Journal Pre-proof

Novel centrality measures and applications to underground networks

L. Mussone, H. Viseh, R. Notari

PII: S0378-4371(21)00863-3
DOI: https://doi.org/10.1016/j.physa.2021.126595
Reference: PHYSA 126595

To appear in: *Physica A*

Received date : 14 March 2021
Revised date : 4 November 2021

Please cite this article as: L. Mussone, H. Viseh and R. Notari, Novel centrality measures and applications to underground networks, *Physica A* (2021), doi: https://doi.org/10.1016/j.physa.2021.126595.

NOVEL CENTRALITY MEASURES AND APPLICATIONS TO UNDERGROUND NETWORKS

L. Mussone[a], H. Viseh[b], R. Notari[c]

[a] DABC, Politecnico di Milano, Milano, Italy

[b] DICA, Politecnico di Milano, Milano, Italy

[c] DMAT, Politecnico di Milano, Milano, Italy

ORCID

L. Mussone 0000 0002 2431 080X

R. Notari 0000-0003-4108-205X

H. Viseh 0000-0002-5477-2677

corresponding authors: lorenzo.mussone@polimi.it, roberto.notari@polimi.it

Abstract:

Two novel centrality indices, PathRank and Icentr, are defined. PathRank is a generalization of the PageRank algorithm, suitable to rank nodes of undirected graphs according to number and weight of paths in the graph. Icentr ranks the nodes of the graph by means of a combination of the weights of nodes and edges, scaled according to the distance from each node, one at a time. We apply the two novel indices to underground transportation networks, since these networks represent an infrastructural backbone for the transportation system of most big cities over the world. The characterization of the most important components of those networks and the simulation of their responses when they stop working properly, are vital for maintaining the mobility service at a desirable level. Since there are different ways to associate a graph to an underground network according to the degree of detail and aims of the study, we describe the methodology we adopted to associate a graph to such a network. The methodology was applied to 34 underground networks of worldwide cities, and the resulting graphs constitute the reference dataset. A detailed study of both Boston network and the dataset is proposed as prototypal for either a graph alone or all graphs in a dataset. Results show how different features of graphs are revealed by the two novel indices.

Keywords: Underground networks, Graph centrality indices, Adjacency matrix, Disruption, Dataset comparison

1. Introduction

The paper originates from the purpose of evaluating the performances of underground networks, an infrastructural backbone for urban transportation. To achieve our goal, we designed two centrality indices, PathRank and Icentr, for undirected, simple and connected graphs, and we applied them to graphs representing underground networks from a topological point of view. PathRank and Icentr can be useful as well to analyze graphs arising from different applications.

The characterization of the most important components of networks and the simulation of their responses when they stop working properly, are vital for avoiding devastating scenarios and maintaining the service at a desirable level.

Since we are mainly interested in the topology of underground networks, we focus on stations and tracks, including and representing into them all other components of networks. The main purpose is then to provide a methodology to locate the most important stations in the working network. We expect that the removal of such important stations, together with the links through them, will cause the highest reduction in the performance of considered networks.

Now, 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.

In the transportation system context, disruption is defined as a deviation from planned operations. Because of its intensity, a disruption can affect performances of stations, links between two stations or even an entire line of the network. Disruption or reduction in operation of any component of a transportation network can be highly expensive from a financial, economical and temporal point of view, for both passengers and operating companies. However, in a complex transportation system, some stations are more critical than others for the functionality of a network.

Disruptions can be classified according to their origins as

- Natural Disaster;
- Intentional or Terrorist Attack;
- Random Failure or Incident.

Now, we describe disruptions in more details according to the previous classification.

The most common natural disaster affecting underground networks is flood. In general, flood causes cancellations of some trips or the stop on some lines. An overall delay is a consequence of the increased congestion on routes passengers are forced to use just to avoid flooded links. Other worth mentioning natural disasters are earthquakes. They may have different effects both for the entity of damages and for time needed to recover them. While smaller earthquakes may only force the underground train to slow down until the system is not completely verified, larger earthquakes can cause structural damages causing the suspension of services, lasting from hours to weeks. Intentional attacks are targeted destructions caused by outside artificial forces, mainly terrorists. 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 terrorists (Yu et al., 2019). It is almost impossible to specify the corresponding destructive power for a random failure, and therefore, a random failure is generally described as a dysfunction of a network due to failure on one or several nodes or edges with random probability of occurrence. Every local failure may affect the normal functionality of more components or even of the whole system. Technical malfunctions such as power, gear, or brake failures, operational mistakes by staff or drivers, temporary suspension of service for special activity, maintenance or safety inspection, are examples of random failures.

In comparison with other means of transportation, underground systems usually have good safety records. However, because of the large number of persons potentially involved, the possible damage in case of disruption is high. E.g., 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 underground systems (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.

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 lesser economical, social and operational costs in case of any disruptive event. Reliability means that the expected additional trip cost due to a disruption is 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 et al., 2006).

Due to the reasons mentioned above, researchers continuously aim to identify useful techniques and quantitative measures to pinpoint the most critical components of a network, and to evaluate the impact that each component has on performance of the whole network.

With that aim in mind, we have built a dataset of 34 graphs derived from such networks, amongst those of most known cities in the world. Such a dataset has been already studied through the four well-known centrality indices, namely Betweenness, Closeness, Degree and Eigenvector centrality, and through the values of Ishortest, an index specifically designed for the purpose (Mussone et al. 2020).

To get the described goal, we propose two further novel indices enabling us to rank nodes of a graph, by identifying those for which exposure to a disruptive event may lead to the most harmful consequences for the network.

The paper is organized as follows. Section 2 reports some applications of centrality indices on performance assessing to transportation networks. In section 3, the two new indices are presented. Section 4 describes how a graph is created from the underground network map and contains some of the main features of the dataset. Section 5 reports a detailed case study based on the Boston (USA) network, and a detailed analysis of the above mentioned dataset. Section 6 concludes the paper resuming principal achieved outcomes and outlines future developments.

2. Centrality indices in transportation

This review does not aim to present a comprehensive list of the enormous amount of work done on applications of centrality indices on transportation networks, but only to recognize their role in the present research through a few significant contributions. Graph-based indices such as centrality indices are among the most commonly used tools enabling researchers to analyze and explain some features of networks. It is worth noting that every index highlights some features of a transportation network, but neglects some others. For example, classical centrality indices, such as Degree or Closeness, highlight some topological properties, but neglect important aspects associated with transportation such as time delay and traffic flow. Since in a transportation network a node can be critical from some, but not all, perspectives, it would be better to consider more indices and to compare their outcomes. In addition, all traditional centrality measures such as Degree, Closeness,

Eigenvector and Betweenness centrality, are vertex based and do not provide any information about edge centrality.

Degree and Eigenvector centrality are measures that date back at least to the beginning of XX century, and aim at measuring the attractiveness of each node. Two more centrality measures are Betweenness and Closeness. They were introduced by Freeman (1978), and are only based on the number of intermediary nodes on shortest paths between couples of nodes. They disregard the node strength. Also in some proposed generalization of Betweenness (Brandes, 2001) and of Closeness (Newman, 2001), the length of the shortest paths only depends on the sum of the reciprocal link weights.

(Opsahl et al., 2010) introduced a new generalization for Freeman (1978) centrality indices in which both the number of edges and their weights are considered. This approach provides us with a better insight regarding the involvement level of each node within the network. Authors designed the new index

$$C_D^{w\alpha}(i) = \left(\sum_{j=1}^{N} x_{ij}\right)^{1-\alpha} \left(\sum_{j=1}^{N} w_{ij}\right)^{\alpha}. \quad (1)$$

In (1), the number of nodes, the ij −element of the adjacency matrix and the weight of the edge between nodes i and j are N, x_{ij}, w_{ij}, respectively. The α parameter needs to be calibrated according to the studied case. For α = 0, one gets Freeman's (1978) measure of an unweighted network. For $\alpha = 1$, one gets a generalization of degree in which only node strength is considered. The authors also proposed a generalization for Closeness and Betweenness. It takes into account both aspects mentioned above, by means of the introduction of α. In particular, the index the author proposed is

$$C_C^{w\alpha}(i) = \left(\sum_{j=1}^{N} d^{w\alpha}(i,j)\right)^{-1}. \quad (2)$$

In (2), the weighted distance $d^{w\alpha}(i,j)$ between nodes i and j is equal to the minimum of $w_{hk}^{-\alpha}$ for every edge hk in a shortest path between nodes i, j.

(Wang and Cullinane, 2016) utilized the centrality measures proposed by (Opsahl et al., 2010) to analyze the maritime transportation network, where container ports are considered as nodes and linear shipping service as edges. In addition, the weekly transportation capacity was applied as weight on each edge.

(Wang et al., 2011a) applied three classical centrality indices, Degree, Closeness and Betweenness, on the air transport network of China (ATNC) to explore the structure of the network and nodal centrality of individual cities. Their case study involved all cities with operating airports in mainland China (excluding Hong Kong, Macao, and Taiwan and some small airports, which do not have regular flights) over six months of observation. The case study network consisted of 144 nodes and 1018 links. The proposed approach allowed the authors to find a high correlation between the values of the considered indices and socio-economic measures of cities such as air passenger volume, population, and gross regional domestic product.

(Wang et al., 2011b) examined the correlation between "street centrality" and land use density in Baton Rouge, Louisiana. They represented the road network of East Baton Rouge Parish and its surroundings, comprising about 1809 miles of roadway, through a network with 12,235 nodes and 17,219 links (a link is a street segment). They used population and employment densities as indicators of land use intensity. Street centrality is obtained through applying Closeness, Betweenness, and Straightness centrality indices. The latter index refers to the hypothesis that the connectivity between two points is better when the path is straight. Finally, the results from

centrality measures and land use density measures are transformed into the same units through a kernel density estimation and floating catchment areas. Acquired results indicate a high correlation between street centrality indices and land use densities showing the significant interplay between land use pattern and transportation network topology.

(Tsiotas and Polyzos, 2015) introduced a new centrality index called "Mobility Centrality" for analyzing traffic flow in a road transportation network. They used this index to measure the propensity of each node to attract network flow. The authors utilized Straightness Centrality (firstly proposed by Crucitti et al.2006),

$$C_i^c = \frac{1}{|V|-1} \sum_{j=1, i \neq j}^{|V|} \frac{d_{ij}^E}{d_{ij}} \quad (3)$$

which is a centrality index that determines the level of each network path deviation from the straight line, and a modified formula of kinetic energy. In Straightness Centrality formula (3), d_{ij}^E is the Euclidean distance, d_{ij} is the travelled distance between nodes i and j, and $|V|$ is the number of nodes.

(Cheng et al., 2015) proposed three new indices, named Commuter Flow, Time Delay and Delay Flow in which the authors consider commuter flow and the amount of delay induced to the passengers due to disruption at each node. In this way, they solve the problem of neglecting flow rate and delay through analyzing the network performance by the three novel centrality indices. The formula of Commuter Flow Centrality for a disrupted node uses the number of commuters per hour affected by that disruption. Time Delay Centrality is defined as time passengers will spend if they would take a different mean to reach their destinations in case of disruption in that node. Delay Flow Centrality is calculated as the ratio between Time Delay centrality and the total commuter flow of the considered network.

According to (Gu et al., 2020), vulnerability of a transportation network can be examined through analyzing the variation in a specific quantitative index. The choice of the index for analyzing the network vulnerability depends on the aim of investigation since different indices evaluate vulnerability of the system from different perspectives. Authors categorized the concept of vulnerability according to the utilized index into *connectivity vulnerability, accessibility vulnerability, and capacity vulnerability.* With this approach in mind, (Gu et al., 2020) proposed two measures as examples of topology-based indices. The former quantifies the relative change in the average shortest distances between network nodes, the latter considers the reversed average distance between nodes of the network as a measure of efficiency of the network in conveying passengers.

In addition, other vulnerability indices based on system features providing with information regarding the network performance, can be obtained by applying topology-based indices with different inputs. As an example, travel cost can be utilized instead of the average shortest travel distance in the previous index.

(Kumar et al., 2019) suggested a methodology for ranking the links in a road network. According to authors, in a transportation network, some links can be more critical considering daily functionality of the network, evacuation planning and emergency operations. Closure of even one critical link can alter the whole travel pattern significantly, leading to major changes in the travel pattern. They consider three factors for developing a link criticality indicator, including link traffic volume, connectivity to important facilities, and number of origins and destinations served by the link.

Complex Network theory has proved to be a powerful tool in many research fields. Among the many, see (Fei Xiong et al., 2021) for an application to epidemic dynamic and diffusion processes, and (Fei Xiong et al., 2020) for a recommendation model that includes evolutionary opinion interactions. In transportation, complex network theory provides a different point of view from which to explore properties of transportation networks. (Latora and Marchini, 2002) studied in detail the Boston underground network, and proved that it has the small world property. Moreover, they proposed some efficiency measures. Later, in (Derrible and Kennedy, 2010), the authors proved that metro networks are quite often scale-free and small world, and made suggestions on how to improve robustness of such networks.

