UCLouvain

École polytechnique de Louvain

European air and rail transport networks between main cities

A comparative analysis

Author: **Damien LEPAGE**
Supervisor: **Jean-Charles DELVENNE**
Readers: **Vincent BLONDEL, Éric CORNELIS, Bart WUYTS**
Academic year 2019–2020
Master [120] in Mathematical Engineering

Abstract

International transport networks are at the core of our economies and societies as they shape the way we travel and the way we trade. At European level, the global air and rail transport networks allow passengers to travel between most of the cities in a short time. Low-cost airlines offer affordable flights between most of the large European cities and the high-speed rail (HSR) infrastructure is continuously expanding. The last decade has however seen a growing concern in the way we travel and the impact it has on the environment, as flying remains one of the most carbon-intensive activity one can do on an hourly basis. This phenomenon has been reflected by policy initiatives proposing to reduce the number of domestic flights, by the Swedish movement of *flygskam* (shame of flying) or by the recommissioning of some night train routes.

In this thesis, we propose to study the European Air Transport Network (EATN) and the European Rail Transport Network (ERTN) using complex networks theory. Transport systems can indeed be modelled as networks with nodes (airports, train stations) and edges linking them (travel routes). The particular feature of this thesis is to study both networks in parallel with essentially the same nodes representing main cities in Europe.

Using measures of complex networks theory, we show that the EATN has the properties of *small-world networks*, with the average number of flights required to travel between any two cities being as low as 1.81. On the other hand, the ERTN is characterized by much longer paths and travelling between two cities requires taking 5.77 trains on average.

Using directed graphs, we model door-to-door journeys between cities taking into account transfer times to the airport/station, checkin-in and connecting times. We find that for 6% of the journeys between main cities in the EU, the door-to-door travel time is shorter by train than by plane, and that in June 2017 at least 4 million passengers took a flight on a journey where the train is faster. In addition, we evaluate that the train is usually competitive with the plane up to a distance of 444 km, and up to 704 km when the journey can be made on high-speed trains.

We then assess the impact on train lines following a potential shift of those 4 million passengers from air to rail. Our analysis shows that high-speed rail will have to carry the biggest share of additional passengers resulting from a modal shift, and that the main hubs are not always the same in the air and rail networks.

Our network model is then used for two case studies. Firstly, we evaluate the feasibility of night train services between main cities in Europe. We find that night train routes could in theory be operated between 22% of the cities, and up to 38% if we implement high-speed night trains as in China. Secondly, we assess the potential in terms of passenger traffic on the future high-speed line between Lyon and Turin and find that the market share that could be gained from the air transport sector is rather modest.

Finally, we consider the problem of the optimal network design for the HSR network in Western Europe using methods developed by Micheal Gastner and Mark Newman. Several optimal network candidates are generated with some of them exhibiting up to 75% similarity with the actual network.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Context

In June 2019, two bills were introduced at the French Assemblée Nationale to reduce domestic flights in France. The first suggested to ban domestic flights when an alternative taking less than 5 hours by train exists. The second proposition suggested to ban flights when the same journey by train is at most two and a half hours longer. This extra time was meant to take into account the time needed to get to the airport and go through the security checks when travelling by plane. An article in Le Monde (Vaudano and Sénécat, 2019) assessed that the first and second proposition would reduce the number of domestic flights by 51% and 25% respectively. This would represent a decrease of 13.8 or 6.6 million flights per year depending on the proposition.

At the time my supervisor and myself were quite surprised to hear that so many journeys were still performed by plane when credible train alternatives exist. It also seemed obvious that an analysis like the one in Le Monde was to be performed at European scale. In 2017, passengers on national flights only represented 17% of the total passengers of the air transport sector. On the other hand, national and intra-EU flights represented together almost two thirds of the total traffic in the air transport (see Fig. 1.1[1]). For which flights in Europe is there a reliable alternative by train? What share of the air passengers traffic does this represent at European level? How many additional passengers will rail lines have to carry if those flights are banned? There are currently no clear answers to those questions.

Both bills failed to be approved by the Assemblée Nationale but the subject still remains topical. In May 2020, the French government agreed to help Air France in the Covid-19 crisis if the company stopped operating domestic routes where the plane competes with high-speed trains (Briginshaw, 2020). In its last Transport White Paper (Kallas, 2011), the European Commission sets as objective to triple the length of the existing high-speed rail network. Furthermore, by 2050 the majority of medium-distance passenger transport should go by rail. The global target is to reduce the greenhouse gas emissions in the

[1]Source: Eurostat

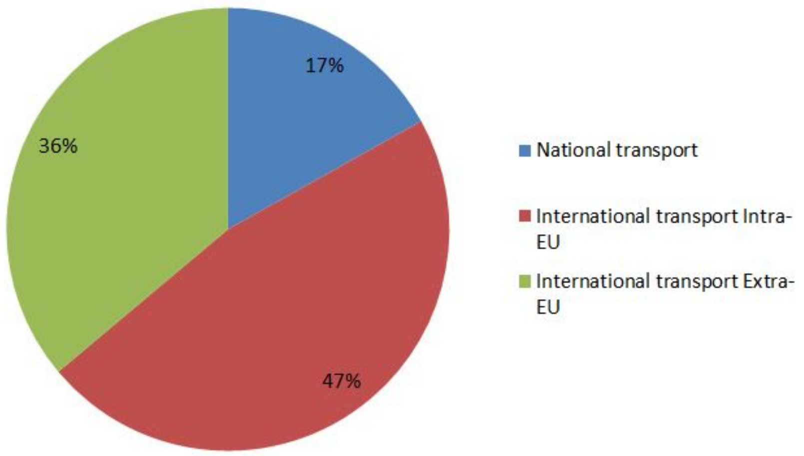


Figure 1.1: Overview of air passenger traffic in the EU, 2017.

transport sector by 60%.

Besides the political and economical aspect, the last decade has seen a growing interest in the way we travel and the impact it has in terms of pollution. In the last two years, we have seen the emergence of the the Swedish concept of *flygskam*, the shame of flying. As the world needs to dramatically cut greenhouse gas emissions, flying is one of the most carbon-intensive activity on an hourly basis (Miles, 2017). Even if commercial flights represent merely 3% of the global CO_2 emissions from fossil fuel, these are growing fast. Between 2013 and 2018 the commercial aviation sector has seen a 32% increase of its emissions (Brandon Graver and Rutherfor, 2019).

Low-Cost Carriers have played an important role in the expansion of aviation over the past 20 years. Between 2001 and 2012, the market share of low cost airlines in terms of seats has raised from 14% to 38% (Rozenberg et al., 2014). As a consequence, in 2019 Ryanair became one of Europe's top 10 polluters (BBC, 2019).

In the meantime, growing awareness on the pollution generated by the aviation industry is reinvigorating rail travel. SJ which is the main Swedish train operator reported a 12% increase in passenger numbers in the first months of 2019. Indeed, travelling by train often generates much less emissions than by plane. A journey by train from Stockholm to Copenhagen generates for instance more than 50 times less carbon dioxide per passenger than by plane [2]. On a journey between Brussels and Marseille, the train and the plane

[2]Source: EcoPassenger.org

emit respectively 8 and 139 kg CO_2 per passenger (see Fig. 1.2).

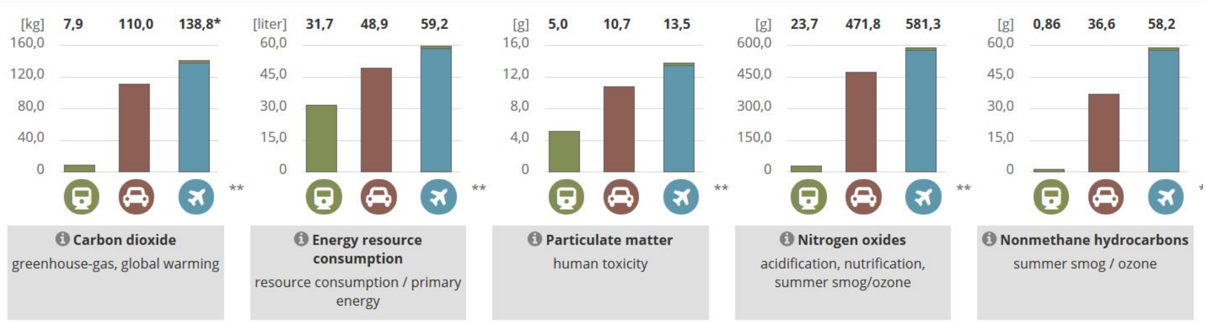


Figure 1.2: Emissions and energy consumption for a journey from Brussels to Marseille. Source: EcoPassenger.org.

The last 50 years have seen the emergence of high-speed rail (HSR) in Europe. After the first line was opened between Rome and Florence in 1977, France quickly launched its "Train à Grande Vitesse" (TGV) followed by Germany (Intercity Express) and Spain (Alta Velocidad). At the end of 2017, the EU had 9067 km of high-speed lines and 1671 km were under construction. Passengers traffic is growing as well: the number of kilometers passengers travelled by HSR grew from 15 billion in 1990 to 124 billion in 2017 (ECA, 2019).

Night trains are also rebounding in several European countries. A new connection between Brussels and Vienna was inaugurated in January 2020 and now runs twice a week. Sweden never stopped operating night trains from Stockholm to the cities of Northern Sweden and is now considering opening a new connection to Brussels. Once considered old-fashioned, night trains offer a new alternative to the plane or the car for long journeys where no high-speed infrastructure is available.

1.2 Literature review and specificity of this thesis

We propose to study the rail and air transport networks at European level using complex networks theory. The particular feature of this work is to study both networks in parallel with essentially the same nodes representing the main cities of Europe. This makes it possible to compare door-to-door travel times between transport systems, to assess until which distance the train is competitive with air travel or to analyse the impact of a modal shift from air to rail on specific routes.

Since the emergence of complex networks theory in the late 1990's and early 2000's (Albert and Barabási (2002), Newman (2003)), large-scale networks like computer networks, social networks, biological networks or transport networks have received a growing interest. Networks are abstracted as *graphs* $G = (V, E)$ with V the set of *vertices* (also called *nodes* or *points*) and E the set of *edges* (also called *links*, *routes* or *lines*). Two nodes are said to be neighbors if there is an edge linking them. Transport networks belong to the category of *spatial networks*, with vertices occupying particular positions in space as opposed to non-spatial networks like the Internet or social networks. At network level, complex graph theory provides new ways of understanding the structure and properties of networks. Local measures at node level allow for instance to quantify the importance or centrality of nodes, thereby identifying hubs or vulnerabilities (Newman, 2018).

These new measures allow to define two types of networks frequently encountered in real-world systems: small-world networks and scale-free networks. *Small-world networks* in the sense of (Amaral et al., 2000) are defined by a high *clustering coefficient* (probability that neighbors of a node are themselves neighbors) and a low *diameter* (longest distance between nodes). Such graphs can for instance be obtained by rewiring a small number of edges in a regular ring-like network (Watts and Strogatz, 1998). In *scale-free* networks, the *degree* distribution (number of neighbors of a node) follows a power-law distribution. This means that many nodes are poorly connected while a few have an important number of routes, such nodes are often called hubs. This feature can be explained by a preferential attachment mechanism: as the network grows new nodes tend to link themselves to already well-connected nodes (Barabasi, 2016). As a consequence of those findings, various transport systems could possibly be shaped by the same mechanisms and thus have similar topological properties.

Compared to other networks, air transport networks can be easier to model since they are only constrained by the location of airports, as opposed to rail or road networks which rely on track infrastructure to connect nodes. As a result, and thanks to a relatively good availability of data, air transport networks are probably among the most studied transport networks. Using a weighted complex network methodology (routes are weighted by average daily passenger traffic), the U.S. domestic airline network has been shown to be a small-world network (Xu and Harriss, 2008). The Chinese air transport network turns out to have a degree distribution captured by an exponential function, and displays some small-world network properties with a low average path length and a high clustering coefficient (Wang et al., 2011). At the worldwide level, the air transportation network appears to be a scale-free and small-world network (Guimerà et al., 2005).

In air transport networks, nodes (airports) are connected if an airline operates a route between the airports. Flights are mostly direct, i.e. usually requiring no stopovers. In contrast, in rail transport networks most of the trains stop between their departure and arrival points. Sometimes trains also pass through a station (node) without stopping. As a consequence there are several ways of modelling rail transport networks. In (Dasgupta et al., 2003), two stations are connected if at least one train stops at both stations. This approach is known as the *space-of-changes* (also called *P-space*) (Kurant and Thiran, 2006). With this modelling, the Indian railway network is shown to have properties of a small-world network. Using the same method, the Chinese railway system is shown to have properties of both small-world and scale-free networks (Li and Cai, 2007).

Another approach is the *space–of–stops* (also called *L-space*) where two stations are connected if they are two consecutive stops on the route of at least one train. In (Sienkiewicz and Hołyst, 2005) both definitions of network topology (L-space and P-space) are used to analyse the transport systems of 22 Polish cities. It is shown that the degree distribution of networks follow a power law in the L-space and an exponential distribution in the P-space. Finally, in the *space–of–stations* approach, two stations are connected only if they are physically directly connected (with no station in between).

Several applications on real-world systems have thus confirmed the small-world and/or scale free properties of some transport networks. However, perhaps not all transport systems can be classified as small-world or scale-free networks. The Chinese high-speed Rail network has a bimodal degree distribution which reflects a two-regime structure, but is not found to have the properties of small-world or scale-free networks (Wei et al., 2019).

Regarding rail transport networks, studies have nonetheless been restricted by the poor availability of data on infrastructure and mainly on passenger traffic. Most of the time, the traffic is roughly approximated as the number of trains passing through a station. This is slowly starting to change with the availability of new datasets on train ticketing, for instance in China (Wei et al., 2017).

Despite the significant interest in transport networks in the last decades, the study presented here is, to the best of our knowledge, the first time that air and rail transport networks are studied at European level. The only study with a similar scope that we note is that of (Kurant and Thiran, 2006) where the railway network formed by the main train lines and stations in countries of central Europe is extracted from timetables. The case of the European railway network is rather taken as an example to illustrate Kurant's method for analysing topology and traffic of transport networks from timetables. It is therefore not very clear what the exact scope of the study was, in terms of countries and stations considered, and only a brief analysis on the network topology was developed. As far as we know, it is in addition the first time that the European rail and air transport networks are studied in parallel in terms of complex graph theory. We have not found any example of such comparative study on rail and air networks for any other region.

Besides the topological study of complex networks, another field of research is the design and models of growth for such networks. One of the problems addressed is the optimal design for spatial networks where the purpose is to choose the best design of the network, i.e. the design minimizing a given objective function. Several choices are possible when it comes to the definition of the objective function that captures the best the cost of the network. Most often the objective function seeks a compromise between construction and travel costs (i.e. convenience of travelling in the network). Two types of methods can be used to generate optimal or near-optimal networks: generative and degenerative processes (Ducruet and Lugo, 2011). In generative methods, we start from an unconnected network and add successively the edges decreasing the most the network cost. Degenerative methods start from the complete network (all edges possible) and successively remove the less valuable edges. Studies on the subject include (Schweitzer et al., 1997), (Barthelemy and Flammini, 2006) and (Gastner and Newman, 2006b), with an application on the U.S. air transport network in (Gastner and Newman, 2006b) revealing a similar structure to the actual network. In the last chapter of this thesis, we adapt Gastner's approach to the European high-speed rail system to compute near-optimal networks.

1.3 Executive summary

In Chapter 2 we define the scope of the study and the data used to model the European rail and air transport networks both in terms of routes and traffic.

Chapter 3 studies the topological properties of the unweighted transport networks using complex graphs theory. Both the European Air Transport Network (EATN) and the European Rail Transport Network (ERTN) are shown to have a degree distribution which is best fitted by an exponential law. The implication is that the EATN and the ERTN are dominated by a few nodes having a high number of links (hubs) while many nodes only have a few. The degrees of the main hubs are nonetheless lower than in scale-free networks where a few nodes with an even higher degree have a significant fraction of the edges.

Both networks are characterized by a high clustering coefficient but diverge in terms of average path length. In the EATN, the average path length is 1.81 and 95% of the journeys can be completed in at most 2 hops (flights). The ERTN has an average path length of 5.77 reflecting thereby the high number of connections needed when travelling between cities in the network. We conclude that the EATN has properties of a small-world network which is essentially not the case for the ERTN due its strong dependence on the Euclidean distance.

In Chapter 4 we then weight the network edges by their travel duration to compute shortest paths between cities. We use directed graphs to model door-to-door journeys taking into account the travel time to the airport/station and connecting times. Using this approach, it turns out that for 6% of the departure-arrival city pairs, the train is faster than the plane, and for 13% of the city pairs the train is at most 2 hours longer. We then put the travel time in relation with the distance between cities in both transport systems. It is found that on average the train is faster than the plane when the distance between cities by train is shorter than 444 km. Looking only at city pairs that are fully connected by high-speed trains, the rail network is competitive over the air network up to a distance of 704 km. To conclude Chapter 4, we measure the accessibility of cities in both networks using the closeness centrality of nodes. We find that on average, cities are better connected to other nodes in the EATN than in the ERTN due to long travel times between geographically remote nodes in the latter.

The next chapter focuses on passengers traffic. Having identified routes where a viable alternative by train exists, we compute how many passengers could possibly switch from air to rail. Considering only journeys where the train is faster, at least 4 million passengers could switch from air to rail and this numbers grows with the extra time passengers are willing to spend on train travel. We then study the impact of such modal shift on train lines by computing the passenger flows resulting from the switch of passengers to the ERTN. We highlight the vital importance of high-speed rail when it comes to carrying these additional passenger flows, with the majority of the kilometres being covered on high-speed trains. Next, we use these flows to compute the betweenness centrality of nodes and identify hubs. We find that major hubs are not always the same in the EATN than in the ERTN, the geographical position of nodes playing a more important role in the latter.

To conclude Chapter 5, we use our new network model for two case studies. We first evaluate the feasibility of night train services between main cities in Europe. We find that night train routes could in theory be operated between 22% of the cities, and up to 38% if we implement high-speed night trains as in China. Secondly, we assess the potential in terms of passenger traffic on the future high-speed line between Lyon and Turin and find that the market share that could be gained from the air transport sector is rather modest.

Finally, Chapter 6 addresses the optimal network design problem for the European high-speed rail. We use the approach in (Gastner, 2005) where the cost of networks is defined by a mix of construction costs and traffic costs. Near optimal-optimal networks minimizing the total cost can be obtained by a greedy degenerative algorithm or by a simulated annealing algorithm following a Metropolis-Hastings rule. Gastner's methods are adapted to compute near-optimal candidates for the European high-speed rail network. Depending on the importance given to traffic costs, different optimal design candidates are generated. We then highlight the similarities and the discrepancies with the real-world network.

2 Scope and data

2.1 Cities and countries of interest

The countries of interest (CoI) for this study are the members of the European Economic Area (EEA). This includes Switzerland, the United Kingdom and Norway in addition to the EU Member States. Iceland is also a member of the EEA but was not taken into account in this study since its geographical position and the absence of rail infrastructure makes it non relevant for a comparison between rail and air transport.

Air and high speed rail networks operate between main cities in each country. Data on the population of these cities is extracted from Eurostat City Statistics from 2016. For each city, the population considered is that of the so called "urban area" encompassing the suburbs.

The quite large scope of this study as well as the absence of a common European system regrouping all the infrastructure and flow data on transport makes it so that several datasets were needed to model the rail and air transport systems. The different sources of information, their scope and purpose are described below.

2.2 European air transport network (EATN)

2.2.1 Eurocontrol

Eurocontrol is the central organisation for coordinating and planning of air traffic control for all of Europe. Access was granted to their archive of operational flows of data for R&D which covers all historic commercial flights in four sample months. The period we used is June 2017, which is the most recent sample available. It is also a month which we expect to exhibit "normal" flight patterns, unlike holiday months like July or August.

A total of 912 787 flights were accounted for in June 2017. Since we focus on commercial flights performed by passenger airlines; business, military and cargo flights records were excluded. Furthermore, we only consider flights between airports in the CoI which leaves us with 450 725 flights in total. For this study, the relevant data is the departure and arrival airports for each flight as well as the flight duration and distance flown.

This information allow us to identify which commercial routes exist between airports. We consider that a route exists if a given aircraft (AC) operator performs a flight at least once a week between two airports. The flight duration of the route is then the average of all the flights that ensured the connection, the same goes for the distance flown.

2.2.2 Eurostat air transport measurement – passengers

Eurostat provides air transport data regarding passengers on national and international flights, intra and extra-EU. Data are transmitted to Eurostat by EU Member States and partner countries.

From Eurostat database we have extracted for each airport the traffic in the number of passengers for 2017. Since our focus is on comparing travel times between main cities of Europe, only airports with more than 500 000 passengers per year are taken into account. This represents 194 airports all over the CoI, around a hundred million passengers for the month of June 2017 and 83% of the total passenger traffic in the CoI.

Eurostat database also contains records of passenger traffic on the flight routes. This allows us to associate passenger flows with the routes mentioned in the section above, which will be very important to evaluate the impact of deleting or "banning" a given route. Passenger flows are given per month and from departure to arrival airport. For Czech Republic, the information on the destination airport is unfortunately missing and only the destination country is known. Sometimes, only the destination city is given instead of the airport (e.g. London instead of London Luton Airport). In these cases, the flow is divided evenly among the airports of the city/country. This represents however only a tiny portion of the routes.

