017) "Hydraulic reliability analysis of a real network with remotely real-

time controlled pressure control valves" In: "Computing and Control for the Water Industry (CCWI 2017)" – paper F67.

11. BERARDI, L., LAUCELLI, D. B., **SIMONE, A.**, RASPATI, G., UGARELLI, R.M., GIUSTOLISI, O., (2017) "Mechanical reliability analysis of a real network to support the design of Isolation Valve System" In: "Computing and Control for the Water Industry (CCWI 2017)" – paper F68.

12. BERARDI L., LAUCELLI D.B., **SIMONE A.**, GIUSTOLISI O., (2017) "Optimal pump scheduling strategies accounting for background leakages and energy cost" in International Conference on ENERGY and ENVIRONMENT (CIEM) October 2017.

13. **SIMONE A**., LAUCELLI D.B., BERARDI L., GIUSTOLISI O., (2017) "Modularity index for optimal sensor placement in WDNs" (2017) in SimHydro 2017: Choosing the right model in applied hydraulics 433–447 in: Gourbesville P., Cunge J., Caignaert G. (eds) Advances in Hydroinformatics. Springer Water.

14. LAUCELLI D.B., **SIMONE A.**, BERARDI L., UGARELLI R.M., GIUSTOLISI O., (2017) "Optimal DMA design based on the WDN-oriented modularity index: a real case application" in Leading Edge Sustainable Asset Management of Water and Wastewater Infrastructure Conference (LESAM), Trondheim.

15. **SIMONE A**., BERARDI L., LAUCELLI D. B., GIUSTOLISI O. (2016) "A proposal of optimal sampling design using infrastructure modularity" in XVIII International Conference on Water Distribution Systems, WDSA2016, Procedia Engineering 186, 559 – 566.

16. BERARDI L., LAUCELLI D.B., **SIMONE A**., MAZZOLANI G., GIUSTOLISI O., (2016) "Active leakage control with WDNetXL" in 12th International Conference on Hydroinformatics, HIC 2016 Procedia Engineering, 154 (2016) 62 – 70.

17. LAUCELLI D.B., **SIMONE A.**, BERARDI L., GIUSTOLISI O., (2016) "Optimal Design of District Metering Areas" in International Conference on Efficient & Sustainable Water Systems Management toward Worth Living Development, 2nd EWaS 2016. Procedia Engineering 162, 403 – 410.

18. LAUCELLI D., BERARDI L., UGARELLI R., **SIMONE A.**, GIUSTOLISI O., (2016) "Supporting real-time pressure control in Oppegård municipality with WDNetXL" in 12th International Conference on Hydroinformatics, HIC 2016 Procedia Engineering, 154, 71 – 79.

19. BERARDI L., **SIMONE A.**, LAUCELLI D.B., GIUSTOLISI O., (2016) "Feasibility of mass balance approach to Water Distribution Network model calibration" in XVIII International Conference on Water Distribution Systems, WDSA 2016, Procedia Engineering 186, 551 – 558.

20. MAZZOLANI G., BERARDI L., LAUCELLI D.B., MARTINO R., **SIMONE A.**, GIUSTOLISI O., (2016) "A methodology to estimate leakages in water distribution networks based on inlet flow data analysis" in International Conference on Efficient & Sustainable Water Systems Management toward Worth Living Development, 2nd EWaS 2016. Procedia Engineering 162, 411 – 418.

21. DI NARDO A., DI NATALE M., GIUDICIANNI C., MUSMARRA D., RODRIGUEZ VARELA J.M., SANTONASTASO G.F., **SIMONE A.**, TZATCHKOV V. (2016) "Redundancy Features of Water Distribution Systems" in 18th Conference on Water Distribution System Analysis, WDSA 2016, Procedia Engineering 186, 412 – 419.
22. DI NARDO A., DI NATALE M., GIUDICIANNI C., MUSMARRA D., SANTONASTASO G.F., **SIMONE A.**, (2015) "Water distribution system clustering and partitioning based on social network algorithms" in 13th International Conference on Computing and Control for the Water Industry (CCWI2015) Procedia Engineering 119, 196 – 205.

