Entropy of centrality values for topological vulnerability analysis of water distribution networks

Seyed Ashkan Zarghami, Indra Gunawan and Frank Schultmann
Entrepreneurship, Commercialisation and Innovation Centre (ECIC), The University of Adelaide, Adelaide, Australia

Abstract

Purpose – The increased complexity of water distribution networks (WDNs) emphasizes the importance of studying the relationship between topology and vulnerability of these networks. However, the few existing studies on this subject measure the vulnerability at a specific location and ignore to quantify the vulnerability as a whole. The purpose of this paper is to fill this gap by extending the topological vulnerability analysis further to the global level.

Design/methodology/approach – This paper introduces a two-step procedure. In the first step, this work evaluates the degree of influence of a node by employing graph theory quantities. In the second step, information entropy is used as a tool to quantify the global vulnerability of WDNs.

Findings – The vulnerability analysis results showed that a network with uniformly distributed centrality values exhibits a lower drop in performance in the case of partial failure of its components and therefore is less vulnerable. In other words, the failure of a highly central node leads to a significant loss of performance in the network.

Practical implications – The vulnerability analysis method, developed in this work, provides a decision support tool to implement a cost-effective maintenance strategy, which relies on identifying and prioritizing the vulnerabilities, thereby reducing expenditures on maintenance activities.

Originality/value – By situating the research in the entropy theory context, for the first time, this paper demonstrates how heterogeneity and homogeneity of centrality values measured by the information entropy can be interpreted in terms of the network vulnerability.

Keywords Information entropy, Betweenness centrality, Closeness centrality, Eigenvector centrality, Vulnerability analysis, Water distribution networks

Paper type Research paper

1. Introduction

Water distribution networks (WDNs) have been identified by many countries as national strategic assets that should be highly valued and efficiently managed. However, with the ageing of these networks, numerous problems are emerging such as pipe bursts, leakage, water quality degradation, water supply interruption and loss of hydraulic performance. Consequently, maintenance expenditures are becoming the major driver of capital expenditure in the water sector. This coupled with budget constraints for the maintenance activities reinforces the need to accurately identifying vulnerabilities and prioritizing water pipeline assets (Zamenian et al., 2017).

Vulnerability analysis of WDNs has been an active area of past research. Shuang et al. (2014a) evaluated the nodal vulnerability of WDNs under cascading failure by monitoring pressure in nodes and flows in pipes during the cascading process. Fragiadakis and Christodoulou (2014) and Fragiadakis et al. (2016) performed a seismic hydraulic vulnerability assessment of urban water networks using survival analysis. Shuang et al. (2015) suggested different recovery strategies of WDNs, focusing on the vulnerability of nodes due to exceeding their hydraulic (pressure) capacity. Laucelli and Giustolisi (2015) evaluated the vulnerability of WDNs under seismic actions using a hydraulic modeling paradigm taking into account unsupplied demand to customers.

Such studies approach the vulnerability analysis of WDNs from a hydraulic perspective, which is concerned with satisfying flow and pressure requirements. These studies take into
consideration factors such as failures due to demand variation, undersized pipes, storage capacity, insufficient pressure or a combination of these factors (Zhuang et al., 2013). However, due to the complex interactions among a large number of subsystems and components, the exclusive hydraulic analysis of WDNs just partially describes the network performance (Gunawan et al., 2017). On this premise, the topological vulnerability analysis, as a complementary approach, provides a robust model, thereby assessing the vulnerability of WDNs more accurately (Yazdani et al., 2011).

There is a small body of literature, which analyzes the vulnerability of WDNs from the topological point of view. The topological vulnerability analysis refers to analyzing the configuration of the network based on graph theory techniques (Di Nardo and Di Natale, 2011). Perelman and Ostfeld (2011) constructed a topological connectivity matrix aimed at clustering the nodes in WDNs based on their connectivity, thereafter identifying weakly and strongly connected clusters. Yazdani and Jeffrey (2011, 2012) examined the vulnerability of WDNs to the failure of individual components by identifying the critical components using metrics from graph theory. Sheng et al. (2013) adopted a complex network-based model for exploring the malfunction of WDNs by measuring the spectral properties and subsequently identifying the isolated communities.

The approaches discussed above assess the vulnerability of WDNs by adopting very generic topological properties of the network within which mainly the vulnerability problem at the local level is addressed. While identifying the critical components through performing a vulnerability analysis at a specific location is of great importance, studies on how to quantify the vulnerability of a WDN as a whole remain scant. As the literature review revealed, little effort has gone to developing quantitative metrics through which the global vulnerability can be measured. A global vulnerability index can provide insight into the current vulnerability of the network. In fact, this index can be used as a benchmark to determine the vulnerabilities in WDNs. It also allows for measuring the extent of improvement in the reliability of the network after the implementation of maintenance strategies.

The theme of this paper is the development of a novel methodology for local and global vulnerability analysis of WDNs. The proposed methodology includes two stages. In the first stage, the graph theory quantities known as closeness, eigenvector and betweenness centrality are employed in order to prioritize the nodal components in WDNs. This stage evaluates the vulnerability of a WDN at the local level based on the importance of its components. In the second stage, the potential of using Shannon (information) entropy as a means to measure the homogeneity and heterogeneity of the centrality values is explored. Accordingly, this work develops a new vulnerability index and demonstrates how heterogeneity and homogeneity of the centrality values measured by the information entropy can be interpreted in terms of the network vulnerability.

The rest of the paper is structured according to the following plan. Section 2 recalls the basic definitions of the three most commonly used centrality measures. Section 3 develops the proposed method and details the procedural steps to evaluate the vulnerability of WDNs. In Section 4, the paper presents two case studies, a real-world WDN of an Australian town and a network from the literature, as illustrations of the proposed method. Conclusions are drawn in Section 5, followed by a discussion of the avenues for future research.

2. Centrality analysis
The centrality of elements in a network is concerned with the identification of the elements with a more central role than others (Qi et al., 2012). In recent years, a number of centrality measures have been devised to evaluate the importance of nodes and links in a network, within which different dimensions of the intuitive notion of the centrality are addressed (Brandes et al., 1999). Centrality measures have mainly focused on solving the problem of revealing the importance of elements by measuring the various network topologies.
Broadly classified by Boldi and Vigna (2013), these measures can be categorized into the three following types:

1. Geometric measures where the importance of an element is a function of distances. The examples of centrality measures in this category are indegree, closeness, Lin’s index and harmonic centrality.

2. Spectral measures such as Seeley’s index, Katz’s index, PageRank and eigenvector centrality, which are based on the left dominant eigenvector of a network adjacency matrix or some other matrix derived from it.

3. Path-based measures that assess the importance of a component based on the number of paths or shortest path passing through the component. Stress, betweenness and k-path centrality are the most well-known examples of this category.

A review of the existing literature indicates the presence of confusion in what centrality measure scores best compared to others. In order to capture various topological characteristics of a network, this research identifies the most central nodes in a network by adopting an exemplary centrality measure from each of the abovementioned categories. These centrality measures are closeness centrality as a geometric measure, eigenvector centrality as a spectral measure and betweenness centrality as a path-based measure. These three centrality measures are the most widely used centrality measures and set the basis for the development of other mathematically related measures (Lozares et al., 2015).

Closeness centrality was first introduced by Bavelas (1950) and can be explained as a mean distance from a given node to other nodes in a network. Closeness centrality indicates how close a given node is with respect to the entire network and can be formulated as follows:

$$C_C(i) = \frac{n-1}{\sum_{j \neq i} d_{ij}}$$  \hspace{1cm} (1)

where $C_C(i)$ represents the closeness centrality, $n$ is the number of nodes in the network and $d_{ij}$ denotes the shortest path lengths between node $i$ and $j$.

Betweenness centrality is based on the idea that a given node is central if it lies between many other nodes (Cadini et al., 2009). Betweenness centrality of node $i$, $C_B(i)$, is defined as the number of shortest paths between pairs of nodes that pass through a given node and can be stated by the following formula:

$$C_B(i) = \frac{1}{(n-1)(n-2)} \sum_{s \neq r \neq i} n_{sr}(i)$$  \hspace{1cm} (2)

where $n$ is the number of nodes in the network, $n_{sr}(i)$ denotes the number of shortest paths between $s$ and $r$ passing through $i$ and $n_{sr}$ represents the number of shortest paths between $s$ and $r$. $C_B(i)$ takes on values between 0 and 1 and attains its maximum value when node $i$ falls on all shortest paths between two nodes.

