# Analysis of the Segment Graph of Water Distribution Networks

Tamás Huzsvár[1*], Richárd Wéber[1], Csaba Hős[1]

[1] Department of Hydrodynamic Systems, Faculty of Mechanical Engineering, Budapest University of Technology and Economics,
  H-1521 Budapest, P. O. B. 91, Hungary
[*] Corresponding author, e-mail: thuzsvar@hds.bme.hu

## Abstract

One of the basic infrastructures of every settlement is the water distribution system, which provides clean and potable water for both private houses, industrial consumers and institution establishments. The operational robustness and vulnerabilities of these networks is an essential issue, both for the quality of life and for the preservation of the environment. Even with frequent and careful maintenance, unintentional pipe bursts might occur, and during the reparation time, the damaged section must be isolated hydraulically from the main body of the water distribution network. Due to the size and complexity of these networks, it might not be trivial how to isolate the burst section, especially if one wishes to minimize the impact on the overall system.  This paper presents an algorithmic method that is capable of creating isolation plans for real-life networks in a computationally efficient way, based on the graph properties of the network. Besides this segmentation plan, the topological behavior of the structural graph properties was analyzed with the help of the complex network theory to create a method for the quantitative topology based categorization of the water distribution networks.

## Keywords

segment graph, isolation plan, structural robustness

## 1 Introduction

Water distribution networks (WDNs) are an essential infrastructure of every settlement that provide clean drinking water for the industrial and residential area. Even in the case of careful maintenance, various failures (e.g. pipe burst) might occur, and the responsible utility company must be capable of creating a hydraulic isolation plan that allows quick separation of the area while minimizing the number of affected consumers, which can be achieved via the closure of isolation valves. Because of the complexity, heterogeneity and size of the WDNs, the optimal location of these isolation valves is far from being trivial. Therefore, the time delay between the identification of the pipe burst and the proper segregation of the area can exceed over hours due to the search of the correct, functioning valves. During these hours many liters of drinking water are wasted, moreover, because of the presence of soil erosion, the damage caused to the property or transport infrastructure (notably roads) can increase drastically. The purpose of this paper is to develop a method to build such isolation plans that result in the smallest isolated segment and hence minimizes the affected area. The proposed technique uses graph properties that are capable of describing the topological type of a WDN without its graphical image or hydraulic simulation.

In this study, we classify WDNs based on their topology as either an urban district or suburban area, see Fig. 1 and Fig. 2. As can be seen, the main difference between them is the redundancy and the number of the loops: an urban-like network such as the one depicted in Fig. 1 is grid-like, the majority of the pipes are short, and the connectivity between the nodes is high. Such networks are redundant, i.e. if a pipe segment (edge) is closed, there are still several alternative routes to reach the nodes. On the contrary, suburban networks are mostly linear in the sense that the connectivity between the nodes is small and often, by closing an edge, some part the system becomes isolated from the water source. Purely urban or suburban WDSs are rare, and we wish to develop measures to quantify the topology of real-life networks. We will use 10 real-life networks from Sopron Waterworks (Hungary) with increasing complexity and size, whose properties are listed in Table 1.