Posters

1. BERARDI L., LAUCELLI D.B., **SIMONE A.**, GIUSTOLISI O., (2018) "Investigation of DMA consumption by visibility algorithms", 13° International Conference HIC 2018 (Hydroinformatics Conference) - Palermo, Luglio 2018.

Bibliography

1. Ackley, J.R.L., Tanyimboh, T.T., Tahar, B. Templeman, A.B. (2001). Head-driven analysis of water distribution systems, Ulanicki, B.(ed.) Proc. of Computer and Control in Water Industry, *Water software systems: theory and applications*, Research Studies Press, UK, 183-192.

2. Albert R., Jeong H., Barabasi A-L., (2000). Error and attack tolerance of complex networks, *Nature*, 406(6794), 378-382.

3. Albert, R., A.L. Barabasi (2002). Statistical mechanics of complex networks, *Reviews of Modern Physics*, 74(1), 47–97.

4. Alvarez-Socorro, A.J., Herrera-Almarza, G.C., González-Díaz, L.A., (2015). Eigencentrality based on dissimilarity measures reveals central nodes in complex networks, *Scientific Reports*, 5, 17095.

5. Alvisi, S., Franchini M., (2014). A heuristic procedure for the automatic creation of district metered areas in water distribution systems, *Urban Water Journal*, 11(2), 137–159.

6. Barabási, A.-L., Albert, R., (1999). Emergence of scaling in random networks, *Science*, 286, 509-11.

7. Barthélemy, M., (2011). Spatial networks. *Physics Reports*, 499(1-3), 1-101.

8. Barthélemy, M., (2018). *Morphogenesis of Spatial Networks*. Springer International Publishing AG, Springer, Cham.

9. Berardi L., Simone A., Laucelli D.B., Ugarelli R.M., Giustolisi O., (2017). Relevance of Hydraulic Modeling in Planning and Operating Real-Time Pressure Control: Case of Oppegård Municipality, *Journal of Hydroinformatics*, 20(3),535–550.

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

11. Bhave, P.R., (1981). Node flow analysis of water distribution systems, *Transportation Engineering Journal*, 107(4), 457-467.

12. Bianconi, G., Barabási, A-L. (2001). Competition and multiscaling in evolving networks, *Europhysics Letters*, 54 (4), 436–442.

13. Bonacich, P., (1987). Power and centrality: a family of measures, *American Journal of Sociology*, 92(5), 1170-1182.

14. Borgatti, S. P., (2005). Centrality and network flow, *Social networks*, 27(1), 55-71.

15. Buhl, J., Gautrais, J., Reeves, N., Sole, R. V., Valverde, S., Kuntz, P., Theraulaz, G., (2006) Topological patterns in street networks of self-organized urban settlements, *The European Physical Journal B*, 49(4), 513-522.

16. Cauchy A. L. (1813). Recherche sur les polyèdres. *Journal de l'École Polytechnique*, 9(16), 68–86.

17. Chapela, V., Criado, R., Moral, S., Romance, M., (2015). *Intentional Risk Management through Complex Networks Analysis*, SpringerBriefs in Optimization.

18. Davidson, J., Bouchart, F., Cavill, S., Jowitt, P., (2005). Real-time connectivity modelling of water distribution networks to predict contamination spread, *Journal of Computing in Civil Engineering*, 19 (4), 377–386.

19. Deuerlein, J. W., (2008). Decomposition Model of a General Water Supply Network Graph, *Journal of Hydraulic Engineering*, 134(6), 822-832.

20. Diao, K., Zhou, Y., Rauch W., (2013). Automated Creation of District Metered Area Boundaries in Water Distribution Systems. *Journal of Water Resources Planning and Management*, 139(2), 184-190.

21. Dijkstra, E., (1959). A Note on Two Problems in Connexion with Graphs, *Numerische Mathematlk*, 1(I), 269-271.

22. Easley, D., Kleinberg, J., (2010). *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*, Cambridge University Press.

