UCLouvain

École polytechnique de Louvain

Passenger Flows, Backbone and Robustness in Subway Networks

Author: **Pierre Beltjens**
Supervisor: **Jean-Charles Delvenne**
Readers: **Eric Cornelis, Julien Hendrickx**
Academic year 2019–2020
Master [120] in Mathematical Engineering

Abstract

The importance of subway systems in urban areas is constantly increasing. Having a thorough understanding of them is essential to build and operate them properly.
In this work, we present a new way of computing passenger flows through those networks. This is done by using a multilayer modelling and a new centrality measure that is based on the amount of passengers that enter each station in the network. The obtained passenger flows are then compared with the ones that are obtained without the new modelling and measure.
Using the passenger flows through the network, the busiest stations and edges are identified. Lines are also compared based on the total amount of stops that are done on each one of them. The impact of modelling choices on these amounts of stops is also analyzed.
Two backbones of the Paris subway network are then extracted by combining multiple sparsification filters using passenger flows as weight. The first one is composed of edges and the second one is composed of lines. This second backbone shows that the lines in the Paris subway are split into two distinct groups.
A different approach on robustness based on passenger flows is then suggested that analyzes the impact of a line or station failure on the passenger flows in the network.
Finally, the impacts of the line 14 extension and the deconfinement plan in the Paris subway are analyzed.

I sincerely wish to thank Prof. Jean-Charles Delvenne for his time and his availability. Thank you for your ideas and suggestions as well as for the time spent reviewing this work.

I would also like to thank Prof. Julien Hendrickx and Prof. Eric Cornelis for agreeing to spend time reading my thesis.

Finally, I am grateful to all those that supported me during the writing of this document but also during the last 5 years.

Contents

Notations

$\mathbf{V}$: A set of vertices of size N

$\mathbf{D}$: A set of layers

$\mathbf{E}$: A set of edges of size M with $E \subseteq V \times V \times D$

$\mathbf{G(V,E,D)}$: A multilayer graph

$\mathbf{W}$: A set of weights

$\mathbf{G(V,E,D,W)}$: A weighted multilayer graph

$\mathbf{A}$: The $N \times N$ or $N \times N \times |D|$ adjacency matrix

$\mathbf{k_i}$: The degree of node i

$\mathbf{DC_i}$: The degree centrality of node i

$\mathbf{d_{ij}}$: The distance between node i and j

$\mathbf{Ecc_i}$: The eccentricity of node i

$\mathbf{CC_i}$: The closeness centrality of node i

$\sigma_{\mathbf{st}}$: The number of shortest paths from node s to node t

$\sigma_{\mathbf{st}}(\mathbf{i})$: The number of shortest paths from node s to node t through node i

$\sigma_{\mathbf{st}}(\mathbf{i}, \ell)$: The number of shortest paths from node s to node t through node i and layer ℓ

$\mathbf{BC_i}$: The betweenness centrality of node i

$\mathbf{BC_{ij}}$: The betweenness centrality of edge (i,j)

$\mathbf{IBC_i}$: The input betweenness centrality of node i

$\mathbf{IBC_{ij}}$: The input betweenness centrality of edge (i,j)

$\mathbf{IBCLine}_\ell$: The input betweenness centrality of line ℓ

$\alpha_{\mathbf{ij}}$: The significance score of an edge (i,j) according to a specific sparsification criterion

α : A fixed threshold

Introduction

In the last few years, there has been a turning point in mobility[1]. While the policy during most of the 20th century was to favor cars above all other transport modes, this is changing rapidly in the 21th century. The all-car doctrine is outdated especially in cities where road congestion is a daily occurrence. Some cities like Brussels[2] and Paris[3] have already started banning cars and are planning to restrict them even more in the close future.

To replace cars, other more sustainable types of transport are now encouraged. Amongst these, urban subway systems become more and more important amidst ever growing cities. They are supposed to substitute cars for most urban journeys and become a pillar in sustainable cities.

Contrary to some other urban transport networks like busses, subway systems are practically impossible to restructure once they are built[4]. Removing or reshaping parts of the network once it is built represents an immense added cost.
All possible details should thus be considered before the construction starts when building or expanding them.
It is thus essential to analyze existing subway systems not only to better understand and operate them but also to create better ones in the future.

In order to analyze subway systems, they must first be modelled and this can be done mathematically using networks and graph theory. Multiple types of representations as a network exist like the L-space representation in which a station becomes a node with edges to the neighboring stations[5], the P-space representation where a station becomes a node and edges exist between any 2 stations that belong to a same line[5] or the reduced L-space configuration which is similar to the L-space configuration but only includes termini and transfer stations[6].
In this work, a multilayer representation to fit the multiple lines in subway networks will be presented and used.

Many studies already analyze subway networks in Madrid[4], Paris[7], London[8] or Shanghai[9] and focus on the structural properties of the network.
In this work, the focus will be on the estimation of the passenger flows through the network and the applications of it. For this, a new way of computing passenger flows will be presented.

When planning new networks, estimating future passenger flows accurately is essential to avoid bottlenecks and line segments with a very low amount of passengers. In existing networks where unexpected failures can happen, being able to quickly estimate how the passengers will be redistributed over the network is also important. In the case of an extension of the network[10, 11], its impact should also be measurable just as for new networks.
A better understanding of the passenger flows through the subway network can lead to a better operation of it.

Multilayer networks, centrality measures and sparsification methods will be defined in chapter 1. Using a multilayer network and the centrality measures, a modelling and a new betweenness centrality measure will be suggested to estimate the passenger flows through the network in chapter 2. The subway network of Paris which will be used as main example throughout this work will also be introduced in chapter 2.
In chapter 3, centrality results and passenger flows in the Paris subway network will be presented. The impact of some modelling choices will also be analyzed.

Further applications will then be presented in chapter 4, 5 and 6.
Chapter 4 will present a way of defining the backbone of a subway network based on sparsification methods.
Chapter 5 analyzes robustness of a subway network in terms of passenger flows to line and station failures.
Finally, chapter 6 examines two specific situations in the Paris subway network: the 2020 extension of line 14 and the deconfinement plan of May 2020.

Chapter 1

Complex Networks

In this first chapter, some existing concepts and methods that will be useful during this report are presented.

1.1 Single and Multilayer Networks

1.1.1 Single Layer Networks

A graph (i.e. a single-layer network) is a tuple $G = (V, E)$, where V is the set of vertices and $E \subseteq V \times V$ is the set of edges that connect pairs of nodes. If there is an edge between a pair of vertices, then those vertices are adjacent to each other[12].

Vertices are also called nodes, there is no difference in meaning between both terms. The amount of vertices or nodes $|V|$ is also written as N.

The adjacency matrix of an unweighted and undirected graph is a matrix A such that:

$$\mathrm{A}_{ij} = \begin{cases} 1, & \text{if there exists an edge from vertex } i \text{ to vertex } j \\ 0, & \text{else} \end{cases} \tag{1.1}$$

The degree of a vertex in a single layer undirected network is defined as the amount of edges that are connected to the vertex or: $\mathrm{k}_i = \sum_j \mathrm{A}_{ij}$.

A weighted graph is a tuple G = (V, E, W) where W assigns a weight to each edge.

A shortest path between 2 nodes is a path that minimizes the sum of the weights of the edges in the path. In an unweighted graph, the shortest path is the one that minimizes the amount of edges in the path. A shortest path does not have to be unique.

1.1.2 Multilayer Networks

While single layer networks are already well known concepts, multilayer or multidimensional networks are more recent:
To accommodate the presence of more than one type of link, a multidimensional network is represented by a triple G=(V,E,D)[1], where D is a set of dimensions (or layers), each member of which is a different type of link, and E consists of triples (u,v,d) with u,v $\in$ V and d $\in$ D[13].

[1] Can also be weighted. Then it becomes G=(V,E,W,D).

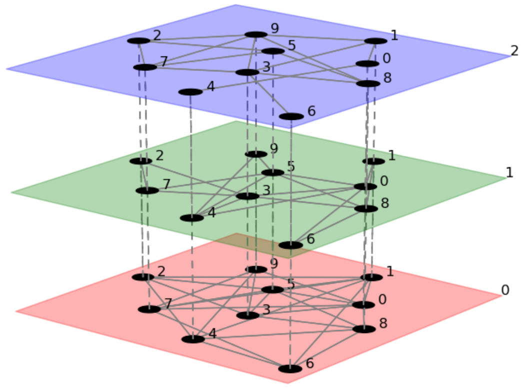


Figure 1.1: A multilayer network[14]

Multilayer networks are useful to represent networks that have different types of edges. Nodes can for example be representing cities while each layer represents the flights operated by different companies between those cities.
In a transportation network, nodes can represent stations while the different layers represent the different lines being operated in the network.

Instead of using an adjacency matrix like in common graphs, multilayer networks will use an adjacency tensor.
This will be a $|V| \times |V| \times |D|$ tensor A such that:

$$\mathrm{A}_{ij\ell} = \begin{cases} 1, & \text{if there exists an edge from vertex } i \text{ to vertex } j \text{ in layer } \ell \\ 0, & \text{else} \end{cases} \tag{1.2}$$

The degree of a node in a multilayer network can be defined in multiple ways but in this document it will always be defined as the total amount of edges that are connected to the node all layers combined. Mathematically the degree will thus be $\mathrm{k}_i = \sum_\ell \sum_j \mathrm{A}_{ij\ell}$

The shortest path between 2 nodes in a multilayer network can also be defined in multiple ways.
One of these ways is to flatten the multilayer network into a single layer and compute the shortest path in the single layer network. This has the advantage to be easy but it completely neglects the complexity of the multilayer network.
In this report, the methodology that is chosen is to allow the path to go from a certain edge e_1 that is incident to node n in layer ℓ_1 to another edge e_2 that is incident to node n in layer ℓ_2 but with a distance penalty (see chapter 2).
There will be a penalty at each change of layer. A path of 10 edges that goes through 3 different layers will thus be considered longer than a path of 10 edges that stays in the same layer.

Multilayer networks can be used to model multiple types of networks like: social networks[15, 16, 17], co-authorship networks[18, 19] or transportation networks[20, 21, 22].

1.2 Centrality Measures

Knowing which parts of the network are central and which ones are less central leads to a better understanding of the network and has many applications in different fields. This is why there exist centrality measures.
Unfortunately, there is no universal definition of what a central edge or a central node is. Depending on the situation and the application, the most central parts of the network could be completely different.
There thus exist multiple types of centrality measures that are each based on a different definition of centrality.

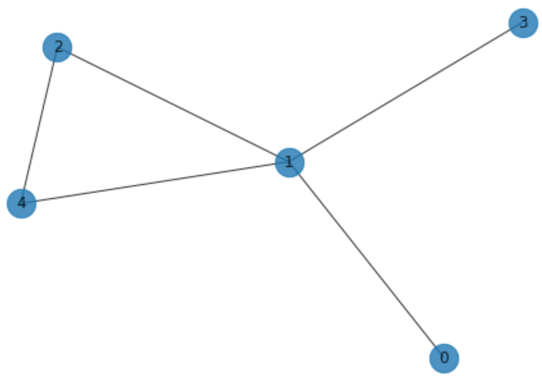


Figure 1.2: Undirected, unweighted, single layer network

1.2.1 Degree Centrality

The first centrality measure that is also the most basic one is the degree centrality.
The centrality of a node will in this case be its degree and the most central node will be the one with the highest degree.
In figure 1.2, the most central node will thus be node 1 with a centrality measure of 4 while the least central nodes will be node 3 and node 0 with a centrality measure of 1.

Node	0	1	2	3	4
Degree Centrality	1	4	2	1	2

Application : In a network where nodes are people and edges are friendships, a node with a high degree centrality represents a person with many friends.

1.2.2 Eccentricity

The eccentricity of a node in a graph is defined as the maximal distance towards any other node in the graph or:

$$\mathrm{Ecc}_i = \max_j d_{ij} \tag{1.3}$$

where d_{ij} is the distance from node i to node j.

The diameter of the graph is:

$$\mathrm{Diameter} = \max_i \mathrm{Ecc}_i = \max_{ij} d_{ij} \tag{1.4}$$

The diameter of the graph in figure 1.2 is thus 2 and the eccentricities are as follows:

Node	0	1	2	3	4
Eccentricity	2	1	2	2	2

1.2.3 Closeness Centrality

The closeness centrality measure of a node in a graph is defined in this report as the average distance towards the other nodes or:

$$\mathrm{CC}_i = \frac{\sum_j d_{ij}}{|V| - 1} \tag{1.5}$$

where d_{ij} is the distance from node i to node j.

In the graph of figure 1.2, the closeness centrality measure of node 1 is thus: $\frac{1+1+1+1}{4} = 1$, the closeness centrality of nodes 2 and 4 will be $\frac{1+1+2+2}{4} = 1.5$ while the one of nodes 0 and 3 will be $\frac{1+2+2+2}{4} = 1.75$.
The lower the centrality measure is, the more central the node is. In the example the most central node is thus node 1 again.

Node	0	1	2	3	4
Closeness Centrality	1.75	1	1.5	1.75	1.5

Application : In a subway network where nodes are stations and edges are connections between stations, living next to a station that has a low closeness centrality allows fast access to the rest of the city using the subway.

1.2.4 Betweenness Centrality

Betweenness centrality is a measure of the influence of a vertex over the flow of information between every pair of vertices under the assumption that information primarily flows over the shortest paths between them[23].
The betweenness centrality measure of a node is the amount of shortest paths between any 2 nodes that go through it or:

$$\mathrm{BC}_i = \sum_{s \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}} \tag{1.6}$$

with σ_{st} the total amount of shortest paths between s and t and $\sigma_{st}(i)$ the amount of shortest paths between s and t that go through i.
If there are multiple shortest paths between 2 nodes, we thus add the proportion of them that go through the node whose betweenness centrality measure is being computed.

Back to figure 1.2, the BC of node 1 will be 18 because there are 20 shortest paths in the graph and only 2 of them (2-4 and 4-2) do not use node 1. On the other hand, the BC of nodes 0, 2, 3 and 4 will only be 8 because only the shortest paths going to and from them will go through them.
The higher the BC measure, the more central the node is, node 1 is thus once again the most central node.

Node	0	1	2	3	4
Betweenness Centrality	8	18	8	8	8

We can notice that nodes 0,2,3 and 4 all have the same BC while nodes 2 and 4 have a higher CC and DC. This shows that different centrality measures might give different results.

Even though betweenness centrality is here defined on vertices, it can be defined in exactly the same way on edges as :

$$\mathrm{BC}_{ij} = \sum_{s \neq t} \frac{\sigma_{st}(i,j)}{\sigma_{st}} \tag{1.7}$$

Application : The betweenness centrality measure has applications in social networks[24, 25, 26], power grids[27, 28], routing[29] and many others.

1.3 Sparsification methods

(a) Full graph (b) Coarse-graining technique (c) Filter based technique

Figure 1.3: Different types of sparsification techniques

Sparsification methods are methods used on large networks with the aim to simplify them by extracting their backbone. An analysis can then be operated on the backbone instead of on the whole network which will reduce the complexity. Sparsification methods can also be used to discover which parts of the network are the most important according to specific criteria.

There are two types of sparsification methods[30]:

- The first type are called coarse-graining techniques[31, 32, 33]. *These techniques gather nodes sharing common properties into groups and then build a new graph considering each group as one single node. A major property of coarse-graining techniques is that they produce a rescaling, in the sense that the resulting network components are not at the same scale than the initial components: they no longer have the same meaning.[30]*
  As we can see on figure 1.3(b), the amount of nodes and the amount of edges is reduced.

- The other type of techniques are called filter-based methods[34, 35]. These methods create a backbone of the network by keeping only the most significant edges. Nodes are thus not added, removed or changed.
  As we can see on figure 1.3(c), the amount of nodes and their position does not change but only some of the edges are kept.

  Mathematically the problem is defined as finding a subgraph $G'(V, E', W')$ of $G(V, E, W)$ such that $|E'| \leq |E|$ and $\forall (ij) \in E' : \alpha_{ij} \geq \alpha$ with α_{ij} being the significance score of the edge (i, j) and α being the chosen threshold.

Both kinds of methods are used but the focus in this work is on the filter based methods that only keep the most important edges.

Once again, what edges are left in the backbone will depend on what criterion is used to determine the significance score of each edge in the network. There exist many types of filter based methods that sometimes give completely different results.
Depending on the application, a different technique will be considered.

1.3.1 Weight Threshold

A basic sparsification method that can be applied is the weight threshold [36, 37]. The method only consists in keeping the edges that have a weight higher than a predefined threshold[2]. The significance score α_{ij} of each edge will thus be w_{ij}.

The extracted part of the network G=(V,E,W) is thus G'=(V,E',W') where $|E'| \in |E|$ and $\forall (ij) \in E'$, $w_{ij} \geq \alpha$.

In a multilayer network, it is possible to apply different thresholds in each layer but in this report a single threshold will be applied globally.

1.3.2 Salience Filter

As will be pointed out later in this document, the weight threshold filter using betweenness centrality measures as weights tends to favour the edges that are in the center of the network since the significance scores will generally be higher in the center of the network than at its extremities. More details about this will be given in chapter 4.
To avoid this bias, a method that does not use the weight of the edges can be used and the Salience filter is one of them.

The Salience filter does the shortest path algorithm on each node consecutively. For each node n, the salience of a link s_{ij} is incremented by 1 if it is on a shortest path from n to another node. Whether it is only on one shortest path starting from n or on all of them, the salience of the link is only incremented by 1. The mathematics are as follows[38]:

σ is first defined[3]:

$$\sigma_{ij}(x,y) = \begin{cases} 1, & \text{if edge } (i,j) \text{ is on a shortest path between } x \text{ and } y \\ 0, & \text{else} \end{cases} \tag{1.8}$$

Then, Grady and al[38] define a matrix $\mathrm{T}(x)$ that defines all edges that are on a shortest path starting from x:

$$\mathrm{T}_{ij}(x) = \begin{cases} 1, & \text{if } \sum_y \sigma_{ij}(x,y) > 0 \\ 0, & \text{else} \end{cases} \tag{1.9}$$

Finally, the salience of each link is defined as the amount of matrices T(x) in which it is present and thus:

$$\mathrm{s}_{ij} = \sum_x \mathrm{T}_{ij}(x) \tag{1.10}$$

These values can be normalized to become:

$$\mathrm{s}_{ij} = \frac{1}{|V|} \sum_x \mathrm{T}_{ij}(x) \tag{1.11}$$

[2]The method only makes sense in a weighted network
[3]σ is defined slightly differently to how it is defined in the BC formula where (i,j) and (x,y) are inverted.

As an example of how the Salience filter acts differently than the weight threshold with the BC as weight, we look at the significance scores that each edge will have under both filters in the following graph inspired by[30]:

Figure 1.4: Salience filter vs weight threshold

Edge	(0,1)	(1,2)	(2,3)	(3,4)	(4,5)
BC weight threshold	10	16	18	16	10
Normalized salience	1	1	1	1	1

While the most important edges according to the weight threshold are the central ones, the Salience filter gives an equal importance to all edges in this example.

Intuitively, the salience of an edge can be seen as a measure of how important the edge is for the paths that go through it. As will be seen in chapter 4, if the paths that go through a certain edge can easily be changed slightly to not go through that edge anymore, the salience of the edge will be low. On the other hand, if the paths that go through a certain edge need to be modified significantly to avoid the edge, the salience will be high.
In the example, there are no alternatives for any of the edges which is why the salience is maximal for all of them.

1.3.3 Disparity Filter

The third type of filter that will be used is the disparity filter[30, 35]. This filter uses the weights of the edges like the weight threshold filter but in a different way.
While the weight threshold simply keeps the edges that have the highest weights, the disparity filter will favour edges that have a high weight compared to its incident edges. An edge can have a high weight but if all the edges around it have even higher weights, it will still not be extracted while an edge with a low weight could be extracted if its incident edges have even lower weights.
The filter thus aims at extracting the edges that have a high local weight.

For each node i, the incident edges (i, j) will get a relative weight

$$p_{ij} = \frac{w_{ij}}{\sum_k w_{ik}} \tag{1.12}$$

p_{ij} is thus only equal to p_{ji} if $\sum_k w_{ik} = \sum_k w_{jk}$!

These values will then be compared to a null hypothesis which states that the p_{ij}^{NH} corresponding to a certain node i of degree k_i are produced by random assignments from a uniform distribution.

β_{ij} can then be defined as the probability that the p_{ij} in the real network is higher than the one in the null-hypothesis or $\beta_{ij} = P(p_{ij} \geq p_{ij}^{NH})$ thus:

$$\beta_{ij} = 1 - (1 - p_{ij})^{k_i - 1}$$

Since p_{ij} is often not equal to p_{ji} and k_i is often not equal to k_j, β_{ij} and β_{ji} will also be different in most scenarios.
Since every edge can only have one significance score, there are 3 possibilities when using the filter. The one that is chosen depends on the application:

- The significance score α_{ij} is $max(\beta_{ij}, \beta_{ji})$
- The significance score α_{ij} is $min(\beta_{ij}, \beta_{ji})$
- The significance score α_{ij} is $avg(\beta_{ij}, \beta_{ji})$ or $(\beta_{ij} + \beta_{ji})/2$

Whichever option is taken, the filter extracts the edges (i, j) of which the maximum, the minimum or the average of β_{ij} and β_{ji} is higher than a certain threshold.

For example, if we use the first option, the extracted network will be G'(V,E',W') with

$$\forall (i, j) \in E' : max(\beta_{ij}, \beta_{ji}) = \alpha_{ij} \geq \alpha$$

With this filter, an edge that has a high weight compared to its incident edges will get a high relative weight. The probability that this weight is higher than what it would be with a random assignment will thus be high and the significance score of the edge will thus be high as well.

Chapter 2

Description and Modelling

While the modelling approach that will be presented in this chapter is not dependent on one particular city or one particular transportation network, it will be implemented mostly on the subway network in Paris. It is thus worthwhile to start with a quick introduction on that particular network.

2.1 Transportation in Paris

Transportation in Paris is centered around its subway network which is over a 100 years old and is still growing today. With 16 lines, 302 stations, around 1.5 billion yearly passengers and over 200 kilometers of tracks, it is one of the biggest and busiest subway networks in the world.

Compared to other cities, the Paris subway is special by its high density of stations. The choice was made originally to have stations that would be close to each other. Today, the average distance between stations is 562 meters which is less than in other subway networks like Madrid that for the same amount of stations has 288km of lines instead of 214km[39].

2.1.1 History

Construction of the Paris metro started around 1900 which makes it one of the oldest networks in the world. Since then, subway lines have been continually added while existing lines were also being extended.

Starting in the 1960's and due to the subway technology being outdated and its continued expansion being complicated, the construction of the RER[1] began. The purpose of the RER being to better serve the suburbs of the city where more and more people started to live. The RER lines C,D and E are operated by SNCF which is the french public train operator while the 2 other lines A and B are partially operated by the SNCF and partially operated by the RATP[2] which is the subway operator in Paris.

While the tramway in Paris had been sacrificed for the subway and cars in the middle of the 20th century, it made its reappearance in the 90's. The lines were constructed in the close suburbs where there are less passengers than in the center of Paris.

In 1998 subway line 14 was introduced which was the first fully automated line in the network. Later, in 2013, subway line 1 also became fully automated and works are currently being done on subway line 4 to automate it as well.

[1] French: "Réseau Express Régional" or Regional Express Network
[2] French: "Régie autonome des transports parisiens"

Automated lines are more reliable because they are less dependent on strikes for example. Human mistakes being out of the way, the security distance that is kept between 2 vehicles can also be smaller and thus the frequency of the line can be higher. Line 14 for example has a frequency of one vehicle every 85 seconds at peak hour.

As of 2020, there are 14 main subway lines from 1 to 14 and 2 small lines 3bis and 7bis for a total of 16 lines. There are also 5 RER lines that connect Paris to its close and further suburbs, 11 tramway lines in the suburbs of Paris, train lines that go towards Paris from all around the region, but also from the whole of France and hundreds of bus lines in and around Paris.

