International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

An approach based on ranking elements to form supply clusters in water supply networks as a support to vulnerability assessment

J. A. Gutiérrez-Pérez[1], M. Herrera[1], J. Izquierdo[1], R. Pérez-García[1]
[1]Fluing-IMM, Universitat Politècnica de València, Spain.
e-mail: {joagupre, mahefe, jizquier, rperez}@upv.es

Abstract: We propose an approach as a support to the vulnerability analysis of Water Supply Networks (WSNs). The method is based on graph measurements such as the relative importance (ranking) and the degree of the vertices of a graph. We calculate an index taking into account both measurements and, which represents the topological vulnerability of the network. The index gives information about the accessibility and exposure in case of external threats or hazards. In addition, we divided the network into clusters by the spectral clustering algorithm taking into account the vulnerability index. This division allows characterizing the water supply network into small sub-networks and simplifying the security analysis.

***Keywords**:* Water Supply Networks, clustering process, ranking, PageRank

1 INTRODUCTION

In recent years, there has been an increased awareness of the vulnerability of drinking water networks infrastructures, to accidents and deliberate acts. Due to several events that have demonstrated the fragility of water supply networks infrastructure, there is a growing interest in methods that help to screen and prioritize their components in order to facilitate the vulnerability assessment in Water Supply Networks (WSNs) management.
Ranking elements in a WSN is a recent proposal. Graph-based ranking algorithms are used for deciding the importance of an element in a graph by taking into consideration global information drawn from the graph structure. Addressing this issue, Herrera et al. [2011] and Gutiérrez-Pérez et al. [2011], demonstrated the possibilities through the adaptation of the PageRank algorithm to calculate the importance of the nodes. Izquierdo et al. [2008] focused on pipes and assessed their relative importance in the water distribution process. Moreover, Yazdani [2011] and Grubesic et al. [2008] demonstrated the scope of graph theory and complex networks principles in the analysis of WSN vulnerability and robustness. These works reviewed metrics of indexing nodes based on statics and spectral analysis. Michaud and Apostolakis [2006] evaluated the importance of the nodes through a methodology based on risk scenarios.
In the present work, we use the spectral-clustering methodology to obtain supply clusters, also known as District Meter Areas (DMAs), [Herrera, 2010]. The spectral clustering is a methodology in data analysis that improves the straightforward application of K-means, works well in non-convex spaces, and takes into account the underlying graph structure under study. Spectral-clustering uses information obtained from computing the eigenvalues and eigenvectors of the Laplacian matrices obtained from partitioning the graphs. The clusters are obtained according

to similar characteristic, or based on a specific measure. Our purpose is to introduce relative important measurements as a starting point to form the DMAs into a water network. For this purpose, we use the PageRank algorithm [Brin and Page, 1998] and the node degree. The PageRank is a very useful relative important measure that helps understand the connectivity structure and the performance of a network. Also, it enables to identify critical elements into the network.
This will help manage critical points of a Water Supply Network (WSN), and ascertain how they affect the nodes they are connected to in case of threat or hazard. In addition, the hydraulic analysis of a network results simpler by a suitable network division. A case study of a real WSN is presented to illustrate our methodology.

2 RELATIVE IMPORTANCE MEASURES OF NETWORK TOPOLOGY

2.1. Ranking algorithm: PageRank

The PageRank algorithm [Brin and Page, 1998] is the initial calculation method that Google founders use to classify web pages by their importance. This algorithm has been improved and adapted to several fields of study. The objective of this method is to obtain the pagerank vector that provides the relative importance of the web pages. The PageRank integrates the impact of incoming and outgoing links into the single model, and therefore it produces only one set of scores.
To formulate the concepts, the Web network is represented as a directed graph $G=(V,E)$ where V is the set of vertices (i.e. the set of all pages in the Web) and E is the set of directed edges between a pair of nodes (i.e. the hyperlinks). Let the total number of pages on the Web be $n(n=|V|)$. Therefore, the pagerank value of each vertex i, V_i, is defined as:

