Environmental Modelling & Software 65 (2015) 1–14

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Automated sub-zoning of water distribution systems

Lina Sela Perelman [a, *], Michael Allen [b], Ami Preis [c], Mudasser Iqbal [c], Andrew J. Whittle [a]

[a] Department of Civil and Environmental Engineering, MIT, Cambridge, MA, USA
[b] Faculty of Engineering and Computing, Coventry University, UK
[c] Singapore-MIT Alliance for Research and Technology, Singapore

ARTICLE INFO

Article history:
Received 27 February 2014
Received in revised form
24 November 2014
Accepted 26 November 2014
Available online

Keywords:
Water distribution systems
Community structure
Graph clustering and partitioning

ABSTRACT

Water distribution systems (WDS) are complex pipe networks with looped and branching topologies that often comprise thousands to tens of thousands of links and nodes. This work presents a generic framework for improved analysis and management of WDS by partitioning the system into smaller (almost) independent sub-systems with balanced loads and minimal number of interconnections. This paper compares the performance of three classes of unsupervised learning algorithms from graph theory for practical sub-zoning of WDS: (1) Global clustering – a bottom-up algorithm for clustering n objects with respect to a similarity function, (2) Community structure – a bottom-up algorithm based on the property of network modularity, which is a measure of the quality of network partition to clusters versus randomly generated graph with respect to the same nodal degree, and (3) Graph partitioning – a flat partitioning algorithm for dividing a network with n nodes into k clusters, such that the total weight of edges crossing between clusters is minimized and the loads of all the clusters are balanced. The algorithms are adapted to WDS to provide a practical decision support tool for water utilities. Visual qualitative and quantitative measures are proposed to evaluate models' performance. The three methods are applied for two large-scale water distribution systems serving heavily populated areas in Singapore.

1. Introduction

The water sector worldwide is facing growing challenges in providing adequate quantities and quality of water to a continuously growing population. The finite water resources, climate change, and population growth exercise an increasing impact on water resources and the water industry. To address this problem, various optimization-based formulations have been developed both at a regional and local scales. At regional scale, these include the design and operation of water systems under future demand and water availability uncertainty (Chung et al., 2009; Housh et al., 2013), creation of new water resources by producing high quality water from saline or brackish water (Avni et al., 2013) and waste water reclamation (Zhang et al., 2013). At local scale, previous works focus on water distribution system design under demand (Babayan et al., 2005; Perelman et al., 2013) and hydraulic model uncertainty (Fu and Kapelan, 2011; Laucelli et al., 2012), operation for leakage control (Giustolisi et al., 2008; Ulanicki et al., 2008; Price and Ostfeld, 2014), and reduction of potable water demand by on-site graywater reuse (Penn et al., 2013).

The creation of new water resources and re-use of waste water involve high economic costs and environmental impacts, hence, conservation of water through efficient end-use and active loss-control have attracted much interest both in research and practice (Mutikanga et al., 2013). Water conservation, traditionally tends to focus largely on the end user (e.g. installing water efficient fixtures in the home and the workplace (Kunkel, 2003)). Whereas, water utilities traditionally operate without consistent standards for water accounting and water loss control.

Water loss is the difference between the volume of water input into the system and all the authorized billed and unbilled (e.g. firefighting) water consumption (Kingdom et al., 2006). Water losses are characterized by real and apparent losses. Real losses are physical losses (e.g. leakages, bursts, tank overflows) that represent a waste of water resources. Apparent losses include meter inaccuracy, billing errors, and unauthorized use. Water losses, both real and apparent, constitute a major inefficiency in water supplies because water and energy resources are wasted, operating costs are increased, and water revenue is reduced. Water loss control requires a wide range of technologies supporting both *re-active* and

* Corresponding author.
E-mail address: linasela@mit.edu (L. Sela Perelman).

http://dx.doi.org/10.1016/j.envsoft.2014.11.025
1364-8152/© 2014 Elsevier Ltd. All rights reserved.

pro-active approaches to reduce water losses including: (1) monitoring, (2) detection, (3) localization, (4) response, (5) pressure management, (6) leakage management, and (7) demand management.

Network sub-zoning is one of the tools for leakage and pressure management for water loss control. The requirement of sub-zoning is to define the properties of the sub-zones within a network (e.g. size limit, total demand), to identify their boundaries (i.e. pipes or valves), and to monitor these boundaries for leakage and/or pressure control (with a limited number of meters). For example, the management of district metered areas (DMAs), has proven highly successful for leakage management (Thornton et al., 2008; Kunkel, 2003). A DMA for a water distribution system is a specifically defined area, in which the quantities of water entering and leaving the district are metered (Morrison, 2004). The subsequent analysis of flow calculates the level of leakage within the district. According to Kunkel (2003) up to 85% of the measured leakage in the UK has been eliminated through a national water loss control program based on DMA's.

Recently, network sub-zoning has attracted many researchers with a variety of applications. Zheng et al. (2013) and Zheng and Zecchin (2014) utilized network decomposition for an optimal design of water distribution systems. The full network is decomposed into sub-networks followed by a solution of a set of subproblems representing each of the sub-networks using evolutionary algorithms. Perelman and Ostfeld (2011) and Deuerlein (2008) used graph decomposition methods to analyze network structure and connectivity. Diao et al. (2014) used network decomposition method to accelerate the hydraulic simulation process by subdividing the network into smaller sub-networks and solving the hydraulic equations of each of the sub-networks independently. Ulanicki et al. (2008) applied pressure management schemes to DMA by scheduling the set-point (output) pressures of boundary pressure relief valves which control the inlet pressures to the DMAs. Low operational pressures result in reduced leakage and minimization of the risk of bursts. Furnass et al. (2013) developed a data-driven methodology to identify cause–effect linkages of known water quality anomalies through mining the large volumes of water quality, hydraulic and asset data collected by water utility companies utilizing network internal partition having few inlets and outlets.

This work focuses on network sub-zoning as a mitigation tool for water scarcity by facilitating water loss control. Urban water distribution systems (WDS) can reach a substantial size of hundreds to thousands of nodes (i.e. consumers) and links (i.e. pipes, valves). The layout of WDS is typically looped having multiple flow paths from the water sources to consumers. The looped layout of WDS, which provides a high level of reliability to the system supply in the event of mechanical failures (e.g. pipe breaks, valves malfunctions), imposes difficulties on water loss control. Due to the complexity of WDS, the re-design of an existing network can impair water supply, system reliability, and water quality (Grayman et al., 2009). A number of methods for re-designing existing WDS into independent areas, by the closure of existing valves or disconnection of pipes, have been suggested. These vary from manual trial and error approaches, involving identification of water mains, manual division into districts, and hydraulic simulations (Murray et al., 2010), to highly sophisticated automated tools integrating network analysis, graph theory and optimization methods. The partition of the network is typically achieved by using graph algorithms, e.g. breadth first search and depth first search (Deuerlein, 2008; Perelman and Ostfeld, 2011; Ferrari et al., 2014; DiNardo et al., 2014), multilevel partitioning (DiNardo et al., 2013), community structure (Diao et al., 2013), and spectral approach (Herrera et al., 2010). The selection of pipes that need to be disconnected is found by iterative procedures (Ferrari et al., 2014; Diao et al., 2013) or genetic algorithms (DiNardo et al., 2013, 2014).

Despite the recent developments, the application of proposed sub-zoning methods to real large-scale water distribution systems has been found to be generally limited (Mutikanga et al., 2013). Furthermore, there is a lack of consensual quantitative measures for evaluating system partition, hence the results are generally analyzed qualitatively.

This work presents: (1) a generic framework for simplifying the full-scale WDS by partitioning the system into smaller balanced sub-zones (within a range of specified size limits) with minimum number of inter-connecting pipes/valves and (2) qualitative and quantitative measures for evaluating the performance of the network decomposition models. Three types of unsupervised learning algorithms are compared: global clustering – representing a more naive approach given limited information about the WDS, community structure – adopted from social studies with similar previous application to WDS sub-zoning (Diao et al., 2013), and network partitioning – adopted from distributed computed and previous similar application (DiNardo et al., 2013). In graph theory, these algorithms aim at grouping similar or closely connected vertices such that the set of nodes in each group has better connections to the nodes belonging to the same group than to the remaining nodes in the network.

Following the position paper of Bennett et al. (2013) for characterizing the performance of environmental models, this paper is structured as follows: Section 2 introduces the graph theory methods; Section 3 defines visual and quantitative performance criteria; Section 4 demonstrates the methods and their performance using an illustrative example; Section 5 shows the application to two large-scale water distribution systems serving heavily populated areas in Singapore; Section 6 evaluates and compares the different methods based on the performance measures for sub-zoning; Section 7 summarizes current work and suggest direction for future research.

2. Methods

The application of the sub-zoning to WDS involves defining a full network model based on the available data, formulating a decomposition method based on the network graph, evaluating the performance of the method based on visual, qualitative and quantitative measures, and providing a practical decision support tool for water utilities.

Many of the processes in physical, cyber, and social systems are described by complex networks or graphs. The water distribution network can be naturally represented as a graph $G = G(V,E)$ over a set of vertices (nodes) V and a set of connecting edges (links) E, where the vertices represent consumers, sources, and tanks and the edges – pipes, pumps, and valves. The graph can be characterized by nodal weights $w_i, i \in V$ (e.g. demand, elevation), link weights $w_i, i \in V$ (e.g. diameter, flow), and an adjacency matrix A based on network topology. The graph division to clusters is designated by the set $C = (c_1, \ldots c_i, \ldots, c_k)$ where each node i uniquely belongs to one of the clusters $i \in c_k$. In this work, we adopt and adapt popular approaches from the three branches of graph theory to sub-zoning of WDS.

Clustering, community structure, and partitioning are closely related methods for understanding and analyzing complex systems, which have been extensively studied by a large interdisciplinary community over the past few years (Schaeffer, 2007; Fortunato, 2010). Generally, given a data set, the goal of these methods is to divide the data set into clusters such that the elements assigned to a particular cluster are similar or better connected in some predefined sense then to elements in other clusters. Global *clustering* is related to grouping sets of points which are close to each other, with respect to a measure of similarity defined for each pair of points. *Community* algorithms reveal the natural community structure using the concept of edge density (i.e. intra-clusters versus inter-cluster edges). Graph *partitioning* divides the graph into a predefined number of groups such that the number of inter-cluster edges is minimal. The methods differ by their required input, underlying objective, and output. The main features of the three classes of methods in graph-theory are summarized in Fig. 1 and are described next.

2.1. Global clustering

Global clustering is one of the traditional algorithms for clustering n objects with respect to a similarity function (Hastie et al., 2009). It produces a multi-level or