Eigenvector centrality of a node takes into account the combined centrality values of its neighbors based on the philosophy that a given node is more central if its neighbors are also highly central (Joyce et al., 2010). The mathematical expression of the eigenvector centrality of node $i$, $C_e(i)$, can be expressed as per the following equation:

$$C_e(i) = \frac{1}{\lambda} \sum_{j \neq i} C_e(j)$$  \hspace{1cm} (3)

$$\lambda e = Ae,$$
where $A$ is the adjacency matrix, $e$ denotes the eigenvector centrality of all nodes and $\lambda$ is the largest eigenvector of the adjacency matrix.

Centrality measures evaluate the impact of each node on the network performance and provide a numerical indicator to identify the network’s most influential components. However, stand-alone use of these metrics yields insufficient information as to the weaknesses of a network. An attempt is made in the following section to provide a solution to this problem by proposing a vulnerability index based on the joint entropy of the distribution of centrality values.

3. An entropy-based vulnerability index

In this section, the expected level of the network vulnerability is evaluated by computing the Shannon entropy of centrality values. Shannon entropy, introduced by Shannon (1948), is a widely used evaluated measure of choice, uncertainty and heterogeneity of a set of probabilities, which can be expressed by the following equation:

$$H = -\sum_{i=1}^{n} p_i \log_b p_i,$$

where $H$ is the entropy of distribution, $p_i$ is the probability associated with the $i$th outcome, $n$ denotes the number of possible outcomes and $b$ is an arbitrary logarithm base indicating the unit of entropy. For example, for $b = 2$, $b = e$ and $b = 10$, the unit of entropy is, respectively, defined as bit, Napier and decibels (Zarghami et al., 2018). By definition, $0 \leq H \leq \log_b n$. The lower extreme value, $H = 0$, occurs when one of the probabilities is 1 and the rest take on the value of 0, whereas the upper extreme value, $H = \log_b n$, occurs when all the probabilities are of equal value of $1/n$.

Shannon introduced Equation (4) for complete probability distributions, where $\sum_{i=1}^{n} p_i = 1$. In order to follow the definition proposed by Shannon, the normalized form of each centrality value is used by scaling it to the [0, 1] interval.

Let $\text{CC} = \{C_{\text{CC}}(i) : i = 1, 2, \ldots, m\}$, $\text{CE} = \{C_{\text{CE}}(i) : i = 1, 2, \ldots, m\}$ and $\text{CB} = \{C_{\text{CB}}(i) : i = 1, 2, \ldots, m\}$ be, respectively, a set of closeness, eigenvector and betweenness centrality values for a network with $m$ edges. Let assume $C_M(i)$ as a symbolic representation of a given centrality value of node $i$ such that $C_M(i) \in \{\text{CC or CE or CB}\}$. The normalized centrality is then defined as the ratio of a given centrality value to the sum of all values of the given centrality measure, as such $\sum_{i=1}^{n} C_M(i) = 1$. The normalized centrality of node $i$, $\text{PC}_M(i)$, can be stated as follows:

$$\text{PC}_M(i) = \frac{C_M(i)}{\sum_{i=1}^{n} C_M(i)}.$$  

(5)

As can be seen from Equation (5), in order to obtain $\text{PC}_M(i)$, the value of centrality from a given set of centrality is scaled relative to the sum of all centrality values of the set. Therefore, $\text{PC}_M$ provides a numerical indicator to evaluate the relative influence of a node in a network.

By substituting $\text{PC}_M(i)$, Equation (4) can be restated as follows:

$$H_{CM} = -\sum_{i=1}^{m} \text{PC}_M(i) \log_2 \text{PC}_M(i),$$

(6)

where the notation of $H_{CM}$ is a symbolic representation of the entropy of a given set of centrality values, meaning that $H_{CM} = H_{\text{CC}}$ if $C_M(i) \in \text{CC}$, $H_{CM} = H_{\text{CE}}$ if $C_M(i) \in \text{CE}$ and $H_{CM} = H_{\text{CB}}$ if $C_M(i) \in \text{CB}$.
As stated earlier, to achieve a more accurate result for evaluating the global vulnerability of a network, it is now time to combine the entropy of individual centrality measures by means of joint entropy of the three centrality measures.