Along the same line of research, (Dimitrov and Ceder, 2016) considered the Auckland integrated public transport network, and proved that it is a mixture of scale-free and exponential network.

(Xingtan Wu et al., 2018) considered transportation networks from complex networks and graph theory points of view. The authors introduce a novel centrality index, the node occupying probability, and they use it to evaluate the robustness of metro networks.

3. Two novel indices

Each of the indices mentioned above will provide us with a variety of information about the topology of the network or other desirable aspects depending on the considered weights. Moreover, they enable us to rank nodes of the network from different perspectives. However, the obtained results from these indices may also be misleading and may not reflect the node importance or an accurate comprehensive perspective of the network functionality. For example, Degree and Betweenness assign 1 and 0, respectively, to every terminal node. Hence, these two indices do not distinguish between different terminals, while, in underground networks, different terminal nodes have different importance. Closeness only depends on the weights of shortest paths, and so it does not take into account paths whose weight is greater than the minimum. Since such paths do play a role in the description of transportation networks, closeness, as it is, is not suitable for applications to transportation we want to consider. Mending the aforementioned deficiencies, we propose two new indices, which have notable advantages and increase our capability for a more extensive investigation of transportation networks, particularly underground networks.

The first novel index (PathRank) is a generalization of the PageRank index, invented by Lawrence Edward Page, which is a well-known ranking algorithm for Google. For a deep description of PageRank, see (Bryan and Leise, 2006). When running PageRank on undirected graphs, one gets the normalized degree as value at each vertex, and so this index does not give new information on undirected graphs. The idea of the generalization is to consider paths instead of edges. PathRank index provides us with a straightforward vision about each node's accessibility to other areas of the network.

In the second proposed index (Icentr), we use a combination of both vertex weight and link weight to overcome some of the limitations of previous indices in which only one among node weights or link weights were considered for ranking purpose. In this way, we have more reliable and comprehensive results in comparison to previous measures.

Of course, both the absolute values and the normalized values of an index are important. However, since we want to use indices to rank nodes, the absolute and the normalized values carry the same information. On the other hand, normalized values provide us an easier way to compare them. Hence, in the following, we always refer to normalized values for indices we consider. If v is the

value the index takes at a node, and M, m are respectively the maximum and the minimum value the index takes at a node of the graph, the normalized value is

$$nv = \frac{v-m}{\mathrm{M}-m} \tag{4}$$

3.1 PathRank

In this section, we define the new index PathRank. As said, it takes inspiration from the PageRank algorithm, and is suitable for transportation. We first recall the PageRank algorithm and we explain why it is not suitable for the applications we have in mind, and then we define PathRank.

Since there are different approaches to PageRank, we follow the one described in (Bryan and Leise, 2006). Let $G = (V, E)$ be a graph. The goal of PageRank algorithm is to rank vertices according to their importance. To fix ideas, we assume G has n nodes, labelled as $1, 2, \dots, n$, and we set x_i the score of node i. In PageRank, the score x_i depends on the scores of neighbor nodes as follows

$$x_i = \sum_{\substack{j=1, \\ j \neq i}}^{n} a_{ij} \frac{x_j}{d_j}. \tag{5}$$

In the previous equation (5), d_j is the number of edges outgoing from node j and $A = (a_{ij})$ is the adjacency matrix of G. For sake of completeness, we recall that the elements of A are $a_{ij} = 1$ if there is an edge from node j to node i, and 0 otherwise. The order of the nodes does not play an important role. The meaning of the previous equation is that the score x_j of j is equally distributed among the nodes one can reach starting from j, and the score of node i is the sum of all contributions associated to edges pointing to i. If we collect the scores in the matrix X and we define the matrix $\boldsymbol{B} = (b_{ij})$ as $b_{ij} = \frac{a_{ij}}{d_j}$, the previous equation can be rewritten as

$$X = B\,X. \tag{6}$$

Hence, X is an eigenvector of B associated to the eigenvalue $\lambda = 1$. Actually, $\lambda = 1$ is an eigenvalue of B because B is column-stochastic, that is to say, the elements of B are positive, and the elements on the same column sum to 1 for every column of B. The graph G is assumed to be strongly connected without dangling nodes.

In transportation, in general, graphs are undirected, simple and connected. A graph is undirected if there is an edge from node i to node j as soon as there is an edge from j to i, or, equivalently, the adjacency matrix A of G is symmetric. A graph is simple if there is no loop, that is to say, an edge from a node to itself, and there are no multiple edges. Finally, a graph is connected if, for every couple of nodes i, j, there exists a walk from the first to the second one and conversely, where a walk is a finite sequence of nodes and two consecutive nodes in the sequence form an edge. We remark that connectedness and strongly connectedness are the same if the graph is undirected.

When we use PageRank to rank the nodes of such a graph, d_j is nothing but the degree of node j, that is to say, the number of edges containing j because ingoing and outgoing edges coincide, and so it is easy to check that, up to transposition, $(d_1, \dots, d_n)$ is an eigenvector for B associated to $\lambda = 1$. Since the eigenspace $V(1)$ has dimension 1 because of the connectedness of G, we have that PageRank score is proportional to the degree of the node. Hence, PageRank is not suitable to rank the nodes of an undirected, simple and connected graph. A second reason for which PageRank needs adjustment for applications to transportation networks, is that one is interested in paths, that is to say, finite sequences of distinct nodes, more than in edges, when considering connectedness

properties of graphs. For the above two reasons, we modified PageRank so to get an index that is suitable for applications in transportation.

The main idea is to substitute paths to edges, and to construct a suitable column-stochastic matrix B that records the contribution of each path. Since graphs that arise in transportation can be large, the number of paths between two nodes can be huge, making unworkable the task of enumerating all of them. Then, we first fix the integer N, the maximum number of edges in a path. Secondly, we enumerate all paths from node i to node j of length $\leq N$, and we compute the contribution of each one of them according to the following two principles: a path is barely attractive if (1) its weight is big and (2) its weight is bigger than the minimum weight of a path between the same terminal nodes. Before giving equations that translate the principles above, we spend a few words about the principles themselves. In transportation, users choose their paths according to some principles, which generally are based on minimization of (generalized) costs. So, if the weight of a path is big in an absolute sense, only few users will choose that path. This remark justifies the first principle. Moreover, if a user must travel between two far nodes, then he/she will choose a path whose weight is as close as possible to the minimum weight. This leads to the comparison of the minimum weight and the actual one. Hence, we have to consider the second principle, too. We translate the principles above into the following equation

$$\text{path\_contribution } = f\left(\frac{w_{ij}}{w}\right)\frac{1}{w} \tag{7}$$

where w_{ij} is the minimum weight of a path from i to j and w is the weight of the considered path, that is to say, the sum of the weights of edges in the path. The ratio $\frac{1}{w}$ embodies the first principle. To represent the second one, we propose three different functions, according to the importance one wants to give to it. The functions are

$$(1)\ f_1\left(\frac{w_{ij}}{w}\right) = \frac{w_{ij}}{w}, (2)\ f_2\left(\frac{w_{ij}}{w}\right) = \frac{w_{ij}^2}{w^2}, (3)\ f_3\left(\frac{w_{ij}}{w}\right) = e^{-\frac{w^2}{w_{ij}^2}+1}. \tag{8}$$

With a little effort, it is possible to check that $f_1(x) \geq f_2(x) \geq f_3(x)$ for $x = w_{ij}/w \in (0, 1]$. The previous list is not exhaustive: we considered a wider set of possible functions, and we computed PathRank with them all. The three proposed functions have been selected for the following reasons: on tests, they produced mostly uniformly distributed values; the ranking of nodes with other functions was similar to the one we get with one of the three functions in equation (8); the comparison of the three functions is quite simple.

The choice between functions f_1, f_2, and f_3, has to be made according to the importance the second principle has in the application one has in mind. For example, in telecommunication networks, it is not so important if a path is longer than the minimum, because the signal speed is very high, and so f_1 is more suitable than the other proposed functions. On the contrary, when we consider a road network, shortest paths are considerably more appealing due to the highest costs associated with longer paths, therefore f_3 should be preferred. All three functions get the value 1 when $w = w_{ij}$, and tend to 0 when w increases indefinitely.

Thirdly, we determine b'_{ij} to be the sum over all paths from i to j of the contribution of every path, and so we get a symmetric matrix B'. We define PathRank at node j to be

$$PathRank(j) = \sum_{i=1}^{n} b'_{ij},$$

that is to say, it is the sum of the elements on the same column. Hence, larger is the sum of the contributions of the paths starting from a node, higher is the ranking of that node.

The relationship between PathRank and PageRank stays on the following result.

Theorem. Let B be the square matrix whose elements are $b_{ij} = \frac{b'_{ij}}{PathRank(j)}$. B is column-stochastic and $\big(PathRank(1), \dots, PathRank(n)\big)^T$ is eigenvector of B for the eigenvalue $\lambda = 1$.

Proof. The first claim is trivial because of the definition of PathRank. To check the second, let us compute the product between B and $\big(PathRank(1), \dots, PathRank(n)\big)^T$, a row at a time.

$$\sum_{j=1}^{n} b_{ij}\ PathRank(j) = \sum_{j=1}^{n} b'_{ij} = \sum_{i=1}^{n} b'_{ij} = PathRank(i)$$

because B' is a symmetric matrix. ■

As a further remark, our assumptions on the graph assure that the eigenspace $V(1)$ has dimension 1. The proof of this last claim is the same as in PageRank case, and so we do not report it. The interested reader can find the proof in (Bryan and Leise, 2006).

To clarify the way we compute matrix B', we consider the graph in Fig. 1. To make the computation simpler, we assume every edge has weight 1 that is to say, we only consider the topology of the graph and its adjacency matrix. Moreover, we set $N = 5$, and we select f_1.

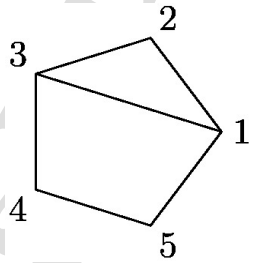


Figure 1. The graph above has 5 nodes and 6 edges. Moreover, it is evident that nodes 1 and 3 play the same role, as well as nodes 4 and 5. Hence, we expect that the values at 1, 3 and at 4, 5 are the same for every index.

In the graph above, the paths from node 2 to node 1 are $21, 231, 23451$. All three paths contribute to the value of b'_{12}, because their length is smaller than N. The three paths above have weights $1, 2, 4$, respectively, and so $b'_{21} = 1 + \frac{1}{4} + \frac{1}{16} = \frac{21}{16}$. Analogously, the paths from 3 to 1 are $31, 321, 3451$, those from 4 to 1 are $451, 431, 4321$, and finally the ones from 5 to 1 are $51, 5431, 54321$. Then,

$b'_{31} = 1 + \frac{1}{4} + \frac{1}{9} = \frac{49}{36}, b'_{41} = \frac{2}{4} + \frac{2}{4} + \frac{2}{9} = \frac{11}{9}, b'_{51} = 1 + \frac{1}{9} + \frac{1}{16} = \frac{169}{144}$.

The first column sum is $\text{PathRank}(1) = \frac{365}{72}$.

Paths to node 2 are $12, 132, 15432, 32, 312, 34512, 432, 4312, 4512, 45132, 512, 5132, 5432,$ and 54312, and so

$$b'_{12} = \frac{21}{16}, b'_{32} = \frac{21}{16}, b'_{42} = \frac{77}{72}, b'_{52} = \frac{77}{72}.$$

Their sum is $\text{PathRank}(2) = \frac{343}{72}$.

For symmetry, the paths to node 3 can be obtained from the one to node 1, and so the third column of B' is equal to the first one, up to swapping the first and third elements, and the fourth and fifth ones.

Paths to node 4 are $154, 134, 1234, 234, 23154, 2134, 2154, 34, 3154, 32154, 54, 5134, 51234$, and so

$$b'_{14} = \frac{11}{9}, b'_{24} = \frac{77}{72}, b'_{34} = \frac{169}{144}, b'_{54} = \frac{169}{144}.$$

Their sum is $\text{PathRank}(4) = \frac{167}{36}$.

Because of the symmetry of the graph, the fifth column of B' is equal to the forth one, up to swapping the first and third elements and the fourth and fifth ones.

In conclusion, nodes $1, 3$ are the most important, node 2 is the next one, while nodes $4, 5$ have the least score. We remark that the scores reflect the symmetries of the graph. Finally, PathRank is able to make differences between nodes $2, 4,$ and 5, while PageRank is not, since these three nodes have the same degree.