23. Erdös P., Rényi A., (1959). On random graphs, *Publicationes Mathematicae*, 6, 290–297.

24. Erdös P., Rényi P., (1960). On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci.* 5, 17–60.

25. Euler, Leonhard (1736). *Solutio problematis ad geometriam situs pertinentis*, Typis Academiae, St. Petersburg.

26. Everett, M. G., Borgatti, S. P., (1998). Analyzing Clique Overlap, *Connections*, 21(1), 49-61.

27. Fortunato, S., Barthélemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences of the United States of America*, 104(1), 36–41.

28. Fortunato, S. (2010), Community detection in graphs, *Physics Reports*, 486(3–5), 75–174.

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

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

31. Germanopoulos, G., (1985). A technical note on the inclusion of pressure dependent demand and leakage terms in water supply network models, *Civil Engineering Systems*, 2(3), 171-179.

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

33. Giudicianni, C., Di Nardo, A., Di Natale, M., Greco, R., Santonastaso, G.F., Scala, A., (2018). Topological Taxonomy of Water Distribution Networks, *Water*, 10(4), 444.

34. Giustolisi, O., Ridolfi, L., Simone, A. (2020). Embedding the intrinsic relevance of vertices in network analysis: the case of centrality metrics. *Scientific Reports*, 10, 3297.

35. Giustolisi, O., Ridolfi, L., Simone, A., (2019). Tailoring Centrality Metrics for Water Distribution Networks, *Water Resources research*, 55(3), 2348–2369.

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

37. Giustolisi O., Ugarelli R. M., Berardi L., Laucelli D.B., Simone A. (2017b). Strategies for the Electric Regulation of Pressure Control Valves, *Journal of Hydroinformatics*, 19 (5), 621–639.

38. Giustolisi, O., Savic, D.A., Kapelan, Z. (2008). Pressure-driven demand and leakage simulation for water distribution networks, *Journal of Hydraulic Engineering*, 134 (5), 626-635.

39. Giustolisi, O., Kapelan, Z., Savic, D.A. (2008b). An algorithm for automatic detection of topological changes in water distribution networks, Journal of Hydraulic Engineering, 134 (4), 435-446.

40. Giustolisi, O., and Walski, TM., (2012). Demand components in water distribution network analysis, *Journal of Water Resources Planning and Management*, 138(4), 356-367.

41. Giustolisi, O., and Ridolfi, L. (2014a). A new modularity-based approach to segmentation of water distribution network, *Journal of Hydraulic Engineering*, 140(10), 04014049.

42. Giustolisi, O., and Ridolfi, L. (2014b). A novel infrastructure modularity index for the segmentation of water distribution networks, *Water Resources Research.*, 50, 7648–7661.

43. Gupta, R., Bhave, P.R., (1996). Comparison of methods for predicting deficient network performance, *Journal of Water Resources Planning and Management*, 122(3), 214-217.

44. Gupta, R., Prasad, T. D., (2000). Extended Use of Linear Graph Theory for Analysis of Pipe Networks, *Journal of Hydraulic Engineering*, 126(1), 56-62.

45. Gutiérrez-Pérez J. A., Herrera M., Pérez-García R., Ramos-Martínez E., (2013). Application of graph-spectral methods in the vulnerability assessment of water supply networks, *Mathematical and Computer Modelling*, 57(7), 1853–1859.

46. Haggett P., Chorley R. J., (1969). *Network analysis in geography*, London, Edward Arnold.

47. Hawick K.A., (2012). Water Distribution Network Robustness and Fragmentation using Graph Metrics, *Proceedings of the IASTED African Conference on Water Resource Management*, (AfricaWRM 2012), 304-310.
48. Jacobs, P., Goulter, I.C., (1988). Evaluation of methods for decomposition of water distribution networks for reliability analysis, *Civil Engineering Systems*, 5(2), 58-66.

49. Jacobs, P., Goulter, I. C., (1989). Optimization of Redundancy inWater Distribution Networks Using Graph Theoretic Principles, *Engineering Optimization*, 15(1), 71-82.

50. Jeong, H., Mason, S., P., Barabási A., Oltvai, Z., N., (2001). Lethality and centrality in protein networks, *Nature*, 411, 41–42.

