Journal of Water Resources Planning and Management

Optimal Rehabilitation of Water Distribution Systems using a Cluster-based Technique

--Manuscript Draft--

Manuscript Number:	WRENG-2713
Full Title:	Optimal Rehabilitation of Water Distribution Systems using a Cluster-based Technique
Article Type:	Technical Paper
Abstract:	Optimal rehabilitation of large water distribution system (WDS) with many decision variables is often time-consuming and computationally expensive. This paper presents a new optimal rehabilitation methodology for WDSs based on graph theory clustering concept. The methodology starts with partitioning the WDS based on its connectivity properties into a number of clusters (small sub-systems). Pipes which might have direct impact on system performance are identified and considered for rehabilitation problem. Three optimisation-based strategies are then considered for pipe rehabilitation in the clustered network: 1) rehabilitation of some of the pipes inside the clusters; 2) rehabilitation of pipes in the path supplying water to the clusters; 3) combination of strategies 1 and 2. In all optimisation strategies, the decision variables are the diameters of duplicated pipes; the objective functions are to minimise the total cost of duplicated pipes and to minimise the number of nodes with pressure deficiency. The performance of proposed strategies was demonstrated in a large WDS with pressure deficiencies. The performance of these strategies were also compared to the full search space optimisation strategy and engineering judgement based optimisation strategy in which all pipes or selection of pipes are considered as decision variables respectively. The results show that strategy 3 is able to provide the best Pareto optimal front. The results also demonstrate that the cluster-based approach can significantly reduce the computational efforts for achieving optimum rehabilitation compared to the other optimization strategies.

Optimal Rehabilitation of Water Distribution Systems using a Cluster-based Technique

Karwan Muhmmed, Raziyeh Farmani*, Kourosh Behzadian, Kegong Diao, David Butler

Karwan Muhmmed, Research Graduate, Centre for Water systems, University of Exeter, Exeter, UK, E-mail: kam223@exeter.ac.uk (Lecturer, Irrigation Department, Sulaimani, Kurdistan Region, Iraq);

Raziyeh Farmani*, Associate Professor, Centre for Water systems, Department of Engineering, University of Exeter, Exeter, EX4 4QF, UK, E-mail: r.farmani@exeter.ac.uk

Kourosh Behzadian, Senior Lecturer, University of West London, London, UK, E-mail: kourosh.behzadian@uwl.ac.uk

Kegong Diao, Lecturer, De Montfort University, Leicester, UK, E-mail: k.diao@dmu.ac.uk

David Butler, Professor, Centre for Water Systems, University of Exeter, Exeter, EX4 4QF, UK, E-mail: d.butler@exeter.ac.uk

* Corresponding author, Email: r.farmani@exeter.ac.uk , Tel: +44 (0) 1392 723630

Optimal Rehabilitation of Water Distribution Systems using a Cluster-based Technique

Abstract

Optimal rehabilitation of large water distribution system (WDS) with many decision variables, is often time-consuming and computationally expensive. This paper presents a new optimal rehabilitation methodology for WDSs based on graph theory clustering concept. The methodology starts with partitioning the WDS based on its connectivity properties into a number of clusters (small sub-systems). Pipes which might have direct impact on system performance are identified and considered for rehabilitation problem. Three optimisation-based strategies are then considered for pipe rehabilitation in the clustered network: 1) rehabilitation of some of the pipes inside the clusters; 2) rehabilitation of pipes in the path supplying water to the clusters; 3) combination of strategies 1 and 2. In all optimisation strategies, the decision variables are the diameters of duplicated pipes; the objective functions are to minimise the total cost of duplicated pipes and to minimise the number of nodes with pressure deficiency.

The performance of proposed strategies was demonstrated in a large WDS with pressure deficiencies. The performance of these strategies were also compared to the full search space optimisation strategy and engineering judgement based optimisation strategy in which all pipes or selection of pipes are considered as decision variables respectively. The results show that strategy 3 is able to provide the best Pareto optimal front. The results also demonstrate that the cluster-based approach can significantly reduce the computational efforts for achieving optimum rehabilitation compared to the other optimization strategies.

Keywords*:* Water distribution systems; optimal rehabilitation; graph theory, clustering

Introduction

Growing water demand and ageing or inadequate infrastructure are some of the challenges that water distribution systems (WDS) are facing in a lot of countries. These challenges can lead in delivering water that does not satisfy some requirements such as minimum pressure, quality etc. or is delivered at high costs due to operational costs or water losses. Another challenges is financial constraints which do not allow major rehabilitations to be considered. Therefore a proper strategy for maintenance and rehabilitation needs to be developed to ensure an efficient and reliable operation. The strategy should be cost-effective while ensuring key WDS performance indicators (e.g. hydraulic, water quality and serviceability) are within required limits for current and future conditions. Due to the advancements in computer modelling tools and processing technologies, optimisation models have received a lot of attention for developing rehabilitation strategies in the recent decades (Deuerlein 2008). The key advantage of using optimisation models is their ability to consider a large number of decision variables for rehabilitation and efficiently search potential combinations of rehabilitation strategies (Savic and Banyard, 2011).

A large number of optimal rehabilitation strategies have been developed for WDS by many researchers and practitioners in the recent decades, in which a wide range of decision variables have been considered in the optimal rehabilitation models such as pipe rehabilitation (Kim and Mays, 1994; Giustolisi et al. 2006), tank sizing and sitting and pump operation schedules (Farmani et al. 2005a). Usually the problem is set as a multi-objective optimisation problem considering objectives such as minimising total capital and operation cost, leakage and maximising reliability and resilience (Kim and Mays, 1994; Farmani et al. 2005a; Fu et al. 2012; Wang et al. 2014). A trade-off exists between conflicting objectives in these optimisation problems which can be obtained, using multi-objective evolutionary algorithms (MOEAs), as a Pareto front of non-dominated solutions. Each solution in the Pareto front can represent an individual rehabilitation plan with specific objective values. Finding optimal Pareto front for a WDS with a large number of the potential pipes for rehabilitation is a major challenge due to a large size of decision space (Kadu et al. 2008). Different techniques have been employed to alleviate the complexity and computational burden of optimal rehabilitation problem such as path method (Kadu et al. 2008), global sensitivity analysis (Fu et al. 2012) and sequential multi-stage MOEAs (Rahmani et al. 2015).

Cluster based analysis is another efficient technique for reducing the complexity of water distribution system analysis. It divides a network into a number of sub-systems (i.e. clusters) with vertices and edges (Schaeffer, 2007). The resulting cluster structure simplifies the network layout and hence more explicitly reveals the network structure and interactions between components. Several clustering techniques have been applied to WDSs. Tzatchkov (2006) applied a depth-first and breadth-first based graph algorithm for WDS decomposition. Perelman and Ostfeld (2011) used the same algorithms to divide the system into strongly and weakly connected subgraphs according to the flow directions in pipes. Deuerlein (2008) developed a graph decomposition model that simplifies a network into a graph consisting of two main elements, called forests (tree structure) and cores (looped structure). Diao (2012) used modularity-based approach (Clauset et al., 2004; Newman, 2006) for WDS segmentation. Giustolisi and Ridolfi (2014) modified modularity-based approach by developing a new modularity index which was used in multiobjective optimisation in order to generate a variation of decomposition results for a WDS.

One of the main applications of the clustering-based decomposition is for district metered areas (DMAs) planning (Sempewo et al., 2008; Di Nardo and Di Natale, 2011a, 2011b; Fernandez, 2011; Scibetta et al., 2013; Ferrari et al., 2013; Diao et al., 2012). Swamee and Sharma (1990, 2008) developed a method for decomposing a multi-source WDS with predefined locations and influence zones of all water sources into single-source subsystems, which can be separately designed and then linked together. Using a similar approach, Zheng and Zecchin (2014) recently proposed an efficient network decomposition-based dual-stage multi-objective optimization method, in which each of decomposed independent subsystems is optimized individually and they are combined for an entire system optimization. Diao et al. (2015) proposed a twin-hierarchy decomposition to reformulate optimization of the whole WDS into that of backbone mains and communities. In their work as communities are independent from each other, their optimal design could be carried out individually. Other applications of clustering in water distributions systems include: cluster-based hydraulic computation (Zecchin et al., 2012; Diao et al., 2013), analysis of water quality events (Mandel et al. 2015), sensor placement (Perelman and Ostfeld, 2011), evaluation of redundancy (Yazdani and Jeffrey, 2010), vulnerability analysis (Kessler et al. 1990), and identification of the most critical pipes in a real-world WDS (Diao et al 2014).

Despite a plethora of recent advances of different clustering-based approaches applied to WDSs, to the best of the authors' knowledge, none of the previous works has been applied for optimum rehabilitation of WDSs. This paper presents a new methodology, based on the graph clustering and decomposition concepts (Schaeffer 2007; Fortunato 2010), for optimum rehabilitation of WDSs. The proposed methodology aims to substantially reduce the number of decision variables for optimisation by integrating hydraulic knowledge gained from each subsystem. The performance of a number of graph based optimisation strategies are analysed and compared in the rehabilitation of a water distribution system. Next sections present the proposed methodology and its application to a case study. The results are then presented and discussed. Finally, the key findings are summarized and future recommendations are made.

Methodology

The proposed methodology for rehabilitation of water distribution systems aims to include pipes which have direct impact on system performance as design variables in the optimisation process. This is deemed as a benefit in solving highly complex water distribution systems with large number of potential rehabilitation options. The methodology consists of two main stages. First, the network is partitioned based on its connectivity properties into a predefined number of clusters (sub-systems). In the second step, pipes which might have direct impact on system performance are identified and considered as design variables in the optimum rehabilitation of WDSs. Three different rehabilitation strategies are considered each with its own set of pipes as decision variables as described in the following section.

Problem Formulation

Pipe duplication is considered to be the only rehabilitation option here and thus the decision variables in the optimisation problem are pipe diameters. A two-objective optimisation problem is formulated for WDS rehabilitation where the objectives are: 1) minimisation of total capital cost of duplicated pipes (F_1) and 2) minimisation of the total number of demand nodes with pressure below minimum pressure requirement (F_2).

$$\text{Min } (F_1) = \sum_k C(D_k, R_k) \times L_k \qquad k\text{=1,..., number of duplicated pipes} \tag{1}$$

$$\text{Min } (F_2) = \sum_i N_i \qquad i\text{=1,…, number of demand nodes with pressure deficiency} \tag{2}$$

where $C(D_k, R_k)$ is the unit cost of pipe *k* which is function of diameter D_k and road type R_k (major or minor); L_k is the length of pipe *k*; N_i = Node *i* with pressure deficit (pressure below minimum pressure requirement).

Pressure deficit in an individual node is calculated as follows:

$$PD = \left| P_i - P_{\min} \right| \qquad \textit{if } \; P_i < P_{\min} \tag{3}$$

where PD is the pressure deficit at node *i*; P_i is the pressure at node *i*, $P_{\min}$ is the minimum pressure requirement. Note that pressure deficit is only calculated for demand nodes. In addition, in order to speed up the process of identifying hydraulically feasible rehabilitation solutions, a constraint is considered for nodes with negative pressures.

EPANET software (Rossman, 1999) is used for hydraulic simulation of the WDS. A GANetXL (Savic et al. 2011) based multi-objective optimisation algorithm is used to carry out optimisation of the system. GANetXL is an Excel-based add-in of NSGA-II (Deb et al. 2000). NSGA-II is a non-dominated sorting genetic algorithm which has been widely used to optimise large WDS.

WDS Clustering

A variety of approaches is available for network clustering. In this study, a modularity-based method (Clauset et al., 2004; Newman, 2006) is applied due to its competency for fast and reliable decomposition of large-scale complex systems.

The WDS clustering is conducted in two steps in this paper: 1) Mapping WDS into graph: the WDS is mapped into an undirected graph in which the vertices represent the consumers, sources, and tanks and the edges represent the connecting pipes, pumps, and valves (Perelman and Ostfeld, 2011). 2) Modularity-based clustering (Clauset et al., 2004; Newman, 2006) is used to divide the WDS graph into clusters with stronger internal connections than external connections (Figure 1). The modularity index, a metric to be maximised during clustering, is defined as:

$$Q = \frac{1}{2m}\sum\nolimits_{\upsilon\omega}\left[A_{\upsilon\omega} - \frac{k_\upsilon k_\omega}{2m}\right]\delta(c_\upsilon, c_\omega) \tag{3}$$

where $A_{\upsilon\omega}$ is an element of the adjacency matrix of the network; $k_\upsilon = \sum\nolimits_\omega A_{\upsilon\omega}$ is the sum of the number of edges connected to vertex υ; c_υ is the cluster to which vertex υ belongs, $\delta(c_\upsilon, c_\omega)$ is 1 if $c_\upsilon = c_\omega$, and 0 otherwise and $m = \frac{1}{2}\sum_{\upsilon\omega} A_{\upsilon\omega}$ is the number of edges in the graph.

A summary of the general concepts of WDS clustering used in this paper is described here but further details of this method and WDS application can be found in Clauset et al (2004), Blondel et al. (2008), and Diao et al (2012).

The clustering method is implemented using "Gephi", an open source and free software widely used for graph network visualization and manipulation (Gephi, 2014). First an input file of the WDS compatible with Gephi is generated. The level of decomposition, i.e. the number of clusters, is controlled by the "Resolution" parameter in the modularity settings. The default "Resolution" is set to one, higher values lead to fewer clusters and vice versa. The proper level of decomposition for any WDS analysis could be determined based on trial-and-error using different resolution values.

Rehabilitation Strategies

After clustering the network, a number of pipes are selected as potential pipes for rehabilitation. The pipes in the areas of the WDS which have no hydraulic performance issues, nor participate in water transmission to other areas, will have little contribution towards reducing or eliminating deficiency in the system. Hence, they are discarded from the pool of potential pipes for rehabilitation. As a result, the potential pipes for rehabilitation are selected from either inside a cluster with pressure deficiency, feed pipelines or pipes in the path between sources and those clusters. A feed pipeline is defined as a pipe which transports potable water between two clusters. The way that the potential pipes are selected for rehabilitation specifies one of the three strategies as described below.

Strategy 1: Rehabilitation of pipes within clusters

The aim of this strategy is to rehabilitate the potential pipes only located inside clusters with pressure deficiencies and inter-cluster water transmission. This is based on the assumption that if the pressure deficiency of the demand nodes is due to high pressure head loss of existing pipes (i.e. resulted from large pipe roughness or small pipe diameter) in close proximity, rehabilitation of those pipes will remove or alleviate pressure deficiencies. Hence, the pipes upstream of the nodes with pressure deficiency are considered as decision variables for rehabilitation taking into account their flow direction, capacity and length (e.g. considering a pipe with high flow rate with short length). In addition, a pipe which is the only feasible feed pipeline for a number of pipes in a tree network, either within a cluster or between clusters, is considered as a decision variable (e.g. a pipe that is the only link between a cluster with deficiency and the rest of the system).