The joint entropy of three variables, \(H(X_1, X_2, X_3)\), with a set of joint probability of \(\{p_{ijk}: i = 1, 2, \ldots, N; j = 1, 2, \ldots, M; k = 1, 2, \ldots, P\}\), is defined as:

\[
H(X_1, X_2, X_3) = - \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{k=1}^{P} p_{ijk} \log(p_{ijk}) \leq H(X_1) + H(X_2) + H(X_3). \tag{7}
\]

Since \(CC(i), CE(i)\) and \(CB(i)\) are statistically independent variables, their joint entropy can be obtained from the following mathematical expression:

\[
H_{CT} = H_{CC} + H_{CE} + H_{CB}. \tag{8}
\]

Intuitively, failure of a junction node with a high centrality value results in the disruption of the service for many nodes in the network due to its central location. Therefore, in the case where all junction nodes are of equal value of centrality, the debilitating effect on the network performance due to the failure of each individual node will be minimum. This intuitive description is very reminiscent of the principle of Shannon entropy, which is a decreasing function of scattering of random variables, and attains its maximum value when all the outcomes are equally likely (Maszczyk and Duch, 2008).

In a network with \(m\) junction nodes, when all centrality values are equally likely, \(H_{CT}\) is maximum when \(PC_C = PC_E = PC_B = 1/m\), thus:

\[
H_{CC, \text{max}} = H_{CE, \text{max}} = H_{CB, \text{max}} = -\log_2 m. \tag{9}
\]

By substituting these values into Equation (8):

\[
H_{CT, \text{max}} = H_{CC, \text{max}} + H_{CE, \text{max}} + H_{CB, \text{max}} = -3\log_2 m. \tag{10}
\]

The vulnerability index, VI, can be constructed based on the fractional differences between \(H_{CT}\) and maximum achievable \(H_{CT}\). Thus, VI is defined as one minus the relative entropy as follows:

\[
VI = 1 - \frac{H_{CT}}{H_{CT, \text{max}}}. \tag{11}
\]

The vulnerability index of the network falls within the range of \([0, 1]\), where a higher value of VI indicates the higher vulnerability, whereas a lower value implies the lower vulnerability. VI represents the comparative heterogeneity of the centrality values defined by \(H_{CT}\) with respect to the maximum possible entropy value where all values are uniformly distributed (Singh, 2013). VI attains its minimum value (VI = 0), when \(\{PC_M(i)|i = 1, 2, \ldots, m\}\) is uniformly distributed. Theoretically, this case corresponds to the situation where all components in the network are equally central.

VI describes how severe the consequences of random failures may be. It refers to the likely magnitude of failures. That is, in a homogeneous case where the nodes in the system are almost equally central, the severity of the random failure of a node is lesser than that of the case when some nodes are highly central and others are peripherals. In other words, when a very few central nodes dominate the network, the failure of each of these nodes leaves a large number of nodes disconnected, which implies the severity of the failure and consequently a high vulnerability of the network.
4. Application
In order to illustrate the proposed vulnerability assessment method and to evaluate the effectiveness of the new vulnerability index, this section presents two contrasting case studies. These two case studies represent two different extreme layouts of WDNs: a strongly looped layout and a branched configuration of pipeline assets. These two contrasting case studies are used to investigate the possibility to generalize the application of the proposed method to different WDNs with different topological characteristics. An open-source graph analysis software, igraph, is used to compute the closeness, eigenvector and betweenness centrality values for each network. The vulnerability index of each network is then calculated by using Equation (11). After computing the vulnerability index, this section compares the vulnerability of two case studies.

4.1 Case Study 1
The first case study, as shown in Figure 1, is a looped WDN taken from the literature (Islam et al., 2014; Shuang et al., 2014b). As a means to illustrate the proposed vulnerability index, the case study is mapped into an undirected graph with a node set of size 27 and an edge set of size 40. Water is supplied from two reservoirs connected to nodes 1 and 3.

The values of betweenness, eigenvector and closeness centrality for all nodes using igraph software are now obtained. These values are presented in Table I. The gray columns report the ranking of each node based on its corresponding centrality score.

In order to provide better visualization of the results in Table I, the resulting rankings of the nodes are plotted in Figure 2.

As noted earlier, \( C_C \), \( C_E \) and \( C_B \) measure closeness, eigenvector and betweenness centrality importance score, respectively. As reported by Table I and Figure 2, betweenness centrality places nodes 16 and 11 at the first and the second positions, respectively. This is because these two nodes are centrally located in the network, thus when compared to other nodes, they participate in a higher number of shortest paths between any given pair of nodes.
As expected, nodes 19 and 25 take on the lowest betweenness centrality values, indicating that these nodes play a role in very few shortest paths.

Using eigenvector centrality, node 11 is scored 1.00 and all other nodes have lower scores ranging downwards toward 0.1534. The highest value of $C_E$ is for node 11 because the sum of the eigenvector centrality values of its immediate neighbors (nodes 6, 10, 12 and 16) is relatively high.

When closeness centrality is applied, the highest ranks are, respectively, attained for nodes 11 and 16 because the mean distance from each of these two nodes to all other nodes in the example network is relatively low. On the other side, closeness centrality places nodes 25 at the lowest position since the average farness of this node to all other nodes is high.

It can be observed that closeness and eigenvector centrality are equal on their impact on the ranking of the nodal elements. A slight difference between the rankings of nodal elements can be observed when closeness and eigenvector centrality are applied. For example, nodes 11, 12, 13, 18, 19, 24 and 25 have exactly the same rankings when these two centrality measures are used. Furthermore, the differences in the ranking of other nodes using closeness and eigenvector centrality are not conspicuous. This can be interpreted as the evidence that betweenness centrality differs from other centrality measures. A node can have no important neighbors or can be a long way from all other nodes on average, and still has a high betweenness score.

Up to this point, the importance of each node using various centrality measures has been assessed. It is now possible to calculate the vulnerability index described in Section 3.

Using Equation (11), the normalized centrality values are presented in Table II.

Using Equation (6), the entropy of the normalized centrality values obtained from Table II is:

$$H_{CC} = -4.6245, \quad H_{CE} = -4.4888, \quad H_{CB} = -4.2192.$$
By substituting the results into Equation (8):
\[ H_{CT} = -13.3325. \]

Given \( m = 25 \), using Equation (10):
\[ H_{CT,\max} = -3\log_2{25} = -13.9317. \]

By using Equation (11) and substituting the values of \( H_{CT} \), and \( H_{CT,\max} \):
\[ VI = 0.0430. \]

As can be seen, the resulting vulnerability index for this case is rather low, which implies a relatively low risk to disruption of the service in the network. This can be ascribed to a highly homogenous distribution of the normalized centrality values. This result conforms to the intuition of the vulnerability concept because in this case study water can flow from source nodes to junction nodes through many pathways; as a result, the nodes have a comparable influence on the overall performance of the network.

4.2 Case Study 2
The second case study is a real-world WDN of Price, a small town in South Australia, located 140 km west of Adelaide, Australia. The network is a directed tree-shaped WDN, represented
by 18 nodes connecting 17 pipes (Figure 3). The layout for this case study has been obtained from the official website of South Australia Water company (http://sawater.maps.arcgis.com).

Applying the same procedure as the previous case study, the values of betweenness, eigenvector and closeness centrality of nodes for the second case study are presented in Table III, where the gray columns report the ranking of each node based on its corresponding centrality score.

Examination of Table III and Figure 4 reveals a high correlation between eigenvector and betweenness centrality for the tree-shaped case study. These two centrality measures identify nodes 2, 5, 6, 13 and 16 as the top 5 critical nodes. When betweenness centrality is applied, nodes 5 and 13 are, respectively, placed at the first and the second positions. In fact, these nodes are centrally located in the network, as such, a high number of shortest paths passes through them. Similarly, eigenvector centrality posits these two nodes at the exact same positions. This can be attributed to the importance of the immediate neighbors of these nodes.

Using betweenness centrality, nodes 4, 8, 9, 11, 12, 14, 15, 17 and 18 take on the value of 0. In a similar vein, eigenvector centrality considers the lowest criticality for these nodes. The fact that these nodes play no role in any shortest paths as well as the relatively low value of the sum of the eigenvector centrality of their immediate neighbors make the results unsurprising.

Using closeness centrality, node 1 turns out to be the most critical element, whereas nodes 11 and 12 are placed in the same position as the least critical nodes. This is because the closeness centrality of a junction node in a directed graph is defined by the inverse of the average length of shortest paths to/from all other nodes; therefore, unlike an undirected graph, the total number of nodes may not be used in Equation (1). Intuitively, the failure of node 1 leaves all other nodes without water supply, while the failure of nodes 11 or 12 does not affect other nodes in the network.

<table>
<thead>
<tr>
<th>Node</th>
<th>(C_C)</th>
<th>(PC_C)</th>
<th>(C_E)</th>
<th>(PC_E)</th>
<th>(C_B)</th>
<th>(PC_B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2500</td>
<td>0.0359</td>
<td>0.3775</td>
<td>0.0289</td>
<td>0.0953</td>
<td>0.0314</td>
</tr>
<tr>
<td>2</td>
<td>0.2796</td>
<td>0.0386</td>
<td>0.4734</td>
<td>0.0362</td>
<td>0.0833</td>
<td>0.0275</td>
</tr>
<tr>
<td>3</td>
<td>0.2708</td>
<td>0.0374</td>
<td>0.3836</td>
<td>0.0295</td>
<td>0.073</td>
<td>0.0291</td>
</tr>
<tr>
<td>4</td>
<td>0.2549</td>
<td>0.0352</td>
<td>0.3404</td>
<td>0.026</td>
<td>0.0201</td>
<td>0.0067</td>
</tr>
<tr>
<td>5</td>
<td>0.3132</td>
<td>0.0432</td>
<td>0.7040</td>
<td>0.0538</td>
<td>0.1796</td>
<td>0.0592</td>
</tr>
<tr>
<td>6</td>
<td>0.3421</td>
<td>0.0472</td>
<td>0.8519</td>
<td>0.0651</td>
<td>0.1999</td>
<td>0.0659</td>
</tr>
<tr>
<td>7</td>
<td>0.3291</td>
<td>0.0454</td>
<td>0.7291</td>
<td>0.0557</td>
<td>0.1884</td>
<td>0.0621</td>
</tr>
<tr>
<td>8</td>
<td>0.2708</td>
<td>0.0374</td>
<td>0.3622</td>
<td>0.0277</td>
<td>0.0229</td>
<td>0.0076</td>
</tr>
<tr>
<td>9</td>
<td>0.2737</td>
<td>0.0378</td>
<td>0.4574</td>
<td>0.035</td>
<td>0.0560</td>
<td>0.0185</td>
</tr>
<tr>
<td>10</td>
<td>0.3421</td>
<td>0.0472</td>
<td>0.8323</td>
<td>0.0636</td>
<td>0.198</td>
<td>0.0658</td>
</tr>
<tr>
<td>11</td>
<td>0.3768</td>
<td>0.052</td>
<td>1.0000</td>
<td>0.0764</td>
<td>0.216</td>
<td>0.0796</td>
</tr>
<tr>
<td>12</td>
<td>0.3611</td>
<td>0.0498</td>
<td>0.8879</td>
<td>0.0678</td>
<td>0.2176</td>
<td>0.0717</td>
</tr>
<tr>
<td>13</td>
<td>0.2921</td>
<td>0.0403</td>
<td>0.5066</td>
<td>0.0387</td>
<td>0.0667</td>
<td>0.022</td>
</tr>
<tr>
<td>14</td>
<td>0.2653</td>
<td>0.0366</td>
<td>0.3880</td>
<td>0.0297</td>
<td>0.0621</td>
<td>0.0205</td>
</tr>
<tr>
<td>15</td>
<td>0.3291</td>
<td>0.0454</td>
<td>0.6782</td>
<td>0.0518</td>
<td>0.2127</td>
<td>0.0701</td>
</tr>
<tr>
<td>16</td>
<td>0.3611</td>
<td>0.0408</td>
<td>0.8397</td>
<td>0.0641</td>
<td>0.2907</td>
<td>0.0859</td>
</tr>
<tr>
<td>17</td>
<td>0.3467</td>
<td>0.0478</td>
<td>0.7936</td>
<td>0.0606</td>
<td>0.2299</td>
<td>0.0758</td>
</tr>
<tr>
<td>18</td>
<td>0.2826</td>
<td>0.039</td>
<td>0.4785</td>
<td>0.0366</td>
<td>0.0821</td>
<td>0.0271</td>
</tr>
<tr>
<td>19</td>
<td>0.2149</td>
<td>0.0297</td>
<td>0.1882</td>
<td>0.0144</td>
<td>0.0033</td>
<td>0.0011</td>
</tr>
<tr>
<td>20</td>
<td>0.2549</td>
<td>0.0352</td>
<td>0.2539</td>
<td>0.0194</td>
<td>0.0276</td>
<td>0.0091</td>
</tr>
<tr>
<td>21</td>
<td>0.2921</td>
<td>0.0403</td>
<td>0.3831</td>
<td>0.0301</td>
<td>0.0546</td>
<td>0.018</td>
</tr>
<tr>
<td>22</td>
<td>0.2826</td>
<td>0.039</td>
<td>0.5015</td>
<td>0.0383</td>
<td>0.1508</td>
<td>0.0497</td>
</tr>
<tr>
<td>23</td>
<td>0.2385</td>
<td>0.0329</td>
<td>0.3323</td>
<td>0.0254</td>
<td>0.0622</td>
<td>0.0205</td>
</tr>
<tr>
<td>24</td>
<td>0.2261</td>
<td>0.0312</td>
<td>0.1920</td>
<td>0.0147</td>
<td>0.0217</td>
<td>0.0715</td>
</tr>
<tr>
<td>25</td>
<td>0.1970</td>
<td>0.0272</td>
<td>0.1534</td>
<td>0.0118</td>
<td>0.0039</td>
<td>0.0013</td>
</tr>
</tbody>
</table>

Table II. Normalized centrality values for the first case study

by 18 nodes connecting 17 pipes (Figure 3). The layout for this case study has been obtained from the official website of South Australia Water company (http://sawater.maps.arcgis.com).

Applying the same procedure as the previous case study, the values of betweenness, eigenvector and closeness centrality of nodes for the second case study are presented in Table III, where the gray columns report the ranking of each node based on its corresponding centrality score.

The resulting rankings of nodes for this case study are plotted in Figure 4.

Examination of Table III and Figure 4 reveals a high correlation between eigenvector and betweenness centrality for the tree-shaped case study. These two centrality measures identify nodes 2, 5, 6, 13 and 16 as the top 5 critical nodes. When betweenness centrality is applied, nodes 5 and 13 are, respectively, placed at the first and the second positions. In fact, these nodes are centrally located in the network, as such, a high number of shortest paths passes through them. Similarly, eigenvector centrality posits these two nodes at the exact same positions. This can be attributed to the importance of the immediate neighbors of these nodes.

Using betweenness centrality, nodes 4, 8, 9, 11, 12, 14, 15, 17 and 18 take on the value of 0. In a similar vein, eigenvector centrality considers the lowest criticality for these nodes. The fact that these nodes play no role in any shortest paths as well as the relatively low value of the sum of the eigenvector centrality of their immediate neighbors make the results unsurprising.

Using closeness centrality, node 1 turns out to be the most critical element, whereas nodes 11 and 12 are placed in the same position as the least critical nodes. This is because the closeness centrality of a junction node in a directed graph is defined by the inverse of the average length of shortest paths to/from all other nodes; therefore, unlike an undirected graph, the total number of nodes may not be used in Equation (1). Intuitively, the failure of node 1 leaves all other nodes without water supply, while the failure of nodes 11 or 12 does not affect other nodes in the network.
### Table III.

Numerical values of centrality metrics for the second case study.

<table>
<thead>
<tr>
<th>Node</th>
<th>( C_r )</th>
<th>( C_t )</th>
<th>( C_b )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2787</td>
<td>1</td>
<td>0.2275</td>
</tr>
<tr>
<td>2</td>
<td>0.2742</td>
<td>2</td>
<td>0.5834</td>
</tr>
<tr>
<td>3</td>
<td>0.0588</td>
<td>7</td>
<td>0.2683</td>
</tr>
<tr>
<td>4</td>
<td>0.0556</td>
<td>10</td>
<td>0.1046</td>
</tr>
<tr>
<td>5</td>
<td>0.1717</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>0.0664</td>
<td>6</td>
<td>0.5834</td>
</tr>
<tr>
<td>7</td>
<td>0.0588</td>
<td>7</td>
<td>0.2683</td>
</tr>
<tr>
<td>8</td>
<td>0.0556</td>
<td>10</td>
<td>0.2275</td>
</tr>
<tr>
<td>9</td>
<td>0.0556</td>
<td>10</td>
<td>0.1046</td>
</tr>
<tr>
<td>10</td>
<td>0.0588</td>
<td>7</td>
<td>0.46</td>
</tr>
<tr>
<td>11</td>
<td>0.0555</td>
<td>17</td>
<td>0.1794</td>
</tr>
<tr>
<td>12</td>
<td>0.0555</td>
<td>17</td>
<td>0.3656</td>
</tr>
<tr>
<td>13</td>
<td>0.0821</td>
<td>4</td>
<td>0.9374</td>
</tr>
<tr>
<td>14</td>
<td>0.0556</td>
<td>10</td>
<td>0.3656</td>
</tr>
<tr>
<td>15</td>
<td>0.0556</td>
<td>10</td>
<td>0.2622</td>
</tr>
<tr>
<td>16</td>
<td>0.0667</td>
<td>5</td>
<td>0.6724</td>
</tr>
<tr>
<td>17</td>
<td>0.0556</td>
<td>10</td>
<td>0.2622</td>
</tr>
<tr>
<td>18</td>
<td>0.0556</td>
<td>10</td>
<td>0.2622</td>
</tr>
</tbody>
</table>

**Figure 3.**

Case Study 2: price water distribution network.

**Water Transmission Line**

Topological vulnerability analysis of WDNs.
It is now possible to calculate the vulnerability index proposed in this work. Using Equation (5), the normalized centrality values for this case study are presented in Table IV.

Using Equation (6), the entropy of the normalized centrality values obtained from Table IV is:

\[ H_{CC} = -3.8387, \quad H_{CE} = -3.8989, \quad H_{CB} = -2.6301. \]

By substituting the results into Equation (8):

\[ H_{CT} = -10.3677. \]

Given \( m = 18 \), using Equation (10):

\[ H_{CT,\max} = -3\log_2 18 = -12.5098. \]

By using Equation (11) and substituting the values of \( H_{CT} \) and \( H_{CT,\max} \):

\[ \text{VI} = 0.1712. \]

The high value of VI in this case study describes how significant the likely consequences of failures may be. This can be interpreted as the evidence that due to the heterogeneous distribution of the nodal centralities in this case study, failure of a highly central node (e.g. node 5) leads to a significant loss of the performance in the network.
4.3 Comparison of results

The previous discussions reveal the relationships between closeness, eigenvector and betweenness centrality in assessing the importance of nodes within two contrasting case studies. Overall, the results show a positive correlation between all centrality measures. Correlation between closeness and eigenvector centrality is more evident in the strongly looped layout, whereas the tree-shaped case study exhibits a relatively high correlation between eigenvector and betweenness centrality. However, despite the similarity in concept, the strict linear relationships in any of the cases are not observed. For example, closeness centrality does not convey the same information for two case studies. In the looped network, closeness centrality counts the total number of nodes for measuring the mean distance from a node to other nodes, while in the tree-shaped network, it does not use the total number of nodes and distinguishes between upstream and downstream nodes.

What is particularly striking about the contrast between vulnerability indices generated for two case studies is that the first case study presents a rather homogeneous distribution of the nodal centralities, whereas the second case study produces a highly heterogeneous distribution of the centrality values. The resulting vulnerability index in the first case study is therefore far lesser than that of the second case study. This proves that the proposed vulnerability index captures the distinctions between the tree-shaped networks, where water can take only one pathway from the source to the households, and the looped WDNs, where water flows from the source node to the households through many pathways. In fact, VI measures the risk to the satisfactory level of water supply service. The larger the vulnerability index, the larger magnitude of the failure, as such the consequences of disruptive events on the network performance in the second case study have been precisely captured by a higher vulnerability index when compared to the first case study.

5. Conclusion

The methods presently used for vulnerability analysis of WDNs have lagged far behind capturing various topological attributes of these networks. The existing literature on this problem captures very generic topological features of WDNs and analyzes the vulnerability in a local sense. The present paper, first, evaluates the degree of influence of a node by employing graph theory quantities known as closeness, eigenvector and betweenness centrality.
This work then extends the vulnerability analysis further to the global level by generating a new vulnerability index. The new index is developed from the information entropy based on the distribution of the normalized centrality values.

Using two case studies, a tree-shaped WDN and a looped network from the literature, this paper has demonstrated the effectiveness of the proposed method. As the previous discussion attests, the proposed vulnerability analysis method is consistent with the intuitive notion of vulnerability and is capable of capturing the distinctions between various layouts of WDNs.

This work provides two types of practical implications. First, the maintenance strategy based on the vulnerability assessment model proposed in this work enables water service providers to rank and prioritize the deteriorating pipes. This, in turn, allows for least cost decisions to make for renewal and rehabilitation of pipeline assets. Second, the global vulnerability analysis, proposed in this work, provides water utilities with the opportunity to determine the consequences of unexpected events on the overall performance of the network. This can be used to establish a risk management plan that deals with prevention, decision-making, action taking, crisis management and recovery.

This paper contributes to vulnerability analysis of WDNs by generating new knowledge in the area of vulnerability analysis through coupling the centrality analysis and the entropy theory. However, the conventional centrality measures rely only on the topological information. Therefore, these measures only partially describe a network structure and cannot entirely characterize its properties. Further research might seek to develop a domain specific centrality metrics by taking into account the topological along with the hydraulic attributes of the nodes in the network.

References


Further reading


Corresponding author

Seyed Ashkan Zarghami can be contacted at: ashkan.zarghami@adelaide.edu.au

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com