2.2.3 OpenFlights airport database

OpenFlights is an open database that provides data on flights, airlines, and airports around the world. The airport database was used in this study to get relevant information about airports, including latitude, longitude, time zone, country and city. OpenFlights is considered as a reliable source of information and is used by several research projects, for instance to study accessibility of the Arctic region (Reid et al., 2015).

2.3 European rail transport network (ERTN)

Building the rail transport network turns out to be a much harder task than the air network. In the EATN, a flight goes from one airport to another and the route taken is of no importance in this study. In the ERTN, the route is on the contrary dependent on the rail infrastructure, with possible intermediate stops. Unfortunately, such information on the infrastructure is difficult to obtain at European scale. The European Union Agency for Railways (ERA) has a system called Register of Infrastructure (RINF) which gathers the data on infrastructure in most of the CoI, but not all. This system is however rather aimed at Rail Infrastructure Managers (RIM) (e.g. Infrabel) with a level of detail including junction points, border points, switches, freight only lines, etc. The level of complexity of RINF turns out to be too high for the purpose of this study, instead a more simple but well-known tool was used: Google Directions API.

2.3.1 Google Directions API

The Directions API is a service that calculates directions between locations with different modes of transportation including transit. This allows us to find the shortest path between two cities by train. Many cities have several main train stations (e.g. Gare du Nord, Gare de l'Est, Gare de Lyon and Gare Montparnasse for Paris). The advantage of Google Maps API is that we do not need to specify the exact departure or arrival train station, the best station depending on the journey is picked automatically. Furthermore, the results are based on current timetables which reflects the present state of the network. The itinerary queries were performed between February and May 2020.

The drawback of using Google's API is that train schedules can be impacted by current circumstances. Several international connections were for instance deleted in the last months due Covid-19 sanitary situation. The corresponding links were then added manually using past schedules from Deutsche Bahn database[3]. The query period used was then January 2020.

[3]Deutsche Bahn turns out to provide the most complete database regarding national and international train schedules in Europe, with the possibility of looking at past schedules. Thanks to Wouter Janssens from SNCB/NMBS for pointing that out.

2.3.2 Passenger traffic

Unlike for the EATN, there is no data available on the amount of passengers between stations in Europe. Nothing is recorded when a passenger gets into a train or steps out of it. Records of train tickets sold may be a good indicator, but such data is not available at the scale of the CoI. Moreover, this would probably not take into account monthly or yearly subscriptions.

2.4 Summary

To summarize, the scope of this study are the countries of the European Economic Area, minus Iceland. Routes and passenger traffic data regarding the EATN are those for June 2017.

Routes for the ERTN are obtained using Google Directions API which reflects the current state of train schedules in January-May 2020. Passenger flows are not available for the ERTN.

3 Topology of networks

Before comparing the air and rail transport networks in terms of paths and journey lengths, it is meaningful to understand the topological properties of those networks. The theory of complex networks provides several tools to understand, measure and compare real large-scale networks.

3.1 Methods

The EATN and ERTN are abstracted as networks $G = (V, E)$ with V the set of *vertices* (also called *nodes* or *points*) and E the set of *edges* (also called *links*, *lines* or *routes*). As a first step, cities are considered as nodes which means that airports or stations belonging to the same city are merged into 1 node. A similar approach was used in (Wang et al., 2011) to study the topology of the Chinese air transport network. An edge (u, v) represents a travel route between two cities u and v. The number of nodes is $n = |V|$, $m = |E|$ is the number of edges or links.

3.1.1 Complementary cumulative degree distribution (CCDF)

In the ERTN, it seems clear that an edge between u and v can also be used to go from v to u. In other words, the graph is undirected. On the other hand in the EATN, it is not because an airline operates a route between Paris and Athens that the same routes exists in the other way. However, (Li and Cai, 2004) shows that we can expect a very strong in-out-degree correlation in a directed air transport network. We will thus consider the EATN as undirected as well.

The *degree* of a given node i is defined as the number of edges from this node to other nodes of the network. These nodes are called the *neighbors* of i. For a network with n nodes, if n_k of them have degree k, the network *degree distribution* $p(k)$ is defined as the fraction of these k-degree nodes, i.e. n_k/n. The *complementary cumulative degree distribution (CCDF)* (often simply called *cumulative degree distribution*) is then defined as

$$P(k) = \sum_{k'=k}^{\infty} p(k'). \tag{3.1}$$

In other words, $P(k)$ is the fraction of nodes having degree k or more.

3.1.2 Average path length (APL)

The *shortest path length* d_{ij} is the shortest distance between two nodes i and j along the edges of the graph. In an unweighted network, d_{ij} is simply the number of edges on the shortest path between i and j. However, in transport networks edges are often given a weight expressed in distance or travel time. The *shortest path* is then the path minimizing the sum of edge weights along the path and the distance d_{ij} is then the sum of those weights.

Remembering that the graphs are undirected, we can then compute the *average path length* L for all the departure-arrival city pairs in the graph as

$$L = \frac{2}{n(n-1)} \sum_{i>j} d_{ij}. \tag{3.2}$$

In this chapter edge weights are not taken into account, L represents thus the average number of flights/trains (also called hops) that a traveller will need to take to go from one city to another. The *diameter* of the graph is the length of the longest shortest path, i.e. the maximum number of links that one will need to take to go from one node to another. In Chapter 4 however, edges will be weighted according to the travel time and the average path length is then the average time needed to go from one city in the network to another.

3.1.3 Clustering coefficient (CC)

The *local clustering coefficient* C_i measures the density of edges among the neighborhood of a node i. In a social network for instance, the clustering coefficient is the fraction of a user's friends which are also friends of one another. It is computed as the number of edges E_i between the k_i neighbors of i over the total number of edges possible between those neighbors (Newman, 2018)

$$C_i = \frac{E_i}{\frac{1}{2}k_i(k_i-1)}. \tag{3.3}$$

For nodes of degree zero or one C_i is conventionally set to 0. Taking the average of the local clustering coefficients over the nodes yields the network *clustering coefficient*

$$C = \frac{1}{n} \sum_{i \in V} C_i. \tag{3.4}$$

Note that this definition from (Watts and Strogatz, 1998) is different from the definition commonly used in graph theory where the clustering coefficient is computed as the fraction of paths of length two in the network that are closed.

3.1.4 Network models

A *random network* is a network model in which the values of certain properties are fixed, but the network is in other respects random (Newman, 2018). The fixed properties are typically the number of vertices n and the number of edges m, with the edges then being randomly distributed among the vertices. Random networks typically have a low clustering coefficient due to the random distribution of the edges. Coming back to the social network example, if a user's friends are chosen randomly in the network it is unlikely that those users will also be friends one of another (except if the graph is dense, i.e. almost all users are friends one of another). It is however not how social networks usually work, in fact they will typically have a high clustering coefficient. It can be shown that the diameter of random networks grows as $\ln(n)$, i.e. they typically have a small diameter.

Several real-world networks as the Internet or social networks were found to have a small diameter. However, real-worlds networks also tend to have a high clustering coefficient which does not fit the random network model. The *small-world network* model is introduced in (Watts and Strogatz, 1998) to illustrate how the properties of high clustering coefficient and low diameter can coexist in various real-world networks. In (Amaral et al., 2000), the small-world model is defined more formally as networks with a (1) a high clustering coefficient (typically much higher than that of random networks with the same number of vertcies and edges) and (2) a low diameter, increasing logarithmically with the number of vertices. Several transport networks were shown to be small-world networks as developed in Section 1.2.

Scale-free networks are networks for which the cumulative degree distribution (CCDF) follows a power-law, i.e.

$$P(k) \sim k^{-\alpha} \tag{3.5}$$

where α is known as the *exponent* of the power law. A known example of scale-free network is the World Wide Web[4]. In scale-free networks, the few nodes at the tail of the distribution have a significant fraction of the links. For example, in the World Wide Web 50% of the hyperlinks go to 2% of the most referenced pages (Newman, 2018).

3.2 European Air Transport Network

The European Air Transport Network has 184 nodes (cities) and 3927 edges (routes). We consider that there is a route between two airports if a given airline operates a flight at least once a week between those airports. As a reminder, when a city has several airports those are merged into one node. The resulting network is illustrated on Fig. 3.1 where nodes have a size proportional to their passenger traffic. The five biggest cities in terms of passenger flow are, in decreasing order, London, Paris, Amsterdam, Frankfurt and Madrid. Together they account for 28% of the total traffic.

[4]Difference between the Internet and the World Wide Web: the Internet is a global network of networks while the Web, also referred formally as World Wide Web (www) is collection of information which is accessed via the Internet. Another way to look at this difference is; the Internet is infrastructure while the Web is service on top of that infrastructure. Source: GeeksForGeeks.

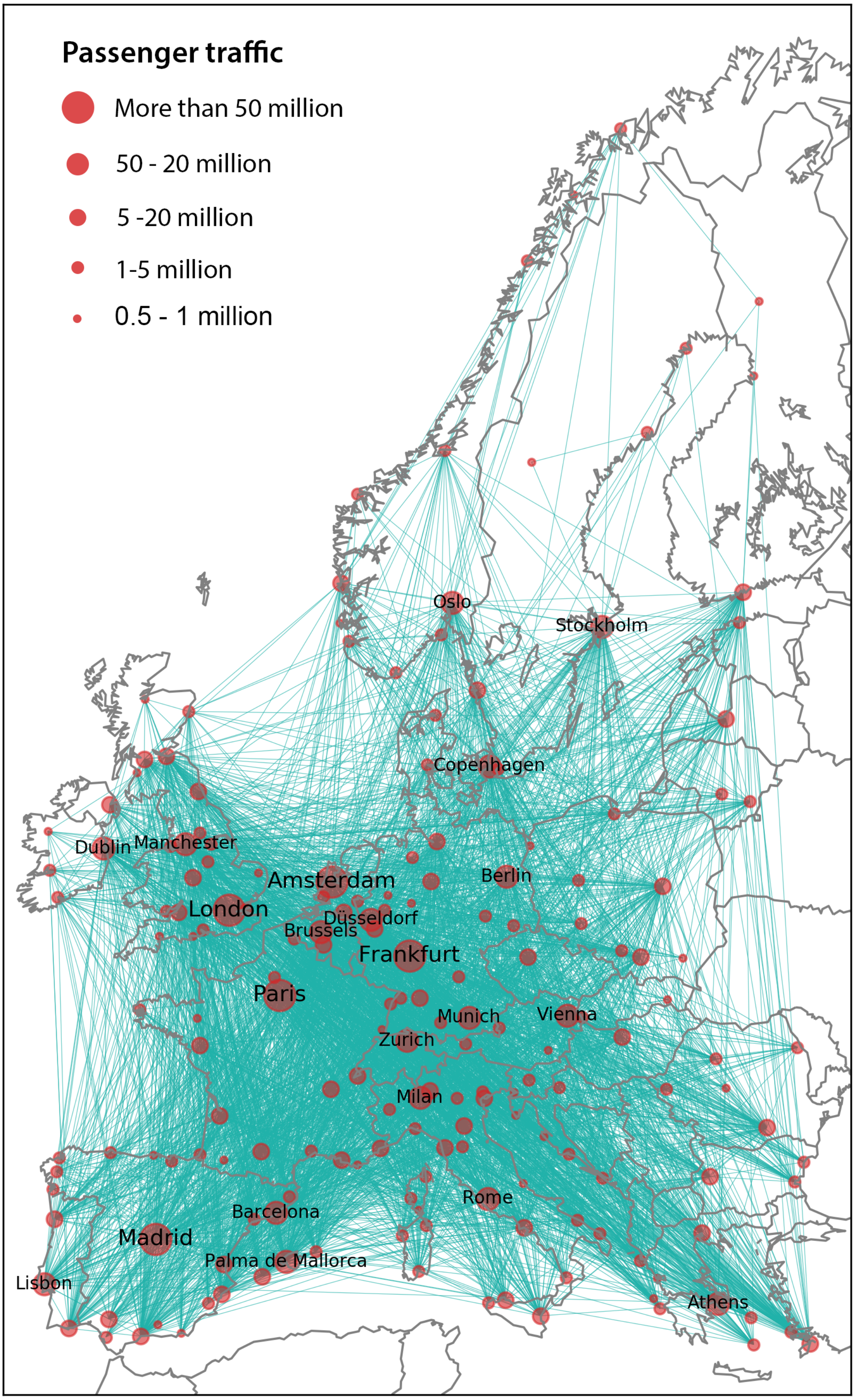


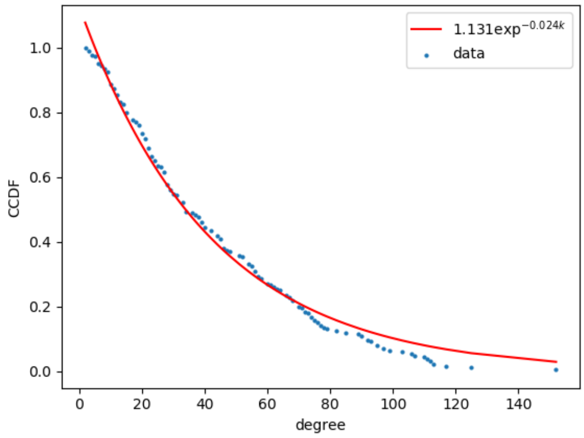
Figure 3.1: European Air Transport Network, cities by air passenger traffic.

The cumulative degree distribution (CCDF) is quite well fitted by an exponential rule as shown on Fig. 3.2a. R-squared (R^2) is a measure the goodness of a fit, expressed as the fraction of the variance captured by the model over the total variance in the data. For $P(k) = 1.131e^{-0.024k}$ we get $R^2 = 0.942$. We also tried to fit the cumulative degree distribution with a power-law but it turns out to captures rather poorly the observed distribution. The best fit using non-linear least squares is indeed given by $P(k) = 2.046k^{-0.452}$ and has $R^2 = 0.588$.

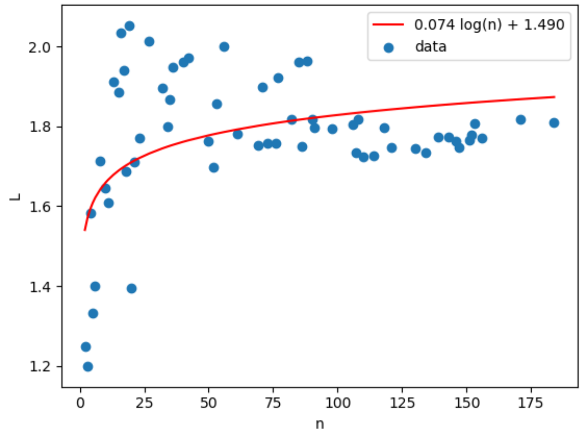
The exponential decay of the cumulative degree distribution indicates that a few cities have a high degree while many cities only have a few direct links. The cities with the most routes are London, Amsterdam, Paris, Barcelona and Palma de Mallorca. These busy cities act as hubs, redirecting passenger flows between regions. This structure can be partially explained by the preferential attachment process, i.e. new nodes prefer to connect to well connected nodes in the graph. Networks growing solely according to a preferential attachment mechanism however tend to have a cumulative degree distribution which is better fitted by a power-law rather than an exponential distribution (Barabasi, 2016). In such networks (referred to as scale-free networks, see Section 3.1.4) the degree of the main hubs is even higher as reflected by the long tail of the power-law, and these hubs have a significant fraction of the network routes. In the EATN the dominance of hubs is not so extreme as reflected by the exponential distribution of the degrees. The highest degree is 152 (London), just a couple times the average node degree (42.68). The median is 33 and the mode 9, showing again that the cumulative degree distribution is right skewed.

The average path length (APL) (Eq. 3.2) is a good measure of the efficiency of a network. For the EATN, the average shortest path has length 1.81, which is similar to the US air network (between 1.84 and 1.93, Xu and Harriss (2008)) but lower than the Chinese air network for instance (2.23, Wang et al. (2011)). The diameter of the graph is 3, with 23.32% of the city pairs reachable in 1 flight, and 95.75% of the pairs reachable in at most 2 flights.

Random and small-world networks are characterized by a short diameter, increasing logarithmically with the number n of vertices. The topological size of the network can also be characterized by the average path length between two vertices (Amaral et al.,



(a) Cumulative degree distribution in the EATN.



(b) Average path length in the EATN as a function of the number of nodes.

2000). We would therefore like to see how the APL varies with the number of vertices. To that end, we compute the APL in the air transport subnetwork of each of the countries of interest. We then extend the national subnetworks to the neighboring countries to obtain larger networks and we recompute the average path length on those networks. We repeat the process by adding the neighbouring countries of the extended subnetworks until the fourth "neighboring degree", where the subnetworks span most of the original network. The average path length L as a function of the number of nodes n in the graph is plotted on Fig. 3.2b together with the best logarithmic fit (in the least-square sense). If the results obtained do not fit completely a logarthmic increase with the number of nodes (probably because n is small compared to other large-scale networks like the Internet), the APL appears to grow at most slowly with the number of vertices.

The network clustering coefficient (CC) is 0.60 which is much higher than that of a uniform random graph with the same number of nodes and edges (0.23). With a small topological size (growing slowly with the number of nodes) and a high CC, the EATN has the two main properties of small-world networks in the sense of (Amaral et al., 2000). As discussed in Section 1.2, it turns out that many transport networks like the EATN tend to be small-world networks (Newman, 2003).

Another phenomenon to be observed is what we call the rich-club phenomenon: the formation of cliques between the most busy cities. A *clique* is a set of nodes which are fully connected, i.e. there is an edge between each pair of nodes. The 22 cities with more than 20 million passengers per year have a clustering coefficient of 0.99 which means that they almost form a clique. High connectivity between hubs is probably one of the mean reasons behind the efficiency of air transport networks.

3.3 European Rail Transport Network

For the rail network, we take as nodes those same cities than in the EATN with more than half a million air passengers per month. We then add edges according to the trains currently operating between cities. More precisely, two cities are connected if a train stops in both cities. This approach is referred to as the space-of-changes or P-space in (Kurant and Thiran, 2006). In this framework, the number of edges on the shortest path between two cities is the number of changes the passenger has to make.

Unlike the EATN, the rail network is not fully connected: islands like Ireland, Mallorca, Sicilia, etc. are not connected to the mainland train network. Some cities like Pula in Croatia are also not connected to the main train network, these are not taken into account. It also turns out that the railway systems of Greece, Estonia and Finland are (currently) not connected to the main network[5]. There exists a section of line between Sofia (Bulgaria) and Thessaloniki (Greece) but it has not been operated since 2017 due to technical problems. Since we need a connected network to compare shortest paths, these countries will not be taken into account for the rest of this study.

The purpose here is not to span whole train network in the countries of interest (for which we do not have the complete information) but only the part of it that operates the connections between the cities considered. Many national train systems contain short and local lines which are not relevant when it comes to compare international transport modes. These are thus not taken into account to build the ERTN shown on Fig. 3.3. The network has 139 main nodes (corresponding to cities) and 1104 edges. As a reminder, in this chapter edges represent train routes between cities and not the real path taken by the train. For instance, since there are a daily trains that stops in Lille and Marseille we add an edge that goes straight between the two cities even if in reality the path goes

[5]As of May 15th 2020, a link has been reopened between Riga (Latvia) and Tallinn (Estonia), reconnecting Estonia to the main network component. It is however operated by local trains taking more than 6 hours and making 29 intermediate stops on the way. For this reason and because of the close deadline of this thesis we have decided not to add Tallinn to the ERTN.

through Paris. Like in the air network, edges represent operated routes and not real paths or infrastructure.