The rest of this paper is organized as follows. Section 2 begins with the theoretical toolbox of the research and also summarizes how other researchers addressed the issue. The third section is concerned with the methodology used for the creation of the segment graphs. The fourth and fifth sections present the key findings of the
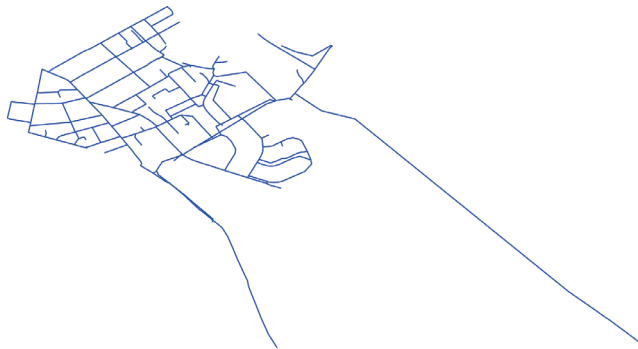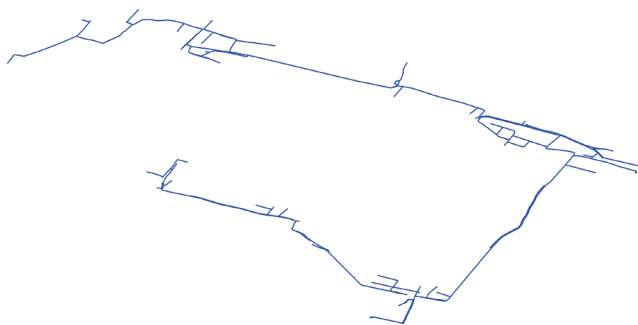
**Fig. 1** Urban district (Network 9)



**Fig. 2** Suburban area (Network 5)

**Table 1** The properties of the investigated networks

| ID | Node No. | Type | Sum of pipe length [m] | Sum of demand [m³/h] |
|---|---|---|---|---|
| 1 | 6450 | Urban | 111920 | 535.5 |
| 2 | 6410 | Suburban | 122651 | 51.95 |
| 3 | 2704 | Urban | 39459 | 50.47 |
| 4 | 2673 | Suburban | 54696 | 311 |
| 5 | 2282 | Suburban | 48335 | 30.57 |
| 6 | 1912 | Suburban | 36393 | 10.38 |
| 7 | 1806 | Suburban | 34592 | 36.06 |
| 8 | 1513 | Urban | 30841 | 78.19 |
| 9 | 1383 | Urban | 27179 | 19.44 |
| 10 | 1201 | Urban | 19551 | 6.153 |

research, Section 4 focuses on the issue of shutdown plans, while Section 5 discusses the findings of the graph analysis. Finally, Section 6 gives a brief summary and discussion of the findings.

**2 Literature survey**

Let us imagine that one closes all the isolation valves in a WDS at once. The smallest connected parts containing pipelines, tanks, pumps, etc. of a hydraulic system are called **segments** [1]. In [2] and [3] has been suggested that ideally, the number and the placement of the isolation valves is such that there are $N$ or $N\text{-}1$ valves around all nodes, where $N$ stands for the number of connecting pipelines at that node (we shall refer to these suggestions as $N$ and $N\text{-}1$ rules). Using this method every single pipe can be isolated individually. However, real WDNs cannot follow these rules due to financial reasons. For finding an optimal set (in the Pareto sense) of solutions in [4] has been published a method where the total cost and the average demand shortfall compete against each other.

WDSs are usually represented by the edges (pipes, valves, pumps, etc.) and nodes, which we will refer to as *link-node representation*. In this paper, however, we will use another representation, where the set of nodes, pipes, pumps, etc. are the nodes and the isolation valves connecting them will be the edges, resulting in a *segment graph*. One of the first algorithms capable of building such a segment graph was presented in [5]. Later in [6] has been developed a method that handled one-way devices as well (e.g. check valves) and demonstrated its capabilities on two real-life WDNs. By improving the Floyd-Warshall algorithm [7] defined a computationally efficient technique that was used in the city of Tianjin containing more than 700k nodes.

There are several research papers trying to capture the robustness of WDNs against pipe failures. In [8] is defined the Global Resilience Analysis (GRA) based on extended hydraulic calculations and demonstrated its use on four artificial networks, assuming that the isolation valves are located according to the $N$-rule (although this is far from the real-life cases) and that the repair time of a pipe failure is three hours (which might also be optimistic). The GRA was also extended to handle excess demands (e.g. firefighting) and contamination. In [9], the authors presented different approaches for placing isolation valves (e.g. $N$, $N\text{-}1$ rule, limited, scarce) and the failure of a single isolation valve was investigated based on the number of isolation elements that are required for segregation, the segment lengths, and the system supply shortfall. The theory was capable of determining the critical valves; moreover, the authors concluded that the critical valves are often associated with key segments, on the other hand they are related to the topology of the network layout. Having said that, the cost of the valve itself and that of the necessary auxiliary labor (earthwork, repavement cost, etc.) was not considered.

