# Incremental Versus Optimal Design of Water Distribution Networks - The Case of Tree Topologies

Vivek Anand[1], Aleksandar Pramov[1], Stelios Vrachimis[3,4], Marios Polycarpou[3,4], and Constantine Dovrolis[1,2](✉)

[1] Georgia Institute of Technology, Atlanta, GA, USA
constantine@gatech.edu
[2] The Cyprus Institute, Nicosia, Cyprus
[3] KIOS Research and Innovation Center of Excellence, Nicosia, Cyprus
[4] Electrical and Computer Engineering Department, University of Cyprus, Nicosia, Cyprus

**Abstract.** This study delves into the differences between incremental and optimized network design, with a focus on tree-shaped water distribution networks (WDNs). The study evaluates the cost overhead of incremental design under two distinct expansion models: random and gradual. Our findings reveal that while incremental design does incur a cost overhead, this overhead does not increase significantly as the network expands, especially under gradual expansion. We also evaluate the cost overhead for the two tree-shaped WDNs of a city in Cyprus. The paper underscores the need to consider the evolution of infrastructure networks, answering key questions about cost overhead, scalability, and design efficacy.

**Keywords:** Water distribution networks · incremental design of tree topologies · cost overhead of incrementally-designed networks

## 1 Introduction

Cities live. Cities evolve. Throughout human history, urban agglomerations have changed due to many factors in and out of city planners' control such as migration, economic growth or depression, natural disasters, etc. In the modern age, a city is expected to provide several service infrastructures, including water, sewage, electricity, garbage collection, transportation, etc. Infrastructure networks should be periodically redesigned to adapt to their evolving urban environments. The design approach is typically *incremental*, that is, the cost of network modification is minimized in each design phase [5,6]. Incremental design can potentially lead to suboptimal network topologies [16]. On the contrary, an *optimal* design approach assumes that the network can be redesigned from scratch, and thus the total network cost is minimized.


H. Cherifi et al. (Eds.): COMPLEX NETWORKS 2023, SCI 1144, pp. 251–262, 2024.
https://doi.org/10.1007/978-3-031-53503-1_21

Water Distribution Networks (WDNs) can be divided into transport networks, which are made up of large pipes that transport water from sources to consumer areas, and district networks that connect pipes of smaller diameter to households. District networks often grow in *tree-like topologies*, while transport networks have a looped topology to increase resilience to faults and avoid disruption of service [7,11,17]. However, some cities with hilly terrain may have tree-shaped transport networks to better manage pressure variations.

In this paper we focus on tree-shaped transport WDNs because this simplification allows the derivation of simple and insightful analytical expressions. Understanding the evolution of tree-shaped WDNs represents a reasonable first step, given the simplicity of this topology and the fact that it is often adopted in suburban or rural environments.

The paper focuses on the following fundamental questions:

1. How does a tree-shaped WDN evolve over time, and how does this evolution compare to optimally designed tree topologies?
2. What is the "price of evolution," i.e., the cost overhead of an incremental design relative to the corresponding optimal design for tree-shaped WDNs?
3. How does this cost overhead depend on the way the city expands spatially? In particular, how does random expansion compare to gradual expansion?

In the tree topology, the objective of the design process is to create a network that interconnects a given set of water demand locations, aiming to minimize cost-related objective functions, such as construction, maintenance, and energy consumption. Most of these cost factors are directly related to the length of the pipes. For this reason, and to simplify the optimization problem, we focus on a distance-based cost formulation, where the cost of a network is the total length of its edges (pipes).

Our main findings include analytical asymptotic derivations showing, for instance, that the cost of incremental and optimal designs for tree-shaped networks scales with $\sqrt{N}$, where $N$ is the number of nodes, under random expansion. Additionally, in our quantitative analysis for the transport WDN of a city in Cyprus, we find that the actual water network has only *6% higher length* than the corresponding optimal network.

We review the related work in Sect. 2. In Sect. 3, we present mathematical formulations for both optimal and incremental design problems, applying them to the specific context of topology design for tree-shaped water distribution networks. In Sect. 4, we derive expressions for the evolvability and cost overhead of the incremental design process, comparing different expansion models. In Sect. 5, we examine how the incremental design of a real WDN in Cyprus compares to an optimal tree-shaped WDN.

