A TOPOLOGICAL ANALYSIS OF UNDERGROUND NETWORK PERFORMANCE UNDER DISRUPTIVE EVENTS

Lorenzo Mussone
Politecnico di Milano, DABC
Hiva Viseh
Politecnico di Milano, CERM
Roberto Notari
Politecnico di Milano, DMAT

Abstract. Every disruptive event in a transportation network has costly consequences both for users and for managing companies. To prevent large declines in the desired level of mobility, it is important to locate the most important nodes of the network, and to strengthen them. In this paper, we focus on underground networks. We introduce a methodology based on graph theory to locate the most important nodes, and we apply it to Washington D.C. network as a case study, and to the underground networks of 34 cities all over the world.

1. INTRODUCTION

Transportation systems are nowadays one of the main pillars on which our society is built. From time to time, a disruption affects performances of the various transportation systems. A disruption is usually classified as random failure, intentional attack, or natural disaster, in dependence of its origin. According to the considered transportation mode, the resulting capacity reduction can be addressed by more or less efficient alternatives. E.g., when one or two lanes on highways are closed, the remaining ones can still maintain a certain level of service; in an aviation network, a disruptive event in an airport or air route may lead to great perturbations or indeed to a complete interruption of service.

Generally, to describe the capability of a transportation network to face a disruptive event, researchers consider three features, namely *resiliency*, *reliability* and *robustness*. Resiliency is defined as the ability of a transportation network to absorb disruptive events easily and to return back to the prior level of service, or to a higher one, within an acceptable time frame; higher resiliency of a system means less economical, social and operational costs in case of any disruptive event. Reliability means that the expected additional trip cost due to a disruption are acceptable even if users are extremely pessimistic about the state of the network (Bell, 2000). Robustness is the property of a transportation network to maintain its functionality unchanged or nearly unchanged, when exposed to disturbances in various accident scenarios (Scott, 2006).

The paper focuses on underground networks, an infrastructural backbone for urban transportation, and proposes some ways to evaluate their performances, from topological, accessibility and economical point of view. The main purpose is to provide a methodology to locate the most important stations, whose removal is expected to cause the highest reduction in the performance of the underground network.

Firstly, we examine the main causes of disruptive events occurred in underground networks all over the world and how they modified the level of service of the infrastructure. Then, we construct graphs of 34 underground networks, amongst those of most known cities in the world. Graphs are commonly used to represent and analyse networks. They easily capture the topological structure of a network and some features of it, like distance or number of stations between nodes. We study them through calculating the values of four well-known centrality indices, namely betweenness, closeness, degree and eigenvector centrality, and through the values of Ishortest, a new index we have designed. The above mentioned indices enable us to rank nodes of a graph, by identifying those stations for which exposure to a disruptive event may lead to the most harmful consequences for the network. The new index Ishortest measures the impact of a complete closure of a station. In this respect, it is a relatively novel index in the study of transportation networks.

To represent the results, we plot them on the graph and pinpoint each node by a circle whose radius is proportional to node's index value.

The paper is organized as follows. In section 2, we describe the main disruptive, potential and occurred, events in underground networks. Section 3 explains how we build a graph from an underground network. In section 4, we discuss four centrality indices and present the novel index Ishortest. Section 5 reports a detailed case study. We use Washington D.C. underground network as a case study. Furthermore, we report the most critical stations of all 34 networks, and we infer some features on them. Section 6 concludes the paper resuming principal achieved outcomes and outlines future developments.

2. MAIN DISRUPTIVE EVENTS IN UNDERGROUND NETWORKS

Underground networks in metropolitan cities are the busiest modes of transport. Disruption in any components of these networks can be highly expensive from a financial or a temporal point of view, for both passengers and operating companies. Due to significant number of passengers who travel through underground on a daily basis and the forecasted increase, boosting the resiliency and reliability of these networks is very crucial. A disruption in an underground system can be classified as:

- "Natural Disaster",
- "Intentional or Terrorist Attack", or
- "Random Failure or Incident".

2.1 Disruption due to "Natural Disasters"

Since underground networks are usually underground structures, limited kinds of natural disasters have impact on these types of systems. The most concerns regarding the natural hazards which threat underground networks are related to flood and earthquake.

Flood impact on underground network

In case of heavy rain fall, increasing the sea level, or by tsunamis the water can flow into the stations and tunnels throughout the station accesses and tunnels. Among the many direct consequences of a flooded underground, we highlight:

1. Cancellation of some trips because either the origin station or the destination station is flooded.
2. Cancellation of some trips because flooding makes links impassable.
3. Significant delay in several trips. This may happen either because travelers are forced to take indirect routes from origin to destination to avoid flooded links, or as a result of increased strain on passable links.

In addition to direct losses, there are indirect socio-economic losses including:

1. Economic loss due to damaged ventilation systems, cleaning up the debris after flood also suspension of the service during the recovery time.
2.
3. Increasing in road traffic and, as a consequence, increase of emissions into the environment.

Earthquake impact on underground network

Earthquakes with different magnitude influence the underground systems in different ways. Usually a smaller earthquake may only force the underground train slow down or stop and then rail operations center put on restricted speed to allow the train operator to inspect the tunnels and the signaling and electronic systems for any signs of trouble or damage. But a larger earthquake can cause structural damage to underground stations, tunnel collapse and failure of traffic signals and as a result, temporary suspension of underground services, lasting from hours to weeks considering the resiliency of the system.

Disruption due to “Intentional and Terrorist Attacks”

Intentional attacks are defined as targeted destructions caused by outside artificial forces. Reports on known cases show that underground systems are one of the preferred targets of terrorists. It is primarily due to the potential for disruption, destruction, openness, accessibility, lack of passenger identification, and weakness of security caused by the significant number of passengers using the network on a daily basis (Jenkins and Gersten, 2001). According to the statistics of the attack cases, placing explosives, suicide bombings and release poisonous gases are the three main types of underground attack tactics used by terrorist (Yu et al., 2019). The result of such attacks on a hub station would be devastating since they can become very crowded especially during peak hours. Costs which are associated to an intentional attack in a underground network include those borne by the victims, psychological trauma to the victims, their families and friends and infrastructure and property destruction.

Disruption due to "Random Failure"
It is almost impossible to specify the corresponding destructive power for a random failure, and therefore, we describe a random failure as dysfunction of a network due to failure on one or several nodes or edges with random probability of occurrence. Due to linkages between the components of a underground network, the failure of a component may affect the normal functionality of other components or even the whole system.

Technical malfunctions such as power failure, gear failure, brake failure, operational mistakes by staff or drivers, temporary suspension of service for special activity or maintenance or safety inspection, are examples of random failure.

Fire due to intentional attack or random failure in the underground network
In comparison with other means of transportation, underground systems usually have good safety records. However, because of the large number of persons potentially involved in the case of a fire incident, the possibility of damage is high. Kings Cross (London, 1987, 31 fatalities), Baku metro (Azerbaijan, 1995, 286 fatalities), Kaprun funicular tunnel (Austria, 1996, 155 fatalities), and Daegu metro (South Korea, 2003, 192 fatalities) are examples which demonstrate the high damage possibility in case of severe metro accidents, particularly for large fires involving several trains (Bettelini, 2019).

Power line failure, mechanical equipment failure and arson are among the top three causes of fire in the underground system (Yu et al., 2019).

The required recovery time in case of any of the above-mentioned incidents depends on the resiliency of the underground network as well as the extent of incident diffusion, of exposed population and infrastructures and their vulnerability.

3. REPRESENTATION OF UNDERGROUND NETWORKS AS GRAPHS

An underground system is a physical network, which consists of stations (nodes or vertices) and rail tracks (links). Stations and rail tracks provide measurable network properties. In this study, we analyzed thirty-four underground networks of most known cities in the world. For each network, we construct its correspondent undirected graph G (V, E). For our aims, it has to represent the topological structure of the network. In order to represent an underground network as a graph, we apply the following principle: a node of G is either a transfer station or a terminal. So, passing stations (which have degree two) do not appear in the graph. There are exceptions to the previous rule: if two different links connect the same couple of nodes, we add an additional node (which has degree two) in order to distinguish the different links. Furthermore, if the node is the terminus of two lines, we insert a fictitious node so to avoid the node to have degree two. When considering weights, we assign the length of fictitious links equal to 1 kilometer (Figure 1).

Afterwards, for each graph, we prepare three matrices: the adjacency matrix and two other matrices weighting the edges of the graph. In the first weighted matrix, we measure distances between each couple of vertices; in the second one, we count the number of stations between each couple of vertices, endpoints of the edge included. Hence, the weight of an edge is at least 2.