3.2 Icentr

In this section, we define another Index of CENTRality, Icentr for brief, designed to evaluate the performance of transportation networks, that takes into account both node weights and edge weights at the same time. To authors' knowledge, most centrality indices consider either edge weights, or node weights.

As before, we consider an undirected simple graph $G = (V, E)$ that represents the network we take into consideration, and we assume that both nodes and edges are weighted. To fix notation, the nodes are $1, 2, \dots, n,$ and the edges are $e_1, e_2, \dots, e_r$. We recall that an edge is a set containing two different nodes and two distinct edges contain at most a common node. Given an edge, each node in it is a neighbor of the other one. The notion of neighbor is at the core of the definition of Icentr. The weights of the nodes are $x_1, x_2, \dots, x_n$ while the ones of the edges are $w_1, w_2, \dots, w_r$, respectively. We remark that all node and edge weights are equal to 1 if the graph is unweighted, or it can happen that there exist x, w such that $x_i = x, w_j = w$ for every $i = 1, \dots, n$, and every $j = 1, \dots, r$. In such a case, nodes and edges are uniformly weighted. Of course, it is possible that the nodes are uniformly weighted and the edges are not so, or conversely, the edges are uniformly weighted and the nodes are not so. In general, neither nodes nor edges are uniformly weighted. Finally, we remark that one can use the values whatever centrality index assigns to each node as weights for nodes.

Now, we explain how to compute the value Icentr takes on a node. Since the value depends on the connected component containing the node only, we assume G to be connected. Let i_0 be the node we are interested in. By using an adapted Breadth-First Search algorithm, we divide nodes and edges in levels, as follows. i_0 is the only level 0 node. The level 1 nodes are the neighbors of i_0, and they are $i_{11}, i_{12}, \dots, i_{1n_1}$. The level 2 nodes are the neighbors of the level 1 nodes that are not in a previous level, and they are $i_{21}, i_{22}, \dots, i_{2n_2}$. In general, a node is in level h if it is neighbor of a

level $h-1$ node, and is not a neighbor of a node in level k for some $k < h-1$. Level h nodes are $i_{h1}, i_{h2}, \ldots, . \ i_{hn_h}$. Because of the previous construction, no node belongs to two different levels. Moreover, since we are working under the assumption that the graph G is connected, every node belongs to a level. To divide edges in levels, we remark that an edge $e = \{i, j\}$ connects two nodes either in different levels, or in the same level. In the first case, the nodes belong to consecutive levels and we set

$$lev(e) = \max\big(lev(i), lev(j)\big) \tag{9}$$

In the second case, we set

$$lev(e) = 1 + lev(i) = 1 + lev(j) \tag{10}$$

This choice of levels for the edges corresponds to the order in which they can appear in a path starting from i_0. In fact, if an edge connects two nodes in different levels and it is in a path from i_0, then it is at the place corresponding to the maximum level of a node in the edge. On the other hand, if an edge connects two nodes in the same level and it is in a path from i_0, then it is at the place corresponding to the next level with respect to the level of the nodes in the edge.

We recall that $w(e)$ is the weight of the edge e.

If e connects two nodes at different levels, let $x(e)$ be the weight of the node at the maximum level in e. Then, the contribution of e to the value Icentr takes at i_0 is

$$ic(e) = \frac{x(e)}{2^{lev(e)-1}}\, w(e) \tag{11}$$

If $e = \{i, j\}$ connects two nodes at the same level whose weights are x_i, x_j, respectively, the contribution of e to the value Icentr takes at i_0 is then

$$ic(e) = \frac{x_i + x_j}{2^{lev(e)}}\, w(e) \tag{12}$$

The final value is simply the sum of the partial contributions, and so we have

$$Icentr(i_0) = \sum_{j=1}^{r} ic\big(e_j\big) \tag{13}$$

Now, we compute the values of Icentr on the nodes of the graph in Figure 1, and we assume that the weights of nodes and edges are equal to 1, so to consider the topology of the graph, only. Of course, because of the symmetries of the graph, we have $Icentr(1) = Icentr(3), Icentr(4) = Icentr(5)$, and so, it is enough to compute $Icentr(1), Icentr(2), Icentr(4)$. When we start from node 1, the nodes are partitioned as follows: 1 is the only level 0 node, 2,3,5 are level 1 nodes, and 4 is the only level 2 node. Hence, edges $12, 13, 15$ are at level 1, and $23, 34, 45$ are at level 2. Then,

$$Icentr(1) = ic(12) + ic(13) + ic(15) + ic(23) + ic(54) + ic(34) = 1 + 1 + 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = 4.5.$$

When we start from node 2, the nodes are partitioned as follows: 2 is the only level 0 node, 1,3 are level 1 nodes, and 4,5 are level 2 nodes. Hence, edges $21, 23$ are at level 1, edges $13, 34, 15$ are at level 2, and finally, edge 45 is at level 3. Then,

$$Icentr(2) = ic(12) + ic(23) + ic(13) + ic(34) + ic(15) + ic(45) = 1 + 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{4} = 3.75.$$

When we start from node 4, the nodes are partitioned as follows: 4 is the only level 0 node, 3,5 are level 1 nodes, and 1,2 are level 2 nodes. Hence, edges $34, 45$ are at level 1, edges $13, 23, 15$ are at level 2 and edge 12 is at level 3. Then,

$$Icentr(4) = ic(34) + ic(45) + ic(13) + ic(23) + ic(15) + ic(12) = 1 + 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{4} = 3.75.$$

In conclusion, when we consider topology only, nodes 1,3 have the highest ranking, while nodes 2,4,5 have the same smallest ranking according to Icentr. We remark that Pathrank and Icentr rank the nodes of the graph G in Figure 1 in a different order, and so they are definitively different indices.

To end the section, we spend a few words on the principles behind the definition of Icentr. If we imagine to explore the graph starting from the node i_0, the partition of nodes in levels correspond to the minimum number of steps needed to reach another node, that is to say, if the node j is in level h, the shortest path, not necessarily the one with smallest weight, from i_0 to j contains h edges. A similar argument holds for edges. Then, we modify the weights of the edges according to the weights of the nodes in the edge and to the distance from i_0. We consider the distance from i_0 as a wide penalty, and the weight of the node as a positive attribute (see the discussion in Pathrank section on penalties and their levels). The partial contribution of a single edge states the above principles. From the discussion above, it follows that, if we assume that every node and every edge has weight 1, a node has a higher ranking if the edges of the graph are closer to the node, according to the levels we construct to explore the graph starting from the given node. We remark that the partition of nodes in levels is the same needed to construct a spanning tree, rooted at i_0, so to look at the graph from the point of view of the first node, in some sense. When computing a spanning tree, however, we do not consider edges connecting nodes in the same level. On the contrary, in the computation of Icentr, we consider all edges in the connected component containing the node i_0.

4. The dataset

In big cities, if available, the favorite means of transport is the metro, because of its higher speed if compared to surface transport (buses or tramways) or private cars. An underground train system is a complex physical network, whose simplified description reduces it to stations and rail tracks. Number of stations, length of rail tracks and fares provide easily measurable properties of such a network. Passengers per hour or total number of them by day or year, or also most required origin-destination stations are much more demanding measurable properties. In this study, we analyzed thirty-four underground networks of most known cities in the world. We associate an undirected graph $\mathrm{G} = (\mathrm{V}, \mathrm{E})$ to each network, where V is the set of nodes, and E is the set of edges. By definition, an edge is the set of the two nodes linked by the edge.

For our study, the graph we associate to a network has to represent the topological and functional properties of the network. Then, not every station of the network needs to be a node of the graph. In fact, a station is a node of G if it is either a transfer station or a terminal. In particular, 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 two lines have the same terminal station, we insert a fictitious node so to avoid the node having degree two. When considering weights, we assign the length of fictitious links equal to 1 kilometer.

As an example of the methodology we have adopted, we insert both a Boston metro network map, and the associated graph. In Boston metro network, there are 149 stations, but only 29 of them are nodes in the associated graph. Hence, about 20% is needed to represent the topology of the network.

As a general remark, from the graph it is not possible to recover the lines of the network. If this information is needed, one has to attach a label to each edge, so to distinguish between different lines.

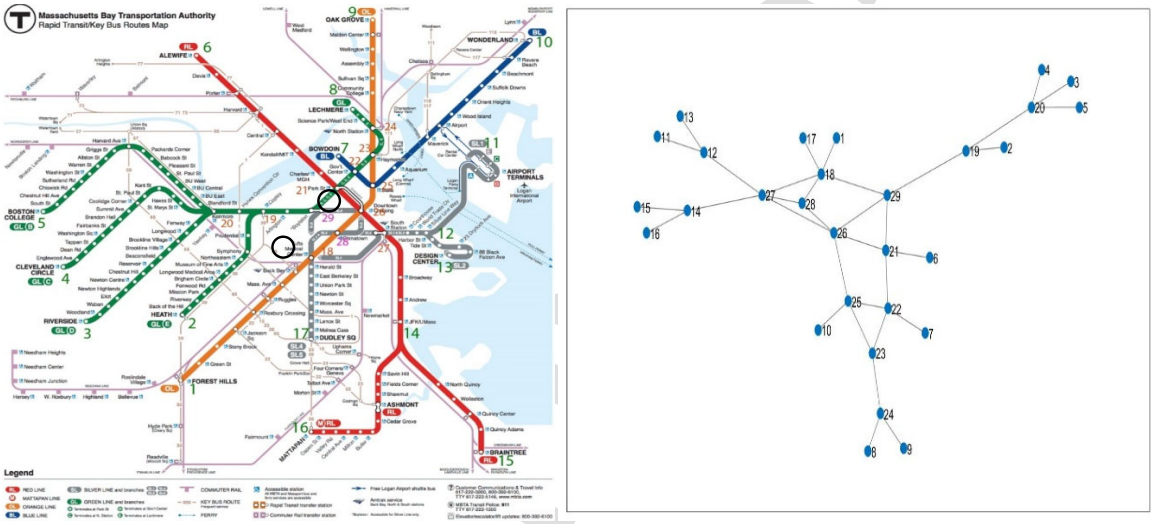


Figure 2: Boston metro map on the left, as found on the web (MBTA site). On the right, the associated graph: the nodes generally represent stations that are either terminal or junction. After a rotation of almost 180°, one can overlap the graph to the map.

We summarize some information on the networks and the associated graphs in the following Table 1.

Afterwards, for each graph, we assigned weights to edges in two different ways. The first weight is the number of stations on the edge, terminal stations included. The second weight is the distance measured along the tracks between the terminal stations. We do not analyze the weighted graphs in this paper, and we plan to report the analysis of the dataset of weighted graphs in a future work. More weights could be considered, so to analyze every graph from different perspectives. For example, the number of passengers per hour or by other aggregated interval (day, year), the origin-destination matrix and fare policy are very important information to analyze every transportation network under study.

Table 1: List of cities and number of nodes and edges of corresponding graphs.

Underground Network	Number of Nodes	Number of Edges	Total Length (km)	Underground Network	Number of Nodes	Number of Edges	Total Length (km)
Athens	12	14	85.8	Mexico City	42	62	195.8
Barcelona	36	53	131.2	Milan	21	24	91.3
Beijing	78	121	632.6	Montreal	11	12	61.6
Berlin	34	45	141.9	Moscow	65	106	393.1
Boston	29	34	114.7	New York	85	124	418.6
Brussels	9	10	35.9	Osaka	39	51	215.7
Bucharest	13	14	66.2	Paris	86	136	312.5
Buenos Aires	18	24	54.4	Prague	9	9	64.0
Cairo	12	14	88.8	Rome	10	10	62.7
Chicago	19	19	161.7	Saint Petersburg	17	20	117.9
Delhi	45	59	365.4	Seoul	119	194	1089.6
Hong Kong	35	39	205.7	Shanghai	88	142	683.5
Lisbon	13	15	40.6	Singapore	34	57	237.1
London	75	112	405.5	Stockholm	20	20	105.7
Lyon	10	10	27.7	Tokyo	65	110	279.0
Madrid	56	90	296.2	Toronto	12	12	75.9
Marseilles	7	7	21.1	Washington DC	26	32	218.2

5. Application and analysis of outcomes

In this section, we customize PathRank and Icentr for transportation applications, and we show how to use indices defined in section 3 in two different scenarios: the analysis of a single network, and the comparison of networks in a dataset. The first application, namely the analysis of a single network, is the most common when using centrality indices, and we include it so to show the potentialities of PathRank and Icentr. The second application, namely the comparison of different networks, is more challenging. We show that PathRank and Icentr can be used to cluster networks in the dataset, finding similarities between networks that can look very different at a first glance. E.g., the number of nodes and edges can be different. As a single network, we study the graph associated to Boston underground system in full details. The case study can be easily adapted to other transportation networks. As a dataset, we use the one described in section 4 that contains underground networks of worldwide cities. It is worth noting that this second case study can be easily adapted to a wide range of different datasets, too.