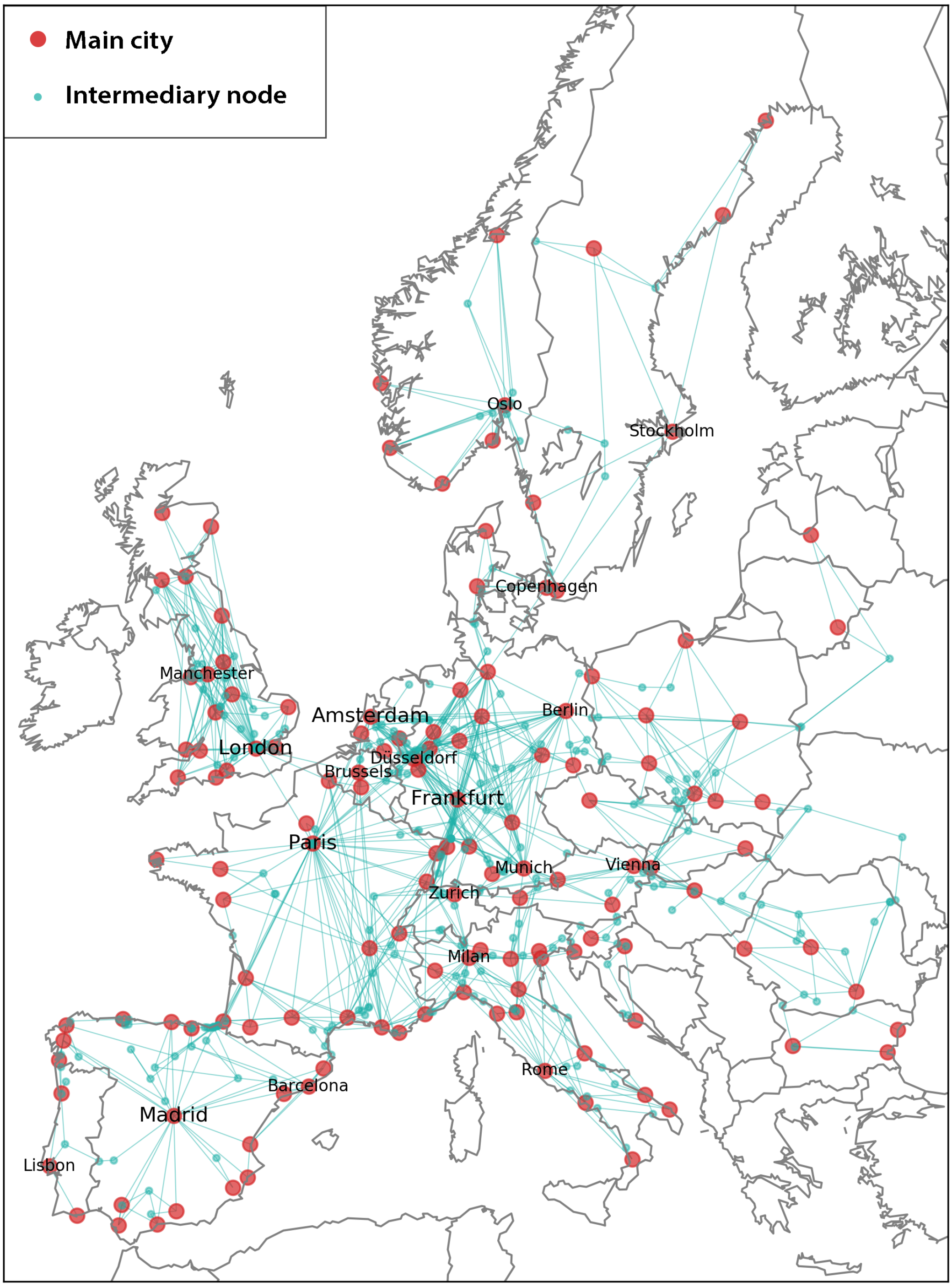


Figure 3.3: European Rail Transport Network.

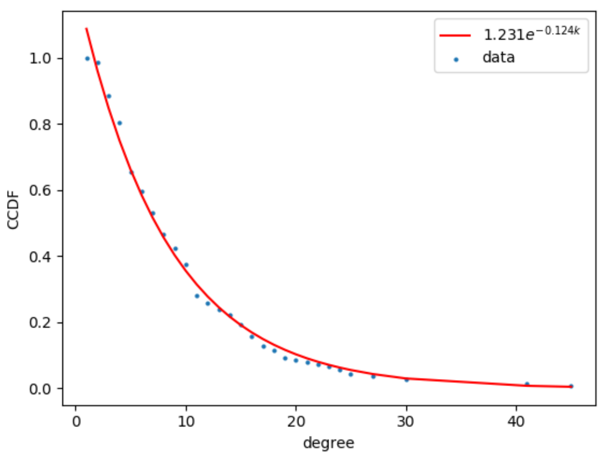
When travelling by train, it is sometimes not possible to go from one city to another without changing train at an intermediary station. Such nodes (drawn in turquoise on Fig. 3.3) do not belong to the set of cities considered but are just added to connect main cities when there is no direct path. They are often smaller towns with no significant airport, e.g. Antwerp (Belgium) which is on the shortest itinerary between Brussels and Eindhoven. There are 309 such intermediary nodes in our model of the ERTN.

In the European rail network, the average path length between (APL) main cities is 5.77. This means that a traveller going from one city to another in the network will need to take 6 edges on average, i.e. change trains 5 times. This is more than 3 times the average path of the air network where more than 95% of the journeys can be performed in at most 2 flights. In the ERTN only 13% of city pairs can be connected in at most 2 trains. The high number of connections needed will play an important role on the competitiveness of the train over the plane when travelling over long distances. The longest journey in terms of hops (i.e. the diameter of the network) is from Split (Croatia) to Faro (Portugal) which requires taking 18 trains. This is of course only relevant for the topology analysis of the network, not many people would go for such a journey in real life.

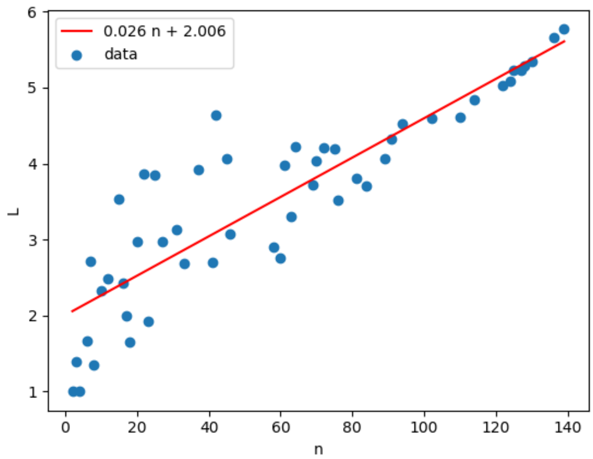
As for the EATN, we would like to see how the topological size of the network varies with the number of nodes n. As for the air network, we start by computing the average shortest paths at national level and then extend the subnetworks to neighbouring countries to obtain larger networks. The APL is computed over the new networks and the process is then repeated until the fourth neighbouring degree. We observe that the APL seems to grow faster than logarithmically with the number of nodes, in fact the observed data appears to be better fitted by a linear function as presented on Fig. 3.4b.

With a relatively high diameter (18) and a topological size which appears to grow faster than $\ln(n)$, the ERTN does not fit the small-world network model as defined by (Amaral et al., 2000). The short path lengths of small-world networks rely on the possibility of building "shortcut edges", i.e. edges linking nodes that are geographically far away in the network. With a few such edges the average path length can be reduced significantly while keeping a high clustering coefficient (Watts and Strogatz, 1998). Due to its very spatial nature, the rail network is constrained by the Euclidean distance and significant shortcuts are not possible. For example, an airline can easily operate a flight route between Lisbon and Warswaw but it is much harder to operate a train connecting those cities. Slower travel speeds and infrastructure requirements make trains less convenient to operate on long distances.

Figure 3.4a shows the cumulative degree distribution for the main nodes. Intermediary nodes are not taken into account since we do not look for all the links they may have (we only span the links on shortest paths between main cities). As for the EATN the node degree sequence seems to follow an exponential distribution which is best fitted for $P(k) = 1.231e^{-0.124k}$ ($R^2 = 0.974$). The best power-law distribution in terms of least-square residuals yields a much poorer fit ($R^2 = 0.646$ for $P(k) = 1.297k^{-0.620}$). As in the EATN, a few nodes have a high degree (hubs) while most of the nodes only have a few links. The decay is even faster than in the EATN which in practice will lead to an even higher hubness phenomenon (this is developed further in section 5.3). The average degree is 9.17, the median is 7 and the mode is 4 which shows again a right skewed distribution.



(a) Cumulative degree distribution in the ERTN.



(b) Average path length in the ERTN as a function of the number of nodes.

As for the degree distribution, we do not take intermediary nodes into account to compute the network clustering coefficient (for these nodes only the links on shortest paths between main cities are added). We get a clustering coefficient (CC) of 0.51 which is much higher than for a random graph with the same number of nodes and edges (0.03). The high clustering coefficient value is for some part a consequence of the space-of-changes approach used to model the ERTN. Since all nodes where a given train stops are connected, cities on the line operated by the train form a clique (in a clique, all the neighbors of a node are linked which yields a high clustering coefficient).

3.4 Summary

In this section we have seen that European Air Transport Network (EATN) is characterized by a low diameter and a high clustering coefficient which are characteristics of a small-world network. As in other real world networks, this enables a fast spread of information through the network (passengers in this case) reflected by a short average path length (APL). Flight networks benefit from a high spatial flexibility, only constrained by facility locations and air corridors.

On the other hand, the ERTN is a profoundly spatial network. Links are usually fixed and rely on track infrastructure. Lower travel speeds make distance a crucial factor when travelling in the ERTN. Shortcut edges in the graph are not physically doable, even if this tends to change a little with high-speed rail. As a consequence, the ERTN has a much longer diameter and the APL appears to grow faster than logarithmically with the number of nodes. Although it shows a high clustering coefficient, the ERTN does therefore not qualify as a small-world network.

Network	n	m	APL	diameter	CC ([6])	Shape of CCDF
EATN	184	3927	1.81	3	0.60 (0.23)	exponential
ERTN	139 + 309[7]	1104	5.77	18	0.51 (0.03)	exponential

Table 3.1: Main topological properties of networks.

[6]Between brackets, the clustering coefficient in a random network with same number of number of nodes and edges.

[7]Main cities + intermediary nodes.

4 Shortest paths between cities

4.1 Door to door journeys

With the topology of both the European air and rail transport networks in mind, we can now start computing shortest paths between cities. When you ask a traveller why he/she prefers the plane over the train the answer will often be "because it's faster". Of course, the flight time is typically shorter than the time you would spend on the train but what should be considered is the door-to-door travel time.

In terms of graph theory, this means that edges are now weighted by their travel time. The shortest path between two nodes is then the one minimizing the sum of edge weights along the path. For the rest of the study, we will only take into account as departure and arrival points the main cities that are connected to both the train and the flight network. In this manner, we can compare travel time between cities in both transport systems.

In this thesis, we consider the door-to-door travel time as the time needed to go from city center to city center. When travelling by plane this includes:

1. Going from the city center to the airport. The time needed is computed using Google Directions API with the fastest transport mode.

2. Going through check-in and security checks at the airport. The time needed for this can vary a lot from airport to airport and whether the passenger has to check in a luggage or not. In this study, a standard duration of 2 hours was allocated for this step[8].

3. Flying: the flight time is extracted from Eurocontrol flight records and averaged over all flight operating between the two airports.

4. Going from the destination airport to the city center, this is again computed using Google Directions API.

In case of a travel journey containing several flight segments, a minimum connection time

[8] Airline companies typically advise their customers to arrive 2 hours before departure at the airport. This may seem a lot for small airports. On the other hand, no time is added in this study to pick up luggage and exit the airport. However, a more precise evaluation of the check-in time for each airport could be a worthy further development.

of 1 hour is allocated between flights.

When travelling by train, the door to door journeys include:

1. Going from the city center to the train station.
2. Catching the train. A standard time of 15 min is allocated for this (standard connecting time used by the Belgian national rail operator).
3. Taking the train.
4. Going from the destination train station to the city center.

As above, travel time between the city center and the train station is computed with Google Directions API using the fastest transport mode. The journey time by train is also given by the API. When the traveller has to change trains, a connecting time of 15 min is allocated.

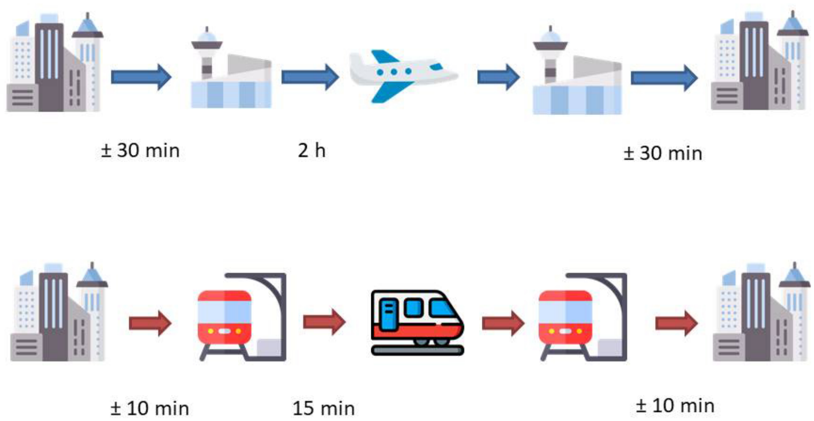


Figure 4.1: Door-to-door journeys in air and rail transport networks.

Door to door journeys can be modelled with graphs using directed edges as illustrated on Fig. 4.2. Directed edges can only be taken in one direction: from source node to destination node. The EATN and ERTN thus become directed graphs. Airports are splitted in three nodes: a check-in, a departure and an arrival node. Passengers arrive at the check-in node and then need to get to departure node to catch their flight. Flights run from the departure node of an airport to the arrival node of another airport. Finally, when a passenger arrives at the destination arrival node, he/she still needs to get to the city center node. Note that the check-in node is not necessary here, we could also have drawn an edge directly from the city center to the departure node with as duration the journey time to the airport plus the check-in time of 2 hours. It is however necessary to have a directed graph, otherwise it would be faster in order to take a flight to go from the city center to the arrival node and then the departure node.

The ERTN works in a similar way with only two nodes per station: a departure and an arrival node. Regarding air and rail routes, the new graphs are only directed in appearance since edges from the undirected graphs have simply been copied in both directions.

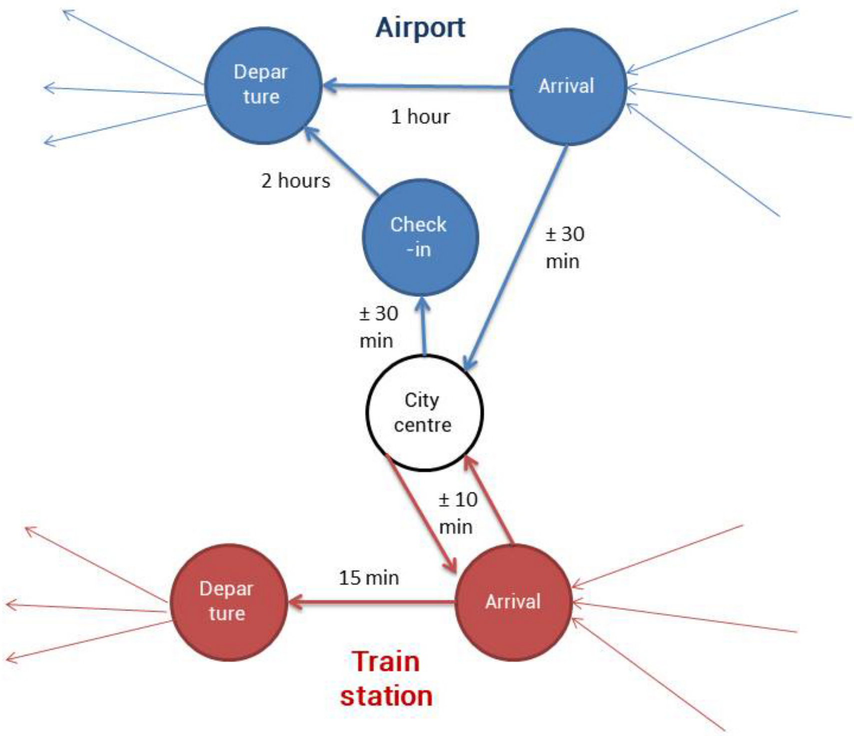
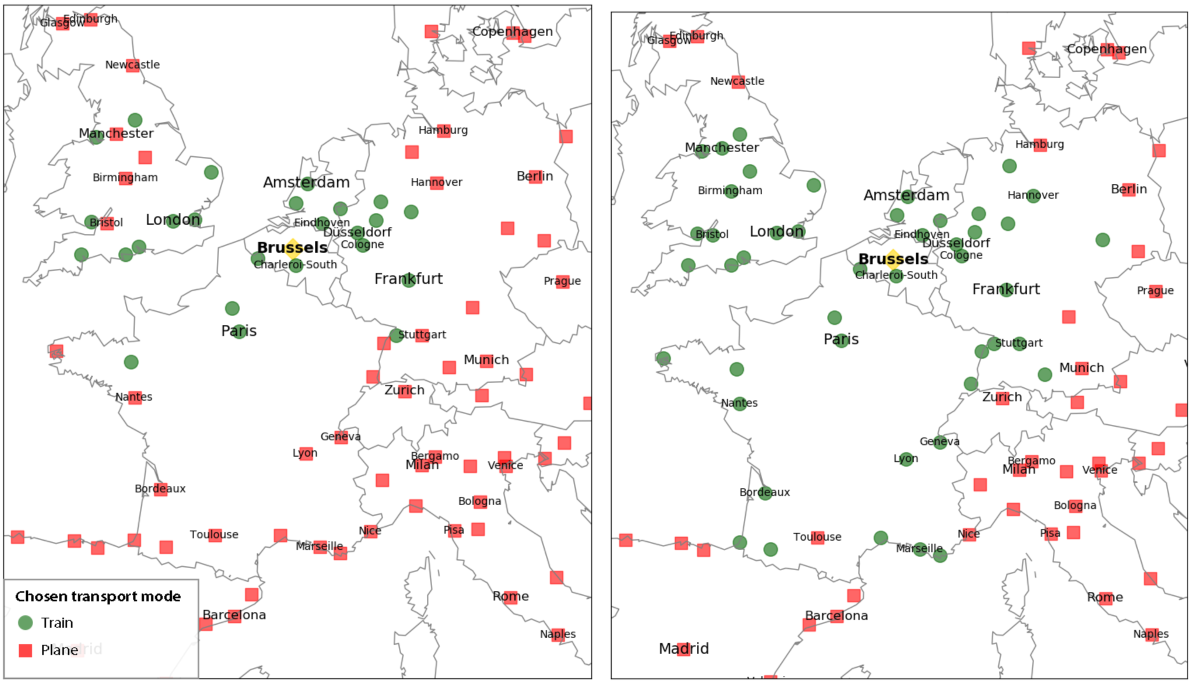


Figure 4.2: Nodes in door-to-door graph

If flying is usually faster than taking the train, airports are often located further from the city center than train stations. Furthermore, stepping into a train is much faster than catching a flight. For these reasons, taking the train can sometimes be faster than flying. Using door to door journey times, it turns out that 6.58% of the city-to-city journeys are faster by train (among the city reachable as well by train as by plane). As a reminder, here we only look at which city-to-city journeys are faster by train, not how many passengers actually take those journeys. The passenger traffic analysis on those routes will be the subject of the next chapter.

High-speed rail is considered a comfortable, safe, flexible and environmentally sustainable mode of transport (ECA, 2019). Sometimes, people are thus willing to spend more time travelling to take the train. In this regard, 9.40% of the journeys take less than 1 hour more by train and 13.03% take less than 2 hours more by train. As an example, Fig. 4.3 shows the best transport mode from Brussels if the passenger takes the fastest transport mode or if he/she is ready to spend up to 2 hours more to take the train. All city-to-city journeys which are faster by train are represented as a network in Appendix A2.



(a) When Brussels' passengers choose the train if the journey is faster than by plane.

(b) When Brussels' passengers choose the train if the journey is at most 2 hours longer than with the plane.

Figure 4.3: Chosen transport mode per city from Brussels.

In the first scenario (when passengers choose the train only if it is faster than the plane from door to door), the train is competitive on journeys to the Netherlands, Northern France, Eastern Germany and the Southern part of the UK. Even though the number of cities for which the train is chosen over the plane seems quite restricted, a significant number of flights were still operated between Brussels and those cities in June 2017. For that month, 274 commercial flights were for instance recorded to Frankfurt and 148 to Amsterdam.

If a passenger is ready to spent 2 more hours travelling to take the train, the train will be the preferred option on journeys to several other cities, including almost all main cities in France. In Fig. 4.3b, we start to see importance of high-speed rail (HSR) when it comes to compete with the plane on longer journeys, with cities like Marseille or Bordeaux displaying a travel time at most 2 hours longer by train. For other cities in Southern France which are not connected to the HSR like Toulouse or Nice, the travel time is typically over two hours longer by train. The importance of HSR in the modal shift from air to rail is discussed further below and in the next chapter.

A question that arises is "Up to which distance is the train competitive with the plane ?". To this end, we compute the travel distance by train between each pair of cities and put it in relation with the shortest path duration in both networks. This yields 19 182 distance-duration data points for each transport network. The travel distance by train is then binned in intervals of 50 km and averaged over each interval. The corresponding graphs are presented on Fig. 4.4 for distances of less than 1500 km. It turns out that, on average, the train is faster when the travel distance between cities is shorter than 444 km. Conversely, when the distance by train between two cities exceeds 444 km, the plane is on average faster.

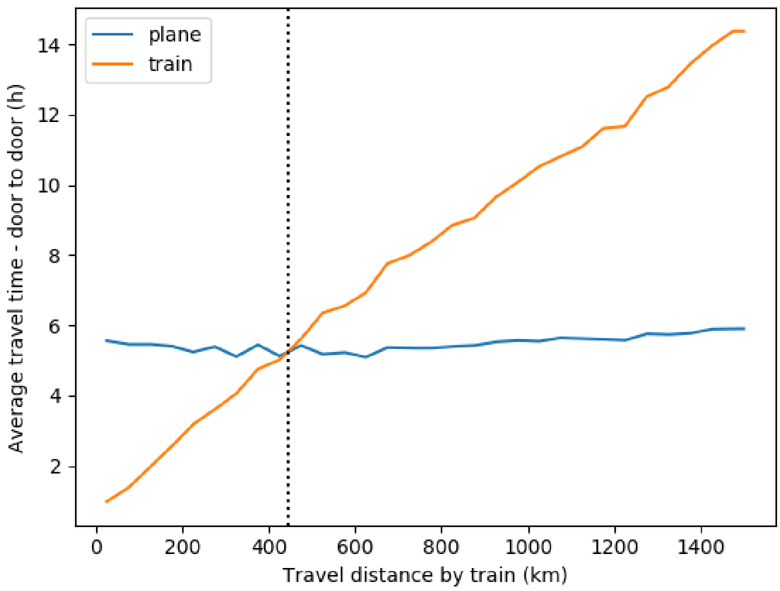


Figure 4.4: Average journey duration as a function of the train distance.