51. Kansky K. J., (1963). *Structure of transportation networks: relationships between network geometry and regional characteristics,* University of Chicago, Department of Geography.

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

53. Kesavan, H. K., Chandrashekar, M., (1972). Graph theoretic model for pipe network analysis, *Journal of Hydraulic Engineering*, 98(2), 345-364.

54. Kessler A., Ostfeld A., Sinai G., (1998). Detecting Accidental Contaminations in Municipal Water Networks, *Journal of Water Resources Planning and Management*, 124(4).

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

56. Latora, V., Nicosia, V., Russo, G., (2017). *Complex Networks Principles, Methods and Applications*, Cambridge University Press.

57. Laucelli, D., Simone A., Berardi L., Giustolisi O., (2017). Optimal Design of District Metering Areas for the Reduction of Leakages, *Journal of Water Resources Planning and Management*, 143(6).

58. L'Huillier, S.-A.-J., (1861). *Mémoire sur la polyèdrométrie*, Annales de mathématiques pures et appliquées, Tome 3, pp. 169-189.

59. Mays, L.W., (2000). *Water Distribution Systems Handbook*, New York: McGraw-Hill.

60. Mazzolani G., Berardi L., Laucelli, D., Simone A., Martino R. Giustolisi O., (2017). Estimating Leakages in Water Distribution Networks Based Only on Inlet Flow Data, *Journal of Water Resources Planning and Management*, 143(6).

61. Milgram S., (1967). The Small World Problem, *Psychology today*, 60-67.

62. Moody, J., (2001). Race, School Integration, and Friendship Segregation in America. *American Journal of Sociology*, 107(3).

63. Nazempour, R., Monfared, M.A.S., Zio, E., (2018). A complex network theory approach for optimizing contamination warning sensor location in water distribution networks. *International Journal of Disaster Risk Reduction*, 30(B), 225-234.

64. Newman M.E., (2002). Assortative mixing in networks. Physical Review Letters, 89, 208701.

65. Newman, M.E.J., Girvan, M., (2004). Finding and evaluating community structure in networks, Physical Review E, 69, 026113.

66. Newman M., Barabási A-L., Watts D. J., (2006). *The Structure and Dynamics of Networks*, Princeton University Press, Princeton.

67. Newman, M.E.J., (2004). Fast algorithm for detecting community structure in networks, *Physical Review E*, 69, 066133

68. Newman, M., (2010). *Networks: An Introduction*. Oxford, UK: Oxford University Press.

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

70. Norcini Pala, L., (2020). *Social... Mente. Come si formano le idee e l'opinione pubblica, tra rete e social*. San Paolo Edizioni.

71. Ostfeld A., Shamir U., (1996). Design of Optimal Reliable Multi-Quality Water-Supply Systems, *Journal of Water Resources Planning and Management*, 122(5), 322-333.

72. Padgett, J. F., Ansell C.K., (1993). Robust Action and the Rise of the Medici, 1400-1434, *The American Journal of Sociology*, 98(6), 1259–1319.

73. Page, L., Brin, S., Motwani, R., Winograd, T., (1998). The PageRank Citation Ranking: Bringing Order to the Web, *WWW 1999*.

74. Pelino, V., (2013). *Cyber intelligence: dark networks, in book Cyberworld capire, proteggersi e prevenire gli attacchi in rete*, HOEPLI.

75. Perelman, L., Ostfeld, A., (2011). Topological clustering for water distribution systems analysis, *Environmental Modelling & Software*, 26(7), 969–972.

76. Piraveenan M.R., (2010). *Topological analysis of complex networks using assortativity*, School of Information technologies, University of Sydney.

77. Rochat, Y., (2009). Closeness centrality extended to unconnected graphs: The harmonic centrality index, *Proceedings of the Conference ASNA*, Zürich.

78. Sabidussi G., (1966). The centrality index of a graph, *Psychometrika*, 31(4), 581-603.

79. Scibetta, M., Boano, F., Revelli, R., Ridolfi, L., (2013). Community detection as a tool for complex pipe network clustering, *Europhysics Letters*, 103(4), 48001.