## 2 Related Work

The study of water distribution networks (WDN) optimal design, has been the subject of extensive research [13]. However, the exploration of incremental network design, particularly in contrast to optimal design, remains relatively unexplored.

*Optimal WDN Design:* Most of the existing literature in the domain of WDNs centers on optimal network design. Various methodologies and algorithms have been proposed (see e.g. [8–10,19]) to achieve optimized solutions based on different constraints and objectives, such as capital cost minimization, reliability enhancement, and operational performance optimization (e.g., reduction of pump energy consumption). Notable works include the application of linear programming for optimal design [8], the use of particle-swarm harmony search [10], and an exploration of the trade-off between cost and reliability [9]. However, these studies predominantly overlook the incremental design approach, often considering it merely as a stepping stone to obtain a nearly-optimal solution.

*Incremental vs Optimal Network Design:* Two key studies in this area by Bakhshi and Dovrolis [2,3] focus on the "price of evolution" in ring and mesh networks in the context of telecommunication networks. These two papers provide the motivation and analytical framework for our study of tree-shaped WDN networks.

Incremental network design, also referred to as "multi-period design," has been also explored in other contexts, such as the design and evolution by source shifts [15], and on biologically inspired adaptive network design [18].

These works propose algorithms and optimization frameworks for incremental network design under diverse constraints and objectives. However, none of them directly compares incremental designs with corresponding optimized designs, nor do they consider different expansion models.

The comparison between incremental and optimized design in the context of WDNs is a relatively unexplored area of research. One relevant study is [14], which introduced a multi-objective optimization model for incremental design, considering topology and capacity expansion.

## 3 Framework and Metrics

In this section, we present the analytical framework for comparing *incremental* and *optimal* network designs. This framework is quite similar to the one developed for the comparison of ring networks in the context of telecommunication infrastructures [2]. An important difference is that we adapt that framework in the context of tree networks (Fig. 1).

The WDN of any city gradually changes with time. To model this, we consider a discrete time mode, where the $k$'th step is referred to as the *k'th environment*. We assume that new nodes are added to the network at each environment. There are two fundamentally different approaches to design a WDN. In the *clean slate* or *optimal* approach we aim to minimize, in every environment, the total cost of the network subject to the given constraints; we refer to the resulting network as the *optimal* design. The other approach is to *minimize the modification cost relative to the network of the previous environment*; we refer to that as the *evolved* or *incremental* design.

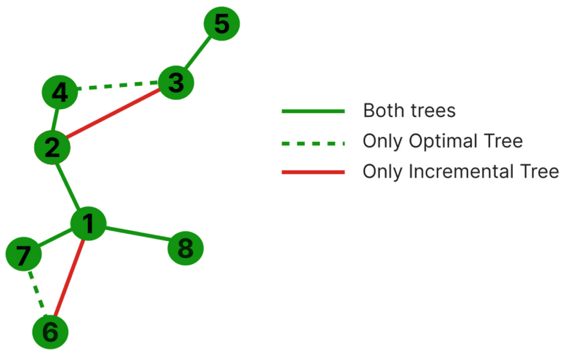



**Fig. 1.** An illustration of the suboptimality (in terms of total length) of an incremental tree (335.91) compared to the optimal one (304.56). The node labels indicate the order in which nodes are added to the graph.

More formally, let $\mathcal{N}(k)$ be the set of *acceptable networks* at environment $k$, i.e. those networks that provide the desired function and meet the constraints of environment $k$. Let the cost of a particular network $N \in \mathcal{N}(k)$ be $C(N)$. In the optimal design, we seek to find a network $N_{opt}(k)$ from the set $\mathcal{N}(k)$ that has minimum cost $C_{opt}(k)$ at environment $k$. Here, $N_{opt}(k)$ is the optimal network at environment $k$.

$$C_{opt}(k) \equiv C(N_{opt}(k)), \quad N_{opt}(k) = \arg\min_{N \in \mathcal{N}(k)} C(N) \tag{1}$$

In the incremental design approach however, we design the new network $N_{evo}(k)$ based on the network $N_{evo}(k-1)$ from the previous environment $k-1$ aiming to minimize the *modification* cost $C_{mod}(N_{evo}(k-1); N(k))$ between the evolved network $N_{evo}(k-1)$ and $N(k)$. We denote the previous modification cost as $C_{mod}(k)$, which is defined as the cost of new design elements in $N(k)$ but not present in $N_{evo}(k-1)$, and the initial evolved network at time 0. We assume that we know the cost of the design elements at time 0.

Now, we can formulate the incremental design problem as

$$C_{evo}(k) \equiv C(N_{evo}(k)), \quad N_{evo}(k) = \arg\min_{N \in \mathcal{N}(k)} C_{mod}(k) \tag{2}$$

Thus, we can express the cost of the evolved network recursively as ($k \geq 1$)

$$C_{evo}(k) = C_{evo}(k-1) + C_{mod}(k) \tag{3}$$

Expanding 3, we get

$$C_{evo}(k) = C_{evo}(0) + \sum_{i=1}^{k} C_{mod}(i) \tag{4}$$

This means that the total cost of the evolved network at environment $k$ is the cost of the initial network plus all the edges that were incrementally added in the last $k$ environments.

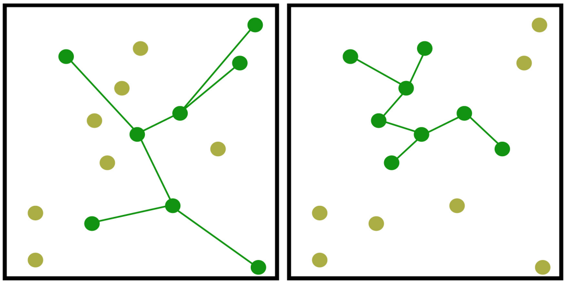
*Metrics:* We use three metrics to compare the two design approaches:

1. *Cost Overhead* $v(k)$ is the cost of the evolved design $N_{evo}(k)$, relative to the corresponding optimal design $N_{opt}(k)$ at environment $k$:
$$v(k) = \frac{C_{evo}(k)}{C_{opt}(k)} - 1 \geq 0 \tag{5}$$
The bigger the cost overhead gets, the more expensive incremental networks are compared to the corresponding optimal networks.
2. *Evolvability* $e(k)$ represents the cost of modifying the evolved network from environment $k-1$ to $k$ relative to the cost of redesigning the network from scratch
$$e(k) = 1 - \frac{C_{mod}(k)}{C_{opt}(k)} \leq 1 \tag{6}$$
If evolvability is close to 1, that means it is much less expensive to modify the existing network than to redesign the network from scratch.
3. *Topological Similarity* $t(k)$ between the optimal $N_{opt}(k)$ and evolved $N_{evo}(k)$ networks is defined as the Jaccard similarity coefficient of the two corresponding adjacency matrices. In other words, $t(k)$ is the fraction of distinct links in either network that are present in both networks.

*Expansion Models:* We consider two specific models for how the environment changes over time, i.e. the set of new nodes added to the network at each environment $k$. We consider the simplest type of expansion in which only one new node is added at each environment. Multi-node expansions can be decomposed as a series of single node expansions.

The two expansion models are *random* and *gradual*. For the former, each new location is selected randomly across a set of all possible new locations $\mathcal{L}$. In gradual expansion, however, we iteratively add the closest new location of $\mathcal{L}$ to the existing nodes of the network (Fig. 2).

The case of random expansion would be more relevant when, for instance, the city planners decide that the WDN of an existing city needs to connect with a village or suburb that is not geographically adjacent to the existing city infrastructure. The case of gradual expansion, on the other hand, would be more relevant when the city grows by adding new developments next to the existing infrastructure, as it often happens when a city grows. Obviously, we expect that the modification cost in gradual expansion will be quite low relative to random expansion (Table 1).



(a) Random Expansion (b) Gradual Expansion

**Fig. 2.** Comparison of the incremental trees under Gradual and Random expansion for the same set of points. Green nodes are the points that have already been added to the tree by the current timestep, while yellow nodes are those that are yet to be added to the tree.