Of course, the competitiveness of the train depends on the speed of the lines which is heterogeneous when looking at all the ERTN. In this regard we looked at the connected subnetwork of the High Speed Rail (HSR) in Western Europe (see Fig. 4.5). That is, we only keep the nodes (cities) connected to HSR and only take into account journeys that can be made entirely with high-speed trains except for the transit from city center to the train station. Another exception is made when a passenger has to move between train stations inside a city, e.g. from Gare du Nord to Gare Montparnasse in Paris. Even though they have high speed rail, the HSR networks of Spain and Italy are not included here since they are not directly connected to the main component of the network (France, Belgium, Netherlands, Western Germany, London). Similarly, to travel to Eastern Germany trains have to use some regular track sections so we restrict ourselves to the German cities of Frankfurt, Cologne, Düsseldorf and Stuttgart.

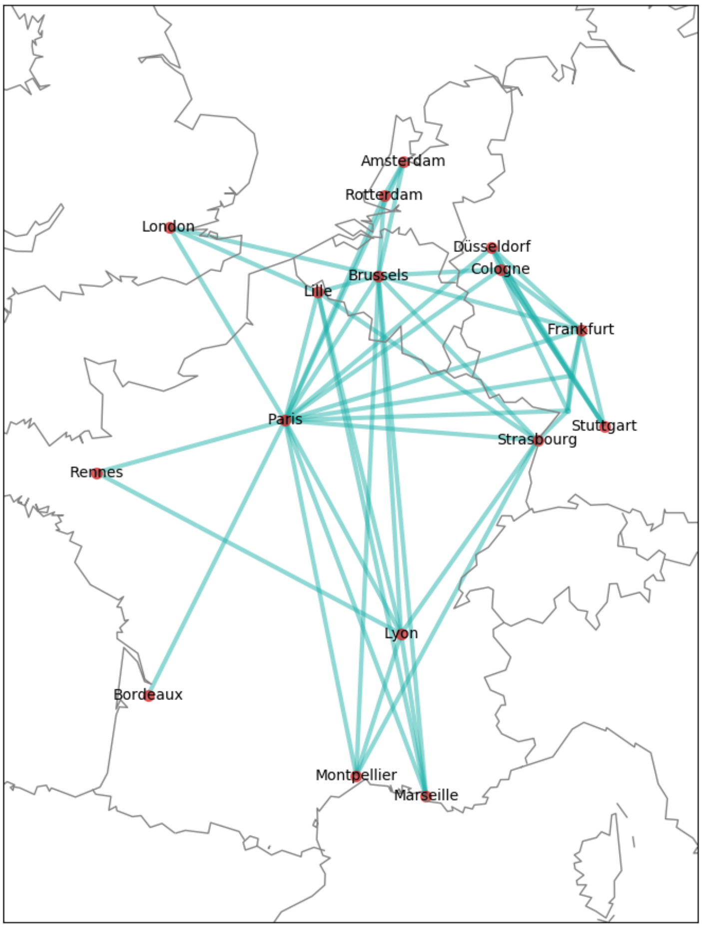


Figure 4.5: High speed rail subnetwork in Western Europe.

Using the HSR-only journeys, we get the distance-to-duration plot presented on Fig. 4.6. Here, the data is binned in intervals of 100 km due to a more restricted number of data points (186). The treshold under which the train is faster than the plane is less clear cut in this case but we conclude that, on average, high-speed rail is competitive over air transport up to a travel distance of 704 km. This result appears to be similar to assessments made on other HSR networks. In Spain, the Alta Velocidad Española (AVE) was found to be competitive up to 400 miles (643 km) (Albalate and Bel, 2012) and the Japanese Shinkansen is competitive up to 438 miles (704 km) (Taniguchi, 1992). To the best of our knowledge it is however the first time that such threshold distance is computed for the European rail network (as well for the whole network as for the HSR network in Western Europe).

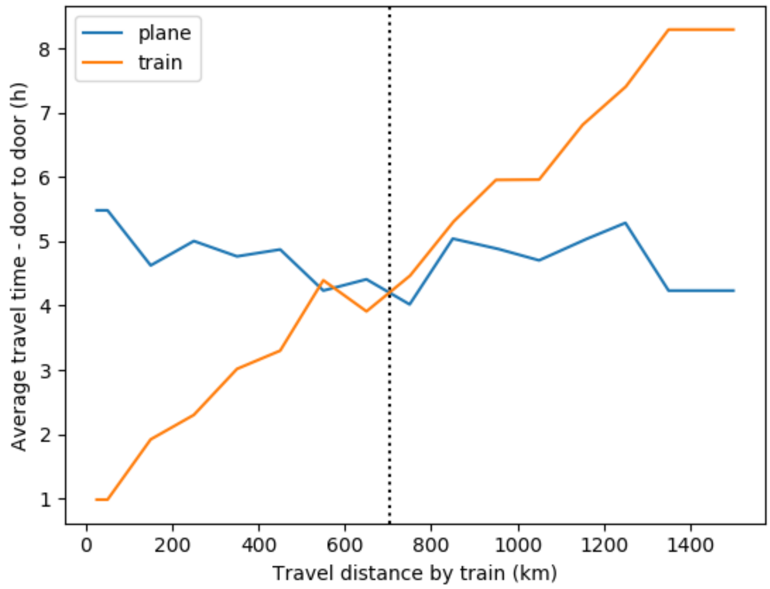


Figure 4.6: Average journey duration as a function of the train distance for cities connected to the HSR.

4.2 Closeness centrality

Some cities are well connected to the air network but poorly to the rail network, or vice-versa. An example of such a city is Charleroi in Belgium. With more than 7 million passengers a year, Charleroi airport is the second most important airport in Belgium. On the other hand, the city is poorly connected to the ERTN with only local train lines passing through it and no direct connection to any other of the main nodes considered except for Brussels.

Closeness centrality is a measure of the accessibility of a node in a graph. It is inversely proportional to the sum of the shortest distances to all other nodes in the graph. It is defined as

$$C_C(i) = \frac{n-1}{\sum_{j \neq i} d(i,j)} \tag{4.1}$$

where $d(i,j)$ is the shortest path length between nodes i and j and n is the number of nodes in the graph. The distance here is computed in terms of the travel duration from city center to city center. The higher the closeness centrality of a node, the faster it is to reach other nodes in the network from this node.

Closeness centrality values are computed for nodes connected to both networks and then normalized by the highest centrality value (see Fig. 4.7). The average node centrality is much lower in the ERTN than in the EATN, respectively 0.29 and 0.81. In the rail network, travel time to nearby nodes is short but grows fast for further nodes due to slow travel speed and many connections. The accessibility of all other nodes from a given node in the graph is thus relatively poor on average. On the other hand, once a passenger is at the airport and has been through security checks, most destinations are accessible very fast. Fast travel speed and a high number of direct routes make all nodes in the graph very accessible.

Centrality values have similar standard deviation in both networks (~ 0.07). However, in the ERTN the closeness centrality seems to be very dependent on the geographic position of the city, which is much less the case for the EATN. Cities far from the geographic center of the graph have a poor closeness centrality because most train journeys to other cities are long. This is again a consequence of the strong spatial nature of the ERTN constrained by the Euclidean distance. On the other hand in the EATN, cities like Malaga (Spain) or Gdansk (Poland) have a high closeness centrality despite their eccentric position. As seen in Section 3.2, the air transport network has characteristics of a small-world network with "shortcut edges" in the graph allowing for short paths even to remote nodes. As a consequence, even eccentric nodes are easily accessible from other nodes in the graph.

Finally, the correlation between node closeness centrality in the air and rail network is 0.55 which indicates that cities tend to be connected in a relatively similar way to both networks.

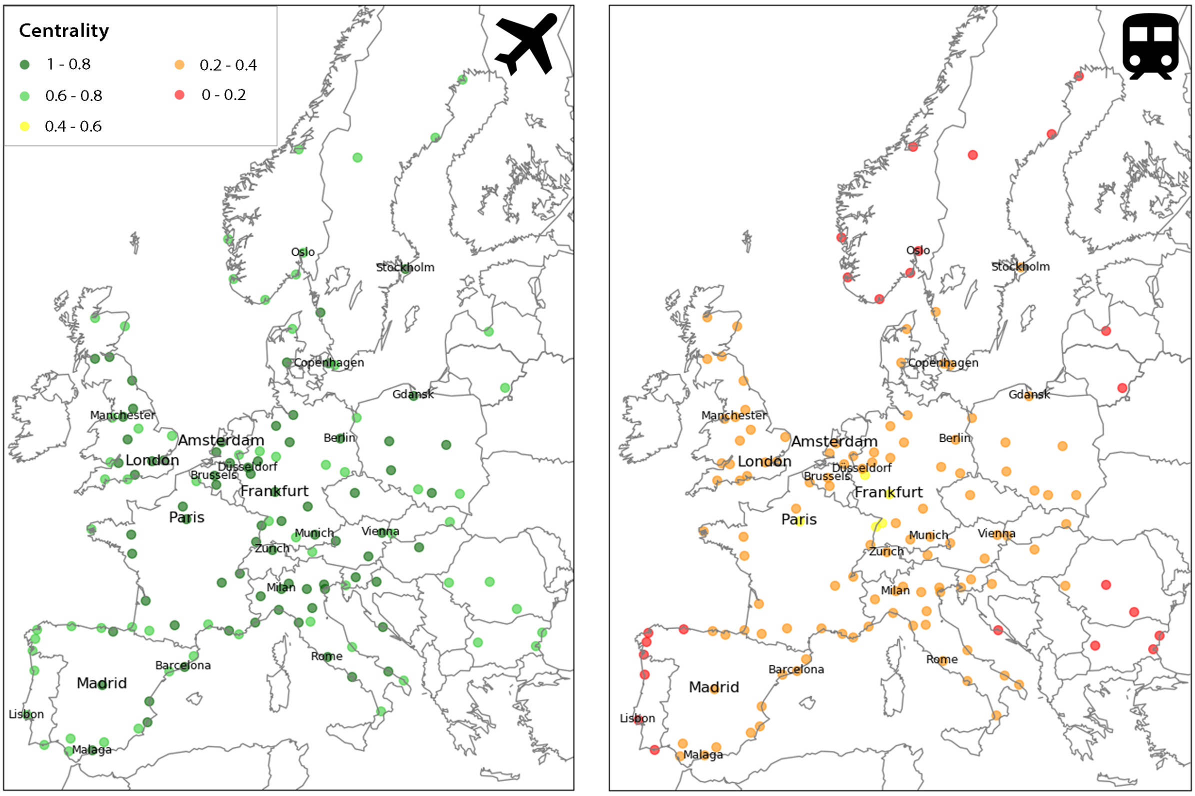


Figure 4.7: Closeness centrality.

4.3 Summary

In this section it has been shown how directed networks can be used to model city center to city center journeys. Using weighted edges, we can compute shortest paths in terms of shortest travel time. It turns out that 6.58% of the journeys between the cities considered are faster by train, and 13.03% do not take more than 2 hours extra by train than by plane. In terms of distance, the train turns out to be competitive until a distance of around 444 km. When considering only the HSR network in Western Europe, the train is competitive over the plane up to 704 km on average.

Closeness centrality was studied as a measure of node accessibility in both transport systems. Nodes in the EATN tend to be much more accessible than in the ERTN, the latter being strongly constrained by the Euclidean distance.

5 Passenger traffic and modal shift

5.1 Passenger transfer from air to rail

In the last section we identified which door-to-door journeys are shorter by train than by plane (Appendix A2). Using the passenger traffic data on air routes from Eurostat, we can then link this with the number of passengers who could possibly switch from the plane to the train. In June 2017, more than 4 million passengers took a direct flight between two cities for which the train is faster from city center to city center. This represents 5.61% of the traffic between the airports considered, 2.81% of the total air traffic[9] and 21 thousand flights.

With only global traffic data per flight route, we are however not able to assess the number of passengers taking the plane when there is no direct flight between the two cities considered. The journey from Lille (France) to Newcastle (UK) is for instance faster by train than by plane (5:21 and 5:56 respectively). The fastest path in the flight network goes from Lille-Lesquin airport to Paris Charles de Gaulle Airport and from there to Newcastle International Airport. We are however not able to evaluate which share of the passengers on the flight route between Lille airport and Paris Charles de Gaulle are taking a journey between Lille and Newcastle, and the same goes for the Paris-Newcastle flight route. To the best of our knowledge such refined data is unfortunately not available yet. The exact number of passengers who took the plane on journeys where the train is faster from door to door is therefore probably higher than the 4 million cited above.

As mentioned earlier, passengers are sometimes also willing to spend more time travelling to take the train rather than the plane. The more time passengers are willing to spend travelling by train, the more traffic can be taken off from the EATN, in terms of passengers as well as in number of flights (Fig. 5.1). If passengers took the train when the travel time is at most one hour longer than with the plane, the number of air passengers per month would decrease by 11 million (again, between cities where there is a direct flight). We observe a linear-like increase of the number of passengers that could operate a modal shift with the number of acceptable extra travel hours. For each additional hour, about

[9] All flights departing from the CoI: national, international intra-EU and international extra-EU.

5 million passengers are transferred from air to rail. The same relation can be observed regarding the number of flights, with about 25 thousand flights less per additional hour spent in the train. Detailed results are gathered in Appendix A3.

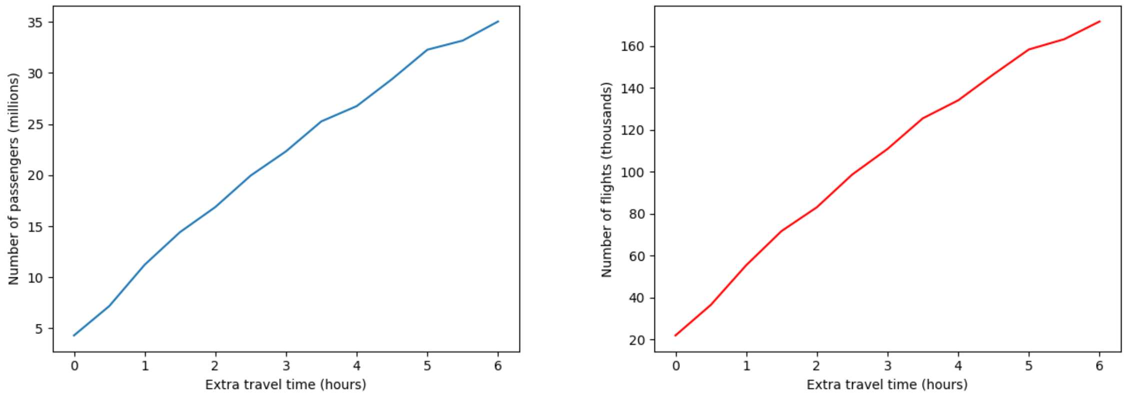


Figure 5.1: Passengers and flights drop per month in the EATN as a function of the number extra hours that passengers are willing to spend on train travel.

5.2 Impact on rail lines

We have now identified some of the flight routes that could be replaced by train journeys, the next step is to assess what the impact would be of transferring those passengers to the ERTN. Each flight route removed from the air network brings a given number of passengers which enter the rail network at a given node (departure city) and need to exit the network at another node (arrival city). To travel from one node to another, passengers will use the shortest path between the two nodes since it is for this specific path that the train is competitive with the plane. All edges along the shortest path will be assumed to transport a flow equal to the number of passengers from the original flight route. Let us recall that rail lines are duplicated in two directed edges, one in each direction. We are thus able to compute the flow on each line for both directions.

Adding the passengers of the "dropped" or deleted flight routes successively to the ERTN, we obtain the total passenger traffic throughput or flow on each edge. The flows are per month since we have air traffic data for one month. For the rest of this chapter we will study 2 scenarios:

1. Air passengers from door-to-door journeys which are faster with the train are transferred to the ERTN (Scenario 1).

2. Air passengers from door-to-door journeys which take less than 2 hours more by train than by plane are transferred to the ERTN (Scenario 2).

As above, we can only evaluate the number of passengers between the two cities considered if there is a direct flight between them. If the shortest path includes several flight routes, it is not possible to assess the number of passengers taking the specific city-to-city journey from the global traffic data per route.

As presented in Appendix A3, 4.31 million passengers are transferred to the ERTN in the first scenario and 16.91 million in the second scenario. Fig. 5.2 illustrates how the flows arising from this modal shift are distributed among the network in both cases.

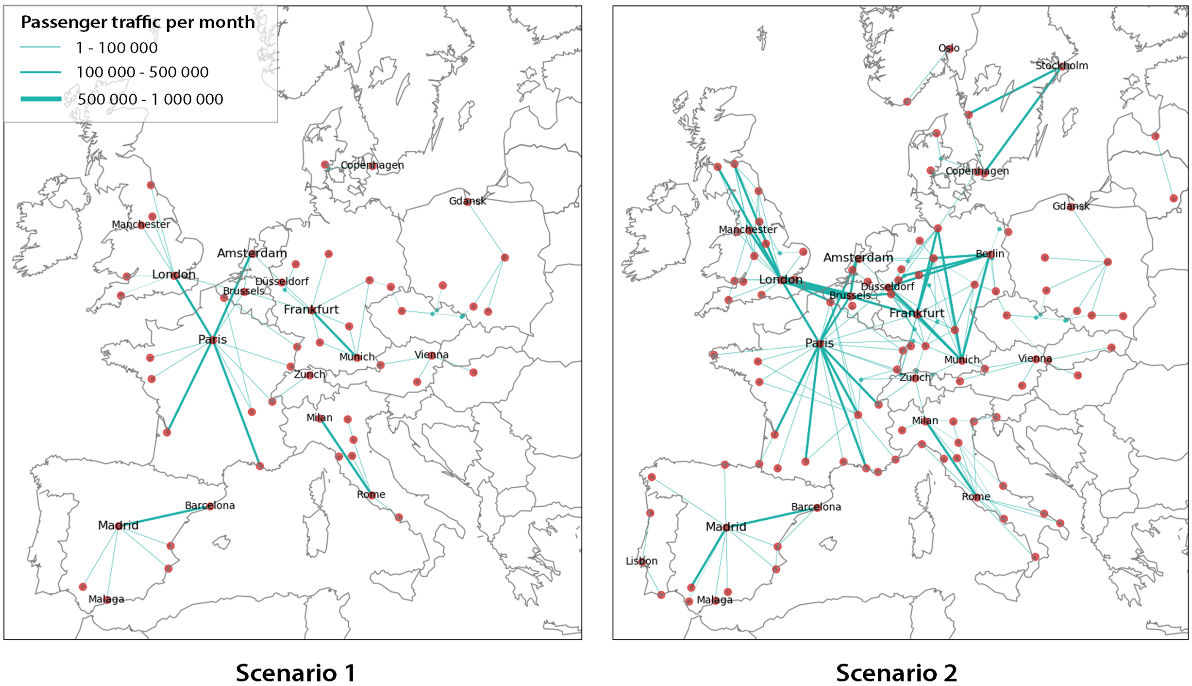


Figure 5.2: Additional traffic in ERTN due to passenger transfer.

When comparing with the HSR subnetwork in Fig. 4.5, it seems that a significant part of the additional passengers is transferred to high speed-rail lines. To measure the share of kilometers travelled on HSR, we used the passenger-kilometer theoretical measure from Eurostat. A passenger-kilometre (pkm) represents the transport of one passenger by a defined mode of transport. Here, we use a more general definition of high-speed rail as lines where the average speed is higher than 150 km/h. This is the definition used in (Pyrgidis, 2016) without the condition on the maximum achievable speed since we unfortunately do not have no such data. This allows to include additional edges where the whole line is not equipped with high-speed infrastructure. For example, the Paris - Nantes connection runs on high-speed tracks from Paris to Le Mans and then on regular tracks from Le Mans to Nantes. However, with an average speed of 170 km/h, this link can be considered as HSR and is indeed operated by high-speed trains. On the other hand, it is noteworthy that in practice not all HSR routes can achieve an average speed of 150 km/h. For example, high-speed trains operating on the line between Brussels-South and Cologne have an average speed of 128 km/h.

With this definition of high-speed rail in mind (average speed $>$ 150 km/h), we can compute the fraction of additional passenger-kilometers (pkm) travelled on HSR. Results for both scenarios are summarized in Table 5.1. It turns out that high-speed rail represents a big share of the additional pkm travelled: 73% in the first scenario and 48% in the second. This highlights the vital importance of HSR when it comes to transferring passengers from air to rail.

Scenario	Max additional travel hours	Total pkm (billions)	HSR pkm (billions)	Share of HSR (%)
1	0	1.96	1.44	73.60
2	2	9.83	4.74	48.26

Table 5.1: Additional pkm per month on ERTN due to passenger transfer from air transport.

Tables 5.2 and 5.3 give the lines with the highest additional traffic per month in both scenarios. For each line, the traffic in both directions turns out to be almost equal with a difference of at most 1000 passengers/month. Results are thus given per line, with the traffic applying in both directions. The share of passengers for each line is the ratio of the number of passengers using that line over the total number of passengers transferred from the EATN.

In both cases, the five most busy lines are high-speed rail lines. We unfortunately lack data on the capacity and current use of those lines. However, the European Court of Auditor special report on high-speed rail (ECA, 2019) indicates that a significant share of HSR lines in Europe have an insufficient number of passengers, causing sustainability issues. It is thus reasonable to think that some of these lines have the capacity to carry additional passengers.

A likely bottleneck would probably be the HSR line in the Channel Tunnel. For this particular line we see the limitations of the space-of-changes approach where we link two nodes if a train stops in both cities, regardless of the infrastructure. In our graph, 3 edges cross the English Channel to London: one from Paris, one from Brussels and one from Lille. In fact, those three lines join in Calais to cross the sea. The passenger flow that needs to pass through the tunnel is thus the sum of the edge flows: 257 thousand passengers in the first scenario and 1.37 million in the second scenario. This represents respectively 5.95% and 8.11% of the additional number of passengers per month.

Departure station	Destination station	Additional passengers per month (thousands)	Share of passengers (%)
Madrid-Puerta de Atocha	Barcelona-Sants	214	4.97
London St-Pancras	Paris Gare du Nord	196	4.55
Paris Gare Montparnasse	Bordeaux Saint-Jean	154	3.57
Paris Gare de Lyon	Marseille Saint-Charles	154	3.56
Paris Gare du Nord	Amsterdam Schiphol	116	2.69

Table 5.2: Train lines with most additional traffic, scenario 1.

Departure station	Destination station	Additional passengers per month (thousands)	Share of passengers (%)
London St-Pancras	Brussels-South	806	4.77
London St-Pancras	Paris Gare du Nord	449	2.67
Brussels South	Amsterdam Schiphol	407	2.41
Madrid-Puerta de Atocha	Barcelona-Sants	357	2.11
London King's Cross	Edinburgh Waverley	315	1.87

Table 5.3: Train lines with most additional traffic, scenario 2.

5.3 Betweenness centrality

The previous section identified the most busy train edges when transferring passenger flows from air to rail. Regarding nodes, the importance of a given node to connect other nodes from the network can be computed using *betweenness centrality*. Formally, the betweenness of a node i is the ratio of all shortest paths passing through it:

$$C_B(i) = \sum_{i \neq j \neq k} \frac{\sigma_{kj}(i)}{\sigma_{kj}} \tag{5.1}$$

where $\sigma_{kj}(i)$ is the number of shortest paths from k to j passing through i and σ_{kj} is the number of all shortest paths from k to j.

In our transport networks, we typically have only one shortest path between each pair of nodes. However, we can weight Eq. 5.1 with the flows by defining $\sigma_{kj}(i)$ as the number of passengers passing through i on the shortest path between k and j. In other words, $\sigma_{kj}(i)$ will be equal to the number of passengers from k to j if i lies on the shortest path between them and 0 otherwise. The betweenness centrality is then a measure of the share of passengers that pass through a node i on shortest paths between other nodes.

The higher the betweenness centrality of a node, the higher its importance as a hub to transfer passengers in the network. High betweenness values also point out bottlenecks of the network: airports or stations many passengers have to go through and where important flows thus need to be managed. When one of those nodes experiences a technical problem, a significant share of the network flow is likely to be impacted.

Fig. 5.3 illustrates the betweenness centrality of airports and stations using the flows from scenario 2. This means that the passenger flows in the graph represent the traffic on journeys where the train is at most 2 hours longer. We observe similar betweenness centrality values in both networks, but not always for the same cities. In the EATN, the most central airports are Paris Charles de Gaulle, Frankfurt and London Heathrow with respectively 13%, 13% and 10% of the traffic passing through these nodes.

In the ERTN, the most central stations are St-Pancras International (London), Brussels-South and Frankfurt main station. Respectively 16%, 11% and 10% of the passengers pass through these stations. This points out the importance of Brussels as an important link between Germany, the Netherlands and the UK in the ERTN. In comparison, only 3.3% of the traffic passes through Brussels Airport in the EATN (3.5% if we add Charleroi-Brussels-South airport). No train station in Paris is among the 3 biggest hubs despite its very good connection to HSR. This is due to the fact that the traffic is divided among 3 stations: Gare Montparnasse, Gare du Nord and Gare de Lyon which are respectively 5^{th}, 6^{th} and 7^{th} most central stations.

The mean betweenness centrality among the 10 largest hubs is 9% higher in the ERTN than in the EATN which indicates a slightly higher hubness phenomenon in the rail network. As before, this is probably due to the strong spatial constraints in the ERTN. If one goes from mainland Europe to the UK by train, the station of London St-Pancras is a necessary waypoint (end station of the Eurostar). On the other hand, when travelling by plane London is not a mandatory spot if one wants to go to Manchester for instance. Similarly, almost all journeys from Belgium, the Netherlands or the UK to the main cities of France pass through one of the several train stations in Paris.

Betweenness centrality also shows that main network hubs and bottlenecks are not always the same in both transport systems. Beyond line capacities, station capacities need to be taken into account when transferring passengers from air to rail. Central cities like Brussels (geographically speaking) may gain importance as transport hubs following a modal shift while geographically eccentric cities may lose passenger traffic.

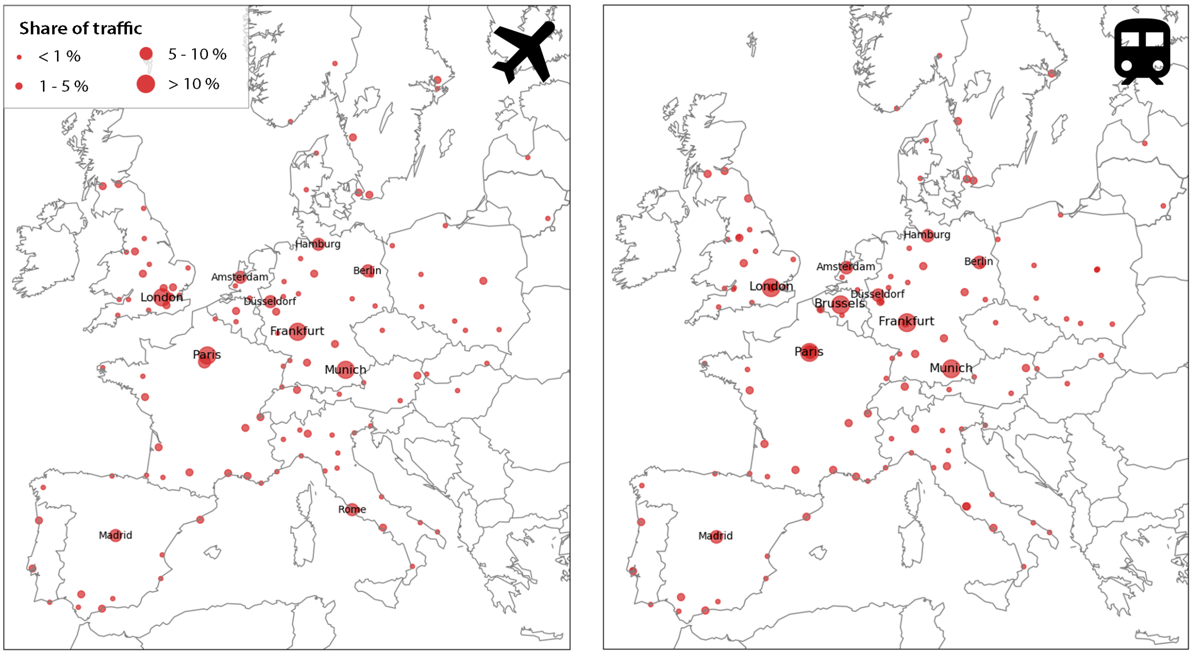


Figure 5.3: Betweenness centrality using flows from senario 2.

5.4 Applications

For the rest of this chapter we present two applications of our city-to-city network model for the European air and rail transport networks. We first study night trains as an alternative to the plane on journeys where high-speed infrastructure is not available. We then assess the potential traffic on a future high-speed rail line between Lyon (France) and Torino (Italy).

5.4.1 Night trains

The number of passenger night trains operating in Europe has fallen since around 2010 due to an increased competition from low-cost airlines, high-speed trains and overnight coaches (European Parliament, 2017). In the last years, the market for night train services has however picked up somewhat, mainly in response to growing concern on the pollution generated by the air transport industry. For example, Austria's ÖBB Nightjet service has recently opened a new route from Vienna to Brussels and the Swedish rail operator SJ is considering a night route from Malmö (Sweden) to Brussels.

For journeys longer than 440 km, where day trains are usually not competitive anymore with the plane in terms of travel time (see Section 4.1), night trains could be an alternative to air transport when high-speed infrastructure is not available. In France for example, SNCF Intercités de Nuit operates a daily night route between Paris and Toulouse (which is not connected to high-speed infrastructure). A consortium of railway undertakings also operates a route between Paris and Milan as the French and Italian HSR networks are currently not connected.

Night trains usually leave the departure station in the evening and reach their destination station on the morning of the following day. They are usually equipped with sleeping berths allowing passengers to rest during most of the journey. The CO_2 emissions of a passenger travelling by night train depend on several factors, including the energy source and load factor of the train, but they have been shown to be usually much lower than when travelling by plane. The International Union of Railways (UIC) evaluated that the average CO_2 emissions per passenger were between 11 and 13 g/km on a night train, compared to 185-215 g/km on a plane (Interational Union of Railway, 2013).

A certain number of overnight train routes are still operated and we have gathered information on some of them in the table below. Average speeds are given between departure and arrival stop, including stopping times at intermediate stops.

Route	Duration	Distance (km)	Average speed (km/h)
Paris-Toulouse	7h56	713	90
Stockholm-Lulea	12h22	1056	85
Watford Junction (London) - Inverness	12h74	964	75
Milano-Lecce	11h40	1012	87
Bucarest-Timisoara	9h44	535	55

Table 5.4: Some night train routes in Europe.

The routes described in Table 5.4 have relatively low average speeds, this is in fact typical for overnight train services. It is due to several factors including the fact that night trains need a minimum journey time (more on this later) and that they sometimes need to adapt to the low operating speed of freight trains which run mostly at night. We therefore limit the travel speed in the network at 90 km/h to reflect the real operating conditions.

They are currently no high-speed trains in Europe operating night services. One of the main reasons is that maintenance operations on high-speed infrastructure mostly take place at night. Some high-speed lines also carry freight traffic during the night. Moreover, track access charges on HSR are often considered too high for night train services (Interational Union of Railway, 2013). As a first step, we will therefore not consider the possibility of using high-speed rail for night train services. Whenever we have a high-speed route we thus replace it in our network by a conventional train route, with an operating speed of 90 km/h. We thereby assume that, whenever there is a high-speed route, there is also a conventional train route operating on roughly the same itinerary. This generally makes sense since these conventional tracks were usually built prior to the HSR infrastructure and used to operate the routes between main cities before the development of high-speed trains. A brief analysis on the French rail network confirms that our hypothesis usually holds.

A possible exception may be the Channel Tunnel where only high-speed trains are currently operated for passenger transport. However, freight trains run through the tunnel and a project for overnight passenger trains (Nightstar) was even considered in the 1990's, before being abandoned in 1997 for lack of commercial viability. We therefore assume that operating night trains under the Channel is technically feasible as well.

Night trains usually operate in 3 phases (Interational Union of Railway, 2013):

1. A *boarding phase* during which the train stops at several stations to pick up passengers (up to around midnight).
2. The *travel phase* during which the train does not stop.
3. The *de-boarding phase* where the train stops at several stations to drop off passengers (usually starting at 6 a.m.)

As mentioned earlier, night trains therefore need a minimum travel time of about 6 hours. Travel times can not be excessively long neither given that their benefit lies for a large part in the fact that passengers can sleep through the night on the train. In this study, we will therefore consider that the journeys on night trains can be between 6 and 14 hours long. In practice several routes are however longer than that, e.g. SJ's night train from Stockholm to Riksgränsen (Northern Sweden) which lasts for more than 17 hours.

With these assumptions in mind (that we recall below), we can compute which city-to-city journeys could be operated by night trains at European level.

Main assumptions for operating night trains in the ERTN

- The operating speed is limited at 90 km/h.
- High-speed routes can be replaced by regular train routes operating at 90 km/h.
- Travel time should be between 6 and 14 hours.

Fig. 5.4 shows to which cities night train routes departing from Brussels could be operated. For some cities like Vienna, night train services already exist. However, for other cities like Berlin or Milan, for which we count more than 50 thousand passengers per month from Brussels airport, there is currently no night train service available. We note that in both cases it is possible to make the journey in two legs by combining a high-speed train and a night train. For Berlin, one could take a high-speed train to Frankfurt and from there a night train to Berlin. For Milan, the same goes with an intermediate stop in Paris where the passenger needs to change from train station. We note that schedules then need to be synchronized, especially since in both cases the two legs of the journey are not operated by the same company. Even so, changing trains is usually less convenient for passengers.

We also note that for some cities like Lyon, Bordeaux, Marseille, Manchester or Hannover, train routes during the day (mostly high-speed) already enable a travel time which is not much longer than with the plane (see Fig. 4.3b).

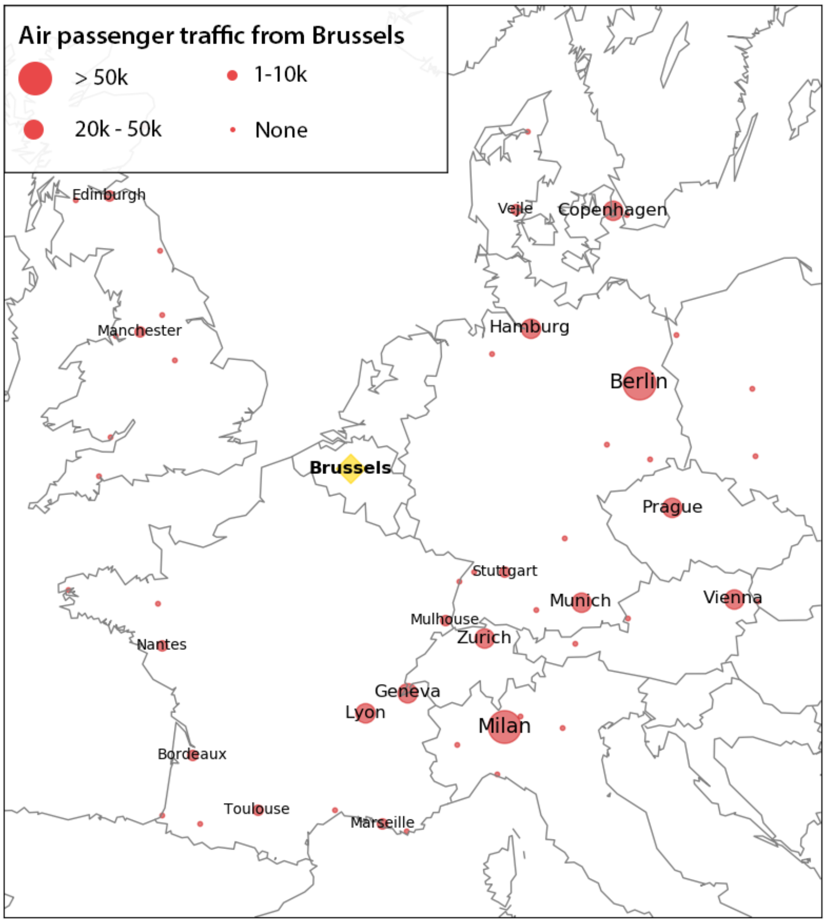


Figure 5.4: Cities for which night trains could be operated from Brussels, weighted by the monthly air passenger traffic.

At the European level, we find that 22.71% of the journeys between the city pairs considered (those connected as well to the train as to the rail network) could in theory be operated by night trains. In the radical scenario where all the passengers from the air transport sector would take night trains on those journeys, the air traffic between the cities considered would drop by 41.68%.

We now take a step further and assume that high-speed rail could be used for night train services. This possibility, known as Very Long Distance Night Train (VLDNT) service, was analysed by the Interational Union of Railway (2013). The study shows that the technical difficulties of VLDNT services could be overcome (night maintenance, freight services), but that the main obstacles are tracks access charges which are the main cost drivers for overnight high-speed train services. It is worth noting that overnight high-speed trains are already operated in China. For this second scenario, we also remove the speed limitation of 90 km/h on regular tracks.

As a consequence, the area reachable by night train is extended, especially on journeys where high-speed tracks can be used. For example, we see on Fig. 5.5 that several cities including Madrid, Barcelona, Nice and Rome are now reachable from Brussels by night train within the limit of 14 hours. Several cities do on the contrary not fit the night train service framework anymore since the travel time is shorter than 6 hours, e.g. Lyon, Bordeaux, Hannover.

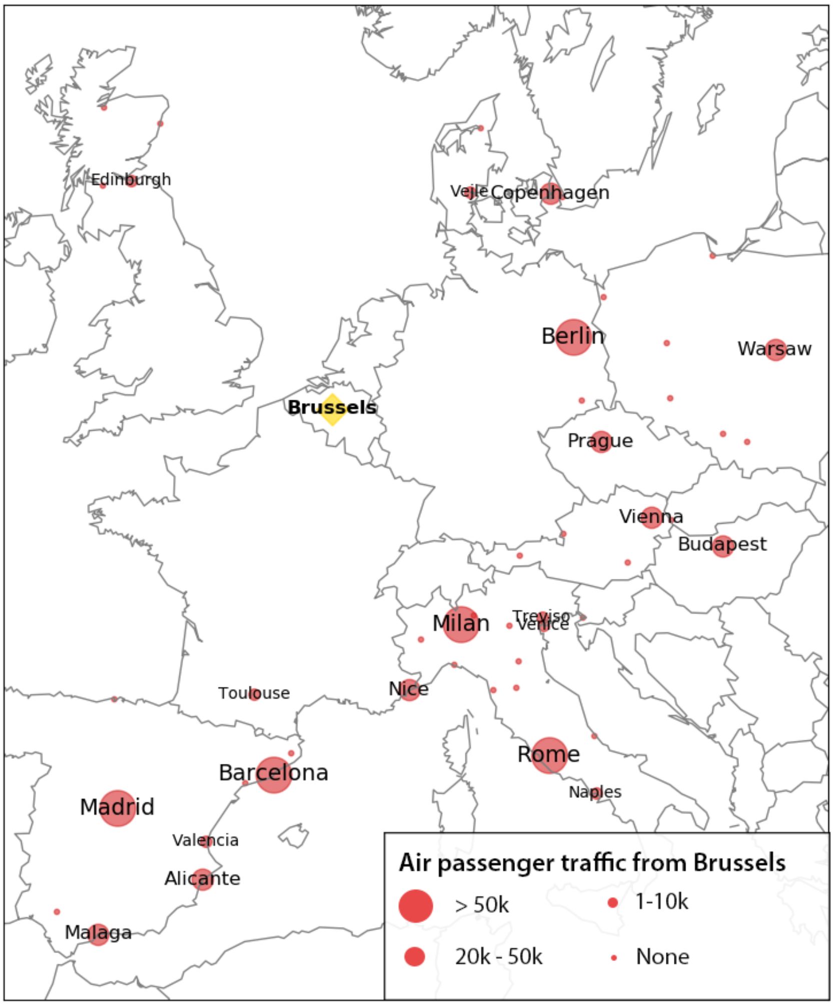


Figure 5.5: Cities for which night trains (including VLDNT) could be operated from Brussels, weighted by the monthly air passenger traffic .

At European level, 33.88% of the city-to-city journeys could then be performed using night trains. If overnight train services were used instead of the plane on all those journeys, the air passenger traffic between the cities considered would drop by 43.86%. This is only a 2% increase compared to the scenario where we do not use high-speed trains for night service. The mean reasons are probably that

1. High-speed infrastructure is mainly restricted to some countries of Western Europe, with often not more than a few lines per country.
2. Since we use high-speed trains on overnight routes, some journeys take less than 6 hours and therefore do not qualify anymore for night train services.

In both cases (with and without high-speed trains), it appears that night trains could be a viable alternative to the plane on some of the journeys where day trains are not competitive with air transport in terms of travel time. Night trains services however require adjustments on the traffic management at night, and their success lies for a part in the willingness of passengers to use night trains instead of planes when travelling over longer distances.

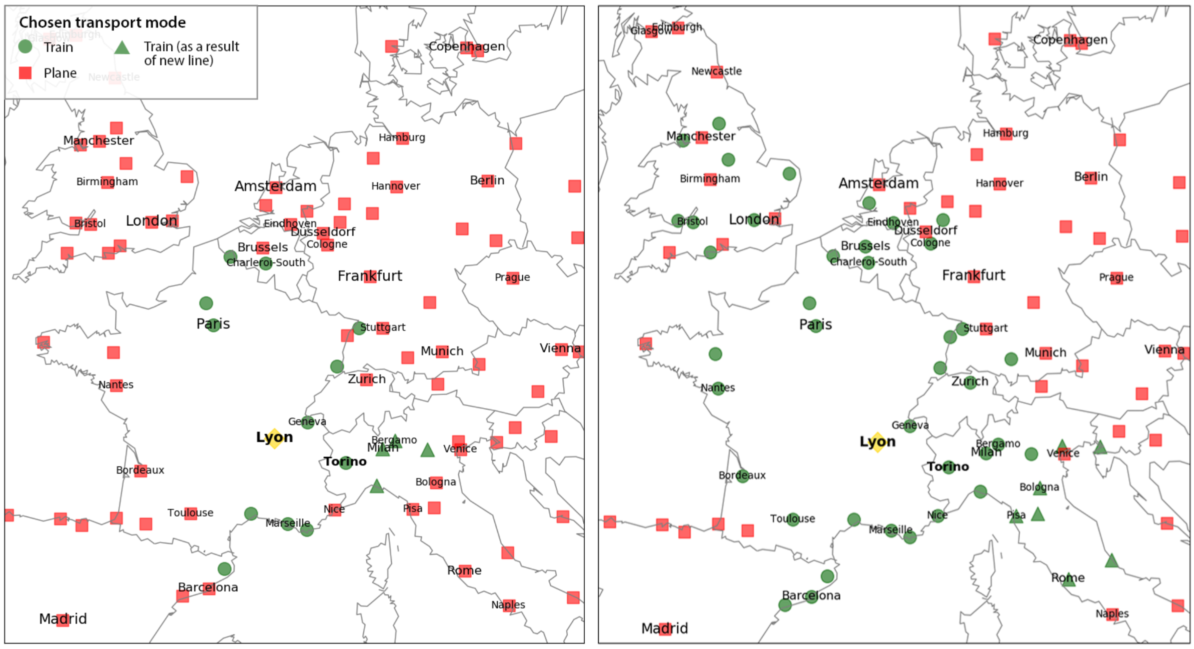
5.4.2 Lyon-Torino: a controversial high-speed line

In 2001, a project was launched to build a high-speed rail line between Lyon (France) and Torino (Italy). With a speed of up to 220 km/h, the new line will link Torino with Lyon in 1h45 (compared to 4h21 at the moment) and by that connect the French and Italian HSR networks. The new segment will connect to the already existing HSR infrastructure between Lyon and Paris on the French side, and between Torino and Milan on the Italian side, thereby putting Paris at 4 hours from Milan.

The new line is composed of 3 sections: the French section, the Italian section and a common section joining the national sections through a tunnel under the Alps. The common section, funded jointly by the French and Italian governments and the EU, is by far the most expensive as it includes a 57 km tunnel under the mountains. The construction works for the tunnel have started in 2016 on the French side but were delayed several times on the Italian side due to deep disagreements in the Italian government. The project was mainly criticized for its important construction costs and its low expected revenues. With the EU threatening to cancel the funding of the project, the Italian government finally agreed in 2019 to the construction of the tunnel, and thereby the whole line. The entry into service is scheduled for 2030.

We decided to evaluate the impact of the future line in terms of travel times and passenger flows between main cities. To do so, we simply add the new route in our rail network model and we recompute the door-to-door shortest paths between cities. As in the previous sections, we use two scenarios. In the first scenario, passengers take the train on journeys where it is faster than the plane. In the second scenario, passengers are eager to spend at bit more time travelling to take the train rather than the plane, for example because they are concerned about their carbon footprint. They will therefore take the train on any journey where the travel time is at most 2 hours longer than with the plane.

Figure 5.6 shows the chosen transport mode from Lyon when taking into consideration the new line in the ERTN. The cities for which the train is now the preferred transport mode (compared to situation without the new HSR line) are indicated with a triangle. With the new line, door-to-door journeys from Lyon to Milan, Bologna, Genoa and Verona are now shorter by train. In the second scenario (when passengers are ready to spend up to 2 hours more travelling to take the train), the cities for which the train is the preferred transport mode from Lyon include Rome, Pisa or Bologna.



(a) When Lyon' passengers choose the train if the journey is faster than by plane.

(b) When Lyon' passengers choose the train if the journey is at most 2 hours longer than with the plane.

Figure 5.6: Chosen transport mode per city from Lyon with the new Lyon-Torino line.

The possible impacts of the Lyon-Torino line at European level are summarized in Table 5.5. With the introduction of the new line, 11 door-to-door journeys between city pairs become faster by train. All these journeys link a French and an Italian city. Furthermore, if all passengers on those journeys were to take the train instead of the plane, the air traffic would decrease by an additional 28 960 passengers per month (compared to the possible decrease of passengers without the new line).

In the second scenario, the train becomes the preferred transport mode on 34 additional journeys following the commissioning of the new line. If all passengers on those journeys were to be transferred to the rail, the air traffic would decrease by an additional 444 thousand passengers per month.

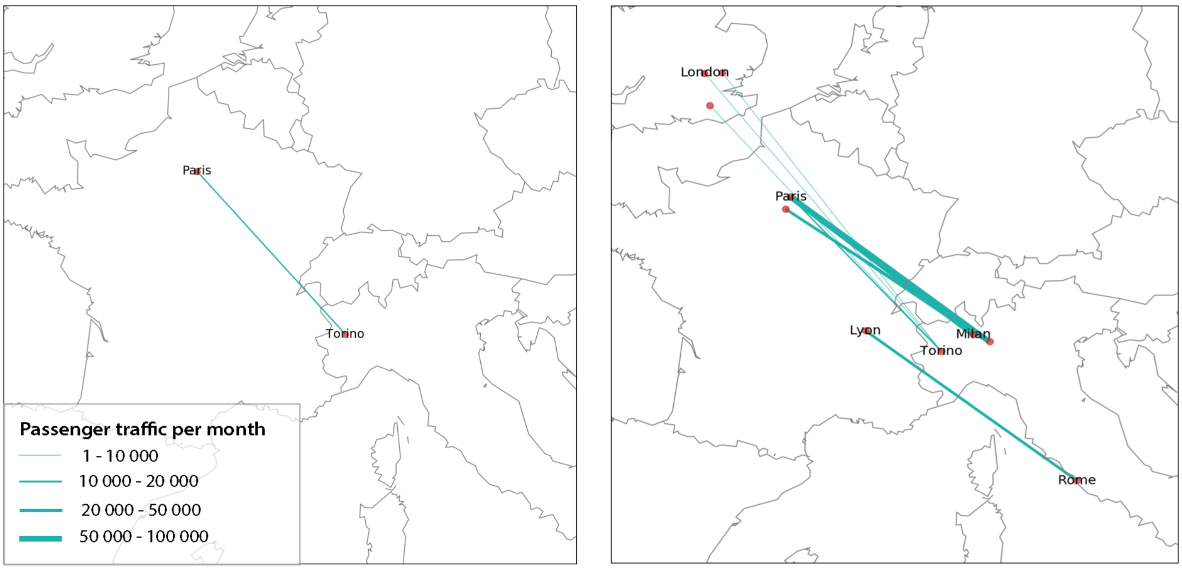
We then evaluate the potential traffic on the new line by recomputing the flows in the ERTN resulting from the transfer of passengers from air to rail, on all routes where the train is the chosen transport mode (as in Section 5.2). We find that the traffic on the Lyon-Torino line is exactly the same as the number of additional passengers that could be transferred from air to rail following the introduction of the line. In other words, no air passenger traffic was carried over to the conventional line between Lyon and Torino in Section 5.2. If it was the case, the additional flow on the conventional line would have been transferred to the new high-speed line since it offers a shorter path in terms of travel time. The main reason is probably that, at the moment, the train is usually much longer on journeys from France to Italy due to the low speed on the conventional lines (average speed $\sim$ 65 km/h between Lyon and Torino, $\sim$ 55 km/h between Nice and Genoa). A few such journeys are however faster by train (e.g. Lyon-Torino, Nice-Genoa) but no passenger flight routes are operated between those cities, meaning that other transport modes than the plane are already used on those journeys.

Scenario	Additional journeys between cities made by train	Additional air passengers decrease per month	Traffic on new line (per month)
1	11	28 960	28 960
2	34	444 105	444 105

Table 5.5: Impact of the new Lyon-Turin high-speed line on shortest paths and passenger traffic.

Depending on the scenario, the market share that could be gained on the air transport sector, and thereby travel on the new line, represents 0.3 or 5.3 million passengers per year. A study on 11 high-speed lines in Europe has shown that, to be successful, a high-speed line should ideally have nine million passengers per year (ECA, 2019). It thus makes sense that concerns were raised regarding the economic viability of the line, especially since its constructions costs are high due to the geography of the route. As shown on Fig. 5.7, in the first scenario the Lyon-Torino high-speed line could gain market share from the air transport sector on only one flight route (Paris Charles de Gaulle-Torino). In the second scenario, passengers from additional flight routes (Paris-Milan, London-Torino, Lyon-Rome) would increase the number of passengers on the new HSR line but the potential traffic is still below the sustainability treshold of 9 million passengers per year.

Moreover, when computing the potential traffic carried by the new line, we consider that all passengers operate a modal shift from air to rail on the routes where the trains is the preferred transport mode, which in practice will not be the case (except if those flight routes are forbidden as already considered at national level in some countries). However, the current traffic on the conventional train lines between France and Italy should also be taken into account, as the new line will offer a much faster alternative. Passenger traffic can also be gained from other modes of transport as the car or coaches. Finally, the line is also meant to carry freight, most of which currently passes through the Mont-Blanc tunnel or the Fréjus tunnel.



(a) Scenario 1 **(b)** Scenario 2

Figure 5.7: Flight routes from which passenger traffic could be transferred to the Lyon-Turin line.

To make from the Lyon-Torino line a success in terms of passenger transport, the analysis above nevertheless shows that the range of city-to-city journeys that could make use of the new infrastructure should be expanded. In terms of air passenger flows, Milan (2 airports, 31 million passengers/year) is much bigger hub than Torino (4 million passengers/year). What is really at stake in terms of passengers with the new Lyon-Torino junction is thus the connection of Milan to an HSR network as large as possible, so that high-speed trains can compete with some of the flight routes operated from the airports of Milan. For example, 5 million passengers take the plane every year between London and Milan. Even with the new HSR line, high-speed trains are currently not competitive with the plane on that journey due to the important number of connections needed. Going from London to Lyon requires changing trains in Lille or Paris, before taking the new route from Lyon to Torino and finally the existing line between Torino and Milan. However, if a direct high-speed route were to be operated between London and Milan (without any connection), one could take the door-to-door journey in 6h58 which is just about 2 hours more than with the plane. If the new high-speed line could recover just a share of those 5 million yearly passengers, its sustainability would certainly be improved. Therefore, the success of the new Lyon-Torino line depends in part on the implementation of fast and direct routes to some important cities, and on the willingness of passengers to spend some additional time travelling to take the train.

5.5 Summary

In this section we have studied the possibility of transferring passengers from the EATN to the ERTN, on city-to-city journeys where the train is faster but also on journeys where the train is a few hours longer. Considering only journeys where the train is faster, at least 4 million passengers could switch from air to rail and this numbers grows with the number of additional hours passengers are ready to spend travelling by train.

We have then studied the impact of such modal shift on train lines by computing the passenger flows corresponding to the transfer of passengers in the ERTN. This showed the vital importance of HSR when it comes to transferring passengers, with a big share of the passenger-kilometers travelled on high-speed trains.

The betweenness centrality was then used as a measure of the hubness of airports and train stations in each network. It turned out that major hubs are not always the same depending on the transport system, with the ERTN being again constrained by the

Euclidean distance. As a result of the potential modal shift, Brussels was for example shown to become an important train travel hub between Germany, the Netherlands and the UK, much more important than its position as air travel hub.

Finally, we used our model for transport networks between main cities to perform two short case-studies on night trains and on the new Lyon-Turin line. We showed that night trains could provide a viable alternative to air transport on a substantial number of journeys at European level, as long as passengers are willing to use night train services instead of the plane. An analysis of the traffic on the future Lyon-Turin line showed that the market share that could be gained from the air transport industry was rather modest. In order for the line to be a commercial success, new long-distance routes between large cities will probably need to be established. Once more, the potential traffic depends for a large part on the willingness of passengers to spend a couple more hours travelling to take the train.

6 Optimal networks

A significant part of this chapter is inspired by the excellent work of Michael T. Gastner on spatial networks between 2004 and 2006, with some reformulations and adjustments to the European rail network. We will first discuss the notion of optimal network together with methods to compute optimal network candidates (also called near-optimal networks). We will then present the adaptations that have been made to Gastner's original method to compute optimal candidates for the European high-speed rail network. Finally, we will review the near-optimal networks obtained and highlight similarities and discrepancies with the real-world network.

6.1 Methods

6.1.1 Optimal networks: balancing construction costs and efficient journeys

Most real-world networks grow incrementally. A sewage network grows by connecting new houses to the sewage system as the city expands; an Internet network grows as new computers connect to the network and new links are added to connect them. This is also the case for transport networks: from a limited set of vertices and edges, a transport system expands by successively adding nodes (stops) and edges (lines) to the network to connect more people or increase the travel speed.

Far from being ineffective, real-worlds networks can be quite efficient despite an usual lack of information when the original network was designed. In (Gastner and Newman, 2006a), two criteria are used to measure the efficiency of networks with one "root" or "source" node (in a sewage system, the root node would for instance be the city's sewage treatment plant; in a transport system the root is the central or main station).

1. The sum of edge lengths

$$T = \sum_{i<j} A_{ij}\ell_{ij} \tag{6.1}$$

where ℓ_{ij} is the length of the edge between i and j and A_{ij} is the adjacency matrix of the network ($A_{ij} = 1$ if there is an edge between vertices i and j, 0 if not).

The construction and maintenance costs of a network can be considered roughly proportional to the total sum of the edge lengths which should thus be kept as low as possible.

2. The route factor

$$q = \frac{1}{n} \sum_{i=1}^{n} \frac{\ell_{i0}}{s_{i0}} \tag{6.2}$$

where ℓ_{i0} is the distance from node i to the root node along the edges of the network while s_{i0} is the direct Euclidean distance between node i and the root (in other words, the distance as the crow flies) and n is the number of nodes in the graph. The lower the route factor, the more efficient the network is at in transporting flows (passengers, wastewater, etc.).

There are two graphs which respectively minimize each of these criteria.

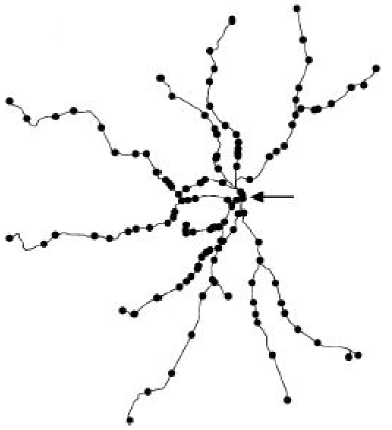
1. The minimum spanning tree (MST) is the graph connecting all nodes minimizing the sum of edge lengths.

2. The star graph has an edge from the root node to each node and thus a root factor of 1.

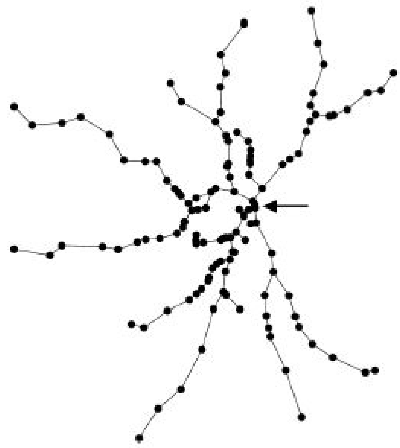
When it comes to designing a real-world network, none of these solutions are however viable. Coming back to our transport system example, an MST network would mean that there is only one train line that goes from the main station to all other stations and that trains stop at each station. On the other hand, a star graph corresponds to a scenario where a direct train line links every station to the central station without any intermediary stop. In more formal terms, the tree shape of the MST makes it inefficient to travel in the network and connecting all nodes directly to the root as in the star graph is often not practically doable.

An efficient transport network, as defined in (Gastner, 2005), is thus a reasonable compromise between construction costs and travel costs. As shown in (Gastner and Newman, 2006a), real-world networks tend to be remarkably good at balancing these criteria. The Boston commuter rail network is for instance shown to have a root factor of 1.14 and a total edge length of 1.12 times that of the MST graph.

One may however wonder what the optimal graph would be if the geographical position of all nodes and the flows between them were known in advance. This is referred to as the optimal design problem for spatial distribution networks. Even if full networks are seldomly built from scratch, knowing the optimal or near-optimal network configuration can be used to assess the efficiency of the real-world network. For instance, it turns out that the actual Boston commuter rail network is highly similar to the optimal network which explains its efficiency in terms of route factor and total edge length (see Fig. 6.1).



(a) Actual network



(b) Optimal network as computed by (Gastner and Newman, 2006a)

Figure 6.1: Actual and optimal networks for Boston commuter rail (Gastner and Newman, 2006a). The arrow indicates the central station which is the root node.

In this chapter, we will therefore look at the optimal design problem for the European rail transport network. First of all, we need to formalize the concept of "best compromise" between construction costs and short travel distances. The construction cost T is defined as in Eq. 6.1 and, in a connected and undirected graph with several sources or roots, the travel cost is defined as

$$Z = \sum_{i<j} w_{ij} d_{ij} \tag{6.3}$$

where d_{ij} is the shortest path length between nodes i and j in the network, and w_{ij} is a weight depending on the amount of traffic between the nodes. In some sense, Z is simply a generalization of the route factor q (Eq. 6.2) for a network where every node i can be a source, and the importance of the shortest path length between i and j is then determined by the weight w_{ij}.

The total distance is however not the only factor that affects convenience of travelling. In an air transport network, the time needed to go through security checks and board into the airplane, as well as the time needed to catch a connecting flight has a significant impact on the travel time. To put it in terms of graph theory, the number of hops or legs in a path is also an important factor beyond the travel distance. This can be taken into account by redefining the effective edge length when computing shortest paths as

$$\tilde{\ell}_{ij} = (1-\delta)\ell_{ij} + \delta \tag{6.4}$$

where δ is a parameter determining the user's preference for measuring length in terms of distance or legs. The transport cost Z is then redefined as

$$Z = \sum_{i<j} w_{ij}\tilde{d}_{ij} \tag{6.5}$$

where $\tilde{d}_{ij}$ is the shortest path length between nodes i and j when using edge lengths $\tilde{\ell}_{u,v}\ \forall\ (u,v)\ \in E$.

The network cost is then given by

$$C = T + \gamma Z \tag{6.6}$$

where $\gamma \geq 0$ is a parameter that sets the relative importance of travel costs. When $\gamma = 0$, only construction costs matter and the network minimizing the cost is the MST. As γ increases more importance is given to travel costs and when $Z \to \infty$ the optimization tends towards a fully connected graph or star graph.

For a given γ, the optimal network with n nodes is the one minimizing C over all possible networks of n nodes. We will now see how to generate such optimal or near-optimal graphs.

6.1.2 Computing near-optimal graphs

A way of finding the optimal spatial network connecting n vertices is to generate all possible networks and choose the one with the lowest cost. An undirected graph with n nodes has at most $\frac{n(n-1)}{2}$ edges, we would thus have to generate $2^{\frac{n(n-1)}{2}}$ graphs and compute their costs which includes computing shortest paths for each of them. This in practice not feasible for complex graphs like transport systems.

(Billheimer and Gray, 1973) proposes a generic and heuristic algorithm to find optimal network candidates (also called near-optimal networks) minimizing some cost in a network. The algorithm works in a greedy and degenerative way, with 3 main steps:

1. Start with all the $\frac{n(n-1)}{2}$ edges (full graph).
2. Edge elimination phase : find the existing edge leading to the greatest cost reduction, remove it and go back to step 2 . If no such edge exists, go to step 3.
3. Edge insertion phase: among all non-existing edges, add the one decreasing the cost the most and go back to step 2. If no such edge exists, return the current graph.

Each time the graph is perturbed (i.e. an edge is deleted or added), the cost of the new graph has to be computed. When $\gamma > 0$, this implies computing the shortest path between each pair of nodes which has complexity $O(mn \log n)$ for a graph with m edges. However adding or removing an edge does sometimes not change the shortest paths in the graph. (Gastner, 2005) uses the dynamic all-shortest paths algorithm developed in (Ramalingam and Reps, 1996) that allows to recompute only the paths impacted by the graph perturbation. The cost of newly generated graphs is then computed much faster than recomputing all paths at each iteration.

The implementation of such an algorithm is outside the scope of this thesis but an easy speed-up of the algorithm can be obtained by keeping track of the shortest paths in the graph. When an edge is removed, we can recompute only the shortest paths which were using the deleted edge. When an edge is inserted all the shortest paths need to be recomputed. However, the structure of the algorithm (using the full graph as initial network and starting by removing edges which reduce the cost) makes it so that much more edge deletions than insertions are performed.

The main drawback of the greedy algorithm is that it may get stuck in a local minimum. (Gastner, 2005) proposes a simulated annealing algorithm which in theory does not suffer that problem. The idea is to generate new graphs by adding or deleting edges randomly and then accept new candidates according to a Metropolis-Hastings rule. If the cost of the current graph is C_{cur}, the probability of accepting a new candidate with cost C_{new} is

$$p_{cur,new} = \begin{cases} \exp -[\beta(C_{new} - C_{cur})] & \text{if } C_{new} > C_{cur}, \\ 1 & \text{otherwise} \end{cases} \tag{6.7}$$

with β a parameter that increases after each iteration. The algorithm is ergodic which means that we can in theory reach the optimal network from any initial network in a finite number of iterations. This requires however to increase β very slowly which in practice is too time-consuming. We will therefore use a faster increase of β, with the potential risk of still being trapped in a local minimum. As we will see, the simulated annealing algorithm however usually generates networks with a slightly lower cost than the greedy algorithm.

In (Gastner, 2005), a good starting value for β is found to be $\frac{0.1}{C_{MST}}$ where C_{MST} is the cost of the minimum spanning tree. The value of β is multiplied by $1 + 3 \cdot 10^{-6}$ at each iteration to slowly lower the probability of a cost increase. If the current graph has not changed for 10^6 iterations, we consider it as sufficiently optimized.

6.1.3 Application to an optimal high-speed rail network

To apply this to an optimal European high-speed rail network, we choose as nodes the main cities in Europe. As a first step, we take into account the 17 cities connected to the HSR in Western Europe as presented on Fig. 4.5. For simplicity, we assume that each city has only one main train station. In a second time, we take into account all cities from a subset of European countries with a urban population above 1 million. The populations are extracted from Eurostat city statistics database from 2016.

For each city, the population is normalized by the highest population among the cities considered. If city i has a normalized population p_i and city j has a normalized population p_j, the traffic weight in the travel cost (Eq. 6.5) is given by $w_{ij} = p_i \cdot p_j$. A similar choice of weights is used in (Gastner and Newman, 2006b) to compute the optimal air transport network in the US.

In the previous chapters we systematically measured distance between cities in terms of travel time which is more relevant than the travel distance when looking at door-to-door journeys. The edge length ℓ_{ij} is thus measured in hours for our implementation of the optimal design problem. In the work of Michael T. Gastner, distances between cities in the network are measured in km or in hops.

Moreover, working with durations instead of distances allows to use different types of rail tracks with different speeds. For this study we will consider 2 types of tracks:

1. High-speed tracks (HST) allowing an average speed of 200 km/h.
2. Very high-speed tracks (VHST) allowing an average speed of 300 km/h.

The price per kilometer is considered 10% cheaper for HST than for VHST (Armitt and Houghton, 2018). In reality, many standards exist and there is no single definition of high-speed rail or very-high-speed rail. Considering two types of tracks is however a good start for our optimal design problem.

Allowing different types of edges involves adapting Billheimer's greedy algorithm to allow speed upgrade and downgrade and not only line insertion or elimination. Between two nodes u and v there can thus be:

1. No edge.
2. A high-speed line with average speed of 200 km/h.
3. A very high-speed line with average speed of 300 km/h.

As previously, we start with a fully connected graph using only VHST. During the edge elimination phase, we consider removing edges but also downgrading edges. The removal/downgrade resulting in the greatest cost reduction is performed. Similarly, in the edge insertion phase we consider adding both a VHST and an HST between each pair of non-connected nodes, or upgrading an existing line. The insertion/upgrade leading to the greatest cost reduction is implemented.

From a graph point of view, upgrading or downgrading a track amounts to changing the weight of an edge, with the weight being the travel time in the ERTN framework. Removing an edge can be seen as setting its weight to $+\infty$, and inserting one is setting the weight to the travel time which depends on the type of tracks installed. A unitary perturbation of a graph is thus changing the weight of one of its edges[10]. A complete dynamic shortest paths algorithm allows to recompute only the shortest paths which are

[10]Besides edge deletion and insertion, edge rewiring is also used in (Gastner, 2005). An edge (i, j) is then rewired to (i, k) if k is not already a neighbour of i. With such operations 2 edge weights are modified.

impacted by the unitary perturbation. Without such an algorithm, simulated annealing turns out to be computationally too heavy when working with several types of tracks. As mentioned before, implementing such an algorithm is however out of the scope of this thesis. When working with VHST and HST, only the greedy algorithm will thus be used. As we will see, the near-optimal networks generated by the greedy algorithm are often almost as good as with the simulated annealing algorithm.

In the previous chapters on the ERTN, a fixed connecting time of 15 min was allowed to change trains. Although very useful to compare network structures as in (Gastner and Newman, 2006b), redefining the edge length with a parameter δ as in Eq. 6.4 does not work to add a minimum connecting time (MCT). Instead, we simply redefine the edge length as

$$\tilde{\ell}_{ij} = \ell_{ij} + MCT. \tag{6.8}$$

The minimum connecting time can be seen as the time needed to "step into an edge", in other words to step into the train.

Finally, some sections of line can be more expensive than others. Subsections consisting mainly of tunnels and viaducts cost on average 2.5 times more than other sections (Armitt and Houghton, 2018). This is typically the case for a section of line that would have to cross the Channel (e.g. between Paris and London) or Zeeland[11] (e.g. between Lille and Rotterdam). The construction cost of a line between cities i and j is redefined accordingly as

$$c_{ij} = \ell_{ij} + 1.5t_{ij} \tag{6.9}$$

where t_{ij} is the total track length that has to be built with tunnels or viaducts. The construction cost is then given by

$$T = \sum_{i<j,(i,j)\in VHS} A_{ij}c_{ij} + 0.9 \sum_{i<j,(i,j)\in HS} A_{ij}c_{ij} \tag{6.10}$$

where VHS and HS are the sets of very high-speed and high-speed tracks respectively.

[11] Province of the Netherlands consisting mainly of islands and peninsulas.

6.1.4 Comparing networks

We compare the obtained networks in terms of construction cost T and transport cost Z which are a by-product of the optimization algorithms. One can also compute the total track length in km as the sum of edge lengths (Eq. 6.1). This is typically a bit lower than the construction cost T since some lines have sections with tunnels or viaducts. We also compute the average path length (APL)

$$L = \frac{2}{n(n-1)} \sum_{i<j} \tilde{d}_{ij} \tag{6.11}$$

where $\tilde{d}_{ij}$ is the shortest path length between nodes i and j (taking into account connecting times). The APL gives an indication of how efficient it is to travel in the network without weighting by the populations of cities i and j (unlike Z).

We would also like to compare how similar an optimal network G_{opt} is to the real-world network G_{real}. This will be computed using the Jaccard index

$$J = \frac{a}{a+b+c} \tag{6.12}$$

where

- a is the number of edges that are in G_{opt} and in G_{real},
- b is the number of edges that are in G_{opt} but not in G_{real},
- c is the number of edges that are in G_{real} but not in G_{opt}.

Moreover, we find that weighting the edges in a, b and c by their length yields a more relevant index. We call the new index $\tilde{J}$.

6.2 Results

In the high-speed rail network of Section 4.1 (Fig. 4.5), we used the space-of-stations approach where 2 nodes are linked if there is a train stopping in both cities. For instance, we had an edge between Lille and Strasbourg which corresponds to a regular TGV running between the two cities. In reality, there are however no direct tracks between Strasbourg and Lille and the train passes through Paris. In this chapter, we only keep links corresponding to real tracks in order to make valid comparisons with optimal networks who only build the actual lines. In other words, two nodes are linked if they are physically connected, this is referred to as the space-of-stations (Kurant and Thiran, 2006). Furthermore, the cities of Avignon and Mannheim which were only considered as intersections in Section 4.1 were also added to the network to ensure valid comparisons with the real-world network (Jaccard index computation). The corresponding real-world network is shown on Fig. 6.2a.

Three optimal networks corresponding to $\gamma = 100, 200, 300$ are plotted on Fig. 6.2b, 6.2c and 6.2d respectively. In a first time, we consider only one type of tracks enabling an average speed of 300 km/h. For $\gamma = 100$, the network still has some kind of tree structure which indicates that the total cost is still dominated by the construction cost. For $\gamma = 200$ and $\gamma = 300$, more links are created as the weight given to transport costs increases. Results are gathered in Table 6.1 for these 4 networks as well as for the minimum spanning tree ($\gamma = 0$), the star graph ($\gamma = +\infty$) and an additional scenario where even more weight is given to transport costs ($\gamma = 500$).

Network	APL (h)	Tracks length (km)	T	Z	$\tilde{J}$
Real	3.08	3340	3452	37	1
MST ($\gamma = 0$)	4.34	2790	2904	49	0.40
Near-optimal, $\gamma = 100$	3.46	2816	2929	40	0.66
Near-optimal, $\gamma = 200$	3.13	3375	3488	36	0.63
Near-optimal, $\gamma = 300$	3.04	3738	4010	34	0.75
Near-optimal, $\gamma = 500$	2.94	4263	4597	32	0.56
Star graph ($\gamma = \infty$)	1.98	79215	81259	25	0.04

Table 6.1: Comparison between real and near-optimal networks (1).

As expected, the average path length (APL) and the transport costs Z decrease as γ increases while the tracks length and the construction costs T increase. In terms of graph structure, the optimal graph with $\gamma = 300$ seems to be the most similar to the real-world network with a Jaccard index $\tilde{J} = 0.75$. The main difference is a direct link between Paris and London, i.e. a second tunnel under the English Channel despite high construction costs (105 km of tunnel on the as-the-crow-flies route). This is explained by the high traffic weight (w_{ij} in Eq. 6.5) given to the path between the two cities since together they account for 49% of the population of all considered cities. It is interesting to note that in Section 5.2, the line running under the Channel was already the first or second most affected line (depending on the scenario) by a modal shift from air to rail (by most affected we mean having to carry the biggest share of additional passengers). In some sense, this "second" tunnel stresses the importance of the link between the UK and continental Europe. The near-optimal network obtained is however based on a very rough estimation of the cost of tunnels which in practice varies depending on the project.

A second difference is that to the East, Paris is connected to Mannheim instead of Strasbourg. This can probably be explained by the fact that the former has a larger population and a more central position among the cities of the Eastern part of the network. Mannheim then acts as a hub for the cities of Strasbourg, Stuttgart and Frankfurt.

In terms of costs, the closest scenario to the real-world network is the one with $\gamma = 200$. It is interesting to note that with roughly the same construction costs, different choices on the lines to build have been made. We first note that in both networks, Paris acts as an important hub for all passengers from or to cities of Southern or Western France. In the near-optimal network, the Paris-Strasbourg line known as "LGV Est" has not been implemented. Located at the end of the network, Strasbourg and the cities of Western Germany (Frankfurt, Mannheim, Stuttgart) are therefore not directly connected to Paris or other cities in France. Instead they constitute a kind of small world centered in Mannheim, with from there a connection to Brussels. On the other hand, the choice was made to link Paris directly with Brussels, allowing from there a faster transfer of passengers to the cities of the North and North-East part of the network. Even if the detour via Lille as in the real-world does not seem that significant, it often induces a change of train or a stop at the station. In this regard, let us recall that in our model a fixed connection time of 15 min is counted when passing through a city. Together, the cities located North and North-East of Paris (Amsterdam, Rotterdam, Brussels, Düsseldorf, Cologne) account for 21% of the total population while the cities located East of Paris (Strasbourg, Stuttgart, Mannheim, Frankfurt) represent 14% of the population. It is thus probable that the choice of a better connection to Brussels instead of Strasbourg has been made to decrease the travel time of the greatest number of passengers. Looking back at the near-optimal network for $\gamma = 300$, it is noteworthy that here too Paris is connected straight to Brussels and not to Lille.

Finally, we also note that to the East Brussels is linked to Düsseldorf instead of Cologne in the real world. If Cologne has a larger population, Düsseldorf is closer to Brussels in terms of Euclidean distance. For this particular choice it therefore appears that the construction costs prevailed over the travel costs.

The networks on Fig. 6.2 are obtained with the simulated annealing algorithm. On those 3 scenarios, simulated annealing yields networks with a total cost between 0.3% and 1.5% lower than the greedy algorithm.

In a second time, we consider a different set of nodes for our optimal network design problem. We will take as nodes all cities in France, Belgium, Germany, the Netherlands, Switzerland and Northern Italy with more than 1 million inhabitants (urban area), regardless whether they have an airport or not. We also add the city of London in view of the importance of the connection between the UK and mainland Europe. The purpose is now to design connected near-optimal networks using two types of tracks: very high-speed tracks and high-speed tracks, enabling an average speed of 300 km/h and 200 km/h respectively. The latter have 10% cheaper construction costs per kilometer. As discussed above, only the greedy algorithm is used in this case. The simulated algorithm turns out to be computationally too heavy without a dynamic all-shortest algorithm.

We generate near-optimal networks for γ =10, 100, 200 and 300. Results in terms of average path length, total tracks lengths, construction and travel costs are gathered in Table 6.2. The near-optimal networks are displayed on Fig. 6.3. As γ increases, travel costs grow in importance and consequently they decrease together with the average path length. In the meantime, construction costs and the total track length increase as edges are added/upgraded. Contrarily to the scenario above, construction costs can be lower than the total tracks length due to cheaper cost of high-speed lines limited at 200 km/h.

Network	APL (h)	Tracks length (km)	T	Z
Near-optimal, $\gamma = 10$	4.64	4108	3977	163
Near-optimal, $\gamma = 100$	3.47	5611	5687	119
Near-optimal, $\gamma = 200$	3.28	6406	6641	112
Near-optimal, $\gamma = 300$	3.00	8174	8471	103

Table 6.2: Comparison between real and near-optimal networks (2).

For $\gamma = 10$, the edges built are those corresponding to the minimum spanning tree. The trunk of the network connects the largest cities (London, Paris, Berlin, Milan) and is built using very high-speed tracks. The leafs are built with high-speed tracks and connect mostly smaller cities to the main branches. On the contrary, the near-optimal network with $\gamma = 100$ has only very high-speed tracks. The network consists essentially of two main rings connecting most of the nodes. One ring links the main cities of Eastern France, Belgium, Western Germany, Switzerland and Northern Italy, the other one consists essentially of most of the cities of Germany. The absence of long shortcuts crossing the rings shows that the optimization is still partially dominated by transport costs. Such shortcuts start to appear for $\gamma = 200$ (Lyon to Zurich, Paris to Mannheim, Amsterdam to Düsseldorf). As above, we see that a direct link under the Channel is created between London and Paris which are still the two largest cities of the network. The near-optimal network for $\gamma = 300$ multiplies long edges. Hubs appear with many direct connections extending radially from the corresponding nodes (Frankfurt, Paris, Lyon, Zurich, Hannover, Brussels).

We note that the Lyon-Turin line studied in Section 5.4.2 is part of the near-optimal solution in each of the scenarios studied. If its economic viability is often called into question, it nevertheless seems to be the prevailing option to link the French and Italian high-speed networks. Once again, we use a highly simplified model and a more realistic analysis should imply a more refined evaluation of the costs depending on the choice of the route and the infrastructure.

We also find that the cheaper high-speed tracks are only used when γ is very low (i.e., when travel costs have a relatively low importance). As travel costs become more significant, existing lines are first upgraded before new lines are added. Therefore, reducing the speed by one third does not turn to be worth it if construction costs decrease only by 10%.

In this case, a rigorous comparison with the real-world network can hardly be achieved due to its heterogeneity. Many lines consist of several sections with different infrastructures, and standards vary from one country to another. The European Court of Auditors goes so far as to say that the European high-speed rail network is not a reality but an ineffective patchwork.

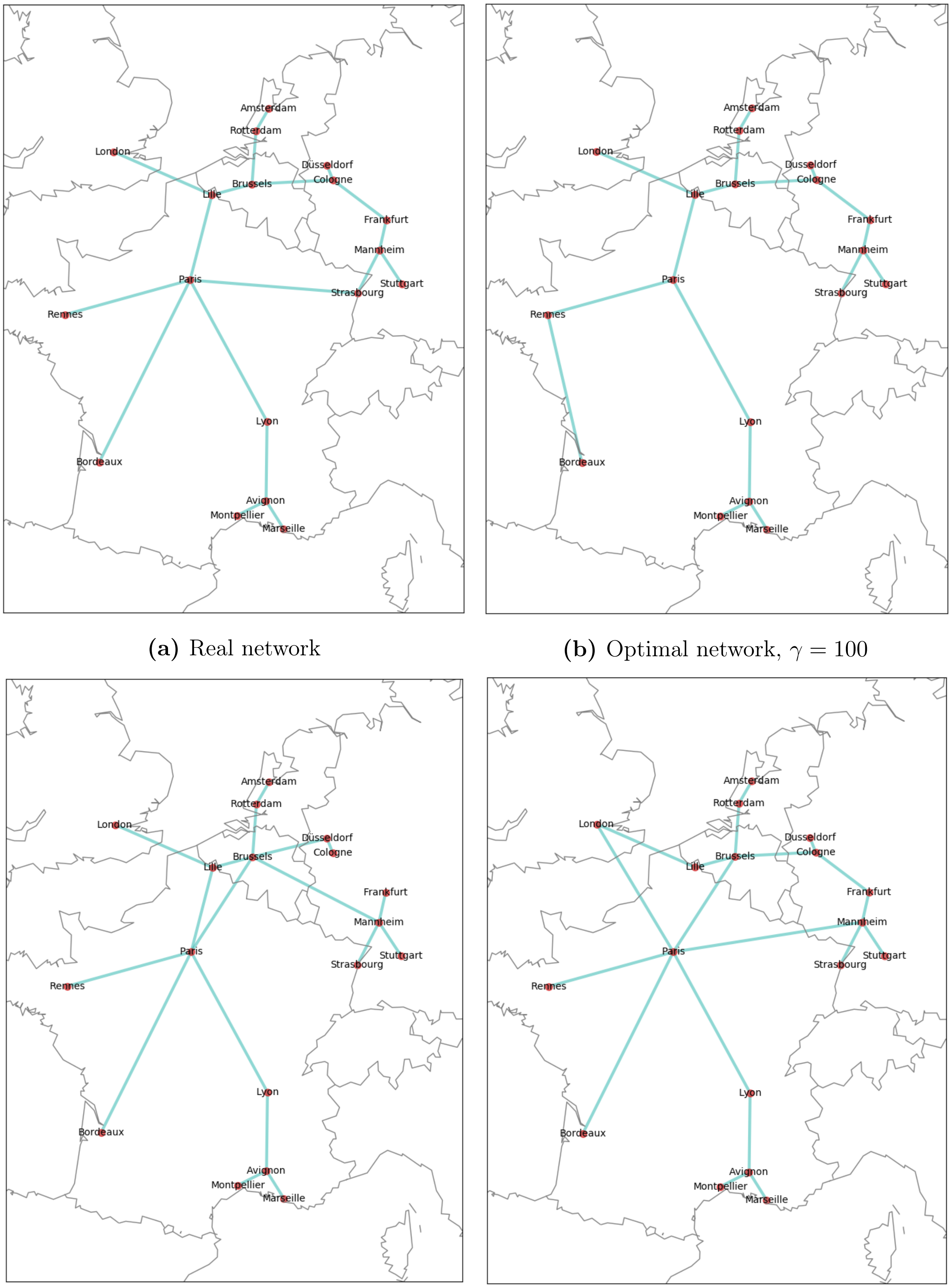
The near-optimal networks computed below are based on highly simplified assumptions, one of those is that they are not sensitive to national borders. Very often, choices regarding high-speed rail have been made at national level, leading to a multitude of infrastructure standards and sometimes to poor connections between the networks of neighbouring countries. Joint projects are not always easy to implement as they can lead to tensions and disagreements between countries, as has been demonstrated by the Lyon-Turin line. In our mathematical model, such barriers do not exist and international lines are therefore sometimes given preference over national lines. An example of such alternative choice is the Paris-Mannheim line on Fig. 6.2d, instead of the actual Paris-Strasbourg line. As pointed out by the European Court of Auditors (ECA, 2019), such choices on high-speed infrastructure should probably take place more often at European (or multi-national) level than at national level.

6.3 Summary

In this section we have seen how we can define optimal transport networks as a good balance between construction costs T and travel costs Z. Construction costs are roughly proportional to the sum of edge lengths while travel costs can be computed as the sum of shortest path lengths weighted by the amount of traffic. A parameter γ sets the relative importance of travel costs.

Near optimal-optimal networks minimizing the total cost can be obtained by a greedy degenerative algorithm or by a simulated annealing algorithm using a Metropolis-Hastings rule.

Gastner's methods were adapted to compute near-optimal candidates for the European high-speed rail network. Adaptations included weighting edges by the travel time, adding a minimum connecting time and allowing different types of tracks. Even if they are based on highly simplified models, near-optimal networks provide an interesting comparison with the real-world network. For $\gamma = 300$, the near-optimal network was for instance shown to have a similarity of 75% with the actual high-speed network in Western Europe. Design choices in the computed networks also brought a new perspective on some results from previous chapters, such as the importance of the link between mainland Europe and the UK or the strategic choice of a new line between Lyon and Torino to connect the French and Italian HSR networks. Some discrepancies between the near-optimal and the real-world networks were also highlighted, such as the implementation of a line between Paris and Mannheim (instead of the actual Paris-Strasbourg line), probably due to the absence of national barriers in our network model.