80. Sheng N., Jia Y. W., Xu Z., Ho S. L., Kan C. W., (2013). A complex network based model for detecting isolated communities in water distribution networks, *Chaos*, 23(4), 043102.

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

82. Simone A., Ciliberti F. G., Laucelli D. B., Berardi L., Giustolisi O., (2020). Edge Betweenness for Water Distribution Networks domain analysis, *Journal of Hydroinformatics*, 22(1), 121–131.

83. Simone A., Ridolfi L., Laucelli D.B., Berardi L., Giustolisi O., (2018) Complex Network Theory for Water Distribution Networks analysis in *13° International Conference HIC 2018* (Hydroinformatics Conference) - Palermo- *EPiC Series in Engineering*, 3(2018), 1971–1978.

84. Simone A., Giustolisi O., Laucelli D.B., (2016). A proposal of optimal sampling design using a modularity strategy, *Water Resources Research*, 52(8), 6171–6185.

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

86. Steinhaeuser, K., Chawla, N.V., (2010). Identifying and evaluating community structure in complex networks, *Pattern Recognition Letters archive*, 31(5), 413-421.

87. Taaffe, E., Morrill, R., Gould, P., (1963). Transport Expansion in Underdeveloped Countries: A Comparative Analysis, *Geographical Review*, 53(4), 503-529.

88. Todini, E., Pilati, S., (1988). A Gradient Algorithm for the Analysis of Pipe Networks. *Computer Applications in Water Supply* 1 (System Analysis and Simulation), John Wiley & Sons, London, 1, 1-20.

89. Todini, E., (2003). A more realistic approach to the "extended period simulation" of water distribution networks, *Advances in Water Supply Management*, Maksimovic C., Butler D. and Memon F.A. (eds), CRC Press.

90. Torres, J.M., Duenas-Osorio, L., Li, Q., Yazdani A., (2016). Exploring Topological Effects on Water Distribution System Performance Using Graph Theory and Statistical Models, *Journal of Water Resources Planning and Management* ,143 (1) 04016068.

91. Travers J., Milgram S., (1969). An experimental study of the Small World Problem, *Sociometry*, 32(4), 425-443.

92. Van den Heuvel, M. P., Sporns, O., (2011). Rich-club organization of the human connectome. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 31(44), 15775–15786.

93. Van Mieghem. P., (2014). *Performance Analysis of Complex Networks and Systems*. Cambridge University Press.

94. Wagner, J.M., Shamir, U., Marks, D.H., (1988). Water distribution reliability: simulation methods, *Journal of Water Resources Planning and Management* ,114(3), 276-294.

95. Walski, T. M., (1983). Technique for calibrating network models, *Journal of Water Resources Planning and Management*, 109(4), 360–372.

96. Watts D.J., Strogatz D.H., (1998). Collective dynamics of small-world networks, *Nature*, 393-440.

97. Xiao, F.W., Guanrong, C., (2003). Complex networks: small-world, scale-free and beyond, *IEEE Circuits and Systems Magazine*, 3(1), 6-20.

98. Xu, X., Zhang, J., Sun, J., Smal,l M., (2009). Revising the simple measures of assortativity in complex networks, *Physical Review E*, 80(5), 056106.

99. Yang, S.-L., Hsu N.-S., Loule P.W.F., Yeh W.W.-G., (1996). Water distribution network reliability: connectivity analysis, *Journal of Infrastructure Systems*, 2(2), 54-64.

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

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

102. Yazdani, A., Jeffrey, P., (2012a). Robustness and Vulnerability Analysis of Water Distribution Networks Using Graph Theoretic and Complex Network Principles, *12° Annual Conference on Water Distribution Systems Analysis* (WDSA).

103. Yazdani, A., Jeffrey, P., (2012b). Applying Network Theory to Quantify the Redundancy and Structural Robustness of Water Distribution Systems, *Journal of Water Resources Planning and Management*, 138(2).

104. Zhou, S., (2004). *Parameterising and Modelling the Internet Topology*, Thesis submitted to the Mary Queen University of London for the degree of Doctor of Philosophy.