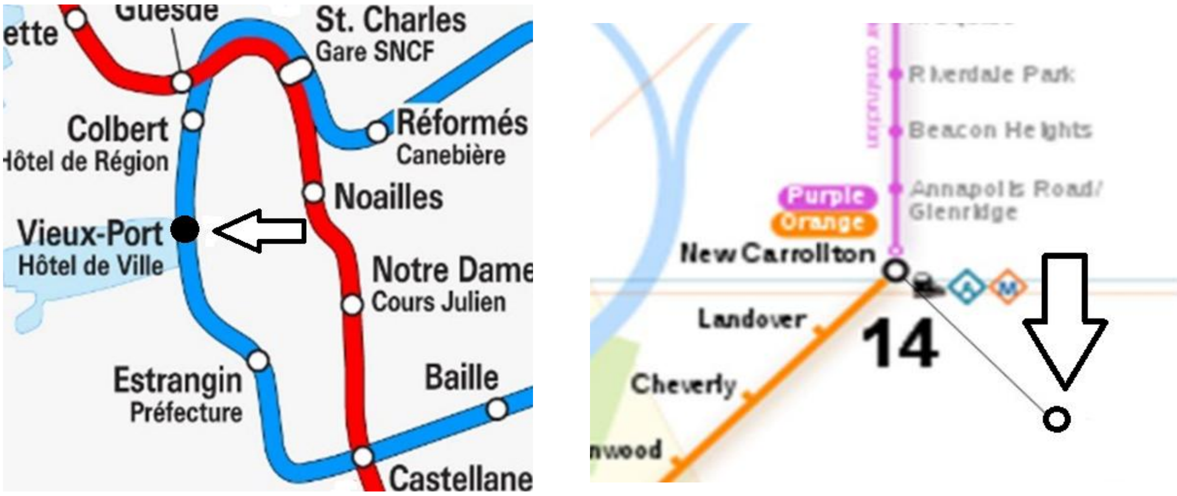

Figure 1. Building the graph. (A) Vieux-Port station, Paris (marked in black) is considered to distinguish the blue from the red path. (B) The fictitious vertex 14 is added to represent the terminus role of New Carrolton station, Washington D.C.

Since weights can be considered as an attribute with both a positive or negative interpretation, we use the inverse of the distance or the inverse of the number of stations when we want to give a negative (in the sense of costs) meaning to weights. In fact, in general, both travelling time and cost increase almost proportionally to distance. For example, by increasing the length of an edge, it becomes more costly and then less attractive to use it. Then, if we interpret distance in this way, it is more appropriate to weight edges by the inverse of distance. On the contrary, just for the sake of example, if we consider the value of an infrastructure, it is again directly related to distance, but distance now assumes a positive interpretation.

4. CENTRALITY MEASURES AND ISHORTEST

It is a standard practice in analyzing graphs to evaluate indices and to infer conclusions on graphs from their values. In the present paper, at first, we focus ourselves on centrality measures, that is to say, on indices designed to locate the most central nodes in the graph. Here, central means most important with respect to the topology of the graph itself if we perform the analysis on a non-weighted graph. When we consider weighted graphs, the importance of a node is a byproduct of the interplay between topology and weights. Later, we introduce a newly designed index, Ishortest, and we compare its values with the centrality measures. Ishortest is designed to look for nodes whose removal from the graph has the largest outcome in terms of lengths of shortest paths.

We assume the readers to be familiar with the basic definitions and results on graphs. In the present paper, by graph, we mean a connected, undirected, simple graph. Hence, edges work in both directions to connect their endpoints, there is no edge from a node to itself, that is to say, no loop, and there is at most one edge between two vertices, if any. Finally, we assume that there is a path between every couple of nodes.

4.1 Degree

The degree of a node is the number of its neighbors, or equivalently, the number of edges that contain the node. We denote the degree of v as $deg(v)$. The degree of a node depends on the topology of the graph only, and in literature, weights do not affect its definition. However, since we assign weights to the edges, we compute weights for nodes from the ones of the edges, as follows: given a node, its weight is the sum of the weights of the edges that contain the considered node. Going back to the topology of graphs we construct from underground networks, because of the methodology we use, we have very few degree 2 nodes.

The degree of a node allows an observer to draw immediately a picture of a graph in a neighbor of a node. From this respect, this index belongs to the set of indices suitable to describe visually a graph.

According to the degree, there are two different kinds of graphs. The first kind of graph consists of the ones in which a few nodes have very large degree. They present a hub structure, where a hub is a node with large degree. When a new node is inserted in such a graph, one usually connects the new node to an existing hub, so that the degree of the hub increases. The second kind of graph consists of the ones in which almost all nodes have the same degree. This is the most common structure of graphs associated to metro networks of large cities.

In both cases, central nodes with respect to the degree are the ones that have largest degree. For example, if Δ is the maximum degree of a node in a graph, we say that the central nodes are the ones whose degree is at least $\Delta - \delta$, where δ depends on the graph. For example, if the graph has not a hub structure, $\delta = 0, 1$ or 2 is a good choice. On the contrary, if the graph has a hub structure, all hubs must be central nodes, and so the value of δ has to be chosen accordingly.

Since the values of the degree are only a few integers for a given graph, in general, it is not so useful to normalize this index. For example, in the graphs we construct from underground networks, it is quite rare to have a node of degree 10. However, we propose two possible normalizations for degree, according to different scenarios in which we will use them.

The first normalization is

$$\mathrm{ndeg_1(v)} = \frac{\mathrm{deg(v)} - \mathrm{d_m}}{\mathrm{d_M} - \mathrm{d_m}}, \qquad (1)$$

where d_m, d_M are the minimum and the maximum degree of a node in the given graph.

The second normalization is

$$\mathrm{ndeg_2(v)} = \frac{\mathrm{deg\,(v)}}{\mathrm{D_M}}, \qquad (2)$$

where D_M is the maximum degree of a node in a graph which belongs to a given reference set. For example, if we consider the set of graphs with n nodes, then D_M is equal to $n-1$, and is the degree of a node that is connected by an edge to every other node in the graph.

The first normalization has to be preferred when comparing different indices evaluated on the nodes of the same graph, whereas the second one has to be chosen when comparing the same index evaluated at different graphs sharing a common feature.

4.2 Betweenness

A different centrality measure is betweenness, proposed by L. Freemann (Freemann, 1977). The basic idea of betweenness is that a node is central if it is on many shortest paths, where shortest means a path with least weight. We remark that, in an unweighted graph, we can say that every edge has weight 1. Hence, when moving on a graph, it is very likely that one has to go through some nodes that are central with this respect. The definition of betweenness $g(v)$ of a node v is

$$\mathrm{g(v)} = \sum_{\mathrm{v \neq x \neq y}} \frac{\sigma_{\mathrm{xy}}(\mathrm{v})}{\sigma_{\mathrm{xy}}}, \qquad (3)$$

where x, y are different nodes, σ_{xy} is the number of shortest paths between x, y and $\sigma_{xy}(v)$ is the number of shortest paths between x, y containing the node v.

To illustrate the differences between degree and betweenness, we can consider the graph representation of a single metro line, where each station is a node. Hence, the associated graph is a path. Apart the terminal nodes that have degree 1, all the other nodes have degree 2. On the contrary, the betweenness of a node is the product of the distances from the terminals, and so the betweenness of two nodes is different, except if they are symmetric with respect to the terminals, and increases together with the distance from the terminals. E.g., in the case we have 6 nodes, the betweenness of the terminals is 0, then it becomes 4 and then 6. In conclusion, the central nodes with respect to degree are all but the terminals, with respect to betweenness are the most inner ones (one or two according to the parity of the nodes).

To compare different graphs with respect to betweenness, it can be useful to have a normalized index. There are some different ways to normalize an index, each one having pros and cons.

A first normalization is

$$\mathrm{ng}_1(\mathrm{v}) = \frac{\mathrm{g(v)} - \mathrm{g_m}}{\mathrm{g_M} - \mathrm{g_m}}, \tag{4}$$

where g_M, g_m are the maximum and the minimum value of the betweenness for a given graph. In this way, for whatever graph, one gets values in the range $[0, 1]$. On the other hand, it is not possible to compare easily the betweenness values for nodes in graphs sharing some common feature, as the same number of nodes. Another criticality for unweighted graphs is the following. If a node has degree 1, its betweenness is equal to 0. After normalizing the betweenness as in (4), some nodes will have normalized betweenness equal to 0 in each graph, and so it is not possible to detect the existence of degree 1 nodes from the normalized betweenness.

A second normalization is

$$\mathrm{ng}_2(\mathrm{v}) = \frac{\mathrm{g(v)}}{\mathrm{g_M}}, \tag{5}$$

where g_M is the maximum of the betweenness for a given graph. As before, one gets values in the range $[0, 1]$ but now we can detect the existence of degree 1 nodes from $ng_2(v)$. It is not easy, however, to compare graphs from $ng_2(v)$ when the graphs share some common feature, as the same number of nodes, for example.

A third normalization, the one we adopt, is

$$\mathrm{ng}_3(\mathrm{v}) = \frac{\mathrm{g(v)}}{\mathrm{G_M}}, \tag{6}$$

where G_M is the maximum of the betweenness for all graphs sharing the same feature. For example, when we consider all graphs with the same number of nodes $G_M = \binom{n-1}{2} = (n-1)(n-2)/2$. The number G_M above is the betweenness of the central node of a star-shaped graph with n nodes.

4.3 Closeness

Another centrality measure is closeness. Closeness was first introduced by A. Bavelas (Bavelas, 1950), and later reintroduced as it is or slightly modified by various authors. The basic idea of closeness is that a node is central if it is not too far from every other node in the graph. The distance between two nodes is the length of the shortest path joining the two nodes, where the length of a path is the sum of the weights of the edges in the path. For unweighted graphs, as previously explained, we assume that every edge has weight 1.

The definition of closeness $C(v)$ of a node v is

$$\mathrm{C(v)} = \left(\sum_{\mathrm{v} \neq \mathrm{x}} \mathrm{w(x, v)}\right)^{-1}, \tag{7}$$

where $w(x, v)$ is the distance between the nodes x, v.

Of course, since we are interested in nodes for which the sum of distances is small, central nodes with respect to closeness are the ones for which $C(v)$ is large. In a modified closeness index, the authors propose to use the average distance instead of the sum of distances.

We normalize closeness by computing the following

$$\mathrm{NC(v)} = \frac{\mathrm{C(v)}}{\mathrm{C_M}}, \quad (8)$$

where C_M is the maximum of closeness for graphs sharing a common feature. For example, if we consider unweighted graphs with the same number n of nodes, then $C_M = 1/(n-1)$. In fact, in a star-shaped graph with n nodes, a node is a neighbor of all the others, and so its closeness is actually C_M. Of course, since the sum of distances is never smaller than $n-1$, then $C(v) \geq C_M$, for every node v in every graph with n nodes. We remark that the normalized closeness $NC(v)$ is exactly the same index we get by taking the average distance.

Closeness on paths provides results that are similar to betweenness (and then it is different from degree). In order to understand their difference, consider a path (a chain graph), and add a new terminal at one of the two far ends, so it becomes a fork with two teeth, obtaining still a tree. Whereas betweenness is zero at all three degree 1 nodes, closeness is the same at the terminals on the fork, but different at the last terminal.

4.4 Eigenvector centrality

The fourth and last centrality measure we consider is eigenvector centrality. The basic idea of eigenvector centrality is that a node is central if its neighbors are central, too. In more details, assume that $x(v)$ is the value of eigenvector centrality at the node v. Then, if $u_1, \dots, u_t$ are the neighbors of v, we have

$$\lambda\, \mathrm{x(v)} = \mathrm{x(u_1)} + \cdots + \mathrm{x(u_t)} \quad (9)$$

for a suitable constant λ not depending on v, and for every node v. We can rewrite the above equation by using the adjacency matrix A of the graph, and we have

$$\lambda\, \mathrm{X} = \mathrm{A\, X} \quad (10)$$

where X is a $n \times 1$ matrix whose elements are the eigenvector centrality values of the n nodes of the graph. Of course, we require that $x(v) > 0$ for every node. The Peron-Frobenius theorem guarantees that the adjacency matrix, no matter whether weighted, has a maximum eigenvalue of multiplicity one, and whose eigenvectors have positive elements, up to the multiplication by a scalar. Hence, the choice for the eigenvector centrality falls on the unique eigenvector with positive elements and norm equal to 1. It is evident that the values of the eigenvector centrality are normalized.

This fourth centrality measure is independent from the previous three ones, and it is easy to construct graphs whose nodes are differently ordered with respect to the four measures.

We will discuss application of these four measures on metro graphs.

4.5 Ishortest

The above centrality measures aim at finding the most important nodes in a graph, where 'important' depends on the index one considers. Such nodes are the ones whose cancellation is expected to largely modify the performances of the graph, even if no one can guarantee that. In fact, there is no way to compare easily the values of a centrality measure on a graph before and after removing a node and all edges containing it.

To overcome this difficulty, we propose a novel designed index, Ishortest. Its values depend on both a graph and all graphs obtained by cancelling a node at a time. In fact, to compute the values of Ishortest, we compare the lengths of the shortest paths before and after removing a node. Since the length of a path is well defined in graphs, no matter whether weighted or unweighted, we can compute Ishortest for nodes in every graph.

Let $\mathrm{G} = (\mathrm{V}, \mathrm{E})$ be a graph, where V is the set of nodes, and E is the set of edges. We assume that there are at least three nodes in V. Given two nodes x, y, we denote $\mathrm{w}(\mathrm{x}, \mathrm{y})$ the length of the shortest path from x to y. Of course, $\mathrm{w}(\mathrm{x}, \mathrm{y}) = \mathrm{w}(\mathrm{y}, \mathrm{x})$ and $\mathrm{w}(\mathrm{x}, \mathrm{x}) = 0$ for every $\mathrm{x}, \mathrm{y} \in \mathrm{V}$. The total length of shortest paths in G is then equal to

$$\mathrm{w}(\mathrm{G}) = \sum_{\mathrm{x}<y} \mathrm{w}(\mathrm{x}, \mathrm{y}) \tag{11}$$

where we suppose to order the nodes, so that $\mathrm{w}(\mathrm{x}, \mathrm{y})$ contributes only once to the sum above.

Now, we select a node v, and we compute the total length of shortest paths from v, that is to say,

$$\mathrm{w}(\mathrm{v}, \mathrm{G}) = \sum_{\mathrm{x}} \mathrm{w}(\mathrm{x}, \mathrm{v}). \tag{12}$$

Moreover, we construct the new graph $\mathrm{G}_{\mathrm{v}} = (\mathrm{V}', \mathrm{E}')$, where $\mathrm{V}' = \mathrm{V} - \{\mathrm{v}\}$ and E' is the subset of edges in E that do not contain v. According to G and v, the subgraph G_{v} may happen to be connected or not. The computation of the value of Ishortest at v varies according to the connectedness properties of G_{v}.

If we assume that G_{v} is connected, we compute $\mathrm{w}(\mathrm{G}_{\mathrm{v}})$, that is to say, the total length of shortest paths in G_{v}, and we define (in next sections, we use IS1, for brief)

$$\mathrm{Ishortest}(\mathrm{v}) = \frac{\mathrm{w}(\mathrm{G}_{\mathrm{v}})}{\mathrm{w}(\mathrm{G}) - \mathrm{w}(\mathrm{v}, \mathrm{G})}. \tag{13}$$

Of course, since there are at least three nodes in V, the denominator is strictly positive. Furthermore, when we take two nodes in V', then $\mathrm{w}(\mathrm{x}, \mathrm{y}) \leq \mathrm{w}_{\mathrm{v}}(\mathrm{x}, \mathrm{y})$ where $\mathrm{w}_{\mathrm{v}}(\mathrm{x}, \mathrm{y})$ is the length of the shortest path in G_{v} from x to y, because some edges in G are no more edges in G_{v}. More precisely, if all shortest paths between x, y in G pass through v, then the two lengths are different, otherwise the two lengths are equal. Hence, $\mathrm{w}_{\mathrm{v}}(\mathrm{G}_{\mathrm{v}}) \geq \mathrm{w}(\mathrm{G}) - \mathrm{w}(\mathrm{v}, \mathrm{G})$, that is to say, $\mathrm{Ishortest}(\mathrm{v}) \geq 1$. For example, if v is a terminal, then G_v is connected and $w(x,y) = w_v(x,y)$ because no shortest path passes through a terminal, unless it starts from there. Then, it follows that $w(G_v) = w(G) - w(v,G)$ that is to say, $\text{Ishortest}(v) = 1$ for every terminal v. To distinguish between non-terminal nodes with $\text{Ishortest}(v) = 1$ and terminals, we set $\text{Ishortest}(v) = 0$ for every terminal v.

On the other hand, we now assume that G_v has two or more connected components, and so v is a cut-vertex, as those nodes are referred to in literature. Of course, there is no path at all from x to y when the two nodes belong to different connected components, and so there is a loss of connections when the cut-vertex is removed. Moreover, when x, y belong to the same connected component, the length of the shortest path in G_v is not less than the length of the shortest path in G. Both the loss of connections and the increase in distance deserve to be considered. For this reason, Ishortest evaluated at a cut-vertex is a couple. From the definition of betweenness, we recall that $\sigma(x,y)$ is the number of shortest paths from x to y in a graph. Hence, we define $\sigma_v(x,y)$ the number of shortest paths between x, y in G_v: if $\sigma_v(x,y) > 0$ then the two nodes are in the same connected component, $\sigma_v(x,y) = 0$ otherwise. A good measure of the loss of connections is

$$\text{Ishortest}_1(v) = \frac{\sum_{\sigma_v(x,y)=0} w(x,y)}{w(G) - w(v,G)} \tag{14}$$

because the numerator is exactly the total length of paths that do not exist in G_v. In next sections, we use IS21 for brief.

To evaluate the increase of total length, instead, we take into account the sizes of the connected components that arise when v is removed. So, we assume that G_v is the disjoint union of $G_1, \dots, G_t$, and that G_i has n_i nodes. Of course, $n = n_1 + \cdots + n_t + 1$ is the number of nodes in G. Let

$$w_v(G_i) = \sum_{\substack{x<y \\ x,y \in G_i}} w_v(x,y) \tag{15a}$$

$$w(G_i) = \sum_{\substack{x<y \\ x,y \in G_i}} w(x,y) \tag{15b}$$

be the total length of shortest paths between nodes in G_i evaluated in G_i (15a) and in G (15b). An estimate of the increase of lengths of shortest paths is given by (in next sections, we use IS22, for brief)

$$\text{Ishortest}_2(v) = \sum_{\substack{i=1 \\ n_i \geq 2}}^{t} \frac{n_i \, w_v(G_i)}{n \, w(G_i)}. \tag{16}$$

The above equation is the average increase of lengths of shortest paths, weighted on the size of each connected components with at least two nodes. Finally, we set

$$\text{Ishortest}(v) = \big(\text{Ishortest}_1(v), \text{Ishortest}_2(v)\big). \tag{17}$$

The first component is always smaller than 1, because the numerator is a summand of the denominator. Of course, the larger the first component, the bigger the loss in connections due to the cancellation of the node.

We expect the second component to be close to 1 when there is one very large connected component and a few small ones. On the contrary, when there are at least two connected components of comparable sizes, we expect that component to decrease. In the case of weighted graphs, its value may increase much more than 1, depending on weights of edges on the new shortest paths. It is worth noting that this component is always equal to 1 if v (the cut-vertex) has degree two.