In the late 90's a new discipline appeared, namely network theory, where the scientists analyzed the topology of huge networks (e.g. world wide web, human relations, electrical networks) and tried to understand the processes based on the topology (e.g. effect of cyber-attacks, spreading of infections). For example, the effect of a random

failure and direct attack was analyzed in [10] in case of random and "scale-free" networks. A detailed overview of the complex network principles can be found in [11]. In recent years graph theory started to infiltrate into WDN analysis. In [12] has been calculated structural properties and degree distribution of four WDNs (two artificial, two real) and tried to define an objective indicator for robustness and vulnerability. However, it was also mentioned that caution is required since topological quantities can only partially describe a network. In [13] has been made an attempt to quantitatively categorize WDNs based on graph metrics, defined as branch index.

Another approach of recent papers is to find a connection between topological properties and hydraulic metrics in a WDN. [14] found a relationship between the vulnerability of a network and the betweenness centrality. [15] and [16] both created artificial WDNs (but in a completely different way) and searched for a relation between structural values and hydraulic indicators. This connection was used to tune two different regression models to predict hydraulic values (maximum of system unit head loss, average water age, and containment concentration) in [15]. [16] also found relations quantitatively, however, it is important to emphasize that both articles investigated artificially generated WDNs, while real-life cases might differ from those.

Therefore, our aim is to investigate the segment graph of ten real life water distribution system from the perspective of complex network theory to represent the possibilities what this literaturely underrepresented field of study behold for robustness analysis.

## 3 Segment graph

Nowadays, numerous algorithms are available in the literature to create the segment graph of a water distribution system, most of which is based on the $A$ adjacency matrix of the network, that is:

$$A = \begin{cases} 1, \text{if } i\text{-th and } j\text{-th nodes are connected}, \\ 0, \text{ if } i\text{-th and } j\text{-th nodes are not connected}. \end{cases} \quad (1)$$

It is important to notice that the adjacency matrix is an $N \times N$ *sparse* matrix, and in order to keep the computational effort and storage requirement as low as possible, special sparse matrix algorithms and representation should be used. The main novelty of the approach proposed here is that it is based on the list of the edges, that is a 2M long vector containing the indices of the nonzero values of the adjacency matrix, i.e.

$$v = \left[ s_0, e_0, s_1, e_1, \ldots, s_k, e_k, \ldots, s_M, e_M \right] \quad (2)$$

where the k-th edge is connecting the nodes with $s_k$ and $e_k$ indices. In overall, the presented method does not require sophisticated sparse matrix-related software (or programming library) and yet it is advantageous in terms of computational effort.

The sketch of the proposed algorithm can be seen in Fig. 3. The input data are (1) the list of the connected pipelines (in the form of Eq. (2)), (2) the data of the nodes and (3) the position of the isolation valves. Using these the algorithm is defined as follows.