**Table 1.** Notation

| Symbol | Explanation |
|---|---|
| $N_{evo/opt}$ | The evolved/optimal network |
| $C_{opt}(k)$ | The cost of the optimal network at environment $k$ |
| $C_{opt}^{rnd/grd}(k)$ | Optimal cost under random/gradual expansion |
| $C_{mod}(k)$ | Modification cost for added edges in $N(k)$ |
| $C_{mod}^{rnd/grd}(k)$ | Modification cost under random/gradual expansion |
| $v(k)$ | Cost overhead |
| $e(k)$ | Evolvability |
| $t(k)$ | Topological similarity |

## 4 Tree Networks

We assume that all possible nodes of the network, as it expands, are located in a given 2D Euclidean area. This assumption is not necessarily true in real life due to the 3D curvature of the earth and other topographical concerns but it serves as a first-order approximation. Additionally, we assume that the cost of the network is the sum of the edge costs, and that each edge cost is the geodesic distance between the corresponding two nodes.

The optimal tree network can be easily computed using a *minimum spanning tree* (MST) algorithm, such as Kruskal's or Prim's methods [4].

We rely on an asymptotic expression for the length of the MST, derived in [12]. Specifically, *the length of the minimum spanning tree connecting $n$ points in a 2D square of area $A$ scales with $\sqrt{A\,n}$.*

For random expansion, the $n$ points can be anywhere in a given square of area $A$, though these results can be extended to any convex polygon, and so the cost of the MST increases as

$$C_{opt}^{rnd}(n) \sim \sqrt{A\,n} \sim \sqrt{n} \tag{7}$$

In other words, $A$ is viewed as a constant in this case.

In the case of gradual expansion, the area in which the $n$ nodes are located increases with every new node. If we assume that all possible nodes are uniformly distributed with point density $\sigma$, then the area in which $n$ nodes are located is $A = n/\sigma$. So, the cost of the MST under gradual expansion scales as

$$C_{opt}^{grd} \sim \sqrt{A \times n} = \sqrt{n^2/\sigma} = n$$

Let us now calculate the modification for incremental design. If the new node added at environment $k$ is $z$, then the modification cost of the new MST will be at most the distance between $k$ and the closest node to $z$ in the existing MST. More formally, this can be written as

$$C_{mod}(k) \leq \min_{x \in N_{evo}(k-1)}(||z - x||) \tag{8}$$

### 4.1 Random Expansion

To calculate the modification cost under random expansion, we need to recall the nearest neighbor problem: given a set $S$ of $n$ points and a new point $z$ find the closest neighbor of $z$ in $S$. When $n$ points reside in a square of size $A$ with point density $\sigma_{rnd}$, the expected value of the nearest neighbor distance is $1/\sqrt{\sigma_{rnd}} = 1/\sqrt{n/A}$ (see e.g. [1]).

So, the modification cost under random expansion scales as:

$$C_{mod}^{rnd}(n) \sim \frac{1}{\sqrt{n}} \tag{9}$$

If we add the modification cost over $n-1$ node additions, we get the cost of the incrementally designed network under random expansion:

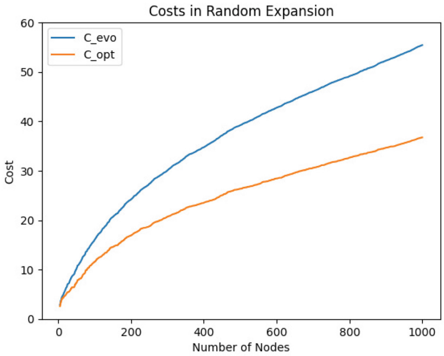
$$C_{evo}^{rnd}(n) = \sum_{i=2}^{n} C_{mod}^{rnd}(i) \sim \sqrt{n} \tag{10}$$

Based on these asymptotic expressions, we see that the cost overhead under random expansion is expected to be constant, at least for large networks, because the cost of both the optimal and incremental trees scales with $\sqrt{n}$,

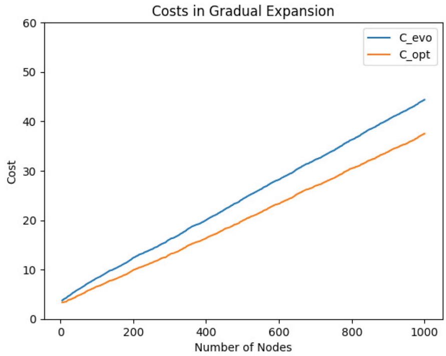
$$v^{rnd}(n) \sim \text{constant}. \tag{11}$$

In addition, from the evolvability definition, we see that the evolvability of the incremental network converges to 1:

