

Uso de teoría de grafos para determinar zonas de vulnerabilidad en redes de abastecimiento de agua

Joanna Gutiérrez-Pérez, Manuel Herrera, Rafael Pérez-García, Joaquín Izquierdo

Grupo Fluing - Instituto de Matemática Multidisciplinar (IMM) – Universidad Politécnica de Valencia, Camino de Vera s/n, 46022, Valencia – España.

{joagupre, mahefe, rperez, jizquier}@upv.es

1 Introducción

Las infraestructuras críticas, como lo son las redes de abastecimiento de agua, están formadas por redes con estructura compleja. Su estructura se caracteriza por tener elementos altamente interdependientes, cuyo grado de complejidad depende de las configuraciones y el nivel de interacción entre sus elementos. Un aspecto importante a considerar en el análisis de la topología de redes complejas, es que la estructura afecta el funcionamiento y por lo tanto, la arquitectura de la red puede ser usada para entender y predecir procesos dinámicos que afecten su desempeño y la distribución del flujo, así como también, su tolerancia y estabilidad.

En el contexto del análisis de la vulnerabilidad de redes complejas, las técnicas de teoría de grafos facilitan la representación y su análisis. Los grafos son una abstracción útil que permiten simular diversos escenarios reales. Por lo tanto, dependiendo del tipo de red que se analice, los nodos y arcos pueden ser definidos de diferentes maneras (nodos de consumo, tanques, tuberías o estaciones de bombeo).

En relación con lo anterior, debido a las perspectivas que proporcionan las técnicas de simulación y optimización basadas en teoría de grafos, éstas han ganado importancia en el análisis de las vulnerabilidades potenciales de las redes de infraestructuras crítica. Entre los trabajos realizados con técnicas de teoría de grafos, se pueden encontrar estudios enfocados a la evaluación de la vulnerabilidad y el análisis de la estructura de redes de transporte (Masucci *et al.*, 2009; Grubestic *et al.* 2008; Zio y Sansavini, 2007), de redes de energía eléctrica (Bompard *et al.*, 2009; Holmgren, 2006; Crucitti *et al.*, 2005), de redes informáticas de distribución de información (Aoun *et al.*, 2006; Albert *et al.*, 1999).

En cuanto a trabajos sobre redes de abastecimiento de agua, se puede citar a Jacobs y Goulter (1989, 1988), quienes demostraron que las redes que no son vulnerables a fallos, son grafos regulares con igual número de arcos incidentes en cada nodo, mientras que, la relación inversa no necesariamente es cierta, debido a la presencia de puentes y nodos de articulación. En Kessler *et al.* (1990) desarrollaron una metodología basada en teoría de grafos para obtener el diseño de menor coste de redes de abastecimiento de agua no vulnerables y fiables. Ostfeld y Shamir (1996) y Ostfeld (2005) introdujeron el concepto de *backups* (subsistemas dentro de la red que sobreviven a la ocurrencia de un fallo y cuyo funcionamiento es explícitamente tomado en consideración en las fases de trazado, diseño y operación) y utilizaron grafos para analizar su selección en el nivel uno de redundancia del sistema, tomando como caso de estudio un sistema de abastecimiento de agua sometido a fallos. En Yazdani y Jeffrey (2010), los sistemas de abastecimiento de agua son considerados como grafos planos con características de redes complejas y con una importante relación topológica entre los elementos de la red. En su trabajo utilizan técnicas determinísticas de redes complejas y teoría de grafos para cuantificar la conectividad y estimar la redundancia. Por otro lado, en trabajos de análisis de fiabilidad de redes, se demostró que el uso puramente topológico de la teoría de grafos tiene alcances limitados. Tal es el caso de Walski (1993), quien demostró que la representación de la conectividad de los nodos, no explica la importancia de las válvulas en una red de abastecimiento de agua. Por su parte, Ostfeld y Shamir (1993) clasificaron los fallos en dos tipos, los fallos de los componentes del sistema y los fallos en satisfacer la demanda de los consumidores, y señalaron que esos dos tipos de fallos no se deben considerar de forma separada, ya que están fuertemente relacionados. En los temas de análisis de la conectividad, fiabilidad y riesgo en el diseño de redes de abastecimiento, se pueden mencionar los trabajos realizados por Ostfeld (2004), Babayan *et al.* (2005) y Kapelan *et al.* (2006). Otro aspecto analizado en redes de abastecimiento de agua es la sectorización. Tzatchkov *et al.* (2006) abordan ese tema presentando algoritmos basados en teoría de grafos para la obtención de sectores independientes, así como, para la obtención del conjunto de nodos que pertenecen a cada sector, el conjunto de nodos desconectados y el nodo fuente de contribución. En la misma línea, Herrera *et al.* (2010a y 2010b) desarrollan una metodología basada en aprendizaje semi-supervizado para la conformación de *clusters* de abastecimiento de agua. Dicha metodología utiliza tanto información gráfica como vectorial.