$$PR(V_i)=(1-d)+d\sum_{v_j\in In(v_i)}\frac{PR(V_j)}{|Out(V_j)|} \quad (1)$$

where for a given vertex V_i, $In(V_i)$ is the set of vertices that point to it (predecessors), and $Out(V_j)$ is the set of vertices that vertex V_i points (successors). And, where d (named damping factor) is a parameter that is set between 0 and 1. Usually, d takes the value of 0.85 [Brin and Page, 1998].
To adapt the algorithm to WSN, the physical structure of a WSN is considered as a mathematical graph $G=(V,E)$, where V is the set of all graph vertices which represents tanks, water sources, and nodes. The set of graph edges, E, represents pipes, valves, and pumps. By understanding a WSN as a graph of special characteristics, it is possible to abstract the concept of a web page and see it as a consumption node in a WSN. Links between pages are now understood as pipes connecting different nodes. Thus, it is possible to demonstrate the possibilities of the ranking algorithm through its adaptation to measure the importance of the nodes of a WSN.

2.2. Vertex degree distribution

Among the graph based measurements, the vertex degree is a good indicator of its topological importance (or connectivity). The degree k_i of a vertex i is the number of edges incident with the vertex and is defined in terms of de adjacency matrix A as:

$$k_i=\sum a_{ij}\,. \quad (2)$$

A directed graph has two components [Boccaletti et al., 2006]: the number of outgoing links $k_i^{out}=\sum a_{ij}$ (i.e., the out-degree of the vertex), and the number of

ingoing links $k_i^{\text{in}} = \sum a_{ij}$ (i. e., the in-degree of the vertex). Therefore, the total degree is defined as: $k_i = k_i^{out} + k_i^{in}$.

The basic topological characterization of a graph G , can be obtained in terms of the degree distribution. The vertex degree is characterized by a distribution function $P(k)$, which defines the probability that a randomly selected vertex has k edges. Equivalently, as the fraction of vertices in the graph having degree k . In the case of directed networks it is necessary to consider two distributions $P(k^{in})$ and $P(k^{out})$. Information on how the degree is distributed among the vertices of a undirected graph can be obtained either by a plot of $P(k)$, o by the calculation of the moments of the distribution. The n-moment of $P(k)$ is defined as [Boccaletti et al., 2006]:

$$\langle k^n \rangle = \sum k^n P(k), \tag{3}$$

where the first moment $\langle k \rangle$ is the mean degree of G , and the second moment, measures the fluctuations of the connectivity distribution.

2.3. Index based on Ranking and topological importance of the vertices

In graph based methods, there are several measurements proposed to characterize networks. However, we considered that both the accessibility and the physical exposure can be measured based on the relative importance (ranking) and topological importance of the vertices. In addition, the pagerank is a significant concept in current studies of the complex networks.
Therefore, based on the two measures discussed in sections 2.1 and 2.2 (the pagerank and the vertex degree distribution) the ranking-topological index can be introduced an expressed as follows:

$$RD_i = PR(v_i)P(k_i) \tag{4}$$

where RD_i represents the relative topological importance of the vertex i , based on its ranking and the connectivity importance. $PR(v_i)$ is the pagerank of the vertex i , and $P(k_i)$, is the probability that the vertex i has k edges connected.

3 CLUSTERING PROCESS

3.1 Spectral-clustering

The graph clustering process consists of grouping vertices of graph into clusters while taking into consideration the edge structure of the graph in such a way that there are many edges within each cluster and relatively few between clusters. There are several methods for finding a solution in graph clustering; among the most important is the spectral-clustering. This method is based on the eigenvalues and eigenvectors of a block-diagonal matrix that is associated with the graph [Ng et al, 2001].
During the application of the spectral-clustering process, the Kernel matrix of the normalised Laplacian is used. In order to make a complete analysis, it is possible to add this kernel matrix with others kernel matrices which are associated to different variables. However, the kernel matrices addition is not directly applied. First, their respective similarity matrices must be calculated, and then, these matrices must be transformed into symmetric and positively defined. Thus, each matrix can be understood as a kernel matrix.
The kernel matrices have two basic properties, an additive property (the sum of kernel matrices is another kernel matrix), and another property which affirms that a kernel matrix multiplied by any scalar higher than zero, it remains a kernel matrix.

Based on these considerations it is possible to analyze the problem by a jointly way. On the one hand, the information associated to the graph (through the normalised kernel Laplacian) and by the other side, other important inputs to construct the clusters (through the kernel matrices related to other data, such as the graph coordinates). Therefore, the final kernel matrix on which work is defined as:

$$K = \lambda_A K_A + \sum_{i \in I} \lambda_i K_i \quad (5)$$

where K is the kernel matrix for clustering process, K_A is the kernel matrix related to the affinity graph, and K_i, is the matrix associated to the inputs of our interest in the process of building the clusters into the network. Finally, λ_A and λ_i, $i \in I$, are the weights entering the linear combination. The parameters λ_A and λ_i, can be calibrated within the spectral-clustering process to improve the results, or propose them according to the importance of each input. Starting from the information arranged in a suitable kernel matrix it is possible to perform a detailed analysis as the spectral-clustering method.

4 EXPERIMENTAL STUDY

To demonstrate the discussed approach we considered a real case of WSN. The structure of the network is presented in Figure 1. The WSN is made of 107 consumption nodes, 134 pipes, and three tanks.

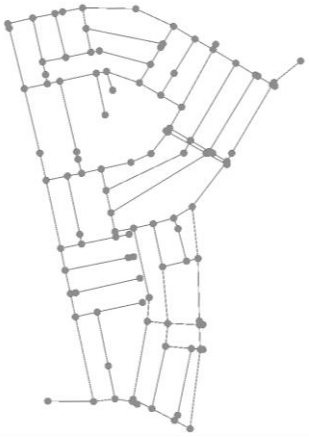


Figure 1. General scheme of the WSN

The aim of our proposal is to divide the WSN into DMAs taking into account all the available information of the network, and the RD_i index calculated by the pagerank and the node degree distribution. This information will be constructed and treated as input matrices. In the first instance, the calculation was made without the RD_i index. The affinity graph matrix of the WSN is transformed into a kernel matrix, carrying out the correspondence kernel abstraction of the essential characteristics of the WSN. Likewise, the dissimilarity matrix of the RD_i index is transformed into a kernel matrix. Following, spectral-clustering techniques are applied to these new matrices. A general description of the spectral-clustering process is summarized in Table 1. A detailed discussion can be consulted in Herrera [2011].

Table 1. The Spectral-clustering process

Clusters into WSN by spectral-clustering process
1. Abstraction of the WSN as a graph
2. Construction of the Laplacian graph
3. Transformation of the affinity matrix into kernel matrix
4. Calculating the spectrum of the matrix

5. k-means on the first c eigenvectors
6. re-allocation of the results to the original data

The application of the graph-spectral algorithm to our case study consisted in introducing a WSN as a special case of undirected and weighted graph. The available information of the network is introduced and conformed as the input matrix. Then the transformation and calculus of the kernel matrices were made with the R Language [R-Development-Core-Team, 2010] and the postprocessing with the NetLogo [Wilensky, 1999] plataform. A scheme of the overall process is shown in Figure 2.

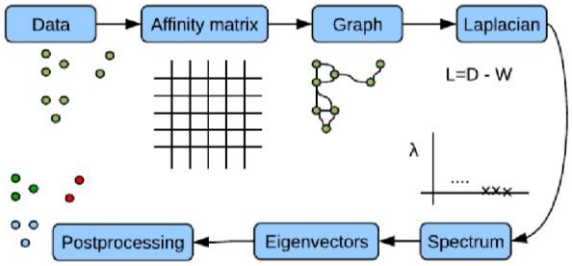


Figure 2. Scheme of the process of spectral-clustering [Vejmelka, 2009]

Regarding the results of PageRank algorithm, in Figure 4(a) we can observe the relative importance distribution of the nodes into WSN. In general, there is a middle pagerank level. The principal data of the ranking nodes are in Table 2. The node degree distribution $P(k)$ of the WSN, is displayed in Figure 3, and shows its topological characteristic. The degree of a node define the number of pipes (edges) the node is connected to, which has further insights on the connectivity properties of the network. For the WSN, most of the nodes have low degree, about 90% have degree less than or equal to 3. However there are some high-degree nodes.

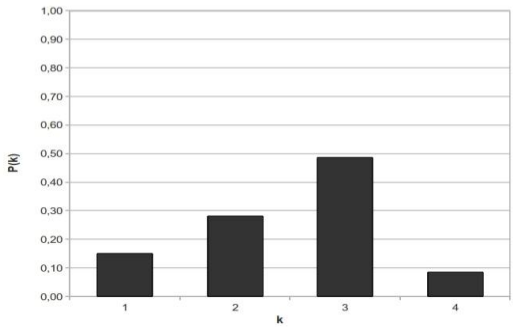
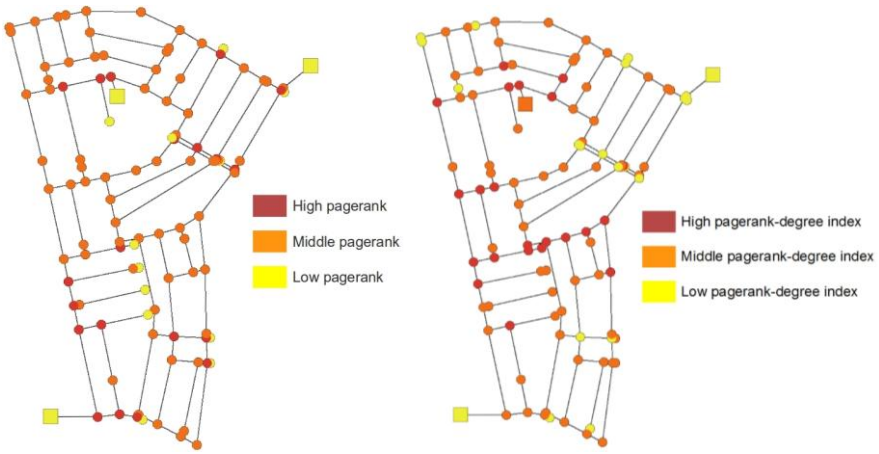


Figure 3. Degree distribution of the network

About Figure 4(b), it is represented the RD_i distribution which is calculated taking into account the pagerank, and $P(k_i)$. Table 3 summarize the principal data of the results obtained.

Table 2. Principal data of the ranking levels

pagerank level	Representative value	Average	Number of nodes
High	0,01640	0,0129	20
Middle	0,00935	0,0093	71
Low	0,00434	0,00483	16



(a) **(b)**

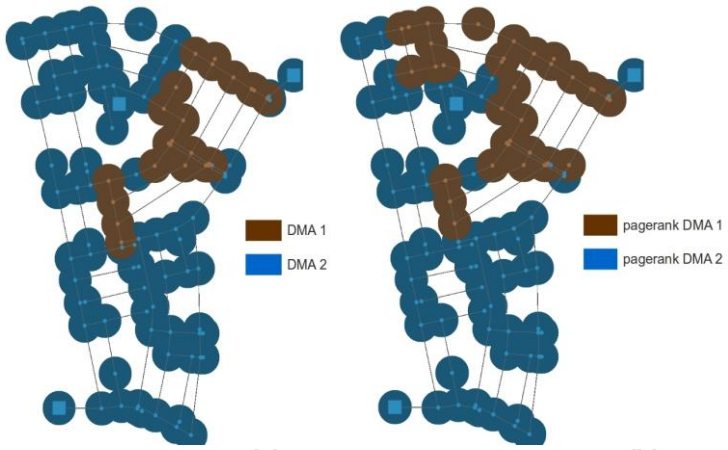
Figure 4. Scheme of ranking distribution of the nodes (a). Scheme of the RD_i index distribution of the nodes (b)

Table 3. Principal data of the RD_i index levels

RD_i	**Representative value**	**Average**	**Number of nodes**
High	0.02423	0.00931	22
Middle	0.00489	0.00217	64
Low	0.00094	0.00062	21

After the application of the graph-spectral clustering process, we obtained the division of the WSN presented in Figure 5(a). Then, we introduce the pagerank measurement and build the clusters again (Figure 5b). It is observed that some nodes are included into the DMA1. In Table 4 the most important data of these clusters are detailed.

As mentioned above, the main point of this work is to form clusters by taking into consideration all the information of the WSN and the RD_i index. To achieve this purpose, we use the dissimilarity matrix of the index as an input, and as it is discussed above we applied the spectral-clustering process. The results obtained are presented in Figure 6(a). In order to improve these results, we applied a multi-agent method which allows including the nodes that are out of the convenience cluster (Figure 6b).

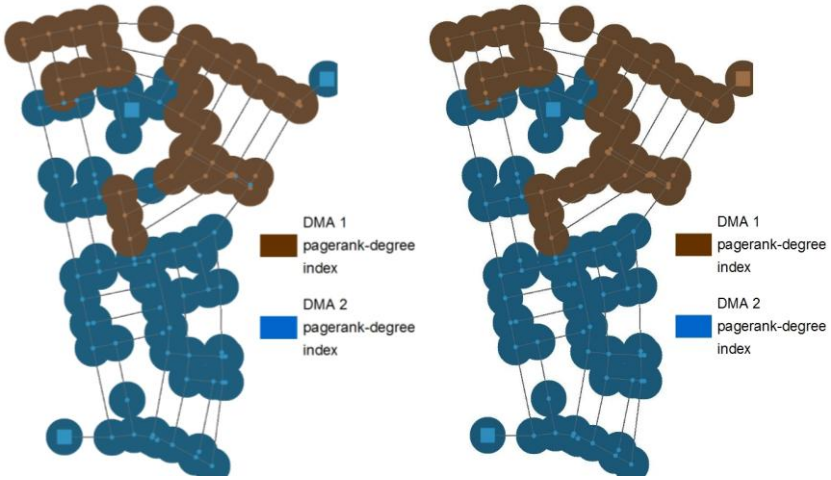


(a) **(b)**

Figure 5. Scheme of the division into clusters by the spectral-clustering method (a). Scheme of the division into clusters by the spectral-clustering method, taking into consideration the PageRank algorithm (b)

Table 4. Principal data of the pagerank clusters

Pagerank cluster	Representative value	Average	Number of nodes
DMA 1	0.01763	0.00576	35
DMA 2	0.05143	0.01109	72



(a) **(b)**

Figure 6. Scheme of the division into clusters by the spectral-clustering method taking into consideration the RD_i index (a) Scheme of the division into clusters by the spectral-clustering method, taking into consideration the RD_i index and modified by multi-agent method (b)

Table 5. Principal data of the RD_i index clusters

Cluster	Representative value	Average	Number of nodes
DMA 1	0.00602	0.00177	44
DMA 2	0.02423	0.00443	63

5 CONCLUSIONS AND RECOMMENDATIONS

It has been demonstrated that the spectral-clustering uses both graphical and vector information and is more efficient and robust than other methodologies. The flexibility of including various inputs in the study is another advantage of the methodology.

In this work, we propose an index by taking into consideration the pagerank score and the node degree distribution for generating a topological importance measurement to form clusters in a WSN. Also, these measurements help us to understand the structure of the network about their critical elements, specifically the nodes of the network. Moreover, hydraulic analyses of network results can be simplified with a suitable network division. Aspects of operational management, such as vulnerability analysis, can be complemented by using the index. In addition, we consider that the results obtained using these methodologies can be applied as another hydraulic criterion for further network division. Future works could be focused on developing these approaches. For example, the authors of this paper have recently proposed some modifications of the spectral-clustering algorithm, in order to apply it in other water supply management fields. They proposed the division of the WSN to locate sensors into the network to detect contaminations events. This type of researches helps in vulnerability assessment analyses.

ACKNOWLEDGMENTS

This work has been supported by the project IDAWAS, DPI2009-11591, of the Dirección General de Investigación of the Ministerio de Ciencia e Innovación of Spain and the complementary support ACOMP/2011/188, of the Consellería de Educació of the Generalitat Valenciana.

REFERENCES

Boccaletti, S., V. Latora, Y. Moreno, M. Chavez, and D-U. Hwang, Complex networks: structure and dynamics, *Physics Reports*, 424, 175-308, 2006.

Brin, S. and L. Page., The anatomy of a large-scale hypertextual web search engine, *Computer Networks and ISDN Systems,* 30, 1-7, 1998.

Grubesic, T.; T. Matisziw; A. Murray; D. Snediker, Compartive approaches for assessing network vulnerability, *International Regional Science Review*, 31(1), 88-112, 2008.

Gutiérrez-Pérez, J. A., M. Herrera, R. Pérez-García and E. Ramos-Martínez, Application of graph-spectral methods in the vulnerability assessment of water supply networks, *Mathematical and Computer Modelling*, doi:10.1016/j.mcm.2011.12.008, 2011.

Herrera, M., Improving water network management by efficient division into supply clusters. Ph.D. thesis, Universitat Politècnica de València, España, 2011.

Herrera, M., J. A. Gutiérrez-Pérez, J. Izquierdo and R. Pérez-García, Ajustes en el modelo PageRank de Google para el estudio de la importancia relativa de los nodos de la red de abastecimiento, paper presented at X Seminario Iberoamericano de planificación, proyecto y operación de sistemas de abastecimiento de agua (SEREA). Morelia, México, 2011.

Herrera, M., S. Canu, A. Karatzoglou, R. Pérez-García, and J. Izquierdo, An approach to water supply clusters by semi-supervised learning, paper presented at iEMSs-2010, International Congress on Environmental Modelling and Software, Ottawa, Canada, 2010.

Izquierdo, J.; I. Montalvo, R. Pérez-García and M. Herrera, Sensitivity analysis to assess the relative importance of pipes in water distribution networks. *Mathematical and computing Modeling*, 48, 268-278, 2008.

Michaud, D. and G. E. Apostolakis, Methodology for ranking the elements of water supply networks, *Journal of infrastructure Systems*, 12(4), 230-242, 2006.

Ng, A. Y., M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In Advances in Neural Information Processing Systems 14, pages 849–856, 2001.

R-Development-Core-Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 54, 88, 97, 101, 133, 144, 155, 161, 172, 173, 185, 193.

Vejmelka, M. (2009). Spectral graph clustering. In Seminar z Umele Inteligence, Prague. 44

Wilensky, U. (1999). Center for Connected Learning and Computer Based Modeling, Northwetern University, Evanston, IL. 54, 68, 69, 88, 97, 98, 101, 133, 144, 155, 156, 185, 193.

Yazdani, A. And P. Jeffrey, Complex network analysis of water distribution systems, Chaos, 21, 016111, doi:10.1063/1.3540339, 2011.