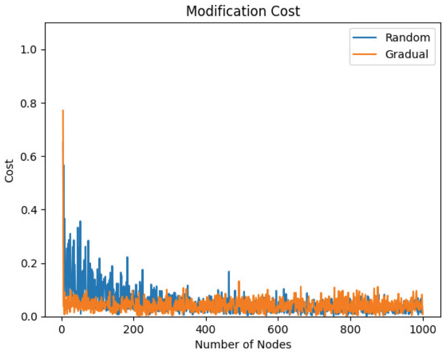
$$1 - e^{rnd}(n) \sim \frac{1}{n} \tag{12}$$



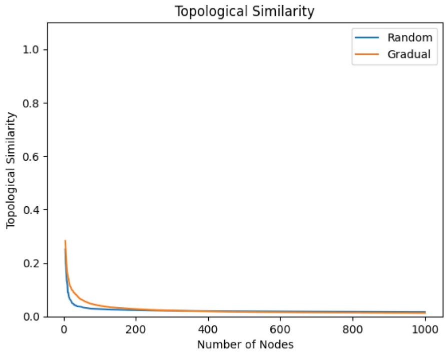

(a) Costs in Random Expansion



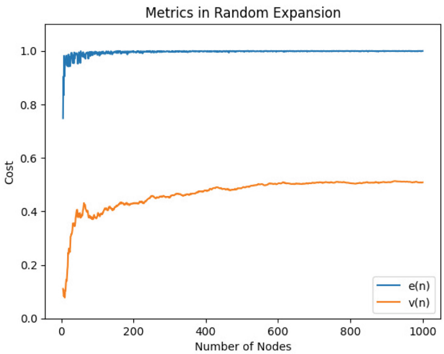

(b) Costs in Gradual Expansion



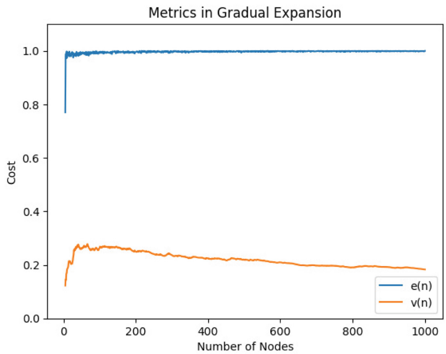

(c) Modification Cost




(d) Topological Similarity




(e) Metrics in Random Expansion




(f) Metrics in Gradual Expansion

**Fig. 3.** Comparison of Metrics between Random and Gradual Expansion during Single Node Expansion

We have confirmed these asymptotic expressions with numerical experiments. We pick points uniformly within a 2D circle before incrementally adding nodes to the minimum spanning tree. The optimal MST is calculated using Prim's Algorithm. Figure 3a shows the evolved and optimal network costs. Figure 3c shows the modification cost whereas Fig. 3e shows the evolvability and the cost overhead. An interesting point is that the topological similarity is close to 0, as shown in Fig. 3d, even for $n < 100$, suggesting that there are many different almost-optimal trees.

## 4.2 Gradual Expansion

Under gradual expansion, we know that the new node is selected as the closest location to the nodes of the existing tree. Consequently, we can again rely on the previous nearest neighbor expression to bound the modification cost. If the potential locations of nodes are uniformly distributed with density $\sigma_{grd}$, then the nearest neighbor is expected to be at distance $\sqrt{1/\sigma_{grd}}$, which does not depend on $n$.

As we only need to add one edge from the existing tree to the new node, the modification cost under gradual expansion does not depend on $n$:

$$C_{mod}^{grd}(n) \sim \text{constant} \tag{13}$$

Using Eq. (4), we see that the cost of the evolved network under gradual expansion scales linearly with $n$,

$$C_{evo}^{grd}(n) = \sum_{i=2}^{n} C_{mod}^{grd}(i) \sim n. \tag{14}$$

So, the cost overhead under gradual expansion remains constant, at least for large network sizes,

$$v^{grd}(n) \sim \text{constant} \tag{15}$$

The evolvability under gradual expansion scales in the same way as with random expansion

$$1 - e^{grd}(n) \sim \frac{1}{n} \tag{16}$$

The difference between random and gradual expansion is more evident in the numerical results shown in Fig. 3. Specifically, the optimal and incremental network costs scale with the square-root of the network size under random expansion, and linearly under gradual expansion – but in absolute magnitude they are much lower under gradual expansion. Additionally, the cost overhead under gradual expansion is significantly lower than that of random expansion. In both cases however, the cost overhead does not increase with network size.

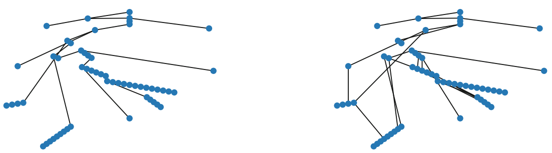
## 5 Case Study

In this section, we examine the transport component of the WDN of a city in Cyprus. We ask, how would this network compare to an optimal design connecting exactly the same set of nodes.

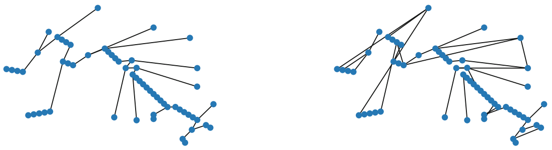
The nodes in this network consist of water sources and water sinks; typically, for water transport networks, sources are water reservoirs and sinks are consumer areas where the consumption is measured, referred to as District Metered Areas (DMAs). Node parameters include their coordinates and elevations. Link parameters, that is the pipes connecting those nodes, include pipe lengths and diameters.

The city's WDN has two special features which make it distinct: it is split in two 'zones', each with their own water source. Second, both zones use a tree topology for the main transport network. If we denote the two networks as $G_{evo}^{Zone_1}$ and $G_{evo}^{Zone_2}$, we are thus interested in finding their optimal counterparts $G_{opt}^{Zone_1}$ and $G_{opt}^{Zone_2}$.

We find the optimal trees $G_{opt}^{Zone_1}$ and $G_{opt}^{Zone_2}$ by first calculating the pairwise distance between all nodes in each zone separately, using the geodesic distance approximation. Then we calculate the minimum spanning tree over the entire pairwise distance matrix using Kruskal's algorithm [4].



(a) Zone 1: Optimal network (b) Zone 1: Actual network



(c) Zone 2: Optimal network (d) Zone 2: Actual network

**Fig. 4.** Visualization of optimal and actual trees for Zone 1 and Zone 2.

In Table 2, we see that for both Zone 1 and Zone 2, the actual network is at most 6% longer than the corresponding optimal network. The topological similarity for both Zone 1 and Zone 2 is almost 60%. In Fig. 4, we see that the two trees for the optimal and actual network of each Zone have some similarities but there are also many differences in the connectivity. The rationale for these differences is something we plan to further explore, consulting with the engineers of the city's water utility.

**Table 2.** Comparison of actual and optimal water networks for a city in Cyprus. The costs are measured in meters. We show the cost of the actual network $G_{evo}$ both based on the length of the actual pipes (always placed under large roads) and based on geodesic distances. For the optimal network $G_{opt}$ we can only calculate costs based on geodesic distances.

| Zone | # Nodes | Cost of $G_{evo}$ (pipe lengths) | Cost of $G_{evo}$ (geodesic) | Cost of $G_{opt}$ (geodesic) | Cost Overhead | Topological Similarity |
|---|---|---|---|---|---|---|
| 1 | 56 | 12006 | 11171 | 10540 | 0.06 | 0.61 |
| 2 | 59 | 10875 | 10519 | 10221 | 0.03 | 0.59 |

This city in Cyprus has expanded significantly since 1990, with major new developments and population increase. The low cost overhead shown above suggests that this expansion was probably closer to the gradual model, with new WDN nodes added close to existing ones. The fact that the topological similarity of the optimal and actual tree networks is close to 60%, on the other hand, suggests that the networks are significantly different, as there are probably many different 'close-to-optimal' trees with about the same cost.

## 6 Conclusion, Limitations and Future Work

The paper contributes to the evaluation of incrementally designed networks relative to their optimal counterparts, in the context of tree topologies applied for water distribution in urban settings. Through simple and insightful analytical derivations, we show that the cost overhead of incrementally designed trees under random expansion can be significant but, it is expected to asymptotically remain constant. Under gradual expansion on the other hand, the cost overhead also does not increase with network size, and it is significantly lower in absolute terms than under random expansion.