Strategy 2: Rehabilitation of feed pipelines

The second strategy explores potential for rehabilitation of pipes in the path between source(s) and the clusters or feed pipelines between the clusters with pressure deficiencies. The assumption is that if the bottlenecks, which are causing pressure deficiencies in clusters, originate from pipes in the path between source/s and deficient cluster, rehabilitation of small number of pipes can cost-effectively address the problem. Note that the pipes in the path between a source and a cluster with deficiency is selected with respect to capacity and total length.

Strategy 3: Combination of strategies one and two

This strategy considers the impact of both aforementioned strategies. Thus, decision variables in this strategy are the potential pipes considered in strategies 1 and 2, i.e. pipes within the clusters with deficiency and the feed pipelines between these clusters and the source(s). Simultaneous consideration of rehabilitation of some paths and pipes within the clusters with performance deficiency may allow to identify the best combination of pipes that have the most contribution in reducing pressure deficiency in the system.

Case Study

The proposed rehabilitation strategies are demonstrated here in EXNET water distribution system as shown in Figure 2 (Farmani et al. 2005b; CWS, 2016). The network serves an approximately 400,000 customers and needs to be rehabilitated to meet the projected demands in 2020 and diminish pressure deficit. The network has 1891 nodes, 2462 pipes, two main reservoirs 3001 and 3002, with total heads of 58.4 and 62.421 m respectively. It also has five nodes (3003, 3004, 3005, 3006 and 3007) that supply water to the system from adjacent systems with base demand values of -63, -1388, -10.78, -926, -26.1 l/sec and elevation values of 11, 73.54, 30.5, 33 and 16 m respectively. The total system demand is 3245.81 l/sec. The main causes of deficiency in the network are relatively small pipe diameters; limited number of transmission mains; and high difference between elevations of demand nodes. Minimum nodal-pressure head has been set as 15m and the existing network has 534 nodes (28% of total demand nodes) with pressure deficiencies. The existing network is unable to satisfy future water demands (Farmani et al. 2005b).

The proposed methodology is applied to the EXNET water distribution system to identify the optimal rehabilitation solutions (with a trade-off between cost and number of nodes with pressure deficiency) that can meet future water demand.

Clustering EXNET WDS

One of the main objectives of optimum rehabilitation in this work is to minimize the overall pressure deficiency at demand nodes. Hence, understanding interactions between pressure deficient nodes and other components is critical. Based on clustering, this can be simplified to analyse interactions between clusters with pressure deficiency and other components of the system. An undirected graph is used in this work which represents the network topology and connectivity, without considering edge direction as a function of flow direction at a given time step (i.e. directed graph) (Perelman and Ostfeld, 2011).

A proper level of clustering is case specific. In this work the level of clustering is identified by trial-and-error considering the following characteristics, where possible:

1) clusters containing no nodes with deficiency,
2) clusters containing nodes with pressure deficiency between 0 and 15 m,
3) clusters containing nodes with pressure below zero,
4) minimum number of deficient clusters,
5) equal cluster sizes as much as possible (in terms of number of nodes).

The appropriate level of clustering for EXNET was 16 with different levels of pressure deficiency in the clusters as shown in Figure 3.

Setting of optimal rehabilitation problem

For each pipe duplication ten pipe diameter options are available, each with a specific pipe roughness coefficient and unit cost based on road type (i.e. major and minor road) (Farmani et al. 2005b). One additional option is defined as ‘do nothing if no duplication is required’. The total number of pipes which can potentially be considered as decision variables is 2462 and hence the size of full search space is equal to $11^{2462} = 8.11\times10^{2563}$. A total of 248 pipes were selected for rehabilitation (decision variables) for strategy 1 from the clusters with pressure deficiencies (Figure 4.

In strategy 2, a total of 149 pipes in the paths between the source and the clusters with pressure deficiencies and feed pipelines between deficient clusters were identified as decision variables as shown in Figure 5. The number of decision variables for strategy 3, which combines strategies 1 and

2, is 349 pipes including 248 pipes in strategy 1 plus 149 pipes in strategy 2 minus 48 pipes which were similar for the two strategies.

To carry out a fair comparison between the three strategies, a number of trial runs were conducted to identify the best parameter settings for the optimisation algorithm. Consequently, the following parameters were determined for all strategies: population size of 50; binary tournament selection operator; simple-by-gene mutation with the probability equal to the inverse of the length of decision variables corresponding to each strategy; and single-point crossover with the probability of 0.95. The optimisation algorithm was allowed to run for 10,000 generations.

Results and Discussion

Characteristics of the clusters

Table 1 summarizes the statistical and hydraulic characteristics (e.g. total number of nodes and pipes and initial percentage of deficiencies) of all the clusters. The number of decision variables considered for rehabilitation in the strategies is also shown. Percentage of deficiency in each cluster (column 5) was determined using the number of deficient nodes in the cluster (column 4) and the total number of nodes in the cluster. Clusters 0, 1, 4, 6, 13, and 14 are the most deficient clusters (percentage of deficiencies are more than 50%). Hence, as shown in Table 1, a high percentage of decision variables are considered for rehabilitation in those clusters for the strategy 1.

Results of Strategies 1, 2 and 3

Figure 6 shows the solutions on the Pareto-front generated for strategies 1, 2 and 3. It can be observed that although solutions on the Pareto-front for strategy 2 are generally dominating those for strategy 1, strategy 2 is unable to attain solutions with no pressure deficiency or with a small number of pressure deficiencies. Solutions on the Pareto-front for strategy 3 dominate solutions on the

Pareto-fronts for strategies 1 and 2. Comparison of the Pareto-fronts also shows that strategy 3 has the best performance in generating a solution with no pressure deficiency that has a very low cost.

Details of the cost, number of deficient nodes in the clusters, and percentage of pipes rehabilitated for the generated best solutions in each strategy are given in Table 2. Solution 1 of strategy 1 (Figure 6) has a total cost of £5.650 million and no demand node with pressure deficiency. This solution has more pipes reinforced in comparison with the other two solutions in strategies 2 and 3; 161 pipes (14 feed pipelines and 147 pipes inside the clusters) are duplicated which is 7% of the total pipes in the network. The six most deficient clusters, i.e. 0, 1, 4, 6, 13, and 14, contain larger number of rehabilitated pipes. Reinforcing feed pipelines between clusters demonstrates the hydraulic interaction between them. For instance, as shown in Figure 7, for this solution there is an interaction between clusters 8 and 14. Two feed pipelines have been reinforced (duplicated), indicating an increase in the capacity of pipes to transfer water from two water sources (e.g. reservoir 3001 and adjacent system through node 3004) in cluster 8 to cluster 14.

Solution 1 of strategy 2 (Figure 6) has a cost of £4.919 million with 13 nodes, in the cluster 11, not satisfying the minimum pressure requirement. This solution has fewer rehabilitated pipes than solution 1 of strategy 1 which has 47 duplicated pipes (12 feed pipelines and 35 pipes in the path between sources and clusters) which is 2% of the total pipes in the network.

Solution 1 of strategy 3 (Figure 6) is a solution with no pressure deficiency and has cost of £3.05 million which is cheaper than solutions 1 of strategies 1 and 2 by approximately 46% and 38% respectively. Figure 8 shows the pipe characteristics for this solution (i.e. pipe locations and diameters). A total of 76 pipes are duplicated including 6 feed pipelines between clusters and 70 pipes inside clusters (3% of the total pipes in the network) which is 50% less than the pipes rehabilitated in solution 1 of strategy 1. Most of the duplicated pipes have small diameter sizes of 110 mm (represented by least think lines in Figure 8).