- First, every pipeline with a built-in isolation valve is identified. After that, these valves are closed virtually, which means that the corresponding valve is replaced by two unconnected nodes as it can be seen in Fig. 4 a.
- The virtual valve closure break the network into small groups of pipes and nodes, that are the segments of the water distribution system. In the next step, the algorithm chooses the first pipe in the edge list and searches for all of its connections. Once a connected edge is found, it is removed from the edge list and added to the segment specified edge list. The algorithm restarts the search from the beginning of the modified edge list, but comparing the element of the reduced edge list to the extended segment specified edge list. It means that in the next cycle the connections of the newly found edge are also searched. This algorithm is called breadth-first search [11]. If the loop could not found any new connection, a segment is defined and saved then a new search is started at the beginning of the edge list.
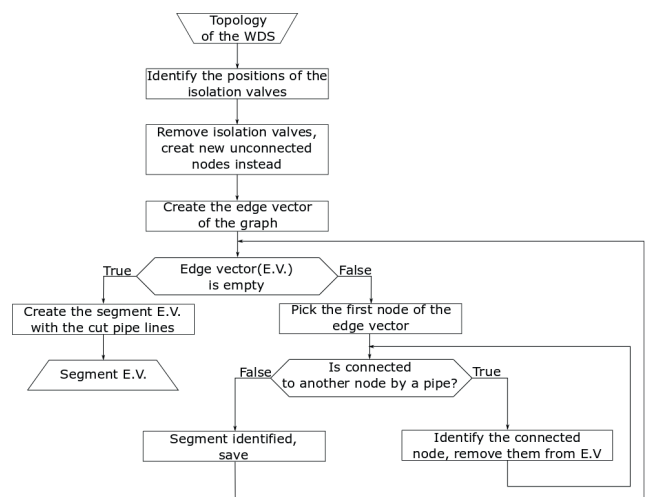


**Fig. 3** The sketch of the segment graph creator algorithm.

- When the size of the edge list is zero (there are no more unvisited edges), it means all of the segments have been identified. The connections between the segments are derived by means of the location of the removed isolation valves. The flowchart segment graph building is explained in Fig. 4 c).

After testing the algorithm on the test networks listed in Table 1, it can be concluded that the number of nodes shows a linear correlation with the run-time of the segmentation algorithm. Moreover, even in case of our largest presented network (that contains more than 6000 nodes), the computation time was below one second while it was running on an average PC.

## 4 Isolation plan

From the practical point of view, the most important outcome of the segment graph analysis is the automatic generation of a shutdown plan, that is the hydraulic segregation of the damaged part of the network for the time of the reconstruction. In the case of a pipe burst, after the location of the damaged pipe was identified by the maintenance staff (i.e. its identifier has been found based on the network plans), the shutdown plan can be determined by searching which segment contains the problematic pipeline and then listing the isolation valves that are connecting to this segment with their exact position.
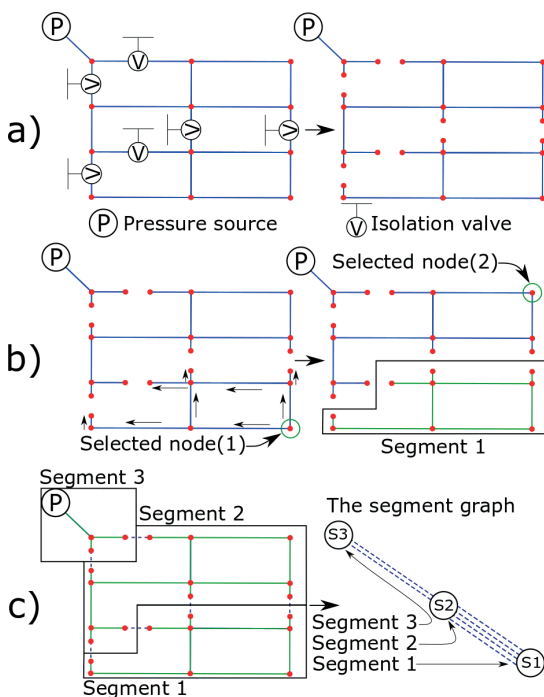


**Fig. 4** The method of the segmentation.

A remarkable advantage of this method is that the segment graph needs to be calculated only once, and through that, the real-time identification of the isolation valves required for the segregation of the damaged area is possible even in case of large networks. On the other hand, as the connected segments are identified, the effect of the segment segregation can easily be predicted; the segments (consumers) which are cut off from the flow source (e.g. pumps, tanks) can be identified easily as well, so the locations of water outage can be forecasted.