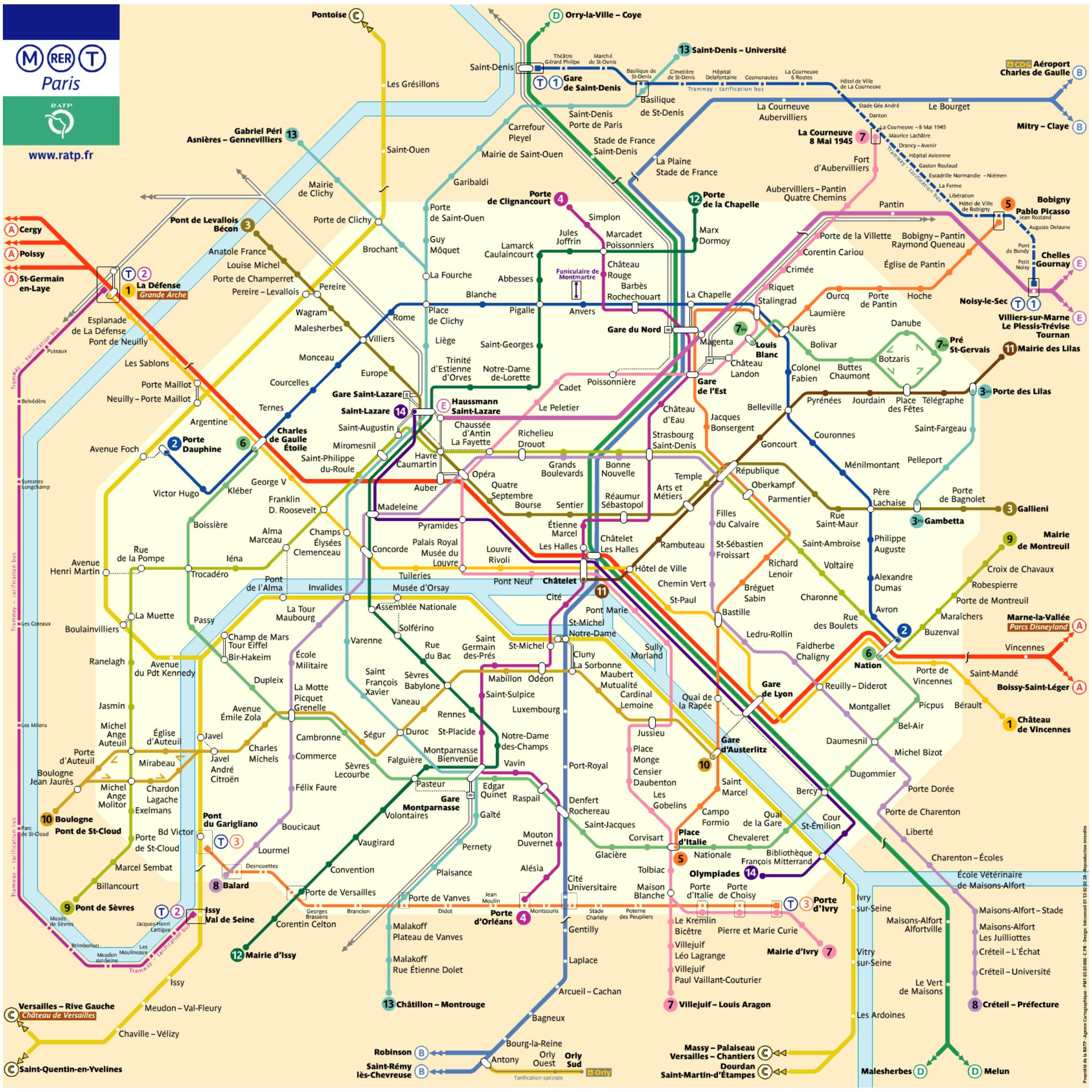


Figure 2.1: Paris subway[40]

2.1.2 Organization

A journey in the region of Paris often relies on 2 steps. The first step is to take a train, a tram or a RER from the close or further suburbs towards Paris and the second step is then to step over to the subway to go to the desired location.
It is possible for inhabitants of Paris to make journeys by using only the subway but for most people working in Paris or travelling to Paris, another type of transport has to be used first. The main Paris airport[3] for example is only accessible by train or RER.
The fact that many journeys are organized in a 2-step way is important because it creates big transfer hubs between the subway and the other transport modes. It is thus essential to keep this fact in mind while analyzing the Paris subway.

Even though there are also many busses that drive around the center of Paris, they do not attract that many passengers[4]. They are used either to get to a subway station or to go straight to the desired destination.

2.2 Modelling as a Complex Network

Modelling a transportation system as a complex network is rather straightforward. The basic modelling is the one where the mapping of the network is done in the L-space[5]. Each station is a node/vertex and there exists an edge between tho vertices if the two stations that the vertices represent are subsequent stations on a subway line.
A transportation system consisting of one line with six stations will thus be modelled as a network with six nodes and five edges:

Figure 2.2: 1 line with basic model

Two different improvements are suggested to model the network more realistically: multiple layers and using the load of the network.

[3] Paris Charles-de-Gaulle
[4] 350 Million users annually compared to 1.5 billion for the subway

2.2.1 Load of the Network

Most studies that analyze Network properties of subway systems whether it is Paris[6, 7, 41, 42], Madrid[4, 5, 6] or London[6, 8, 42] do not focus on the load of the network.

To compute most properties in a network, the load is indeed not necessary, but to keep track of passenger flows, the load of the network should be considered in its modelling.

We define the input of a station as the amount of passengers that enter a specific station annually. Passengers doing a transfer at that station are not counted.
In all transport networks, some stations will have a high passenger input while others will have a low passenger input and Paris is no exception[5]. Due to a journey often starting with a train or a RER and then moving forward with a subway, big hubs exist that represent important entry points into the subway system.
Similarly to the input, the output is defined as the amount of passengers that exit the network at a specific station annually.

To analyze the flow of passengers through a network, it is chosen to use betweenness centrality measures.
With the current definition of betweenness centrality defined in chapter 1:

$$\mathrm{BC}_i = \sum_{s \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}},$$

a station is assigned a betweenness centrality measure depending on how many shortest paths the station lays on with no distinction being made if these shortest paths are frequently used or not.
In the city of Paris for example, this means that for a station, being on the shortest path between 'Pelleport' and 'Volontaires' that are two small stations where less than 3 million passengers take the subway annually will contribute to the betweenness centrality as much as being on the shortest path between 'Gare du Nord' and 'Saint Lazare' which are both stations with an input of over 40 million passengers.

Passenger Betweenness Centrality

To overcome this problem, a new centrality measure should be defined. The passenger betweenness centrality[9] or PBC can be defined in the following way with $\mathrm{NbPass}(s \to t)$ being the amount of passengers going from s to t:

$$\mathrm{PBC}_i = \sum_{s \neq t} \frac{\sigma_{st}(i)\mathrm{NbPass}(s \to t)}{\sigma_{st}} \tag{2.1}$$

For each combination of stations s and t, the passenger betweenness centrality of node i will be increased by the proportion of shortest paths between s and t that go through i multiplied by the amount of passengers going from s to t.
Assuming passengers always use the shortest path, the passenger betweenness centrality of i will be the passenger flow through i.

The main problem with this measure is that it requires the data of travelling passengers between any 2 stations in the network to be available which it rarely is because it either simply does not exist or cannot be published due to confidentiality issues. In order to use this measure on most transportation networks like the Paris subway network, it thus has to be approximated.

[5] The standard deviation of the input is 7.5 million in the London subway compared to 9 million in the Paris subway for a similar mean of 6.3 million in London and 5.8 million in Paris

To do the approximation, we start by rewriting equation 2.1:

$$\text{PBC}_i = \sum_{s \neq t} \frac{\sigma_{st}(i)\text{Input(s)}P(s \to t|s)}{\sigma_{st}} \tag{2.2}$$

with $P(s \to t|s)$ the probability that a passenger entering the network in s, exits it in t.
Among passengers entering the network at station s, we then assume, as a simple estimate, that the fraction leaving towards station t is given by the proportion of passengers that exit the network at station t, or output(t), out of all the passengers except the ones exiting at s or:

$$P(s \to t|s) \approx \frac{\text{Output(t)}}{\text{TotOutput-Output(s)}}$$

The output of s is subtracted from the total output because a passenger will never start and end his journey in the network at the same station.

The total output and total input of the network is the same. Assuming that the output of a station is equal to its input as well, we get:

$$P(s \to t|s) \approx \frac{\text{Input(t)}}{\text{TotInput-Input(s)}} \tag{2.3}$$

This assumption makes sense in a subway network where a journey going from A to B is often followed by a journey back from B to A. The input and output of stations will thus have similar values.

Further work could try to approximate $P(s \to t|s)$ more accurately by taking factors as proximity into consideration.

Input Betweenness Centrality

Using equations 2.2 and 2.3, the passenger betweenness centrality can be approximated by a new measure that we'll call input betweenness centrality or IBC that only requires the input of each station to be known and is defined as:

$$\text{IBC}_i = \sum_{s \neq t} \frac{\sigma_{st}(i)\text{Input(s)Input(t)}}{\sigma_{st}(\text{TotInput-Input(s)})} \tag{2.4}$$

For edges this formula becomes:

$$\text{IBC}_{ij} = \sum_{s \neq t} \frac{\sigma_{st}(i,j)\text{Input(s)Input(t)}}{\sigma_{st}(\text{TotInput-Input(s)})} \tag{2.5}$$

If the input of stations is not known and cannot be approximated either, using the IBC becomes impossible.

2.2.2 Multilayer aspect

Another significant difference between the basic model explained above and a subway network is that a subway network is based on lines while the basic model does not consider them.
Unfortunately, both the notions of closeness centrality and betweenness centrality that are essential in this document are based on shortest paths which are not computed properly when lines are not considered.

The problem will be made clear with this simple example in the north of Paris:

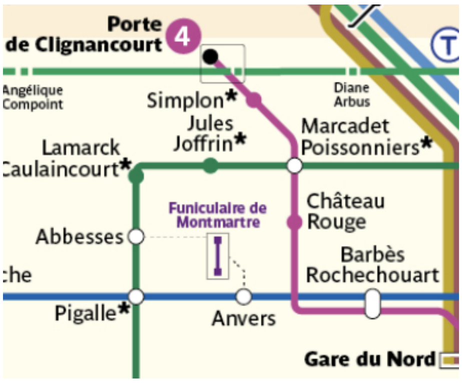


Figure 2.3: Importance of transfer penalties

If one were to calculate the shortest path between ‘Marcadet-Poissonniers’ and ‘Pigalle’ using the basic model thus in a single layer network and neglecting the fact that a subway system is composed of lines, there would be 2 different results of 4 stops:

- The one using only the green line
- The one using the purple line to ‘Barbès-Rochechouart’ and then the blue line

Even though both of these paths consist of the same amount of stops, nobody would take the second path since a transfer between 2 lines requires time to walk between the two platforms and then to wait for the new vehicle.
Making one extra stop takes less time if you can stay on the same line than if you have to change lines.

We thus establish a penalty in the shortest path algorithm each time a change of line is done! If a penalty of one stop is established for example, the first path will still have a length of 4 stops while the second one will now have a length of 5 stops. There will thus only be one shortest path left which is more realistic in this situation.

To implement these penalties, the modelling will use a multilayer network where each layer represents one single line. Each time a path goes from one layer to another its length is penalized accordingly.
It is thus possible that the shortest path between two stations is not the one that has the least stops.
Some other path could have fewer stops but would still be considered longer due to more transfers.

Penalizing transfers too hard would not be realistic either, the shortest path will thus contain few stops and have few transfers.

2.2.3 Why Paris subway

Now that the input at stations and the multilayer aspect have been introduced, it can be explained why the Paris subway network was chosen as example in this document:

1. As stated previously, the differences in input between stations in the Paris subway are higher than in other subway networks.
   Taking the input into account will modify the results more when the differences are larger. It is thus even more important to consider the input of stations in the Paris subway than in other cities.

2. The input data of all stations is accessible[43].

3. The Paris subway network is very dense with a large amount of lines crossing each other in the center of the city. Therefore, many different paths with almost the same length are often possible between two stations. A small penalty on transfers can thus easily change which path is the shortest.
   By modifying more shortest paths, using the transfer penalty in a dense network like Paris will have more impact than in other cities.

2.2.4 Scope of the Paris network modelling

As stated before, the Paris subway network will be used as the main example. More precisely, it will be the 16 subway lines and the two parts of the RER lines that are operated by the RATP (RER A and RER B) that are used. The modelling will thus be a multilayer network with 18 layers.
The reason why only those 18 lines will be modelled is that they are the only ones for which the amount of passengers entering at each station is available[43]. This data is important to implement the input of the stations in the model. The data of 2018 is used in this work.
The data as documented in [43], is given under the form of a closed model. This means that the input numbers consist of all passengers entering the station and all passengers transferring from a line that is not in the 18 lines.
A few examples:

- A passenger going from ‘La Défense’ on subway line 1 to ‘Mairie de Montrouge’ on subway line 4 and transferring at ‘Châtelet’ between line 1 and line 4:
  This passenger is given in the data as entering at ‘La Défense’ but not at ‘Châtelet’.

- A passenger going from ‘Creil’ on RER line D to ‘Mairie de Montrouge’ on subway line 4 and transferring at ‘Châtelet’ between RER D and line 4:
  This passenger is given in the data as entering at ‘Châtelet’ but not at ‘Creil’ because the RER D is not part of the 18 lines.

The input of a station is thus incremented each time a passenger enters onto one of the 18 considered lines while not coming from one at that specific station.

The focus will thus be on the 16 subway lines, the RER A and the RER B because the available data[43] creates a closed system consisting of these 18 lines!

As seen below, the only parts of the RER lines that will be considered are the parts in the center of Paris. The 2 lines are thus significantly cropped and the amount of passengers entering the RER at the extremes in the model is adapted consequently. This is done in order to keep the visualization compact.

Even though the model does not only consist of subway lines but also of 2 RER lines, it will often be referred to as: “the subway network” even though RER lines A and B are not strictly speaking subway lines.

Referring to “the subway network” is less confusing than referring each time to “the subway network and the 2 RER lines” or “the 18 considered lines”.
Starting from here, the RER A and RER B lines are thus considered to be part of the subway network of Paris.

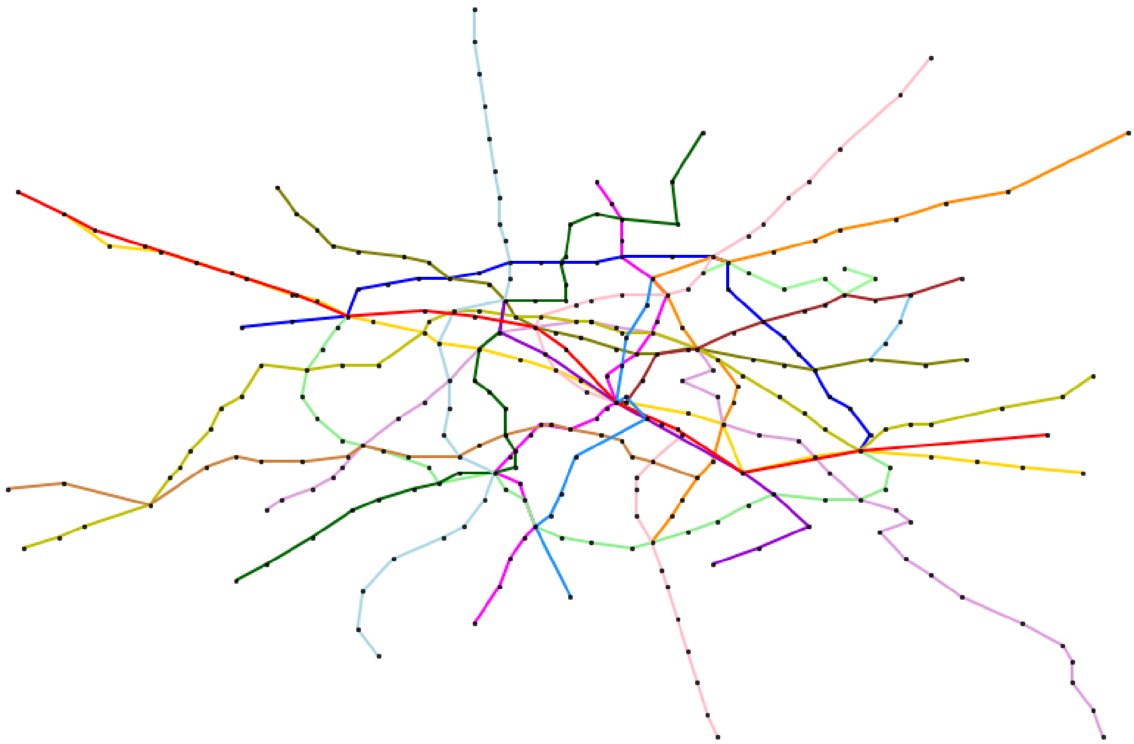


Figure 2.4: Network representing the 18 lines in the Paris subway[44]

Some extra notes about the representation:

- The RER A (the red line from left to right) is cropped to only be in the zone where it has transfer options with the subway.
- The RER B (the short vertical blue line in the middle) is cropped in the south after its last transfer with the subway and is cut of in the north at ‘Gare du Nord’ because the operation of the line is then taken over by the SNCF. All passengers coming from the north that continue on the RER B after ‘Gare du Nord’ are in the data of passengers entering the network at ‘Gare du Nord’ so this does not cause a problem, the system remains closed.
- For clarity, the north part of line 13 (the long light blue one from north to south) and the south part of line 7 (the long pink one from north to south) are not separated in 2 branches as they are in reality and in figure 2.1. This has no impact on passenger flows or other properties in the network since incoming passengers of both branches are added up and the branches do not connect with other lines.

2.2.5 Duration of each stop

Most subway stops have a similar duration of around 90 seconds and public transport passengers usually look at a journey in the subway in terms of stops and not in terms of minutes. All subway stops are thus considered to last as long and the unit for the computation of the shortest paths will be an amount of stops and not an amount of minutes.
RER stops on the other hand are longer than subway stops and amidst themselves they have larger differences. Edges in the RER layers of the network are thus considered to be between 1 and 5 times longer than the ones in the subway layers of the network.

2.2.6 Transfer penalty

As stated in a previous section, there should be a time penalty in a path when a transfer is done between lines.
This penalty is a combination of the time to move to the new platform and the time to wait for the new vehicle. Both of these extra times will depend on the hour and the day but are selected here to be at peak hour:

- From personal experience, it is estimated that moving from one platform to the other often takes around two minutes on average during peak hours.
- The frequency during peak hours of the subway is around 30 per hour for all lines thus the average waiting time will be around one minute[6].

Combined, these two times are of 3 minutes which is the equivalent of 2 stops since a stop is around 90 seconds.
Both times but especially the waiting time are higher when the transfer is towards an RER. The penalty in that case will be of 5 stops during peak hour.

Each time a path changes layer to go to a subway layer its length will be increased by 2 stops and each time it goes to a RER layer in will be increased by 5 stops.

2.3 Validity of adding multilayer and input

Now that the model is defined, it is worthwhile to analyze the impact of having multiple layers instead of one and adding an input to the network has on results. To do this, a new notion is first introduced: The input betweenness centrality of a line or IBCLine. Values are then compared in the Paris subway network.

2.3.1 Input Betweenness Centrality of a line

The IBCLine is defined as the total amount of stops that are done on a certain line by the passengers:

$$\text{IBCLine}_\ell = \sum_{i\in\ell}\sum_{s\neq t} \frac{\sigma'_{st}(i,\ell)\text{Input(s)Input(t)}}{\sigma_{st}(\text{TotInput-Input(s)})} \tag{2.6}$$

[6]The real waiting time is in fact randomly distributed but a passenger will select a path through the network based on an expected waiting time for a subway because the random waiting time is not known yet at that moment. When waiting times become large, passengers will start to plan their journey based on schedules. A modelling as a temporal network[45] is then necessary.

where σ_{st} is still the the total amount of shortest paths between s and t but $\sigma'_{st}(i,\ell)$ is the amount of shortest paths that go from station s to station t through station i **and** having used line ℓ as last line to get to i.

If we'd just sum the IBC of each node on a certain line and thus get $\sum_{i\in\ell}\sum_{s\neq t}\frac{\sigma'_{st}(i)Input(s)Input(t)}{\sigma_{st}(TotInput-Input(s))}$, many paths would be taken into account because they go through a station of the line without actually using the line. Using $\sigma'_{st}(i,\ell)$ enables to only consider paths that really used the line.

If a passenger does 5 stops on a certain line, it will be counted 5 times in the input betweenness centrality of the line because there are 5 stations through which the passenger went while using the line to get to that station.

The input betweenness centrality of the line is thus not the amount of passengers that use the line but the total amount of stops on the line that are done by passengers.

2.3.2 Comparison

Now that the IBCLine is defined, it is possible to compare its values when: the precise input or an uniform input is used for all stations and when transfer penalties are used or not. This way, it is possible to compare the new model as defined in chapter 2.2 with the basic single layer network that does not consider the input.
Ideally, it would be possible to compare the IBCLines as found by the models with the real IBCLines but this data is usually not available.
The comparison will thus be done based on two things.

- Some data of line usage is published by [46]. This data contains the amount of passengers in million that start their journey with that specific line. Passengers are thus only counted once in the value of the first line they use.
  This data thus does not give the amount of stops done on each line but there is a correlation.

1	1 Ligne 1	180,7	2018
2	4 Ligne 4	152,2	2018
3	9 Ligne 9	143	2018
4	7 7bis Lignes 7 et 7 bis	136,2	2018
5	13 Ligne 13	135,7	2018
6	6 Ligne 6	115,1	2018
7	5 Ligne 5	114	2018
8	2 Ligne 2	109,1	2018
9	8 Ligne 8	107,8	2018
10	3 3bis Lignes 3 et 3 bis	98,9	2018
11	14 Ligne 14	87,4	2018
12	12 Ligne 12	85,4	2018
13	11 Ligne 11	47,8	2018
14	10 Ligne 10	46,4	2018
TOTAL		1 559,5	

Figure 2.5: Entering passengers per line in million in 2018[47]

- Passenger experience can also be used as a comparison point. Tools as [48] use passenger experience to know which lines are busy or not. Again, this is not perfect since it is not because a line is the busiest that it transports the most passengers. It could simply be due to lower frequencies or frequent failures for example.

The penalty is the usual one of 2 stops for a transfer to a subway line and 5 stops for a transfer to a RER line.
In order to have comparable data, each station in the models that do not consider input differences was given an input of 5.82 Million which is the average input. This way, the total amount of passengers is the same in each situation.
The following table shows the IBCLine in million of each line for a couple of different models:

Line	Basic Model	Penalty	Input	Penalty + Input	Entering Passengers
1	1146M	1424M	1583M	1757M	181M
2	1297M	1428M	1065M	1283M	109M
3	945M	1047M	678M	727M	99M
3bis	92M	39M	49M	11M	/
4	1860M	1637M	1747M	1777M	152M
5	896M	1266M	696M	1078M	114M
6	1203M	1666M	903M	1130M	115M
7	1230M	1652M	1200M	1486M	136M
7bis	173M	151M	50M	36M	/
8	2738M	2808M	1770M	1714M	108M
9	959M	1675M	683M	1194M	143M
10	1011M	935M	572M	495M	46M
11	608M	788M	363M	512M	48M
12	812M	1210M	548M	751M	85M
13	1543M	1530M	1618M	1656M	136M
14	1469M	850M	1660M	1088M	87M
RERA	838M	856M	1625M	2122M	/
RERB	1041M	440M	1554M	1022M	/

It can be directly noticed that there are some huge differences between the different approaches. Line 1 and RER A for example have a much higher IBCLine when the penalty and the input is added while line 8 and 10 see their IBCLine being seriously reduced.
For some other lines like line 13 for example, the differences are rather small.

It can also be noted that both adding the penalty or adding an input have a significant impact on their own which are sometimes in opposite directions.

Even though it is impossible to compare the numbers with the real values, some analysis is still possible:

- Line 1 is a line that is famous for having a high number of passengers and having often full vehicles. It is the line with the highest frequency after line 14 and was also the first one to be transformed into an automatic one in order to achieve a higher frequency. Last but not least, it is the subway line with the most entering passengers.
  Knowing this, it is likely that its IBCLine would be higher than 1146 million and probably even higher than 1757 million.

- Line 4 will be the second line to be automatized[49] and is also the line with the second highest amount of entering passengers.
  Its high IBCLine in every model is in accordance with that.

- Line 8 is a long line which passengers tend to use to do many stops. It thus makes sense that the total amount of stops done by passengers on this line, which is the IBCLine, would be high. Apart from that, the line is not known to be particularly busy and its amount of entering passengers is average. An IBCLine that is almost twice as high as any other line seems overestimated. An IBCLine of 1714 million, that is in the highest ones but not the highest one, seems more realistic again.

- Line 10 is the one with the least entering passengers, its IBCLine is thus likely to be closer to 495M than 1011M.

- Line 14 is a busy line that has always been automatic and that even has a slightly higher frequency than line 1[7]. The line is really short though, only 8 stops. This means that even with the improved model, the IBCLine of line 14 is 136 million per stop compared to 73 million for line 1.
  This difference can be explained by the fact that line 1 has crowded and less crowded parts while the whole of line 14 is very busy. Its average is thus higher. Nonetheless, the difference stays consequent and it should definitely not be larger.
  An IBCLine of 1088 million on line 14 still seems a bit much, but it is already significantly lower than the original 1469 million.