It can be seen that clusters 13, 8 and 0 have a higher contribution (in terms of the number of duplicated pipes) 23, 10 and 10 pipes respectively towards reducing the level of deficiency in the network than other clusters. The cluster 13 has 23 nodes with pressure deficiency. This cluster is located at the downstream of the network at high elevation away from water sources and it is the only connection between the cluster 6 and the rest of network. Deficiencies of clusters 6 and 14 have been eliminated by reinforcing their adjacent clusters (i.e. clusters 13, and 8).

Most of the pipes in the paths between sources and clusters or feed pipelines, which have been rehabilitated for solution 1 of strategy 3 are located in the clusters with water sources (e.g. cluster 8) or in the clusters between water sources and deficient clusters. For instance in the cluster 8 there are two water sources, 3003 (reservoir) and 3004 (supplying water form adjacent systems) as shown in Figure 8. Two feed pipelines are reinforced, one for cluster 0 and another for cluster 1. One feed pipeline near the water source 3007 has been rehabilitated which increases system’s capacity towards the cluster 13.

4.3 Comparison of Strategy 3 with other Methods

In order to verify robustness of the performance of the strategy 3, two other methods were considered. The first methodology considers all the pipes as potential design variables for rehabilitation and second methodology considers a subset of the pipes (567 pipes represented by tick solid lines in figure 9) that were selected based on engineering judgement to address the network’s performance deficiency.

Figure 10 shows Pareto fronts generated by both methodologies which are dominated by the Pareto-front generated by strategy 3. The figure also shows that the solutions with no pressure deficiency found for these two methodologies are more expensive than the one identified in strategy 3. This demonstrates that the efficiency of the proposed methodology in identifying the optimum rehabilitation plans for large WDSs problem by reducing the size of the search space and therefore extensive computational efforts.

Conclusions

A new methodology for optimum rehabilitation of WDS, based on graph theory clustering, was proposed. The methodology uses the graph theory principles and algorithms for clustering the network into a predefined number of clusters (subsystems). The problem was posed as a multi-objective optimization problem with minimizing total cost and the number of demand nodes with pressure deficiency as the two main objective. The design variables were the diameter of duplicated pipes. Three optimisation-based strategies were considered for pipe rehabilitation in the clustered network: 1) rehabilitation of some of the pipes inside the clusters; 2) rehabilitation of pipes in the paths supplying water to the clusters; 3) combination of the strategies one and two

The results show that the methodology is able to identify the payoff characteristics between the total cost and the number of demand nodes with pressure deficiency. The strategy 3 generated a solution with minimum cost and no pressure deficiency in comparison with the other two strategies. The performance of the strategy 3 was assessed in comparison with two additional methodologies (i.e. whole search space and engineering judgement based optimisation strategies). The results indicate that the strategy 3 outperformed these two methodologies as well.

It can be concluded that using the cluster-based method helps to identify the most problematic areas in complex water networks, and reduce the complexity of redesign and rehabilitation problem. The methodology is most efficient in finding optimal solutions. Future improvement of the methodology could include considering various purposes of clustering of WDS (e.g. installing flow devices to measures and control flow) in modularity analysis before performing the rehabilitation strategies.

Acknowledgements

The work reported is supported by the UK Engineering & Physical Sciences Research Council (EPSRC) project Safe & SuRe (EP/K006924/1).

References

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

Bandes U., Eiglsperger M., Herman I., Himsolt M. and Marshall M. (2001). "GraphML progress report: structural layer proposal". International Symposium on Graph Drawing (GD '01), New York, NY: Springer-Verlag, 501-512.

Blondel, V.,Guillaume, J., Lambiotte, R. and Lefebvre, E. (2008). "Fast unfolding of communities in large networks". Journal of Statistical Mechanics: Theory and Experiment, 2008(10), 10008.

Blondel, V.,Guillaume, J., Lambiotte, R. and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), 10008.

Cheung, P.B., Reis, L.F., Formiga, K.T., Chaudhry, F.H. and Ticona, W.G., 2003, January. Multi-objective evolutionary algorithms applied to the rehabilitation of a water distribution system: A comparative study. In *Evolutionary Multi-Criterion Optimization* (pp. 662-676). Springer Berlin Heidelberg.

Clauset, A., Newman, M. and Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, 70(6), 066111.

Deuerlein, J. (2008). Decomposition model of a general water supply network graph. *Journal of Hydraulic Engineering*, 134(6), 822-832.

Diao, K., Zhou, Y. and Rauch, W. (2012). Automated creation of district metered area boundaries in water distribution systems. Journal of Water Resources Planning and Management, 139(2), 184-190.

Diao, K., Wang, Z., Burger, G., Chen, C., Rauch, W., and Zhou, Y. (2013). "Speedup of water distribution simulation by domain decomposition". Environmental Modelling & Software. 52, 253-263.

Diao, K., R. Farmani, G. Fu, M. Astaraie-Imani, S. Ward and D. Butler. (2014a). "Clustering analysis of water distribution systems: identifying critical components and community impacts". *Water Science & Technology*, 70(11), 1764-1776.

Diao, K. G., Fu, G. T., Farmani, R., Guidolin, M. and Butler D. (2015). "Twin Hierarchy Decomposition for Optimal Design of Water Distribution Systems". *Journal of Water Resources Planning and Management*, C4015008.

Di Nardo, A. and Di Natale, M. (2011). A heuristic design support methodology based on graph theory for district metering of water supply networks. *Engineering Optimization*, 43(2), 193-211.

Clauset A., Newman M. E. J. and Moore, C. (2004). "Finding community structure in very large networks". Physical Review E, 70(066111).

CWS, (2014).EXNET network benchmarks at centre for water systems, Available at: http://emps.exeter.ac.uk/engineering/research/cws/resources/benchmarks/expansion/exnet.php [Accessed 17 Jun. 2014].

Engelhardt, M., Skipworth, P., Savic, D., Saul, A. and Walters, G. (2000). Rehabilitation strategies for water distribution networks: a literature review with a UK perspective. *Urban Water*, 2(2), 153-170.

Farmani, R., Walters, G. A., and Savic, D. A. (2005a). "Trade-off between total cost and reliability for Anytown water distribution network." *J. Water Resource. Plann. Manage*., 131(3), 161-171

Farmani, R., Savic, D.A. and Walters, G.A., (2005b). Evolutionary multi-objective optimization in water distribution network design, *Engineering Optimization*, vol. 37, issue 2, pages 167-183

Ferrari, G., Savic, D. and Becciu, G. (2014). A Graph Theoretic Approach and Sound Engineering Principles for Design of District Metered Areas. *Journal of Water Resources Planning and Management*. 140(12),

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

Fu, G., Kapelan, Z. and Reed, P. (2012). Reducing the Complexity of Multi-objective Water Distribution System Optimization through Global Sensitivity Analysis. *Journal of Water Resources Planning and Management*, 138(3), 196-207.

Gephi.org, (2014). Gephi, an open source graph visualization and manipulation software. [online] Available at: https://gephi.org/ [Accessed 22 Jun. 2014].

Giustolisi, O., Laucelli, D., and Savic, D. A. (2006). "Development of rehabilitation plans for water mains replacement considering risk and cost-benefit assessment." *Civil Eng. and Environ. Syst*., 23(3), 175-190

Giustolisi, O. and Ridolfi, L. (2014). "New Modularity-Based Approach to Segmentation of Water Distribution Networks". Journal of Hydraulic Engineering, 140(10), 04014049.

Goldberg, D. (1989). *Genetic algorithms in search, optimization, and machine learning*. 1st ed. Reading, Mass.: Addison-Wesley Pub. Co.