To identify the segregated segments, the same method was used as before, but this time the starting node is the source. The difference in the visited paths with/without the segregation gives the segments which are cut off from the source under the reconstruction of the broken pipe. It means that, if the technical terms are available, the maintenance staff not just easily finds the isolation valves needed for the segregation, but they can also send an automated warning about the problem to the consumers located in the cut-off segments.

## 5 Degree distribution

As it was introduced in the literature survey, nowadays many research focuses on the analysis of the water distribution systems from the network theory approach. The presented studies provide evidence that there is a connection between the hydraulic reliability and the topological properties of the network [12, 15]. We are investigating now the degree distribution of the *segment graph*. The degree of a node is the number of the connections between itself and any other node in the network, i.e. in case of the segment graph, it gives how many isolation valves must be closed in order to cut off a segment. For the determination of the degree distribution we used the C package IGRAPH [17].

The results of degree distribution analysis can be seen in Fig. 5. We found that the all investigated networks - except the smallest one - can be characterized with the same degree distribution independently from its size. The average node degree is 2.33 with a standard deviation of 0.1 (excluding the smallest network), based on Fig. 5, in overall $\bar{n}_D = 2.33 \pm 0.1$. On the other hand, it does not follow any specific network degree distribution, e.g. scale-free [11]. However, the comparison of the structural properties (e.g. diameter, cluster coefficient, degree distribution) and considering this graph is necessarily planar, also fully connected i.e. there are no crossing edges without a node, also there are no segregated islands, we can say it is mostly similar to a planar, connected random graph. This means
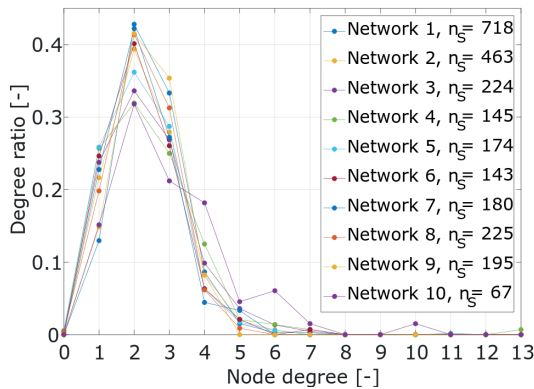
**Fig. 5** The degree distribution of the segment graphs, where $n_S$ is the number of the segments in the network. For the details of the networks, see Table 1.

that the probability of having a node with high degree is negligible, thus the networks are robust against a random pipe failure. In overall, we can say that the investigated networks can be described with the same (but not specific) degree distribution, and based on the structural properties they behave as a planar, connected random graph.

Although the degree distributions are similar, a significant difference can be observed: the number of the segments with degree 1 and with degree 3 or higher. We will refer to the objective quantity as $n_{1/3+}$. This quantity was calculated for every network.

If segments with high connection number dominate the network (e.g. $n_{1/3+}$ = 0.287 in case of network 10) it is a clearly an urban network, and in the opposite case, if the end segments - with 1 degree - dominate the network (e.g. $n_{1/3+}$ = 0.700 in case of network 4) it is a suburban network, see Fig. 6.

$$n_{1/3+} \begin{cases} < 0.45, & \text{if the network is urban-like} \\ \approx 0.45 .. 0.55, & \text{if the network is a transition,} \\ > 0.55, & \text{if the network is suburban-like.} \end{cases} \quad (3)$$

With the help of the $n_{1/3+}$ ratio, it is possible to characterize the ratio between the extension and the density/redundancy of a network. If one has to cope with numerous WDSs (e.g. a regional water utility company serving over a large area) it is possible to use this method that is capable of identifying the weak zones where a pipe burst have a larger impact to the water distribution.