An important limitation of this study is that we focus only on the topology of the network. A real WDN also has to meet construction constraints (e.g., pipes are typically under large roads) or hydraulic constraints (e.g., minimum or maximum water pressure). These constraints are met in practice through optimized pipe sizing and placement of water pumps and tanks, while also considering the effect of changing water demands. Additionally, the optimization of a WDN considers not only the length of the pipes but also operational factors such as the electricity consumed by pumps. Further, many transport WDNs in larger cities have mesh-like structures, for reliability and redundancy. In future work, we plan to expand this investigation considering these more pragmatic constraints.

**Acknowledgments.** This work was co-funded by the European Research Council (ERC) under the ERC Synergy grant agreement No. 951424 (Water Futures), and supported by the European Union's Horizon 2020 Teaming programme under grant agreement No. 739551 (KIOS CoE), and the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy.

## References

1. Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R., Wu, A.Y.: An optimal algorithm for approximate nearest neighbor searching fixed dimensions. J. ACM (JACM) **45**(6), 891–923 (1998)
2. Bakhshi, S., Dovrolis, C.: The price of evolution in incremental network design (the case of ring networks). In: Hart, E., Timmis, J., Mitchell, P., Nakamo, T., Dabiri, F. (eds.) BIONETICS 2011. LNICST, vol. 103, pp. 1–15. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-32711-7_1
3. Bakhshi, S., Dovrolis, C.: The price of evolution in incremental network design: the case of mesh networks. In: 2013 IFIP Networking Conference, pp. 1–9. IEEE (2013)
4. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms. MIT press (2022)
5. Creaco, E., Franchini, M., Walski, T.M.: Accounting for phasing of construction within the design of water distribution networks. J. Water Resour. Plan. Manag. **140**(5), 598–606 (2014)
6. Cunha, M., Marques, J., Creaco, E., Savić, D.: A dynamic adaptive approach for water distribution network design. J. Water Resour. Plan. and Manage. **145**(7), 04019026 (2019)
7. Diao, K., Sweetapple, C., Farmani, R., Fu, G., Ward, S., Butler, D.: Global resilience analysis of water distribution systems. Water Res. **106**, 383–393 (2016)
8. Eiger, G., Shamir, U., Ben-Tal, A.: Optimal design of water distribution networks. Water Resour. Res. **30**(9), 2637–2646 (1994)
9. Farmani, R., Walters, G.A., Savic, D.A.: Trade-off between total cost and reliability for anytown water distribution network. J. Water Resour. Plan. Manag. **131**(3), 161–171 (2005)
10. Geem, Z.W.: Particle-swarm harmony search for water network design. Eng. Optim. **41**(4), 297–311 (2009)
11. Giudicianni, C., Di Nardo, A., Di Natale, M., Greco, R., Santonastaso, G., Scala, A.: Topological taxonomy of water distribution networks. Water **10**(4), 444 (2018). https://doi.org/10.3390/w10040444
12. Jaillet, P.: Rate of convergence for the Euclidean minimum spanning tree limit law. Oper. Res. Lett. **14**(2), 73–78 (1993)
13. Kapelan, Z.S., Savic, D.A., Walters, G.A.: Multiobjective design of water distribution systems under uncertainty. Water Resour. Res. **41**(11), 1–15 (2005)
14. Marques, J., Cunha, M., Savić, D.A.: Multi-objective optimization of water distribution systems based on a real options approach. Environ. Model. Softw. **63**, 1–13 (2015)
15. Prakash, R., Shenoy, U.V.: Design and evolution of water networks by source shifts. Chem. Eng. Sci. **60**(7), 2089–2093 (2005)
16. Saleh, S.H., Tanyimboh, T.T.: Optimal design of water distribution systems based on entropy and topology. Water Resour. Manage **28**(11), 3555–3575 (2014)
17. Shuang, Q., Liu, H.J., Porse, E.: Review of the quantitative resilience methods in water distribution networks. Water **11**(6), 1189 (2019). https://doi.org/10.3390/w11061189
18. Tero, A., et al.: Rules for biologically inspired adaptive network design. Science **327**(5964), 439–442 (2010)
19. Yazdani, A., Dueñas-Osorio, L., Li, Q.: A scoring mechanism for the rank aggregation of network robustness. Commun. Nonlinear Sci. Numer. Simul. **18**, 2722–2732 (2013)