Herrera Fernandez, A. M. (2011). Improving water network management by efficient division into supply clusters. PhD thesis in Hydraulic Engineering and Environmental Studies. [Tesis doctoral no publicada]. Universitat Politècnica de València. Available at: http://riunet.upv.es/handle/10251/11233 [Accessed 17 Jun. 2014].

Hartigan, J. and Wong, M. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C Applied statistics*, Vol. 28, No. 1, 100--108. Available: (stable URL from JSTOR): http://www.jstor.org/stable/2346830 [Accessed 17 Jun. 2014].

Kadu, M. S., Gupta, R., and Bhave, P. R. (2008). *"*Optimal design of water networks using a modified genetic algorithm with reduction in search space*." J. Water Resour. Plann. Manage.*, 134(2), 147-160.

Kessler, A., Ormsbee, L. and Shamir, U. (1990). A methodology for least-cost design of invulnerable water distribution networks. *Civil Engineering Systems*, 7(1), 20-28.

Kim, J. and Mays, L. (1994). Optimal rehabilitation model for water-distribution systems. *Journal of Water Resources Planning and Management*, 120(5), 674-692.

Lambiotte, R., Delvenne, J. and Barahona, M. (2008). Laplacian dynamics and multiscale modular structure in networks. *arXiv preprint arXiv*:0812.1770. 1-29.

Mandel, Pierre, Marie Maurel, and Damien Chenu. (2015): "Better Understanding Of Water Quality Evolution In Water Distribution Networks Using Data Clustering". *Water Research* 87. 69-78.

Newman, M. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 026113.

Newman, M. E. J. (2006). "Modularity and community structure in networks". PNAS, 103(23), 8577-8582.

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

Rahmani, F., Behzadian, K., Ardeshir A, (2015). "Rehabilitation of a water distribution system using sequential multiobjective optimization models." *Journal of Water Resources Planning and Management,* C4015003.

Rossman, L. (1999). The EPANET programmer's toolkit for analysis of water distribution systems. Erin M. Wilson (ed.) *29th Annual Water Resources Planning and Management Conference:* 1-10.

Savic, D. and Banyard, J. (eds.) (2011). *Water distribution systems*. 1st ed. London: ICE Pub.

Savić, D. A., Bicik, J., & Morley, M. S. 2011 A DSS Generator for Multiobjective Optimisation of Spreadsheet-Based Models. *Environmental Modelling and Software*, 26(5), 551-561

Schaeffer, S. (2007). Graph clustering. *Computer Science Review*, 1(1),27-64.

Shamir, U. and Howard, C. (1979). An Analytic Approach to Scheduling Pipe Replacement. *American Water Works Association*, 71(5), 248-258.

Tzatchkov V, Alcocer-Yamanaka V, Ortiz V (2006). Graph theory based algorithms for water distribution network sectorization projects. In 8th Annual Water Distribution Systems Analysis Symposium, 27–30 August, Cincinnati, Ohio (USA).

Walski, T. M. (2001). “The wrong paradigm–Why water distribution optimization doesn ’t work.” *Journal of Water Resources Planning and Management*, 127(4), 203 –205.

Wang, Q., Guidolin, M., Savic, D., & Kapelan, Z. (2014). Two-Objective Design of Benchmark Problems of a Water Distribution System via MOEAs: Towards the Best-Known Approximation of the True Pareto Front. *Journal of Water Resources Planning and Management.*

Yazdani, A. and Jeffrey, P. (2010). A complex network approach to robustness and vulnerability of spatially organized water distribution networks. *arXiv preprint arXiv*:1008.1770.

Zecchin, A.C., Thum, P., Simpson, A.R. and Tischendorf, C. (2012), Steady-state behavior of large water distribution systems: algebraic multigrid method for the fast solution of the linear step, Water Resources Planning and Management, Vol. 138, No. 6, Nov./Dec., doi: 10.1061/(ASCE)WR.1943-5452.0000226.

Table 1. Characteristics of clusters and pipes considered for rehabilitation for strategies 1 and 2

Cluster characteristics					**Pipes considered for rehabilitation**					
					Strategy 1			**Strategy 2**		
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Cluster	Total number of nodes	Total number of pipes	Number of nodes with pressure deficiency	**Deficient nodes (%)**	Number of feed pipelines between clusters	Number of pipes inside cluster	**Decision variables (%)**	Number of feed pipelines between clusters	Number of pipes in the path between sources and clusters	**Decision variables (%)**
C0	173	226	**112**	**64**	4	55	26	3	5	4
C1	61	75	**40**	**65**	4	13	23	3	2	7
C2	70	95	13	18	1	5	6	4	3	7
C3	134	172	0	0	1	0	0.6	1	10	6
C4	100	146	**89**	**89**	5	40	31	5	1	4
C5	191	256	2	1	1	0	0.4	5	10	6
C6	49	56	**34**	**69**	1	17	32	1	0	2
C7	66	92	7	10	1	6	8	3	2	5
C8	232	313	35	15	5	11	5	7	34	13
C9	89	118	0	0	0	0	0	0	0	0
C10	115	154	0	0	0	0	0	2	8	7
C11	111	149	15	13	2	12	9	2	2	3
C12	80	102	30	37	4	7	11	6	5	11
C13	104	79	**67**	**64**	3	35	48	3	9	15
C14	86	117	**57**	**66**	3	12	13	4	7	9
C15	232	315	33	14	5	17	7	6	19	8
Total	**1893**	**2465**	**534**	**28**						

Table 2. Solutions with minimum costs for strategies 1, 2 and 3

Cluster	Total number of nodes	Total number of pipes	**Strategy 1 (Solution 1)**					**Strategy 2 (Solution 1)**					**Strategy 3 (Solution 1)**				
			Total cost (£ million)				5.650	Total cost (£ million)				4.919	Total cost (£ million)				3.05
			Number of nodes with pressure deficiency	**Pressure deficiency (%)**	Number of feed pipelines between clusters	Number of pipes inside cluster	**Rehabilitated pipes (%)**	Number of nodes with pressure deficiency	**Pressure deficiency (%)**	Number of feed pipelines between clusters	Number of pipes in the path between sources and clusters	**Rehabilitated pipes (%)**	Number of nodes with pressure deficiency	**Pressure deficiency (%)**	Number of feed pipelines between clusters	Number of pipes inside cluster and pipes in the path between sources and clusters	**Rehabilitated pipes (%)**
C0	173	226	0	0	2	32	15	0	0	3	5	4	0	0	1	10	5
C1	61	75	0	0	1	7	11	0	0	1	2	4	0	0	1	0	1
C2	70	95	0	0	1	4	5	0	0	0	0	0	0	0	0	3	3
C3	134	172	0	0	0	0	0	0	0	0	1	0.6	0	0	0	2	1
C4	100	146	0	0	3	19	15	0	0	1	0	0.7	0	0	2	5	5
C5	191	256	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C6	49	56	0	0	1	16	30	0	0	1	0	2	0	0	0	0	0
C7	66	92	0	0	1	6	8	0	0	0	0	0	0	0	0	3	3
C8	232	313	0	0	0	8	3	0	0	0	12	4	0	0	0	10	3
C9	89	118	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C10	115	154	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.7
C11	111	149	0	0	1	10	7	13	12	0	0	0	0	0	0	6	4
C12	80	102	0	0	0	3	3	0	0	0	0	0	0	0	0	2	2
C13	104	79	0	0	2	23	32	0	0	2	9	14	0	0	1	23	30
C14	86	117	0	0	2	10	10	0	0	2	3	4	0	0	0	0	0
C15	232	315	0	0	0	9	3	0	0	2	3	2	0	0	1	5	1.9
Total	**1893**	**2465**	**0**	**0**	**14**	**147**	**7**	**13**	**0.7**	**12**	**35**	**2**	**0**	**0**	**6**	**70**	**3**

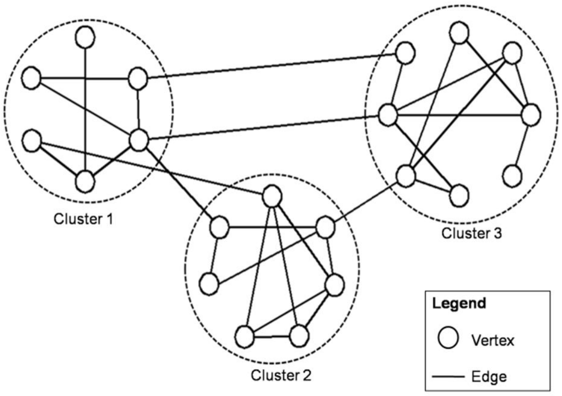

Figure 1. An example of modularity-based clustering (Reproduced from Fortunato (2010)).

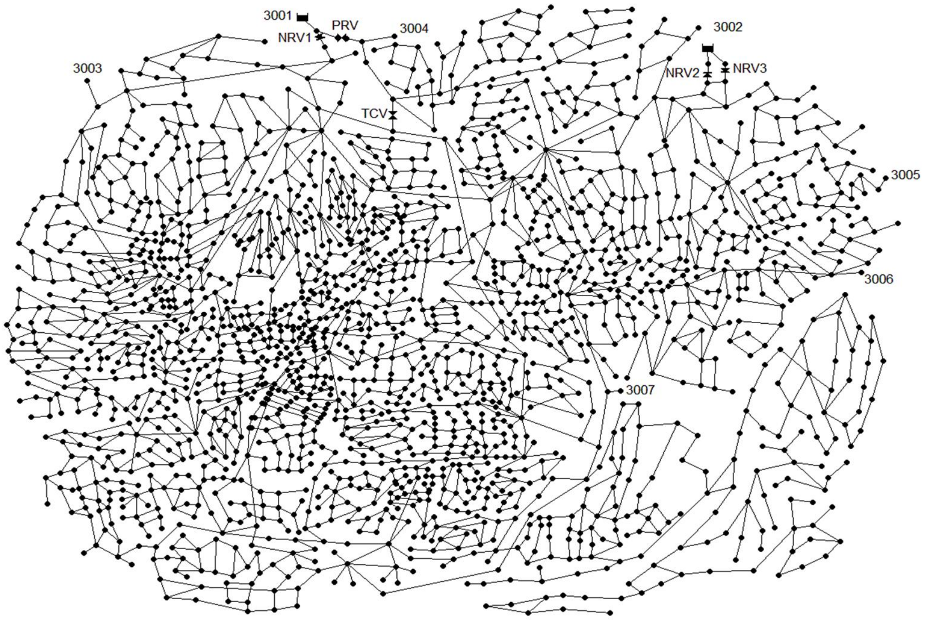

Figure 2. EXNET Water Distribution System

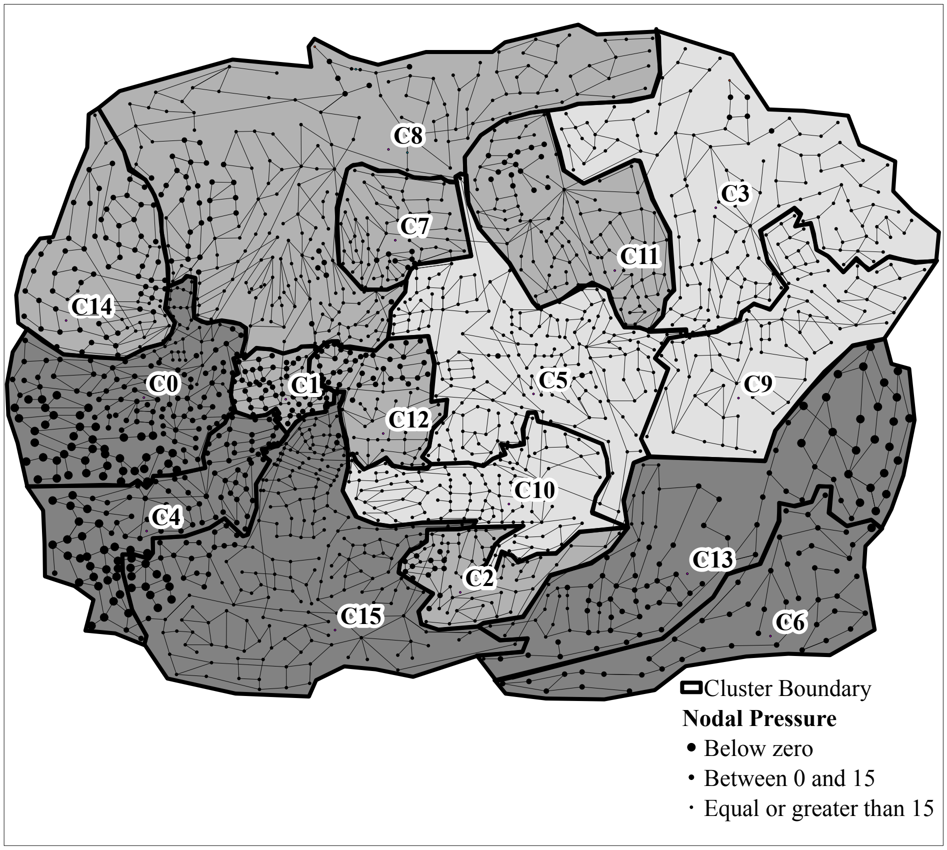

Figure 3. EXNET clusters and their pressure deficiencies

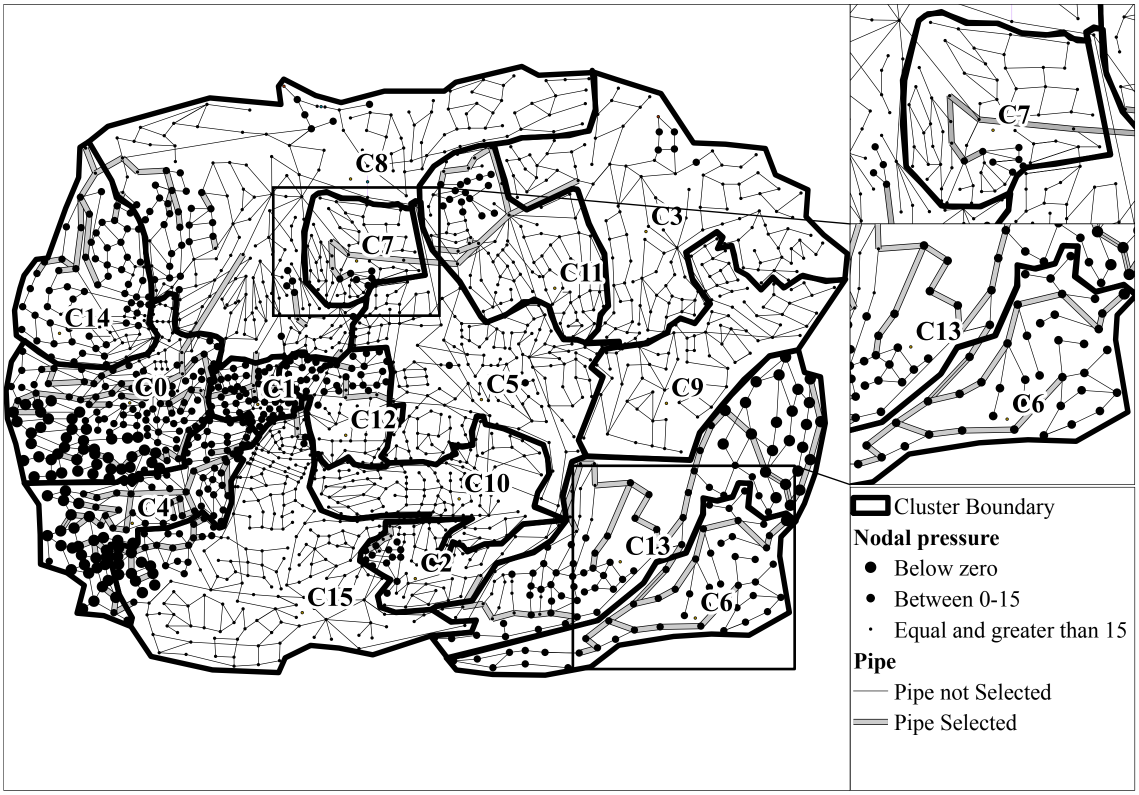

Figure 4. Pipes considered for rehabilitation in strategy 1

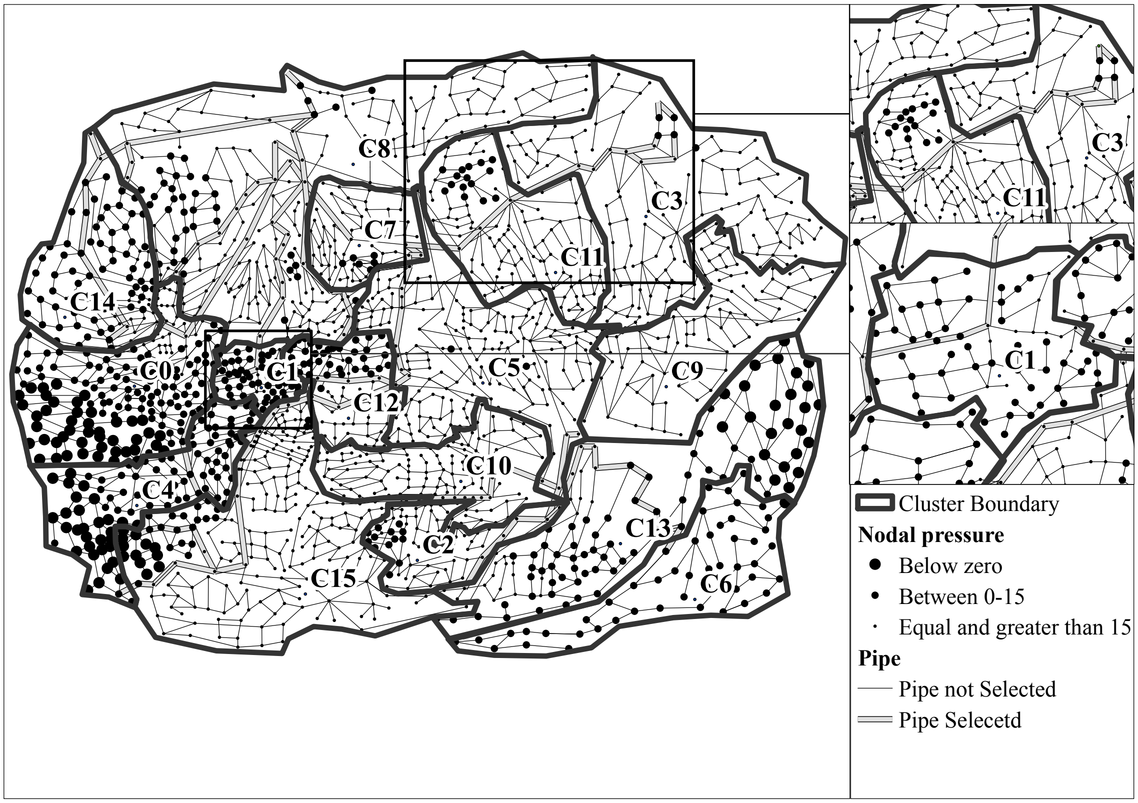

Figure 5. Pipes considered for rehabilitation in strategy 2

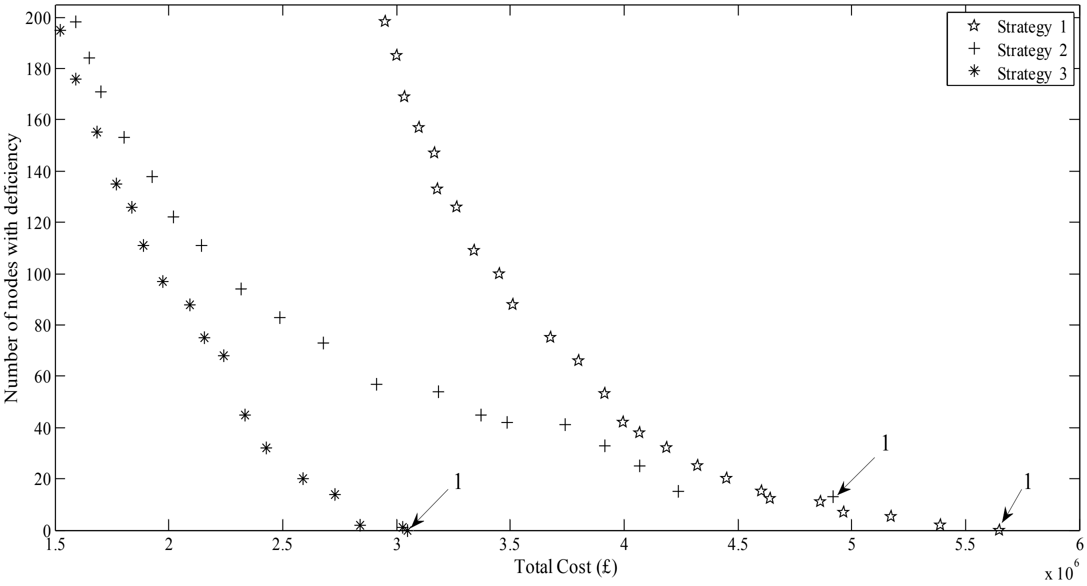

Figure 6. Pareto-fronts of strategies 1, 2 and 3

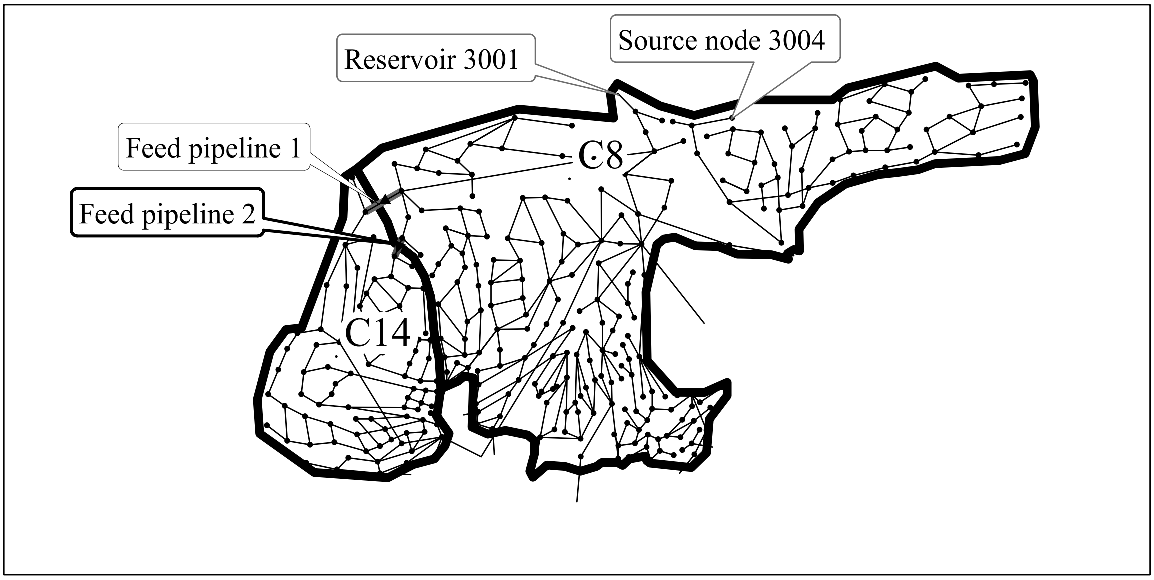

Figure 7. Duplicated feed pipelines between clusters 8 and 14 (Solution 1, Strategy 1)