5.1 Customize the indices

Indices PathRank and Icentr, as defined, depend on choices and parameters to be settled according to the application.

PathRank

Let us consider PathRank first. Its definition depends on the maximum length N of paths, and on the choice of the function that embodies the second principle used to design the index (see eq. (8) in section 3.1).

Since graphs associated to underground networks are finite, also the maximum length of paths in them is finite. Hence, in principle, we can settle N large enough so that every path has length smaller or equal to N. In correspondence of such an integer, once we select a penalty function, we get asymptotic values for PathRank at each node, that is to say, even if we increase N, both the matrix B and its eigenvector associated to the eigenvalue $\lambda = 1$ do not change. For this reason, we call N_{asy} that value for N. From a computational point of view, for big graphs, N_{asy} can be quite large, and the total number of paths is so huge that it is not possible to actually compute them all in a reasonable time. It is then meaningful to estimate N in such a way that the ranking of the nodes is approximately the same as for N_{asy} even if the value of PathRank at each node is different from the asymptotic one. Of course, when the penalty function changes, the asymptotic values for PathRank change as well, and so the estimate for N_{asy} does. In the following Figures 3, 4 and 5, we highlight the differences between the asymptotic values and the values taken by PathRank at each node for maximum path length N when N grows, penalty function by penalty function, for Boston underground network.

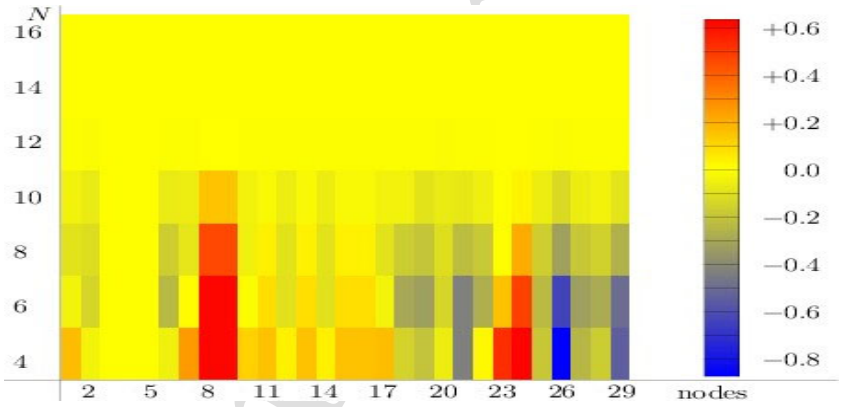


Figure 3: Differences between the asymptotic values and the ones for PathRank at each node of Boston graph, for N varying from 4 to 16, for the harmonic penalty function f_1.

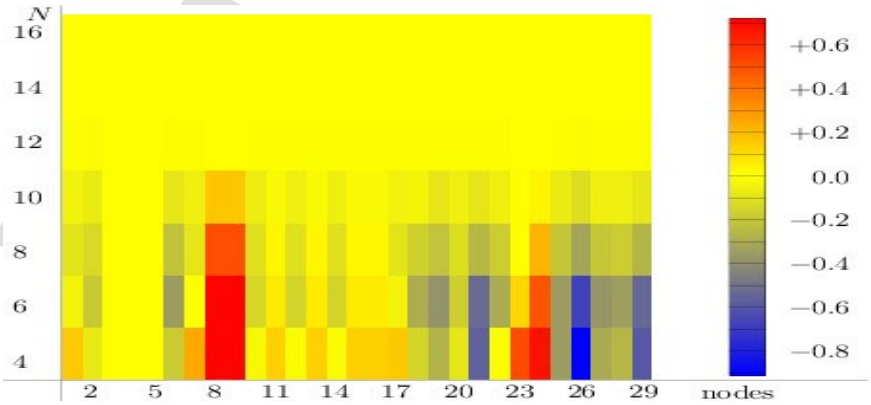


Figure 4: Differences between the asymptotic values and the ones for PathRank at each node of Boston graph, for N varying from 4 to 16, for the quadratic penalty function f_2.

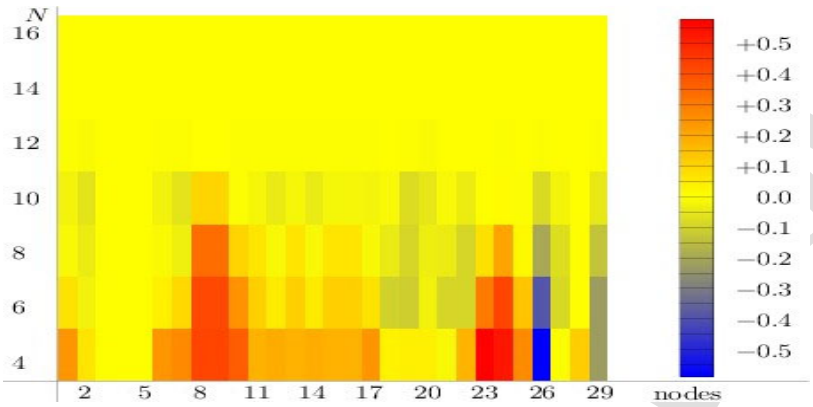


Figure 5: Differences between the asymptotic values and the ones for PathRank at each node of Boston graph, for N varying from 4 to 16, for the exponential penalty function f_3.

As those Figures 3, 4 and 5 show, no matter the penalty function we select, only a few nodes do not change their PathRank value on N, while nodes 8,9,21,23,24,26, 29 are the ones that get a value at the beginning that is very far from the final one. Nodes 8, 9, and 26 are the slowest to get their asymptotic value. On the other hand, the values at some nodes fluctuate very much. E.g., look at nodes 10, 12, 14, and 22. However, for $N \geq 12$, almost all nodes have their final ranking. Hence, the above Figures 3, 4 and 5 show that, even if $N_{asy} = 16$ for Boston graph, an acceptable estimate for the maximum path length is $N = 12$.

The main difference between the three penalty functions is that the differences between the values for $N \geq 4$ and the asymptotic ones are in a smaller range if we select f_3 for computing PathRank. This means that, independently from the value we fix as maximum path length, the error in using PathRank for N smaller than the asymptotic one is in general smaller than when using different penalty functions. This is the reason why we select f_3 in studying the dataset in next sections.

We explain the behavior of penalty function f_3 as follows. When the maximum path length grows, we can construct not only paths between nodes that are far in the graph, but also much more paths joining the same couple of near nodes. Since f_3 strongly penalizes paths longer than the minimum, the contribution of longer paths between the same couple of near nodes is smaller than for other penalty functions. It follows that, when we select the penalty function f_3, the eigenvector for $\lambda = 1$ of matrix B is barely influenced by the many longer paths between near nodes, and so its variation on N is small.

Of course, because of the same argument, f_1 and f_2 are of greater interest in applications in which path weight does not play an important role.

To finish the section, we make some further remarks on the maximum length N of paths. When we select a very small maximum path length, PathRank and Eigenvector Centrality give very similar results. On the other hand, the choice of N strongly affects the computation of all paths with length $\leq N$ in a graph. Hence, one has to choose N according to the computer on which PathRank runs, and to the size of the graph, where the size depends on the number of nodes, of edges, and of the topology of the graph itself. When analyzing the dataset, we settle $N = 12$, because of computational limitations on the computer on which the Matlab script to evaluate PathRank ran.

Finally, we remark that, when analyzing a single graph, one can choose the penalty function f according to the application, and then can compute PathRank values letting N vary, so to explore the ranking of nodes when the maximum path length grows. On the contrary, when analyzing a dataset, we suggest to settle both the penalty function f and the maximum length N to be the same for all graphs in the dataset, so to have PathRank values computed under uniform constrains.

As said, in next sections, we choose function f_3 as penalty, and $N = 12$ in analyzing the dataset.

Icentr

Now, we consider Icentr. This index depends on the weights of nodes and edges. If the graph is weighted, then it is natural to use the given weights. When the nodes of the graph do not have a weight, there are many different possible choices to assign weights to nodes. E.g, uniform weights, or the outcomes of a centrality index such as Degree, Betweenness, Closeness, or Eigenvector centrality, or also random weights, where the values are randomly chosen in a suitable range (in the case of Boston graph, $[0, 30]$ is a suitable range since there are 29 nodes). Of course, also PathRank can be used. When using a centrality index as weights for the nodes, Icentr can be considered as a second level centrality index. Different choices for the node weights produce different values for Icentr, as it can be easily expected and checked. As the following Figure 6 shows, we essentially get the same values for Icentr when we select one among Closeness, Degree and Uniform as node weights (some values are nil because of normalization).

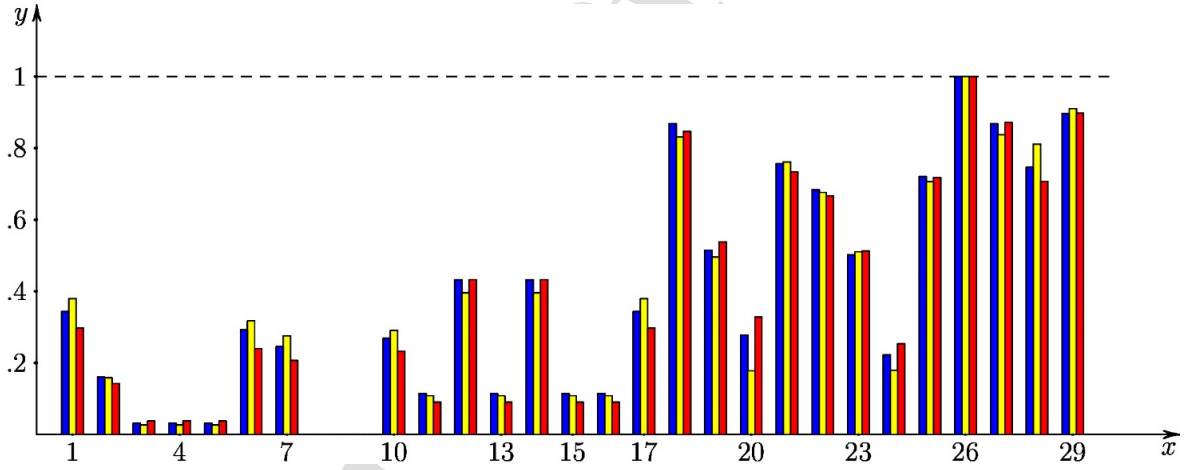


Figure 6: Icentr indices for Boston graph and three types of node weights. On the x-axis, the nodes of Boston graph, from 1 to 29. On the y-axis, the value of normalized Icentr. The blue bars on the left represent Icentr when we select Closeness as node weights, the yellow bars in the middle are obtained when selecting Degree, and finally, the red bars on the right draw Icentr for Uniform node weights.

By observing Figure 6, we can see that only nodes 1, 6, 7, 10, 17, 20 and 28 have considerably distinct index values; all other nodes have more or less the same values. Nevertheless, the highlighted dissimilarities have an impact on the ranking of some nodes. In particular, the seven nodes 1, 6, 7, 10, 17, 20, 24 change their ranking, as well as nodes 19, 23, and 21, 25, 28, and finally, nodes 18, 27, when we order the nodes according to Icentr computed with one of the above mentioned centrality indices.
Now, we compare Icentr values computed by selecting Eigenvector Centrality and Betweenness with the ones computed by selecting Closeness, taken as a representative of the previous selections.

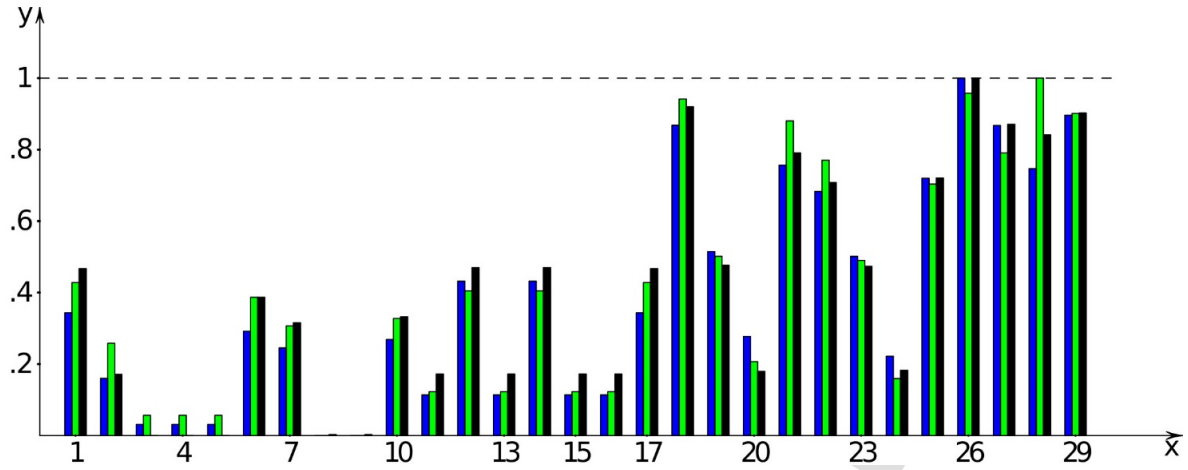


Figure 7: Icentr indices for Boston graph and other three types of node weights. On the x-axis, the nodes of Boston graph, from 1 to 29. On the y-axis, the value of normalized Icentr. The blue bars on the left represent Icentr when we select Closeness as node weights, the green bars in the middle are obtained when selecting Betweenness, and finally, the black bars on the right draw Icentr for Eigenvector centrality node weights.

