Generation of Synthetic Water Distribution Data Using a Multiscale Generator-Optimizer

Ahmad Momeni, S.M.ASCE[1]; Varsha Chauhan[2]; Abdulrahman Bin Mahmoud, A.M.ASCE[3]; Kalyan R. Piratla, A.M.ASCE[4]; and Ilya Safro[5]

Abstract: Rare or limited access to real-world data has widely been a stumbling block for the development and employment of design optimization and simulation models in water distribution systems (WDS). Primary reasons for such accessibility issues could include data unavailability and high security protocols. Synthetic data can play a major role as a reliable alternative to mimic and replicate real-world WDS for modeling purposes. This study puts forth a comprehensive approach to generate synthetic WDS infrastructural data by (1) employing graph-theory concepts to generate multitudinous WDS skeleton layouts through retaining the critical topological features of a given real WDS; and (2) assigning component sizes and operational features such as nodal demands, pump curves, pipe sizes, and tank elevations to the generated WDS skeleton layouts through a multiobjective genetic algorithm (GA)–based design optimization scheme. Thousands of such generated-optimized networks are statistically analyzed in terms of the fundamental WDS characteristics both collectively and granularly. An outstanding novelty in this study includes an automatedly integrated algorithmic function that attempts to (1) simultaneously optimize the generated network in a biobjective scheme, (2) rectify pipe intersections that violate pipeline embedding standards, and (3) correct the unusual triangular loops in the generator by honoring the conventional square-shaped loop connectivity in a WDS. The proposed modeling approach was demonstrated in this study using the popular Anytown water distribution benchmark system. Generation and optimization of such representative synthetic networks pave the way for extensive access to representative case-study models for academic and industrial purposes while the security of the real-world infrastructure data is not compromised. **DOI: 10.1061/JPSEA2.PSENG-1358.**

Author keywords: Water distribution systems (WDS); Optimization; Graph generation; Multiscale method; Infrastructural data; Synthetic networks; Graph theory.

Introduction

One of the major challenges in developing analytical models in critical infrastructures (CIs) including water distribution systems (WDS) for the purpose of design optimization, simulation, and asset management schemes is the limited access to real-world infrastructural data (Ahmad et al. 2020; Zhang et al. 2017). Specifically, scarcity of real-world WDS data has been a troubling hindrance for representative mimicking of such analytical models (Menapace et al. 2020). Moreover, hardship of collection, validation, and calibration as well as digitalization of such scarce data can be deemed problematic to employ analytical or hydraulic applications in WDS (de Corte and Sörensen 2014). Despite the availability of a few synthetic WDS in the literature, the physical characteristics of these benchmarks are not comparable with and applicable to the enormity and complexity of real WDS in practice (Jaskowski et al. 2012).

It is evident that a significant number of WDS are conveniently simplified and adopted for academic research purposes, which may fall short of capturing the subtleties and intricacies of real-world WDS. As physical and specifically cyber attacks have become more prevalent in critical infrastructure systems including WDS, security protocols have been comprehensively tightened regarding to what extent infrastructural and cyber-monitoring data can be overtly accessible (Tuptuk et al. 2021). Furthermore, this lack of infrastructural data fails to provide ample granularity associated with a calibrated hydraulic model for the research community to implement small-scale reliable steady- or transient-state scenarios.

Addressing this scarcity of real-world WDS infrastructure data, this paper presents an efficient approach for generating synthetic WDS data using a topological network generation mechanism combined with a subsequent physical design paradigm. The input for the presented framework is any one WDS network, and the output entails several representative WDS networks with varying topological and hydraulic features. The approach was demonstrated using the popular widely used Anytown benchmark network (Farmani et al. 2005; Prasad and Tanyimboh 2008), although the original layout is somewhat atypical. Efficient, representative generation of synthetic networks was optimized using novel algorithms that can address oddities such as loop triangularity, node degree of connectivity, and pipe intersections on the go. The generated network optimization also accounts for various parameters, and some essential ones undergo randomization for maintaining stochasticity in the process.

[1]Postdoctoral Associate, Dept. of Civil and Environmental Engineering, Cornell Univ., Ithaca, NY 14853 (corresponding author). ORCID: https://orcid.org/0000-0002-4506-517X. Email: amomeni@cornell.edu

[2]Formerly, Graduate Student, School of Computing, Clemson Univ., Clemson, SC 29634. Email: varsha.chauhan1190@gmail.com

[3]Assistant Professor, Dept. of Civil Engineering, College of Engineering, King Saud Univ., Riyadh 11362, Saudi Arabia. ORCID: https://orcid.org/0000-0002-7193-0737. Email: abinmahmoud@ksu.edu.sa

[4]Liles Associate Professor, Glenn Dept. of Civil Engineering, Clemson Univ., Clemson, SC 29634. ORCID: https://orcid.org/0000-0003-0836-7598. Email: kpiratl@clemson.edu

[5]Associate Professor, Dept. of Computer and Information Sciences, Univ. of Delaware, Newark, DE 19716. ORCID: https://orcid.org/0000-0001-6284-7408. Email: isafro@udel.edu

Note. This manuscript was submitted on April 4, 2022; approved on October 27, 2022; published online on December 14, 2022. Discussion period open until May 14, 2023; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Pipeline Systems Engineering and Practice***, © ASCE, ISSN 1949-1190.

Literature Review

A myriad of studies in WDS industry have been affected by the scarcity of historical data for scenario-based and time-series analyses including background leak localization (Hu et al. 2021; Momeni and Piratla 2021) and periodic system demand or consumption data analysis (Zhang et al. 2017), as well as condition assessment practices like pressure transients over a long-term period (Moslehi et al. 2021). The critical need is to have access to data from real WDS infrastructures or at least have synthetic infrastructural data that are highly representative of real WDS. Although there have been various endeavors as to how synthetic infrastructural data can be generated (Mair et al. 2014; Sitzenfrei et al. 2010, 2013), past studies have encountered some limitations such as unaccounted-for operational analysis (Mair et al. 2014), limited synthetic representation of WDS geometric design (Sitzenfrei et al. 2010, 2011), partial parameterization of hydraulic designs (Creaco et al. 2017), inadequate dimensionality and inaccurate graph diameters, as well as unorthodox edge-to-node proportion and connectivity degree (Giudicianni et al. 2018), and finally selective analysis of synthesization of WDS characteristics such as residual demand patterns (the Overall Pulse method) (Candelieri 2017; Di Palma et al. 2017).

Costa and Rodrigues (2021) created a model for the automatic generation of the EPANET input file from the layout of street arrangements in an Auto-CAD framework and developed a set of algorithms on binary image processing to identify nodes and pipes on roads. Ahmad et al. (2020) presented a model entitled Synthetic Infrastructure that leverages roadway networking, water consumption, and source location to synthesize network characteristics such as topology. They tested their model on City of Tempe, Arizona, and scaled their model to other major cities. Zeng et al. (2017) developed a model that can generate a WDS according to real-world counterparts on some structural properties and compared it with existing models in terms of how realistic demand node distribution are for the simulation of real complex networks. Muranho et al. (2012) put forth an EPANET extension called WaterNetGen for automatic water distribution network generation that can assign network topology and pipe diameters. Hallmann and Kuhlemann (2018) presented a generator for WDS models that can create networks with arbitrary size and some realistic properties. In addition, de Corte and Sörensen (2014) developed an artificial network generating platform known as HydroGen to create networks of arbitrary sizes and varying characteristics in EPANET and later compared the networks with their real-world counterparts using graph-theoretic indices.

Such studies have developed sophisticated and representative demonstrations of generative platforms proportionally, yet they have been found to partly suffer from a systematic and robust employment of hydraulic simulation and optimization of the output synthetic networks to ensure a resilient, cost-effective, and granular replication of real-world WDS (Di Palma et al. 2017; Ahmad et al. 2020; Creaco et al. 2017). Besides, most of the generators in the literature factor in a graph-theoretic approach where characteristics of the network are arbitrarily assigned, which might compromise the reliability of the system. As a result, this paper aims at pushing the boundaries of network generation by tackling the problem of paucity of large and complex WDS infrastructural data by providing an approach for generating thousands of synthetic WDS derived from data of a given real-world WDS through preservation of the critical topological and operational characteristics.

This study adds value and further novelty by introducing an automatedly integrated algorithmic approach in (1) optimizing the graph-theoretic generated network in terms of hydraulic properties in a genetic-algorithm framework, (2) removing pipe intersections by factoring in reservoir, tank, and consumption node coordinates, and (3) presenting a novel triangularity-removing capability that attempts to maintain the conventional square loops (Huzsvár et al. 2019; Choi et al. 2019) in a real-world WDS rather than the unorthodox connectivity among three nodes. The resulting synthetic networks were optimized for their physical design based on cost and resilience objectives. Graph-theory principles (Zhou et al. 2020; Zverovich 2021) were employed to preserve the critical topological properties of WDS. The presented synthetic WDS generator platform yields promising horizons in providing academic researchers and industrial entities with reliable synthetic networks in lieu of scarce and inaccessible real-world infrastructure data.

Methodology

The study methodology entails (1) generation of topology, where a predefined WDS benchmark network undergoes various generative modifications through multiscale graph editing methods before producing over 1,000 networks with new skeleton layouts, (2) WDS-tailored modifications of topological features and assignment of specific WDS configurations to the generated WDS skeleton layouts, (3) incorporation of an automated framework integrated in the generator to remove triangular loops and pipe intersection in a systematic manner, and (4) design optimization of the configured WDS layouts for physical characteristics considering the objectives of minimum cost and maximum system resilience.

Network Generation

Generating synthetic networks is one of the central questions in network science (Newman 2010; Penschuck et al. 2020). Reproducing various aspects of similarity between the synthetic and original networks while maintaining sufficient diversity of generated synthetic networks are among major concerns. One large class of generative approaches suggests the generation of a network with several predefined structural properties (e.g., degree distribution) but with no original reference network. Such approaches usually start with the empty or small-seed network and run an iterative process in which nodes and edges are added while preserving some properties (measured from the real instances). Examples include Barabási-Albert and Chung-Lu models (Barabási and Albert 1999; Aiello et al. 2001). Although tens of such approaches exist, they are typically not successful in replicating spatially engineered infrastructures because of many different domain-dependent details.