Por otro lado, el ordenamiento de los nodos de una red es una propuesta reciente. Trabajos como el de Herrera *et al.* (2011), demuestra las nuevas posibilidades en esta área mediante la adaptación del algoritmo *PageRank* de Google para el estudio de la importancia de los nodos. Grubestic *et al.* (2008) y Yazdani y Jeffrey (2010), investigaron aplicaciones de teoría de grafos y flujo en redes, en el análisis de la vulnerabilidad y robustez de sistemas de abastecimiento de agua. En ambos trabajos se revisaron las distintas medidas existentes para la indexación de nodos de la red, en su gran mayoría basadas en estadística y análisis espectral. Otras referencias básicas sobre el ordenamiento de los elementos en una red de abastecimiento son la de Michaud y Apostolakis, (2006), quienes para evaluar la importancia de cada nodo, introdujeron una metodología basada en estudio de escenarios de riesgo. Izquierdo *et al.* (2008), trabajaron en las tuberías, evaluando su importancia relativa respecto del proceso de distribución del agua a través de la red.

El objetivo de este trabajo es presentar un método innovador para la cuantificación de la importancia relativa de los nodos (*ranking*), por medio de medidas de teoría de grafos, que ayuden a la definición de zonas de vulnerabilidad en redes de abastecimiento de agua. La relevancia de la ordenación de los nodos es su relación con la vulnerabilidad de la estructura de la red, así como su utilidad para el análisis de aspectos como la calidad del agua, la localización de sensores y el estudio de la viabilidad de planes de rehabilitación, entre otros (Herrera *et al.*, 2011). Para lograr tal objetivo, se propone la utilización de dos algoritmos comúnmente utilizados para la ordenación de los nodos en redes, que son, el *PageRank* y *HITS*. La metodología seguida en este trabajo para la obtención del *ranking* de los nodos, está basada en los trabajos de Herrera *et al.* (2011 y 2010a). El presente artículo da una breve introducción a los algoritmos mencionados, posteriormente, se prueba la metodología propuesta en una red de abastecimiento de agua real y se presentan los resultados obtenidos. Finalmente, se cierra el documento con un apartado de conclusiones.

2 Breve descripción de los algoritmos de *ranking* basados en teoría de grafos

El análisis de redes complejas emplea técnicas de teoría de grafos para clasificar diferentes tipos de redes, analizar sus estructuras y cuantificar su vulnerabilidad, robustez y redundancia. La estructura topológica de las redes se describe matemáticamente en términos de un grafo. De manera formal, un grafo simple o no dirigido se define como $G=(V, E)$ donde V es un conjunto finito y no vacío de elementos llamados nodos y E es un conjunto de pares no ordenados de elementos distintos de V llamados arcos. Por lo tanto $G=(V, E)$ denota un grafo finito, no vacío y no dirigido.

La teoría de grafos se basa en un conjunto de medidas para evaluar a las redes. Esas medidas pueden ser clasificadas como medidas de conexión básica, medidas estadísticas y medidas espectrales. Las medidas de conectividad básica, , representan la cohesión y adhesión de la red y su sensibilidad a la eliminación de algunos de sus elementos. Las medidas estadísticas cuantifican las propiedades subyacentes de la red, mediante la construcción de patrones de frecuencia y el establecimiento de subredes. En cuanto a las medidas espectrales, estas provienen de la matriz de autovectores asociada a la matriz de adyacencia del grafo, y relacionan la topología de la red con la intensidad de la conectividad y la cohesión gráfica. Estas mediciones van más allá de las estadísticas y revelan información sobre la conectividad de la red o sus niveles de tolerancia a los fallos.

Considerados dentro de los métodos espectrales, los algoritmos de *ranking* u ordenación basados en teoría de grafos son una manera de determinar la importancia de un nodo dentro de un grafo, en base a información de la estructura del grafo. Algoritmos como el *HITS* (*Hyperlinked Induced Topic Search*) desarrollado por Kleinberg (1999) y *PageRank* (Brin y Page, 1998), han sido utilizados en el análisis de la referenciación, de redes sociales y en el análisis de los enlaces en la estructura de la red mundial de redes (*World Wide Web*). El *PageRank* le da mayor importancia a la normalización de los pesos de los vínculos entre páginas así como a la navegación aleatoria, mientras que *HITS* resalta el reforzamiento mutuo entre páginas *web* denominadas “autoritarias” y “centralizadas”.

En general, tanto el *PageRank* puede como el *HITS*, pueden ser vistos como una medida de la importancia relativa de los nodos en cualquier tipo de grafo, lo que nos permite plantear su funcionamiento en el entorno de los grafos de las redes de abastecimiento de agua (Herrera *et al.*, 2011).

2.1 Algoritmo *PageRank*

El algoritmo *PageRank* es uno de los algoritmos de ordenación más populares. El *PageRank* es el método de cálculo que usaron los fundadores de *Google* para asignar cuantitativamente un *ranking* a las páginas *web* y clasificarlas según su importancia. A diferencia de otros algoritmos, *PageRank* integra el impacto tanto de los enlaces entrantes como salientes de cada nodo V_i y produce como resultado un solo conjunto de valores. Por lo tanto, el valor de *PageRank* para cada nodo está dado por (Brin y Page, 1998):

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

donde d es un parámetro que se encuentra en el rango de $[0, 1]$; sin embargo, el valor recomendado para d es de 0,85 (Brin y Page, 1998). El conjunto de todos los valores forman una distribución de probabilidad sobre el conjunto de todos los nodos de la red, por lo que la suma de los *PageRanks* debe ser igual a 1.

2.2 Algoritmo *HITS*

El algoritmo *HITS* fue desarrollado, al igual que *PageRank*, para dar solución al problema de búsqueda en la *web*. El algoritmo hace uso de la estructura de conectividad de la *web* con el objetivo de encontrar y ordenar páginas relevantes para un tópico particular. *HITS* es un algoritmo iterativo diseñado para la ordenación por importancia de páginas *web* de acuerdo a su grado de “autoridad”. El algoritmo hace una distinción entre “autoridades” (páginas con un gran número de enlaces entrantes a ella) y “centros” (páginas con un gran número de enlaces salientes).

En *HITS*, se produce para cada nodo V_i dos valores que lo caracterizan, el *authority score* ($HITS_A(V_i)$) y el *hub score* ($HITS_H(V_i)$). Se parte de la premisa de que una página con un buen valor de *authority* es señalada por muchas otras con un buen *hub*; y un buen valor de *hub*, se dirige a muchos valores altos de *authority*. La relación que existe es de mutuo reforzamiento y es representada con las siguientes expresiones (Kleinberg, 1999):

$$HITS_A(V_i) = \sum_{V_j \in \text{In}(V_i)} HITS_H(V_j) \quad (2)$$

$$HITS_H(V_i) = \sum_{V_j \in \text{Out}(V_i)} HITS_A(V_j) \quad (3)$$

Cada uno de estos algoritmos inicia asignando valores arbitrarios a cada uno de los nodos de la red, las iteraciones se realizan hasta que converge por debajo de un umbral determinado. Los valores obtenidos después de los cálculos, son asociados a cada uno de los nodos, los cuales representan su importancia dentro de la red. Cabe señalar que los valores finales no son afectados por los valores iniciales, únicamente puede ser diferente el número de iteraciones realizadas para la convergencia.

3 Adaptación de los algoritmos a las redes de abastecimiento

Entendiendo una red de abastecimiento de agua como un grafo de características especiales, se puede abstraer la noción de página *web* para considerarla como un nodo de consumo de la red. Mientras que los enlaces entre páginas, se entienden como las tuberías que unen diferentes nodos. Trabajos como el de Chung y Zhao (2008) propusieron el carácter general que puede tener el algoritmo *PageRank* a la hora de establecer la ordenación por importancia de los vértices de un grafo cualquiera, abriendo posibilidades para el análisis de diferentes tipos de redes.

En base a los trabajos de Herrera, *et al.* (2011 y 2010a), en este trabajo se plantea una adaptación de los algoritmos descritos, para la obtención del *ranking* de los nodos de una red de abastecimiento de agua. La

finalidad del método es la obtención de un vector, también denominado *PageRank* que proporciona los valores relacionados con la importancia relativa de los nodos. Asimismo, los vectores de *HITS*, que corresponde a los valores de *authority* y *hub* de cada uno de los nodos. Con el post-procesamiento de los valores de los vectores resultantes, son conformadas las zonas de vulnerabilidad en la red de abastecimiento.

3.1 Proceso de agrupamiento de los nodos: formación de las zonas de vulnerabilidad

El proceso para la determinación de las zonas de vulnerabilidad en la red de abastecimiento, se basan en los vectores resultado del *PageRank* y el *HITS*. La información de entrada está concentrada en matrices e incluye toda la información disponible de la red.

Para la obtención y el tratamiento de la información se hicieron algunas consideraciones. Primero, que la red de abastecimiento es un grafo particular que incluye tanto información geográfica como de conectividad. Segundo, que los nodos y arcos del grafo corresponden a los puntos de consumo de la red de abastecimiento y a las tuberías de la red, respectivamente. Por último, que los pesos de los nodos están dados por los valores de demanda y los pesos de los arcos por el diámetro de las tuberías. De esta manera se construye la matriz de afinidad del grafo asociada a la red.

El interés particular es incluir la información de los vectores de *PageRank* y *HITS*. Para ello, se construyeron matrices de disimilitud usando los valores de dichos vectores. Cabe mencionar que en este proceso es posible añadir tantas matrices como categorías de información se tengan.

Una vez construidas las matrices con la información de entrada, estas son transformadas a matrices *kernel*. Este proceso de transformación a matrices *kernel*, propone una integración de los datos de entrada, ofreciendo una forma regularizada para disponer de la información y un espacio adecuado de operación. Para esto, los datos son escalados a valores entre 0 y 1. Luego, se introduce en cada una de las matrices que se tenga, una diagonal de 1. Posteriormente, las matrices son reflejadas a través de sus diagonales para ser transformadas a simétricas (es decir, las matrices de entrada son triangulares). (Herrera *et al.*, 2010a). La finalidad de transformar a matrices *kernel* se debe a que se simplifica el tratamiento de datos no vectoriales. Además, gracias a su propiedad de que la suma de dos o matrices *kernel* es igual a otra matriz *kernel*, es posible construir una matriz acumulativa y añadir información de diferentes tipos, como es el caso de los valores del *PageRank* y *HITS*. Esto se define con la siguiente ecuación:

$$K = \lambda_A K_A + \sum_{i \in I} \lambda_i K_i \quad (4)$$

Donde K es la matriz *kernel* para el proceso de *clustering*, K_A es la matriz *kernel* relacionada a la matriz de afinidad del grafo y K_i , $i \in I$, es la matriz asociada a los *inputs* de interés en el proceso de construcción de las zonas. Finalmente, λ_A y λ_i , $i \in I$, son los pesos de la combinación lineal. Aquí, cabe mencionar que se decidió dar un mayor peso a cada una de las matrices *kernel* de los algoritmos, ya que la intención era ver la influencia que estos tenían en la determinación de las zonas.

Después de la transformación de las matrices de datos a matrices *kernel*, aplicamos el proceso de *clustering* a la red de abastecimiento. Para el tratamiento de estos datos, se aplicó el algoritmo de *clustering* semi-supervisado (Herrera *et al.* (2010), con la previa transformación de la matriz de afinidad del grafo a matriz *kernel*. A continuación, se une más información agregando otras matrices *kernel*, que en este caso particular, serían las matrices *kernel* de las matrices de disimilitud del *PageRank* y de *HITS*. En la Tabla 1 se resume el proceso descrito.

Tabla 1 Proceso de *clustering* semi-supervisado, (Herrera, *et al.*, 2010a)

Algoritmo: clusters de abastecimiento de agua mediante aprendizaje semi-supervisado	
1.	Abstracción de la red de abastecimiento como un grafo
2.	Construcción de las matrices de disimilitud (de la variable a analizar, en este caso del <i>PageRank</i> y el <i>HITS</i>)
3.	Transformación de las matrices de datos a matriz <i>kernel</i>
4.	Cálculo del espectro de la matriz (<i>spectral clustering</i>)
5.	Asignación de los <i>clusters</i> a los datos originales

En relación con el punto 4 de la tabla anterior, el cálculo del espectro de la matriz, el *spectral clustering* es un enfoque relativamente nuevo en el tema del agrupamiento de los nodos. En el desarrollo del proceso, se ha utilizado el algoritmo propuesto por Ng *et al.* (2001).

Todos los cálculos para la obtención de las zonas de vulnerabilidad, (la obtención del grafo de la red de abastecimiento, las matrices de afinidad y disimilitud asociadas al mismo, los vectores resultado de los algoritmos y el agrupamiento de los nodos) fueron realizados con el soporte del Lenguaje R (*R Development Core Team*, 2010) mediante sus librerías *igraph* y *kernlab*. EL post-procesamiento para la visualización de las zonas en la red de abastecimiento, se realizó con el programa *NetLogo* (Wilensky, 1999).

4 Caso de estudio

Comenzando con la matriz de afinidad del grafo de la red de abastecimiento, el siguiente paso de la metodología propuesta es la transformación de las tuberías, los nodos y la demás información de los algoritmos *PageRank* y *HITS*, en matrices *kernel*, para lo cual, aplicamos el proceso de *clustering* semi-supervisado. Con el objetivo de mostrar el alcance de la metodología propuesta, en esta sección se muestran los resultados obtenidos de su aplicación en un caso de estudio de una red de abastecimiento real.

El caso de estudio es la red de abastecimiento que se presenta en la Figura 1. La red está formada por 107 nodos, 134 tuberías y 3 fuentes de abastecimiento. La red es de tamaño medio, por lo que se consideró dividirla en dos zonas principales de vulnerabilidad, para conservar cierta homogeneidad en el número de nodos contenidos en cada zona. Para este propósito, aplicamos el procedimiento descrito anteriormente (apartado 3.1).

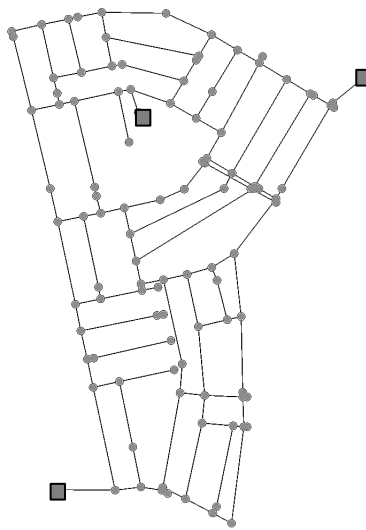


Figura 1 Estructura de la red de estudio

4.1 Resultados

Después del proceso de cálculo de la información para la definición de las zonas de vulnerabilidad, se obtuvieron los siguientes resultados. Primero, se obtuvo una división de la red sin la información de los algoritmos *PageRank* y *HITS* (Figura 2). Los resultados muestran dos zonas, que contienen 25 y 82 nodos. La configuración de estas zonas se basa, tanto en la información de la red de abastecimiento, como en la información de conectividad del grafo. Más adelante veremos, que al introducir una medida de importancia relativa de los nodos, las zonas se modifican.

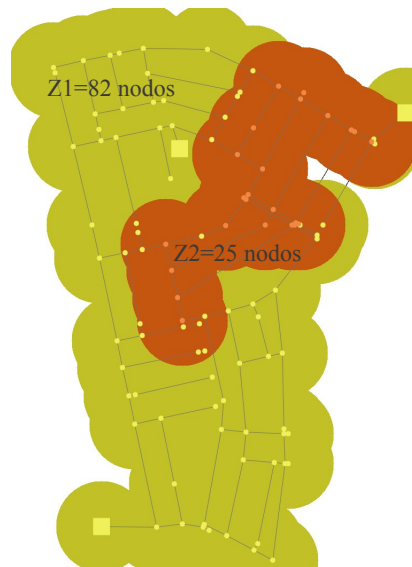


Figura 2 Representación de la división de la red mediante *spectral-lustering*

Para tener una idea más clara de la distribución de los valores de *PageRank* y de *HITS* en la red, a continuación se presentan las Figuras 3(a) y 3(b). En el caso del *PageRank*, 20 nodos registraron los valores más altos, siendo el Nodo 3 el valor más alto (0,01640). Mientras que en *HITS*, los valores más altos se obtuvieron en 5 nodos, el Nodo 21 (0.84580) y el Nodo 40 (1,0) fueron los valores más altos.

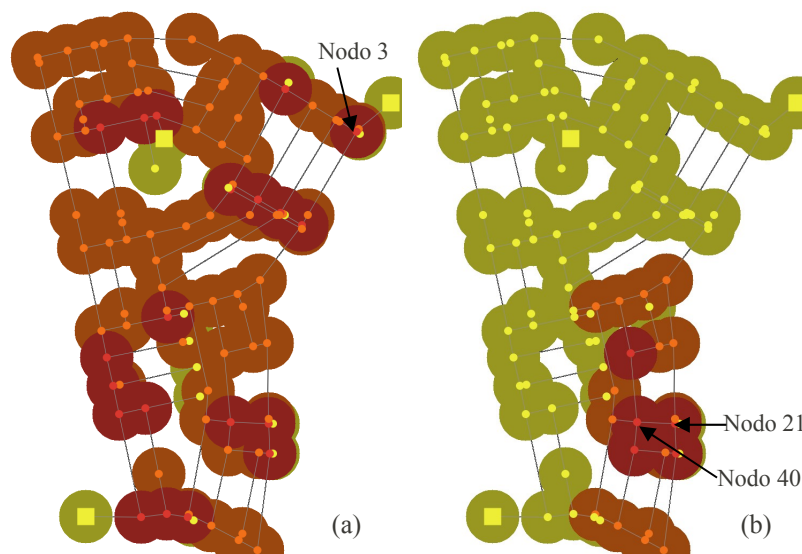


Figura 3 Distribución de los valores de *PageRank* (a) y *HITS* (b) en la red de abastecimiento

Como se indicó, el objetivo planteado es la división de la red en zonas de vulnerabilidad o importancia de los nodos. La división obtenida con el *PageRank* (Figura 4), muestra dos zonas con 59 y 48 nodos cada una. En comparación con la división obtenida sólo con el *spectral-clustering*, se puede observar que hay una variación en la conformación de las zonas, siendo la parte norte de la red la más modificada. En cuanto a las zonas obtenidas con el *HITS* (Figura 5), la configuración es muy similar a la obtenida con el *spectral-clustering*, lo que indica que la influencia de esta medida ha sido menor que el *PageRank*. Las zonas obtenidas contienen 22 y 85 nodos.

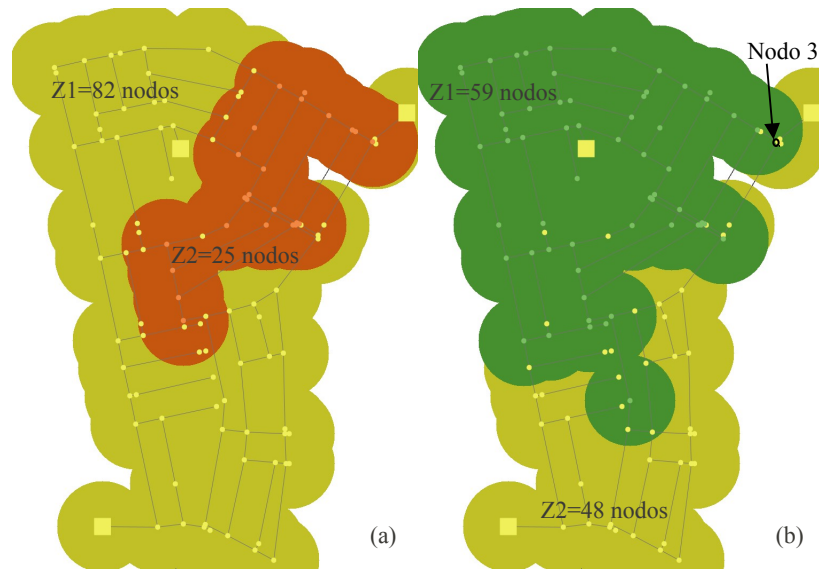


Figura 4 Comparación de la división de la red con *spectral-clustering* (a) y el algoritmo *PageRank* (b)

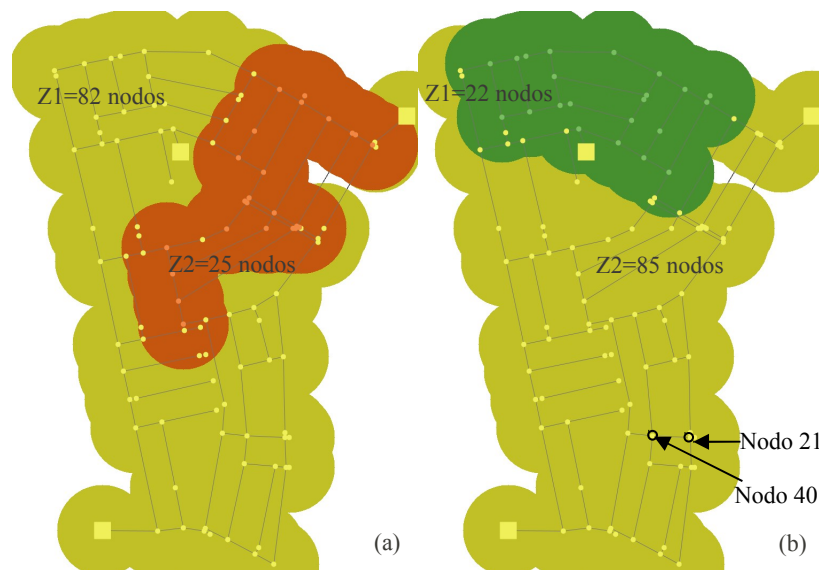


Figura 5 Comparación de la división de la red con *spectral-clustering* (a) y el algoritmo *HITS* (b)

5 Conclusiones

En el presente trabajo se propuso un enfoque para la división de la red mediante medidas de teoría de grafos, específicamente con información de algoritmos de cuantificación de la importancia de los nodos *PageRank* y *HITS*. Por otro lado, la metodología de *clustering* semi-supervisado, ofrece grandes posibilidades a la hora de integrar diferente información. Comparada con otras metodologías, que utilizan únicamente información vectorial o información gráfica, el *clustering* semi-supervisado utiliza ambos tipos de información. Esta flexibilidad nos ha permitido incluir la información de los algoritmos para la formación de las zonas de vulnerabilidad.

La división de la red basada en los algoritmos presentados, es una forma simplificada de identificar los puntos más vulnerables. Cuantificar cuál nodo es más o menos importante, permite saber su influencia en los demás elementos de la red. De esta manera, se podría implementar alguna medida de seguridad en aquellos nodos que tengan más relevancia, con el fin de proteger el funcionamiento de la red de abastecimiento.

Cada uno de los algoritmos ha dado resultados aceptables y no se puede decir que uno sea mejor que otro. Sino que, los resultados dependen, principalmente, de la estructura de la red. En el caso de estudio presentado, el algoritmo *PageRank* ha demostrado tener más influencia en la determinación de las zonas.

Este trabajo es un primer acercamiento a la propuesta de incluir este tipo de medidas, basadas en teoría de grafos, para el análisis de las vulnerabilidades de una red de abastecimiento. Se pretende, en trabajos futuros, realizar un análisis de sensibilidad en cuanto a la variación de los pesos de las matrices kernel. Además, surge el interés de incluir información asociada a parámetros de calidad del agua para obtener zonas de vulnerabilidad a la contaminación.

Agradecimientos

Este trabajo ha sido desarrollado con el apoyo del proyecto IDAWAS (Ref. DPI2009-11591), 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.

6 Referencias bibliográficas

- Albert, R.; Jeong, H.; Barabasi, A. -L., (1999). *Diameter of the World-Wide Web*. Nature (London), No. 401, pp. 130-131.
- Aoun, B.; Boutaba, R., (2006). *Clustering in WSN with Latency and Energy Consumption Constraints*. Journal of Network and Systems Management, Vol. 14, No. 3, pp. 415-439.
- Babayan, A.; Kapelan, Z.; Savic, D.; Walters, G., (2005). *Least-Cost Design of Water Distribution Networks under Demand Uncertainty*. Journal of Water Resources Planning and Management, Vol. 131, No. 5, pp. 375-382.
- Bompard, E.; Napoli, R.; Xue, F., (2009). *Analysis of structural vulnerabilities in power transmission grids*. International Journal of Critical Infrastructures Protection, No. 2, pp. 5-12.
- Brin, S.; Page, L., (1998). *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. [Http://infolab.stanford.edu/backrub/google.html](http://infolab.stanford.edu/backrub/google.html).
- Crucitti, P.; Latora, V.; Marchiori, M., (2005). *Locating critical lines in high-voltage electrical power grids*. Fluctuation and Noise Letters, Vol. 5, No. 2, pp. L201-L208.
- Grubestic, T.; Matisziw, T.; Murray, A.; Snediker, D., (2008). *Comparative Approaches for Assessing Network Vulnerability*. International Regional Science Review, Vol. 31, No. 1, pp. 88-112.
- Herrera, M.; Gutiérrez-Pérez, J.; Izquierdo, J.; Pérez-García, R., (2011). *Ajustes en el modelo Page-Rank de Google para el estudio de la importancia relativa de los nodos de la red de abastecimiento*. Proceedings, X Seminario Iberoamericano de planificación, proyecto y operación de sistemas de abastecimiento de agua (SEREA). Morelia, México.
- Herrera, M.; Canu, S.; Karatzoglou, A.; Pérez-García, R.; Izquierdo, J., (2010a). *An approach to water supply clusters by semi-supervised learning*. International Environmental Modelling and Software Society (iEMSs) 2010 International Congress on Environmental Modelling and Software.
- Herrera, M.; Izquierdo, J.; Pérez-García, R.; Montalvo, I., (2010b). *Water Supply Clusters based on a Boosting Semi-Supervised Learning Methodology*. Proceedings of the Seventh International Conference on Engineering Computational Technology, Civil-Comp Press. Stirlingshire, UK, Paper 22
- Holmgren, Å., (2006). *Using Graph Models to Analyze the Vulnerability of Electric Power Networks*. Risk Analysis, Vol. 26, No. 4, pp. 955-969.

- Izquierdo, J.; Montalvo, I.; Pérez-García, R.; Herrera, M., (2008). *Sensitivity analysis to asses the relative importance of pipes in water distribution networks*. Mathematical and computing Modeling, Vol. 48, pp. 268-278.
- Jacobs, P.; Goulter, I., (1989). *Optimization of Redundancy in Water Distribution Networks Using Graph Theoretic Principles*. Engineering Optimization, Vol. 15, No. 1, pp. 71-82.
- Jacobs, P.; Goulter, I., (1988). *Evaluation of methods for decomposition of water distribution networks for reliability analysis*. Civil Engineering Systems, Vol. 5, No. 2, pp. 58-66.
- Kapelan, Z.; Savic, D.; Walters, G.; Babayan, A.,. (2006). *Risk and Robustness-based solutions to a multi-objective water distribution system rehabilitation problem under uncertainty*. Water Science and Technology, Vol. 53, No. 1, pp. 61-75.
- Kessler, A.; Ormsbee, L.; Shamir, U.,. (1990). *A methodology for Least-Cost Design of Invulnerable Water Distribution Networks*. Civil Engineering Systems, Vol. 7, No. 1, pp. 20-28.
- Kleinberg, J., (1999). *Authoritative sources in a hyperlinked environment*. Journal of the ACM, Vol. 48, pp.604-632.
- Masucci, A.; Smith, d.; Crooks, A.; Batty, M., (2009). *Random planar graphs and the London street network*. The European Physical Journal B, Vol. 71, No. 2, pp. 259-271.
- Michaud, D.; Apostolakis, G. E., (2006). *Methodology for ranking the elements of water supply networks*. Journal of infrastructure Systems, Vol. 12, No. 4, pp. 230-242.
- Ng, A.; Jordan, M.; Weiss, Y., (2001). *On spectral clustering: Analysis and an algorithm*. In Advances in Neural Information Processing Systems, Vol. 14, pp. 849-856. MIT Press.
- Ostfeld, A., (2005). *Water Distribution Systems Connectivity Analysis*. Journal of Water Resources Planning and Management, Vol. 131, No. 1, pp. 58-66.
- Ostfeld, A., (2004). *Reliability analysis of water distribution systems*. Journal of Hydroinformatics, Vol. 6, No. 4, pp. 281-294.
- Ostfeld, A.; Shamir, U., (1996). *Design of optimal reliable multi-quality water-supply systems*. Journal of Water Resources Planning and Management, Vol. 122, No. 5, pp. 322-333.
- Ostfeld, A.; Shamir, U., (1993). *Incorporating reliability in optimal design of water distribution networks-review and new concepts*. Reliability Engineering and Systems Safety, No. 42, pp. 5-11.
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical computing, Vienna, Austria, ISBN: 3-900051-07-0,
- Tzatchkov, V.; Alcocer-Yamanaka, V.; Bourguett, V., (2006). *Graph theory based algorithms for water distribution network sectorization projects*. Proceedings of the 8th Annual Water Distribution Systems Analysis Symposium, Cincinnati, Ohio, USA.
- Walski, T., (1993). *Water Distribution valve topology for reliability analysis*. Reliability Engineering and System Safety, No. 42, pp. 21-27.
- Wilenski, U., (1999). *NetLogo: Center for Connected Learning Comp. Based Modeling*. Evanston, IL: Northwestern Univ. (<http://ccl.northwestern.edu/netlogo/>)
- Yazdani, A.; Jeffrey, P., (2010). *A complex network approach to robustness and vulnerability of spatially organized water distribution networks*, e-print: <http://arxiv.org/abs/1008.1770v2>.
- Zio, E.; Sansavini, G., (2007). *Service reliability analysis of a tramway network*. Proceedings of the European Safety and Reliability Conference 2007, ESREL 2007, Risk, reliability and Societal Safety, No. 1, pp. 907-913.