It is evident that the values of Icentr when we select Eigenvector centrality or Betweenness, almost coincide at half of the nodes, while they both overlap to Closeness for only a few nodes.

Random node weights produce Icentr values that are very different from the others. However, if one is interested in analyzing the graph from a topological point of view, we suggest not to select it, because random weights do not always respect inner symmetries of the graph. E.g., in Boston graph, we can swap places of nodes 3, 4, 5 without changing the layout of the graph, so that we expect their values to be the same. For Random weights that does not happen. In general, it is not easy to determine all possible symmetries of the graph, and so it is hard to assign random weights to nodes that respect all inner symmetries. On the contrary, when we use weighted graphs, topological symmetries are not always respected by weights, and so random weights can be used in this setting. For example, if we weight nodes according to the number of passengers using that station, nodes 3, 4, 5 have no more the same weight. As suggested by the anonymous referee, we have averaged the value of Icentr at each node, v, of Boston graph, after computing it with node weights randomly extracted from the interval [0, b] sufficiently many times. The average value is approximately equal to the value of Icentr at v, when computed with uniform weights equal to b/2. The reason steams on the equalities

$$E\big(Icentr(v)\big) = \sum_{j=1}^{r} \frac{w(e_j)}{2^{lev(e_j)-1}}\, E\left(x(e_j)\right) = \sum_{j=1}^{r} \frac{w(e_j)}{2^{lev(e_j)-1}}\, \frac{b}{2},$$

and

$$E\left(x(e_j)\right) = E(x_h) = \frac{b}{2}$$

for every node v in the graph, where E is the expected value of a random variable.
From a topological standpoint, we believe that the ranking of nodes through Icentr with Eigenvector centrality is the best option among the various possibilities. In fact, the initial stripe contains all terminal nodes, with the only exceptions of nodes 7, 20, and 24. Moreover, node 28 has a higher ranking with respect to nodes 21, 22, and 25. Lastly, nodes 1, 17, and 12, 14 have values that are very close. Eigenvector centrality is the only centrality index that produces those results and is the only classical centrality index suitable for transportation applications. Furthermore, the values of Icentr with Eigenvector centrality can be easily divided in classes, with values smoothly varying.

For those reasons, in next sections, we use Eigenvector centrality when computing Icentr, if we have to select a unique centrality measure to get weights for nodes.
As remarked for PathRank, when considering a graph only, one can vary the centrality measure and compare the different outcomes of Icentr. On the contrary, when analyzing a dataset, we suggest to choose a centrality measure only, and to use it on the whole dataset, to avoid the analysis of the big mess of data produced by computing Icentr with all possible initial options on all graphs in the dataset.

5.2 Boston: a case study

Boston underground network consists of 114.7 km, divided in 3 heavy rail, 2 light rail and 1 bus rapid transit lines. In the network, there are 149 working stations plus 7 under construction. The map and the associated graph are Figure 2 in section 4. In it, we can distinguish a median axis (the path from node 3 to node 26), and two quite symmetric parts, that share the nodes 29 and 26 (an edge only) of the median axis.

The use of centrality indices aims at finding the most important nodes in the graph, from the point of view of each index. When comparing more indices, one can find nodes that are the most important in the graph, no matter the point of view one adopts. Hence, we first analyze the graphs by using PathRank, then by using Icentr. Finally, we compare results to draw conclusions on Boston underground, by considering PathRank, Icentr, and Eigenvector Centrality, a classical centrality index. As previously said at the beginning of section 3, Degree, Betweenness and Closeness are not suitable when analyzing underground networks.

To begin, we compare the asymptotic values that PathRank computes when using the three penalty functions. In fact, in the previous section, penalty function by penalty function, we compared the asymptotic values with the values for a smaller maximum path length, but we did not compare values obtained for different penalty functions. The node ranking changes accordingly to harmonic, quadratic or exponential function.

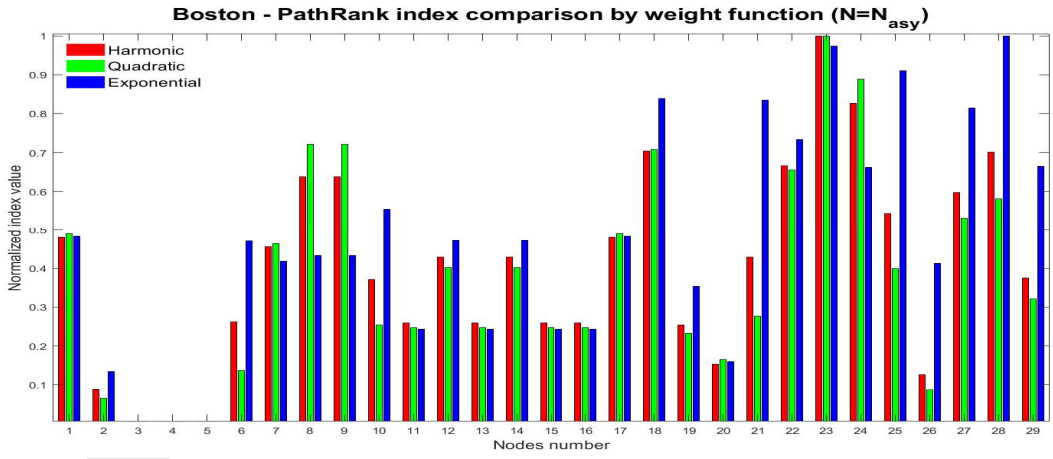


Figure 8: We report the asymptotic values obtained by PathRank. On the x-axis, we have the nodes from 1 to 29, on the y-axis, the normalized values. The harmonic values are in red on the left, the quadratic values are green in the middle, and finally, the exponential values are in blue on the right.

It is evident that the harmonic and quadratic values are very similar at more than half of the nodes, while the exponential values agree with the others for twelve nodes on 29. The ranking is definitely different between exponential and the two others.

Before showing computation results, we describe the methodology we are going to use in analyzing them. As a general remark, we expect that, whatever index suitable to examine one property, the index gets similar values on nodes that have similar features with respect to that property. Hence, it is important to collect together nodes that the chosen index likewise ranks. To make effective the sentence "likewise ranked", we order the values of the index, and compute the difference between consecutive scores. Then, we select the top values in the difference sequence, and collect nodes whose values fall between two consecutive top values. The number of top values to consider is arbitrary: both too few and too many of them will provide a poor clustering. We decided to consider top values in the difference sequence that are about three times the average value $jump_{av} = \frac{\max - min}{n-1}$ where max and min are the maximum and the minimum value the index gets on the nodes, and n is the number of nodes. As we use normalized values, $\max = 1$ and $\min = 0$ so that the numerator of the above ratio is equal to 1.

Now, we present the results of the ranking procedure according to PathRank in some tables. The symbol '—' in a table means that the difference between the nodes in subsequences separated by the symbol is bigger than the threshold.

According to the above procedure for 0.105 as threshold, when using the exponential penalty function, nodes in the top class are (Table 2):

Maximum path length	Nodes in the top classes, ordered in increasing ranking
4	18, 27, 21, 28, 29 — 26
8	29, 22, 21, 18, 27, 25, 23, 28
12	24, 29, 22, 27, 21, 18, 25, 23, 28
16	24, 29, 22, 27, 21, 18, 25, 23, 28

Table 2: Nodes in the top class for 0.105 threshold according to PathRank computed with the exponential penalty function.

It is evident that node 26, top node for $N = 4$, is no more in the top class for the other values of the maximum path length. Moreover, nodes 18 and 21 become more important than node 27 as N grows, while nodes 25, 23, and 28 are the most important for $N \geq 8$.

In the following Tables 3 and 4, we report the nodes in the top class when we use the two other penalty functions.

Maximun path length	Nodes in the top classes, ordered in increasing ranking
4	27, 18, 21, 28, 29, 26
8	10, 6, 26, 19, 12, 14, 7, 1, 17, 24, 29, 21, 25, 27, 22, 28, 18 -- 23
12	10, 29, 21, 12, 14, 7, 1, 17, 25, 27, 8, 9, 22, 28, 18 -- 24 -- 23
16	10, 29, 21, 12, 14, 7, 17, 1, 25, 27, 8, 9, 22, 28, 18 -- 24 -- 23

Table 3: Nodes in the top class for 0.105 threshold according to PathRank computed with the harmonic penalty function.

Node 26, top for $N = 4$, is in the second top class for $N = 8$, and does not belong to the top three classes for bigger N. On the contrary, the ranking of node 24 grows with the maximum path length.

Maximum path length	Nodes in the top classes, ordered in increasing ranking
4	22, 25, 27, 28, 21, 18, 29, 26
8	23
12	24 — 23
16	24 — 23

Table 4: Nodes in the top class for 0.105 threshold according to PathRank computed with the quadratic penalty function.

For the quadratic penalty function, we have differences bigger than the threshold only at the beginning and at the end of the node sequence, and so we get a poor information because almost all nodes belong to the same class. The eight top ranked nodes for quadratic penalty function and $N = 16$, are 27, 28, 22, 18, 8, 9 — 24 — 23, but nodes from 27 to 9 belong to a class different from 24, and they are not the only nodes in that class.

If we compare the results for the three penalty functions and the asymptotic maximum path length, we get that the most important nodes according to PathRank are 24, 22, 18, 23, and 28. It is worth noting that nodes 18 and 28 are linked by an edge, while nodes 22, 23, and 24 make a length 2 path. Nodes 18 and 28 are in the part of the graph opposite to the one that contains nodes 22, 23, and 24 with respect to the median axis. Instead, if we consider $N = 4$, node 26 is the most important in the graph for all penalty functions. Hence, the most important nodes change from one on the median axis for small N to five in different parts of the graph, in symmetric position with respect to the median axis, for N large.

Icentr is the second index we consider. As previously said, we compute the Icentr values at each node by using all but Random as node weights. In this case, too, the threshold to form classes is equal to 0.105, about three times the average difference between two consecutive values in the ordered Icentr sequence. We report the results in the following Table 5.

Node weights	Nodes in the top class, ordered in increasing ranking
Betweenness	25, 22, 27, 21, 29, 18, 26, 28
Closeness	22, 25, 28, 21, 27, 18, 29, 26
Degree	22, 25, 21, 28, 18, 27, 29, 26
Eigenvector Centrality	22, 25, 21, 28, 27, 29, 18, 26
Uniform	18, 27, 29, 26

Table 5: Nodes in the top class for Icentr index, computed with the specified centrality measures as node weights. For all choices but the last, two classes contain all nodes. For the Uniform node weights, the classes are three.

It is evident that the four top ranked nodes are the same for all node weights but Betweenness, even if their order changes according to the centrality measure. In Boston graph, those four nodes form a cycle containing node 28. The cycle has an edge on the median axis, and is on one side of it. If we consider the top ranked nodes without taking into account Uniform node weights, we restore the symmetry around the median axis.

Finally, we compare the results of ranking nodes by PathRank, Icentr and the classical centrality index Eigenvector centrality.