- While the amount of entering passengers is not given for the RER A, other sources say it has 308 million passengers per year[50]. This number includes passengers that use the line in its SNCF part while only the RATP part is used in this work.
  75% of the line is operated by the RATP so we can estimate the amount of passengers in that part at around 225M. This number is still higher than any subway line and the fact that its IBCLine is the highest is coherent with that.

While the values given by the improved model are not perfect, each time there is a significant difference between them and the values from the basic model, they seem to be the most accurate.

2.3.3 Validity of the model

Even though it is impossible to compare the values that are obtained with real world values. It has been shown with the IBCLines in the Paris subway that it is reasonable to assume that the flows of passengers through a transport network are modelled more accurately when using station inputs and transfer penalties in the model.

Unless otherwise stated, the results discussed in this document will thus always be obtained using the model that considers station input and also considers a transfer penalty of 2 for subways and of 5 for RER's.

2.4 Summary

Whichever subway network is analyzed, it is always possible to add penalties for transfers. There is no extra data that is required except the frequencies of the lines. If those frequencies are not known, it is possible to add a small penalty that might not be precise but that will already give more accurate results than if nothing was done.
In Paris, like most other cities, the frequencies of all the lines can easily be found[40].

Giving a weight to each shortest path on the other hand, requires the input at each station to be known. This data can be found for Paris and some other cities like London but is not always available. One can try to estimate them based on some criteria as population density as well as the

[7] 1 subway every 85 seconds at peak hour

amount of employees working in a certain area of the city. This will be complicated but easier than estimating the number of passengers from any s to any t.
When constructing a new subway network or extending an existing one, making such an estimation is necessary to use the IBC.

If the data with the amount of passengers travelling between any 2 stations is available, the amount of passengers going through all edges and stations can be computed more accurately. The passenger betweenness centrality can then be used.

Even though the focus is here on subway systems, the modelling can also be used for other transport systems that are composed of multiple lines as bus networks, train networks or combinations.
Those types of networks might have stop durations that are more diversified than subway networks though. In a train network for example, some stops can last 20 minutes while others only last 3 minutes. In such a situation, it would then be necessary to consider stop duration when computing shortest paths in order to keep it realistic.

Chapter 3

Centrality in the Paris Subway

This chapter analyzes the centrality measures in a subway system using the improved modelling of chapter 2. The focus lies on the Paris subway for the reasons described in section 2.2.3. The aim is to have a better insight of the structural importance of each node and edge in the network.

Some results will be compared with the London tube modelled using the improved modelling and the Madrid subway system modelled with the basic model as defined in chapter 2.2 whose results come from [4]. The fact that 2 RER lines are added to the Paris subway is not a problem for comparisons because both the London tube and the Madrid subway have lines that have the same function as a RER in Paris.

The centrality measures are defined in chapter 1 or in chapter 2 for the IBC measures.

3.1 Degree Centrality

In the network, the degree centrality of each station will represent the amount of directions in which a passenger can go.
The average degree is 2.44. This is high and shows that there are many connecting stations. In Madrid, the average degree is only 2.28 for example[4].
The following stations have the highest degree in the network:

Rank	Station	Degree	Rank	Station	Degree
1	Châtelet	13	11	Jaurès	6
2	République	10	12	Bastille	6
3	Montparnasse-Bienvenue	8	13	Concorde	6
4	Opéra	8	14	La Motte Picquet Grenelle	6
5	Nation	8	15	Gare de Lyon	6
6	Saint-Lazarre	7	16	Denfert-Rochereau	6
7	Charles de Gaulle-Etoile	7	17	Strasbourg-Saint-Denis	6
8	Gare de l'Est	6	18	Place d'Italie	5
9	Madeleine	6	19	Gare du Nord	5
10	Stalingrad	6	20	Palais Royal	4

- 'Châtelet' has the highest degree by a significant margin. 7 different lines can be taken from it!
- Most stations have an even degree which is due to the fact that, except if that station is the terminus, a passenger can go in two directions on a specific line. One more line going through the station thus often increases the degree by 2.

- These results would have been the same if computed with the basic model. Neither the penalties nor the station inputs have an impact on degree centrality.
- On the figure, we see that the stations with highest degree are in the middle part of the network but sill rather spread out. 4 of the 7 stations with highest degree are on the red RER A line.

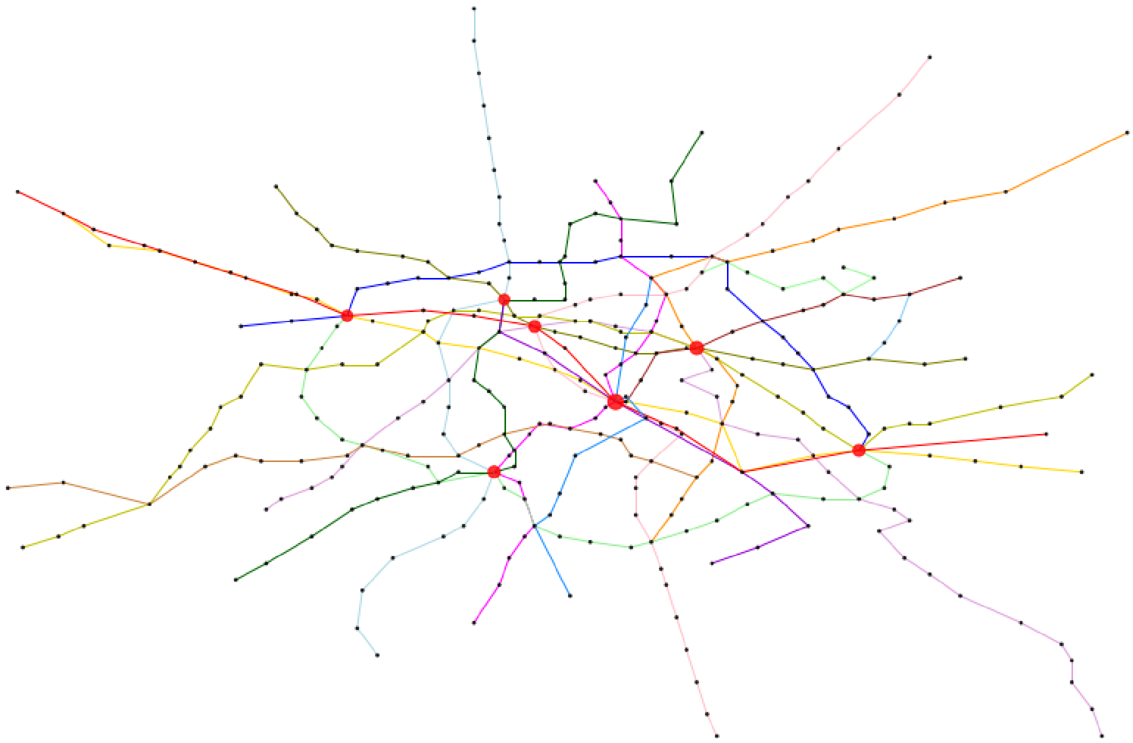


Figure 3.1: Stations with degree ≥ 7

3.2 Eccentricity

The diameter of the network is 42 stops. This is smaller than the diameter of the London tube which is 47 stops. According to [4], the diameter of the Madrid subway is 44 stops but this is without considering the transfer penalties. Madrid's diameter is thus even higher.
The following stations have the lowest eccentricity:

Rank	Station	Eccentricity	Rank	Station	Eccentricity
1	Pyramides	21	11	Cité	23
2	Louvre-Rivoli	22	12	Pont Neuf	23
3	Châtelet	22	13	Bonne Nouvelle	23
4	Madeleine	22	14	Chaussée d'Antin la Fayette	23
5	Palais-Royal	23	15	Saint-Paul	24
6	Hotel de Ville	23	16	Opéra	24
7	Gare de Lyon	23	17	République	24
8	Saint-Lazarre	23	18	Réaumur-Sebastopol	24
9	Strasbourg-Saint-Denis	23	19	Saint-Michel	24
10	Les Halles	23	20	Place d'Italie	24

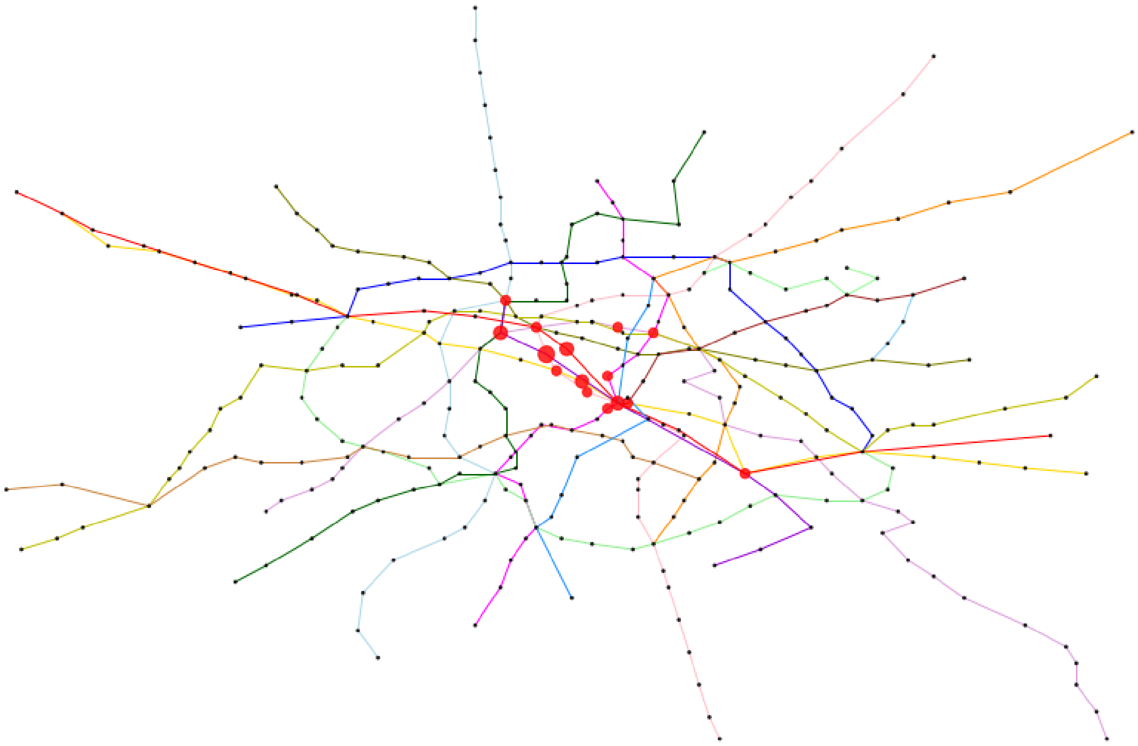


Figure 3.2: Stations with eccentricity ≤ 23

- There seems to be almost no correlation between degree centrality and eccentricity. ‘Pyramides’ that has the lowest eccentricity only has a degree of 4.

- The stations with lowest eccentricities are all close to the middle of the network which is to be expected.
  The larger the highlighted node is on the figure, the lower its eccentricity.

- ‘Pyramides’ and ‘Opéra’ are adjacent nodes but have an eccentricity of 21 and 24. This seems strange at first but is due to the fact that transfer penalties are taken into account to compute the shortest paths and thus the eccentricities!
  Considering the input of stations on the other hand does not have an impact on the eccentricity.

- While interesting, both the eccentricity and the diameter give limited information. The network only needs one line that stretches far out of the city to increase all eccentricities.
  Eccentricities in Paris are highly affected by line 8 in the south east of the city and the green line 9 in the south west of the city.

3.3 Closeness Centrality

While the eccentricity gives information about the furthest other station for each node, the closeness centrality of a station will give information about the average path length towards other stations.
The average closeness centrality which is thus the average length of a any path in the network is of 14.9 stops. This is less than the 16.9 stops in London. It is a bit higher than the 14.7 stops in Madrid[4] but this is, once again, without considering transfer penalties. Counting one transfer on average per path, the average path length in Madrid would already increase by 2 and thus be higher than the one in Paris as well.
The most central stations according to this measure are:

Rank	Station	CC	Rank	Station	CC
1	Châtelet	8.732	11	Strasbourg-Saint-Denis	10.75
2	Opéra	9.487	12	Louvre-Rivoli	10.88
3	Madeleine	9.596	13	Chemin d'Antin La Fayette	10.91
4	Pyramides	9.948	14	Cité	10.96
5	Saint-lazarre	9.987	15	Richelieu-Drouot	10.99
6	Gare de Lyon	10.05	16	Bercy Bastille	10.99
7	Concorde	10.26	17	Champs-Elysées-Clémenceau	11.00
8	Hotel de Ville	10.62	18	Invalides	11.01
9	Palais Royal	10.65	19	Les Halles	11.03
10	République	10.73	20	Havre Caumartin	11.05

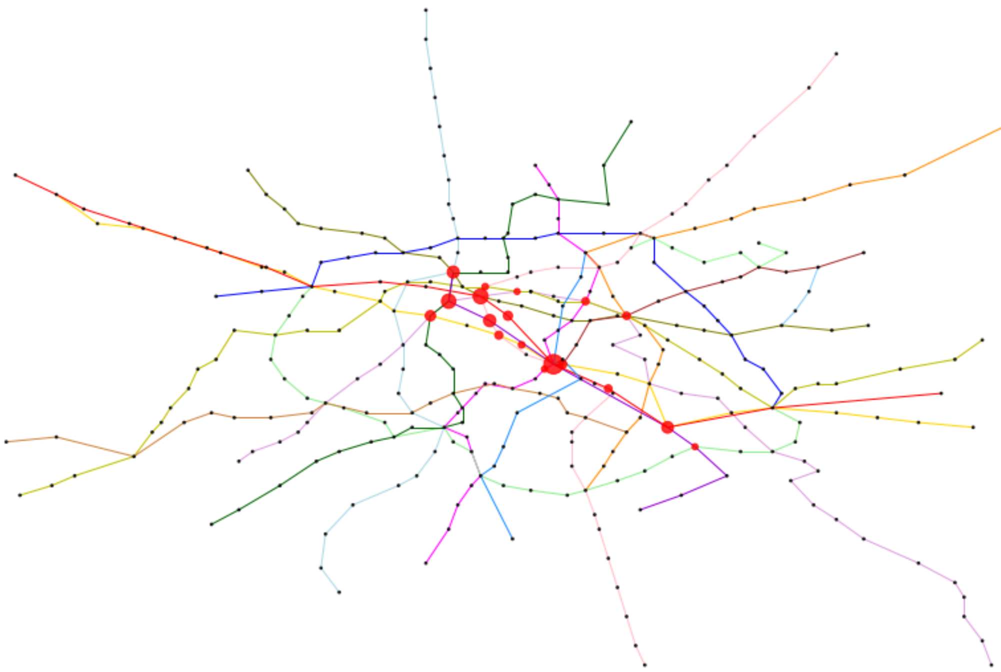


Figure 3.3: Stations with closeness centrality ≤ 11

- The stations that have the lowest closeness centrality are almost the same as the ones that have a low eccentricity. The order in which they are in the ranking is different though.
  Both figures have mostly the same nodes highlighted but the size of these highlighted nodes

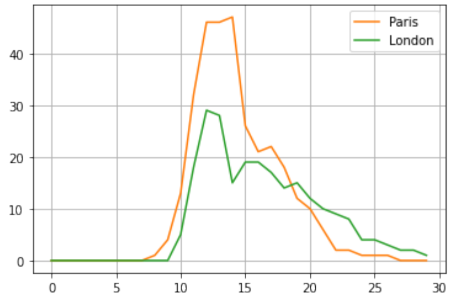


Figure 3.4: Closeness Centrality distribution in Paris and London

differs.
The larger the highlighted node in figure 3.3, the lower its closeness centrality is.

- Just as for the eccentricity, the fact that transfers are penalized has an impact, but the input of the stations does not.
- Most stations have a closeness centrality that is around 13. The stations with a closeness centrality higher than 15 are mostly the ones in the branches in the outer part of the network. Generally speaking, the distribution graph shows that most closeness centralities are in the same range and that there are only few outliers.
  The distribution in London is more spread out, there are thus larger differences between stations than in Paris.

3.4 Input Betweenness Centrality

The average input betweenness centrality for each station is 64.1 million and the maximum is at 'Châtelet' with 743.4 million. This is much more than any other station and shows the essential importance of 'Châtelet' in the network.
In London the average is 59.6 million and the maximum is 377 million. There is thus a bit less traffic than in Paris on average but the main difference is that the network relies less on a single station like the network in Paris does[1]!

The 20 stations with the highest IBC in Paris are the following:

Rank	Station	IBC	Rank	Station	IBC
1	Châtelet	743M	11	République	220M
2	Gare de Lyon	423M	12	Montparnasse-Bienvenue	207M
3	Charles de Gaulle-Etoille	298M	13	Gare de l'Est	196M
4	Nation	296M	14	Place de Clichy	195M
5	Madeleine	293M	15	Invalides	188M
6	Opéra	279M	16	Champs-Elisée - Clémenceau	179M
7	Saint-Lazare	260M	17	Denfert-Rochereau	175M
8	Pyramides	231M	18	Stalingrad	171M
9	Concorde	226M	19	Franklin D. Rooseveld	160M
10	Gare du Nord	220M	20	Barbes Rochechouart	151M

[1]See chapter 5 for what happens when 'Châtelet' is removed

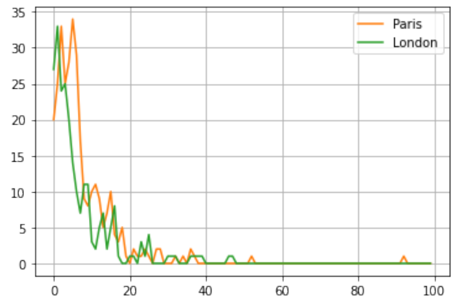


Figure 3.5: IBC distribution in Paris and London (all values divided by 8!)

- With the exception of that one station 'Châtelet', the distribution of IBC in Paris and London is similar.

- When computing the IBC of stations, both the fact of using a multilayer model and using station inputs has an impact on the results[2].

It is not possible using only the IBC of stations to make proper conclusions about the following:

1. The busyness of the lines around a station because it depends on how many there are and how the passengers are distributed between them.
   This is why in the next section, the IBC of edges will be considered.

2. The busyness of the station infrastructure. The amount of people walking around a specific station depends on how many passengers enter/exit the subway network at that station and how many change lines at that station. The IBC of a station does not indicate whether the passenger just passed through the station while staying in the same vehicle or not.
   This is why the amount of transfers at each station will be analyzed as well.

[2]Contrary to the previous measures that were not impacted by the input of the stations

3.4.1 Input Betweenness Centrality edges

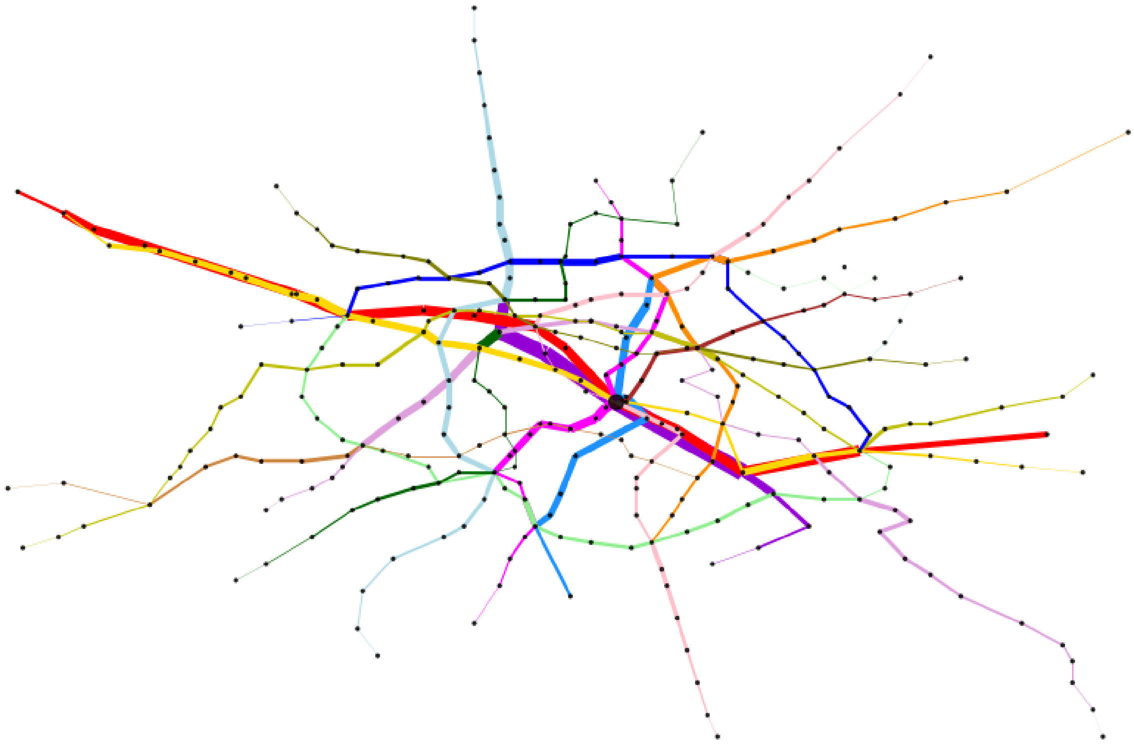


Figure 3.6: IBC of edges in Paris

The thicker the lines are depicted, the higher the IBC of a specific edge is. 'Châtelet', that has the highest IBC can clearly be noticed in the middle of the network with the thicker black dot.

Even though 'Châtelet' has a huge IBC, we see that lines 1 and 11, the yellow and brown ones, are not that busy around 'Châtelet'. This clearly demonstrates the importance to look at the IBC of edges and not only at the IBC of nodes.

The edges on which there is the highest traffic are:

- Most of the edges of line 14 which is the thick purple line are extremely busy with the busiest ones around 'Châtelet' having over 200 million passengers a year.
- After the edges of line 14, the busiest ones are on the red RER A. The edges are not as busy as the ones on line 14 but their IBC still goes up to 160 million passengers a year.
- As can be seen on the figure, most edges with a high IBC are the ones in the middle of the network but even in the middle of the network, there are some huge differences.

3.4.2 Input Betweenness Centrality lines

First of all, it is important to remember that summing up the IBC of all the edges of a specific line does not give exactly the same result as the IBCLine. This is due to some edges being part of 2 different lines.

Ordering the different lines in terms of busyness can be done in two different ways: either the lines can be ordered in terms of their total IBCLine (first table) or they can be ordered according to the edge on which they have the most passengers (second table).

Rank	Line Number	IBCLine	Rank	Line Number	IBCLine
1	RERA	2122M	10	14	1088M
2	4	1777M	11	5	1078M
3	1	1757M	12	RERB	1022M
4	8	1714M	13	12	751M
5	13	1656M	14	3	727M
6	7	1486M	15	11	512M
7	2	1283M	16	10	495M
8	9	1194M	17	7bis	36M
9	6	1130M	18	3bis	11M

Rank	Line Number	Maximum IBC	Rank	Line Number	Maximum IBC
1	14	277M	10	7	71M
2	RERA	163M	11	11	66M
3	4	140M	12	10	65M
4	13	124M	13	6	56M
5	1	123M	14	9	56M
6	8	114M	15	3	55M
7	RERB	99M	16	12	52M
8	2	98M	17	7bis	10M
9	5	84M	18	3bis	5M

Both rankings are similar with the main exception of line 14 that by its short length has an average IBCLine while having edges with very high IBC.

3.4.3 Number of transfers per station

In order to know which stations are the most crowded, we need to know the inputs of the stations but also the amount of passengers that change lines at each station. The station of 'Châtelet' for example will be very crowded even though its input is not that high because so many passengers operate a transfer there.
The most transfers are done in the following stations:

Rank	Station	Transfers	Rank	Station	Transfers
1	Châtelet	378M	11	Bercy	56M
2	Saint-Lazare	134M	12	Place d'Italie	56M
3	Gare de Lyon	128M	13	Stalingrad	54M
4	Nation	120M	14	Gare du Nord	51M
5	Charles de Gaulle-Etoile	103M	15	Place de Clichy	51M
6	Montparnasse-Bienvenue	101M	16	Barbes - Rochechouart	49M
7	République	87M	17	Gare de l'Est	48M
8	La Motte Picquet Grenelle	86M	18	Denfert-Rochereau	47M
9	Opéra	86M	19	Bastille	44M
10	Madeleine	71M	20	Champs-Elysées-Clémenceau	38M

We can then add for each station the input and output which are assumed to be identical, to obtain the total amount of passengers using the station. This is the load of the station.

Rank	Station	NbPass.	Rank	Station	NbPass.
1	Châtelet	470M	11	La Motte Picquet Grenelle	102M
2	Gare de Lyon	279M	12	Gare de l'Est	90M
3	Saint-Lazare	228M	13	Madeleine	85M
4	Gare du Nord	227M	14	Place d'Italie	80M
5	Montparnasse-Bienvenue	163M	15	Bercy	70M
6	Nation	146M	16	Bastille	70M
7	Charles de Gaulle-Etoile	137M	17	Place de Clichy	69M
8	République	123M	18	Stalingrad	68M
9	Opéra	120M	19	Barbes - Rochechouart	67M
10	La Défense	107M	20	Denfert-Rochereau	65M

Including both the results of the IBC of edges and the total amount of passengers frequenting each station, we get the following figure that represents the load on each segment of the network and the load on each station:

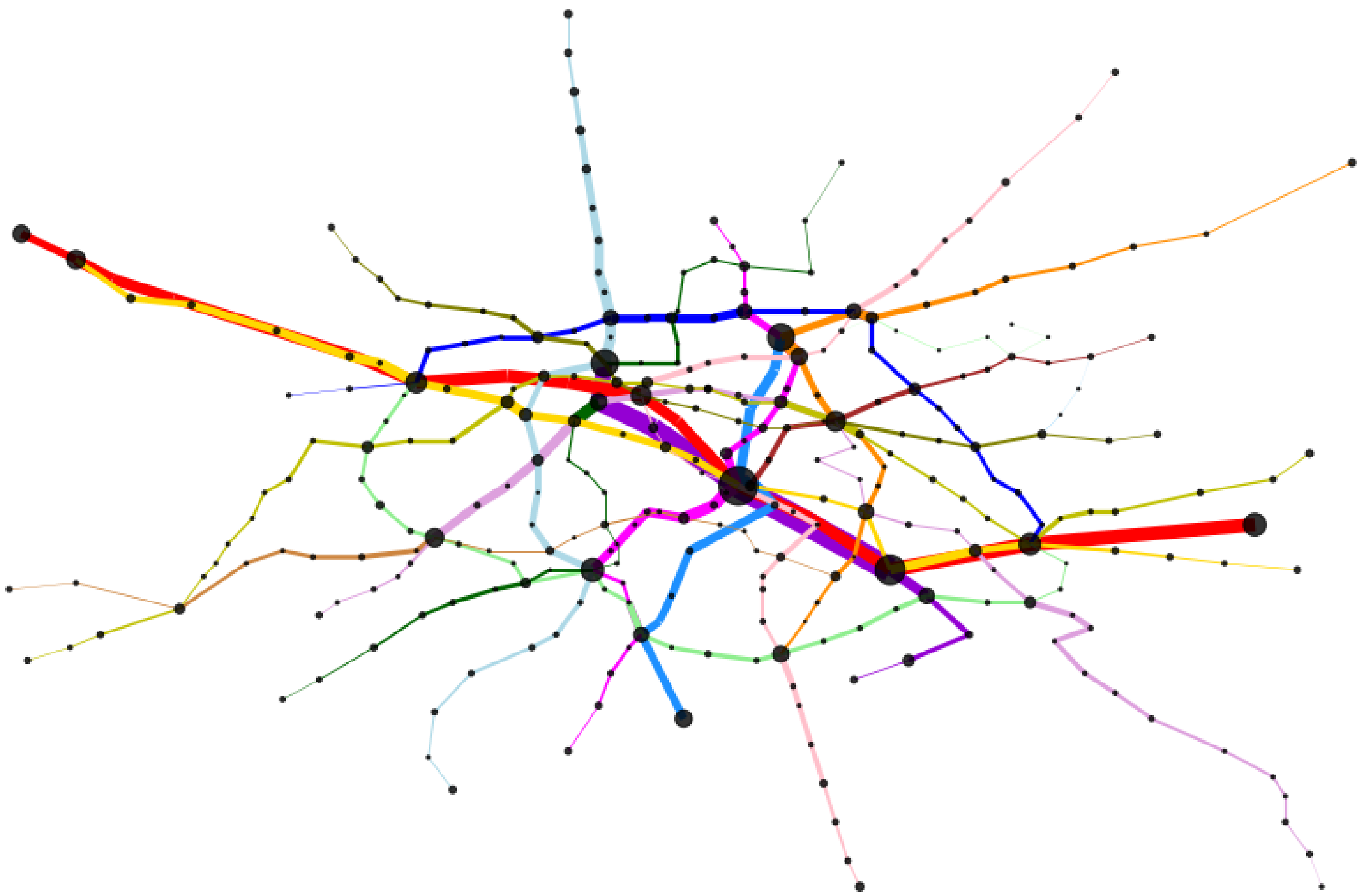


Figure 3.7: IBC of the edges and amount of passengers using each station in Paris

3.5 Impact of the transfer penalty

3.5.1 On Lines

As stated multiple times already, a length penalty is inflicted on every path that changes line in the network. This has an impact on the shortest paths between the multiple stations in the subway network that, in turn, has an impact on the closeness centrality and the IBC of nodes, edges and lines.

All of the previous results were obtained by adding penalties of 2 for a transfer to a subway and 5 for a transfer to a RER. These penalties were chosen to represent real life transfer times during peak hours and thus provide a poor representation of the transfer times during off-peak hours.
All lines will be less crowded during off-peak periods than peak periods but proportionally, their occupation might be different in function of the moment of the day.

The following table shows the IBCLine for every line depending on how many stops the penalty is for the subway and the RER.
Even though the penalties for the subway go from 0 to 10 in the table to give a broader view, the realistic penalties are the ones between 2 and 4 that represent the transfer times that can be expected in the Paris subway where the frequencies are relatively high at any time of the day.
In extreme situations[3], the frequencies can be reduced further and the transfer penalty can then be estimated up to 7[4].
The numbers represent the total amount of stops in million that would be done on each line per year if the penalty was applicable during the whole year.

Line\penalty	0-0	1-3	2-5=peak	3-7=off-peak	4-9=night	5-11	7-15=Covid-19	10-21
1	1583	1340	1757	1892	2054	2124	2187	2254
2	1065	1248	1283	1247	1214	1197	1197	1171
3	678	657	727	814	859	902	963	1033
3bis	49	16	11	8	6	6	6	6
4	1747	1523	1777	1937	2001	2129	2167	2259
5	696	975	1078	1080	1082	1100	1088	1101
6	903	1070	1130	1235	1265	1286	1361	1380
7	1200	1354	1486	1577	1678	1739	1842	1979
7bis	50	47	36	32	33	34	33	33
8	1770	1702	1714	1675	1708	1733	1830	1923
9	683	983	1194	1418	1566	1698	1818	1970
10	572	537	495	477	492	499	498	494
11	363	484	512	517	513	510	501	492
12	548	597	751	819	877	909	943	978
13	1698	1697	1656	1640	1660	1685	1714	1713
14	1660	1307	1088	965	889	823	759	706
RERA	1625	2181	2122	2140	2124	2084	2020	1984
RERB	1554	1269	1022	942	891	832	822	810

It can be seen in the table that changing the penalty has a certain impact. Some lines are thus proportionally more busy during off-peak hours than peak hours. For other lines the opposite is true.

[3] In April 2020 the line frequencies were divided by 3 due to the global pandemic of the Covid-19[51]

[4] 1 subway every 18 minutes thus on average 9 minutes of waiting which represents 6 stops. 1 extra stop is added for the walking time between platforms.

- Lines 1 and 4 which are both already important lines, get even more important when the line frequencies decrease.
- Lines 9 and 7 also get more passengers which can be explained by their length. When transfers are more penalized, passengers will prefer longer lines on which they will do many stops instead of a combination of shorter lines.
- Opposite to lines 9 and 7, line 14 sees its importance significantly decreased due to it being mostly part of multi transfer journeys.

On the figures in appendix A that map out the IBC of the edges and the amount of passengers using each station, the evolution of the importance of each line can clearly be noticed.

3.5.2 On Stations

The transfer penalty has an impact on the IBC of lines and edges but it also has an impact on the amount of transfers that are done at each station and thus on the amount of passengers that use each station.

The table and figure in appendix A show that the differences can be significant depending on the time of the day, but also that the difference between having a penalty and not having a penalty can be huge.

Station 'Madeleine' for example has 294M transfers when no penalty is used, 71M transfers when there is the peak hour penalty of 2 and 39M transfers when there is the night penalty of 4.
While the differences in amount of transfers between peak and night penalty are considerable, it is nothing compared to the differences between no penalty and penalty. Even more than analysing the IBC of lines, analysing the amount of transfers done at each station gets more accurate when transfers are penalized.

3.6 Impact of using input of the stations

Just as adding a penalty to each transfer changes significantly the IBC of the edges and the amount of passengers using each station, so does adding the input of stations.
The differences are not as large but are still considerable.

Visualizations are available in appendix B.

Chapter 4

Backbone of the Network with Sparsification

While the new model presented in chapter 2 enables to obtain the centrality results from chapter 3, it can also be used for other applications. One of these is to extract a backbone with sparsification methods.

Some sparsification methods require a weighted graph and by giving as weight to each edge its input betweenness centrality or IBC as computed in previous chapters, these methods could offer interesting results. Existing methods will thus be used but the extracted backbones will be new.

This chapter starts with a quick reminder on sparsification before applying some filters on the Paris subway network that will once again be used as example. Each filter will extract/select different edges that together create a backbone. A combination of those filters to extract a single backbone is also suggested near the end of the chapter.

The 3 first sparsification methods are defined in chapter 1.

4.1 Sparsification

"Sparsification reduces the size of networks while preserving structural and statistical properties of interest."[52] The aim of sparsification methods is thus to simplify the network and lower its complexity while maintaining most of its accuracy in describing the data. This is often necessary when analysing huge complex networks like influence networks[53], multiscale networks[54] or social networks[52].

Even though sparsification techniques are mostly used to decrease the complexity in huge networks, they can also be applied on smaller networks with the aim to discover what the backbone of the network is[34] and thus have a better understanding of its structure. This second way of using sparsification methods is more interesting here since transportation systems are only medium sized networks.

Since the context is transportation networks which are multilayer networks where edges are grouped in layers or lines, the aim of this section is not only to discover what the most important edges are in the network, but also which lines have the highest structural importance. These will be the ones that have many edges in the extracted network while less important lines will have more of their edges not extracted.
This way, a backbone can be defined as a set of edges or a set of lines.

4.2 Weight Filter

As a reminder, the weight threshold filter simply consists in extracting the edges of the network that have a weight that is higher than a specified threshold.

Using what has been done in chapter 2, the input betweenness centrality or IBC of each edge is chosen as weight for that edge. This is similar to what has already been done in some papers[32, 55] where the BC of edges is chosen as their weight.

The table shows the amount of edges left per line when the weight threshold is set at 50 million and 100 million yearly passengers. The figure shows the extracted edges for the first case.

First of all, it can be noticed that most of the extracted edges are in the center of the network. This makes sense and could already be expected after seeing figure 3.7 where it can clearly be seen that edges in the center are thicker on average and thus busier than the ones at the extremities.
The results also show some differences between the lines. Some of them are completely extracted for the 50M threshold and thus represent important axes. Other lines have no or almost no edges extracted for the same threshold. Unsurprisingly, the correlation between the IBCLine and amount of extracted edges of a line is high.

While the filter was applied on the subway network of Paris here, the busiest parts of most subway networks is in their center thus the filter will mostly extract central edges like in Paris.

Overall, this filter should be used if one wants to make a backbone of a transportation network that includes the busiest parts of the network.

Line	IBCLine	NbEdges	NbEdges with $w_{ij} \geq$ 50M	%	NbEdges with $w_{ij} \geq$ 100M	%
/	/	372	159	43%	45	12%
1	1757M	24	18	75%	5	20%
2	1283M	24	13	54%	1	4%
3	727M	24	3	12%	0	0%
3bis	11M	3	0	0%	0	0%
4	1777M	26	17	65%	5	19%
5	1078M	21	12	57%	2	9%
6	1130M	27	6	22%	0	0%
7	1486M	33	15	45%	0	0%
7bis	36M	7	0	0%	0	0%
8	1714M	37	15	40%	4	10%
9	1194M	36	5	13%	0	0%
10	495M	19	3	15%	0	0%
11	512M	12	6	50%	1	8%
12	751M	28	3	10%	2	7%
13	1656M	25	18	72%	5	20%
14	1088M	8	7	87%	5	62%
RERA	2122M	16	16	100%	15	93%
RERB	1022M	10	10	100%	5	50%

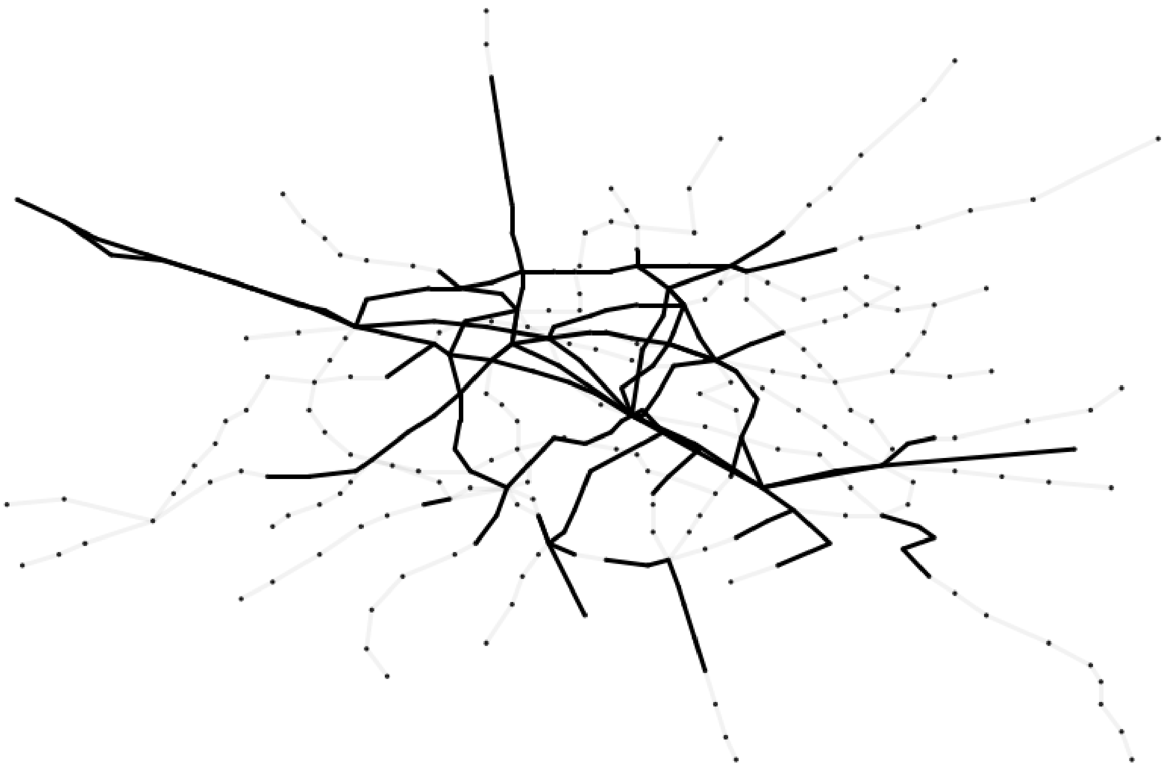


Figure 4.1: Backbone of the Paris subway with a weight threshold of 50M

4.3 Salience Filter

The main problem of the weight filter as applied in the previous section, is that it is directly connected to the IBC of edges and lines. Few new interesting insights can thus be gained using the weight filter when the IBC of edges and lines has already been analyzed.
Furthermore, there are some issues with suggesting that the backbone is only composed of the busiest edges. Some edges are busy but might be easily replaced while less busy edges might be harder to replace in which case they might actually be considered more important.

By not considering the weights of the edges and thus not considering passenger flows, the Salience filter will focus on edges that are hard to replace instead of edges that are busy. It thus does not matter how many passengers use a specific edge but only the importance that that specific edge has for the passengers that use it.
Edges at the extremities of transportation networks will always be extracted because there is no possible alternative for passengers using that edge. Edges in the center of the network will get a lower significance score since transportation networks are often more dense in the center and there will therefore be more alternatives.

The following table and figures show which 162 or 106 edges are extracted by the Salience filter in the Paris subway.

Line	NbEdges	NbEdges with $s_{ij} \geq 0.9356$	%	NbEdges with $s_{ij} \geq 1$	%
/	372	162	43%	106	28%
1	24	10	41%	4	16%
2	24	4	16%	2	8%
3	24	11	45%	9	37%
3bis	3	1	33%	0	0%
4	26	10	38%	6	23%
5	21	12	57%	7	33%
6	27	2	7%	0	0%
7	33	17	51%	14	42%
7bis	7	4	57%	2	28%
8	37	18	48%	17	45%
9	36	14	38%	9	25%
10	19	8	42%	2	10%
11	12	6	50%	1	8%
12	28	13	46%	9	32%
13	25	16	64%	16	64%
14	8	3	37%	3	37%
RERA	16	8	50%	4	25%
RERB	10	6	60%	1	10%

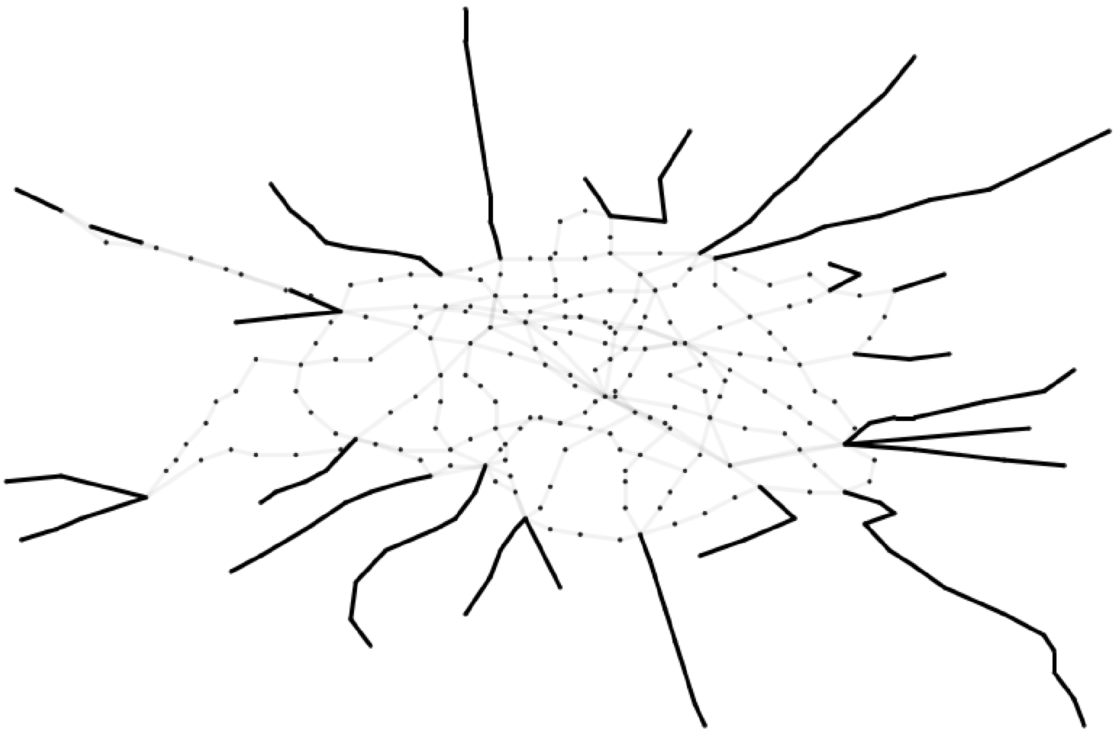


Figure 4.2: Backbone of the Paris subway with the Salience filter with 106 edges left

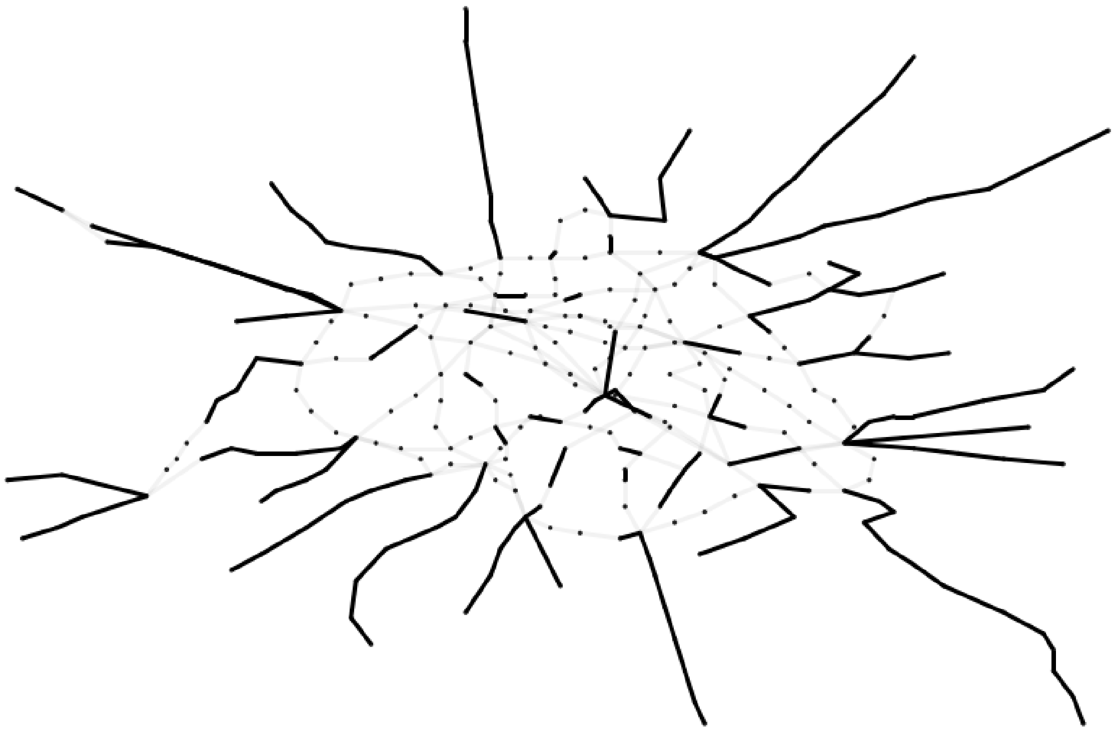


Figure 4.3: Backbone of the Paris subway with the Salience filter with 162 edges left

The first important comment is that the 106 edges that have a normalized salience of 1 which is the highest possible value, are the ones in the arms of the network. 104 out of those 106 edges would disconnect the network in case they were removed. Removing those edges of the network would cause the highest harm to their users because they would simply not be able to reach their destination anymore.
When the amount of extracted edges is increased to 162, some edges are being extracted in the center of the network but the backbone still mainly consists of outer edges. The difference with the backbone of the weight filter in figure 4.1 with almost the same amount of extracted edges is striking.

The lines that have the most extracted edges are mostly the long ones that reach far out of the city like lines 7,8 and 13.
The important lines according to the Salience filter do not match with important ones according to the weight filter. Line 12 for example has many edges in the salience backbone but only few in the weight backbone.

The backbone of a transportation network according to the Salience filter is composed of the edges that are the hardest to replace for their users and is thus mainly composed of off center edges where few alternatives exist.

4.4 Disparity Filter

Whereas the backbone of a subway network according to the weight filter is mainly composed of edges in the center of the network, the backbone according to the Salience filter is composed mostly of edges that are in the outskirts of the network. Both filters thus provide results that are very different.
By its definition, the backbone according to the disparity filter, will give results somewhere in between.

Just as for the weight filter the disparity filter requires that each edge should have a weight and the IBC will thus be used again.
As stated in chapter 1.3.3, the disparity filter will extract edges that have a weight that is high compared to incident edges. While the weight filter keeps edges that have a globally high weight, the disparity filter keeps edges that have a locally high weight!
An edge with a low weight that is surrounded by edges with even lower weights might thus be extracted which would not have been the case with the weight filter. Thanks to this, the edges that are extracted by the disparity filter are more spread out over the network and not only focused on some areas of it.

2 different versions of the filter are used. The one where the threshold is put on the average of β_{ij} and β_{ji}[1] and the one where the threshold is put on the maximum of the 2 values. The choice depends on whether an edge should have a significant relative importance for both of its incident nodes or just for one of them.
With the exception of line 6 where there is a significant difference depending on which version is used, the results are similar with both versions. Henceforth, only the version with the average will be considered because it might make slightly more sense[2].