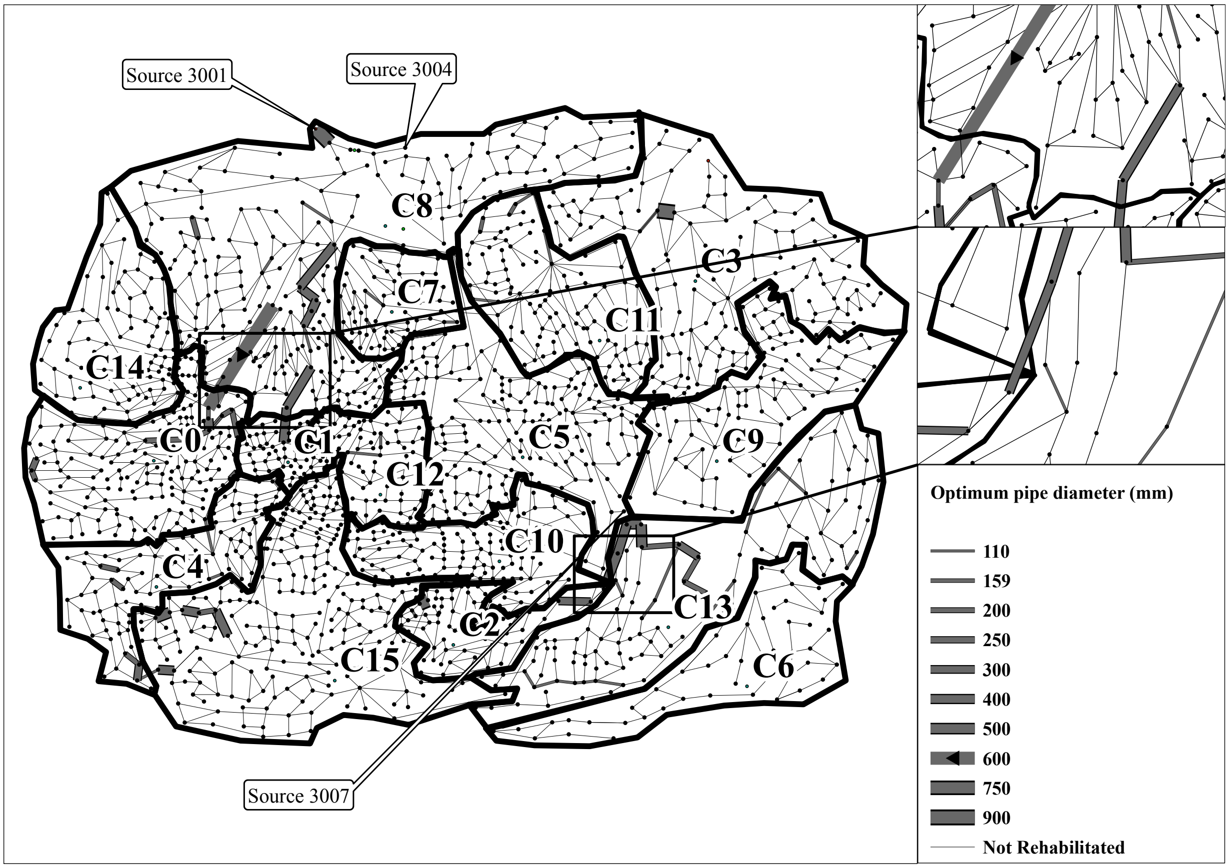

Figure 8. Layout of solution 1 of strategy 3

Figure 9. Pipes selected for rehabilitation based on engineering judgement

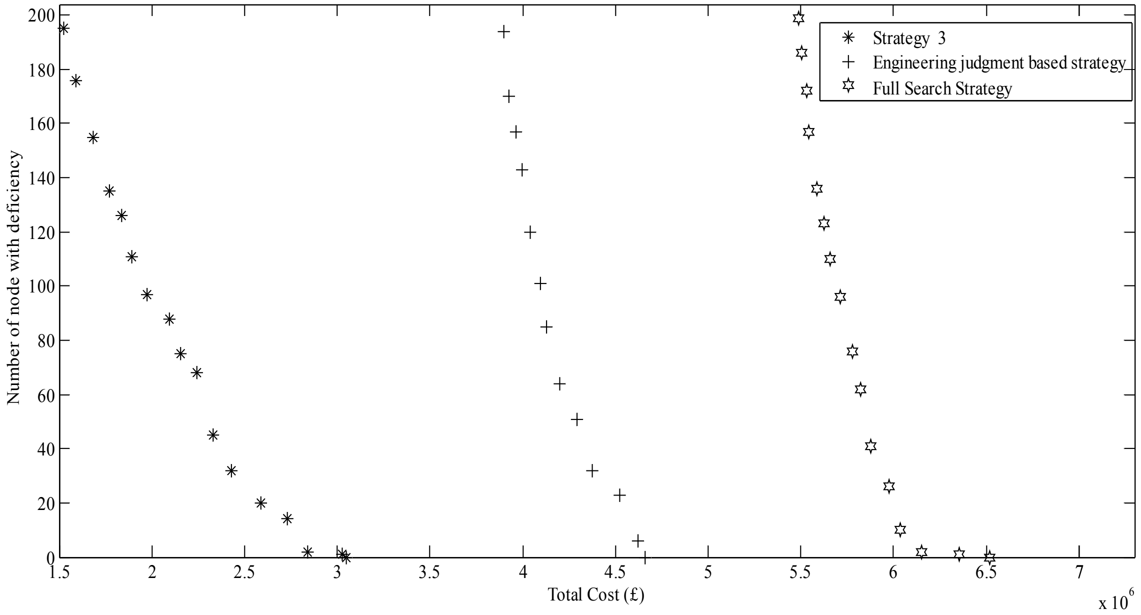

Figure 10. Pareto-fronts of strategy 3, engineering judgement based strategy and full search strategy

Fig. 1. An example of modularity-based clustering (Reproduced from Fortunato (2010)).

Fig. 2. EXNET Water Distribution System

Fig. 3. EXNET clusters and their pressure deficiencies

Fig. 4. Pipes considered for rehabilitation in strategy 1

Fig. 5. Pipes considered for rehabilitation in strategy 2

Fig. 6. Pareto-fronts of strategies 1, 2 and 3

Fig. 7. Duplicated feed pipelines between clusters 8 and 14 (Solution 1, Strategy 1)

Fig. 8. Layout of solution 1 of strategy 3

Fig. 9. Pipes selected for rehabilitation based on engineering judgement

Fig. 10. Pareto-fronts of strategy 3, engineering judgement based strategy and full search strategy