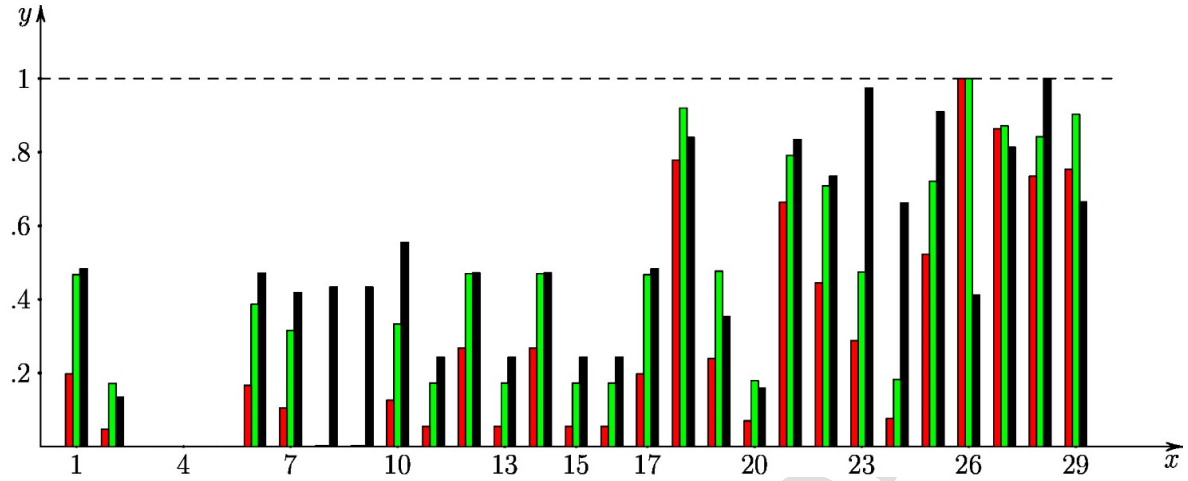


Figure 9: We report the values obtained by Eigenvector centrality (red), Icentr with Eigenvector centrality as node weights (green), and PathRank with f_3 as penalty function (black). On the x-axis, we have the nodes from 1 to 29, on the y-axis, the normalized values.

As the above histogram in Figure 9 shows, the values obtained by Eigenvector centrality and Icentr computed with Eigenvector centrality as node weights are generally different. This shows that Icentr depends not only from node (and edge) weights, but from the topology of graph. Moreover, the differences between PathRank and Eigenvector centrality are evident, even if both indices are based on the computation of eigenvectors. The two indices become more similar if we set the maximum path length to a small value. If we perform an analysis for Eigenvector centrality that looks like the one for the other two indices, we get that the nodes in the two top classes are: 21, 28, 29, 18, 27 in the second top class and 26 in the first one, while twenty-one nodes over twenty-nine belong to the bottom class. This shows that Eigenvector centrality is not able to partition the nodes according to functional similarities.

The top ranked node both for Icentr and for PathRank with asymptotic maximum path length is node 18 that is then the most important in the graph from the two index perspectives. It is in the second top class for Eigenvector centrality, too. If we consider Eigenvector centrality, Icentr and PathRank with maximum path length $N = 4$, node 26 is the most important in Boston network. On the map, node 18 corresponds to Tuft Medical Center and is one of the crossings of Orange and Silver Lines. Node 26 corresponds to Downtown Crossing and sits at the crossing of Red, Orange and Silver lines. The fact that they both sit at crossings of at most two lines is coherent with the methodology explained in section 4 to construct the graph associated to a metro network. Since we are performing a topological analysis of the graph, the results must be considered from that point of view.

5.3 Application to the whole dataset

In this subsection, we use PathRank and Icentr to cluster different networks according to similarities that emerge thanks to the evaluation of the two indices above. The main problem to face when looking for similarities is the fact that graphs associated to different networks in our dataset have different number of nodes and edges. E.g., the number of nodes ranges from 7 in Marseilles network to 119 in Seoul one. In literature, some techniques are available to cluster data, mainly time series, with different lengths. We have used Dynamic Time Warping (DTW, for brief) (Paliwal et al., 1982) for clustering networks in our dataset according to PathRank and Icentr (Figure 10). The

results have not been satisfactory, in our opinion, probably because our data are far from being time series. In particular, the number of nodes in the graph strongly affects the DTW distance between vectors associated to two graphs.

Other standard techniques have shown analogous limitations. Then, taking inspiration from signal processing, we used a different approach we now describe. Even if we refer to PathRank in our description, we use it also for Icentr.

At first, we computed PathRank values for each node for every graph in our dataset. Secondly, we normalized the computed values as follows. For every graph G, we computed the minimum value m_G and the maximum value M_G PathRank gets on the nodes in G. The normalized value at the node x is then $npr(x) = \frac{pathrank(x) - m_G}{M_G - m_G}$. Of course, the normalized values are in the range $[0, 1]$, and the extremal values are attained at the nodes where PathRank gets either the minimum or the maximum value. Thirdly, we order the normalized values, so to get an increasing vector of values from 0 to 1, for every graph in the dataset. Such an ordered vector can be plotted in a plane. The ordering of the vector is equivalent to permute the labels of the nodes, and so to transform the graph in another isomorphic one. Then, we mark equally spaced points on the x-axis, one for each node, in the interval $[0, \pi]$, and for each node, we plot the corresponding value of the index, so that on y-axis we measure amplitudes. It is then natural to transform the vector in a continuous function, symmetric with respect to the amplitude axis, by computing the n-th Fourier polynomial

$$F_m(x) = \frac{1}{2}\, a_0 + \sum_{k=1}^{m} a_k \cos(kx) \tag{14}$$

After a few experiments, we decided to set $m = 3$ so that at each normalized vector, no matter the number of nodes in the graph, we associate the four-tuple (a_0, a_1, a_2, a_3) where

$$a_i = \frac{1}{n-1} \sum_{k=2}^{n} \left(npr(k)\, cos\left(\frac{k-1}{n-1}\, i\, \pi\right) + npr(k-1)\, cos\left(\frac{k-2}{n-1}\, i\, \pi\right)\right). \tag{15}$$

The formula above is a discrete version of the integral $\frac{2}{\pi} \int_0^{\pi} f(x) \cos(ix)\, dx$ that computes the coefficient a_i when one wants to compute the Fourier polynomial of the function $f(x)$, symmetric with respect to the amplitude axis. In the discretization procedure, we use a trapezoidal quadrature formula. As last remark, for $i = 0$, the cosine is evaluated at 0 only, and so it does not contribute to the result.

Once we get the four-tuples associated to the normalized PathRank vector for every graph, we use a clustering algorithm. The squared distance between $A(a_0, a_1, a_2, a_3)$ and $A'(a'_0, a'_1, a'_2, a'_3)$ must be computed as

$$d^2(A, A') = \frac{1}{2}\, (a_0 - a'_0)^2 + (a_1 - a'_1)^2 + (a_2 - a'_2)^2 + (a_3 - a'_3)^2 \tag{16}$$

It represents, up to a scalar, the squared distance between two Fourier polynomials with the third order. Such a procedure allows us to get information on the shape that the graph of the vector of normalized PathRank values has (Figure 11). Furthermore, the comparison between different shapes becomes much easier because each one is codified by means of four coefficients. Results are quite interesting, as the following pictures (figures 12 up to 18) show.

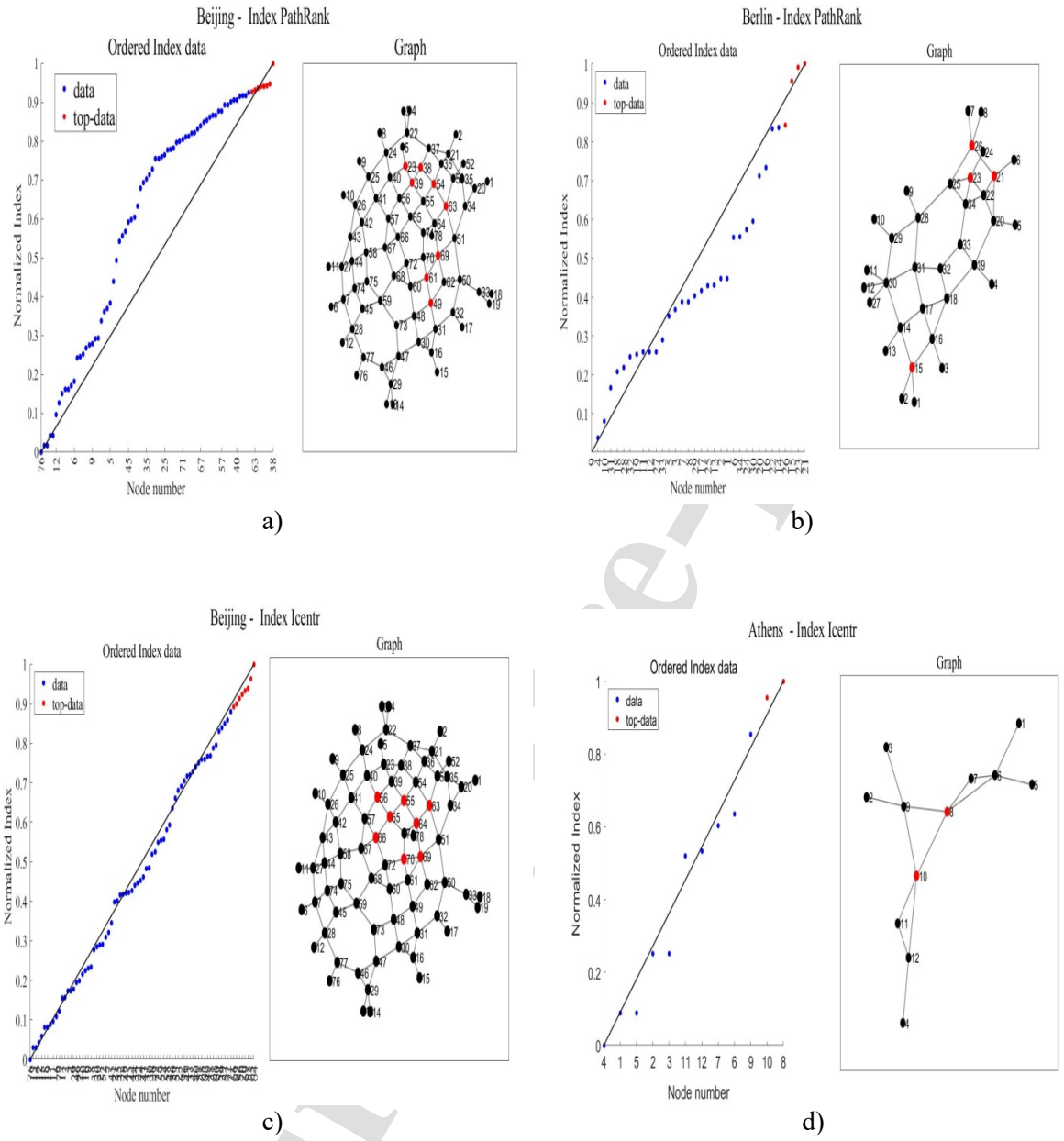


Figure 10: On the first line, the PathRank value graphs for (a) Beijing (left) and (b) Berlin (right). Those two networks are the closest according to DTW distance on PathRank. On the second line, the Icentr value graphs for (c) Beijing (left) and (d) Athens (right). Those two networks are the closest according to DTW distance on Icentr. On the x-axis, we report the node numbers.

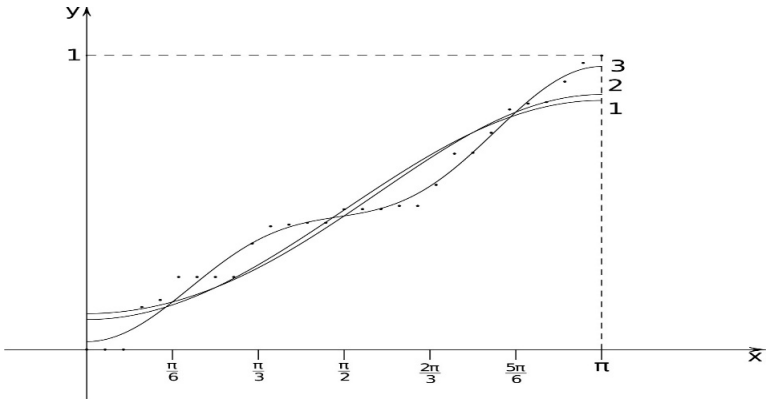
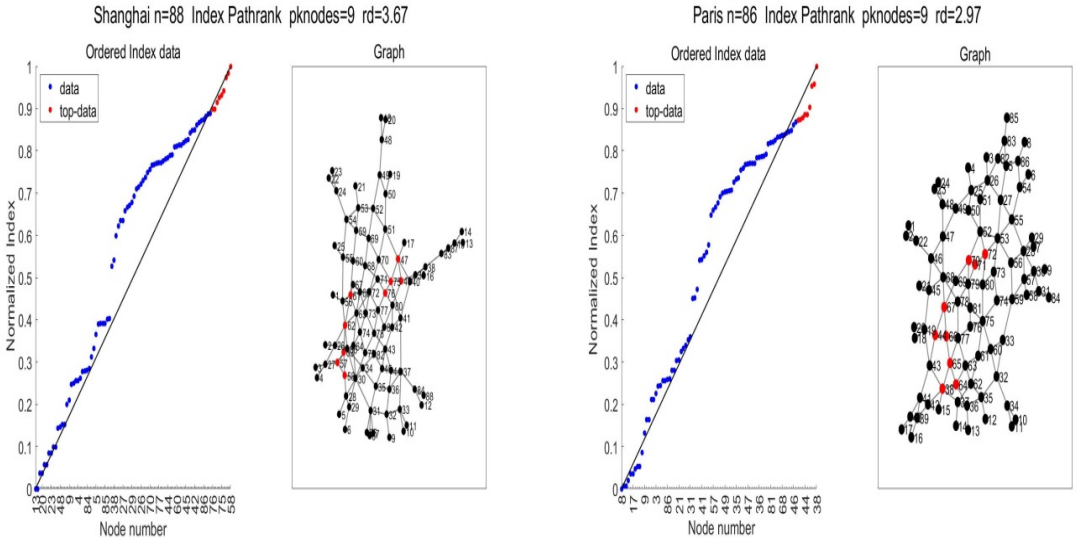


Figure 11. We plotted the ordered normalized PathRank vector of Boston graph, and drew the Fourier trigonometric polynomials with order m=1, 2, 3 to show how the approximation improves according to the order. On the x-axis, the nodes of the graph are equally spaced points in the interval $[0, \pi]$. On the y-axis, both Fourier trigonometric polynomials and the values of the normalized index mostly belong to the interval [0,1].



a) b)

Figure 12: Graphs of the two closest vectors of normalized PathRank values according to the third order Fourier polynomials. (a) Paris and (b) Shanghai network with 86 and 88 nodes, respectively.