**6 Conclusion**

This paper presented a computationally efficient algorithm, that is capable of building the segment graph of a water distribution network. On one hand, the favorable run time - that grows linearly with the size of the network
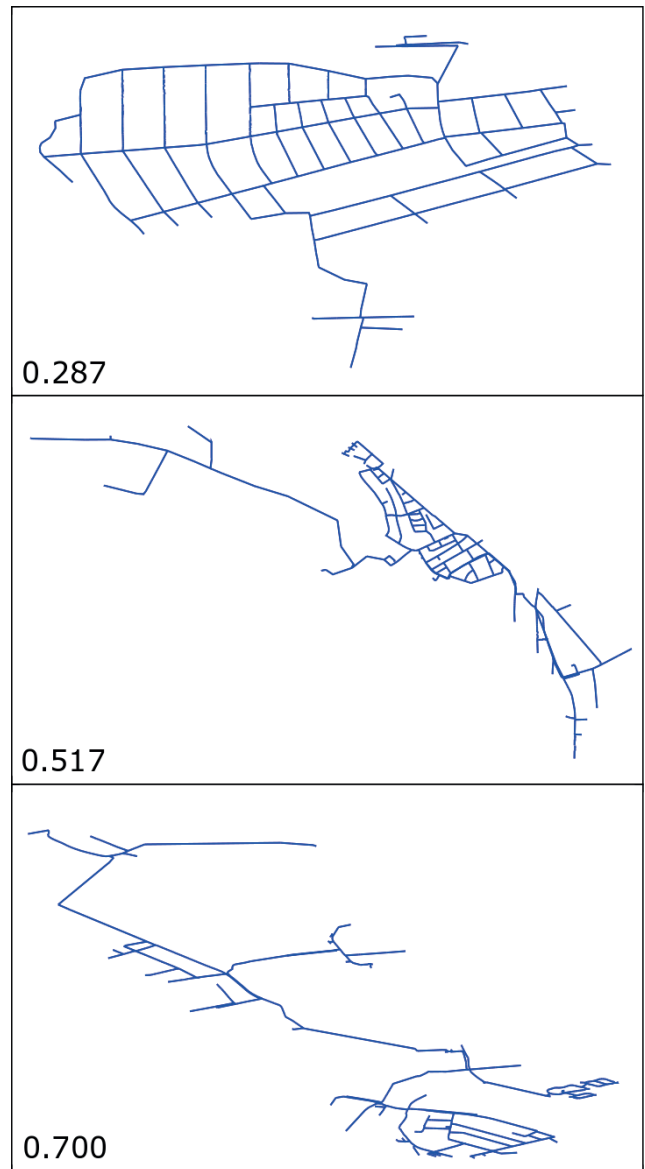


0.287

0.517

0.700

**Fig. 6** Visualizing different type of network topologies (from urban to suburban side) and indicating the $n_{1/3+}$ on Network 10, 8 and 4.

- made it possible to create the isolation plans in real-time for any part of the WDS. On the other hand, the segment graph was examined using the toolbox of complex network theory. Firstly, it was observed that above a certain limit the segment graph of all the test networks (9 real-life WDNs) follows nearly the same degree distribution independent of their size. Secondly, based on their structural properties (e.g. large diameter, low clustering coefficient) the segment graph is similar to a random graph, although in our case the networks are necessarily connected, and planar, i.e. edges (isolation valves) cannot cross each other without a node (segment). Thirdly, we observed that the ratio between the number of nodes having degree one

and degree three (or above) could be a basis for the objective categorization of the water distribution networks for their typical topology (from urban to the suburban/countryside, that is from the grid to tree). Which means that, with the help of this ($n_{1/3+}$) ratio, it is possible to characterize the relation between the extension and the density/redundancy of a network. With the help of this factor possible to create a method which identifies the weak points and through that, the larger coast areas in the network as well. Besides that, this ratio can be an objective function for an optimum searcher algorithm, through that, this factor make possible to increase the robustness of a WDS.

## Acknowledgement

## References

[1] Walski, T. M. "Water distribution valve topology for reliability analysis", Reliability Engineering and System Safety, 42(1), pp. 21–27, 1993.
https://doi.org/10.1016/0951-8320(93)90051-Y

[2] Ysusi, M. "System Design: An Overview", In: Water distribution systems handbook, 1th ed., McGraw Hill Professional, New York, USA, 2000, pp. 74–124.

[3] Liu, H., Walski, T., Fu, G., Zhang, C. "Failure Impact Analysis of Isolation Valves in a Water Distribution Network", Journal of Water Resources Planning and Management, 143(7), Article ID: 04017019, 2017.
https://doi.org/10.1061/(ASCE)WR.1943-5452.0000766

[4] Creaco, E., Franchini, M., Alvisi, S. "Optimal Placement of Isolation Valves in Water Distribution Systems Based on Valve Cost and Weighted Average Demand Shortfall", Water Resources Management, 24(15), pp. 4317–4338, 2010.
https://doi.org/10.1007/s11269-010-9661-5

[5] Jun, H., Loganathan, G. V. "Valve-Controlled Segments in Water Distribution Systems", Journal of Water Resources Planning and Management, 133(2), pp. 145–155, 2007.
https://doi.org/10.1061/(ASCE)0733-9496(2007)133:2(145)

[6] Alvisi, S., Creaco, E., Franchini, M. "Segment identification in water distribution systems", Urban Water Journal, 8(4), pp. 203–217, 2011.
https://doi.org/10.1080/1573062X.2011.595803

[7] Gao, T. "Efficient Identification of Segments in Water Distribution Networks", Journal of Water Resources Planning and Management, 140(6), Article ID: 04014003, 2014.
https://doi.org/10.1061/(ASCE)WR.1943-5452.0000395

[8] Diao, K., Sweetapple, C., Farmani, R., Fu, G., Ward, S., Butler, D. "Global resilience analysis of water distribution systems", Water Research, 106, pp. 383–393, 2016.
https://doi.org/10.1016/j.watres.2016.10.011

[9] Liu, H., Walski, T., Fu, G., Zhang, C. "Failure Impact Analysis of Isolation Valves in a Water Distribution Network", Journal of Water Resources Planning and Management, 143(7), Article ID: 04017019, 2017.
https://doi.org/10.1061/(ASCE)WR.1943-5452.0000766

[10] Albert, R., Jeong, H., Barabási, A. L. "Erratum: correction: Error and attack tolerance of complex networks", Nature, 409(6819), p. 542, 2001.
https://doi.org/10.1038/35054111

[11] Barabási, A. L. "Network Science", 1th ed., Cambridge University Press, Cambridge, UK, 2016. [online] Available at: http://network-sciencebook.com/ [Accessed: 31 May 2019]

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

[13] Hwang, H., Lansey, K. "Water Distribution System Classification Using System Characteristics and Graph-Theory Metrics", Journal of Water Resources Planning and Management, 143(12), Article ID: 04017071, 2017.
https://doi.org/10.1061/(ASCE)WR.1943-5452.0000850

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

[15] Torres, J. M., Duenas-Osorio, L., Li, Q., Yazdani, A. "Exploring Topological Effects on Water Distribution System Performance Using Graph Theory and Statistical Models", Journal of Water Resources Planning and Management, Article ID: 04016068, 2016.
https://doi.org/10.1061/(ASCE)WR.1943-5452.0000709

[16] Meng, F., Fu, G., Farmani, R., Sweetapple, C., Butler, D. "Topological attributes of network resilience: A study in water distribution systems", Water Research, 143, pp. 376–386, 2018.
https://doi.org/10.1016/j.watres.2018.06.048

[17] Csárdi, G., Nepusz, T. "igraph Reference Manual", 2012, [online] Available at: http://igraph.sourceforge.net/doc/igraph-docs.pdf [Accessed: 31 May 2019]