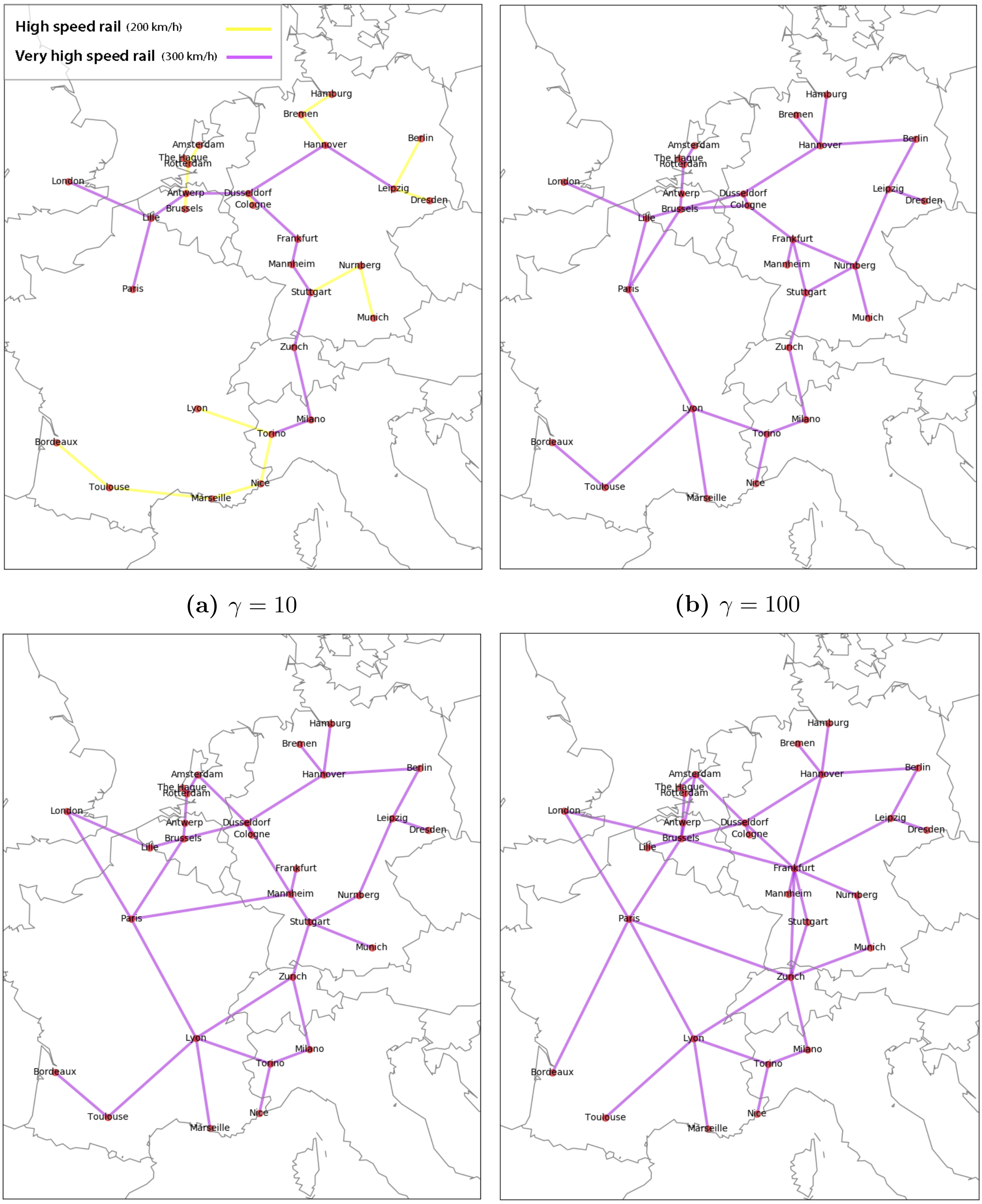
(a) Real network

(b) Optimal network, $\gamma = 100$

(c) Optimal network, $\gamma = 200$

(d) Optimal network, $\gamma = 300$

Figure 6.2: High-speed Rail in Western Europe: real and optimal networks.



(a) $\gamma = 10$ **(b)** $\gamma = 100$

(c) $\gamma = 200$ **(d)** $\gamma = 300$

Figure 6.3: Near-optimal high-speed rail networks between main cities.

7 Conclusions

Transport networks have often been studied as independent systems without taking into account possible inter-dependencies with other transport modes. It is only recently that multi-modal approaches have been used to study transport systems, mostly at the scale of a single city or a region. Such multi-modal systems can be modelled using multiplex networks where each of the transport modes considered (train, bus, metro, pedestrian, etc.) represents a layer of the network. The same nodes are then connected by different types of links depending on the layer. Applications to cities like Paris, London or New York have shown that multi-modal transport systems offer a better coverage of the city and are more robust in case of link failures (Baggag, 2018).

At international level, diverse transport systems also offer different possibilities in terms of routes, travel times, comfort or environmental impact. Those transport systems influence each others and breakthroughs in a transport sector has impacts on other sectors. The development of low-cost airlines has for example severely diminished the market share of night trains. At roughly the same time, the development of high-speed trains has put in competition the air and rail transport systems over longer distances, with HSR now holding 50% of the market on the Madrid-Barcelona route for instance. Beyond technical breakthroughs, political decisions and societal phenomena have been shown to have an impact on the way we travel. Environmental concerns call into question the continuance of flight routes when a credible alternative exists by train. Growing awareness on the emissions generated by the transportation sector is reinvigorating slower but less polluting transport modes like night trains. Passenger traffic should therefore not be considered as fixed and belonging to a specific transport system, but rather as variable flows likely to switch from one transport mode to another depending on the context.

Our case studies on the European Rail Transport Network (ERTN) and the European Air Transport Network (EATN) where we transfer passengers from air to rail when the door-to-door journey is faster by train are, in this sense, just a start. Further analysis should include other transport modes including personal or shared cars and intercity bus services like Flixbus which are booming in some European countries. International transport networks could be modelled as multi-layer networks, allowing for competition but also cooperation between transport modes. Although rare, such cooperation initiatives exist with Air France for instance proposing combined air and rail tickets in partnership with the French railway operator.

The implementation and modelling of such international multiplex transport networks (also called multi-layer networks) however face several challenges. First of all, distinct transport networks are rarely connected in an efficient way. For instance, only a few airports are connected to high-speed rail infrastructure. Secondly, journey planners usually function in distinct silos for each transport mode, making it hard for passengers to plan or book multi-modal journeys. For researchers, data on infrastructure, travel routes and passenger traffic are often dispersed among national operators, making it hard to model multi-modal transport networks at a large scale. The implementation of a database at European level bringing together travel routes and traffic data across transport modes would therefore constitute a huge breakthrough. From a mathematical point of view, the computation of itineraries and passenger flows in multiplex networks taking into account possible delays, travel costs and emissions triggers new interesting problems.

The planning of new infrastructure and services should also take place more often at European level. The analysis in this thesis and in other studies have shown that high-speed networks were too often organized at national level, leading to a an inefficient patchwork of systems (ECA, 2019). Several initiatives point in this direction with, for instance, the development of an European Rail Traffic Management System to harmonize the signaling across European rail infrastructures. The implementation of international train services like night trains also requires coordination between national railway systems. To be economically viable, we have seen that new high-speed rail lines like the Lyon-Turin junction will also need to carry high-speed trains covering several countries. On journeys where connections are needed, the synchronization of schedules among rail operators is crucial to ensure efficient journeys by train.

Rail is however not a viable alternative on all journeys. Even on journeys that can be made fully on high-speed trains, we have shown that the train tends to be slower than the plane above 704 km. Flying will therefore remain an indispensable transport mode when travelling over long distances in Europe. The small-world structure of the EATN enables short paths between main cities, with the average number of flights required to travel between two cities being as low as 1.81. In comparison, the ERTN was shown to have much longer paths and travelling between two cities requires taking 5.77 trains on average.

For some door-to-door journeys, the travel time is however shorter by train. In particular, we evaluated that for the month of June 2017, at least 4 million of passengers took a direct flight between two cities where the train is in theory faster. Policies to encourage taking the train rather than the plane on such journeys should again be implemented at European level, as domestic flights only represent 17% of the commercial air traffic in Europe.

An aspect that has not been taken into account at all in this thesis is the price of the journeys. It is very common that a journey is cheaper by plane than by train. For example, flying turns out to be cheaper than taking the train for 60% of the journeys in the UK (Times, 2019). Among other reason, it appears that one of the factors that make flying so cheap is that the European aviation sector is chronically undertaxed compared to other aviation markets (Transport and Environment, 2019). Fuel for commercial aviation is exempt from taxation and only a few member states impose a tax on tickets. In Belgium for example, taxes on international flight tickets are non-existent.

The idea of a tax on plane tickets reflecting the environmental costs of the aviation sector is probably worth studying. Again, such legislation makes more sense at European level since, if a country decides to apply taxes on air transport unilaterally, airline companies are likely to relocate part of their traffic to neighboring countries. A leaked report by the European Commission and described in (Transport and Environment, 2019) shows that introducing VAT and kerosene taxes on flight tickets could cut emissions by 11%, without any impact on the EU economy and employment. The revenue raised could for example be invested in less polluting transport modes.

References

D. Albalate and G. Bel. *The economics and politics of high-speed rail: Lessons from experiences abroad.* Lexington Books, 2012.

R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, January 2002. doi: 10.1103/RevModPhys.74.47.

L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences*, 97(21):11149–11152, 2000. ISSN 0027-8424. doi: 10.1073/pnas.200327197.

S. J. Armitt and J. Houghton. High speed rail international benchmarking study. Technical Report HS2 Phase Two, PwC, November 2018. URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/755650/high-speed-rail-international-benchmarking-study.PDF.

A. S. Z. T. e. a. Baggag, A. Resilience analytics: coverage and robustness in multi-modal transportation networks. *EPJ Data Sci.*, 7, 2018. doi: https://doi.org/10.1140/epjds/s13688-018-0139-7.

A.-L. Barabasi. *Network Science.* Cambridge University Press, 2016.

M. Barthelemy and A. Flammini. Optimal traffic networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(07):L07002–L07002, July 2006. doi: 10.1088/1742-5468/2006/07/l07002.

BBC. Ryanair one of europe's top polluters, eu data suggests. April 2019. URL https://www.bbc.com/news/business-47783992.

J. W. Billheimer and P. Gray. Network design with fixed and variable cost elements. *Transportation Science*, 7(1):49–74, February 1973. URL http://www.jstor.org/stable/25767685.

K. Z. Brandon Graver and D. Rutherfor. Co2 emissions from commercial aviation, 2018. September 2019. URL https://theicct.org/sites/default/files/publications/ICCT_CO2-commercl-aviation-2018_20190918.pdf.

D. Briginshaw. Air france ordered to curb competition with rail in france. *International Railway Journal*, May 2020. URL https://www.railjournal.com/passenger/high-speed/air-france-ordered-to-curb-competiton-with-rail-in-france/.

S. Dasgupta, A. Chatterjee, P. Sreeram, G. Mukherjee, and S. Manna. Small-world properties of the indian railway network. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 67:36–106, April 2003. doi: 10.1103/PhysRevE.67.036106.

C. Ducruet and I. Lugo. Structure and dynamics of transportation networks: Models, methods, and applications. August 2011. doi: 10.4135/9781446247655.n20.

E. C. o. A. ECA. A european high-speed rail network: not a reality but an ineffective patchwork, 2019. URL https://op.europa.eu/webpub/eca/special-reports/high-speed-rail-19-2018/en/#chapter9.

P. D. f. S. . C. P. European Parliament. Passenger night trains in europe: the end of the line? Technical report, May 2017. URL https://www.europarl.europa.eu/RegData/etudes/STUD/2017/601977/IPOL_STU(2017)601977_EN.pdf.

M. Gastner. *Spatial distributions: density-equalizing map projections, facility location and two-dimensional networks*. PhD thesis, University of Michigan, 2005.

M. Gastner and M. Newman. Shape and efficiency in spatial distribution networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(01):P01015–P01015, January 2006a. doi: 10.1088/1742-5468/2006/01/p01015.

M. Gastner and M. Newman. Optimal design of spatial distribution networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 74:016117, August 2006b. doi: 10.1103/PhysRevE.74.016117.

R. Guimerà, S. Mossa, A. Turtschi, and L. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences of the United States of America*, 102:7794–7799, June 2005. URL https://doi.org/10.1073/pnas.0407994102.

D. B. Interational Union of Railway. Uic-study night trains 2.0: New opportunities by hsr? Technical report, April 2013. URL https://www.wirsindanderswo.de/fileadmin/user_upload/Download-Dateien/study_night_trains_2.0.pdf.

S. Kallas. Roadmap to a single european transport area – towards a competitive and resource-efficient system. 2011. doi: 10.2832/30955.

M. Kurant and P. Thiran. Extraction and analysis of traffic and topologies of transportation networks. *Physical Review E*, 74(3), September 2006. doi: 10.1103/physreve.74.036114.

W. Li and X. Cai. Statistical analysis of airport network of china. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 69:46–106, May 2004. doi: 10.1103/PhysRevE.69.046106.

W. Li and X. Cai. Empirical analysis of a scale-free railway network in china. *Physica A: Statistical Mechanics and its Applications*, 382(2):693 – 703, August 2007. URL https://doi.org/10.1016/j.physa.2007.04.031.

J. Miles. For the love of earth, stop traveling. *The Washington Post*, December 2017. URL https://www.washingtonpost.com/news/theworldpost/wp/2017/11/02/plane-pollution/.

M. Newman. The structure and function of complex networks. *Computer Physics Communications*, 147:40–45, March 2003. doi: 10.1016/S0010-4655(02)00201-1.

M. Newman. *Networks*. Oxford University Press, 2 edition, 2018. ISBN 9780198805090.

C. Pyrgidis. *Railway Transportation Systems: Design, Construction and Operation*. CRC Press, 2016. ISBN 9781482262162.

G. Ramalingam and T. Reps. On the computational complexity of dynamic graph problems. *Theoretical Computer Science*, 158(1):233 – 277, May 1996. URL https://doi.org/10.1016/0304-3975(95)00079-8.

T. Reid, T. Walter, J. Blanch, and P. Enge. GNSS integrity in the arctic. September 2015. URL https://doi.org/10.1002/navi.169.

R. Rozenberg, S. Szabo, and I. Šebeščáková. Comparison of fsc and lcc and their market share in aviation. *International Review of Aerospace Engineering*, 7:149–154, October 2014. doi: 10.15866/irease.v7i5.4439.

F. Schweitzer, H. Rosé, W. Ebeling, and O. Weiss. Optimization of road networks using evolutionary strategies. *Evolutionary Computation*, 5:419–438, December 1997. doi: 10.1162/evco.1997.5.4.419.

J. Sienkiewicz and J. Hołyst. Statistical analysis of 22 public transport networks in poland. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 72:046127, 10 2005. doi: 10.1103/PhysRevE.72.046127.

Taniguchi. High speed rail in japan: A review and evaluation of the shinkansen train. April 1992. URL https://escholarship.org/uc/item/5s48m11f.

T. Times. Flights cheaper than trains for majority of long-distance uk trips. August 2019. URL https://www.thetimes.co.uk/article/flights-cheaper-than-trains-for-majority-of-long-distance-uk-trips-hfvzxbwtg.

Transport and Environment. Leaked study shows aviation in europe undertaxed. May 2019. URL https://www.transportenvironment.org/sites/te/files/publications/2019_05_Tax_report_briefing_web_0.pdf.

M. Vaudano and A. Sénécat. Quels trajets intérieurs en avion pourraient être interdits en cas d'alternative en train ? *Le Monde*, June 2019. URL https://www.lemonde.fr/les-decodeurs/article/2019/06/07/la-majorite-des-vols-interieurs-pourrait-se-faire-en-train-et-ce-serait-bon-pour-l-environnement_5472992_4355770.html.

J. Wang, H. Mo, F. Wang, and F. Jin. Exploring the network structure and nodal centrality of china's air transport network: A complex network approach. *Journal of Transport Geography*, 19(4):712 – 721, July 2011. URL https://doi.org/10.1016/j.jtrangeo.2010.08.012.

D. Watts and S. Strogatz. Collective dynamics of small world networks. *Nature*, 393: 440–442, September 1998. doi: 10.1038/30918.

S. Wei, J. Yuan, Y. Qiu, X. Luan, S. Han, W. Zhou, and C. Xu. Exploring the potential of open big data from ticketing websites to characterize travel patterns within the chinese high-speed rail system. *PLOS ONE*, 12, June 2017. URL https://doi.org/10.1371/journal.pone.0178023.

S. Wei, S. N. Teng, H.-J. Li, J. Xu, H. Ma, X.-l. Luan, X. Yang, D. Shen, M. Liu, Z. Y. X. Huang, and C. Xu. Hierarchical structure in the world's largest high-speed rail network. *PLOS ONE*, 14, February 2019. URL https://doi.org/10.1371/journal.pone.0211052.

Z. Xu and R. Harriss. Exploring the structure of the u.s. intercity passenger air transportation network: A weighted complex network approach. *GeoJournal*, 73: 87–102, October 2008. doi: 10.1007/s10708-008-9173-5.

Appendix

A1 Glossary

- CoI: Countries of Interest
- EATN: European Air Transport Network
- ERTN: European Rail Transport Network
- HSR: High-Speed Rail
- VHSR: Very-High-Speed Rail
- VLDNT: Very Long Distance Night Train

A2 Graph of city to city journeys which are faster by train

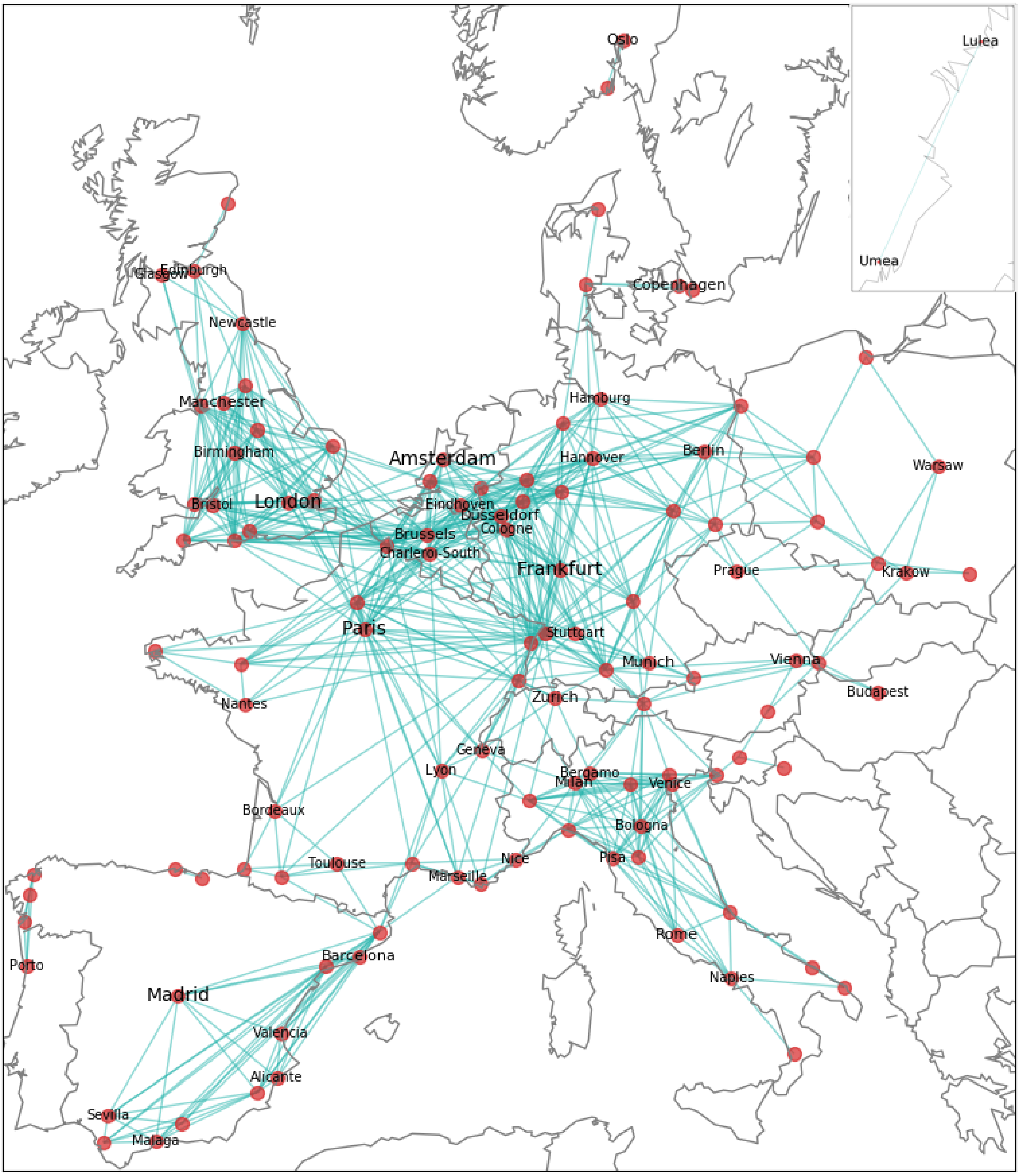


Figure A2.1: Graph of city to city journeys which are faster by train

A3 Passengers transfer from air to rail: results table

Additional travel hours	Number of passengers (millions)	Share of traffic in EATN (%)	Share of total air traffic (%)	Number of flights (thousands)
0	4.31	5.61	2.81	21.83
0.5	7.19	9.35	4.69	36.39
1	11.23	14.61	7.32	55.31
1.5	14.42	18.76	9.41	71.62
2	16.91	22.00	11.03	83.26
2.5	20.01	26.04	13.05	99.10
3	22.37	29.10	14.59	111.11
4	26.79	34.85	17.46	134.31
5	32.31	42.04	21.07	158.58
6	35.07	45.63	22.87	171.93

Table A3.1: Passenger transfer from air to rail and number of flights deleted as a function of the additional travel hours that passengers are ready to spent in the train.

The share of traffic in EATN designates the ratio over the total number of passengers on flights between the airports considered. That is, all flights between airports with more than 500 000 passengers per year in the CoI and from cities which are connected to the ERTN. This gives an idea to which extent the air traffic can be reduced in the European network considered.

The share of total air traffic is the ratio over the total number of passengers departing from the CoI on all flights: national, international to CoI and international outside CoI. This gives an idea to which extent the total air traffic can be reduced in the CoI.

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