The clustering algorithm starts from the two closest points, and substitutes their centroid to the points. Then, it repeats the computation of squared distances. If one of the two closest points is a centroid, the algorithm computes the new position of the centroid of a set of three or more points. At the end, we present the dendrogram of the results. In the following Figure 13, the results of clustering processes for PathRank.

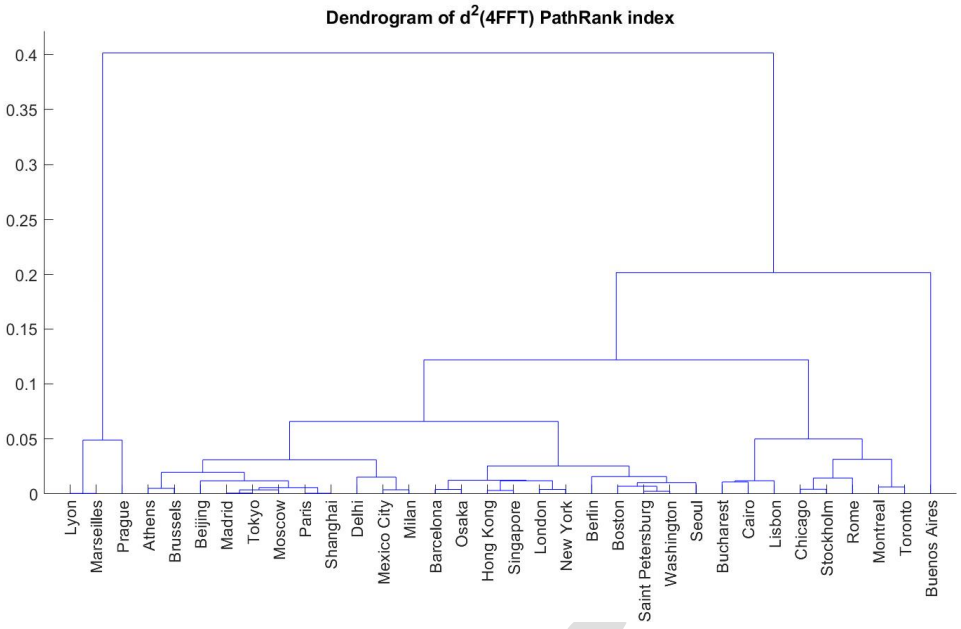


Figure 13: Dendrogram of the networks in the dataset for PathRank index.

For PathRank, when we fix a threshold equal to 0.06, we partition our dataset in 5 classes: the five classes contain 3, 11, 11, 8 and 1 networks, respectively. The number of nodes is not a discriminant to form the classes. For example, the second class contains Athens and Paris, while the third contains Saint Petersburg and Seoul. The first and forth classes are the most homogeneous with respect to the number of nodes of contained graphs, because they contain only small or medium networks. Buenos Aires graph shows itself to be dissimilar to every other.

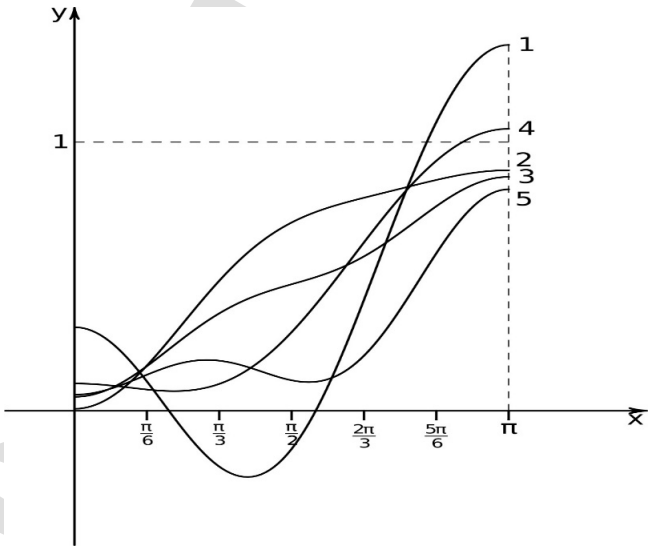


Figure 14. The five drawn curves are the ones corresponding to the five centroids of the classes in which we partition the database. The number corresponds to the class, from left to right in the dendrogram in Figure 13. On the x-axis, the interval $[0, \pi]$.

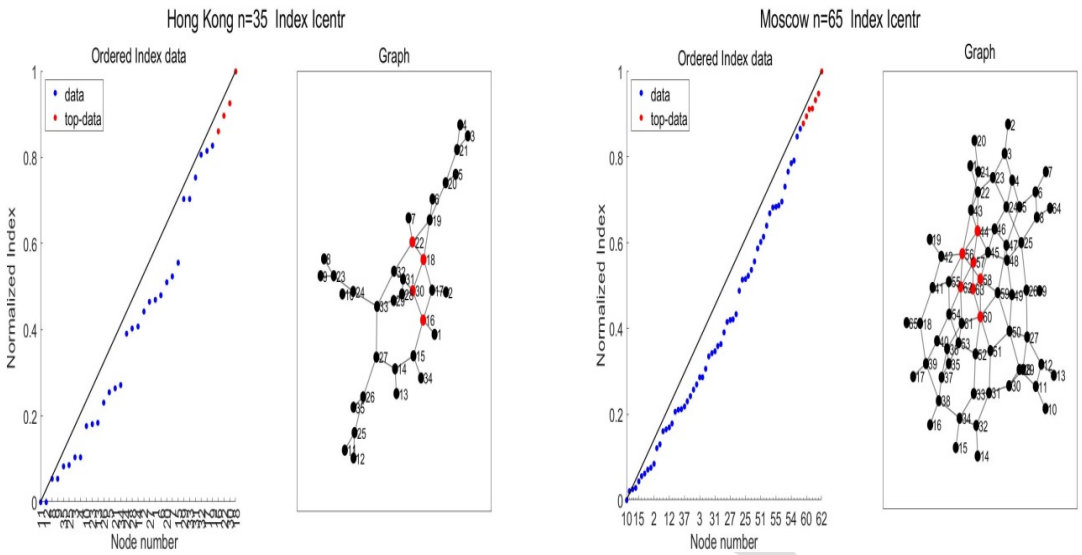
	a_0	a_1	a_2	a_3
1st class	0.693995	-0.55701	0.489510	0.031549
2nd class	1.151829	-0.39900	-0.124710	-0.044960
3rd class	0.932001	-0.34366	-0.004860	-0.065770
4th class	0.867820	-0.49156	0.141562	0.017480
5th class	0.555365	-0.26094	0.164330	-0.12135

Table 6. Fourier coefficients of the curves in Figure 13.

We do not report the four-tuples associated to each graph. However, graphs in the second class are the ones with biggest a_0, Buenos Aires has the minimum for a_0, and graphs in the first class are the ones with smaller a_0, Buenos Aires excluded. It is not easy to distinguish graphs in the third and fourth classes: as a general feature, a_1 coefficient is bigger for graphs in the third class than for graphs in the fourth class. It is possible to give a description of ordered normalized PathRank vectors associated to each centroid. The graphs in the first class have ordered normalized PathRank vectors close to square wave, that is to say, the values of normalized PathRank are mostly 0 or 1, and only a few in between the range. The graphs in the second class have ordered normalized PathRank vectors that, when plotted, are over the straight line connecting the first and last point in the sequence. The graphs in the third class are characterized as the ones whose ordered normalized PathRank vectors swing around the straight line connecting the first and last point in the sequence. The Boston graph fulfils this description, as shown in Figure 11. The graphs in the fourth class have ordered normalized PathRank vectors that, when plotted, are under the straight line connecting the first and last point in the sequence. The Buenos Aires graph has a shape in between the square wave and the graphs in the fourth class.

If we let the threshold grow, the second and the third class join first, then the fourth class joins to their union, then Buenos Aires, and lastly, the first class joins to the others. Hence, Buenos Aires and the graphs in the first class differ very much from all others, from the point of view of PathRank.

We satisfactorily apply the procedure above to analyze the dataset from the point of view of Icentr. In fact, the procedure is able to catch the shape of graphs of vectors of Icentr values, as shown in the following Figure 15.



a) b)

Figure 15: Graphs of the two closest vectors of normalized Icentr values according to the third order Fourier polynomials. (a) Hong Kong and (b) Moscow networks with 35 and 65 nodes, respectively.

Once we have the four-tuples associated to each graph of the dataset, we run a clustering algorithm as for PathRank, and we draw the resulting dendrogram (Figure 16).

For Icentr, when we fix a threshold at 0.04, we partition our dataset into 4 classes: they contain 15, 16, 2, and 1 network, respectively. For Icentr, too, the number of nodes is not a discriminant to form the classes. Moreover, when the threshold grows, the two big classes join first, then, the two small classes join. This shows that the three networks in the last two classes are different from the others, from Icentr point of view.

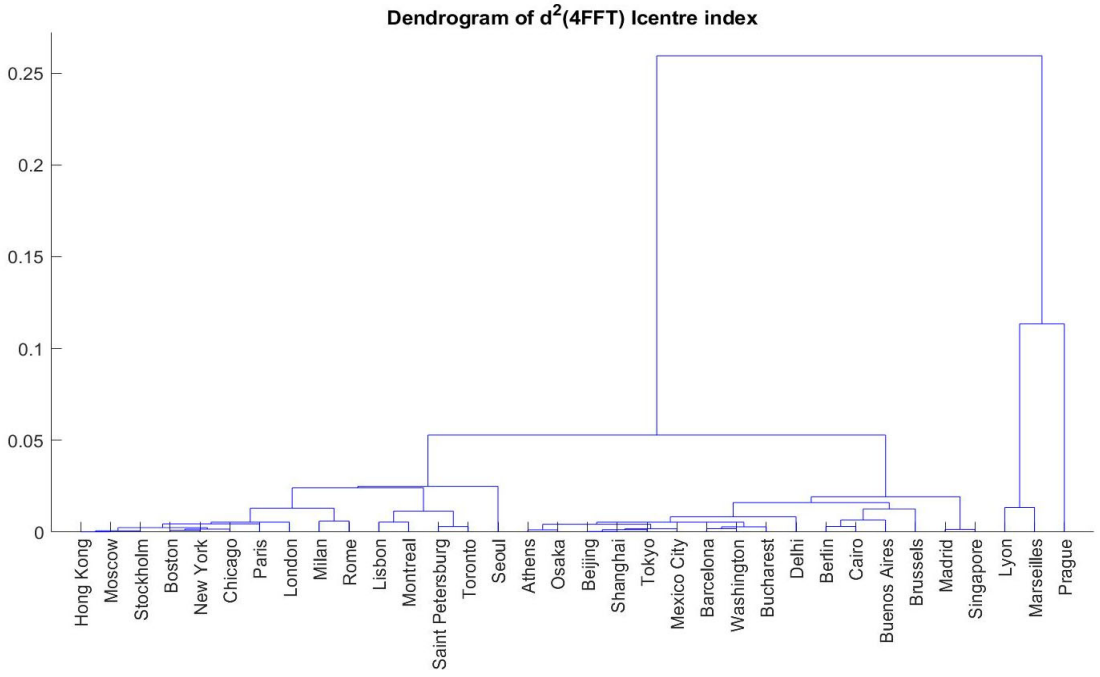


Figure 16: Dendrogram of the networks in the dataset for Icentr index.

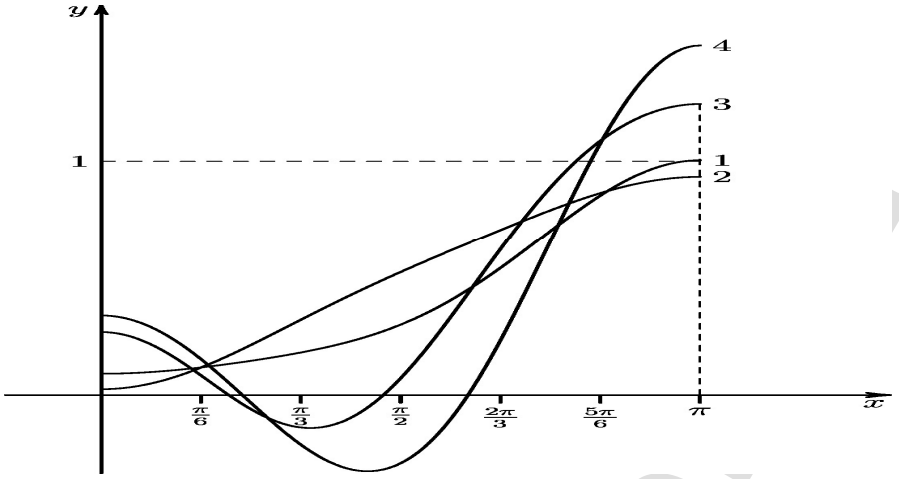


Figure 17. The four drawn curves are the ones corresponding to the four centroids of the classes in which we partition the database. The number corresponds to the class, from left to right in the dendrogram in Figure 16. On the x-axis, the interval $[0, \pi]$.

	a_0	a_1	a_2	a_3
1st class	0.847878	-0.42499	0.123730	-0.031800
2nd class	1.004507	-0.43145	-0.022800	-0.023140
3rd class	0.827827	-0.57917	0.342391	0.091571
4th class	0.625000	-0.53275	0.603553	-0.043890

Table 7. Fourier coefficients of the curves in Figure 16.

Now, we describe the four classes above.

We do not report the four-tuples associated to each graph. However, Prague has the minimum a_0 coefficient, while the graphs in the fourth class have the smallest a_1 coefficient (Table 7). To distinguish between graphs in the first two classes, it is enough to look at a_2 coefficient. In fact, such a coefficient is smaller for graphs in the second class.

The description of ordered normalized Icentr vectors for graphs in each class is like the one for PathRank vectors. We summarize the main characteristic of each class. Graphs in the first class have Icentr vectors that, when plotted (Figure 17), are under the straight line connecting the first and last point in the sequence, while vectors associated to graphs in the second class stay over the above line. Boston graph belongs to the first class, and so its Icentr vector has a shape like the one of vectors in Figure 15. Graphs in the last two classes have ordered normalized Icentr vectors that look like square waves, but, since the number of nodes that get 0 as value is different in percentage between graphs in the third and in the fourth class, the associated vectors do not stay in the same class. As last remark, for Icentr, no vector has a shape swinging around the straight line connecting the first and last point of the sequence.

The above considerations conclude the analysis of our dataset from the point of view of PathRank and Icentr, one at a time. We propose two methodologies to analyze the dataset from the two

perspectives at the same time. The first proposal consists in taking each class with respect to Icentr as a sub-dataset, and to filter it according to PathRank. At the end of the process, we get 20 classes at most associated to each couple (PathRank class, Icentr class). For example, no graph is in the (1,1) class, while Moscow, Paris, and Milan graphs are the only in the (2,1) class. Graphs in each class are homogeneous with respect to both indices. The second proposal consists in comparing the centroids of the two partitions, so to associate graphs with ordered normalized vectors having a similar shape with respect to both indices. Under this second analysis, we get that the first class for PathRank looks like the fourth one for Icentr, PathRank second and third classes looks like Icentr second class, and PathRank fourth class looks like Icentr first class. Finally, the fifth class for PathRank has no correspondence in Icentr classes.

6. Conclusions

In this paper, we define two novel centrality indices, PathRank and Icentr, to rank the nodes of graphs associated to underground networks. After a review of a few papers aiming at studying graphs associated to networks by means of suitable indices, we describe the methodology we adopt to associate a graph to an underground network. We have applied that methodology to 34 underground networks of worldwide cities, and the resulting graphs constitute our dataset. As explained in section 3, the first novel index, PathRank, is a generalization of PageRank algorithm, suitable to rank nodes in a transportation context. The second index, Icentr, ranks the nodes of the graph by means of weights of nodes and edges, scaled according to the distance from the considered node.

They both depend on the topology of the graph, since we consider paths. However, the dependence of PathRank on the topology is partially hidden from the computation of the contribution of each path to the final value.

An improvement of the enumeration of paths into a graph would allow researchers to get asymptotic values for PathRank in every graph, independently from its size. That would allow a comprehensive assessment of the network performance. As explained while studying Boston network, the computation of PathRank for bounded path length is however meaningful, since it is unlikely that metro passengers travel on long routes only. In a different direction, we have translated what we called the first principle, i.e. a long path is barely attractive, into the function $\frac{1}{w}$, while we gave different choices of functions representing the second principle. It would be interesting to compare PathRank values for different choices of functions representing the above first principle.

A possible upgrade of Icentr consists in bounding the number of explored levels when computing it. Nevertheless, it is not interesting to use Icentr outcomes as starting weight for nodes when computing Icentr, since this procedure converges to an asymptotic result very close to the Icentr values obtained at the beginning in 3 or 4 iterations, when one uses normalized Icentr values. Of course, it would be interesting to compare Icentr values for the weights we proposed, with the ones we get when using PathRank as a weighting index. We did not explore that choice because, at today, we are not able to compute the asymptotic PathRank values for all graphs.

After a very general definition of the indices, we have customized them in the first part of section 5. The two remaining parts of that section are devoted to a detailed study of Boston graph, and to the study of the dataset. In both studies, we used novel techniques, since the standard ones had given weak results when applied in that setting. In particular, the use of quantiles in the study of Boston

graph caused the loss of nodes with similar role with respect to the considered index. In the analysis of the dataset, we used Fourier trigonometric polynomials to investigate shapes of normalized vectors, no matter the considered index. Such a change in the perspective allowed us to perform a clustering in the dataset. We believe that both the use of large differences in an ordered vector and the use of Fourier trigonometric polynomials as approximating functions for ordered vectors with different lengths, can be fruitfully applied to study graphs when different indices are computed. As an outcome of the study of Boston graph from the topological point of view, we found the most important nodes in that underground network. Once more, we remark that our analysis is a model of what we mean with "study of a graph". Such study can be performed on other graphs associated to networks, even if not underground. Outcomes of our study on the dataset were the number of classes in which the dataset is partitioned for a given threshold, and their centroids. When adding a new graph to the dataset, one can quickly assign it to one of the classes by computing the squared distance from one of the centroids. Of course, once the new graph is added to a class, one ought to compute the new centroid of the class. A challenging problem is the following: the values of PathRank and Icentr for the graph of an underground network in our dataset produce Fourier trigonometric polynomials that are close to one of the computed centroids. We do not know whether the ordered values for the indices above for a network not in the dataset can be approximated by means of one of our centroids. The problem is equivalent to the robustness of our dataset from a statistical point of view, since it would be able to represent every other graph representing similar networks. In such a case, we could then use the dataset for defining a transportation graph in an abstract sense.

Future research in transportation will concern the extension of this approach to weighted graphs (namely edge length and number of stations per edge) and its use in the study of dynamic features of graphs when weighted by transportation variables such as passenger per hour per edge. In pure mathematics, instead, the investigation of properties of the indices will provide a deeper understanding of the dependence of the indices on the topology of graphs.

References

Bell, M.G.H., 2000. A game theory approach to measuring the performance reliability of transportation networks. Transportation Research Part B: Methodological 34, 533–545.

Brandes U., 2001. A Faster Algorithm for Betweenness Centrality, Journal of Mathematical Sociology 25(2):163-177.

Bryan K., Leise T. 2006, The $25,000,000,000 Eigenvector: The Linear Algebra behind Google, SIAM Review, Vol 48 No 3, pp. 569-581.

Cheng, Y.-Y., Lee, R.K.-W., Lim, E.-P., Zhu, F., 2015. Measuring Centralities for Transportation Networks Beyond Structures. pp. 23–39. https://doi.org/10.1007/978-3-319-19003-7_2

Crucitti P:, Latora V., Porta S., 2006. Centrality measures in spatial networks of urban streets, Phys. Rev. E, 73, 3, pp. 036125-5, https://doi.org/10.1103/PhysRevE.73.036125

Derrible, S.; Kennedy, C. 2010. The complexity and robustness of metro networks, Physica A, 389, pp. 3678-3691, https://doi.org/10.1016/j.physa.2010.04.008

Dimitrov, S.D.; Ceder, A. 2016. A method of examining the structure and topological properties of public-transport networks, Physica A, 451, pp. 373-387. https://dx.doi.org/10.1016 /j.physa.2016.01.060

Fei Xiong; Ximeng Wang; Shirui Pan; Hong Yang; Haishuai Wang; Chengqi Zang 2020. Social Recommendation with Evolutionary Opinion Dynamics, IEEE Transactions on Systems, Man and Cybernetics: Systems, Vol. 50, n. 10, pp. 3804-3816. https://dx.doi.org/10.1109 /TSMC.2018.2854000

Fei Xiong; Yu Zheng; Weiping Ding; Hao Wang; Xinyi Wang; Hongshu Chen 2021. Selection strategy in graph-based spreading dynamics with limited capacity, Future Generation Computer Systems 114, pp. 307-317.

Freeman L. C., 1979. 'Centrality in Social Networks: Conceptual clarification', Social Networks 1, 215-239.

Gu, Y., Fu, X., Liu, Z., Xu, X., Chen, A., 2020. Performance of transportation network under perturbations: Reliability, vulnerability, and resilience. Transportation Research Part E: Logistics and Transportation Review 133, 101809. https://doi.org/10.1016/j.tre.2019.11.003

Kumar, A., Haque, K., Mishra, S., Golias, M.M., 2019. Multi-criteria based approach to identify critical links in a transportation network. Case Studies on Transport Policy 7, 519–530. https://doi.org/10.1016/j.cstp.2019.07.006

Latora, V.; Marchiori, M. 2002. Is the Boston subway a small worls network? Physica A, 314, pp. 109-113.

Mussone, L., Vise, H., Notari, R. 2020. A topological analysis of underground network performance under disruptive events, ETC2020, Milan, pp. 1-25, AET 2020 and contributors.

Newman, M.E.J., 2001. Scientific collaboration networks. II. Shortest Paths, weighted networks, and centrality. Physical Review E, 64, 016132.

Opsahl, T., Agneessens, F., Skvoretz, J., 2010. Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks 32, 245–251. https://doi.org/10.1016/j.socnet.2010.03.006

Paliwal, K. K., Anant Agarwal, and Sarvajit S. Sinha. "A Modification over Sakoe and Chiba's Dynamic Time Warping Algorithm for Isolated Word Recognition." Signal Processing. Vol. 4, 1982, pp. 329–333.

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, 3, pp 215-227.

Tsiotas, D., Polyzos, S., 2015. Introducing a new centrality measure from the transportation network analysis in Greece. Annals of Operations Research 227, 93–117. https://doi.org/10.1007/s10479-013-1434-0

Wang, F., Antipova, A., Porta, S., 2011a. Street centrality and land use intensity in Baton Rouge, Louisiana. Journal of Transport Geography 19, 285–293. https://doi.org/10.1016/j.jtrangeo.2010.01.004

Wang, J., Mo, H., Wang, F., Jin, F., 2011b. Exploring the network structure and nodal centrality of China's air transport network: A complex network approach. Journal of Transport Geography 19, 712–721. https://doi.org/10.1016/j.jtrangeo.2010.08.012

Wang, Y., Cullinane, K., 2016. Determinants of port centrality in maritime container transportation. Transportation Research Part E: Logistics and Transportation Review 95, 326–340. https://doi.org/10.1016/j.tre.2016.04.002

Xingtan Wu; Hairong Dong; Chi Kong Tse; Ho, I.W.H.; Lau F.C.M,. 2018.

Analysis of metro network performance from a complex network perspective. Physica A, 492, pp. 553-563. https://dx.doi.org/10.1016/j.physa2017.08.074

Yu Heng., Yimin Wang, Peiyun Qiu, Jiacheng Chen, 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 (2019), https://doi.org/10.1051/matecconf/201927201031

Site

MBTA, https://www.mbta.com/maps, last visit 15-2-2021

Credit author statement

NOVEL Centrality measures and applications to underground networks

Lorenzo Mussone: Conceptualization, Methodology, Software, Writing, Reviewing and Editing,

Hiva Viseh: Data curation, Writing, Reviewing and Editing,

Roberto Notari: Conceptualization, Methodology, Software, Writing, Reviewing and Editing,

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: