Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 70 (2014) 228 – 237

Procedia Engineering

www.elsevier.com/locate/procedia

12th International Conference on Computing and Control for the Water Industry, CCWI2013

A graph based analysis of leak localization in urban water networks

A. Candelieri[ab] , D. Conti[bc], F. Archetti[ab*]

[a]*Department of Computer Science, Systems and Communication, University of Milano Bicocca, viale Sarca 336, Milan, 20126, Italy*
[b]*Consorzio Milano Ricerche, via Roberto Cozzi 53, Milan, 20126, Italy*
[c]*Department of Operations Research, University of the Andes, Nucleo La Hechicera, Merida, 5101, Venezuela*

Abstract

A graph based analysis is proposed to improve leakage management in water distribution networks. Starting from the model of the network, leakage scenarios, created through hydraulic simulation (EPANET), are considered as nodes in a graph, whose edges are weighted by the similarity between each pair of nodes (scenarios), in terms of pressure and flow variation due to the leak. The graph is then analyzed in the eigenspace of its Normalized Laplacian matrix and specifically into the eigensubspace spanned by the most relevant eigenvectors, allowing Spectral Clustering, which is more effective than traditional techniques but with much higher computational requirements, to be applied also to large scale problems. The results obtained in the eigenspace are eventually mapped back into the physical space where the capability of leakage localization may be further improved through the fusion with leak severity estimation.

Keywords: Water Distribution Networks; Leakage Management; Leakage Localization; Spectral Clustering; Big Data

1. Introduction

Urban water distribution networks suffer, mainly due to the age of their pipeline infrastructure, frequent leaks and failures leading to service disruptions, large amounts of non revenue water, higher energy and rehabilitation costs (Puust, 2010). A more smart management of urban water distribution networks (WDN) is therefore needed to achieve higher levels of efficiency. The International Water Association (IWA) performance indicators (Alegre et

* Corresponding author.
E-mail address: archetti@milanoricerche.it

doi:10.1016/j.proeng.2014.02.026

al., 2006) detail the relevance to improve the leakage management process, generally defined on three different steps: assessment, detection and physical localization (Preis et al., 2010).

Analytical leakage localization tools have been brought to the forefront leading to the proposal of several approaches, one of which posits that leaks can be detected correlating changes in flow and pressure within the real water distribution network to the output of a simulation model whose parameters are then related to both location and severity of the leak. Another relevant research filed is more focused on the application of machine learning strategies to detect leaks and bursts by analyzing the data collected by real-time sensors without using any simulation (Romano et al., 2011).

The approach proposed in this paper belongs to the class of strategies based on the combination of hydraulic simulation of leakages and machine learning. In particular, Artificial Neural Network based approaches have been proposed by Caputo and Pelagagge (2003) and, more recently, by Sivapragasam et al (2007). Their systems use pressure and flow to infer the leak location and severity, through an ANN trained on a dataset generated either by a mathematical model of the network or the hydraulic simulation software EPANET, provided by the Environmental Protection Agency (http://www.epa.gov/nrmrl/wswrd/dw/epanet.html).

Another recent combination between EPANET-based leakage simulation and machine learning has been proposed by Mashford et al (2012), using Support Vector Machines. In this case, the SVM model has been trained on a dataset of leaks simulated on the junctions of the WDN (while most approaches simulate leaks on pipes): the trained SVM classifier is able to infer the leaky junction(s) according to the pressure and flow values. The approach has been tested on a network in the south east Melbourne providing satisfactory results.

Again in combination with the hydraulic simulation software EPANET, other supervised machine learning approaches have been recently proposed, such as Genetic Programming (Lijuan et al., 2012), Bayesian approaches (Poulakis et al., 2003; Xia et al., 2006) and Hidden Markov Mode-based agents (Nasir et al., 2012).

Clustering techniques, mostly applied to group the junctions of the physical network in order to identify suitable sectorization related to District Metering Areas (DMAs) or Pressure Management Zones (PMZs), have been recently proposed also for leakage localization (Xia and Guo-Jin, 2010), Candelieri and Messina (2012) and Candelieri et al. (2013).

This paper investigates the benefits provided by a new clustering methods based on eigenvalues analysis compared to other classical partitioning strategies (*K*-means, *K*-Medoids, etc.). The output of the EPANET simulation for different leak location and severity is a vector of pressure and flow variations with respect to a faultless network, which will be later call “leakage scenario”. These scenarios are the nodes of a graph whose edges are weighted by a similarity measure between each pair of nodes.

The rest of the paper is organized as follows: section 2 describes the overall approach and the different transformations from the physical space, associated to the water distribution network, to the eigenvectors space, associated to the scenarios similarity graph and the Spectral Clustering procedure; in section 3 the Spectral Clustering is detailed; section 4 provides some computational issues for efficient Spectral Clustering in Big Data settings; section 5 reports some experimental results. A discussion about the perspective of the approach is given in section 6.

2. Definition of the overall approach

In the following Fig. 1 the overall workflow of the graph-based analysis of leaks localization is depicted; the relevant difference between traditional and spectral clustering approaches is also highlighted.

More in detail, the first step performs several simulation runs, through EPANET, by placing, in turn, a leak on a pipe and varying its severity in a given range. At the end of each leakage simulation, the EPANET software outputs pressure and flow value at each junction and pipe, respectively. Only the values in correspondence of the position of monitoring devices in the real network are taken into account.

The pressure and flow variations due to each simulated leak are computed with respect to the correspondent values obtained by simulating the faultless network: thus, each simulated leak is stored in a dataset and represented by the pressure and flow variations (features) together with the information related to the affected pipe and the damage severity: each row of the dataset is named “leakage scenario”.

The dataset generation above described allows the transformation from the Physical Space to the Feature Space. A detailed description how the leakage scenarios may be obtained is contained in Candelieri and Messina (2012).

The further step, proposed in this paper, is to build a scenarios (similarity) graph whose edges are weighted by the similarity between each pair of nodes (scenarios), that is similarity in terms of flow and pressure variations induced by two different leaks (different in terms of affected pipe or damage severity).

This step allows us to move from the Feature Space to the Scenarios Network Space and to deal with the problem of grouping similar leakage scenarios as a graph clustering task (Schaffer, 2007).

In general, the aim of graph clustering is to group the nodes of a graph into clusters in order to maximize the sum of the weights on the edges within each cluster (intra-cluster similarity) while minimizing the sum of the weights on the edges connecting nodes in different clusters (inter-cluster similarity).

Spectral Clustering is an effective graph clustering procedure; more details on the algorithm are provided in the following section 3. With respect to the Fig. 1 is important to anticipate that two different Spectral Clustering schemes may be adopted: the recursive bi-partitioning, that initially divides the scenarios graph into two sub-graphs and then is recursively applied on each sub-graph until the desired number of groups (sub-graphs) have been achieved, and the one based on the application of the *K*-means in the space identified by the most relevant eigenvectors of the Normalized Laplacian of the scenarios graph's affinity matrix. Both the schemes are described in section 3.

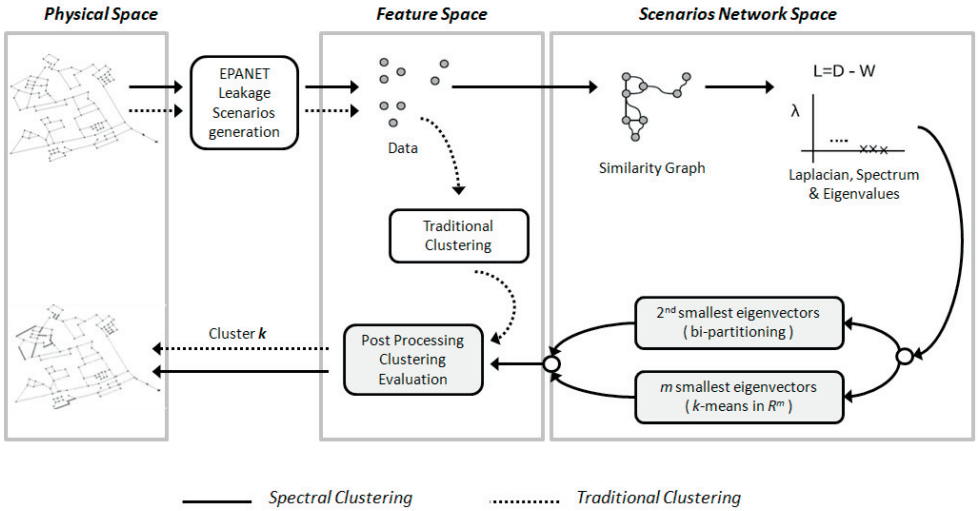

Fig. 1.Overall approach proposed: Spectral versus traditional clustering of leakage scenarios.

As already mentioned, the final goal is the same both for traditional and spectral clustering, that is partitioning leakage scenarios into subsets so that leakage scenarios in a cluster would be more similar than outside the cluster. However, spectral clustering works by taking into account the graph-based structure of the relations (edges) among scenarios (nodes). More in detail, the similarity between two scenarios has been computed as the correlation of the flow and pressure variations, ignoring the features related to pipe and severity.

Although several measures have been proposed for evaluating the internal fitness and intra-similarity of clustering procedures, the evaluation of the leak localization capability has particular features making it necessary the definition of a specific measure, namely "Localization Index". The Localization Index for each cluster (LI_k) is computed as the number of distinct pipes of the scenarios in that cluster with respect to the overall number of pipes in the WDN:

$$LI_k = \frac{|pipes| - |pipes_k|}{|pipes| - 1} \quad (1)$$

where | *pipes* | is the overall number of pipes of the WDN and | $pipes_k$ | is the number of simulated leaky pipes of the scenarios into cluster *k*.

The maximum value of LI_k is $LI_k = 1$ that is obtained when the cluster *k* contains scenarios all related to leaks simulated only on one pipe (i.e., | $pipes_k$ | = 1). On the other hand, the minimum value of LI_k is $LI_k = 0$ that is obtained if the cluster *k* contains scenarios referred to all the pipes of the WDN (i.e., | $pipes_k$ | = | *pipes* |).

The overall localization index of any clustering procedure (*LI*) is computed as the average of LI_k , with *k* varying from 1 to the overall number of clusters.

The effectiveness of different clustering approaches can be ranked by the proposed localization index on the interval [0, 1]. Furthermore, being normalized with respect to the number of pipes of the WDN it can be also adopted to compare results obtained on different WDNs.

It is also relevant to note that the last post-processing step implements the transformation from Scenarios Network Space back to the Feature Space and, finally, back to the Physical Space: when a specific cluster *k* is selected, all the pipes on which a leak has been simulated can be retrieved.

When a possible leak is detected (e.g., with traditional methods, such as Minimum Night Flow analysis, as reported in Liemberg and Farley (2004), Behzedian et al. (2009) and Izquierdo et al. (2011)), the actual pressure and flow values at the monitoring points are compared with those obtained through simulation of the faultless network, to compute the pressure and flow variation due to the possible leak. The vector of pressure and flow variations is then compared, according to the similarity measure (correlation, in this study) with the clusters' centroids in order to identify the most similar one and, consequently, the simulated leaky pipes related to the scenarios of that cluster. These pipes are the ones most probably affected by the leak, according to the pressure and flow variation found.

3. Spectral Clustering

Spectral Clustering (Luxburg, 2007 and Jaakkola, 2006) has recently emerged as an effective graph clustering algorithm. It can be implemented through standard linear algebra but its computational complexity $O(n^3)$ can prevent its application on large dataset.

Although it has been proposed in order to solve graph clustering problems, specifically in the network analysis domain, it very often outperforms traditional clustering algorithms, such as the *K*-means algorithm or other partitioning algorithms, when applied on not relational data points datasets.

In particular, given a set of data points $x_1, ..., x_n$ and some similarity measure $s_{ij} \geq 0$ between each x_i and x_j , traditional clustering approaches identify a partition of the data points into several groups in order to maximize intra cluster similarity and minimize inter clusters similarity.

A possible way of representing the data points consists of building a similarity graph $G=(V,E)$, where vertices v_i are the original data points x_i and edges e_{ij} are weighted by the corresponding s_{ij} of the Affinity matrix (v_i and v_j are not connected by any edge if $s_{ij} = 0$). At this point, the problem can be reformulated as a graph clustering task with the aim to identify a partition of the undirected similarity graph such that the sum of the weights on the edges between different groups is minimal while the sum of the weights on the edges within a group is maximal (i.e., points in different clusters are dissimilar from each other and points within the same cluster are similar to each other).

The solution of this problem can be easily described in the case of bi-partitioning. Given two sets of nodes (clusters), C_1 and C_2, the objective is to minimize:

$$cut(C_1, C_2) = \sum_{x_i \in C_1, x_j \in C_2} s_{ij} \quad (2)$$

A *n*-dimensional vector *p* (i.e., *n* is the number of nodes in the graph) is used to represent the association of each node to cluster C_1 or C_2 :

$$p_i = \begin{cases} +1 & if\ x_i \in C_1 \\ -1 & if\ x_i \in C_2 \end{cases} \tag{3}$$

The graph clustering problem can be formulated as minimization of the following function *f(p)*:

$$f(p) = \sum_{x_i, x_j \in V} L_{ij} \left(p_i - p_j\right)^2 = p^T L\, p \tag{4}$$

Where L_{ij} are the entries of the Laplacian matrix, the core of spectral clustering. Different alternative definitions have been proposed and studied through graph theory (Chung, 1997); the usually adopted definition is:

$$L = D - A \tag{5}$$

Where *A* is the affinity matrix of the undirected graph and *D* is the degree matrix, with each entry defined as:

$$d_{ij} = \sum_j a_{ij}\ , i = j \tag{6}$$

$$d_{ij} = 0, i \neq j \tag{7}$$

The most important properties of the *L* matrix are:

- it is symmetric and positive semi-definite (it has *n* non-negative, real-valued eigenvalues $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, irrespectively to their multiplicity);
- its smallest eigenvalue is 0 (where its multiplicity indicates the number of distinct connected components);

Many applications use Normalized Laplacian matrix instead of the basic one; the most common definition for the Normalized Laplacian matrix is the following:

$$L_{norm} = I - D^{-1/2} A D^{-1/2} \tag{8}$$

The combinatorial complexity of the minimizing (4) can be prohibitive for real world networks. However, a simple algebraic solution to the problem was proposed in (Fiedler, 1973): in particular, he used the result of the Rayleigh theorem and identified the 2nd smallest eigenvector of the Laplacian matrix (usually known as Fiedler vector) as the vector *p* which provides the optimal bi-partitioning of the graph.

This result has permitted to implement recursive bi-partitioning spectral clustering approaches (Hagen and Kahng, 1992) in order to perform partitioning in $K > 2$ groups. However this approach requires the computation of matrices and eigenvalues, as well as the use of the Fielder vector, for each sub-graph until the desired number of clusters is reached.

Another possible schema to solve the *K*-partitioning uses a data representation in the – usually low-dimensional – space of relevant eigenvectors (Luxburg, 2007; Ng et al., 2001). The relevant eigenvectors are the first *l* smallest:

the *l*-th eigenvalue is the one showing a sufficiently large variation in the *eigengap*, that is the difference between two successive eigenvalues in the list of eigenvalues sorted in ascending order.

For example, the *K*-partitioning approach proposed in (Shi and Malik, 2000), consists in selecting the *l* smallest non-zero eigenvalues and performing a traditional *k*-means clustering on the resulting dataset having *n* rows (nodes of the graph) and *l* columns (eigenvectors corresponding to the *l* smallest eigenvalues).

4. Spectral Clustering for "Big Data"

Recently some strategies for reducing memory and time requirements of the spectral clustering on large datasets have been proposed. Three different approximations have been in particular investigated: *(i)* reducing the computational cost of the eigen-decomposition, *(ii)* sparsifying the similarity matrix or *(iii)* performing a preliminary reduction of the original dataset. Moreover, from the technological point of view, some parallel computing solutions have been recently investigated in order to improve efficiency of spectral clustering without requiring any approximation (Chen et al., 2011).

The Nyström method is the widely adopted technique for approximating eigen-decomposition (Williams and Seeger, 2000; Fowlkes et al., 2004; Talwalkar et al., 2008). In particular, a subset of data is selected (randomly or through some "greedy" method); eigen-vectors are then computed on the correspondent sub-matrix and finally used to estimate an approximation of the eigen-vectors of the overall dataset.

Matrix sparsification techniques avoid to store the dense similarity matrix by considering only the more significant relationships between nodes. This can be performed by setting a threshold ε and removing all the edges with weights lower than ε (usually known as ε-neighborhood approach) or by considering only the *t* nearest neighbors of a node (usually known as *t*-nearest-neighbor approach). (Chen et al., 2004)

Analogously to Nyström method, methods performing a preliminary reduction of the data size are based on sampling but, contrary to Nyström, they apply spectral clustering on the selected data without estimating the eigen-vectors of the entire dataset. Fast approximate spectral clustering, proposed by Yan et al. (2009) is based on this idea and two different selection procedure are proposed: random or based on *k*-means. The *k*-means based version performs a preliminary *k*-means on the dataset (feature space) in order to identify a large number of k_0 centroids which spectral clustering is performed on.

As reported in Chen et al. (2004), from the clustering quality perspective, sparsification approaches provide slightly better results than the Nyström approximation: removing small similarity values does not lose much information with respect to the sampling performed by Nyström.

Quality of clustering strictly depends on available dataset: the larger the number of different simulated leaks the higher the probability to correctly localize the leaky pipe. Different scenarios correspond to different pipes and severity without any significant random component, it comes to no surprise that random sampling leads to information losses. Results obtained by using spectral clustering and *k*-means fast approximate spectral clustering are reported in the following section 5.

Efficiency is a really critical issue, not only with respect to spectral clustering, for the proposed leakage localization approach. Time and memory are required in several steps due to the Big Data nature of the problem:

- Building the leakage scenarios dataset requires *n* EPANET runs, with *n* = number of pipes x number of severity values;
- Creating the similarity matrix requires the computation of the similarity between each pair of nodes and also depends on the number of pressure and flow monitoring sensors (features): $O(m*n^2)$, with *m* the number of features;
- Eigen-decomposition of the Laplacian matrix has computational complexity $O(n^3)$.

For the above reasons there has been in the last few years a growing attention on parallel/distributed frameworks based in particular on the Map-Reduce scheme, for the analytical leakage localization.

Big Data issues become more critical with the adoption of smart metering solutions: when Automatic Metering Readers (AMRs) for near-real-time consumption metering are installed, the time series data related to water consumption are the input of the EPANET-based leakage simulation process and variations in flow and pressure are therefore provided as time series, increasing the number of features representing the leakage scenarios in the Feature Space according to time window considered. Currently, most of the WDNs do not use AMRs and consumption data are aggregated values, according to the accounting and billing process. Features representing leakage scenarios are therefore aggregated values, showing a lower level of complexity with respect to the adoption of AMRs.

5. Experimental Results

This section summarizes and compares the results obtained through Spectral Clustering and Fast Approximate Spectral Clustering (*K*-means based implementation) performed on a real WDN, in a little town in the North of Italy (approximately 13 km2), with an elevation ranging from 107 to 118.9 meters. This WDN guarantees the service to about 6300 citizens. As in most of the towns in Italy, water consumption is usually accounted for building and not for single user, therefore the number of consumption points is about 2600, lower than the number of citizens. The number of pipelines in the network model is 931: the simulation of leaks with severity varying on a set of 30 values has generated 931 x 30 = 27930 leakage scenarios. The monitoring sensors deployed into the WDN are 7: 6 acquiring pressure values and 1 acquiring flow values (features).

Several criteria have been proposed in order to evaluate and compare the fitness of different clustering schemes; most of them are based on accuracy, as reported in several studies using benchmark datasets where any example is already associated to a specific group (class). According to aim of leakage localization, clustering fitness has to be evaluated according to a specific index related to the capability to generate clusters of scenarios related to a limited set of pipes. In the following Table 1, the Localization Index (*LI*), as defined in the previous section 2, is reported for:

- a “pure” spectral clustering procedure
- a fast approximate spectral clustering procedure with a 50% reduction of the original dataset size
- a fast approximate spectral clustering procedure with a 75% reduction of the original dataset size

Table 1. Localization Index: comparison between spectral clustering and fast approximate spectral clustering. Selection a and selection b of the Fast Approximate Spectral Clustering are related to a reduction of 50% and 75% of the original dataset size, respectively

	Spectral Clustering	Fast Approximate Spectral Clustering (selection *a*)	Fast Approximate Spectral Clustering (selection *b*)
Mean	0.83	0.48	0.33
Standard Deviation	0.15	0.34	0.24
Min	0.66	0.03	0.00
Max	0.97	0.93	0.67

These results show that the trade-off between complexity and localization capability is rather unfavorable for the specific localization index. As reported in the literature, the full spectral clustering is also, as expected, superior also in terms of accuracy, but the trade-off, in this case, is much better.

The leakage scenarios dataset has been also used to train a regression model able to estimate the leakage severity according to pressure and flow variations at the monitoring points. A simple Least Median Squared Linear

Regression (Rousseeuw and Leroy, 2005) proved to be sufficiently reliable (Relative Mean Absolute Error = 0.8764% and Root Relative Mean Squared Errors = 2.5368%, on 10-fold cross validation). The combination of the leakage localization, based on spectral clustering, and the severity regression model permits to improve the localization capability, supporting the definition of a suitable investigation plan; the overall workflow is depicted in Fig. 2:

- when a possible leakage is assessed, the actual pressure and flow variations at the monitoring points are compared to the centroids of the clusters identified through spectral clustering;
- in parallel, the same pressure and flow variations are used by the regression model to estimate the severity, namely *discharge coefficient*;
- the output of the two previous steps are combined. Fig. 2 depicts an example of the combination process: the pipe most probably affected by a leak is "pipe=14", because it appears in three different scenarios of the identified cluster and also with discharge coefficient equal to the predicted one (C=0.01). Then, "pipe=20" appears twice but with discharge coefficients different from the predicted one, while the "pipe=333" appears only once but with a discharge coefficient equal to the predicted one. Respect to this, a higher priority is given to "pipe=333" than "pipe=20" in the definition of the investigation plan (ranked list of probably leaky pipes depicted in the same figure).

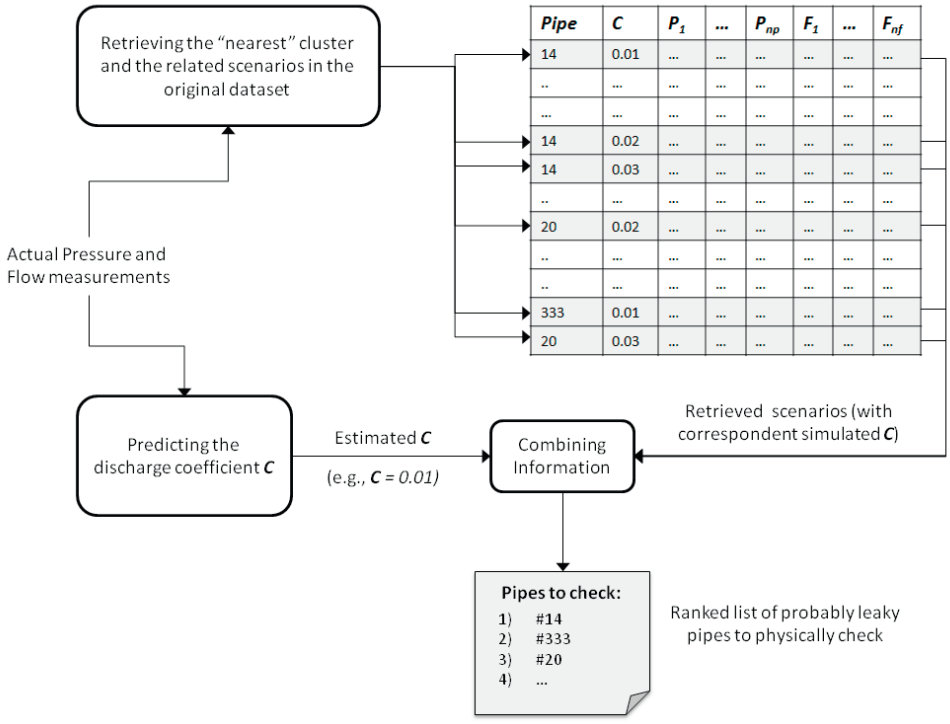

Fig. 2 Fusion of the clustering results and severity estimation.

6. Discussions

This study dealt with the application of graph-based analysis to develop an effective computational leakage localization approach aimed at improving the leakage management process in urban WDN. This approach is based

on a combination of simulation of different leaks, in terms of location and severity, and the graph-based clustering analysis of the pressure and flow variations resulting from each simulation run (leakage scenario).

In particular, Spectral Clustering, in its pure and approximated version, has been investigated according to its widely proved quality.

One important result is the inapplicability of the traditional measures of clustering fitness and the need of the definition of a specific index, namely Localization Index (*LI*).

Another important result stem from the comparison between "pure" and approximated spectral clustering proving that the latter does not provide significant benefits: while the reduction of computational costs affects only slightly general accuracy measures, it affects strongly the water specific measure Localization Index.

Spectral clustering approximations have been considered to tackle the critical issue of Big Data associated to the analytical task. The wider is the set of leakage scenarios generated the higher is the localization capability, however, due to its high computational complexity ($O(n^3)$), spectral clustering is not efficient for large scale problems.

The results obtained comparing "pure" and approximate spectral clustering, together with the opportunity to parallelize the leakage simulation runs, suggest to address future research activities to the design and development of a parallel/distributed framework for the specific problem of computational leakage localization.

Finally, the possibility to combine the graph-based analysis for leakage localization with more traditional machine learning techniques (i.e., regression) for the estimation of leak severity allows us to implement a workflow able to further improve localization and support WDN managers in defining a suitable investigation and intervention plan.

Acknowledgements

This work has been partially supported by the European Union ICeWater project – ICT 317624 (www.icewater-project.eu).

References

Alegre, H., Baptista, J.M., Cabrera, E., Cubillo, F., Duarte, P., Hirner, W., Merkel, W., Parena, R., 2006. Performance Indicators for Water Supply Services, Second Edition, IWA Publishing.

Behzadian, K., Kapelan, Z., Savic, D. A., Ardeshir, A., 2009. Stochastic sampling design using multi objective genetic algorithm and adaptive neural networks. Environmental Modeling and Software 24, 530–541.

Candelieri, A., Messina, E., 2012. Sectorization and analytical leaks localizations in the H2OLeak project: Clustering-based services for supporting water distribution networks management. Environmental Engineering and Management Journal 11(5), 953-962.

Candelieri, A., Archetti, F., Messina, E., 2013. Improving leakage management in urban water distribution networks through data analytics and hydraulic simulation. WIT Transactions on Ecology and the Environment 171, 107-117.

Caputo, A. C., and Pelagagge, P. M., 2003. Using Neural Networks to monitoring piping systems. Process Safety Progress 22(2), 119-127.

Chen, WY, Song, Y., Bai, H., Lin CJ, Chang, E.Y., 2011. Parallel spectral clustering in distributed systems. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(3), 568-586.

Chung, F., 1997. Spectral graph theory. Washington: Conference Board of the Mathematical Sciences.

Fiedler, M., 1973. Algebraic connectivity of graphs. Czechoslovak Mathematical Journal 23, 298–305.

Fowlkes, C., Belongie, S., Chung, F., Malik, J., 2004. Speactral grouping using the Nyström method. IEEE transactions on Pattern Analysis and Machine Intelligence 26(2), 214-225.

Hagen, L. and Kahng, A., 1992. New spectral methods for ratio cut partitioning and clustering. IEEE Transacations on Computer-Aided Design 11(9), 1074-1085.

Izquierdo, J., Herrera, M., Montalvo, I., Pérez-García, R, 2011. Division of Water Supply Systems into District Metered Areas Using a Multi-agent Based Approach, In: Software and Data Technologies, Series Communications in Computer and Information Science, Cordeiro J., Ranchordas A., Shishkov B. (Eds.), Springer Berlin Heidelberg 50, 167-180.

Jaakkola, T., 2006. Course materials for 6.867 Machine Learning, Fall 2006. MIT OpenCourseWare (http://ocw.mit.edu/), Massachusetts Institute of Technology.

Liemberger, R., Farley, M., 2004. Developing a nonrevenue water reduction strategy Part 1: Investigating and assessing water losses. In Proceeding of IWA WWC 2004 Conference, Marrakech, Morocco.

Lijuan, W., Hongwei, Z., and Hui, J., 2012. A Leak Detection Method Based on EPANET and Genetic Algorithm in Water Distribution Systems. Software Engineering and Knowledge Engineering: Theory and Practice – Advances in Intelligent and Soft Computing 14, 459-465.

Luxburg, U., 2007. A Tutorial on Spectral Clustering. Statistics and Computing 17(4), 1-32.

Mashford, J., De Silva, D., Burn, S., and Marney, D., 2012. Leak Detection in simulated water pipe networks using SVM. Applied Artificial Intelligence: An International Journal 26(5), 429-444.

Nasir, A., Soong, B. H., Ramachandran, S., 2010. Framework of WSN based human centric cyber physical in-pipe water monitoring system. 11th International Conference on Control, Automation, Robotics and Vision, 1257-1261.

Ng, A.Y., Jordan, M., Weiss, Y., 2001. On Spectral Clustering: Analysis and an algorithm. Advances in Neural Information Processing Systems 14, 849-856.

Poulakis, Z., Valougeorgis, D., and Papadimitriou, C., 2003. Leakage detection in water pipe networks using a Bayesian probabilistic framework. Probabilistic Engineering Mechanics 18, 315-327.

Preis, A., Allen, M., Whittle, A.J., 2010. On-line hydraulic modeling of a Water Distribution System in Singapore. Water Distribution System Modeling Issues, 1336-1348

Puust, R., Kapelan, Z., Savic, D. A., and Koppel, T., 2010. A review of methods for leakage management in pipe networks. Urban Water Journal 7(1), 25-45.

Romano, M., Kapelan, Z., and Savić, D., 2011. Real-Time Leak Detection in Water Distribution Systems. Water Distribution Systems Analysis, 1074-1082.

Rousseeuw, P.J., Leroy, A.M., 2005. Robust regression and outlier detection. Vol. 589. Wiley. com.

Schaeffer, S.E., 2007. Graph Clustering (survey). Computer Science Review, 27-64.

Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 888-905.

Sivapragasam, C., Maheswaran, R., and Venkatesh, V., 2007. ANN-based model for aiding leak detection in water distribution networks. Asian Journal of Water, Environment and Pollution 5(3), 111-114.

Talwalkar, A., Kumar, S., Rowley, H., 2008. Large-scale manifold learning. In Proceedings of CVPR.

Williams, C.K.I., Seeger, M., 2000. Using the Nyström method to speed up kernel machines. In Proocedings of NIPS, 682-688.

Xia, L., and Guo-jin, L., 2010. Leak detection of municipal water supply network based on the cluster-analysis and fuzzy pattern recognition. 2010 International Conference on E-Product E-Service and E-Entertainment (ICEEE) 1(5), 7-9.

Xia, L., Xiao-dong, W., Xin-hua, Z., Guo-jin, L., 2006. Bayesian theorem based on-line leakage detection and localization of municipal water supply network. Water and Wastewater Engineering 12.

Yan, D., Huang, L., Jordan, M.I., 2009. Fast approximate spectral clustering. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM.