

La regularización del grafo de la red de abastecimiento de agua para la propuesta de su sectorización

Manuel Herrera; Joaquín Izquierdo; Rafael Pérez-García; David Ayala-Cabrera

FluIng-IMM, Universitat Politècnica de València (UPV)

E-mail: {mahefe; jizquier; rperez; daaycab}@upv.es

Resumen

La distribución racional de agua en una red de abastecimiento es un problema complejo, especialmente en sistemas de gran escala. La solución propuesta por la división de la red en sectores es una opción estratégica usada para controlar y operar las redes de abastecimiento buscando la mejora de su gestión al trabajar cada parte como un todo. En la mayor parte de las ocasiones, estas redes no son el producto de un solo diseño, sino la consecuencia de años de historia que han dado respuesta a demandas de agua continuamente crecientes con el tiempo. De esta manera, previo al análisis de las diferentes alternativas de sectorización de una red, cobra importancia llevar a cabo un estudio topológico de la misma. Este artículo propone realizarlo mediante el ensayo de diferentes metodologías de regularización de la matriz Laplaciana asociada a su grafo, tal que represente la estructura de la red de la mejor manera. Para ello será necesario redefinir funciones y operadores matriciales que construyan este nuevo marco de trabajo. El paso siguiente será la aplicación de un algoritmo de cluster sobre grafos que agrupe sus vértices tal que dentro de cada grupo haya un número de arcos suficiente, a la vez que minimice los que separen un grupo de otro.

Trabajar con esta regularización de la matriz de partida proporciona un entorno adecuado para el desarrollo y máximo aprovechamiento de las últimas metodologías de sectorización, basadas en el análisis cluster sobre el espectro del Laplaciano de la red. La aplicación de este procedimiento sobre el distrito hidrométrico de un caso real ofrecerá las directrices y recomendaciones básicas para la propuesta de su sectorización, con un coste mínimo en cuanto a las modificaciones del trazado original de la red.

1. Introducción

La distribución racional del agua en un sistema de abastecimiento constituye un problema complejo. Esta complejidad se acrecienta si dicho sistema es de gran tamaño y se tiene como objetivo un abastecimiento que ofrezca una satisfacción de la demanda con los niveles de presión, calidad y continuidad adecuados. La sectorización, entendida como la partición de la red en subsistemas con entradas y salidas controladas, es una opción estratégica que homogeneiza los elementos, las medidas y los parámetros de diseño de cada sub-red. De esta forma, se gana precisión y se evitan sesgos en los análisis que fundamentarán la toma de decisiones en la gestión del abastecimiento. Además, el conocimiento de los usuarios integrados en cada sector, junto con un adecuado tratamiento de la base de datos, hace fundamental dicha sectorización para la comprensión del comportamiento de la demanda de agua y la propuesta de un Sistema de Gestión de Abonados (SGA) eficiente. Desde el punto de vista que conlleva la división de la red en zonas, la sectorización reduce el área de inspección del sistema, facilitando así la detección, localización y control de anomalías en el suministro.

El concepto de la gestión de las redes de abastecimiento de agua en base a Zonas de Medición por Distritos (DMA), utilizando la división de dicha red en sectores hidráulicos, fue una novedad introducida para la detección y el control de fugas, a primeros de los años 80, por la industria del agua del Reino Unido (*Department of the Environment and the National Water Council, Report 26, 1980*). Desde entonces, su desarrollo se ha basado, principalmente, en su implantación práctica y se cuenta con escasas contribuciones científicas en este campo. Así, cabe destacar el manual que la *UK Water Research Ltd.* publicó en 1999. Más reciente (diciembre de 2007) es la guía práctica de la gestión de las DMA, de la *IWA Water Loss Task Force*. Por otro lado, dentro de un marco más conceptual y de mayor rigor científico, Walski (Walski et al., 2001) propuso el establecimiento de un sistema de submediciones, apoyando la implantación de la sectorización en la red de agua. MacDonald & Yates, 2005, hicieron un estudio aplicado del diseño de DMA en abastecimientos de agua de Norte América. En 2007, Hunaidi & Brothers colaboraron con un artículo donde se busca el tamaño óptimo

de las DMA basándose en diferentes criterios y teniendo en cuenta su coste económico. Tzatchkov et al., 2006, aplicaron teoría de grafos para dividir la red de abastecimiento en diferentes zonas hidráulicas. En 2008, Izquierdo et al., evaluaron la importancia relativa de las tuberías, a partir de la cual se puede establecer un criterio zonal de división de la red. Estos mismos autores propusieron (Izquierdo et al., 2009) una sectorización basada en la aplicación de Sistemas Multi-Agente.

La aplicación de la sectorización hidráulica más desarrollada es la de la localización activa de fugas. Destacan los trabajos de Covas y Ramos (Covas & Ramos, 1999), que desde finales de los 90 y hasta la actualidad, han dedicado a este campo de la aplicación de las DMA. En 2005, Misiunas presenta su tesis doctoral en la que discute diferentes aspectos de los fallos que pueden originarse en una red de abastecimiento de agua dividida en DMA. En este mismo año, Bougadis et al., 2005; plantean la diferencia de comportamiento del consumo de agua potable en diferentes zonas hidráulicas y utilizan una clasificación de éstas en el desarrollo de modelos de predicción de la demanda. Herrera et al. (2010) realizan un completo estudio comparativo de modelos predictivos de la demanda relativos a un DMA.

La división de la red de abastecimiento de agua en sectores hidráulicos no sólo ha tenido el desarrollo científico y técnico expuesto hasta ahora, sino que se ha ido llevando a su implantación práctica en diferentes países de Europa, Asia y América. En el caso concreto de España, se han sectorizado las redes de abastecimiento de ciudades como Barcelona (2004), Madrid (2005), Córdoba (2007), San Sebastián (2008), etc.

En este artículo se pretende hacer la sectorización en base a la información de la conectividad del grafo de la red de abastecimiento (Herrera et al., 2009) recogida por su matriz Laplaciana asociada. El paso siguiente será la aplicación de un algoritmo de cluster sobre grafos que agrupe sus vértices tal que dentro de cada grupo haya un número de arcos suficiente, a la vez que minimice los que separen un grupo de otro. La regularización de la matriz de partida ofrece un mejor y más flexible desarrollo de esta metodología, ya que, aún teniendo en cuenta la especial adaptación del procedimiento para cada caso particular, se contempla su generalización.

El resto del artículo se divide en cuatro partes. En la Sección 2 se presentan las herramientas básicas de análisis para una adecuada sectorización de la red basada en un cluster que tenga en cuenta la estructura y grado de conectividad de la red de abastecimiento. La Sección 3 da un paso más adelante, proponiendo una estrategia que ayude a generalizar los resultados y a hacer más abordable la división de cualquier tipo de grafo. La Sección 4 presenta el caso-estudio de una red de abastecimiento real, donde se propone la aplicación de las técnicas anteriormente presentadas. Por último, una sección de conclusiones y posibles trabajos a desarrollar en el futuro, cierra el artículo.

2. Clustering sobre el grafo de una red de abastecimiento

Antes de comenzar el planteamiento del proceso a seguir, consideraremos la red como un grafo donde los nodos son puntos de consumo y los arcos son tuberías. De esta manera podemos extender procedimientos y resultados de Teoría de Grafos a su aplicación sobre una red de abastecimiento de agua. En particular, adaptaremos un algoritmo basado en *spectral clustering*, con el objetivo de agrupar nodos de consumo y tuberías de carácter similar dentro de cada sector hidráulico.

2.1 Bases de la conceptualización kernel del problema

En los últimos años, una de las formas de abordar problemas complejos ha sido la de proyectar las observaciones en un espacio de características de alta dimensionalidad, donde estos problemas tienen una solución sencilla (usando métodos lineales, e.g.). Una de las ventajas de trabajar en este espacio de características es que, habitualmente, los métodos utilizados únicamente se basan en el resultado de su producto interior. Desde este punto de vista, no será necesario conocer la expresión explícita de estas funciones kernel (Schölkopf & Smola, 2002) y el proceso de análisis será sensiblemente más sencillo. Éste es el ejemplo de los métodos cluster ya que, una vez escogida la función kernel, para trabajar usaremos las propiedades de su reproducción en espacios de Hilbert (RKHS).

Tenemos en cuenta dos propiedades clave de las funciones kernel:

1. Puede capturar la medida aproximada de similitud para cualquier problema y para cualquier dominio. En nuestro caso, nos habilita para trabajar con datos complejos como los provenientes de un grafo.

2. Su evaluación requiere, significativamente, un esfuerzo computacional menor que el necesario en el caso de requerir su expresión explícita en el espacio de características original.

A partir de la información dispuesta en la matriz kernel aplicamos sobre su espectro un algoritmo de clustering que muestre su organización interna y las posibles estructuras que existan en los datos (Shawe-Taylor & Cristianini, 2006).

2.2 Spectral clustering

El análisis de cluster basado en el espectro de una matriz (Ng et al., 2001) es un paradigma de agrupamiento de datos relativamente nuevo. En él se propone usar los autovectores asociados a los autovalores más pequeños (indicativos de partes del grafo fuertemente conectadas) que hayan sido creados por alguna medida de similitud (Karatzoglou, 2006) o, simplemente, calculados a través de la matriz de afinidad del grafo. De esta forma, planteamos resolver una relajación del problema de partición discreta del grafo, que originariamente es *NP-hard*, y así extender la aplicabilidad del algoritmo clásico de *k-medias*. Así, los cortes basados en el segundo autovector¹ garantizan una aproximación al corte óptimo (Chung, 1997). Este análisis puede extenderse a la construcción de clusters de grafos ponderados cuyos nodos correspondan a puntos de datos y los arcos estén relacionados con la distancia entre los puntos.

Existen dos criterios diferentes para dividir los datos en *k* clusters. Uno es usar los dos autovectores asociados a los dos menores autovalores y aplicar sobre ellos el procedimiento de clustering de manera recursiva, hasta obtener los *k* grupos. Otro es hacer uso, directamente, de más autovectores de la matriz kernel original. Resumiendo para este último caso: los *k* primeros autovectores de la matriz de afinidad son usados para formar una matriz $n \times k$, de columnas normalizadas a longitud unidad. Tratando cada fila de esta nueva matriz como un dato usual, podemos usar el algoritmo de *k-medias* (cualquier algoritmo de clustering será válido en esta fase, tomamos el de las *k-medias* por simplicidad) para agrupar estos puntos transformados. Las pertenencias de los puntos originales a los cluster son equivalentes a las pertenencias calculadas sobre los datos transformados.

La Figura 1 muestra el proceso global del algoritmo. Una versión detallada del mismo se encuentra en el trabajo de Ng et al., de 2001.

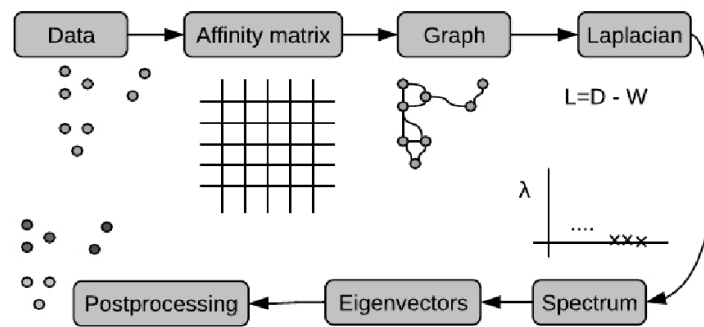


Figura 1. El proceso de spectral clustering. Fuente: Vejmelka, 2009

¹ A lo largo del artículo nos referiremos al orden del autovector según sean menores sus autovalores asociados. De esta manera, el segundo autovector será aquel que está asociado con el segundo menor autovalor.

3. Regularización del grafo de la red

Sea G un grafo consistente en un conjunto de vértices, V , y un conjunto de arcos, E . Definimos la matriz de adyacencia de G como una matriz real, W , tal que $W_{ij} = 1$ si i es vecino de j (es decir, existe un arco que une i con j) $W_{ij} = 0$ en otro caso.

Sea D una matriz diagonal con $D_{ii} = \sum W_{ij}$. El Laplaciano de G se define como $L = D - W$. El Laplaciano normalizado es $\hat{L} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} W D^{-1/2}$, siendo I la matriz identidad. L (y \hat{L}) pueden entenderse como operadores lineales en ciertas funciones, f , y definirse como en la siguiente ecuación:

$$\langle f, Lf \rangle = f^T Lf = -1/2 \sum (f_i - f_j)^2,$$

que se puede contemplar como una semi-norma asociada a f , cuantificando las variaciones locales y la “suavidad” de su dominio (Smola & Kondor, 2003). Además, se puede comprobar que el Laplaciano será invariante a ciertas transformaciones naturales del espacio en el que subyace, convirtiéndole, esencialmente, en único.

Por un grafo regular se entiende aquel que cada vértice posee el mismo número de nodos vecinos (Smola & Kondor, 2003). Una interesante propiedad es que, si el grafo es regular: L , \hat{L} y W comparten los autovectores (Ando & Zhang, 2007). Definimos la clase de funcionales de regularización en grafos como:

$$\langle f, Pf \rangle = \langle f, r(L)f \rangle,$$

donde $r(L)$ es la aplicación del valor de la función $r(\lambda)$ a los autovalores de L , es decir,

$$r(L) = \sum r(\lambda_i) v_i v_i^T,$$

con $\{r(\lambda_i), v_i\}$ que constituye un sistema de autovalores y autovectores de L . De esta manera, es preferible usar el Laplaciano normalizado a L , dado que su espectro está contenido en $[0, 2]$.

Existen ciertas condiciones que la función $r(\lambda)$ debiera cumplir, tal como ser definida positiva y monótonamente creciente, penalizando los mayores autovalores. En el contexto de espectro del grafo, son de especial interés las funciones denominadas de Laplaciano regularizado, $r(\lambda) = 1 + \sigma^2 \lambda$ y de proceso de difusión, $r(\lambda) = \exp(\sigma^2 / 2 \lambda)$. Este interés se basa en que, estas funciones, nos permiten la posibilidad de plantear la solución de las k -medias del proceso de *spectral clustering* a partir de matrices kernel regularizadas, que comparten las soluciones de su sistema de autovalores y autovectores con la matriz original, pero permiten una mejor generalización de los resultados y un adecuado aprovechamiento de características propias del grafo de una red de abastecimiento, como su baja densidad de arcos entre nodos (propiedad de *sparseness*). Así, las matrices kernel a estudio serían: $K = (I + \sigma^2 \hat{L})^{-1}$ en el caso del Laplaciano regularizado y $K = \exp(-\sigma^2 / 2 \hat{L})$ en el del proceso de difusión.

4. Estudio experimental

En la aplicación del proceso de *spectral clustering*, hacemos uso de la matriz kernel del Laplaciano regularizado por el proceso de difusión. Además, para un análisis más completo del sistema, podemos sumar a esta matriz las asociadas a otras variables hidráulicas de interés: como las cotas de los nodos, las coordenadas geográficas y las estimaciones de demanda de agua en cada punto. Esta suma de matrices no es directa. Primero hemos de calcular sus respectivas matrices de similitud, para después, hacerlas simétricas y definidas positivas; pudiéndose entonces entender cada matriz como una matriz kernel. La propiedad aditiva de las matrices kernel (suma de matrices kernel es otra matriz kernel), junto a la que afirma que una matriz kernel multiplicada por un escalar cualquiera, mayor que cero, sigue siendo matriz kernel, nos permite trabajar con una visión conjunta del problema: por un lado, la asociada al grafo (a través del kernel del Laplaciano regularizado, en este caso, por el proceso de difusión) y por otro la que cuenta con otros inputs que pueden ser de interés para construir los sectores (a través de las matrices kernel relacionadas con información de cota y demanda en cada punto, así como sus coordenadas geográficas). La matriz final sobre la que trabajamos será:

$$K = \alpha K_L + (1 - \alpha) \sum \beta_i K_i,$$

combinación lineal de las sumas de matrices kernel. Los parámetros α y los β_i se pueden calibrar dentro del proceso de *spectral clustering*, para la obtención de mejores resultados; o bien, proponer directamente por el interés hidráulico de los diferentes inputs.

4.1 Caso-estudio

Para demostrar la adecuación del presente proceso a la división de la red de abastecimiento en sectores, consideramos el caso de la red real del distrito central de la ciudad de Celaya (perteneciente al estado de Guanajuato, México). El sistema está alimentado por un embalse (*D1*) y cinco depósitos (*E1*, ..., *E5*), con cinco estaciones de bombeo. La red la conforman 479 tuberías y 333 nodos de consumo; su longitud total es de 48 km y la cota de elevación media asciende a 156 metros. La tasa de consumo total de la red es de 91 l/s. La Figura 2 muestra el perfil de la red en Epanet (Rossman, 2000).

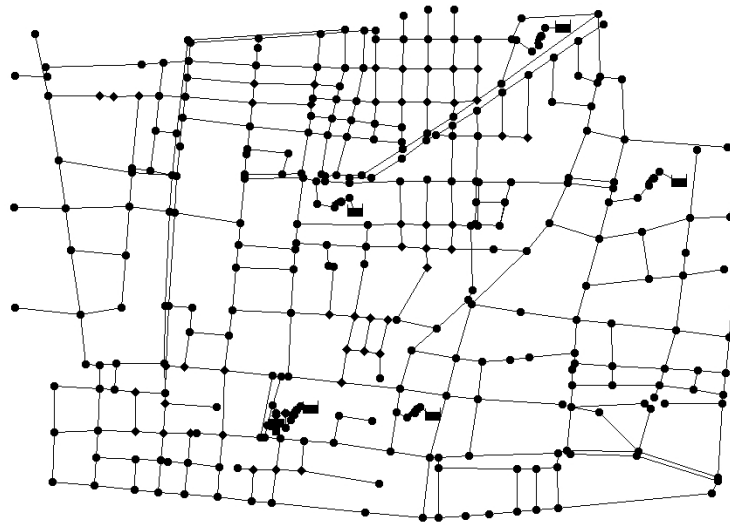


Figura 2. Perfil de la red de abastecimiento de agua usada para este caso-estudio

Nuestro propósito será dividir la red de abastecimiento en 3 sectores. Para alcanzar este objetivo se sigue el algoritmo de *spectral clustering* propuesto en el presente trabajo y se obtienen los siguientes resultados: El tamaño de cada sector, en km de tuberías, es de: 19, 15 y 9.5 km, respectivamente. El diámetro medio (en milímetros) de las tuberías de cada cluster es de: 144, 107 y 130 mm. En esta división, cada sector es abastecido por, al menos, un depósito (los detalles de los puntos de abastecimiento por sector y los datos de cotas y demandas promedio se observan en la Tabla 1). Será necesario el cierre de 26 válvulas para el aislamiento de estos sectores. La función objetivo del algoritmo de cluster garantiza que éste es el número mínimo de válvulas que deben cerrarse para llevar a cabo la partición.

Tabla 1. Descripción de la sectorización resultante del proceso propuesto para el caso-estudio de Celaya

Sector	Nodos	Tuberías	Fuentes de abastecimiento	Cota	Demanda
Sector 1	122	189	E1+E3+D1	157	35
Sector 2	127	131	E4	155	25
Sector 3	84	133	E2+E5	156	31

En la Figura 3 podemos ver la distribución de los tres sectores en la red de abastecimiento. Esta configuración final fue simulada con Epanet, validando así los resultados.

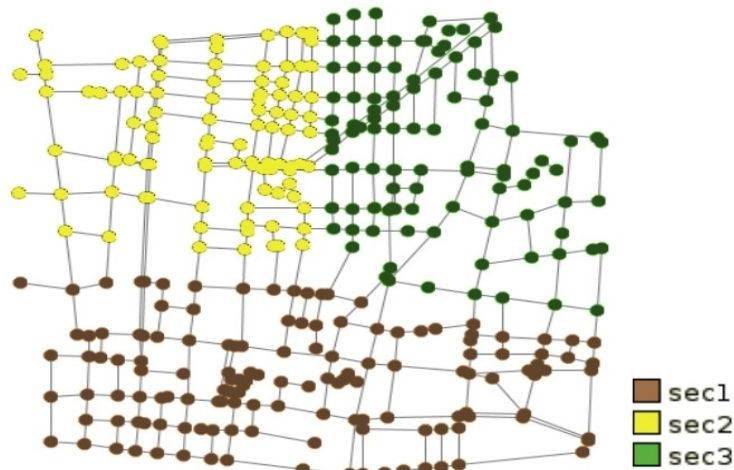


Figura 3. División de la red del caso-estudio en tres sectores

5. Conclusiones

De manera clásica, una división de la red de abastecimiento en sectores tiene como objetivo la mejora de la estrategia de detección de fugas, usando la información de las roturas de tuberías, presiones y demanda. Este artículo propone ampliar la perspectiva de este objetivo, teniendo en cuenta la cantidad de diferente información que puede ser incluida como criterio de división de la red. A las variables utilizadas en este trabajo se le puede añadir cualquier otra que sea de interés (por ejemplo, la inclusión de la edad de las tuberías si queremos sectorizar con fines a establecer un plan de rehabilitación), bastará con seguir la misma metodología de adición de matrices kernel provenientes de matrices de similitudes.

Comparado con otros métodos, que sólo usan información gráfica o vectorial, pero no ambas, el *spectral clustering* toma ventaja por su capacidad para recoger y utilizar información de cualquier tipo. Además, el algoritmo propuesto es computacionalmente eficiente (gracias a la regularización del grafo) y de estructura flexible, ya que está abierto a diferentes configuraciones de pesos con que se suman las matrices kernel. Éste es un posible punto de interés para continuar investigando en esta línea y desarrollar técnicas que calibren la bondad de estas ponderaciones.

Por otro lado, la propuesta de regularización del grafo de la red generaliza el proceso descrito, además de explotar la característica de baja densidad en número de arcos del grafo de una red de distribución, respecto de grafos de otro tipo.

Agradecimientos

Este trabajo ha sido desarrollado con el apoyo del proyecto IDAWAS (Ref. DPI 2009-11591), financiado por el Ministerio de Ciencia e Innovación y fondos FEDER-UE y por la ayuda complementaria ACOMP/2010/146, de la Conselleria de Educación de la Generalitat Valenciana.

Referencias bibliográficas

- Ando, R. K.; Zhang, T. (2007) Learning on graph with Laplacian regularization. NIPS 19, 2007.
- Bougadis, J.; Adamowski, K.; Diduch, R., (2005) “Short-term municipal water demand forecasting”, Hydrological Processes vol. 19, pp. 137 – 148
- Chung, F. (1997) “Spectral graph theory”, CBMS Regional Conference Series in Mathematics
- Covas, D.; Ramos, H.; Almeida, A.B. (1999) “Leakage monitoring control and management of water distribution systems: a challenge for the XXI century”, Proceedings of XXVIII IAHR congress, (Graz, Austria.)

- Herrera, M.; Torgo, L.; Izquierdo, J.; Pérez-García, R. (2010) “*Predictive models for forecasting hourly urban water demand*”, *Journal of Hydrology*, 387, pp. 141-150
- Hunaidi, O.; Brothers, K. (2007) “*Optimum size of District Metered Areas*”, *Water Loss 2007 Conference Proceedings* (Bucharest, Romania)
- Izquierdo, J.; Herrera, M.; Montalvo, I. & Pérez-García, R. (2009) “*Agent-based division of water distribution systems into District Metered Areas*”, *ICSOF 2009: Proceedings of the IV International Congress of Software and Data Technologies* (Sofia, Bulgaria), Vol. 2, pp. 83-90
- Izquierdo, J.; Montalvo, I.; Pérez-García, R.; Herrera, M. (2008) “*Sensitivity analysis to assess the relative importance of pipes in water distribution networks*”, *Mathematical and Computing Modelling*, vol. 48, pp. 268 – 278
- Kondor, R.I.; Lafferty, J. (2002) “*Diffusion kernels on graphs and other discrete input spaces*”, *Proc. of the 19th International Conference on Machine Learning*, pp. 315–322, C. Sammut & A. Hofmann, editors
- MacDonald, G.; Yates, C.D. (2005) “*DMA design and implementation, a North American context*”, *Leakage 2005 Conference Proceedings* (Halifax, Canada)
- Misiunas, D. (2005) “*Failure monitoring and asset condition assessment in water supply systems*”, Ph.D. Thesis, Lund University, Sweden
- Ng, A. Y.; Jordan, M.I.; Weiss, Y. (2001) “*On spectral clustering: Analysis and an algorithm*”, *Advances in Neural Information Processing Systems 14*, pp. 849–856, MIT Press
- Rossman, L. (2000) “*EPANET–User’s Manual*”, United States Environmental Protection Agency (EPA)
- Shawe-Taylor, J.; Cristianini, N. (2006) “*Kernel methods for pattern analysis*”, 462 pp., Cambridge University Press
- Shi, J.; Malik, J. (2000) “*Normalized cuts and image segmentation*”, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, pp. 888–905
- Smola, A. J.; Kondor, R. I. (2003) “*Kernels and regularization on graph*”, *Proc. of the 16th Annual Conference on Computational Learning Theory and the 7th Kernel Workshop*
- Tzatchkov, V.G.; Alcocer-Yamanaka, V.H.; Bourguett-Ortiz, V. (2006) “*Graph theory based algorithms for water distribution network sectorization projects*”, *VIII Annual Water Distribution Systems Analysis Symposium* (Cincinnati, Ohio, USA)
- Vejmelka, M. (2009) “*Spectral graph clustering*”, *Seminar z Umele Inteligence*
- Walski, T. M.; Chase, D.V.; Savic, D.A. (2001) “*Introduction to water distribution modelling*”, *Haestad Methods, Inc.*, 37 Brookside Road, Waterbury, CT 06708 (USA), 441 pp
- Water Loss Task Force – IWA (2007) “*District Metered Areas: Guidance notes – DRAFT*”, (disponible on-line en: <http://www.iwaom.org/wlrf>)
- Zhu, X.; Kandola, J.; Lafferty, J.; Ghahramani, Z. (2006) “*Graph kernels by spectral transforms*”, MIT Press