To address such issues, this paper utilizes the multiscale approach (Gutfraind et al. 2015) that belongs to the class of deviative approaches in which the generation process makes a number of steps to move away from the original network while preserving some structural properties. In a simplest form, such a deviation can be achieved by randomized rerouting of some edges. This study specifically leverages the multiscale approach for generating planar graphs (Chauhan et al. 2019) that are structurally similar to the original one to account for real-world scenarios in terms of synthetic generation of WDS. In particular, the multiscale method used here for network generation belongs to the family of scalable optimization solvers inspired by the algebraic multigrid (Safro et al. 2006; Safro and Temkin 2011; Ron et al. 2011).

Generation of Topology

In order to mass-generate synthetic networks out of a predefined WDS benchmark, multiscale graph editing approach (Chauhan et al. 2019) is hereby leveraged. It is noteworthy that graphs underlying WDS are typically (almost) planar, which is the version that this study will apply instead of the generalized approach (Gutfraind et al. 2015). A known real WDS layout is initially converted to an unweighted, undirected graph with no self-loops or multiedges where water network elements such as junctions, reservoirs, and ground or elevated storage tanks (GSTs and ESTs) are represented as nodes of a graph and pipes; pumps and valves are represented as edges. The generated simple graph is then used as an input to the proposed multiscale graph generator to produce a synthetic graph by introducing random edits at multiple scales of coarseness while preserving structural properties such as degree distribution and clustering coefficient.

There are three main stages in the multiscale solver, namely (1) coarsening, i.e., the process of creating a compressed representation of the original problem, (2) the coarsest scale solution, when the number of degrees of freedom is coarsened to be small enough to be solved with high quality, and (3) uncoarsening, the process of using the coarse-scale solution to obtain the next finer scale solution by interpolating and refining it. In multiscale network generation, the aforementioned second and third stages are replaced with editing the original graph at the given scale by adding and removing its nodes and edges and preserving some structural properties. Because in this work the critical realistic structural property of the water network is planarity, the generator preserves it (Chauhan et al. 2019). The algorithm is explained in detail in the next section.

Coarsening

In this step, a hierarchy of coarse graphs $G_i = (V_i, E_i)$, $1 \leq i \leq k$, is generated from an input graph $G_1 = (V_1, E_1)$ using the weighted aggregation (Safro and Temkin 2011) of nodes and edges. The number of levels in the hierarchy depends on the user input and size of the graph. For example, to introduce only local changes to the network, the editing will be done only at the fine levels without creating coarse levels and going deeper in the hierarchy.

The first step of the coarsening at scale i consists of passing through all nodes and deciding what nodes will seed coarse aggregates. Two sets C and F are primarily considered, where initially $C = \emptyset$ and $F = V_i$ (i.e., contains nodes in the graph at level i). Nodes are then added iteratively from F to C such that newly added nodes are not strongly connected to nodes already in C. The connection strength is the criterion, which is expressed

$$\frac{\sum_{j \in C} w(ij)}{\sum_{j \in V} w(ij)} \leq \alpha \quad (1)$$

where $w(ij)$ = weight of edge between nodes i and j; and α = coarsening threshold bounding the required relative strength of connection between neighbors already chosen to C and the entire neighborhood in V. Typically, $\alpha = 0.5$ ensures graduate coarsening that is not too fast or slow. Various thresholds have been evaluated in many multiscale coarsening schemes including those for graph generation. Too large values of α keep the number of coarse variables large, which increases the number of levels. This makes the computational process less efficient in terms of the number of operations. Too small values of α exhibit an opposite effect because it compresses the entire graph almost immediately. Overcompressing it makes a generated structure less realistic and less similar to the original structure because the introduced randomization tends to change the structure more globally. Because in water networks we are dealing with (nearly) planar graphs, the threshold of 0.5 roughly ensures having approximately $\log|V|$ levels in the hierarchy without overcompressing or undercompressing it.

In general, it is expected that any alpha between 0.4 and 0.6 will exhibit more or less the same robust results that will not make any statistical difference in comparison with 0.5. If extremely high degree of similarity is required by the application, we recommend to sacrifice the computational time and increase α to 0.8–0.9 in combination with keeping the node and edge edit ratio very low. If a considerable deviation from the original network structure is acceptable the value of α can be decreased to 0.3–0.4. It is important to mention that different values of α are not expected to make the optimization of physical parameters more or less difficult. They only control the pace of coarsening and the levels of coarseness in which randomization is introduced.

The final phase of coarsening is computing the connection strength between the coarse nodes.The algebraic multigrid interpolation matrix P of size $|V| \times |C|$ is hereby defined, in which P_{ij} represents the likelihood of i to belong to the jth aggregate. The Laplacian of the coarse graph G_{i+1}, L_{i+1}, can be calculated by the algebraic multigrid coarsening operator $L_{i+1} \leftarrow P^T L_i P$, where L_i is the Laplacian of ith level graph, and

$$P_{ij} = \begin{cases} 1, & \text{for } i \in C, j = i \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

The edge ij, connecting two coarse nodes i and j, is assigned with the weight

$$\sum_{k \neq l} P_{ki} w(kl) P_{lj}$$

and the volume of the ith coarse aggregate is $\sum_j v(j) P_{ji}$. To this end, the $(i+1)$th level graph is generated, and thus the properties of ith level graph can be measured and stored in Q_i.

Graph Structure Uncoarsening and Editing

After the hierarchy of coarse representations is created, the uncoarsening process will be carried out. In contrast to the multilevel optimization solvers in which a coarse problem is locally optimized at each level, in the context of network generation, the uncoarsening consists of graph structure modification and assignment of WDS related parameters to nodes and edges (Chauhan et al. 2019).

At the coarsest level, the graph can only be edited through controlled randomization, whereas at all other levels, the graph that has been modified at the coarse level needs to be uncoarsened. This process begins by uncoarsening unedited nodes and edges. In this case, the coarse aggregates were eliminated, and fine nodes with their forming edges were kept unmodified. In the next stage, edited aggregates are interpolated, and edges that were trapped within these aggregates are uncoarsened and added to connect edited node to preserve degree distribution. The number of edges is preserved from the corresponding original coarsened graph. In the final stage, additional structural editing (i.e., the randomization) is leveraged, in which some nodes and edges are removed and new ones may be introduced. Fig. 1 illustrates the processes associated with the generative scheme in the integrated generator-optimizer framework.

Modification and Assignment of WDS Parameters

Conversion of the plain generated graph into a feasible WDS constitutes the postgeneration step for representativeness of output networks. In this process, fundamental features of the original network in the WDS layouts (e.g., the location of tanks and reservoirs as well as specifications of pumps) are adopted and sometimes retained to finalize the water network configuration according to normal operational and hydraulic specifications of a given real-world WDS.

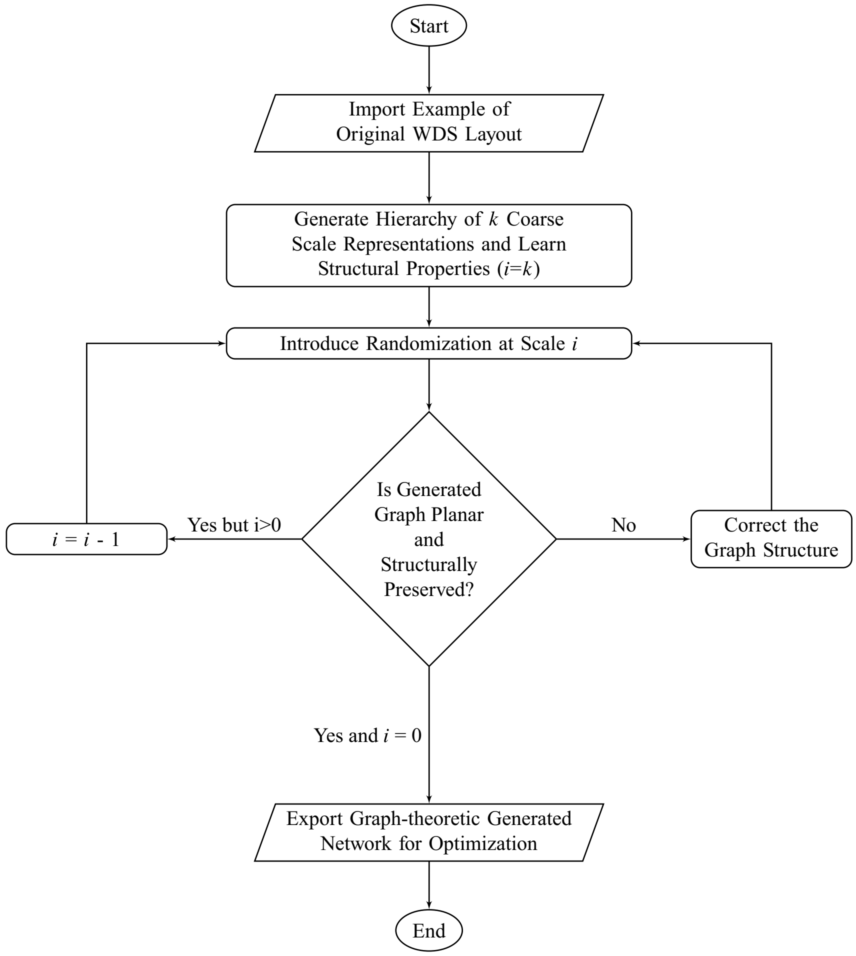

Fig. 1. Algorithmic flowchart of the generative model.

Assignment of Coordinates

The first step in the generation of WDS from a planar topology is associated with the assignment of coordinates of each WDS component. Although the generated topology is planar, embedding of the graph in two-dimensional (2D) space is deemed as a complex procedure; therefore, existing visualization algorithms such as Forceatlas2 (Jacomy et al. 2014) are employed to generate 2D representation of the generated planar graph. The number of iterations of the graph visualization algorithms can be controlled by the end user.

Assignment of Reservoirs

The most essential component in a WDS that can supply water down in the network is a reservoir, and thus ensuring the graph generator can provide a minimum of one reservoir is mandatory. In case the generated network is rescaled such that it is larger than the original network, the number of reservoirs is systematically increased by the rescaling factor while also introducing a decreasing probability coefficient, β, that both adds a randomization factor and restricts assignment of improbable or unrealistic number of reservoirs in the generated network.

In this study, it was hypothesized that the generated reservoir is assumed to be peripherally situated on the generated graph, which ascertains a real-world characterization of system supply by being connected to the main body of the network skeleton. It was also assumed that the generated reservoir is connected to the network through a single edge (link); therefore, candidate nodes whose degree of connectivity is greater than one will be disregarded. Finally, candidate nodes that are selected for the associated reservoirs will be assigned the correct indices accordingly, and certain elevation values will be assigned to each reservoir by reasonably randomizing original network reservoir elevation.

Assignment of Storage Tanks

GSTs and ESTs play a significant role in the WDS operationality. Using GST or EST storage capability, the WDS are designed to be redundant, resourceful, and reliable in case of contingencies. GSTs or ESTs are basically modeled on the peripheries of networks by first dividing the network into clusters such that they mimic the real-world WDS. This was obtained by the minimum edge cut partitioning graph algorithm (Fan et al. 2020). Then, a family of graph partitioning programs known as Karlsruhe High Quality Partitioning (KaHip) (Sanders and Schulz 2013) were recursively employed such that the number of partitions is equal to the number of storage tanks to be allocated. The partitions can be considered as clusters or communities found in real-world water networks, where each cluster consists of a storage tank.

Next, a randomization is presented by introducing probability factor β, which controls addition of unrealistic number of tanks. Within each cluster, a candidate node with a degree of connectivity of one was selected through employment of the random probability factor β. If β turns out to be greater than a defined value (i.e., p), the candidate node will be assigned to a storage tank; otherwise, no storage tank is assumed to be assigned in that partition (cluster).

Table 1. Features of the generator framework

Characteristics	Lower boundary/desirable	Upper boundary/extreme
Degree of node connectivity	One connection	Four connections
Node edit and growth rate	0.02–0.04	0.07–0.09
Edge edit and growth rate	0.02–0.04	0.07–0.09
Tank placement policy	Adjacent to reservoirs/pumps	On network edge
Pump connectivity	Directly connected to reservoirs/tanks	Adjacent to reservoirs/tanks

This process was repeated for each partition assigning tanks in each cluster according to the same probability factor of p.

Finally, fundamental tank characteristics are either randomized based on the original network values or optimized through the biobjective optimization platform presented subsequently in the paper. Such storage tank parameters include minimum level, maximum level, diameter and elevation, whose detail will be discussed subsequently. It is noteworthy that tank levels are correctly placed within a reasonable range of minimum and maximum values, where illogical values will be discarded using a constraint function in the optimization platform.

Assignment of Junctions

Junctions denote the WDS demand locations as consumption nodes where demands should be satisfied through tank or reservoir supply in the system. The assignment of junctions are implemented after tanks and reservoirs have been indexed and added to the network; therefore, the remaining generated nodes will automatically be assigned to demand nodes. In the proposed generation model, junction parametric values were randomized within the range of the demand and elevation of the corresponding nodes from the original WDS. Sets of E and D, which denote elevation and demand values in the input network, respectively, were produced to characterize the fundamental parametric values of each demand node. However, this act of randomization will most likely cause an uneven distribution of elevation and demand values across the network skeleton (e.g., a low-situated elevation/demand node can be in a cluster that mostly entails fairly high elevation/demand nodes), so this may result in a nonrepresentative network layout. Therefore, in order to resolve this issue, an ad hoc iterative smoothening approach (Lei et al. 2020) was employed in the process of junction parameterization such that the neighboring junctions have geologically and geometrically representative characterization of demand and elevation values.

Assignment of Pumps

An elemental property typically found in the real-world WDS includes the placement of pumps in high proximity to supply nodes (i.e., reservoirs and storage tanks). The proposed generator in this study replicated this property by utilizing the identical KaHip recursive partitioning (Sanders and Schulz 2013) as used for assigning tanks and assigning pumps by selecting edges in a partition that is comprised of a tank or reservoir, thus guaranteeing the adjacency of pumps to supply nodes. Ultimately, random values of pump head, flow, and pattern were assigned at the generation phase, which will undergo optimization later by the proposed optimization scheme.

Assignment of Pipes

After pumps were assigned onto the network layout, the remaining edges were assumed to be assigned to pipes to maintain the connectivity and redundancy in the WDS. Fundamental properties of pipes such as length, diameter, roughness, and status were either randomized within a reasonable range of the corresponding values from the original layout or optimized in the optimization platform, which will be presented subsequently in the paper.

Real-World WDS Considerations

This section doubles down on the features of a coarsened generated graph that is meant to satisfy the fundamental characteristics of a real-world WDS. Such considered features in this study include (1) maximum amount of connectivity per node, (2) modification of the generated graph to only account for square loops, (3) modification of the generated graph to avoid pipe intersections, and (4) adjustment of pipe lengths with their associated node coordinates. Such features have been embedded as ad hoc Python function blocks within the generator platform. Table 1 displays the characteristics of the finalized generated networks.

Optimization of Generated Water Distribution System

In order to optimize a water distribution network in terms of operations, there are various approaches proposed in the literature (Farmani et al. 2006; Odan et al. 2015; Ocampo-Martinez et al. 2009). This study considers two objectives to be optimized in a trade-off scheme: (1) a weighted resilience metric, which characterizes the reliability of the WDS produced by the generator; and (2) operational and installation cost, which accounts for the cost of WDS components and their operation over time. Evolutionary algorithms [i.e., Non-dominated Sorting Genetic Algorithm II (NSGA II)] (Shi et al. 2020) in Python 2.7 were coupled with EPANET 2.2 software application input file of the case study in Linux environment to optimize the trade-off between cost and resilience as the two objective functions in this paper (Farmani et al. 2005). Fig. 2 illustrates the algorithmic flowchart associated with the proposed optimization scheme in the proposed generator-optimizer framework.

Resilience Measure for Water Distribution Systems

Resilience of a water distribution system is defined as its ability to recover from local or global failures in that the network is known to be capable of adapting, recovering, and returning to its normal functionality and status quo (Zhang et al. 2020; Lorenz and Pelz 2020; Shin et al. 2020; Diao 2020). The resilience metric R_t at time step t is formulated and can be observed in Eqs. (3) and (4) (Todini 2000; Farmani et al. 2005) as follows:

$$R_t = \frac{\sum_{j=1}^{n} q_{j,t}(h_{j,t} - h^*_{j,t}) + \sum_{k=1}^{T} T^f_{k,t} T^l_{k,t}}{(\sum_{k=1}^{T} T^s_{k,t} T^l_{k,t} + \sum_{r=1}^{S} Q_{r,t} H_{r,t} + \sum_{b=1}^{B} P_{b,t}) - (\sum_{j=1}^{n} q_{j,t} h^*_{j,t} + \sum_{k=1}^{T} T^f_{k,t} T^{\min}_{k,t})} \quad (3)$$

J. Pipeline Syst. Eng. Pract., 2023, 14(1): 04022074

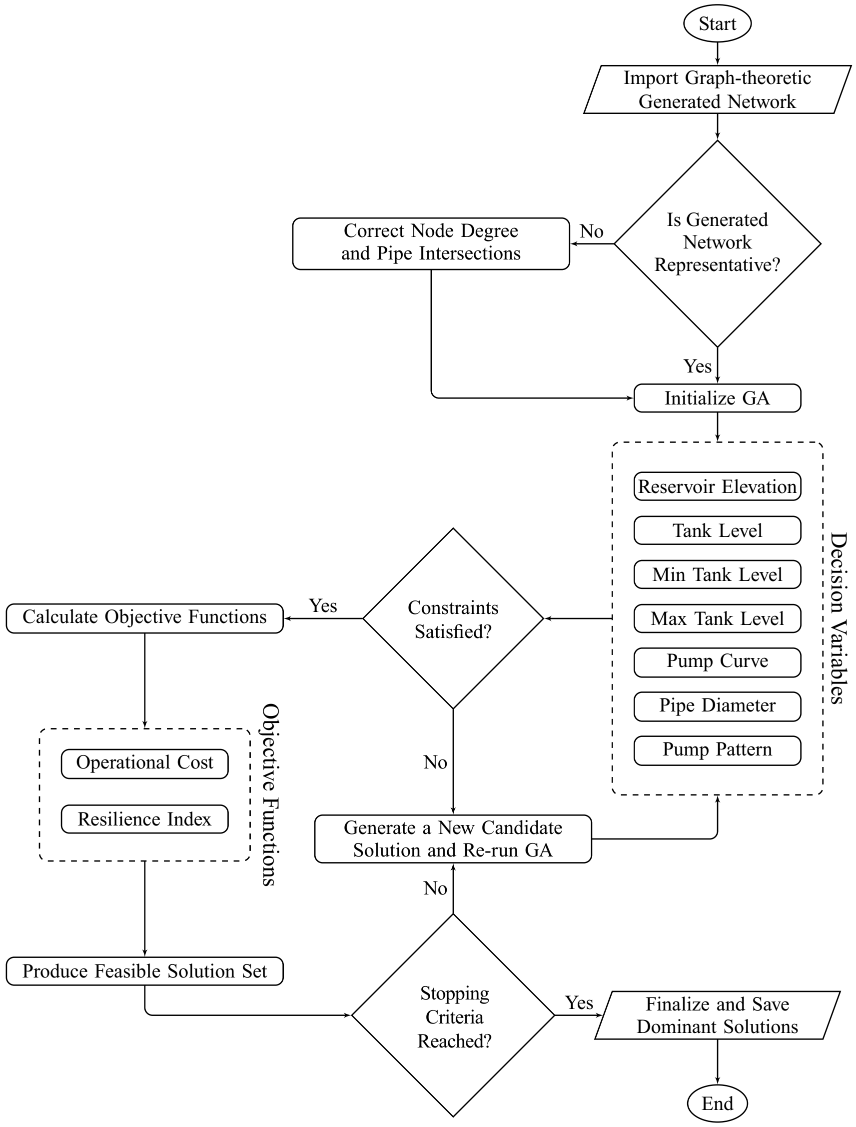

Fig. 2. Algorithmic flowchart of proposed optimization model.

$$R_{\text{tot}} = \frac{\sum_{t=1}^{N} R_t}{N} \tag{4}$$

where R_t = resilience at time step t; n = number of demand nodes; $q_{j,t}$ = demand at node j at time step t; $h_{j,t}$ = head available at node j at time step t; $h^*_{j,t}$ = minimum head required to meet constraints at node j at time step t; S = the number of reservoirs; $Q_{r,t}$ = flow being supplied to the system by reservoir r at time step t; $H_{r,t}$ = head at reservoir r at time step t; $P_{b,t}$ = power produced in the system by pump b at time step t; B = number of pumps; T = number of tanks; $T^f_{k,t}$ = flow from the system used to fill tank k at time step t; T^s_k = flow being supplied to the system by tank k at time step t; T^l_k = level of tank k at time step t; $T^{\min}_{k,t}$ = minimum level of tank k at time step t; R_{tot} = total resilience index for all the time steps; and N = total number of time steps.

Cost Function

The metric representing capital cost that accounts for the installation of pipe sections and pump stations within the network plays an integral part as to how the generated network will be financially optimized. The most elemental parameters that affect the cost turned out to be (1) the length and the diameter of the pipes, (2) pump operational specifications, and (3) tank volumes. The cost of each solution candidate includes the capital costs of pipes, pumps, and tanks as well as the present value of the energy consumed over a specified period (i.e., 20 years). Pump station operating costs are obtained according to a unit cost for energy, which was assumed to be constant throughout a 24-h period, equalling \$0.12/kWh (C^e) (Walters et al. 1999). The aforementioned worth of energy costs takes on an interest rate of 12% and an amortization period of 20 years (Walters et al. 1999). Tank costs were considered a function of unit volume in the cost function (Walters et al. 1999). Eq. (5) summarizes the aggregate cost function associated with the optimization objective function presented in this study as follows:

$$C = \sum_{x=1}^{m} L_x C^p_x + \sum_{k=1}^{T} C^t_k + \sum_{b=1}^{B} C^{pv} N^{op} C^e P_b \tag{5}$$

where

$$C^p_x = -10^{-7} D^3_x + 2 \times 10^{-4} D^2_x + 0.0186 D_x + 5.351 \tag{6}$$

$$C^t_k = -6 \times 10^{-6} V^2_k + 3.61 V_k + 68{,}800 \tag{7}$$

$$V_k = 7.48 T_k^{\max} \pi D_k^2 \quad (8)$$

where C = aggregate cost estimation for the entirety of the system components (\$); L_x = length of pipe x (m); C_x^p = pipe installation cost coefficient for pipe x; D_x = diameter of pipe x (mm); m = number of pipes in the system; C_k^t = tank installation cost coefficient for tank k; V_k = volume of tank k (m^3/h); $T_k^{\max}$ = maximum level of tank k; k = number of tanks in the system; D_k = diameter of tank k; $C^p v$ = present value factor, which equals 0.10367; C^e = electricity cost, which equals 0.12 \$/kW; N^{op} = number of hours of operations in a year, which equals 8,760; P_b = power of pump b (kW); and B = number of pumps in the system.

Decision Variables

Subsequent to the finalization of the generated network, design decision variables in the genetic algorithm (GA) optimization platform are assigned accordingly and are characterized in the optimization objective function as follows:

$$Z = f(R_e, T^l, T^{\min}, T^{\max}, U, D, P_a) \quad (9)$$

where

$$R_e = (r_{e,1}, r_{e,2}, \ldots, r_{e,j}) \quad (10)$$

$$T^l = (t_1^l, t_2^l, \ldots, t_k^l) \quad (11)$$

$$T^{\min} = (t_1^{\min}, t_2^{\min}, \ldots, t_k^{\min}) \quad (12)$$

$$T^{\max} = (t_1^{\max}, t_2^{\max}, \ldots, t_k^{\max}) \quad (13)$$

$$U = (u_1, u_2, \ldots, u_b) \quad (14)$$

$$D = (d_1, d_2, \ldots, d_m) \quad (15)$$

$$P_a = (p_{a,1}, p_{a,2}, \ldots, p_{a,b}) \quad (16)$$

where Z = objective function representation; $f()$ = combinative function of the hydraulic simulation and optimization scheme; R_e = elevation of reservoirs in the network; $r_{e,j}$ = elevation of the jth reservoir; j = number of reservoirs in the generated network; T^l = level of tanks in the network; t_k^l = level of the kth tank; $T^{\min}$ = minimum level of tanks in the network; $t_k^{\min}$ = minimum level of the kth tank; $T^{\max}$ = maximum level of tanks in the network; $t_k^{\max}$ = maximum level of the kth tank; k = number of tanks in the generated network; U = available pump curves for the network; u_b = assigned curve to the bth pump; b = number of pumps; D = set of diameters assigned to the pipes in the network; d_m = diameter of the mth pipe; P_a = pump pattern assigned to each of the pumps; and $p_{a,b}$ = pump pattern of the bth pump. The boundary conditions of each of the decision variables are also demonstrated in Table 2.

According to Table 2, for each pipe, the considered integer values for optimization platform input, i.e., 1–12, correspond to the 12 available discrete pipe diameters, respectively, accounting for 152.4 mm (6 in.), 203.2 mm (8 in.), 254 mm (10 in.), 304.8 mm (12 in.), 355.6 mm (14 in.), 406.4 mm (16 in.), 457.2 mm (18 in.), 508 mm (20 in.), 609.6 mm (24 in.), 762 mm (30 in.), 812.8 mm (32 in.), and 914.4 mm (36 in.).

Design Constraints

Table 3 summarizes the design constraints in the optimization of the generated networks. Specifically, the main constraint for the WDS design is to deliver at adequate amount of pressure that would satisfy customers' demands. The existence of storage tanks and pumps in the network complicate the optimization because these are the two most difficult components to model (Lansey et al. 1989). In this case, tank level variations were optimized according to (1) the optimal operations of pumps, (2) peak-hour-demand level considerations, (3) prevention of exhaustion or overfilling of storage tanks (thus avoiding closing the associated inlet/outlet pipe) to ensure their availability around the clock to be leveraged as surge tanks to avoid water hammer and excessive pressure transients, (4) imposition of a maximum range for initial and final tank levels over the 24-h cycle, (5) penalization of fluctuating tank levels on an hourly basis, and (6) imposition of the start point of storage tanks to fall within the middle third of the range of minimum and maximum tank levels.

The accumulated sum of the mismatch in levels is used as the tank operating level difference (TLD) constraint. Lastly, in order to maintain pump life span and decrease the maintenance costs, it was assumed that a reduction in the number of pump switches results in the reduction of the pump maintenance costs (Lansey and Awumah 1994). Furthermore, including subconstraints such as preventing a pump from running for a single hour improves the life cycle of the pump station collectively (Chen et al. 2021; Cimorelli et al. 2020).

Most importantly, it is vital to note that the most essential constraint was deemed to be the nodal minimum pressure to ensure the satisfaction of water demands; therefore, because pump and tank

Table 3. Design constraints

Constraint function	Constraint form	Constraint limit
NPS at average day flow	≥	172.6 kPa (25 psi)
NPS at instantaneous peak flow	≥	172.6 kPa (25 psi)
NPS at fire flow condition	≥	172.6 kPa (25 psi)
Maximum nodal pressure	≤	1,034.21 kPa (150 psi)
TLD at average day flow	=	0.0
Tank level	≠	Min/max tank capacity
Allowed number of pump switches	≤	4 switches
Minimum successive hours of pump operation	=	2 h

Note: NPS = nodal pressure shortfall; and TDL = tank operating level difference.

Table 2. Decision variable boundaries

Variable type	Lower bound	Upper bound
Pipe size	152.4 mm (6 in.)	914.4 mm (36 in.)
Pump design head	30.48 m (100 ft)	48.7 m (160 ft)
Pump design flow	1,817 m^3/h (8,000 gpm)	5,223.8 m^3/h (23,000 gpm)
Pump status	0 (Off)	1 (On)
Storage tank diameter	9.1 m (30 ft)	30.48 m (100 ft)
Storage tank minimum level	2.74 m (9 ft)	3.04 m (10 ft)
Storage tank maximum level	7.62 m (25 ft)	18.2 m (60 ft)
Reservoir elevation	18.2 m (60 ft)	60.9 m (200 ft)

Table 4. Randomized parameters

Randomized parameters	Lower boundary	Upper boundary
Nodal demands	30% of original values	Original values
Pipe length	30% of original values scaled to the start and end coordinates	80% of original values scaled to the start and end coordinates
Roughness coefficients	120	130
Number of tanks	1	2

operations and maneuvers are optimized simultaneously, some of their associated constraints might be considered as flexible restraints (i.e., marginally violating some of the constraints in Table 3) for a greater good to ensure a reliable water supply and optimized pressure heads.

Randomized Parameters

Apart from the parameters involved in the optimization platform, there are certain other geometric ones that are randomized to account more deeply for a real-world scenario (Kostrzewski 2020; Konstantinov et al. 2019). Table 4 characterizes such parameters along with their randomization range.

To maintain some of the features of the original network to mimic a realistic replication of the baseline system, the number of reservoirs and pumps was considered equal to those of the original network.

Demonstration: Anytown WDS Benchmark

This section aims at presenting a slightly modified Anytown WDS benchmark (Walski et al. 1987; Farmani et al. 2005) to (1) showcase the mass-generation of the input network in the generator model; and (2) investigate the performance of the optimization procedure implemented on each of 1,079 generated networks. The hydraulic and geometric design of Anytown WDS has been adopted from the literature (Farmani et al. 2005), which consists of 43 pipes, 22 nodes, one reservoir, two tanks, and three pumps.

It is noteworthy that a realistic demand pattern in the original version of Anytown WDS has been adopted from the literature and modified accordingly in order to guarantee a meaningful, representative generation. Fig. 3 characterizes the diurnal demand pattern for the original and generated networks because it remains constant through the generation-optimization procedure.

Results and Discussion

To account for all the generated-optimized networks (GONs) in this study, this section presents (1) a discussion of two sample cases of these optimized networks; and (2) tabulation of the main characteristics of the optimized-network pool.

Granular Analysis and Validation of Sample Optimized Networks

For assessing the robustness of the generator-optimizer platform, two sample GONs were randomly selected to be analyzed in terms of their hydraulic, geometric, and operation characteristics as well as their resilience index and cost evaluation.

Hydraulic and Operational Analysis

Fig. 4 depicts the skeleton layouts and collocations of the hydraulic components of two random samples out of the pool of over 1,000 optimized networks whose granular hydraulic and operational specifications are showcased in Figs. 5 and 6 in terms of (1) pressure and demand ranges of five sample nodes, (2) pump operations and scheduling over a 24-h cycle, and (3) tank level variations over a 24-h period.

As can be seen in Figs. 5 and 6, pressure values for both of the sample networks varied roughly between 172 kPa (25 psi) and 586 kPa (85 psi), and most of the pressure variations occurred around demand peak hours in the morning and evening. Moreover, storage tanks were designed to be both drafted around peak hours while experiencing a near-maximum level and filled during regular off-peak hours. The tank fill cycle was well-operated through optimal pump scheduling in each network during midnight hours and early afternoon hours when consumption was found to be at its minimum. This guarantees a maximum life cycle for pump stations.

Resilience and Cost Analysis

To assess the reliability aspects of the GONs, Table 5 displays the trade-off between the resilience index and cumulative installation and operational cost of each of the sample networks.

As can be viewed in Table 5, the resilience and cost values exemplify a fairly resilient and cost-effective operation of these networks, which can be the representatives of the pool of 1,079 GONs. A wider variety of feasible solutions including cheaper/costlier and more/less resilient networks may be produced provided that higher numbers of population size and generations are considered in the optimization framework.

Operational and Hydraulic Assessment of the Optimized Network Pool

Because over 1,000 layout-varied optimized networks have been produced in this study, this section aims at effectively summarizing the assessment and validation of the whole pool collectively. For comprehensively accounting for the validation of the entire pool of networks, this section is therefore categorized into three

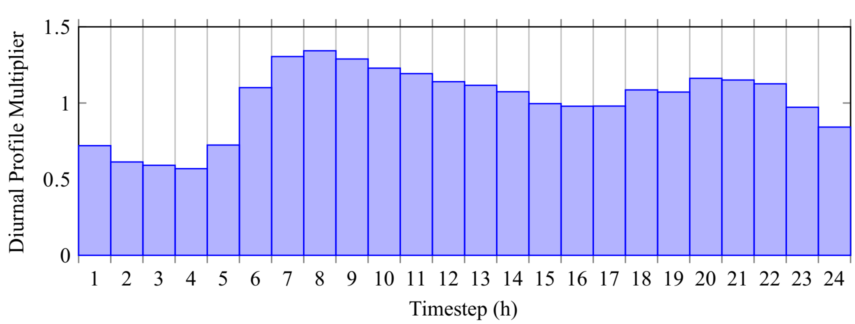

Fig. 3. Diurnal demand profile starting at 12:00 a.m. for original and generated networks.

J. Pipeline Syst. Eng. Pract., 2023, 14(1): 04022074

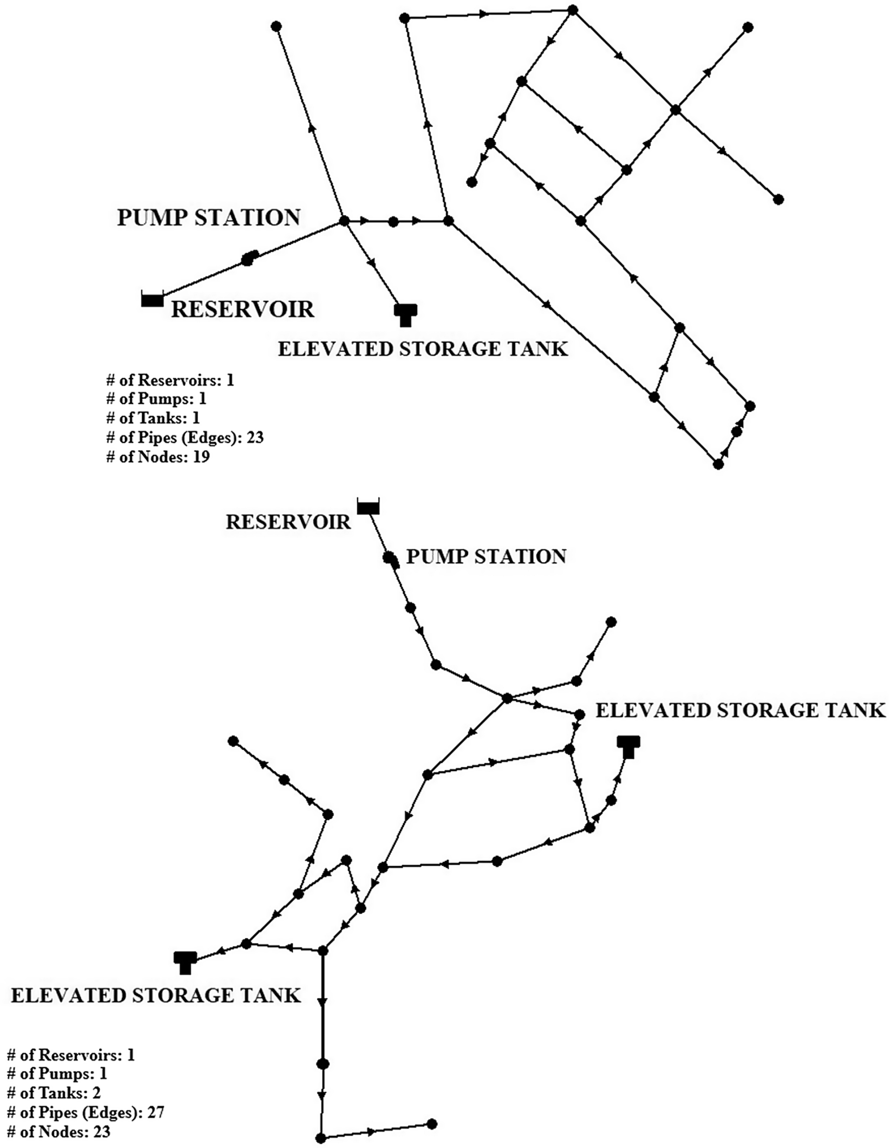

Fig. 4. Two sample generated-optimized network layouts.

subsections covering the assessment of robustness based on (1) distribution analysis of geometric specifications, (2) distribution analysis of hydraulic and operational characteristics, and (3) cost-resilience distribution assessment of the pool.

Geometric Assessment

In this section, six essential geometric characteristics of the GONs are assessed: (1) number of pipes, (2) number of demand nodes, (3) tank elevation/level, (4) demand node elevation, (5) pump power range, and (6) pipe diameter range.

As can be seen in Fig. 7(a), the majority of networks tended to have between 20 to 27 pipes compared with the 43 pipes that the original network contained. This deliberate reduction specifically accounts for the removal of unorthodox features of the original network such as triangular loops as well as degrees of connectivity of up to seven at a single node. Moreover, the pipe and node growth rate factors, listed in Table 1, along with the removal of pipe intersections yielded out of the generator, also contributed to this ultimate reduction of pipes and nodes in the network pool. According to Figs. 7(a and b), it is also noteworthy that the average node-to-pipe ratio of 0.866 in the network pool is indicative of the presence of redundancy loops in them.

Furthermore, according to Figs. 7(c and d), the tank elevation ranged from 36.57 m (120 ft) to 54.86 m (180 ft) for the majority of the networks compared with the corresponding demand node elevation range of 18.28 m (60 ft) to 30.48 m (100 ft). Such values are comparable to the original network tank elevations of 65.5 m (215 ft) and demand node elevation range of 22 m (72 ft). This also suggests that the tanks are designed to either be ESTs at normal level or GSTs at elevated areas for more proper supply of pressure and flow in the systems by including an appropriate hydraulic grade line in the associated pressure zone.

Also, Figs. 8(a and b) demonstrate average operational pump power ranges of 25 to 100 kW in the majority of the networks in the pool along with average pipe diameters of 500 mm (20 in.) to 650 mm (25.59 in.). This dominant concentration of frequency in these ranges suggests a well-consistent distribution of pipe diameter allocation, which both reduces constrictive pipe fittings and thus pressure transients and facilitates the procurement, installation,

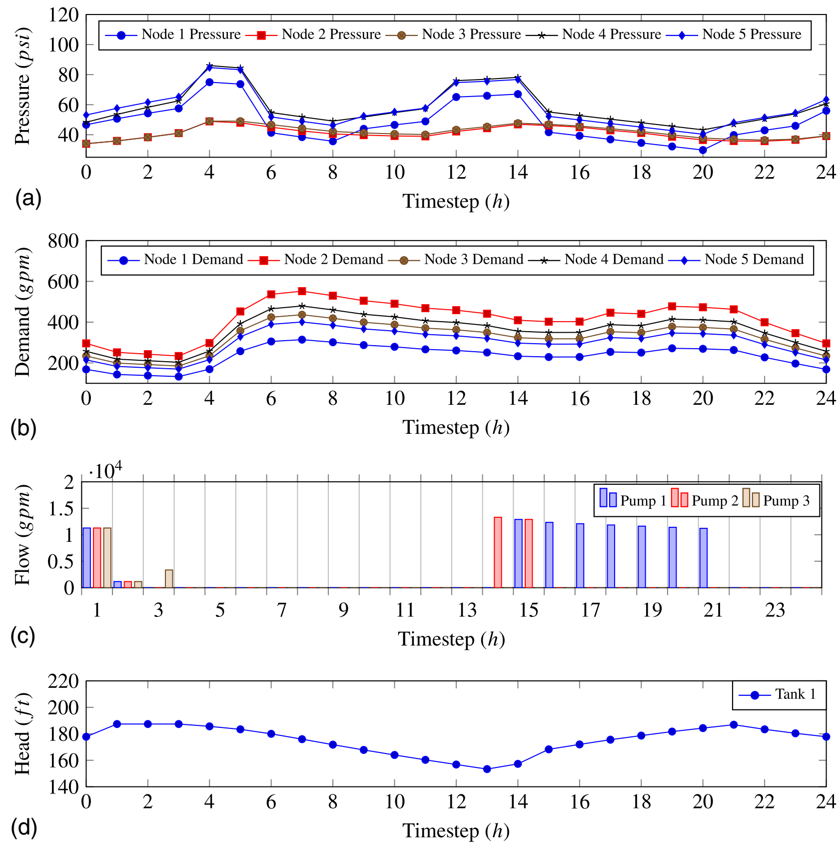

Fig. 5. Sample network 1's 24-h operational analysis: (a) pressure values; (b) demand values; (c) pump flow and scheduling; and (d) tank level variations.

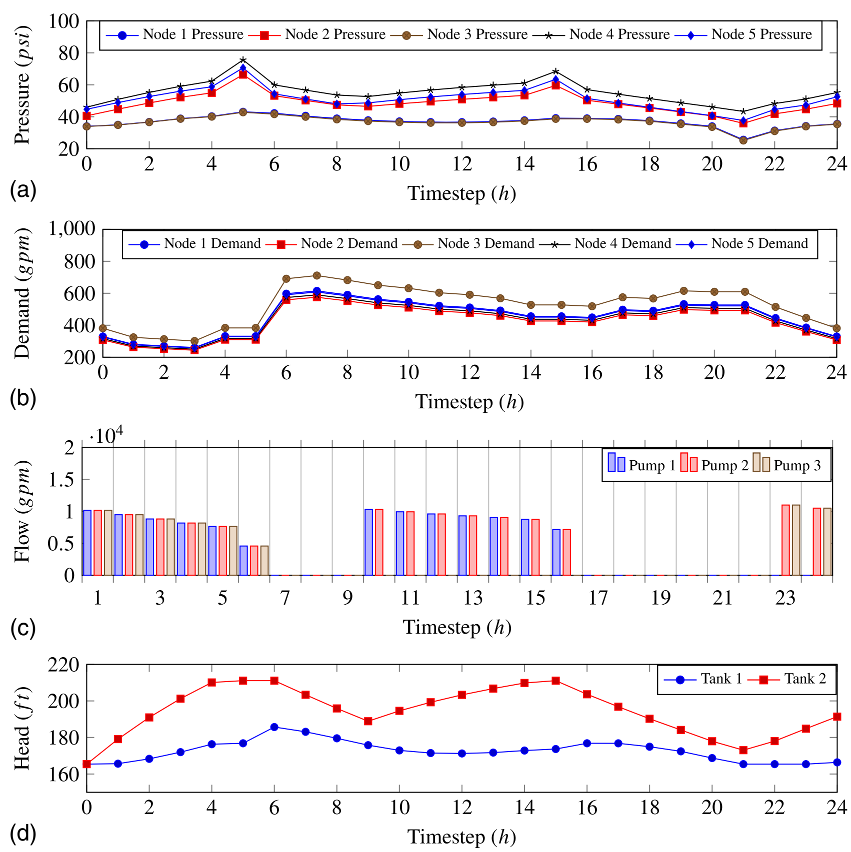

Fig. 6. Sample network 2's 24-h operational analysis: (a) pressure values; (b) demand values; (c) pump flow and scheduling; and (d) tank level variations.

Table 5. Cost-resilience analysis of the sample networks

Network ID	Resilience index	Estimated cost (millions of USD)
Sample network 1	0.792	44.9
Sample network 2	0.873	42.6

and maintenance of similar pipe sections and pump types in an associated municipality.

Operational and Hydraulic Assessment

The hydraulic functionality of the GONs was hereby assessed according to (1) nodal demands; and (2) nodal pressure, which constitute the most vital features a properly operational WDS must maintain optimally. In this respect, Figs. 8(c and d) demonstrates that the majority of nodal demands fell within 36.34 m^3/h [160 gallons per minute (gpm)] to 54.50 m^3/h (240 gpm) on average, and the nodal pressure values stoond between 275.79 kPa (40 psi) to 448.16 kPa (65 psi) on average. These values suggest normally operated networks by maintaining a minimum pressure boundary and making sure extreme pressure transients/spikes are restricted.

Cost-Resilience Assessment

Ultimately, Figs. 9(a and b) represent the ranges of the estimated costs and resilience index for the pool of data, where the majority of the former falls between \$10 million to \$40 million per the installation and operation of each of the networks, whereas the majority of the latter turns out to be between 0.65 and 0.95, both of which covered a wide span of both highly resilient and cost-effective networks.

Discussion: Utility and Variety of the GONs

The proposed generator-optimizer scheme in this study offers promise for a powerful tool for the research and industrial communities to replicate real-world WDS layouts and produce representative operationally optimized benchmarks for various purposes because this tool will be publicly available. One such application is applying the presented GONs as reliable synthetic case studies to the design, modeling, optimization, or rehabilitation studies of WDS. Because scarcity of real-world infrastructural data is evident in the research realm, the network variety that the proposed generator-optimizer platform offers instead provides ample flexibility to generate and optimize tailor-made networks. Such variety and flexibility include (1) tuning geometric parameters such as desired magnitude and ratios of available nodes or links and thus redundancy, (2) tuning operational parameters and fundamental component sizes including but not limited to nodal demands and node elevations, (3) tweaking optimization bounds and parameterization to produce more varied, feasible, and optimal solutions, and (4) fine-tuning the constrainability of the optimization scheme to include or exclude certain boundaries to the operational and geometric design at one's convenience.

The fundamental characteristics in these GONs are essentially preserved to be representative of the original layout; therefore, the convenience and leeway that the framework provides can be taken

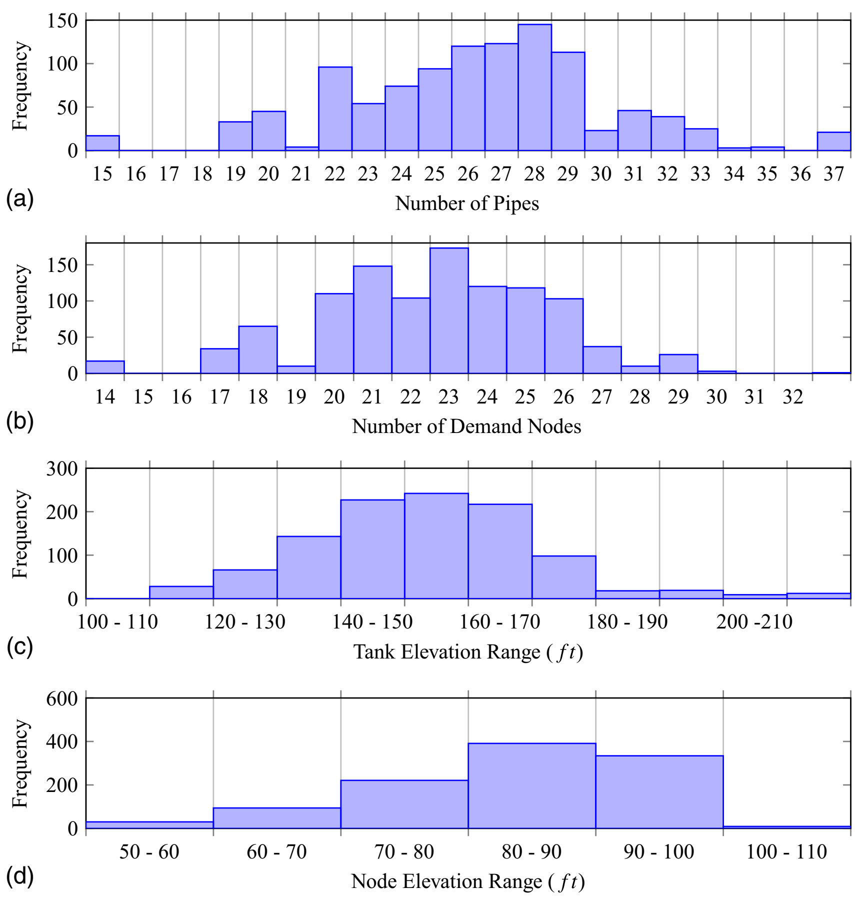

Fig. 7. Frequency of geometric values in the generated-optimized networks: (a) number of pipes; (b) number of demand nodes; (c) frequency of tank elevation ranges; and (d) frequency of the demand node elevation ranges.