Line	Nb Edges	$(\beta_{ij} + \beta_{ji})/2 \geq 0.51$	%	$\text{Max}(\beta_{ij}, \beta_{ji}) \geq 0.675$	%
/	372	167	45%	168	45%
1	24	14	58%	16	66%
2	24	14	58%	12	50%
3	24	9	37%	8	33%
3bis	3	0	0%	2	62%
4	26	19	73%	17	65%
5	21	12	57%	15	71%
6	27	13	48%	7	25%
7	33	12	36%	14	42%
7bis	7	0	0%	1	14%
8	37	14	37%	16	43%
9	36	9	25%	10	27%
10	19	5	26%	5	26%
11	12	10	83%	10	83%
12	28	10	35%	9	32%
13	25	13	52%	13	52%
14	8	7	87%	6	75%
RERA	16	9	56%	10	62%
RERB	10	5	50%	5	50%

[1]See chapter 1.3.3
[2]This is a personal opinion

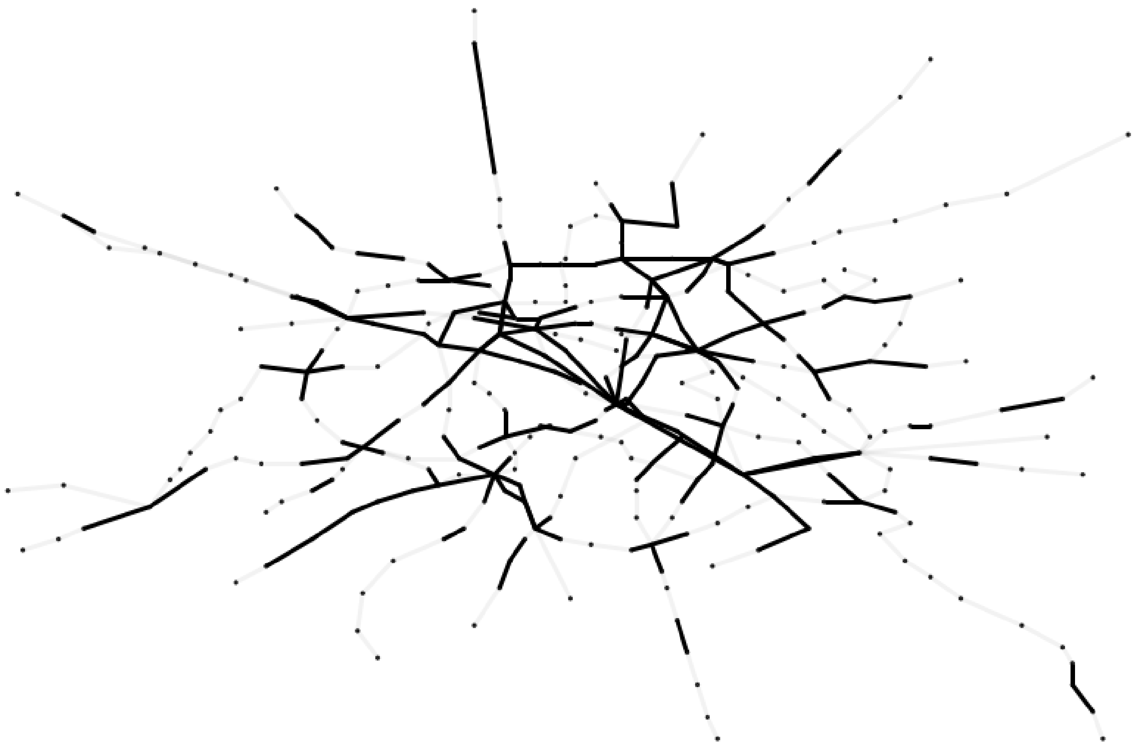


Figure 4.4: Backbone of the Paris subway with the disparity filter (version with average) with 167 edges left

The extracted edges are now more spread out than they were with the weight filter but most of them still remain in the center.
Concerning the lines, line 4 is now the line that has the most extracted edges which it was not with the other filters. Lines 14 and 11 have a high percentage of edges kept which shows that even if those lines are short, they are significant for most stations through which they go.
On the other hand, line 9 which is a very long one, has only a low importance for most stations through which it goes.

4.5 Backbone with a Filter Combination

All 3 of the previous filters extract a backbone of the network based on their specific criterion which results in very different backbones.
In this section it is suggested to combine all 3 previous filters to obtain one single backbone that includes the edges with the highest significance scores of each filter.
The backbone will therefore be composed of the edges with the highest global weights, the edges with the highest local weights and the edges that are the hardest to replace.
Obviously, in order to not have too many edges extracted, it is necessary to use higher thresholds in each of the filters than what is used when the filters are used independently. It is suggested to select around 30% of the edges with the highest significance score for each criterion.

When the 106 edges with the highest salience, the 81 edges with highest local weight and the 117 edges with highest global weight are selected, the following backbone is obtained for the subway network of Paris containing 220 edges[3] out of the original 372.

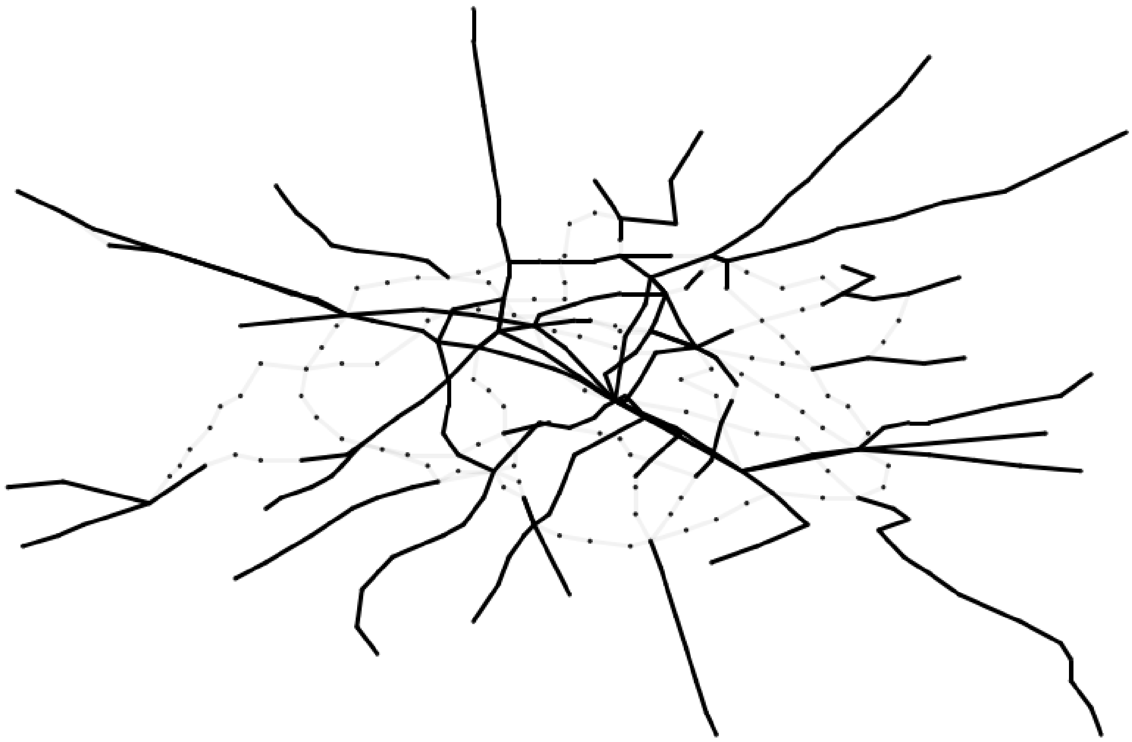


Figure 4.5: Backbone of Paris subway with the combination of filters

This backbone is more representative of the network than the other ones because it includes the most important edges according to multiple criteria. Edges that are not extracted with this combination of filters can fairly be excluded since they are not essential in any way.

[3]Some edges are extracted by multiple filters which is why the number is not 304

4.5.1 Lines

While backbones are usually just a set of edges, it makes sense in a subway network to also think of a backbone as a set of lines. Whole lines are then extracted instead of single edges.

It is possible to use or create filters that specifically extract lines by giving them a significance score as their IBCLine for example.
It is also possible to extract lines based on filters on edges. By giving as significance score to each line the percentage of its edges that are extracted with the combination of filters that was just presented, we can afterwards only keep the lines that are composed of many important edges.

Line	Nb Edges	Nb extracted edges	%
/	372	220	59%
1	24	20	**83%**
2	24	9	37%
3	24	10	41%
3bis	3	0	0%
4	26	23	**88%**
5	21	17	**80%**
6	27	1	3%
7	33	24	**72%**
7bis	7	2	28%
8	37	26	**70%**
9	36	11	30%
10	19	6	31%
11	12	9	**75%**
12	28	11	39%
13	25	25	**100%**
14	8	8	**100%**
RERA	16	16	**100%**
RERB	10	10	**100%**

The higher the amount of extracted edges is per line, the higher its importance in the network.

Some lines already have all of their edges extracted or almost all of their edges extracted while others have almost no extracted edges.
The lines are actually separated in 2 groups. The first group has less than 41% of their edges extracted while the second group has more than 70%.

If one is looking for a backbone consisting of lines and not of edges, the obvious choice is thus to extract the 10 lines of the second group that together consist of 210 out of the 372 edges of the whole network.

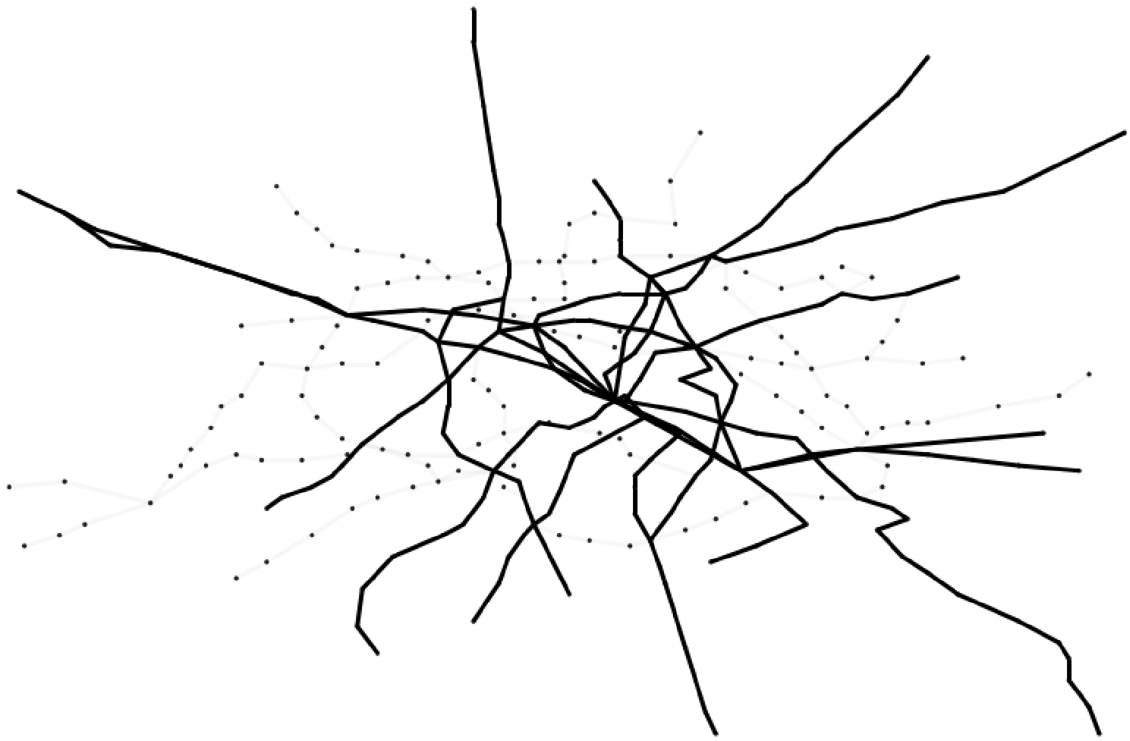


Figure 4.6: Backbone of lines of the Paris subway

4.6 Summary and Discussion

The first filter to be used was the weight filter. This filter does not bring much new information compared to what was already known after chapter 3.
The extracted backbone represents the busiest edges and thus tells us where there are the most passengers in the network.

The Salience filter was then used as a filter that does not use the weight of edges but the extracted edges in the backbone are significantly off center and not connected at all. Much information about the structure of the network cannot be learned from it.

As an intermediate between both previous filters, the disparity filter was used. The resulting backbone which is composed of edges with a locally high weight has the main disadvantage of being composed of many small scattered groups of edges when applied on the Paris subway.
This would probably be less the case for a bigger network with more edges.

Finally, a combination of the 3 previous filters was suggested. This time, the extracted backbone has the spine shape that is often expected from a backbone.
By combining multiple properties, this backbone is more representative than the 3 previous ones and allows to separate the lines into 2 groups. The first group of lines has few structurally important edges while the second group has many of them.

Overall, even if some extra information about the network was discovered, the amount of conclusions that can be made about the structure of the Paris subway with the sparsification analysis is quite low. This is probably due to the fact that the Paris subway is still a relatively small network and that its structure can already be apprehended without sparsifying it.
The sparsification methods would probably give better results when used on transportation networks that are larger than a subway network. If all public transport in and around Paris was considered, there would be hundreds of different lines and thousands of edges in the network. Using the sparsification methods above would then probably help to better understand the structure of the network while they don't seem to be necessary for smaller networks as the Paris subway.

Chapter 5

Robustness in Subway Networks

Robustness, the ability to withstand failures and perturbations, is a critical attribute of many complex systems including complex networks. [56]
This ability to resist random or targeted failures is a recurring field of study for multiple types of networks like power systems[57, 58], internet networks[59, 60], social networks[61, 62], ecological networks[63], transportation networks[64, 65, 66] and many others.

As to subway networks, robustness analyses are also common for cities like Madrid[4, 5], London[42], Paris[42] and Shanghai[9, 67].

Most of these existing papers analyze the robustness of subway networks based on robustness parameters like clustering coefficient, efficiency, connectivity, percolation limits and others[4]. While these parameters are important for all types of networks and thus also for subway networks, they are not specifically designed to reflect the impact of failures and perturbations on a subway network.

Using the modelling presented in chapter 2, this paper suggests to quantify the impact of a failure not only in terms of the usual parameters but also by analyzing how certain failures can increase or decrease the load on other stations or lines. A subway network is in those terms considered as robust when some failures can occur without overloading any lines (by increasing the IBCLines) or stations (by increasing the amount of transferring passengers and thus the amount of users in the station).

Different types of failures than those that are generally considered in robustness analyses are also considered. Link failures for example usually do not happen in subway networks the way they happen in router networks. While in router networks, links or nodes can fail anywhere at anytime, when a link fails in a subway network, the line going through that link will be interrupted and the whole line might thus fail. It thus makes more sense in a robustness analysis of a subway network to take out lines compared to taking out edges.
Sometimes a part of a subway line keeps on functioning when there is a link failure but we will not consider that scenario.
Similarly, stations in a subway network do not fail in the same way as nodes in a router network fail. Robustness to station failure should thus also take this into account.

This chapter starts with an introductory example that shows why load redistribution should be considered when analyzing failures. Results of line and station robustness in the subway network of Paris are then presented.

5.1 Line 14 and RER A failure in the Paris subway network

To illustrate the need to consider network load and passenger flows in the robustness analysis, this section uses an example which is the simultaneous failure of 2 lines in the Paris subway network.

When both line 14 and the RER A fail in the Paris subway network, the following impacts are seen on some global parameters[1]:

- The sum of the inputs of stations that are no longer connected to the network and thus the amount of disconnected passengers is 157 million which represents 9% of the total amount.
- The average degree decreases from 2.444 to 2.397 which is a decrease of 2%.
- The average length of a path between 2 stations or average closeness centrality becomes 15.57 which is a 4% increase.
- The average efficiency[4] of the network goes from 0.1234 to 0.1143 which is a decrease of 7%.
- Finally, the clustering coefficient[4] goes from 0.0187 to 0.0140 which is a more significant decrease of 25%.

With the exception of the clustering coefficient, all parameters change less than 10%. Overall, the structural impact of the failure of line 14 and RER A is limited on the network but this does not mean that there cannot be a high impact on passenger flows.

When lines experience failures in a network, many passengers that use those lines are redistributed over the other lines. If this redistribution is not well spread out across the other lines as is the case here, some lines can saturate.
The table shows how the IBCLine of all lines is impacted when the failure of line 14 and RER A occur at the same time:

Line	Full Network	Line failures	% diff	Line	Full Network	Line failures	% diff
1	**1757M**	**3643M**	**+107%**	8	1714M	1824M	+6%
2	1283M	1352M	+5%	9	1194M	1431M	+20%
3	727M	852M	+17%	10	495M	522M	+5%
3bis	11M	9M	-18%	11	512M	493M	-4%
4	1777M	1743M	-2%	12	751M	872M	+16%
5	1078M	1353M	+25%	13	1656M	1570M	-5%
6	1130M	1339M	+18%	14	1088M	/	/
7	1486M	1655M	+11%	RERA	2122M	/	/
7bis	36M	38M	+6%	RERB	1022M	893M	-13%

As can be seen, the IBCLine of line 1 more than doubles! Line 1 being already one of the busiest lines, we can assume that it will not be able to cope with the extra load.
If the failures happen during peak hours, line 1 will probably become completely saturated and its passengers will be highly impacted.

This may be an extreme case, but it shows the importance of considering the load on the remaining lines when analyzing failures.

Appendix C shows the visualization of the impact of the two failures on the passenger flows through the network.

[1] See chapter 2

5.2 Line Robustness

Now that the way the impact of failures should be quantified has been defined, line and station failures can be analyzed.
The first type of failures that will be considered is line failure. It is frequent in any subway network to have a line whose operation needs to be interrupted for some time and yet only few papers like [4] consider the robustness in terms of line failures.

The focus in this work is on what happens when one line is out of order. Analyzing the failure of multiple lines can be done in a similar way.

As a precision, when nodes are not connected to any lines anymore due to line failures, they are not considered in the network anymore for computation of parameters like average degree.
It is also important to note that most big subway networks have their lines sufficiently interconnected such that the subway network will never be disconnected into two groups of lines due to the failure of 1 or 2 lines[2]. The remaining lines will thus always form a connected network. This is especially the case in Paris where most lines have direct transfers with at least 6 or 7 other lines.

5.2.1 Global Parameters

Before moving on to the analysis of the passenger flows in the network, a few global parameters are first analyzed.
The following table shows some of these global parameters for the Paris subway network.

Deleted line	AvgDC	AvgCC	stdCC	LostPass
/	2.444	14.91	3.122	0M
1	2.389	15.10	3.107	74M
2	2.405	15.12	3.241	56M
3	2.414	15.00	3.168	56M
3bis	2.440	14.86	3.116	2M
4	2.4	15.16	3.147	84M
5	2.409	14.93	3.135	45M
6	2.401	15.17	3.354	52M
7	2.418	14.77	3.060	98M
7bis	2.438	14.82	3.124	5M
8	2.415	14.56	2.842	85M
9	2.414	14.60	3.053	100M
10	2.423	14.86	3.214	32M
11	2.413	15.01	3.155	20M
12	2.428	14.77	3.134	61M
13	2.432	14.75	3.105	99M
14	2.415	15.19	3.12	32M
RERA	2.427	15.15	3.112	125M
RERB	2.442	15.14	3.138	61M
RER	2.425	15.38	3.124	186M

- It can be seen in the first column of this table that whichever line is taken away, the average degree centrality of the network decreases[3]. This shows how much the multiple lines are interconnected in Paris. If there was a long line in the network with few transferring stations,

[2]This is not always the case, example: Madrid
[3]The average degree centrality being computed on the remaining nodes and thus not on the disconnected nodes.

deleting it would increase the average degree centrality. In Madrid for example where the interconnectivity of lines is lower, deleting some lines increases the average degree centrality.

The more the removal of a line decreases the average degree centrality, the more that line is interconnected with the rest of the network.
Line 1 which is the most interconnected one has 12 stations out of 25 where a transfer is possible with another line!

- While it seems at first glance that the higher the average closeness centrality or the average path length between 2 stations becomes after a line has been deleted, the higher the structural importance of that line is, there is a catch.
Deleting line 8 for example reduces the average path length the most because it is a line that stretches far out of the city. The average path length is thus not decreased because the line is unimportant but because stations that are far from the center and that thus have a high closeness centrality are disconnected. They are thus not counted anymore when computing the average closeness centrality which therefore decreases.
At the other extreme, deleting lines 14 or 6 that do not stretch far from the center at all, increases the average path length between stations the most.

  Even though the structural importance of a line will have an impact on the average closeness centrality when the line is removed, the results are too biased by how much the line stretches out from the center of the network to make proper conclusions.

- The same argument as for the average closeness centrality is also valid for the standard deviation of the closeness centrality which is obviously very dependent on the outliers that are off center in the network.

- Finally, the amount of lost passengers is simply the sum of the inputs of all the stations that have been disconnected. Those passengers thus do not have access to the network anymore.
The maximum amount of disconnected passengers happens when the RER A is out of order and consists of 125M passengers on a yearly basis which is 7% of the total. The subway line that disconnects the most passengers is line 9 with 100M passengers which represents 6% of the total.

  For line 9 that was shown to be structurally not so important in the previous chapter, this significant amount of disconnected passengers can actually be considered the main impact of its removal!

Everything considered, the differences on the global parameters are always smaller than 10% when a single line is removed.

5.2.2 Impact on other lines

The conclusion that can be drawn from the global parameters, is that removing 1 line does not have too much of an impact on them. The main drawback of the removal of some lines seems to be the amount of passengers that get disconnected from the network.

As shown in the introductory example, the fact that the impact on global parameters is small does not mean that the overall impact is small since there can be significant impacts on the passenger flows.
Therefore, the following table shows the IBCLine values of all lines when any line is removed[4].

[4]It is assumed that the passengers that only have access to the removed line find another way of transport. Lost passengers are thus not relocated to other surrounding stations. In reality some of them would find a way to get to another subway station. The input of some stations might then be increased slightly which would in turn increase some IBCLines. Taking into consideration the redistribution of station input in case of removed lines might be done in further work.

Line ↓ \ Del. Line →	/	1	2	3	3bis	4	5	6	7	7bis	8
1	**1757**	/	1792	1823	1757	1783	1749	1928	1811	1759	1672
2	**1283**	1195	/	1236	1283	1352	1235	1256	1273	1265	1374
3	**727**	772	873	/	713	763	728	733	736	729	759
3bis	**11**	10	7	48	/	11	11	11	11	11	11
4	**1777**	1805	1812	1791	1777	/	1900	1930	1809	1776	1978
5	**1078**	1071	1205	1067	1077	1206	/	1102	910	1072	1048
6	**1130**	1423	1098	1154	1131	1290	1182	/	1001	1133	1023
7	**1486**	1552	1722	1495	1487	1573	1747	1631	/	1482	1463
7bis	**36**	37	39	43	36	37	33	37	36	/	37
8	**1714**	1848	1759	1689	1710	1884	1721	1738	1814	1711	/
9	**1194**	1313	1296	1194	1194	1213	1190	1257	1220	1192	1499
10	**495**	553	504	501	495	502	502	610	515	496	621
11	**512**	564	557	640	516	483	573	515	518	501	564
12	**751**	735	804	759	751	1036	754	817	774	751	755
13	**1656**	1590	1676	1680	1661	1762	1681	1712	1686	1661	1751
14	**1088**	1133	1134	1039	1089	1199	1124	1027	1091	1089	959
RERA	**2122**	2564	2182	2111	2122	2085	2140	2078	2091	2123	2166
RERB	**1022**	988	1080	1029	1022	1057	1007	1089	1026	1020	1049

Line ↓ \ Del. Line →	/	9	10	11	12	13	14	RERA	RERB	RER
1	**1757**	1652	1833	1748	1753	1703	2216	2996	1841	2797
2	**1283**	1337	1306	1419	1303	1158	1365	1282	1424	1406
3	**727**	741	728	791	728	743	904	767	802	780
3bis	**11**	11	11	30	11	11	12	9	11	9
4	**1777**	1771	1736	1798	1576	1698	1823	1750	2282	2209
5	**1078**	1073	1085	1126	1086	1124	1259	1200	1326	1396
6	**1130**	1075	1161	1153	1037	1294	1260	1167	1131	1201
7	**1486**	1467	1474	1513	1501	1518	1627	1493	1604	1533
7bis	**36**	38	37	61	36	36	37	38	37	38
8	**1714**	1717	1488	1680	1717	1733	1752	1738	1799	1838
9	**1194**	/	1472	1179	1208	1195	1261	1334	1276	1371
10	**495**	399	/	493	496	506	513	511	492	501
11	**512**	529	510	/	506	523	476	516	485	498
12	**751**	755	762	769	/	949	737	753	757	872
13	**1656**	1655	1655	1696	1686	/	1803	1750	1700	1553
14	**1088**	1062	1060	1080	1068	934	/	777	699	766
RERA	**2122**	2089	2156	2119	2159	2221	2533	/	2261	/
RERB	**1022**	1052	1033	1051	1027	1061	903	941	/	/

While removing each line has its specific impact, passenger flows are clearly not distributed in a uniform way over the other lines. Removing line 1 for example increases the IBCLine of line 6 and RER A significantly but decreases the IBCLine of line 2 which thus becomes less busy when line 1 is removed.
Some comments are to be made about this table:

- Removing some lines has only a very limited impact on other lines. This is the case of line 9 for example.
  On the other hand, removing line 14 has a serious impact on many lines.
- Even though removing line 1 does not have a big impact on other lines, removing other lines can have a big impact on line 1. When the RER A is removed, the occupation of line 1 almost

doubles. Since line 1 is already one of the busiest ones, if the RER A has a technical difficulty during peak hour, line 1 will be completely saturated and the same applies in a lesser measure if line 14 is removed.
When both the RER A and line 14 fail at the same time, the load on line 1 becomes even higher as seen in the introductory example.

- Similarly, in case of a failure of the RER B, the load on line 4 increases significantly.
- Further work could try to compute threshold values for the IBCLine of each line above which the line would saturate. These values should be different for each line but would allow to state more accurately which failures will cause a saturation in the network.
- This table was computed using the usual transfer penalty of 2 stops. For different penalties, the impact of line removals will also be different. The table of IBCLines in case of line removals when the transfer penalty is of 4 stops can be found in appendix D.

Based on this table, it is also possible to react accordingly to a line failure by increasing the frequency of the most impacted lines. When line 1 is removed for example, it does not make sense to increase the frequency of line 2 because its occupation decreases but it does make sense to increase the frequency of line 6 if possible.

The current strategy of the RATP[5] is to replace the subway lines by busses in case of failure. This is useful for the passengers that are completely disconnected but passengers that can use another subway route, which will often be possible when the network is dense like in Paris, will choose to do so instead.

It is thus suggested to respond to line failures with busses but also by increasing the frequency of some targeted lines.

5.2.3 Impact on stations

To be complete on the impact a line failure has, the impact on stations should also be considered.
As a reminder, the load of a station is determined by its input, but mostly by the amount of transfers that are done at that station. This amount of transfers will change when lines experience failures as shown in the table.

Overall, the amount of passengers doing a transfer at each station does not change too much when lines experience failures. When a significant change occurs, it is often a decrease in amount of transfers which does not cause any harm.
In some cases though, the failure of a line increases the amount of transfers significantly in certain stations. The failure of line 2 for example increases the amount of transfers in 'Saint-Lazare', 'Opéra' and 'Gare de l'Est' by up to 50%. These kind of transfer increases in a station should be anticipated for organisational and security reasons.

[5] The Paris subway operator

Station ↓ \ Del. Line →	Normal	1	2	3	4	5	6	7	8	9
Châtelet	**378M**	368M	408M	384M	265M	415M	410M	356M	414M	385M
Saint-Lazare	**134M**	151M	**175M**	91M	143M	136M	134M	134M	141M	131M
Gare de Lyon	**128M**	58M	125M	124M	141M	133M	151M	132M	103M	112M
Nation	**120M**	99M	80M	117M	125M	125M	114M	118M	117M	59M
CDG-Etoille	**103M**	81M	51M	104M	101M	102M	68M	109M	112M	106M
Montparnasse-B.	**101M**	95M	103M	102M	35M	101M	111M	102M	110M	100M
République	**87M**	99M	108M	59M	111M	62M	87M	94M	64M	73M
La Motte P. G.	**86M**	89M	88M	85M	94M	86M	54M	87M	16M	79M
Opéra	**86M**	115M	**128M**	58M	106M	82M	87M	45M	53M	99M
Madeleine	**71M**	73M	73M	84M	116M	73M	65M	78M	14M	64M
Bercy	**56M**	88M	57M	55M	58M	62M	0M	53M	45M	58M
Place d'Italie	**56M**	61M	56M	56M	57M	33M	25M	8M	54M	54M
Stalingrad	**54M**	49M	28M	54M	92M	27M	54M	11M	56M	56M
Gare du Nord	**51M**	53M	62M	51M	32M	25M	50M	45M	53M	53M
Place de Clichy	**51M**	56M	0M	59M	52M	48M	52M	50M	56M	53M
Barbes-Rochechouart	**49M**	50M	0M	48M	0M	60M	49M	51M	51M	51M
Gare de l'Est	**48M**	53M	**70M**	48M	38M	38M	48M	8M	47M	43M
Denfert-Rochereau	**47M**	52M	48M	49M	41M	45M	25M	45M	45M	46M
Bastille	**44M**	11M	48M	44M	48M	13M	57M	48M	28M	47M
Champs-Elysées-Cl.	**38M**	0M	44M	42M	38M	39M	47M	41M	43M	43M

Station ↓ \ Del. Line →	Normal	10	11	12	13	14	RERA	RERB	RER
Châtelet	**378M**	378M	337M	372M	376M	287M	286M	233M	186M
Saint-Lazare	**134M**	135M	135M	122M	64M	69M	112M	110M	105M
Gare de Lyon	**128M**	124M	133M	124M	114M	16M	42M	53M	43M
Nation	**120M**	120M	127M	121M	122M	139M	82M	129M	83M
CDG-Etoille	**103M**	104M	106M	98M	114M	108M	67M	113M	82M
Montparnasse-B.	**101M**	107M	98M	69M	47M	116M	110M	103M	93M
République	**87M**	86M	57M	88M	90M	91M	95M	91M	97M
La Motte P. G.	**86M**	28M	85M	83M	97M	78M	82M	81M	88M
Opéra	**86M**	83M	90M	87M	89M	184M	50M	132M	51M
Madeleine	**71M**	56M	71M	53M	77M	8M	35M	32M	37M
Bercy	**56M**	55M	55M	56M	55M	0M	32M	31M	34M
Place d'Italie	**56M**	58M	56M	55M	56M	63M	62M	62M	67M
Stalingrad	**54M**	54M	58M	54M	51M	55M	56M	50M	53M
Gare du Nord	**51M**	51M	60M	47M	55M	43M	54M	14M	16M
Place de Clichy	**51M**	51M	54M	67M	0M	50M	50M	52M	46M
Barbes-Rochechouart	**49M**	51M	48M	50M	41M	55M	59M	59M	66M
Gare de l'Est	**48M**	47M	49M	46M	52M	53M	47M	63M	59M
Denfert-Rochereau	**47M**	46M	48M	46M	48M	56M	59M	11M	14M
Bastille	**44M**	48M	47M	45M	44M	64M	53M	66M	69M
Champs-Elysées-Cl.	**38M**	39M	38M	41M	0M	59M	42M	43M	30M

5.2.4 Simultaneous failures

When multiple failures occur at the same time, the trend is similar as can be guessed with the introductory example.
When two lines fail, the general parameters of the network still mostly do not change much but the differences in load of lines and stations can become more significant by addition.
If line 14 and the RER A fail at the same time, the IBCLine of line 1 goes even higher, if line 1 and 2 experience simultaneous failures, the amount of passengers transferring at 'Opéra' and 'Saint-Lazare' might also become problematic.

5.2.5 Summary

When one line has a failure, the subway network is structurally not too impacted due to the high amount of lines and their very high inter-connectivity in Paris. Yet, there are two types of issues that can arise:

Firstly, many passengers are disconnected from the network when some lines experience failures. This can be partially solved by placing temporary bus lines.

Secondly, in some situations the removal of one line can saturate another line. That particular line is then not able to keep up with the incoming passenger flow anymore. If possible, the frequency of the saturated line should be increased to cope with the problem.

The impact on stations will not be a problem as long as it is anticipated.

5.3 Station Robustness

In most types of networks, the failure of a node comes with the failure of all its incident edges. In a router network for example, when a router is down, so are all the links connected to the router.
In a subway network, things are different. One of the tracks in a station can be blocked in which case only one of the lines is interrupted while the others can continue their operation. We are then in a line failure scenario.
Another possibility is for the station to be closed. The lines can then go through the station but no passenger can use the station. This means that the incoming and outgoing passengers need to find another station and that no transfers can be done at that station anymore for some time.

As an example of such a situation, the subway station of 'Schuman' in Brussels is sometimes closed when there is a European summit[68]. The subway lines then go through the station but do not stop. It is thus impossible to get out at the station or to operate a transfer.

This second possibility is what we will consider as station failure. As with the lines, the focus is on a situation where a single failure occurs.

5.3.1 Direct consequences

When a station needs to closed, there is one obvious consequence: all passengers leaving from or going to that station are affected. The effect will depend on the input of the station where the failure occurred.
For some stations, this impact can already be significant. 'Gare du Nord' has an input of 88 million. The output of the station is in the model assumed to be 88 million as well which means that 176 million passengers are directly impacted. This is almost 10% of the total amount of passengers!

The stations with highest input[6] in 2018[43] were the following ones:

Rank	Station	Input	Rank	Station	Input
1	Gare du Nord	88M	11	Charles de Gaulle-Etoille	17M
2	Gare de Lyon	75M	12	Opéra	17M
3	La Défense	50M	13	Bastille	13M
4	Saint Lazare	47M	14	Nation	13M
5	Châtelet	46M	15	Belleville	12M
6	Montparnasse-Bienvenue	31M	16	Place d'Italie	12M
7	Gare de l'Est	21M	17	Hotel de Ville	11M
8	Bibliothèque F.M.	19M	18	Franklin Rooseveld	11M
9	Les Halles	18M	19	Palais Royal	10M
10	République	18M	20	Gare d'Austerlitz	10M

While the numbers can get high, they are to be put into perspective. Subway stations are rarely too distant from each other especially in Paris. In case of a closed station, most passengers can easily walk towards another station. While being an inconvenience for passengers, alternatives exist.

The second impact of a closed station is for passengers that use the station to operate a transfer. While this has an impact on the passengers, it can also have a serious impact on the passenger flows through the network due to passengers having to choose different routes. The load of lines and stations can thus be changed.
The stations whose failures are likely to cause the most disruption in the network are the following ones in which it was estimated in chapter 2 that the most transfers were made:

[6] See chapter 2.2.1

Rank	Station	NbTransf.	Rank	Station	NbTransf.
1	Châtelet	378M	11	Bercy	56M
2	Saint-Lazare	134M	12	Place d'Italie	56M
3	Gare de Lyon	128M	13	Stalingrad	54M
4	Nation	120M	14	Gare du Nord	51M
5	Charles de Gaulle-Etoile	103M	15	Place de Clichy	51M
6	Montparnasse-Bienvenue	101M	16	Barbes - Rochechouart	49M
7	République	87M	17	Gare de l'Est	48M
8	La Motte Picquet G.	86M	18	Denfert-Rochereau	47M
9	Opéra	86M	19	Bastille	44M
10	Madeleine	71M	20	Champs-Elysées-Clém.	38M

5.3.2 Failure of 'Châtelet'

'Châtelet' is by far the most important station in the subway network of Paris. It has the highest degree centrality, the smallest closeness centrality and the highest IBC. It is also said to be the largest underground station in Europe[69]!
378 million passengers operate a transfer at 'Châtelet' yearly which is thrice as much as any other station in the Paris subway and 20% of the total amount of passengers in the whole network. When the station is closed, all those passengers will need to take an alternative route which might have consequences.

The following results are seen when 'Châtelet' is closed[7].

- The average closeness centrality which, as a reminder, is the average path length in stops between any 2 stations becomes 15.40 which is an increase of 0.49 compared to the previous value of 14.91. This increase is not to be neglected but for a station of such an importance in the network, the increase is reasonable.
  Paths that usually have transfers at 'Châtelet' are now on average 2.5 stops longer[8].

  This limited impact shows once more that the Paris network is very dense in its center. An enormous amount of passengers go through 'Châtelet' because that is the fastest way for them but most of them have other options that are only slightly longer.

- The impact on passengers thus seems acceptable but the impact on line loads is also important:

Line	IBCLine	IBCLine\Châtelet	Line	IBCLine	IBCLine \Châtelet
1	1757M	1673M	8	1714M	1925M
2	1283M	1437M	9	1194M	1366M
3	727M	837M	10	495M	551M
3bis	11M	12M	11	512M	317M
4	1777M	1526M	12	751M	909M
5	1078M	1524M	13	1656M	1904M
6	1130M	1638M	14	1088M	984M
7	1486M	1487M	RERA	2122M	1871M
7bis	36M	43M	RERB	1022M	559M

[7]For modelling purposes, the station is still considered as a normal station apart from its input that is put to 0 and transfers that are not allowed. Paths that go through the station do not have their length decreased by 1 even though they do not stop at that specific station. A path going through 'Châtelet' but without changing lines there will thus have the same length whether 'Châtelet' is closed or not.

[8]20% of the paths have transfers at 'Châtelet' and are thus impacted. Those 20% of paths must thus increase in length of around 2.5 stops on average to increase the global average of 0.5.

The results show that the lines that go through 'Châtelet' (1,4,7,11,14,RER A and RER B) have a decreased IBCLine while the others have an increased IBCLine. Lines 5 and 6 see their IBCLine increase the most.
Lines 8 and 13 are already busy thus the impact is more important on them even though the difference is smaller.

- Concerning the stations, the main differences are at:

Station	Nbtransfers	Nbtransfers\Châtelet
Gare de Lyon	128M	164M
Denfer-Rochereau	47M	102M
Bastille	44M	90M
Bercy	56M	75M

Even though these differences are large especially for 'Bastille' and 'Denfert-Rochereau', having only 2 stations overloaded due to the failure of 'Châtelet' is a good result. As said before, the most important is to anticipate the fact that more people are going to use certain stations.

Overall, even though 'Châtelet' is the station with the highest traffic in the network, its removal would cause only limited harm.

5.3.3 Failure of another station

Even though the subway network does not seem to be dependent on 'Châtelet', the failure of some other important stations need to be considered in order to say that the network is not dependent on any station.
A summary of the impact of the closure of the other stations with more than 100 million transfers is given here.

/	Normal	\Saint-Lazare	\Gare de Lyon	\Nation	\CDG-Etoile	\Montparnasse-B.
CC	14.91	15.08	14.95	**15.24**	15.14	15.08
1	1757M	**2017M**	1886M	1752M	**1961M**	1731M
2	1283M	1341M	1330M	1043M	1011M	1316M
3	727M	722M	753M	802M	799M	732M
3bis	11M	11M	11M	7M	11M	11M
4	1777M	1832	1741M	1808M	1790M	1187M
5	1078M	1100M	1061M	1242M	1092M	1092M
6	1130M	1163M	1224M	1051M	944M	1235M
7	1486M	1507M	1533M	1516M	1527M	1513M
7bis	36M	37M	37M	36M	36M	37M
8	1714M	1750M	1690M	1814M	1764M	1825M
9	1194M	1233M	1248M	**1554M**	1302M	1205M
10	495M	509M	490M	492M	500M	559M
11	512M	507M	511M	641M	516M	476M
12	751M	697M	736M	754M	801M	894M
13	1656M	1726M	1673M	1663M	1739M	1481M
14	1088M	656M	799M	1038M	1139M	1174M
RER A	2122M	2201M	2151M	2023M	1908M	2146M
RER B	1022M	996M	976M	1029M	1032M	1099

With a few exceptions, the closeness centrality and the IBCLines are only slightly impacted by the removal of any of these stations with the main issue being on line 1 when 'Saint-Lazare' is removed.

Concerning the load of stations, the results are the same as for 'Châtelet' but in lesser measure. The flow is redistributed mainly over 2 or 3 other stations that thus get more crowded which should be anticipated.

5.3.4 Multiple station failures

As with lines, when multiple stations fail in a subway network, their impacts pile up. A failure of 3 important stations ('CDC-Etoile', 'Saint-Lazare' and 'Gare de Lyon') for example, brings the closeness centrality up to 15.37 but more importantly might saturate line 1.

An example of multiple station closures at the same time is given in the next chapter.

Chapter 6

Applications

In this document, a new type of subway network modelling is introduced that takes into account the load of the network by using the input of each station and the fact that the network is composed of lines by penalizing transfers in the length of paths.
This new modelling is then used to compute the passenger flows through the network, to create a backbone and to have a different approach on robustness in a subway network.

In this final chapter, a few more applications are shown.

6.1 Extension of line 14 in the Paris subway network

Figure 6.1: Planned extension of line 14[70]

While it is possible to analyze what happens in the network when a line has a failure, it is also possible to analyze what will happen when a new line is added to the network or an existing line is extended.

One of the busiest lines in the network is line 13 with an IBCLine of 1656M. This high IBCLine

is mostly due to the northern part of the line where the edges are the busiest especially before the separation in 2 branches.
While those line segments transport less passengers than some from other lines like line 14, line 14 is a modern and efficient automatic subway with a high frequency while line 13 is an older subway that cannot have as high a frequency and that often experiences technical difficulties. Due to these, the time between two vehicles can get high and the vehicles are often overcrowded.
The result is that line 13 and especially its northern part is often considered to be the worse part in the Paris subway[71, 72] and passengers often have a bad experience!

To solve this issue, it was decided by the RATP to extend line 14 to the north by the end of 2020. They announced 3 main goals with this extension[70]:

1. *Serve existing and future districts, notably the Clichy-Batignolles sector, and the Clichy-la-Garenne and Docks de Saint-Ouen joint development zone (ZAC).*
2. *Strengthen the transfers and transfer points between various transport modes.*
3. *Find a sustainable solution for reducing the saturation of the M13*

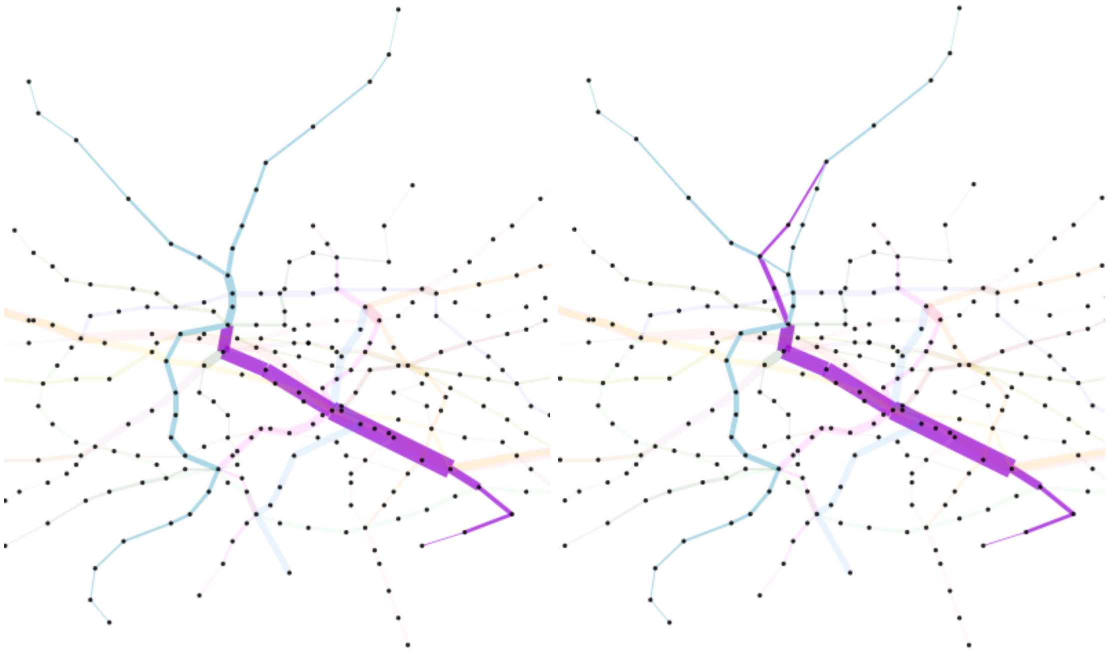
These claims can be verified by modelling the extended network and comparing it with the original one.

1. Obviously, by extending a subway line and creating new stations, more passengers will have access to the network.
2. By extending line 14, the average closeness centrality will go from 14.91 to 14.8. For this global parameter to decrease in such a way, it is necessary for the local closeness centralities to decrease significantly. The objective is thus met as well.
   The closeness centrality of stations is reduced to up to 3 stops for 'Mairie de St Ouen'.
3. Finally, to verify the third and main goal which is reducing the load on line 13, the IBCLine[1] and the passenger flows figures[2] are used.
   As a reminder, the edge thickness on the figure represents the passenger flow through the edge.

Line	L14 Extension	Old IBCLine	Line	L14 Extension	Old IBCLine
1	1754M	1757M	8	1735M	1714M
2	1245M	1283M	9	1181M	1194M
3	730M	727M	10	496M	495M
3bis	11M	11M	11	523M	512M
4	1768M	1777M	12	755M	751M
5	1075M	1078M	13	**1332M**	**1656M**
6	1127M	1130M	14	**1367M**	**1088M**
7	1492M	1486M	RERA	2123M	2122M
7bis	36M	36M	RERB	1029M	1022M

[1]Due to the input of the 2 future new stations being yet unknown, an input of 5 million is given to them for the modelling. This is an average value in the network and is also similar to the surrounding stations.

[2]These figures are slightly different from the usual ones used in this document because the 2 branches of line 13 are depicted on them instead of the usual merger of them. The merger of the 2 branches originally has no impact on the network but this is not valid anymore after the extension since line 14 crosses both branches. As to not alter the results in this section, both branches are thus shown.



(a) Before extension

(b) After extension

Figure 6.2: Impact of line 14 (purple) extension on line 13 (blue)

Clearly, the extension has a negligible impact on the network apart from on line 13 and line 14. The IBCLine of line 13 decreases by 20% and this decrease is mainly visible in the stops that were the busiest before the extension. The IBC of the busiest edges decrease by almost 50%. The IBC of the edge between 'La fourche' and 'Place de Clichy' for example goes from 124M to 65M!

The increase of the IBCLine of line 14 is not an issue since it can clearly be seen on the figure that this increase is due to the extra stops and not to the increase in IBC of existing edges.

It can thus be concluded that all 3 of the claims of the RATP are verified!

6.2 RATP deconfinement plan in May 2020

While in the previous chapter the focus was on removing one station from the network, multiple stations can be removed at once for maintenance reasons for example.
In some situations, the amount of closed stations can even become significant as was the case in the Paris subway network in May 2020.

With the purpose of limiting the amount of passengers that use the subway, multiple decisions were taken starting the 11th of May like diminishing the frequency of some lines and closing around 60 different stations[73, 74, 40] amongst which some important ones like 'République' where 5 subway lines cross each other.
Just as when lines are removed, the removal of stations can easily be implemented in the model to obtain the new passenger flows. It is important to mention that the computations were done based on the usual amount of passengers in the network while the real amount of passengers during that period was lower than it usually is. The numbers should thus be compared in a relative and not absolute way.

These 60 stations have a combined input of 244 million out of the total 1812 million. 13.5% of the passengers thus had to walk to another station or find another transportation method.
The average closeness centrality increased from 14.9 to 15.6 which shows that most removed stations are not important transferring points and the IBCLines changed in the following way:

Line	Deconfinement	Normal IBCLine	Line	Deconfinement	Normal IBCLine
1	1981M	1757M	8	870M	1714M
2	832M	1283M	9	1107M	1194M
3	674M	727M	10	485M	495M
3bis	12M	11M	11	500M	512M
4	1815M	1777M	12	713M	751M
5	828M	1078M	13	1589M	1656M
6	1024M	1130M	14	1053M	1088M
7	1452M	1486M	RERA	2085M	2122M
7bis	45M	36M	RERB	1058M	1022M

Due to the many disconnected stations and thus the lower total input, most IBCLines decreased or remained similar with the exception of line 1 where the IBCLine increased.

Overall, it can be concluded that apart from the direct impact on disconnected passengers, the impact on the network was limited as can also be seen on the 2 figures on the next page that are similar except for the removed nodes.
Each color on the figure represents one line, the edge thickness is the yearly passenger flow through the edge and the node thickness is the yearly amount of passengers using the station.

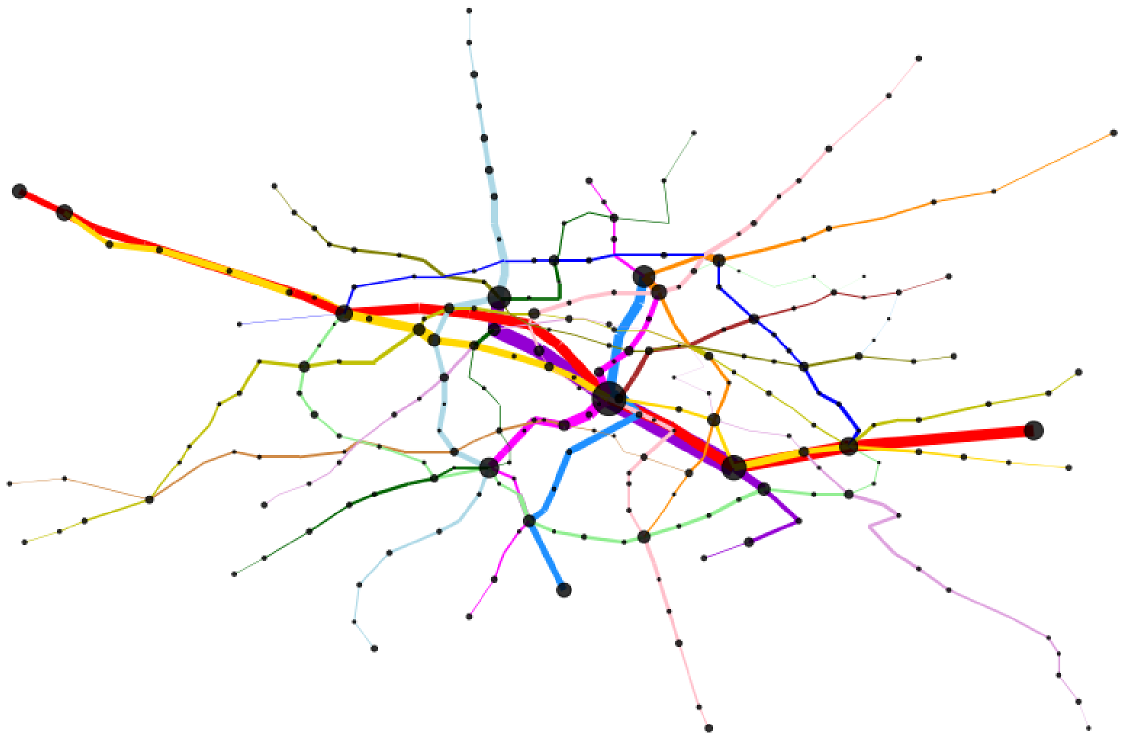


Figure 6.3: Deconfinement plan

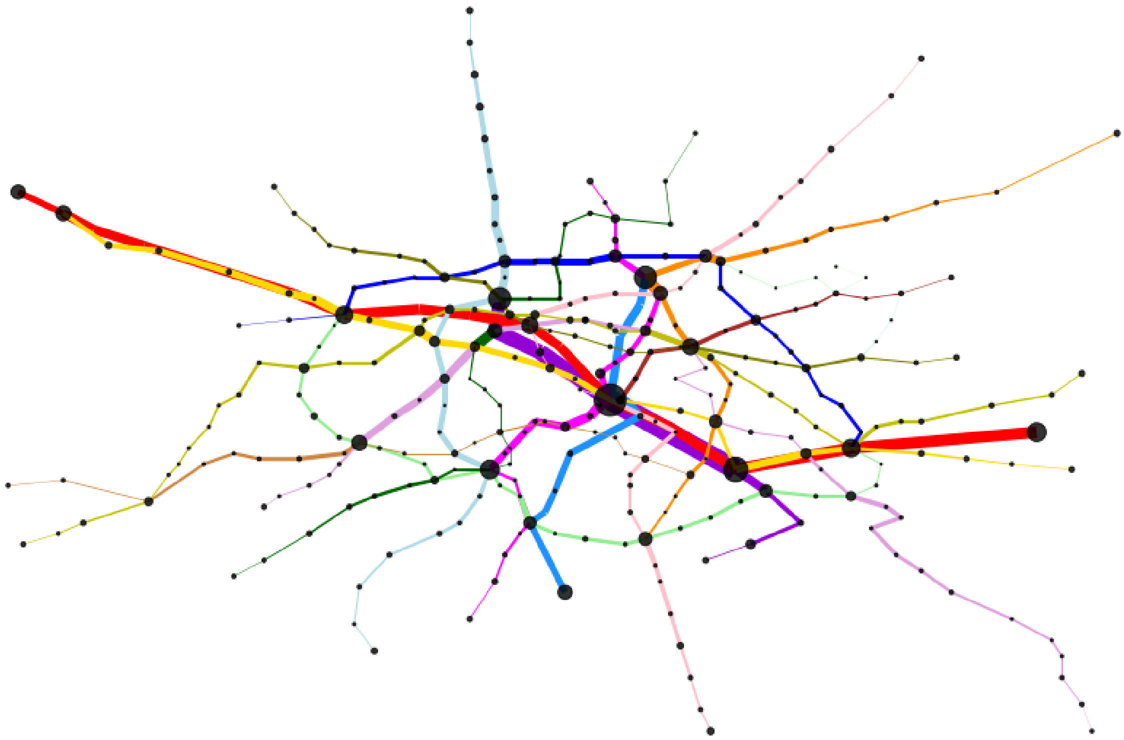


Figure 6.4: Normal Situation

Conclusion

Efficient urban rail transit systems are a pillar of sustainable cities. Analyzing and understanding them is essential.

In this work, a new way of computing passenger flows in subway networks was presented that uses a new centrality measure as well as a multilayer network.
The betweenness centrality measure is fit to compute passenger flows but has the main disadvantage of treating each path equally while paths in subway networks do not all carry the same amount of passengers. To address this issue, we suggested to approximate the passenger betweenness centrality or PBC as defined by [9] by the input betweenness centrality or IBC that can be used more broadly. As long as the input of stations is available, the IBC can be used while the PBC requires the amount of passengers travelling between any 2 stations to be known.
Further work could focus on the estimation of station input when it is not available either. This could then be used to estimate the IBC of future networks before their construction.

To take transfer times between lines into account when computing the path a passenger will take between stations, we suggested to model a subway network as a multilayer network instead of a single layer network. This allows to add a length penalty for each transfer in a path.
We estimated the penalty based on the walking time between platforms but also on the average waiting time for the new vehicle. It will be dependent on the line frequencies and thus not be fixed in time.

While this new way of computing passenger flows was mainly used on the Paris subway in this work, the passenger flows of any subway can be analyzed with it. In order to be able to use it for any transportation system that is based on lines, future work should address the fact that neither the stop lengths nor the transfer penalties are constant. In this work, we only differentiated the subway lines and the RER lines.

Using this new methodology, the centrality results including the passenger flows through the network were computed. Lines were also compared based on their IBCLines which we defined as the total amount of stops that were done by passengers on a specific line.
It was shown based on available numbers of line frequenting in the Paris subway that more accurate results are obtained when using transfer penalties and station input.

In chapter 4, a backbone of the Paris subway network was obtained by combining 3 different sparsification filters: the weight filter with the IBC of edges as weights, the Salience filter and the disparity filter with the IBC of edges as weights as well.
To deal with the different results of each filter, the edges with the highest significance score of each one were extracted to create the backbone.

We found that this backbone separated the lines into 2 groups. Lines 1,4,5,7,8,11,13,14,RER A

and RER B had more than 70% of their edges extracted in the backbone. Lines 2,3,3bis,6,7bis,9,10 and 12 on the other hand had less than 41% of their edges extracted.
Since a subway network can not only be seen as a set of edges but also as a set of lines, a backbone consisting of all the lines of the first group was also suggested.

Due to their relative small size, prioritizing a backbone analysis is not recommended for subway networks. We expect that more insights can be gained from this analysis in larger transportation networks especially if it combines multiple types of transport modes.

The impact of line and station failures on the passenger flows in the Paris subway network was then analyzed. It was concluded that one line failure had a small impact on the structural properties of the network but that some failures could still have a significant impact on the network by saturating lines.
It was also shown that the failure of ‘Châtelet’, which has the highest traffic in the network, only has a limited impact on the passenger flows through the network.

Finally, the extension of line 14 of 2020 as well as the deconfinement plan of May 2020 in the Paris subway network were analyzed as applications. The extension of line 14 will decrease the load on line 13 as claimed by the RATP and the deconfinement plan will not have a significant impact on the passenger distribution in the network.

Two examples were given but the proposed modelling and passenger flow computation can be used on any subway network to estimate the impact of removals or extensions. Subway operators can for example use it during strikes to compare multiple ways of operating the network with fewer employees.

Bibliography

[1] Shannon Bouton, Stefan M Knupfer, Ivan Mihov, and Steven Swartz. Urban mobility at a tipping point. *McKinsey and Company*, 2015.

[2] Low emission zone brussels. https://lez.brussels/mytax/en/.

[3] Diesel vehicles in paris and europe. https://www.autoplus.fr/actualite/Diesel-Interdiction-Paris-Voitures-Europe-1545947.html.

[4] Elisa Frutos Bernal and Angel Martín del Rey. Study of the structural and robustness characteristics of madrid metro network. *Sustainability*, 11(12):3486, 2019.

[5] Mary Luz Mouronte. Topological analysis of the subway network of madrid. In *International Multi-Conference on Computing in the Global Information Technology, Seville, Spain*, pages 9–13, 2014.

[6] Sybil Derrible. Network centrality of metro systems. *PloS one*, 7(7), 2012.

[7] Xi Wang, Qingming Zhan, Philippe Bonnin, and Stephane Douady. Centrality analysis of metro network, a case study of paris. In *5th International Conference on Civil Engineering and Transportation*. Atlantis Press, 2015.

[8] Weisi Guo and Xueke Lu. London underground: Neighbourhood centrality and relation to urban geography. In *2016 IEEE International Smart Cities Conference (ISC2)*, pages 1–7. IEEE, 2016.

[9] Daniel Jian Sun and Shituo Guan. Measuring vulnerability of urban metro network from line operation perspective. *Transportation Research Part A: Policy and Practice*, 94:348–359, 2016.

[10] Line 14 extension goals. https://www.ratp.fr/en/groupe-ratp/metrotrains/m14-extension-mairie-de-saint-ouen.

[11] Austin transit expansion. https://www.fox7austin.com/news/capital-metro-unveils-transit-system-recommendations-including-light-rail-austin-underground.

[12] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 07 2014.

[13] Stefano Boccaletti, Ginestra Bianconi, Regino Criado, Charo I Del Genio, Jesús Gómez-Gardenes, Miguel Romance, Irene Sendina-Nadal, Zhen Wang, and Massimiliano Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122, 2014.

[14] Visualizing networks. http://www.mkivela.com/pymnet/visualizing.html. Accessed: 2020-04-10.

[15] K Lewis, J Kaufman, M Gonzalez, A Wimmer, and NA Christakis. Tastes, ties, and time: A new (cultural, multiplex, and longitudinal) social network dataset using facebook. com in 'social networks. *An international Journal of analysis', Forthcoming*, 2008.

[16] Michael Szell, Renaud Lambiotte, and Stefan Thurner. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*, 107(31):13636–13641, 2010.

[17] Michael Szell and Stefan Thurner. Measuring social dynamics in a massive multiplayer online game. *Social networks*, 32(4):313–329, 2010.

[18] Michele Berlingerio, Michele Coscia, Fosca Giannotti, Anna Monreale, and Dino Pedreschi. Multidimensional networks: foundations of structural analysis. *World Wide Web*, 16(5-6):567–593, 2013.

[19] Michele Berlingerio, Fabio Pinelli, and Francesco Calabrese. Abacus: frequent pattern mining-based community discovery in multidimensional networks. *Data Mining and Knowledge Discovery*, 27(3):294–320, 2013.

[20] Emanuele Cozzo, Mikko Kivelä, Manlio De Domenico, Albert Solé-Ribalta, Alex Arenas, Sergio Gómez, Mason A Porter, and Yamir Moreno. Structure of triadic relations in multiplex networks. *New Journal of Physics*, 17(7):073029, 2015.

[21] Alessio Cardillo, Jesús Gómez-Gardenes, Massimiliano Zanin, Miguel Romance, David Papo, Francisco Del Pozo, and Stefano Boccaletti. Emergence of network features from multiplexity. *Scientific reports*, 3:1344, 2013.

[22] Pablo Kaluza, Andrea Kölzsch, Michael T Gastner, and Bernd Blasius. The complex network of global cargo ship movements. *Journal of the Royal Society Interface*, 7(48):1093–1103, 2010.

[23] Balakrishnan Kannan Sunil Kumar Raghavan Unnithan and Madambi Jathavedan. Betweenness centrality in some classes of graphs. https://www.hindawi.com/journals/ijcom/2014/241723/, 2014. Accessed: 2020-04-10.

[24] Nicolas Kourtellis, Tharaka Alahakoon, Ramanuja Simha, Adriana Iamnitchi, and Rahul Tripathi. Identifying high betweenness centrality nodes in large social networks. *Social Network Analysis and Mining*, 3(4):899–914, 2013.

[25] K-I Goh, Eulsik Oh, Byungnam Kahng, and Doochul Kim. Betweenness centrality correlation in social networks. *Physical Review E*, 67(1):017101, 2003.

[26] Miriam Baglioni, Filippo Geraci, Marco Pellegrini, and Ernesto Lastres. Fast exact computation of betweenness centrality in social networks. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 450–456. IEEE, 2012.

[27] Shuangshuang Jin, Zhenyu Huang, Yousu Chen, Daniel Chavarría-Miranda, John Feo, and Pak Chung Wong. A novel application of parallel betweenness centrality to power grid contingency analysis. In *2010 IEEE International Symposium on Parallel & Distributed Processing (IPDPS)*, pages 1–7. IEEE, 2010.

[28] Ettore Bompard, Enrico Pons, and Di Wu. Extended topological metrics for the analysis of power grid vulnerability. *IEEE Systems Journal*, 6(3):481–487, 2012.

[29] Zhi-Hong Guan, Long Chen, and Tong-Hui Qian. Routing in scale-free networks based on expanding betweenness centrality. *Physica A: Statistical mechanics and its applications*, 390(6):1131–1138, 2011.

[30] Benoît Sluysmans. Complex networks: Analysis, sparsification and visualization. 2016.

[31] David Gfeller and Paolo De Los Rios. Spectral coarse graining of complex networks. *Physical review letters*, 99(3):038701, 2007.

[32] K-I Goh, Giovanni Salvi, Byungnam Kahng, and Doochul Kim. Skeleton and fractal scaling in complex networks. *Physical review letters*, 96(1):018701, 2006.

[33] Stilian Stoev, George Michailidis, and Joel Vaughan. On global modeling of backbone network traffic. In *2010 Proceedings IEEE INFOCOM*, pages 1–5. IEEE, 2010.

[34] Sanjay Chawla, Kiran Garimella, Aristides Gionis, and Dominic Tsang. Backbone discovery in traffic networks. *International Journal of Data Science and Analytics*, 1(3-4):215–227, 2016.

[35] M Ángeles Serrano, Marián Boguná, and Alessandro Vespignani. Extracting the multiscale backbone of complex weighted networks. *Proceedings of the national academy of sciences*, 106(16):6483–6488, 2009.

[36] X Zhang and J Zhu. Skeleton of weighted social network. *Physica A: Statistical Mechanics and its Applications*, 392(6):1547–1556, 2013.

[37] Victor M Eguiluz, Dante R Chialvo, Guillermo A Cecchi, Marwan Baliki, and A Vania Apkarian. Scale-free brain functional networks. *Physical review letters*, 94(1):018102, 2005.

[38] Daniel Grady, Christian Thiemann, and Dirk Brockmann. Robust classification of salient links in complex networks. *Nature communications*, 3(1):1–10, 2012.

[39] Multiple authors. List of metro systems. https://en.wikipedia.org/wiki/List_of_metro_systems. Accessed: 2020-04-13.

[40] RATP. https://www.ratp.fr/. Accessed: 2020-04-13.

[41] Jean-Francois Gleyze. La desserte de paris par le réseau de métro et de rer. 2003.

[42] B Berche, T Holovatch, Yu Holovatch, and C von Ferber. A tale of two cities. vulnerabilities of the london and paris transit networks. *J. Transp. Secur*, 5:199–216, 2012.

[43] Data RATP. https://dataratp.opendatasoft.com/explore/dataset/trafic-annuel-entrant-par-station-du-reseau-ferre-2018/information/. Accessed: 2020-04-13.

[44] Networkx package on python and its documentation. https://networkx.github.io/documentation/stable/.

[45] Petter Holme and Jari Saramäki. Temporal networks. *Physics reports*, 519(3):97–125, 2012.

[46] Omnil website. http://www.omnil.fr/.

[47] Frequentation of lines in 2018. https://fr.wikipedia.org/wiki/Liste_des_lignes_de_métro_parisiennes_par_fréquentation.

[48] Subway busyness google. https://www.blog.google/products/maps/grab-seat-and-be-time-new-transit-updates-google-maps/.

[49] Line 4 automation. https://www.ratp.fr/en/automatisationdela4.

[50] RER A occupation. https://www.ratp.fr/en/travaux-ete-rer/les-chiffres-cles-du-rer.

[51] RATP covid19. https://www.lemonde.fr/sante/article/2020/03/25/coronavirus-trafic-sncf-et-ratp-fortement-reduit-pour-limiter-les-deplacements-non-essentiels_6034389_1651302.html.

[52] Michael Hamann, Gerd Lindner, Henning Meyerhenke, Christian L Staudt, and Dorothea Wagner. Structure-preserving sparsification methods for social networks. *Social Network Analysis and Mining*, 6(1):22, 2016.

[53] Michael Mathioudakis, Francesco Bonchi, Carlos Castillo, Aristides Gionis, and Antti Ukkonen. Sparsification of influence networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 529–537, 2011.

[54] Nicholas J Foti, James M Hughes, and Daniel N Rockmore. Nonparametric sparsification of complex multiscale networks. *PloS one*, 6(2), 2011.

[55] Alireza Abbasi, Liaquat Hossain, and Loet Leydesdorff. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics*, 6(3):403–412, 2012.

[56] Robustness of complex networks. https://en.wikipedia.org/wiki/Robustness_of_complex_networks.

[57] Xi Zhang and K Tse Chi. Assessment of robustness of power systems from a network perspective. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 5(3):456–464, 2015.

[58] MH Khammash, Vijay Vittal, and CD Pawloski. Analysis of control performance for stability robustness of power systems. *IEEE transactions on power systems*, 9(4):1861–1867, 1994.

[59] Duncan S Callaway, Mark EJ Newman, Steven H Strogatz, and Duncan J Watts. Network robustness and fragility: Percolation on random graphs. *Physical review letters*, 85(25):5468, 2000.

[60] Jianwei Wang, Chen Jiang, and Jianfei Qian. Robustness of internet under targeted attack: a cascading failure perspective. *Journal of Network and Computer Applications*, 40:97–104, 2014.

[61] Paolo Boldi, Marco Rosa, and Sebastiano Vigna. Robustness of social networks: Comparative results based on distance distributions. In *International Conference on Social Informatics*, pages 8–21. Springer, 2011.

[62] Paolo Boldi, Marco Rosa, and Sebastiano Vigna. Robustness of social and web graphs to node removal. *Social Network Analysis and Mining*, 3(4):829–842, 2013.

[63] Mercedes Pascual, Jennifer A Dunne, et al. *Ecological networks: linking structure to dynamics in food webs*. Oxford University Press, 2006.

[64] Peng Zhang, Baisong Cheng, Zhuang Zhao, Daqing Li, Guangquan Lu, Yunpeng Wang, and Jinghua Xiao. The robustness of interdependent transportation networks under targeted attack. *EPL (Europhysics Letters)*, 103(6):68005, 2013.

[65] Darren M Scott, David C Novak, Lisa Aultman-Hall, and Feng Guo. Network robustness index: A new method for identifying critical links and evaluating the performance of transportation networks. *Journal of Transport Geography*, 14(3):215–227, 2006.

[66] James L Sullivan, David C Novak, Lisa Aultman-Hall, and David M Scott. Identifying critical road segments and measuring system-wide robustness in transportation networks with isolating links: A link-based capacity-reduction approach. *Transportation Research Part A: Policy and Practice*, 44(5):323–336, 2010.

[67] Jianhua Zhang, Xiaoming Xu, Liu Hong, Shuliang Wang, and Qi Fei. Networked analysis of the shanghai subway network, in china. *Physica A: Statistical Mechanics and its Applications*, 390(23-24):4562–4570, 2011.

[68] Schuman closed. https://www.stib-mivb.be/article.html?l=en&_guid=9042ead7-3ecf-3410-bc8d-937e31e6a1ac.

[69] Chatelet-les-halles. https://fr.wikipedia.org/wiki/Gare_de_Châtelet_-_Les_Halles.

[70] Line 14 extension image. http://www.leparisien.fr/info-paris-ile-de-france-oise/transports/transports-au-coeur-de-l-immense-chantier-qui-va-tripler-la-ligne-14-du-metro-02-04-2019-8045118.php.

[71] Line 13 in paris ranks 5th in the worlds most crowded lines. https://www.blog.google/products/maps/transit-crowdedness-trends-around/.

[72] Why line 13 in paris is considered as the 'hell-line'. https://www.lesinrocks.com/2018/04/25/actualite/societe/pourquoi-la-ligne-13-du-metro-est-consideree-comme-la-ligne-de-lenfer/.

[73] RATP deconfinement plan. https://www.ratp.fr/sites/default/files/inline-files/deconfinement18-2405.pdf.

[74] RATP deconfinement plan. http://www.leparisien.fr/info-paris-ile-de-france-oise/transports/stations-fermees-metro-et-rer-arretes-a-22-h-les-transports-franciliens-au-ralenti-des-jeudi-25-03-2020-8287575.php.

[75] Dijkstra algorithm. https://www.geeksforgeeks.org/dijkstras-shortest-path-algorithm-greedy-algo-7/.

Appendices

Appendix A: Impact of the transfer penalty

The following table shows the amount of transfers done at the 20 stations with the most transfers depending on the penalty that is given.

The figures show the IBC of the edges as well as the amount of passengers using each station for 4 different situations. The input of the stations is always considered.

Station ↓ \Penalty →	0-0	1-3	peak	off-peak	night	5-11	Covid-19	10-21
Châtelet	533M	472M	378M	359M	335M	317M	308M	305M
Saint-Lazare	157M	139M	134M	122M	118M	108M	99M	92M
Gare de Lyon	360M	221M	128M	84M	62M	52M	41M	32M
Nation	152M	127M	120M	113M	105M	103M	100M	100M
Charles de Gaulle-Etoile	146M	124M	103M	96M	88M	85M	82M	76M
Montparnasse-Bienvenue	153M	121M	101M	92M	78M	75M	75M	72M
République	86M	96M	87M	86M	89M	86M	87M	82M
La Motte Picquet Grenelle	109M	98M	86M	67M	57M	52M	42M	39M
Opéra	232M	115M	86M	91M	92M	91M	92M	99M
Madeleine	294M	84M	71M	49M	39M	35M	32M	31M
Bercy	129M	86M	56M	42M	32M	27M	21M	17M
Place d'Italie	80M	59M	56M	47M	52M	41M	36M	34M
Stalingrad	134M	64M	54M	52M	45M	40M	39M	31M
Gare du Nord	198M	105M	51M	38M	32M	24M	23M	19M
Place de Clichy	68M	58M	51M	45M	43M	41M	41M	41M
Barbes - Rochechouart	72M	53M	49M	43M	38M	34M	30M	30M
Gare de l'Est	63M	51M	48M	39M	31M	30M	29M	27M
Denfert-Rochereau	19M	19M	47M	38M	40M	41M	44M	46M
Bastille	46M	34M	44M	44M	42M	43M	38M	39M
Champs-Elysées-Clémenceau	71M	42M	38M	34M	33M	33M	33M	34M

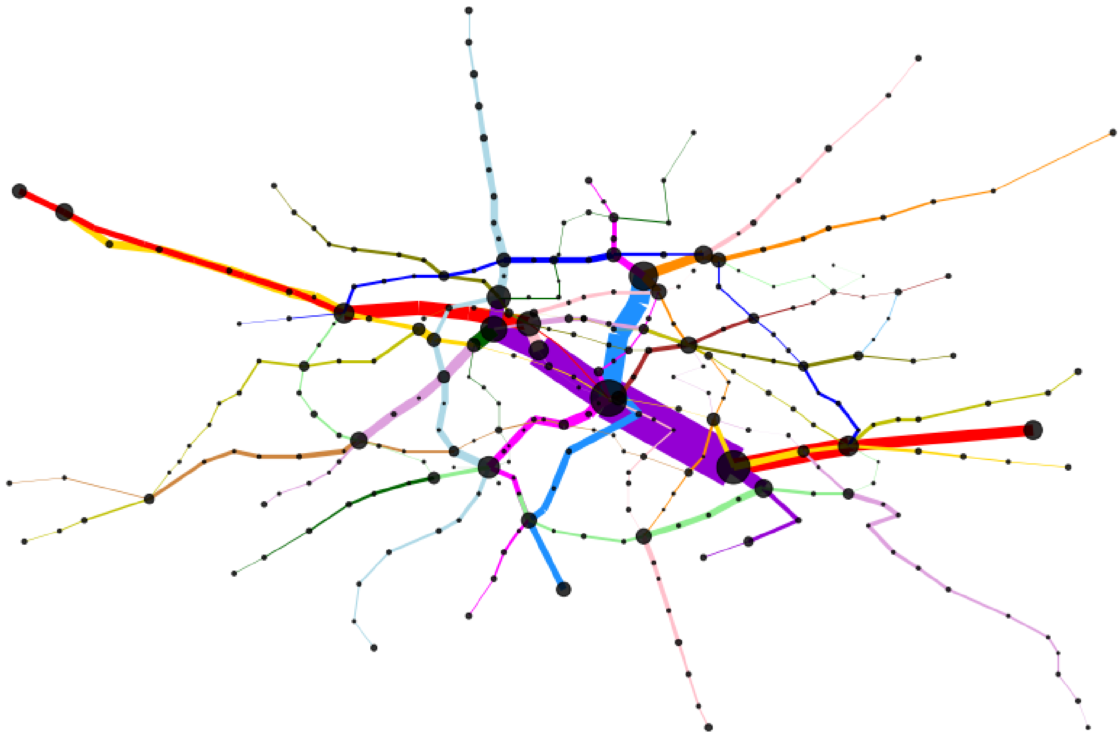


Figure 5: Single layer network - transfer penalty=0

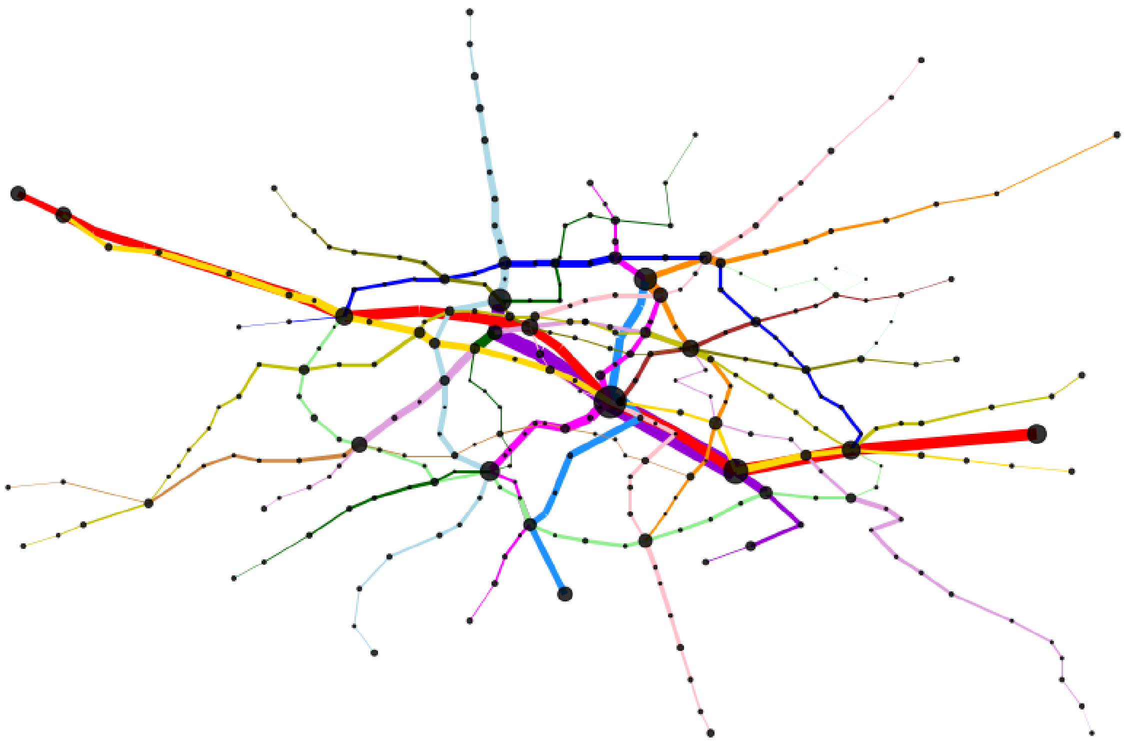


Figure 6: Peak hour - transfer penalty=2

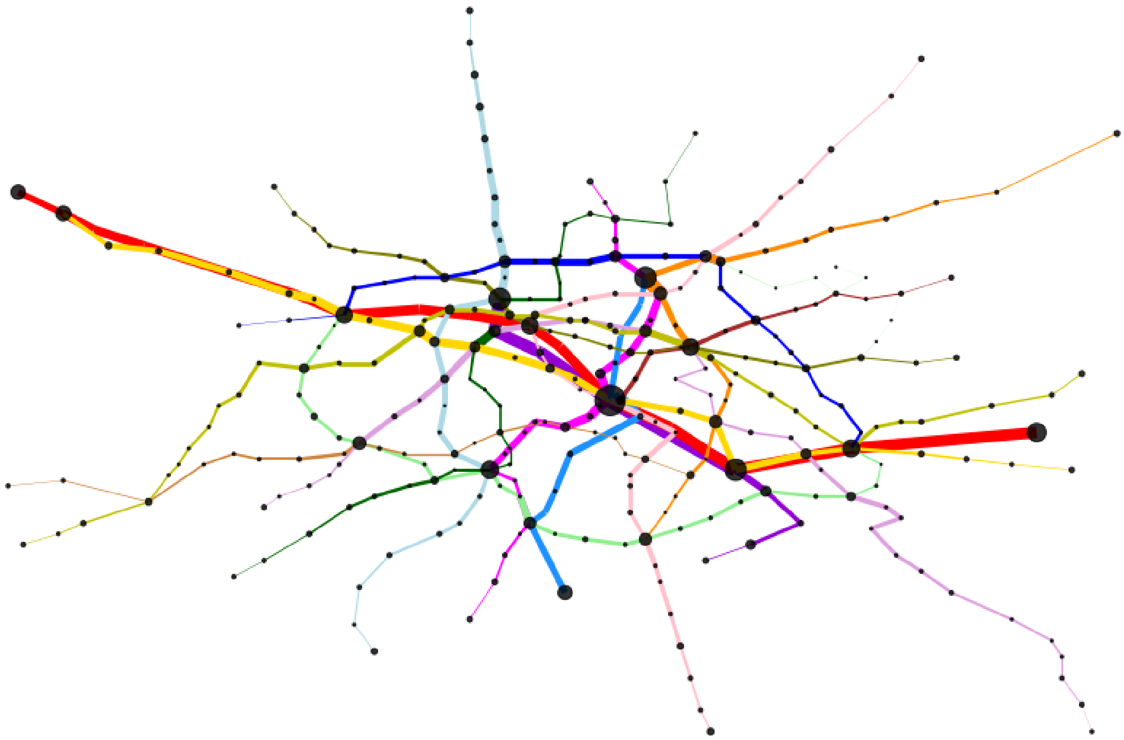


Figure 7: Night - transfer penalty=4

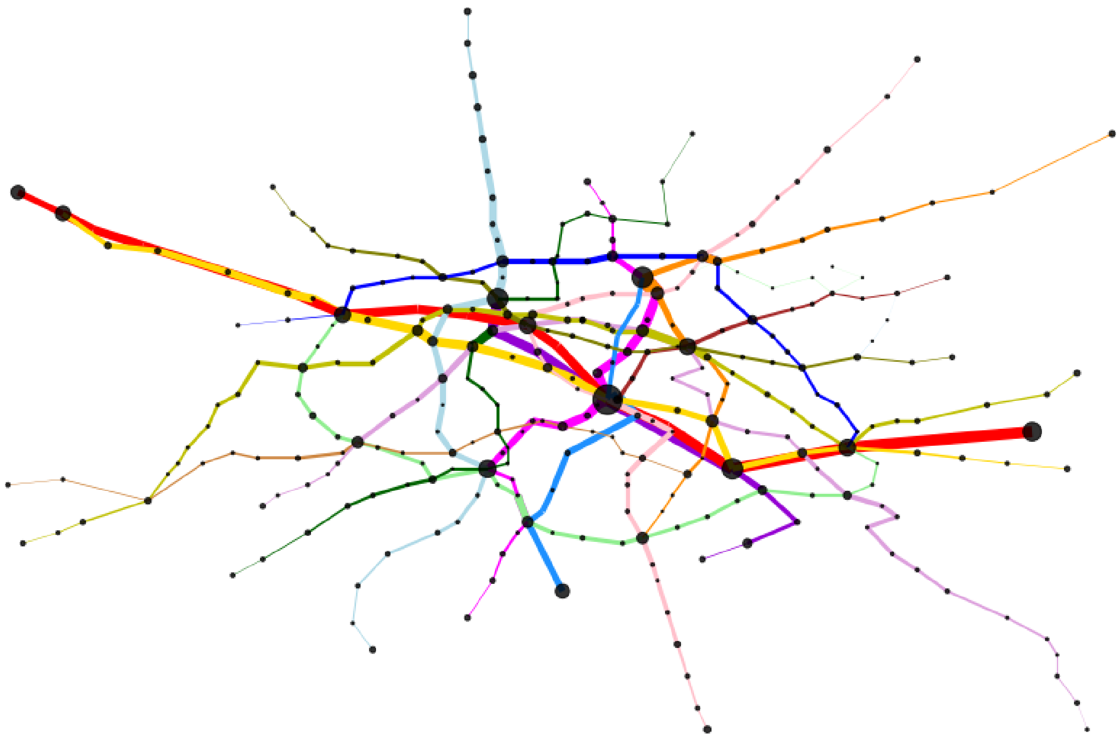


Figure 8: Extreme situation - transfer penalty=7

Appendix B: Impact of using Input of stations

The figures show the IBC of the edges as well as the amount of passengers when the input of stations is considered or not. The transfer penalty is 2 in both cases.

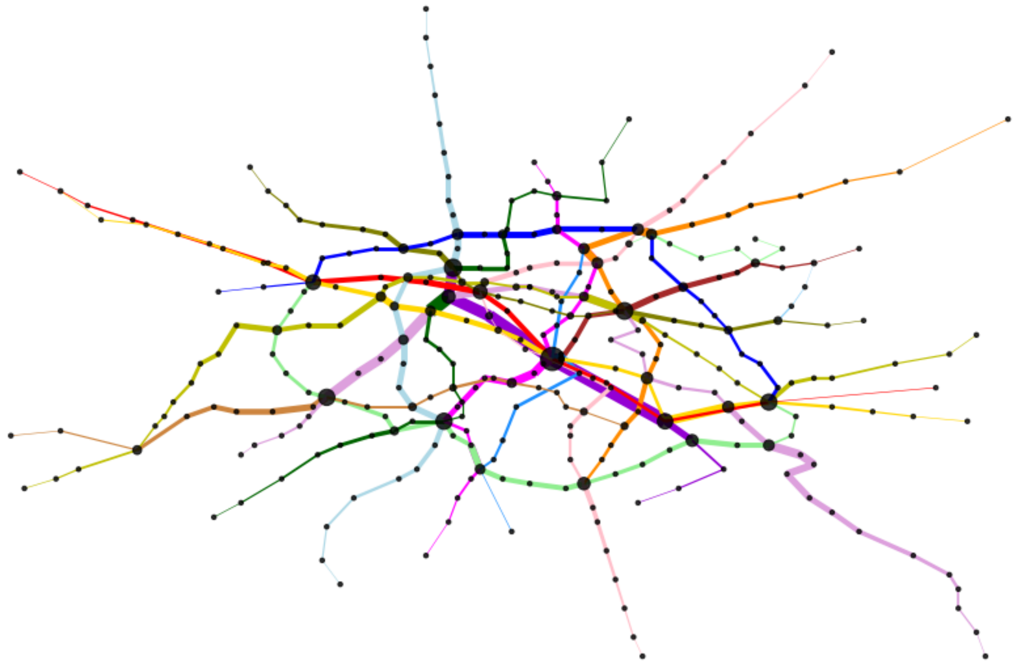


Figure 9: Input not considered

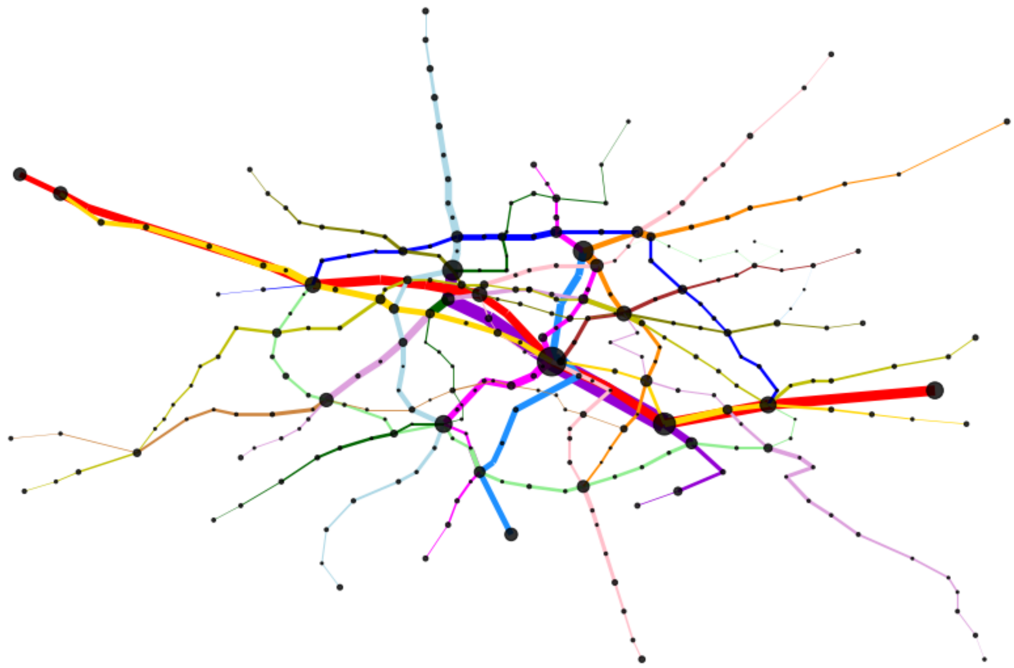


Figure 10: Input Considered

Appendix C: Removal of line 14 and RER A

The figures show the IBC of the edges as well as the amount of passengers when line 14 and the RER A are removed.

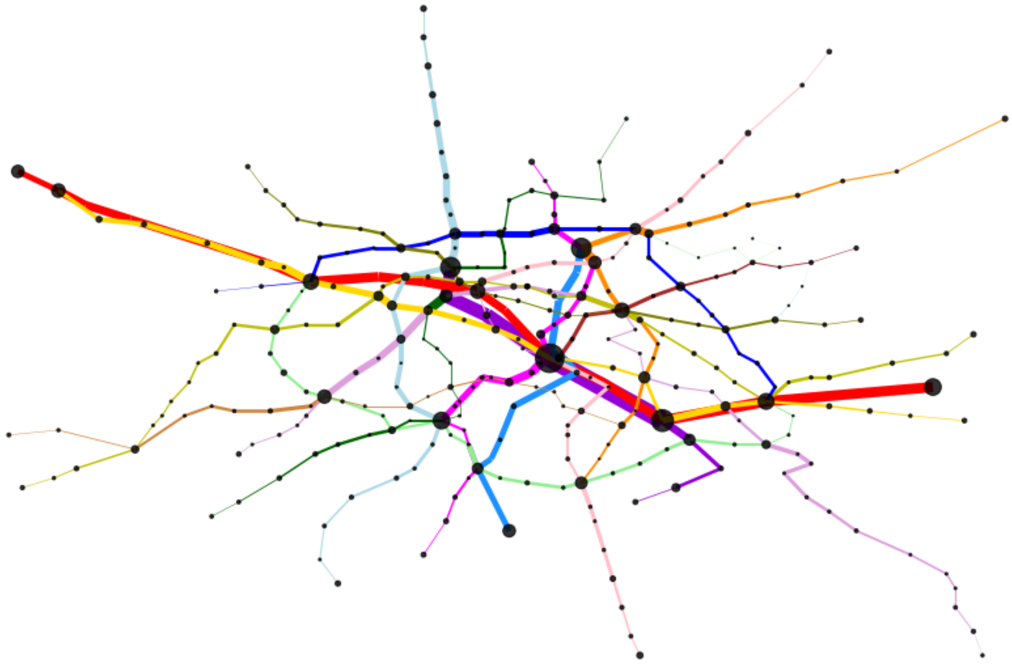


Figure 11: No failures

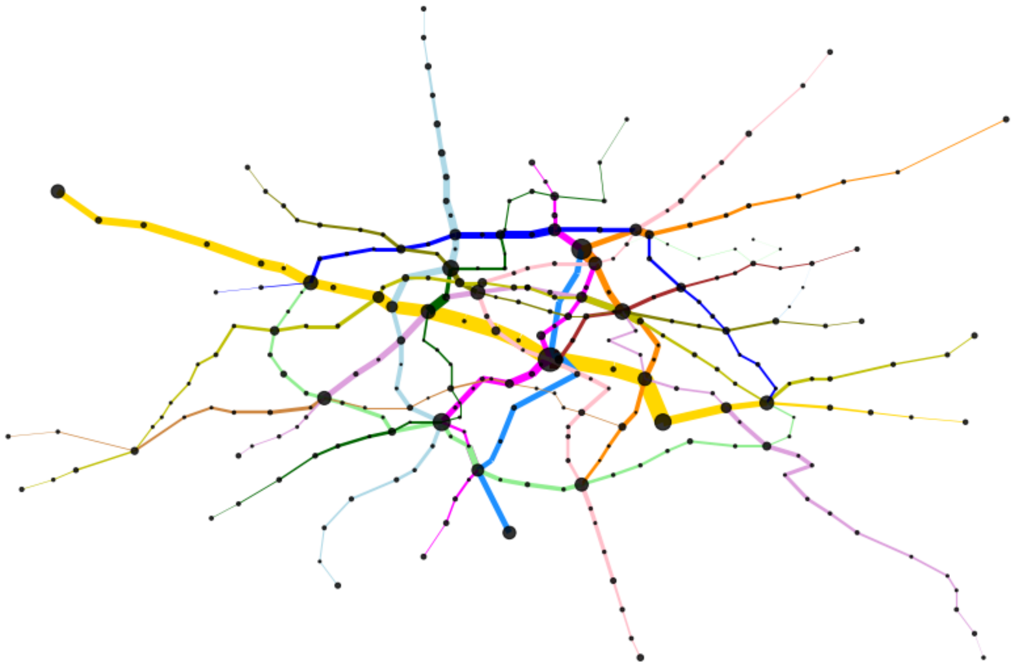


Figure 12: Failure of line 14 and RER A

Appendix D: IBCLines when removing lines with a transfer penalty of 4

Line ↓ \ Del. Line →	Normal	1	2	3	3bis	4	5	6	7	7bis	8
1	**2054**	/	2128	2131	2055	2094	2064	2283	2114	2057	1961
2	**1214**	1156	/	1211	1212	1346	1161	1178	1215	1198	1308
3	**859**	923	1007	/	843	881	873	868	868	860	898
3bis	**6**	6	6	48	/	7	6	6	7	6	6
4	**2001**	2071	2111	2021	2001	/	2131	2176	2033	2001	2147
5	**1082**	1084	1198	1067	1081	1271	/	1073	990	1075	1032
6	**1265**	1561	1210	1303	1266	1438	1318	/	1192	1267	1197
7	**1678**	1730	1850	1687	1678	1766	1917	1830	/	1671	1680
7bis	**33**	33	32	36	32	34	34	33	32	/	33
8	**1708**	1910	1771	1679	1705	1868	1726	1773	1754	1706	/
9	**1566**	1737	1669	1568	1565	1569	1555	1628	1619	1565	1747
10	**492**	557	501	500	493	484	495	559	496	493	588
11	**513**	571	556	656	509	505	563	520	516	496	546
12	**877**	823	921	871	878	1159	888	979	911	876	898
13	**1660**	1603	1688	1664	1660	1750	1676	1717	1691	1660	1722
14	**889**	917	929	875	890	958	909	863	893	890	853
RER A	**2124**	2633	2118	2111	2125	2074	2148	2088	2080	2125	2117
RER B	**891**	845	909	896	891	990	882	952	890	890	924

Line ↓ \ Del. Line →	Normal	9	10	11	12	13	14	RERA	RERB	RER
1	**2054**	1921	2103	2051	2037	2022	2437	3135	2159	2976
2	**1214**	1279	1234	1348	1241	1119	1294	1260	1337	1374
3	**859**	884	860	924	859	846	1020	912	951	926
3bis	**6**	6	6	30	6	6	7	6	7	7
4	**2001**	2018	1954	2033	1856	1997	2097	1926	2404	2262
5	**1082**	1070	1086	1134	1098	1128	1207	1183	1256	1323
6	**1265**	1187	1291	1287	1182	1364	1358	1271	1217	1247
7	**1678**	1693	1666	1701	1695	1712	1747	1659	1723	1671
7bis	**33**	34	33	62	33	32	33	34	33	34
8	**1708**	1767	1571	1675	1711	1757	1818	1807	1814	1875
9	**1566**	/	1713	1541	1588	1575	1643	1695	1657	1744
10	**492**	460	/	495	501	490	510	518	505	519
11	**513**	528	514	/	518	529	481	504	484	483
12	**877**	899	886	898	/	1027	836	877	866	960
13	**1660**	1650	1639	1690	1697	/	1773	1787	1721	1634
14	**889**	907	895	880	853	715	/	605	571	598
RER A	**2124**	2100	2130	2115	2162	2208	2329	/	2163	/
RER B	**891**	916	898	909	898	913	806	829	/	/

Appendix E: Code

The code can be found at the following link: https://github.com/pierrebeltjens/Thesis_Pierre_Beltjens. The code is available as a python notebook and as a normal python file.

Shortest path in subway network

To compute the closeness centrality and IBC, the shortest paths and distances in the network between any 2 nodes need to be known. Dijkstra's algorithm[75] can be used for this.
The main issue is that Dijkstra's algorithm is made for single layer graphs and it doesn't always function properly in multilayer networks. Dijkstra's algorithm is thus not applied on all stations but on all platforms in the network.
A station that is part of 3 lines will be divided in 3 nodes for the algorithm. The shortest paths between all the platforms or nodes will then be computed. The distance between the nodes or platforms from a single station will be the transfer penalty. This way, the new network is simplified to a single layer while keeping the transfer penalty and Dijkstra's algorithm can be applied.
For computational complexity reasons, a single shortest path is computed between any 2 stations. The differences this implies on the IBC of stations and edges is negligible.

Input of model

The input of the model which is given in the code for the subway network of Paris, consists of 3 parts:

- A list of lists 'In':
  Each station in the network is assigned a number between 0 and 310. Each line will then be represented by a list of all the numbers of its stations in order. Line 1 will for example be represented by a list [0,1,...,24], Line 2 by a list [25,26,6,27,28,...,47,20],...
  'In' is the list of all the line representations.

- A list 'Freq':
  This is a list of all the inputs of each station. Freq[61] for example will be the input of station 61 which is 'Sentier' on line 3.

- A list 'Coord':
  Lists the coordinates for all stations to create the visualizations. Coord[2A] will be the y-coordinate of station A and Coord[2A+1] will be the x-coordinate.

Mode

While the code was used to compute most results in this report by altering it slightly each time, 5 modes are build-in and can be chosen when running the whole code:

- 'Basic': This mode computes all passenger flows through the network when neither the inputs nor the multilayer aspect of the network as defined in chapter 2 are considered.

- 'BC': This is the default mode that computes all the passenger flows through the network from chapter 3 when there are no failures. The end result is figure 3.7.

- 'Backbone': This mode computes all the passenger flows through the network as well as the backbone that is obtained by combining the 3 different filters from chapter 4. The visualization at the end is figure 4.5.

- 'L14RERAFailure': This mode computes the passenger flows through the network when line 14 as well as the RER A fail at the same time which is the situation described in chapter 5.1. The end result is figure 12 from appendix D.

- 'DeconfinementPlan': This mode computes the passenger flows for the deconfinement plan described in chapter 6.2. The end visualization represents figure 6.3.

The mode can be chosen in the first lines of the first code cell.

Structure of Main Code

The main code should ideally be opened with a python notebook in order to be able to run separate parts. Nonetheless, the .py is also available.

The structure of the code goes as follows:

- The mode should be selected in the first code cell. The default mode is 'BC'.
- The second code cell contains the input data of the model and computes the adjacency matrix. Removing lines or stations can be done in that block by modifying 'In' or filling 'FailStop' with the numbers of the stations that should be removed.
- The third code cell computes the degree centrality of each station and prints them.
- The fourth code cell computes all the shortest path lengths in the network and the fifth code cell uses that information to compute the closeness centrality of all stations.
- The sixth cell computes all the shortest paths in the network. This is similar to the fourth cell since both cells use Dijkstra's algorithm. Both cells are not combined in order to keep the computation of the closeness centrality and the one of the IBC apart even though they are similar.
  The seventh cell then computes and prints the IBC of stations, edges and lines as well as some extra information like the amount of transfers done at each station.
- The next 4 code cells apply the 3 sparsification filters.
- The last code cell creates the visualization in accordance with the selected mode. The NetworkX and matplotlib.pyplot python packages are used for this.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
École polytechnique de Louvain
Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/epl**