As an example, we consider a graph whose layout is a path (that is a chain) on 5 nodes. The values of $\mathrm{Ishortest}$ at the two terminals is 0. The cancellation of every other node produces a non-connected graph. When we remove the central node, we get $\mathrm{Ishortest(v)} = (\mathrm{IS21}, \mathrm{IS22}) = (\frac{12}{14}, \frac{4}{5})$, while, if we remove one of the two remaining nodes, we get $\mathrm{Ishortest(v)} = (\frac{9}{13}, \frac{3}{5})$. This example shows that when we remove the central node, the loss in connection is larger than when removing another non-terminal node. The analysis of the second component, in this case, is not particularly interesting because shortest paths do not change their lengths when removing a node.

5. WORKING EXAMPLE: WASHINGTON DC UNDERGROUND NETWORK

In this section, we consider the underground network of Washington DC, we construct the associated graph, and we compute the values of the centrality measures and Ishortest for each node of that graph.

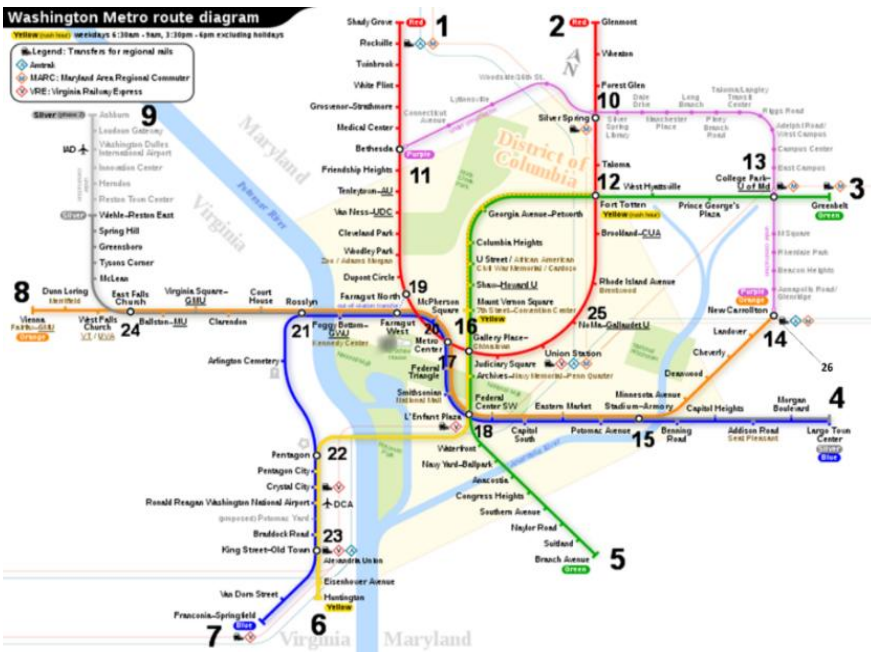

Figure 2. Washington DC underground network map with numbered stations/vertices.

Washington DC metro network (Figure 2) is a six-line system and its total length is over 188 km. The routes include trips from Washington to Maryland (Montgomery and Prince George's counties) and Virginia (Arlington, Fairfax, and Loudoun counties, and the towns of Alexandria, Fairfax and Falls Church).

The topological structure of the Washington DC metro is represented through an undirected graph G with 26 vertices and 32 links (Figure 3). Vertex 26 is a fictitious node added to represent the fact that node 14 is the terminus of both Orange and Purple lines. The length of this fictitious link is settled equal to 1 kilometer. Furthermore, nodes 20 and 25 are inserted because there are two different routes between nodes 17 and 19 and between nodes 12 and 16.

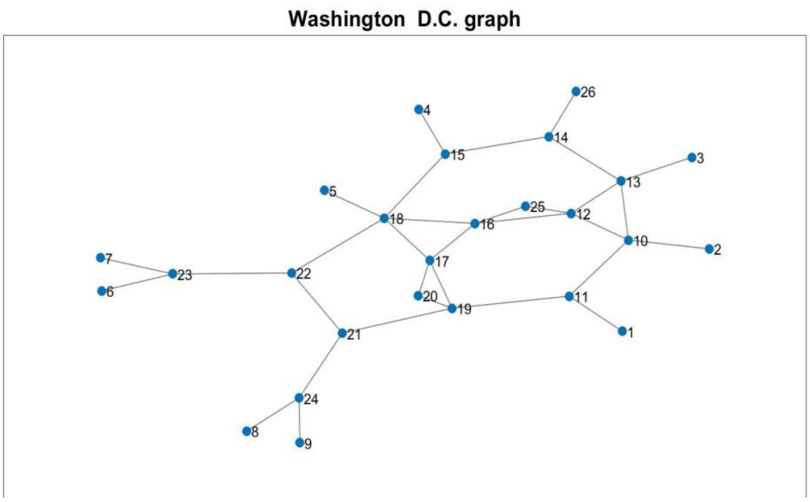

Figure 3. Graph of Washington DC underground network.

5.1 Centrality Indices

We first analyze the degree index. As it follows from Fig. 3, nodes from 1 to 9 and node 26 have degree 1, that is to say, each of them has only one neighbor. In this respect, they are terminal nodes. As previously explained, nodes 1 to 9 exist, while node 26 is fictitious. Nodes 20 and 25 are the only two ones whose degree is 2. All the other nodes have degree larger than two. The top degree node is node 18 that has 5 neighbors, the nodes $5, 15, 16, 17, 22$. In the graph, there are 7 nodes, $10, 12, 13, 16, 17, 18, 19$, with degree larger than 3, and they represent about the 27% of the total number of nodes. Since the top degree is 5, relatively small with respect to the total number of nodes, and other 6 nodes have degree 4, we can say that Washington underground network does not have a hub structure. The normalized degree $ndeg_1$ is equal to $0, 0.25, 0.50, 0.75, 1$ according to the degree in the set { 1, 2, 3, 4, 5}. The second normalized degree $ndeg_2$ is equal to $0.04, 0.08, 0.12, 0.16, 0.20$ according to the degree as above. Since the top degree is small if compared to $D_M = 25$, normalizing factor in the second normalization, $ndeg_2$ gets values in a small range.

Now, we consider the weighted graphs, and we analyze first the graph according to the number of stations. Given a node, we want to compute the number of stations contained in the edges that start from the considered node. Of course, we do not consider the given node in the computation. To achieve this goal, because of the way we weight edges by stations, we have to sum the weights of the edges containing the given node, and we have to subtract the degree of the node itself. The values are contained in the range $[1, 20]$. There are 8 nodes in the first quartile, 7 in the second one, 6 in the third, and 5 in the last one. The nodes in the last quartile are $10, 11, 13, 18, 24$, nodes $10, 13$ with the largest number of stations (20) on edges containing them. By comparing the lists of nodes with degree at least 4 and weight at least 16, we find the three nodes $10, 13, 18$. If we divide the weight of a node by its degree, we get an average number of stations on the edges that contain the node, and this further value can be of some interest to evaluate the accessibility to the underground network in a neighbor of the node.

Lastly, we consider the graph weighted according to the lengths of edges. As previously explained, we consider length as a negative attribute, and so we use the inverse of the length. As for the number of stations, to get the weight of a given node, we sum the weights of the edges containing the nodes. The values we get are in the range $[0.03, 4.15]$. There are 18 nodes in the first quartile, 4 in the second, 2 in the third, and 2 in the fourth and last one. The nodes in the last quartile are $16, 17$. If we compare them with the largest degree nodes, we see that they both have degree 4, while they do not belong to the fourth quartile by stations. This proves that, in this case, the weights by stations and by distance are independent. Once more, if we divide the weight by distance by the degree of a node, we get an average length of edges containing the given node, and this number can be of some interest when evaluating the performances of the network.

Now, we consider betweenness, and repeat our analysis as for degree. As previously stressed, the betweenness of terminals is 0, since a terminal is an inner node on no shortest path. However, when we consider node 20, it is clear from the topology that its betweenness is 0 as well. In fact, given two different nodes x, y, and a path joining them through node 20, we can shorten it by substituting $17, 20, 19$ with $17, 19$, because there is an edge linking node 17 to node 19. Of course, a computation by hand of betweenness is far from our ability, even in the case of a relatively small graph as the one associated to Washington DC underground network. There are 12 nodes in the first quartile, and their betweenness is 0; there are 9 and 2 nodes in the second and third quartile, respectively, and lastly, nodes 18, 21, 22 are in the fourth quartile. When we consider stations as weights, the only nodes in the fourth quartile are 18, 22. We remark that in the first quartile, we have the same 12 nodes as in the unweighted case. On the contrary, when we weight the graph according to the number of stations, we have 5 nodes in the third quartile. Lastly, when we weight the graph according to the inverse of the distance, there are 14 nodes in the first quartile, and 8 in the fourth one (nodes 10, 11, 13, 14, 15, 18, 21, 22). It follows that nodes 18, 22 have the highest betweenness whatever weight function we use. On the contrary, for degree, no node is in the fourth quartile for whatever weight function.

We repeat now the previous analysis for closeness. The values of this index, both in the unweighted case, and in the two weighted cases, are in a quite small range. For example, in the weighted-by-station case, the range is $[0.001; 0.004]$. Hence, when clustering the nodes by quartiles, it is more convenient to normalize the values by using equation

$$\mathrm{NC}_2(\mathrm{v}) = \frac{\mathrm{C(v)} - \mathrm{C_m}}{\mathrm{C_M} - \mathrm{C_m}} \quad (18)$$

where C_m, C_M are the minimum and the maximum value of closeness. When we consider the unweighted graph, there are three nodes in the fourth quartile, and they are nodes $17, 18, 19$. When we weight by station, there are two nodes in the fourth quartile, and they are nodes $16, 18$. When we weight by inverse of the distance, the situation is quite different. In fact, there are 13 nodes in the fourth quartile, that is to say, half of the total number. This shows that the values for this index are very concentrated, clustered around the maximum. The only node in the fourth quartile with respect to whatever weight we use is node 18.

The last centrality measure to consider is eigenvector centrality. There are 4 nodes in the fourth quartile in the unweighted case, and they are nodes $12, 16, 17, 18$. The weighted cases share the same behavior: the most of the nodes are in first quartile, 17 in the weight-by-station case, and 20 in the weight-by-distance one. In both cases, there are two nodes in the fourth quartile, nodes 10, 13 when we weight by the number of stations, and nodes 16, 17 when we use the inverse of distance as weight. It is evident that no node sits in the fourth quartile in all three cases. Moreover, nodes 10, 13 have the largest value in the case of weight by stations.

To summarize the above discussion, we consider all four centrality measures at the same time. In the unweighted case, node 18 is the only that sits in the fourth quartile with respect to all above indices, while node 17 appears in three fourth quartiles, and so they show themselves as the most central ones. In the opposite direction, all terminals are located in the first quartile for every centrality measures. Hence, as expected, terminals are the less central nodes in a graph. When we weight the graph by number of stations, no node sits in all fourth quartiles, while node 18 is the only that belongs to three fourth quartiles over four. Once again, node 18 shows itself as the most central in the graph, also with respect to the number of stations per edge. Lastly, when we weight the graph by the inverse of the distance, the situation is different, again. In fact, no node sits in more than two fourth quartiles, in spite of the fact that the fourth quartile of betweenness contains 8 nodes, and the one of closeness contains 13 nodes. So, when we weight the graph with respect to the inverse of the distance, we can say that no node is actually central. In conclusion, the overall of weight and topology produces effects that are not predictable on the centrality measures.

Table 2: List of four centrality indices for Washington D.C. underground graph.

	Degree			Betweenness			Closeness			Eigenvector		
Vertex	Adjacency	Station	Distance	Adjacency	Station	Distance	Adjacency	Station	Distance	Adjacency	Station	Distance
1	1	7	0.06	0	0	0	0.0101	0.0018	0.0553	0.013	0.032	0.000
2	1	4	0.13	0	0	0	0.0099	0.0021	0.0509	0.019	0.037	0.000
3	1	2	0.26	0	0	0	0.0095	0.0023	0.0448	0.018	0.019	0.000
4	1	6	0.08	0	0	0	0.0102	0.0020	0.0532	0.013	0.016	0.000
5	1	9	0.07	0	0	0	0.0116	0.0020	0.0541	0.028	0.023	0.002
6	1	3	0.50	0	0	0	0.0088	0.0020	0.0303	0.005	0.001	0.000
7	1	3	0.09	0	0	0	0.0088	0.0020	0.0432	0.005	0.001	0.000
8	1	4	0.09	0	0	0	0.0086	0.0019	0.0428	0.004	0.002	0.000
9	1	12	0.03	0	0	0	0.0086	0.0014	0.0459	0.004	0.006	0.000
10	4	24	0.59	52.83	36	85	0.0130	0.0026	0.0602	0.061	0.159	0.005
11	3	20	0.33	54.83	34	83	0.0133	0.0026	0.0601	0.043	0.078	0.005
12	4	18	1.52	32.00	74	32	0.0132	0.0029	0.0543	0.082	0.096	0.066
13	4	24	0.65	45.00	39	88	0.0123	0.0025	0.0618	0.059	0.160	0.003
14	3	14	1.25	41.00	31	82	0.0123	0.0024	0.0602	0.034	0.073	0.000
15	3	18	0.44	56.00	52	80	0.0135	0.0027	0.0597	0.042	0.046	0.007
16	4	16	3.91	45.17	73	0	0.0145	0.0036	0.0357	0.097	0.058	0.190
17	4	10	4.15	31.83	56	11	0.0149	0.0035	0.0399	0.092	0.023	0.238
18	5	24	1.78	110.33	115	102	0.0161	0.0037	0.0597	0.091	0.044	0.082
19	4	15	2.82	67.67	68	70	0.0147	0.0034	0.0588	0.067	0.042	0.162
20	2	4	2.71	0	0	0	0.0122	0.0031	0.0204	0.048	0.007	0.177
21	3	12	0.76	84.00	80	86	0.0139	0.0033	0.0534	0.038	0.012	0.019
22	3	12	0.72	95.33	90	90	0.0145	0.0035	0.0542	0.044	0.009	0.010
23	3	13	0.72	47.00	47	47	0.0112	0.0023	0.0477	0.016	0.004	0.000
24	3	22	0.24	47.00	47	47	0.0109	0.0024	0.0472	0.014	0.009	0.001
25	2	8	0.60	0	0	10	0.0119	0.0031	0.0441	0.054	0.036	0.030
26	1	2	1.00	0	0	0	0.0095	0.0022	0.0246	0.010	0.008	0.000

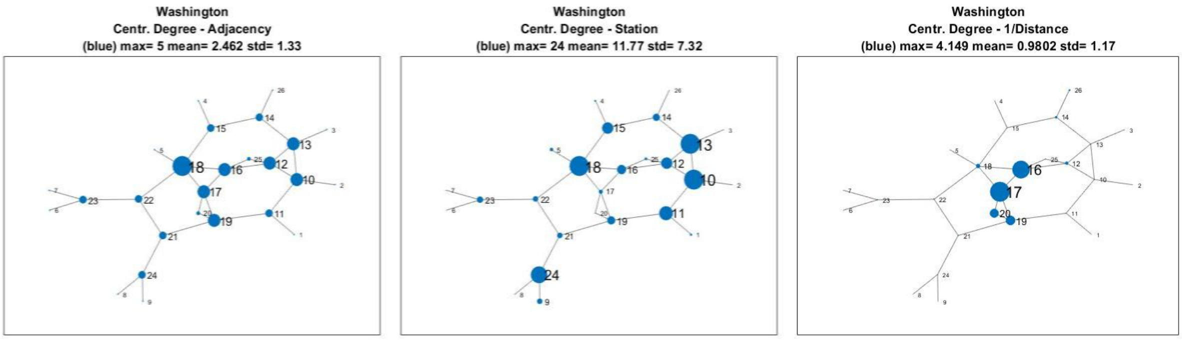

Figure 4: Degree Indices for adjacency graph, and weighted by station and 1/distance graphs.

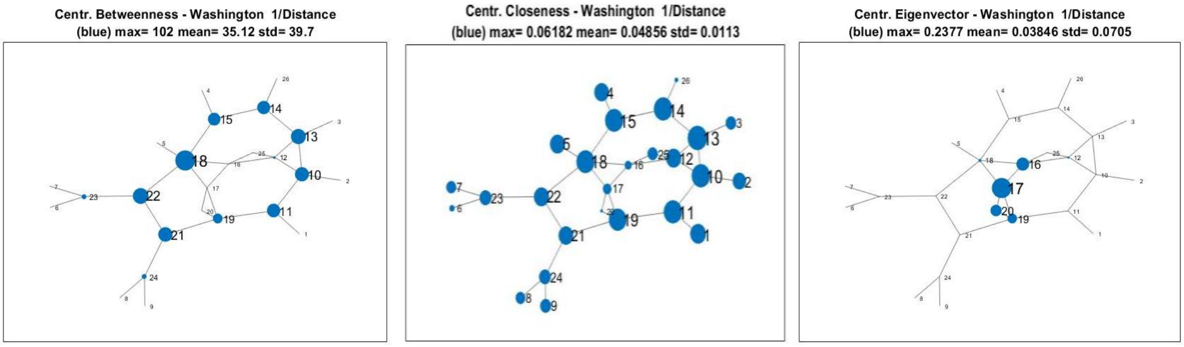

Figure 5: Betweenness, Closeness and Eigenvector Indices for weighted by 1/distance graphs.

5.2 Ishortest

When analyzing Ishortest values we must evaluate separately nodes that do or do not disconnect the graph. Table 3 reports results for adjacency and weighted graphs (note that in this application we use just Distance as weight) and figures 6 shows them graphically (IS1 is drawn in blue, IS21 in red, and IS22 in yellow).

For what concerns non cut-vertices, node 16 is the top value node for Washington graph when we consider weights, while node 19 is the top value node and the only in the fourth quartile when we consider topology only. When we remove node 16, and we weight the graph by distance, the increase of total shortest path length is up to 21%, which is a relevant value, taking into account the global size of the network. Node 12 determines an increase up to 12%, too. Node 17 is generally not significant contrary what occurring for centrality indices.

Now, we consider cut-vertices. They represent most of the nodes (10 out of 16, excluding termini). Looking at them on Fig. 3, one sees that such nodes represent a border of the central part. This is expected, since the graph looks like a central body with arms. These nodes are at the junction of the arms, or on the arms. Node 21 is in the first position for IS21 for all graphs with values 0.29, 0.31, and 0.38 for adjacency, stations, and distance, respectively. From the last value, it follows a loss of total shortest path length up to 37% w.r.t. original total value before removal, that is to say, the 37% of the shortest paths in the original graph do not exist anymore after removing node 21. For IS22 (measuring the increase of sub-graph total shortest path length due to removal) we see that they are rather close to 1. This can be explained as follows. When a cut-vertex is removed, we have a very large connected component and a few very small ones. The increase in length of the shortest paths in the large component is not so relevant, and so the IS22 value is close to 1. Node 18 is the exception with values 1.16, 1.18, 1.15 for adjacency, stations and distance matrices. Since the removal of node 18 produces two connected components, one with 24 nodes (the big one) and the other one with only a node, the actual increases in the big component are 26%, 29%, and 25%, respectively.

Table 1. Ishortest values of Washington DC's graphs (considering distance as weight), terminals excluded.

		Adjacency			Station			Distance		
Vertex	No.subgraphs	Index IS1	Index IS21	Index IS22	Index IS1	Index IS21	Index IS22	Index IS1	Index IS21	Index IS22
10	2		0.0945	1.0376		0.0968	1.0237		0.0898	1.0103
11	2		0.0925	1.0333		0.1110	1.0023		0.1394	1.0000
12	1	1.0283			1.0839			1.1269		
13	2		0.0987	1.0316		0.0891	1.0253		0.0814	1.0172
14	2		0.0987	1.0126		0.0937	1.0072		0.0907	1.0000
15	2		0.0914	1.0373		0.1010	1.0389		0.1015	1.0013
16	1	1.0478			1.1005			1.2182		
17	1	1.0253			1.0293			1.0534		
18	2		0.0792	1.1599		0.0946	1.1790		0.0977	1.1542
19	1	1.0937			1.0575			1.0331		
20	1	1.0000			1.0000			1.0000		
21	2		0.2926	1.0141		0.3091	1.0061		0.3771	1.0002
22	2		0.2833	1.0278		0.2716	1.0230		0.2703	1.0000
23	3		0.2122	1.0000		0.2005	1.0000		0.2063	1.0000
24	3		0.2186	1.0000		0.2439	1.0000		0.3322	1.0000
25	1	1.0000			1.0000			1.0000		

The reported results for Ishortest index show that Washington DC underground network has higher vulnerability according to the proposed measure, at either Fort Totten, Gallery Plaza, L'enfant Plaza and Rosslyn stations (nodes 12, 16, 18 and 21 respectively). It is worth noting that L'enfant Plaza and Rosslyn stations (nodes 18 and 21) are cut-vertices and therefore the loss due to their closure could be more crucial.

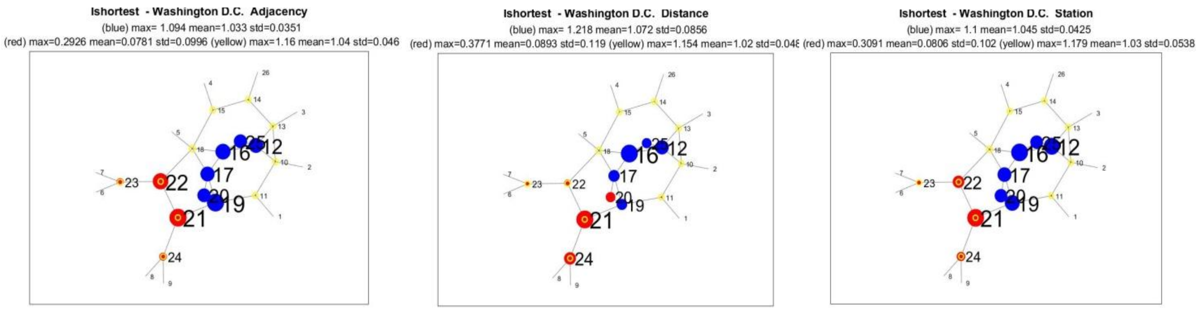

Figure 6: Ishortest indices for adjacency graph (a), and weighted by station (b) and distance graphs (c) (IS1 in blue, IS21 in red, IS22 in yellow).

Table 4. Synthesis of all indices for the 34 underground networks.

City	Number of Vertices	Number of Edges	Total Length(km)	Adjacency Matrix										Station Matrix										Distance Matrix									
				Degree Centrality	Eigenvector Centrality	Betweenness Centrality	Closeness Centrality	Ishortest IS1		Ishortest IS2				Degree Centrality	Eigenvector Centrality	Betweenness Centrality	Closeness Centrality	Ishortest IS1		Ishortest IS2				Degree Centrality	Eigenvector Centrality	Betweenness Centrality	Closeness Centrality	Ishortest IS1		Ishortest IS2			
				v	v	v	v	v	IS1	v	NS	IS21	IS22	v	v	v	v	v	IS1	v	NS	IS21	IS22	v	v	v	v	v	IS1	v	NS	IS21	IS22
Athens	12	14	85.8	6,8,9,10	8	8	8	7	1.000	8	2	0.640	1.000	9	9	8	10	7	1.000	8	2	0.581	1.000	8	8	8	8	7	1.000	9	3	0.600	1.000
Barcelona	36	53	131.2	18,20,30	30	20	30	20	1.139	18	2	0.266	1.002	17,24,33	33	20	27	13	1.100	18	2	0.241	1.000	30	30	20	29	13	1.085	18	2	0.219	1.001
Beijing	78	121	632.6	69	69	51	68,72	47	1.046	50	2	0.096	1.018	29	29	69	61	51	1.043	50	2	0.104	1.008	67	67	51	51	47	1.031	29	3	0.111	1.000
Berlin	34	45	141.9	30	34	28	31	25	1.113	30	4	0.202	1.013	30	30	18	33	18	1.093	30	4	0.265	1.007	23	23	31	31	18	1.066	30	4	0.293	1.004
Boston	29	34	114.7	18,26,17	26	26,29	26	26	1.107	29	2	0.447	1.012	20	20	26,29	26	26	1.051	29	2	0.485	1.006	26	26	26	26	26	1.027	27	3	0.500	1.000
Brussels	9	10	35.9	8	8	8	8	3	1.000	8	2	0.657	1.000	8	8	8	4	3	1.103	8	2	0.705	1.000	3	3	8	8	3	1.096	8	2	0.703	1.000
Bucharest	13	14	66.2	10,13	13	12,13	13	9	1.022	12	2	0.556	1.000	13	13	10,12	13	9	1.051	12	2	0.522	1.000	9	9	13	12	9	1.102	12	2	0.522	1.000
Buenos Aires	18	24	54.4	17	17	17	9,17,18	17	1.156	13	3	0.295	1.000	13	15	17	15	17	1.103	13	3	0.312	1.003	9	9	17	9	17	1.057	13	3	0.330	1.000
Cairo	12	14	88.8	7,8,9,11	11	7,9,11	11	11	1.111	7	3	0.443	1.000	9	9	9	11	11	1.117	7	3	0.451	1.000	11	11	7	11	11	1.054	7	3	0.430	1.000
Chicago	19	19	161.7	8,9,19	8	8	8	-	-	8	3	0.652	1.000	6	6	8	8,19	-	-	8	3	0.592	1.000	8	8	7	7	-	-	8	3	0.605	1.000
Delhi	45	59	365.4	25	28	39	30	25	1.081	39	2	0.411	1.016	38	40	39	27	25	1.067	39	2	0.407	1.003	27	27	39	17	30	1.094	39	2	0.432	1.001
Hongkong	35	39	205.7	16,18,19,22,30,33	18	33	33	32	1.064	19	3	0.406	1.000	24	24	30	30	18	1.099	19	3	0.359	1.000	16	16	33	33	30	1.040	19	3	0.408	1.000
Lisbon	13	15	40.6	8,9,10,11,13	10	9	8,10	10	1.130	13	3	0.395	1.000	8	8	10,11	10	10	1.257	13	3	0.408	1.000	11	11	9	9	10	1.192	13	3	0.415	1.000
London	75	112	405.5	43,57	57	55	55	55	1.232	30	2	0.182	1.000	53	53	55	60	55	1.114	40	2	0.203	1.019	65	65	55	55	55	1.099	40	2	0.280	1.012
Lyon	10	10	27.7	9,10	9	9,10	9,10	-	-	9	3	0.477	1.000	9	9	10	9	-	-	9	3	0.531	1.000	10	10	9	9	-	-	9	3	0.597	1.000
Madrid	56	90	296.2	30	30	37	50	30	1.081	37	2	0.387	1.000	25	25	50	50	46	1.131	37	2	0.409	1.000	51	52	30	38	46	1.072	37	2	0.468	1.000
Marseilles	7	7	21.1	5,6	5	5,6	5,6	7	1.000	5,6	3	0.710	1.000	6	6	5,6	5,6	7	1	6	3	0.752	1.000	5	7	5,6	5	7	1.000	6	3	0.761	1.000
MexicoCity	42	62	195.8	25	25	25	25	17	1.066	36	3	0.131	1.000	22,36	22	23	25	25	1.046	36	3	0.161	1.000	16	27	38	38	15	1.050	36	3	0.185	1.000
Milan	21	24	91.3	12,13,14,16,17,20,21	13	20	14,20	14	1.103	16	3	0.447	1.000	16	16	20	14	13	1.078	16	3	0.489	1.000	13	13	17	14	13	1.054	16	3	0.534	1.000
Montreal	11	12	61.6	8	8	8	8	9	1.000	7	3	0.449	1.000	8	8	8	8	9	1	8	3	0.450	1.044	8	9	7	8	9	1.003	7	3	0.471	1.000
Moscow	65	106	393.1	44,56,60,62	56	56	59	56	1.046	3	2	0.043	1.000	23,25	23	56	58	56	1.041	6	2	0.060	1.002	60	63	4	59	28	1.034	12	2	0.071	1.000
New York	85	124	418.6	27	64	72	72	72	1.053	70	3	0.167	1.032	70	17	27	56	36	1.081	70	3	0.197	1.009	53	53	70	70	36	1.116	70	3	0.210	1.037
Osaka	39	51	215.7	16	33	28	35	35	1.182	28	2	0.322	1.047	17	17	35	38	35	1.202	28	2	0.422	1.023	31	31	28	21	35	1.217	28	2	0.514	1.030
Paris	86	136	312.5	28,38,53,68,75,79	79	53	75	59	1.052	32	2	0.084	1.008	28	30	78	78	59	1.067	32	2	0.090	1.004	79	79	59	60	59	1.052	32	2	0.098	1.001
Prague	9	9	64.0	7,8,9	9	7,8,9	7,8,9	-	-	7	3	0.540	1.000	9	9	7,8,9	8,9	-	-	9	3	0.549	1.000	8	8	7	7	-	-	7	3	0.546	1.000
Rome	10	10	62.7	8,10	8	8	8	-	-	8	3	0.761	1.000	10	10	8	8	-	-	8	3	0.690	1.000	8	8	8	8	-	-	8	3	0.660	1.000
Saint Petersburg	17	20	117.9	16	16	16	16	13	1.102	12	3	0.310	1.000	15	15	16	16	13	1.032	15	3	0.327	1.000	16	16	16	16	13	1.017	12	3	0.321	1.000
Seoul	119	194	1089.6	72,96	96	78	78	17	1.057	29	3	0.114	1.000	43	43	72	103	66	1.073	29	3	0.124	1.000	99	88	17	17	45	1.037	16	2	0.125	1.001
Shanghai	88	142	683.5	30	65	30	70	30	1.058	38	2	0.167	1.000	30	28	30	72	30	1.053	38	2	0.173	1.000	61	61	37	31	51	1.034	38	2	0.173	1.000
Singapore	34	57	237.1	20,29	29	29	29	9	1.067	2	2	0.084	1.009	4,20	4	25	25	22	1.075	2	2	0.118	1.002	20	20	4	4	22	1.158	2	2	0.152	1.000
Stockholm	20	20	105.7	14	14	14	14	20	1.000	14	4	0.796	1.000	17	13	14	14	20	1	14	4	0.736	1.000	14	14	14	14	20	1.000	14	4	0.721	1.000
Tokyo	65	110	279.0	49	49	49	49	26	1.036	37	2	0.197	1.003	21	21	49	49	49	1.03	37	2	0.203	1.002	54	54	30	49	45	1.024	37	2	0.203	1.001
Toronto	12	12	75.9	7	7	7	7	8	1.000	7	3	0.843	1.000	7	10	7	7	8	1	7	3	0.840	1.000	9	9	7	7	8	1.000	7	3	0.848	1.000
Washington DC	26	32	218.2	18	16	18	18	19	1.094	21	2	0.293	1.014	10,13,18	13	18	18	16	1.101	21	2	0.309	1.006	17	17	18	13	16	1.218	21	2	0.377	1.000

Legend : v=vertex, NS=number of connected subgraphs, Isx=index value

5.3 Analysis of the data set

The results for all networks are synthesized in Table 4. Like in every synthesis, many details are lost and, above all, by reporting only the first node or tied ones (v) in the rank for each index, the distribution of their values cannot be disclosed. This could be a significant issue especially in large networks.

When studying a single graph, the interested researcher can repeat an analysis similar to the one we performed in last section. In this section, then, we try to infer some common feature for networks in the dataset we constructed following the rules explained in section 3.

Since we report the nodes with the largest values for the centrality measures, but not the values of the measures, it is not possible to compare graphs by values. Of course, when a comparison is mandatory, the values to be considered are the normalized ones, possibly, with a normalization factor common to all graphs in the set.

From the list of the nodes, it is evident that in all graphs with at most 20 nodes (Lisbon is the only exception), the same node attains the top value for all centrality measures, from a topological perspective, and the same happens for graphs up to 30 nodes (Lisbon, Milan and Washington are the exceptions). When one considers weights, the description is not so evident, as explained in the case of Washington. In conclusion, it is very likely that a graph associated to an underground network with at most 30 nodes has a node that attains the top values for all centrality measures, from a topological perspective. This does not happen when considering weighted graphs.

Now, we analyze the values of Ishortest reported in Table 4. Once more, the analysis is limited from the fact that we report only the top values and the nodes where Ishortest attains it, in a sense that we explain in the following.

In more details, the table reports the index values for Ishortest, distinguished in IS1, for non cut-vertices, IS21 and IS22 for cut-vertices. For IS2x also the number of sub-graphs is reported (NS). If IS1 is missing, then every node in the graph is either a cut-vertex or a terminal. As a general feature, this happens only on small underground networks (like Chicago, Lyon, Prague, Rome). For IS21 and IS22, instead, the top value is the one on IS21, and the node on the left is associated to that larger value. In the IS22 column, we report the value of IS22 for the node written on the left. That value is no more the largest value of IS22. In conclusion, we put first IS21 more than IS22.

Analyzing the IS1 values, we can observe that outcomes are significantly different when we consider the adjacency, station and distance matrices for the same graph. However, a node attains the largest value for adjacency, stations and distance for all graphs with at most 20 nodes. If we consider graphs with at most 30 nodes, then the same happens with Milan as only exception. We remark that the top node for centrality measures is not the same top node for IS1, in most graphs. This is an indirect proof of the fact that Ishortest and the centrality measures give different information on graphs.

If one forgets the nodes that attain the largest values, and considers the IS1 values only, it is standard practice to compute mean value and standard deviation. The three mean values are 1.0726, 1.0736, 1.0676 for adjacency, stations and distances, respectively, with a standard deviation of about 0.06 for all of them. This means that the largest increase in length of the shortest paths is equal to about 7.4%, and so it is not very much affected by the removal of whatever node, among the ones that do not disconnect the graph. Another standard practice is to evaluate the linear correlation between these values. The linear correlation of IS1 of the three couples is in the range (0.50, 0.77). So, even if we are not considering the nodes where IS1 attains its largest value, the largest values themselves do not change in an unpredictable way when moving from a topological point of view, to one of the two weighted perspectives.

When we consider the IS21 column, it is evident that the node that reports the largest value is almost always the same for both adjacency, stations and distance. For example, in the columns for adjacency and stations, the node that attains the largest value is the same with the exceptions of London, Montreal, Moscow, Prague, Saint Petersburg. For stations and distance, the only exceptions are Athens, Beijing, Boston, Montreal, Moscow, Prague, Saint Petersburg, Seoul.

Now, we forget the nodes where IS21 attains its maximum, and we focus ourselves on some standard statistical procedures. The mean values of IS21 for adjacency is equal to 0.381, for stations is equal to 0.372, and for distance is equal to 0.408. The standard deviation is about 0.213 for all three columns. This means that the lost in connections is less than 0.381, in general, for an underground network, when one removes a node, from a topological perspective, and analogously for weighted graphs. For IS21, the linear correlation between its values in the three columns is in the range (0.79, 0.99), and so the values are strongly correlated. This means that weights do not affect very much the analysis made on the topology of graphs in the dataset.

Since the reported values of IS22 are a consequence of the ones for IS21, we do not perform any analysis on them. We only remark that, with very few exceptions, the values of IS22 decrease when moving from adjacency to stations and to distances.

Finally, even if they are not reported in Table 4, IS21 has the highest mean w.r.t. their minimum possible value (0) and with the highest standard deviation. Since IS21 measures the loss in total shortest path length, it means that cut-vertices can be considered the most crucial nodes to preserve service performance.

It must be stressed that these comments are mainly drawn only on values of top rank nodes and do not take into account the whole index distribution of all nodes. This lack is likely to be more relevant for medium/large size graphs than for small ones.

6. CONCLUSIONS AND FURTHER RESEARCH

In the present paper, we have addressed the ambitious aim of designing standard strategies to find the most important stations in terms of functionality of whole underground networks, as well as vulnerability regarding disruption of each station.

After listing the possible disruptive events that may affect an underground network, we have described a methodology to represent such networks as undirected graphs, enlightening their functional properties. We have applied that methodology to thirty-four underground networks of cities all around the world, so to construct a representative dataset. We remark that, for every underground network, we have constructed both the unweighted graph that captures the topology of the network, and two different weighted graphs: as weights, we have used the number of stations on each edge, so to highlight accessibility features to the network, and the distance between two nodes. Since distance can be considered either a positive attribute (e.g. territorial penetration) or a negative one (e.g. travelling time), according to the analysis, we have used 1/distance when we consider it negative, and we have used distance when we consider it positive.

The strategy we have designed consists in evaluating suitable indices on the nodes of graphs of the dataset. Hence, we have recalled definition and properties of four classic centrality measures, namely: degree, betweenness, closeness, and eigenvector centrality. In comparing different graphs or different indices on the same graph, normalization is a very powerful tool, and so we have illustrated pros and cons of some different normalizations of indices above. Finally, we have introduced a new index, Ishortest, designed to evaluate performances of underground networks when a node, and all edges containing it, is removed from a graph. To evaluate Ishortest at a node, we first check whether the node is a cut-vertex. If it is not the case, we compare the total shortest path length of the graph without the node with the total shortest path length between couples of nodes different from the removed one. If the node is actually a cut-vertex, we compute the number of connected components, the weighted average of the total amount of path lengths between nodes in different connected component (in the original graph, of course), and the total path length in the graph without the node. So, from the values Ishortest takes, it is easy to evaluate the importance of a node from a transportation point of view.

To illustrate the four centrality measures and Ishortest, we have performed a detailed analysis of Washington D.C. underground network. That section can be considered a template of possible uses of the centrality measures and Ishortest in finding the most important nodes in an underground network. Even if our main concerns are underground networks, our approach can be easily modified to study networks of different transport modes.

To summarize our analysis on Washington D.C. underground network, we report that L'Enfant Plaza station is the most central station in the network from a functional point of view when using centrality measures only. With functional point of view, we mean that we consider the topology of the graph but not the weights on its edges. When we consider the accessibility to the network, and so we weight edges by number of stations on them, L'Enfant Plaza station is the only that belongs to three fourth quartiles over the four centrality measures, and so L'Enfant Plaza station is the most central one also from that point of view. Instead, when we consider distance as a negative attribute, no station is in more than two fourth quartiles over four and so no station can really be considered central from that point of view. When considering Ishortest, the first remark is that 10 nodes over 26 are terminals, 10 of the remaining 16 are cut-vertices, and only 6 are not. Washington

D.C. network has the shape of a central body with arms. The cut-vertices are on the arms and at the junction of the arms with the body. So, the body is relatively small if compared to the arms. Among the nodes that do not disconnect the graph, Ferragut North is the most important node from the topological point of view, while Gallery Place station is the most important from the accessibility and economical points of view, because their removal causes the largest increase of the total shortest path length. When considering cut-vertices, there are two different data to consider: the loss of total shortest path length, due to the loss of connectivity; the possible increase of the total shortest path length inside each connected component. When we consider the loss in connectivity, Rosslyn station is the most important from all three points of view. Its removal is responsible of a loss of about 37% of the original total shortest path length. When we consider the increase of total shortest path length, L'Enfant Plaza is the most important station .

Finally, we report some data for all networks in the data set, and we infer some common features.

Small networks, such as Marseilles, Prague, Rome and Stockholm, are driven by topological characteristics. In those networks, the ranking of stations depends mainly on the underlying topology more than on weights. Therefore, indices yield rather similar results between both adjacency and weighted graphs and between networks.

In large networks, such as Seoul and Shanghai, each index enlightens different properties of the network and gives us a different station in the top rank position, though analyzing ranking distributions there is a crowded set of stations present in the fourth quartile for all centrality measures. In these networks, the most important stations are located in the central part of the networks (those networks have a circular, rectangular, or polygonal grid in the middle). Besides, the stations that impose the highest cost and are not cut-vertices, are precisely those located in the central part of networks. On the other hand, costly stations that their removal split the network into two or more connected sub-graphs, are the ones connecting the suburban areas to the central part of the network, as in the case of Washington D.C. Some networks amongst the largest ones, e.g. Beijing, Mexico City, Moscow, Seoul, Shanghai and Tokyo, show the least percentage cost in case of disruption in any of their stations. In all these networks, more than half of the stations have degree larger than three. On the contrary, smaller networks, e.g. Chicago, Prague and Rome, show the highest percentage cost in case of any disruption. In those networks, the majority of stations have degree one (they are termini) or three.

The effect of inserting new stations or links between existing stations can be examined through the study of the proposed indices by evaluating how station ranking changes. Then, indices will provide us with information about how the cost due to removal of a station will be reduced by inserting a new station or link in another part of the network. With this approach in mind (a mixed between the "what if" and the "what to" approach), we will be able to decide whether it is worthwhile to invest on a new station or new links or their combination.
We are also planning future studies on Ishortest to analyze the impact of disruption in more than one station or in one or more links at the same time. The selection of stations or

links to be removed at the same time can be done either randomly or by a defined strategy involving different criteria of measuring the effects of removal.

As a final remark, it is worth noting that the above definition of Ishortest deserves more insight. In fact, the present design of Ishortest does not take into account the real use of network (that is demand) but only its potential use (and without any congestion phenomena). An attempt of including demand in it requires a further investigation. A promising direction is the use of weighted graphs to model the problem, where weights are the number of passengers observed or predicted for each origin-destination couple.

BIBLIOGRAPHY

Aydin, N.Y., Duzgun, H.S., Heinimann, H.R., Wenzel, F., Gnyawali, K.R., (2018). Framework for improving the resilience and recovery of transportation networks under geohazard risks. International Journal of Disaster Risk Reduction **31**, 832–843. https://doi.org/10.1016/j.ijdrr.2018.07.022

Ayyub, B.M., (2014). Systems resilience for multihazard environments: Definition, metrics, and valuation for decision making. Risk Analysis **34**, 340–355. https://doi.org/10.1111/risa.12093

Balal, E., Valdez, G., Miramontes, J., Cheu, R.L.,(2019). Comparative evaluation of measures for urban highway network resilience due to traffic incidents. International Journal of Transportation Science and Technology 8, 304–317. https://doi.org/10.1016/j.ijtst.2019.05.001

Bavelas, A., (1950). Communication Patterns in Task- Oriented Groups. The Journal of the Acoustical Society of America **22**, 725–730.

Bell, M.G.H., (2000). A game theory approach to measuring the performance reliability of transport networks. Transportation Research Part B: Methodological **34**, 533–545. https://doi.org/10.1016/S0191-2615(99)00042-9

Bettelini, M., (2019). *Systems approach to underground safety. Underground Space*. https://doi.org/10.1016/j.undsp.2019.04.005

Bonacich, Ph., (2012). Power and Centrality: A Family of Measures. American Journal of Sociology **92**, 1170–1182.

Borgatti, S.P., Everett, M.G., (2006). A Graph-theoretic perspective on centrality. Social Networks **28**, 466–484. https://doi.org/10.1016/j.socnet.2005.11.005

Deng, Y., Li, Q., Lu, Y., (2015). A research on subway physical vulnerability based on network theory and FMECA. Safety Science **80**, 127–134. https://doi.org/10.1016/j.ssci.2015.07.019

Freckleton, D., Heaslip, K., Louisell, W., Collura, J., (2012). Evaluation of Resiliency of Transportation Networks after Disasters. Transportation Research Record: Journal of the Transportation Research Board **2284**, 109–116. https://doi.org/10.3141/2284-13

Freeman, L.C., (1977). *A Set of Measures of Centrality Based on Betweenness*. Sociometry. https://doi.org/10.2307/3033543

Grenzeback, L.R., Lukmann, A.T., (2007). *Case Study of the Transportation Sector ' s Response to and Recovery from Hurricanes Katrina and Rita*. Health (San Francisco) 1–44.

Jamakovic, A., van Mieghem, P., (2008). On the robustness of complex networks by using the algebraic connectivity. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial

Intelligence and Lecture Notes in Bioinformatics) 4982 LNCS, 183–194. https://doi.org/10.1007/978-3-540-79549-0_16

Jenkins, B.M., Gersten, L.N., (2001). Protecting Public Surface Transportation Against Terrorism and Serious Crime : Continuing Research on Best Security Practices San José State University. Transportation.

Lordan, O., Sallan, J.M., Simo, P., (2014). Study of the topology and robustness of airline route networks from the complex network approach: A survey and research agenda. Journal of Transport Geography 37, 112–120. https://doi.org/10.1016/j.jtrangeo.2014.04.015

Mussone, L., Notari, R., (2017). Structure Indicators for Transportation Graph Analysis I: Planar Connected Simple Graphs. Networks and Spatial Economics **17**, 69-106. https://doi.org/10.1007/s11067-015-9318-2

Rodan, S.A., (2011). Choosing the Beta Parameter When Using the Bonacich Power Measure. Journal of Social Structure **12**, 5–11.

Scott, D.M., Novak, D.C., Aultman-Hall, L., Guo, F., (2006). Network Robustness Index: A new method for identifying critical links and evaluating the performance of transportation networks. Journal of Transport Geography **14**, 215–227. https://doi.org/10.1016/j.jtrangeo.2005.10.003

Segarra, S., Ribeiro, A., (2016). Stability and continuity of centrality measures in weighted graphs. IEEE Transactions on Signal Processing **64**, 543–555. https://doi.org/10.1109/TSP.2015.2486740

Serulle, N.U., Heaslip, K., Brady, B., Louisell, W.C., Collura, J., (2011). Resiliency of Transportation Network of Santo Domingo, Dominican Republic. Transportation Research Record: Journal of the Transportation Research Board **2234**, 22–30. https://doi.org/10.3141/2234-03

Sharma, N., Tabandeh, A., Gardoni, P., (2018). Resilience analysis: a mathematical formulation to model resilience of engineering systems. Sustainable and Resilient Infrastructure **3**, 49–67. https://doi.org/10.1080/23789689.2017.1345257

Sun, X., Gollnick, V., Wandelt, S., (2017). Robustness analysis metrics for worldwide airport network: A comprehensive study. Chinese Journal of Aeronautics **30**, 500–512. https://doi.org/10.1016/j.cja.2017.01.010

Yang, Y., Liu, Y., Zhou, M., Li, F., Sun, C., (2015). Robustness assessment of urban rail transit based on complex network theory: A case study of the Beijing Subway. Safety Science **79**, 149–162. https://doi.org/10.1016/j.ssci.2015.06.006

Yu, H., Wang, Y., Qiu, P., Chen, J., (2019). Analysis of natural and man-made accidents happened in subway stations and trains: based on statistics of accident cases. MATEC Web of Conferences 272, 01031.