J. Pipeline Syst. Eng. Pract., 2023, 14(1): 04022074

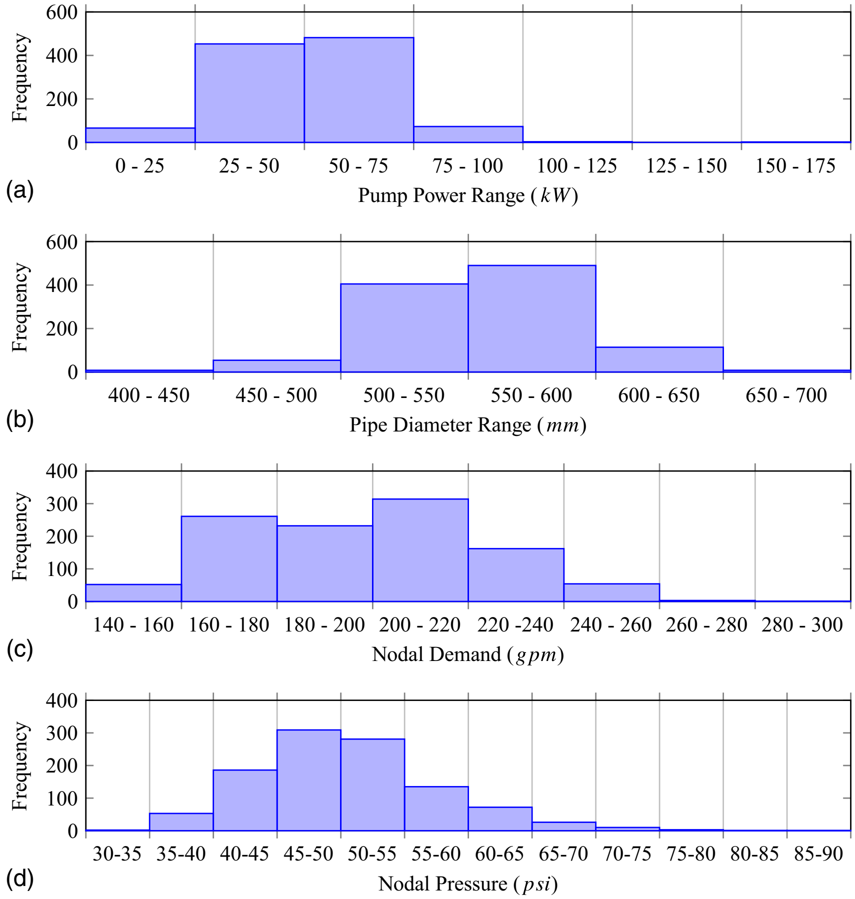

Fig. 8. Frequency of (a) range of pump power; (b) pipe diameter ranges; (c) nodal demand ranges; and (d) nodal pressure ranges.

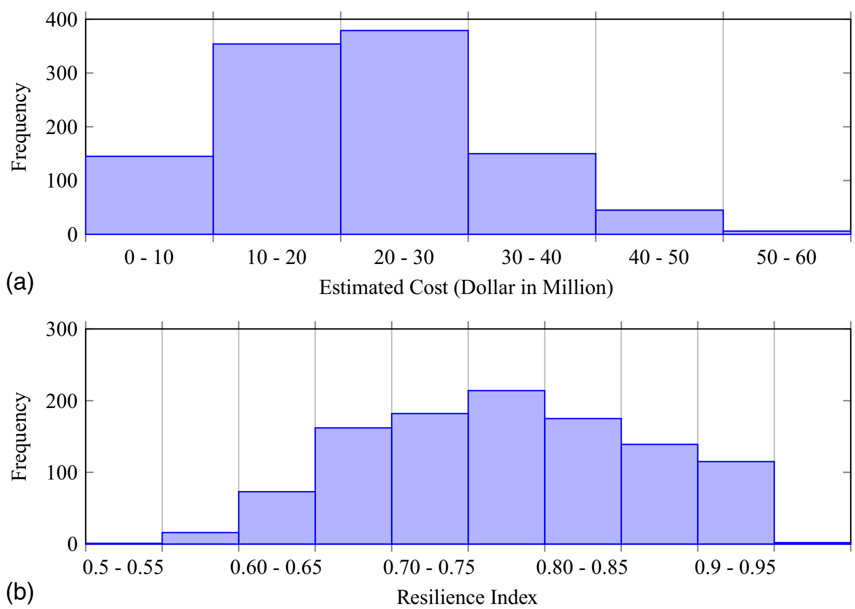

Fig. 9. Frequency of (a) estimated cost; and (b) resilience index.

advantage of to subject these GONs to specific case-study research purposes such as water quality modeling, reliability and resilience studies, or leakage localization and detection modeling. Another essential utility of the proposed platform turns out to be the significant number of varied networks it can produce at a fairly short amount of time. Most of the benchmark-included studies in the literature have utilized only a handful of renowned WDS benchmarks such as Anytown over the years. However, the proposed platform can generate a considerable number of mutually exclusive networks, which promotes the diversity and exclusivity of future research studies in the water industry.

Conclusions and Recommendations

This study presented an interactively robust platform for generating and optimizing synthetic water distribution systems by employing graph theory, the EPANET 2.2 software application, and genetic algorithms in Python to account for inaccessible, scarce, and often unreliable real-world infrastructural data. By leveraging a renowned WDS benchmark from the literature, this paper has (1) introduced a graph theoretic–based generative platform that synthesizes geometric, hydraulic, and operational characteristics of the given real-world WDS benchmark, (2) produced thousands of layout-varied graph-theoretic generated networks that characterize real-world applications and specifications of the original layout, (3) implemented operational and hydraulic optimization of each generated network through a robust and inclusive genetic algorithm framework in Linux using Python 2.7, and (4) provided public access for the research and industrial communities to a pool of over 1,000 generated-optimized networks whose geometric, hydraulic, and operational characteristics were assessed at granular and distributive levels.

Limitations of the study include (1) the fact that the current platform needs to be generalized for any given WDS benchmark, (2) assumption and randomization of some of the network features like node elevations and nodal demands that might affect the resilience and installation cost estimates, and (3) restriction of the node and edge growth and edit rates that could potentially produce more diverse network skeleton layouts. Future work also involves the generalization of the generative platform for other WDS benchmarks as well as reduction in the number of assumptions made in this study. It is essential that the future directions for this study involve a higher level of stochasticity as well as representation and robustness amid being exposed to any given real-world data.

Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request. Preliminary source code, data, and documentation are available at https://github.com/varsha2611/Water-Network-Generation-and-Optimization.

Acknowledgments

This research was supported by the National Science Foundation (NSF) under Grant No. 1745300. This study was also partly supported by King Saud University, Riyadh, Saudi Arabia, through Researchers Supporting Project No. RSP-2021/302. The results and conclusion presented in this paper are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Government. The authors are grateful to the NSF for this support.

References

Ahmad, N., M. Chester, E. Bondank, M. Arabi, N. Johnson, and B. L. Ruddell. 2020. "A synthetic water distribution network model for urban resilience." *Sustainable Resilient Infrastruct.* 7 (5): 1–15. https://doi.org/10.1080/23789689.2020.1788230.

Aiello, W., F. Chung, and L. Lu. 2001. "A random graph model for power law graphs." *Exp. Math.* 10 (1): 53–66. https://doi.org/10.1080/10586458.2001.10504428.

Barabási, A.-L., and R. Albert. 1999. "Emergence of scaling in random networks." *Science* 286 (5439): 509–512. https://doi.org/10.1126/science.286.5439.509.

Candelieri, A. 2017. "Clustering and support vector regression for water demand forecasting and anomaly detection." *Water* 9 (3): 224. https://doi.org/10.3390/w9030224.

Chauhan, V., A. Gutfraind, and I. Safro. 2019. "Multiscale planar graph generation." *Appl. Netw. Sci.* 4 (Jul): 46. https://doi.org/10.1007/s41109-019-0142-3.

Chen, W., T. Tao, A. Zhou, L. Zhang, L. Liao, X. Wu, K. Yang, C. Li, T. C. Zhang, and Z. Li. 2021. "Genetic optimization toward operation of water intake-supply pump stations system." *J. Cleaner Prod.* 279 (Jan): 123573. https://doi.org/10.1016/j.jclepro.2020.123573.

Choi, Y. H., H. M. Lee, J. Choi, D. G. Yoo, and J. H. Kim. 2019. "Development of practical design approaches for water distribution systems." *Appl. Sci.* 9 (23): 5117. https://doi.org/10.3390/app9235117.

Cimorelli, L., C. Covelli, B. Molino, and D. Pianese. 2020. "Optimal regulation of pumping station in water distribution networks using constant and variable speed pumps: A technical and economical comparison." *Energies* 13 (10): 2530. https://doi.org/10.3390/en13102530.

Costa, L. H. M., and G. P. W. Rodrigues. 2021. "Automatic generation of water distribution networks from streets layout." *Water Resour. Manage.* 35 (4): 1299–1319. https://doi.org/10.1007/s11269-021-02785-8.

Creaco, E., A. Campisano, M. Franchini, and C. Modica. 2017. "Unsteady flow modeling of pressure real-time control in water distribution networks." *J. Water Resour. Plann. Manage.* 143 (9): 04017056. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000821.

de Corte, A., and K. Sörensen. 2014. "Hydrogen: An artificial water distribution network generator." *Water Resour. Manage.* 28 (2): 333–350. https://doi.org/10.1007/s11269-013-0485-y.

Diao, K. 2020. "Multiscale resilience in water distribution and drainage systems." *Water* 12 (6): 1521. https://doi.org/10.3390/w12061521.

Di Palma, F., R. Gargano, F. Granata, and R. Greco. 2017. "The overall pulse model for water demand of aggregated residential users." *Procedia Eng.* 186 (May): 483–490. https://doi.org/10.1016/j.proeng.2017.03.260.

Fan, W., M. Liu, C. Tian, R. Xu, and J. Zhou. 2020. "Incrementalization of graph partitioning algorithms." *Proc. VLDB Endowment* 13 (8): 1261–1274. https://doi.org/10.14778/3389133.3389142.

Farmani, R., G. A. Walters, and D. A. Savic. 2005. "Trade-off between total cost and reliability for Anytown water distribution network." *J. Water Resour. Plann. Manage.* 131 (3): 161–171. https://doi.org/10.1061/(ASCE)0733-9496(2005)131:3(161).

Farmani, R., G. A. Walters, and D. A. Savic. 2006. "Evolutionary multi-objective optimization of the design and operation of water distribution network: Total cost vs. reliability vs. water quality." *J. Hydroinf.* 8 (3): 165–179. https://doi.org/10.2166/hydro.2006.019b.

Giudicianni, C., A. Di Nardo, M. Di Natale, R. Greco, G. F. Santonastaso, and A. Scala. 2018. "Topological taxonomy of water distribution networks." *Water* 10 (4): 444. https://doi.org/10.3390/w10040444.

Gutfraind, A., I. Safro, and L. A. Meyers. 2015. "Multiscale network generation." In *Proc., 18th Int. Conf. on Information Fusion: FUSION '15*, 158–165. New York: IEEE.

Hallmann, C., and S. Kuhlemann. 2018. "Model generator for water distribution systems." In *Proc., Operations Research Proceedings 2017*, edited by N. Kliewer, J. F. Ehmke, and R. Borndörfer, 245–251. Cham, Switzerland: Springer.

Hu, Z., B. Chen, W. Chen, D. Tan, and D. Shen. 2021. "Review of model-based and data-driven approaches for leak detection and location in water distribution systems." *Water Supply* 21 (7): 3282–3306. https://doi.org/10.2166/ws.2021.101.

Huzsvár, T., R. Wéber, and C. J. Hős. 2019. "Analysis of the segment graph of water distribution networks." *Period. Polytech. Mech. Eng.* 63 (4): 295–300. https://doi.org/10.3311/PPme.13739.

Jacomy, M., T. Venturini, S. Heymann, and M. Bastian. 2014. "ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software." *PLoS One* 9 (6): e98679. https://doi.org/10.1371/journal.pone.0098679.

Jaskowski, Q., et al. 2012. "Genetic programming needs better benchmarks." In *Proc., 14th Annual Conf. Genetic and Evolutionary Computation*, 791–798. New York: Association for Computing Machinery. https://doi.org/10.1145/2330163.2330273.

Konstantinov, S., A. Diveev, G. Balandina, and A. Baryshnikov. 2019. "Comparative research of random search algorithms and evolutionary algorithms for the optimal control problem of the mobile robot." *Procedia Comput. Sci.* 150 (Apr): 462–470. https://doi.org/10.1016/j.procs.2019.02.080.

Kostrzewski, M. 2020. "Sensitivity analysis of selected parameters in the order picking process simulation model, with randomly generated orders." *Entropy (Basel)* 22 (4): 423. https://doi.org/10.3390/e22040423.

Lansey, K. E., et al. 1989. "Water distribution system design under uncertainties." *J. Water Resour. Plann. Manage.* 115 (5): 630–645. https://doi.org/10.1061/(ASCE)0733-9496(1989)115:5(630).

Lansey, K. E., and K. Awumah. 1994. "Optimal pump operations considering pump switches." *J. Water Resour. Plann. Manage.* 120 (1): 17–35. https://doi.org/10.1061/(ASCE)0733-9496(1994)120:1(17).

Lei, T., C. Luo, J. E. Ball, and S. Rahimi. 2020. "A graph-based ant-like approach to optimal path planning." In *Proc., 2020 IEEE Congress on Evolutionary Computation (CEC)*, 1–6. New York: IEEE. https://doi.org/10.1109/CEC48606.2020.9185628.

Lorenz, I.-S., and P. F. Pelz. 2020. "Optimal resilience enhancement of water distribution systems." *Water* 12 (9): 2602. https://doi.org/10.3390/w12092602.

Mair, M., W. Rauch, and R. Sitzenfrei. 2014. "Improving incomplete water distribution system data." *Procedia Eng.* 70 (Apr): 1055–1062. https://doi.org/10.1016/j.proeng.2014.02.117.

Menapace, A., A. Zanfei, M. Felicetti, D. Avesani, M. Righetti, and R. Gargano. 2020. "Burst detection in water distribution systems: The issue of dataset collection." *Appl. Sci.* 10 (22): 8219. https://doi.org/10.3390/app10228219.

Momeni, A., and K. R. Piratla. 2021. "A proof-of-concept study for hydraulic model-based leakage detection in water pipelines using pressure monitoring data." *Front. Water* 3 (Aug): 93. https://doi.org/10.3389/frwa.2021.648622.

Moslehi, I., M. Jalili-Ghazizadeh, and E. Yousefi-Khoshqalb. 2021. "Developing a framework for leakage target setting in water distribution networks from an economic perspective." *Struct. Infrastruct. Eng.* 17 (6): 821–837. https://doi.org/10.1080/15732479.2020.1777568.

Muranho, J., A. Ferreira, J. Sousa, A. Gomes, and A. Sá Marques. 2012. "WaterNetGen: An EPANET extension for automatic water distribution network models generation and pipe sizing." *Water Supply* 12 (1): 117–123. https://doi.org/10.2166/ws.2011.121.

Newman, M. 2010. *Networks: An introduction*. New York: Oxford University Press.

Ocampo-Martinez, C., V. Puig, G. Cembrano, R. Creus, and M. Minoves. 2009. "Improving water management efficiency by using optimization-based control strategies: The Barcelona case study." *Water Sci. Technol. Water Supply* 9 (5): 565–575. https://doi.org/10.2166/ws.2009.524.

Odan, F. K., et al. 2015. "Real-time multiobjective optimization of operation of water supply systems." *J. Water Resour. Plann. Manage.* 141 (9): 04015011. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000515.

Penschuck, M., U. Brandes, M. Hamann, S. Lamm, U. Meyer, I. Safro, P. Sanders, and C. Schulz. 2020. "Recent advances in scalable network generation." Preprint, submitted March 2, 2020. http://arxiv.org/abs/2003.00736.

Prasad, T. D., and T. T. Tanyimboh. 2008. *Entropy based design of "Anytown" water distribution network*. Reston, VA: ASCE.

Ron, D., I. Safro, and A. Brandt. 2011. "Relaxation-based coarsening and multiscale graph organization." *Multiscale Model. Simul.* 9 (1): 407–423. https://doi.org/10.1137/100791142.

Safro, I., D. Ron, and A. Brandt. 2006. "Graph minimum linear arrangement by multilevel weighted edge contractions." *J. Algorithms* 60 (1): 24–41. https://doi.org/10.1016/j.jalgor.2004.10.004.

Safro, I., and B. Temkin. 2011. "Multiscale approach for the network compression-friendly ordering." *J. Discrete Algorithms* 9 (2): 190–202. https://doi.org/10.1016/j.jda.2010.09.007.

Sanders, P., and C. Schulz. 2013. "Think locally, act globally: Highly balanced graph partitioning." In Vol. 7933 of *Proc., 12th Int. Symp.: Experimental Algorithm*, 164–175. Rome: Springer.

Shi, S., Y. Ge, L. Chen, and H. Feng. 2020. "Four-objective optimization of irreversible Atkinson cycle based on NSGA-II." *Entropy (Basel)* 22 (10): 1150. https://doi.org/10.3390/e22101150.

Shin, S., S. Lee, S. J. Burian, D. R. Judi, and T. McPherson. 2020. "Evaluating resilience of water distribution networks to operational failures from cyber-physical attacks." *J. Environ. Eng.* 146 (3): 04020003. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001665.

Sitzenfrei, R., S. Fach, M. Kleidorfer, C. Urich, and W. Rauch. 2010. "Dynamic virtual infrastructure benchmarking: Dynavibe." *Water Sci. Technol. Water Supply* 10 (4): 600–609. https://doi.org/10.2166/ws.2010.188.

Sitzenfrei, R., M. Mair, M. Möderl, and W. Rauch. 2011. "Cascade vulnerability for risk analysis of water infrastructure." *Water Sci. Technol.* 64 (9): 1885–1891. https://doi.org/10.2166/wst.2011.813.

Sitzenfrei, R., M. Möderl, and W. Rauch. 2013. "Automatic generation of water distribution systems based on GIS data." *Environ. Modell. Software* 47 (Sep): 138–147. https://doi.org/10.1016/j.envsoft.2013.05.006.

Todini, E. 2000. "Looped water distribution networks design using a resilience index based heuristic approach." *Urban Water* 2 (2): 115–122. https://doi.org/10.1016/S1462-0758(00)00049-2.

Tuptuk, N., P. Hazell, J. Watson, and S. Hailes. 2021. "A systematic review of the state of cyber-security in water systems." *Water* 13 (1): 81. https://doi.org/10.3390/w13010081.

Walski, T., et al. 1987. "Battle of the network models: Epilogue." *J. Water Resour. Plann. Manage.* 113 (2): 191–203. https://doi.org/10.1061/(ASCE)0733-9496(1987)113:2(191).

Walters, G. A., D. Halhal, D. Savic, and D. Ouazar. 1999. "Improved design of 'Anytown' distribution network using structured messy genetic algorithms." *Urban Water* 1 (1): 23–38. https://doi.org/10.1016/S1462-0758(99)00005-9.

Zeng, F., X. Li, and K. Li. 2017. "Modeling complexity in engineered infrastructure system: Water distribution network as an example." *Chaos* 27 (2): 023105. https://doi.org/10.1063/1.4975762.

Zhang, Q., Z. Y. Wu, M. Zhao, J. Qi, Y. Huang, and H. Zhao. 2017. "Automatic partitioning of water distribution networks using multiscale community detection and multiobjective optimization." *J. Water Resour. Plann. Manage.* 143 (9): 04017057. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000819.

Zhang, Q., F. Zheng, Z. Kapelan, D. Savic, G. He, and Y. Ma. 2020. "Assessing the global resilience of water quality sensor placement strategies within water distribution systems." *Water Res.* 172 (Apr): 115527. https://doi.org/10.1016/j.watres.2020.115527.

Zhou, Y., Y. Shi, H. Wu, Y. Chen, Q. Yang, and Z. Fang. 2020. "Evaluation method for water network connectivity based on graph theory." In *Mobile wireless middleware, operating systems and applications*, edited by W. Li and D. Tang, 110–118. Cham, Switzerland: Springer.

Zverovich, V. 2021. *Modern applications of graph theory.* Oxford, UK: Oxford University Press.