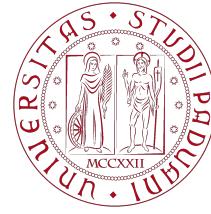


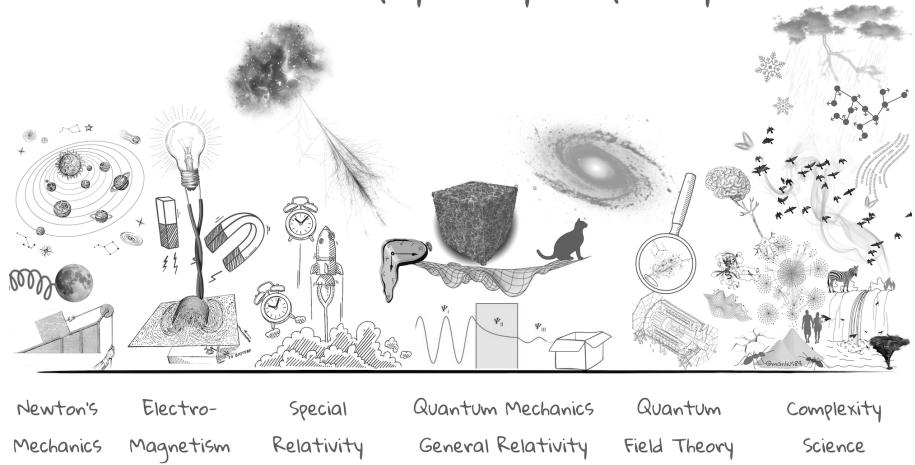
# Final Report

dynamics



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

## Areas of physics by complexity



## Projects:

#16: Traffic congestion

#45: European transportation networks I

Biliato, Pietro

Last update: February 15, 2026

# Contents

---

<b>1</b>	<b>Traffic congestion</b>	<b>1</b>
1.1	A brief review of the sources . . . . .	1
1.2	Methods . . . . .	1
1.3	Results . . . . .	2
1.4	Microscopic analysis . . . . .	5
<b>2</b>	<b>European transportation networks</b>	<b>7</b>
2.1	Data Description and Pre-processing . . . . .	7
2.2	Topology Reconstruction . . . . .	8
2.3	Analysis and Results . . . . .	8
<b>3</b>	<b>Bibliography</b>	<b>15</b>

# 1 | Traffic congestion

---

The aim of this task is to approach routing policies based on traffic awareness by investigating 3 reference papers [2, 4, 3] reproducing their key findings.

## 1.1 | A brief review of the sources

---

The reference literature addresses the mechanisms of packet delivery and the onset of jamming through three perspectives. [2] establishes a foundational model for communication within hierarchical branching structures ("trees"), characterizing analytically the phase transition between free-flow and congested regimes via a specific order parameter. [4] address routing *efficiency* on Scale-Free networks, introducing a tunable "traffic-awareness" parameter, first assessing the dependence of the efficiency on the topological characteristics of the network, and then determining an optimal routing strategy ( $h$  value) for an empirical network. Finally, [3] focuses on the critical behavior of the system, using the same traffic-aware protocol on the same empirical network, proving that the specific protocol can alter the nature of the jamming transition, from continuous to discontinuous, and modify the microscopic distribution of load across the network.

## 1.2 | Methods

---

The task focused on investigating the interplay between the *network structure* and the *traffic dynamics*, of the networks and protocols presented in the sources, in influencing the critical behavior of the system.

**Networks:** Tree networks, to reproduce the hierarchical branching effects, Scale-Free (SF) networks using the Barabási-Albert model, characterized by a degree distribution  $P(k) \sim k^{-\gamma}$  like the empirical networks used in [4, 3], Random networks (Gilbert) (RN) as a null model to control for the effects of structural order. All networks were ensured to be connected by extracting the largest giant component where necessary.

**The routing dynamics:** The discrete-time "traffic-aware" protocol proposed in [4] was used: for a packet at node  $i$  destined to node  $j$ , the next hop is selected from  $i$ 's neighbors by minimizing the cost function:

$$\delta_n = h \cdot d_{n,j} + (1 - h) \cdot c_n \quad (1.1)$$

where  $d_{n,j}$  is the shortest-path from neighbor  $n$  to destination  $j$ ,  $c_n$  is the current queue size at neighbor  $n$ , and  $h \in [0, 1]$  is the tunable parameter controlling the degree of "traffic-awareness", a local information.

The time evolution of the system was simulated through a synchronous update loop: at each time step  $t$ , new packets are introduced into the network at a rate  $p$  (Poisson process), with uniformly random source and destination nodes. Nodes operate forwarding exactly one packet per time step from their First-In-First-Out queue. Packets reaching their destination are removed from the network, and their travel times are recorded. The simulation tracks the total number of active packets  $N(t)$  to monitor the macroscopic state of the system.

The critical behavior of the system was analyzed performing a parameter sweep over the packet generation rate  $p$  computing the **order parameter**  $\rho$ :

$$\rho = \lim_{t \rightarrow \infty} \frac{N(t + \tau) - N(t)}{\tau p} \quad (1.2)$$

with  $\tau$  the observation time, here set to 1. Since a value of  $\rho > 0$  signals a congested phase, the critical point  $p_c$  was identified as the onset of non-zero  $\rho$ . The behavior of  $\rho$ , and thus of  $p_c$ , was monitored over 12 different configurations: 4 topological realizations of each network, using 3 values of  $h$  to explore each. Runs were repeated 5 times and the results averaged.

In addition, as in [4, 3] the microscopic distribution of the load was analyzed, by correlating the queue sizes of the nodes with their Betweenness Centrality. This allows verifying whether congestion is localized at hubs (typical for  $h = 1$ ) or distributed among lower-degree nodes (typical for  $h < 1$ ).

### 1.3 | Results

---

**Tree:** About topology influence: from Figure 1.1(a) one can see that for  $h = 1$  all the curves collapse into a single one. For smaller values, the compatibility is not complete and follows unclear patterns, with one notable exception: for a fixed connectivity  $z$ , if the number of nodes is sufficiently small ( $m = 3$ , with  $m = \#$  layers), using a protocol that on a Tree network would look inefficient ( $h = 0.3$ ), it actually reduces congestion and makes the network more efficient (Figure 1.4). For larger network sizes and connectivities, however, the expected behavior reappears: although the curves in Figure 1.1(b) are all compatible,  $h = 0.3$  leads to less efficient communication (Figure 1.4); curious remains the behaviour of the network for  $h = 0.8$ , from (Figure 1.4).

**SF:** Figure 1.2 shows precisely the results from [3]: the right-shift of  $p_c$  by using  $h < 1$  (only for high enough  $h$ ), as well as the change of the order of the phase transition-like behavior of  $\rho$ ; also that the order of the transition is independent of  $h$ , if  $< 1$ . We can also see the influence of topology: the degree ( $M$  in the plots) is the more relevant factor, but also the number of nodes  $N$  influences the results (for a high enough  $M$ ). Figure 1.5 compares them all in a single graph. There may be an emergence of a tricritical point, in fact the behavior of the configuration "incrM" and "incrNM" for  $h = 0.3$  seems to not match with neither a first order nor a second order phase transition. Thus, the interplay between topology and dynamics is non trivial, and requires further investigation to be further assessed.

**RN:** Figure 1.3 shows results that are analogous to Figure 1.2, with the crucial differences that here it is  $N$  the topological factor that most influences the results, and that the "tricritical-like" phenomenology is already evident for the "standard" topological configuration among those considered.

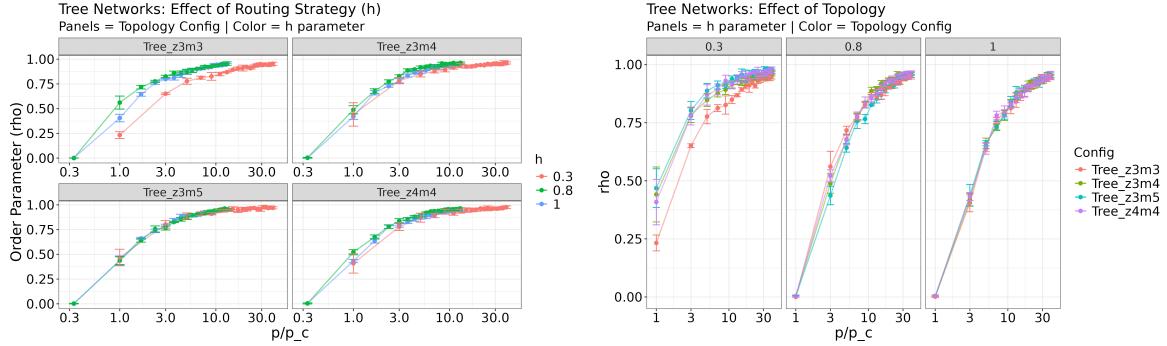


Figure 1.1: Influence of topology (left, a) and  $h$  (right, b) on Tree networks. The x-axis is in log-scale.  $p/p_c$  was used since the focus is on the collapse of the curves, rather than on the estimate of  $p_c$ .

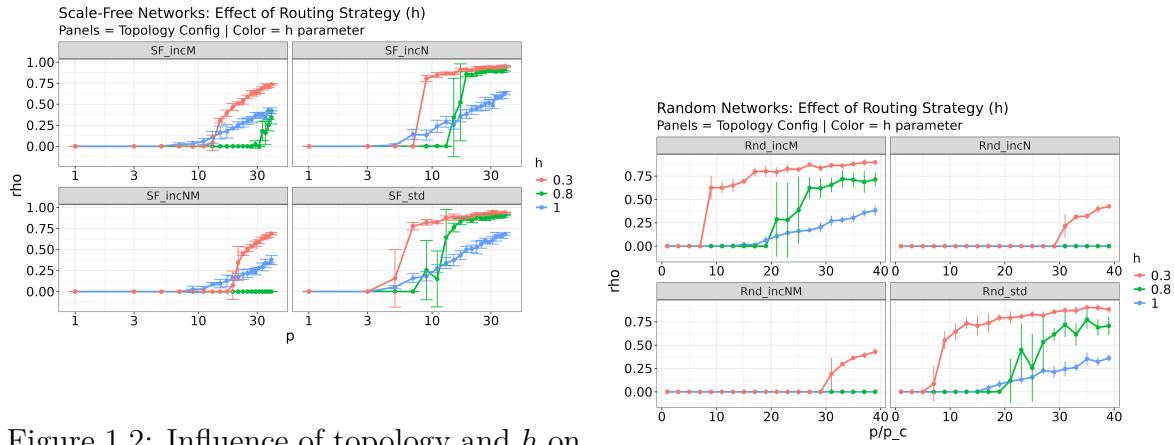


Figure 1.2: Influence of topology and  $h$  on SF networks. The x-axis is in log-scale.  $p$  was used since the focus is both in studying the change of nature of the phase transition and on the shift of  $p_c$ , both due to changes in  $h$ .  $M$  is the degree,  $N$  the number of nodes.  $N_{\text{std}} = 200$ ,  $M_{\text{std}} = 2$ ,  $N_{\text{incr}} = 400$ ,  $M_{\text{incr}} = 4$ .

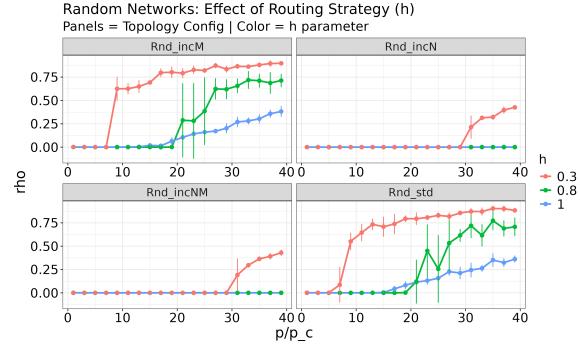


Figure 1.3: Analogous to Figure 1.2 but for RN. We still talk about  $M$  because the link probability was determined as  $2*M/(N-1)$

# Appendix

---

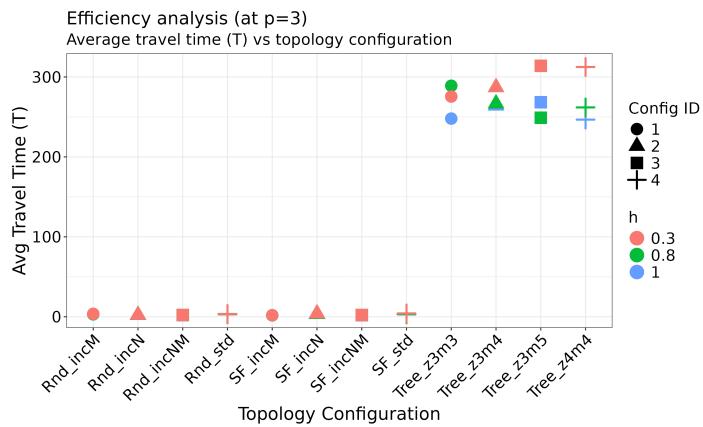


Figure 1.4: Average packet travel time  $\langle T \rangle$  of the various configurations tested, at  $p = 3$

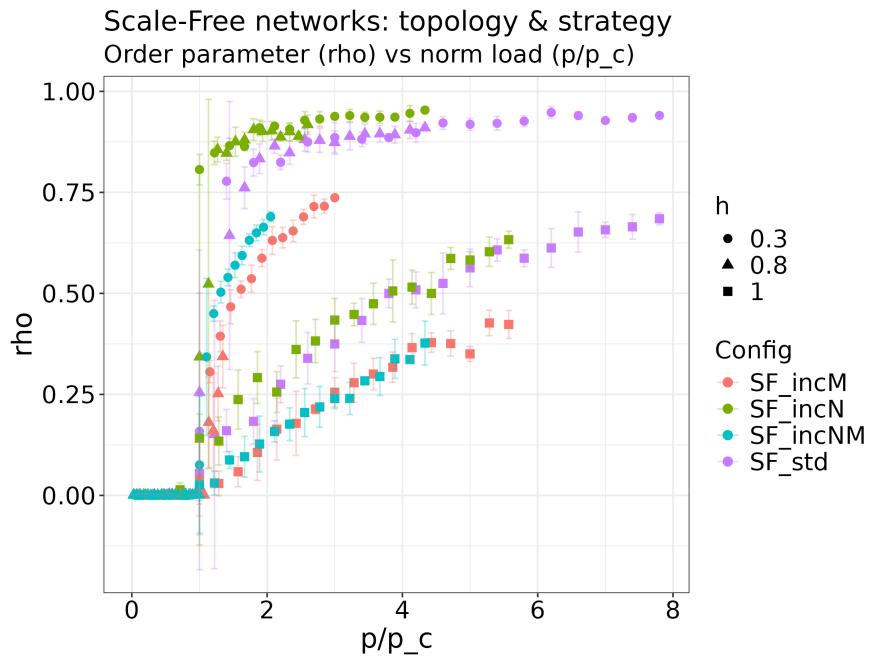


Figure 1.5: Influence of topology and  $h$  on SF networks. The x-axis represents  $p/p_c$  to check the consistency of the estimates of  $p_c$ .

## 1.4 | Microscopic analysis

---

Out of the 12 configurations analyzed, the topological ones that turned out to be more sensitive to changes in  $h$  were selected to try to better understand the microscopic cause of the observed behaviors. These configurations resulted in  $z=3$ ,  $m=3$  for the Tree,  $N=400$ ,  $m=4$  for the SF,  $N=200$ ,  $m=2$  for Random ( $p_{\text{link}}=0.02$ ). It was computed the queue size of each node at stationarity as a function of their betweenness, at  $p=10$  (which is a value at which some configurations may be congested and some others may not). Figure 1.6 shows these quantities, while Table 1.1 computes some statistics on them to better understand them. WAQ is the Weighted Avg Queue (average queue size weighted by the node's centrality), CCM the Congestion Center Mass (average betweenness centrality of the network weighted by the packet load), Corr the Spearman Correlation between  $x$  and  $y$ .

$$\text{CCM} = \frac{\sum_{i \in V} q_i b_i}{\sum_{i \in V} q_i} \quad (1.3)$$

$$\text{WAQ} = \frac{\sum_{i \in V} q_i b_i}{\sum_{i \in V} b_i} \quad (1.4)$$

where:

- $V$  is the set of all nodes in the network.
- $q_i$  is the size of the queue (number of packets) at node  $i$ .
- $b_i$  is the betweenness centrality of node  $i$ .

**Scale-Free Networks:** The SF results further validate the routing protocols of [3]: for  $h = 1$ ,  $\text{CCM} = 0.1345$ , which, compared to the two other values, indicates that congestion is localized at nodes with high betweenness.  $h = 0.8$  reduces drastically the load (total packets = 20), clearing the congestion; in fact  $\text{CCM} = 0.051$ . So in the standard protocol packets travel via the shortest paths, which naturally funnel through hubs. Traffic awareness redistributes instead the load to lower-betweenness nodes.

**Tree Networks:** The Tree network shows insensitivity to the routing parameter  $h$ : #total packets remains extremely high ( $\approx 8200$ – $8500$ ), indicating a collapsed state, as well as the  $\text{CCM}$ , and the Spearman correlation between queue size and betweenness is strong ( $> 0.64$ ) with respect to the others. This phenomenology is explained by presence of one unique path between any pair of nodes. However, the slight increase in #total packets (8500 vs 8215) and decrease in correlation (0.64 vs 0.79) at  $h = 0.3$  suggests that low  $h$  induces "wandering", as observed in the earlier discussion of the results.

**Random Networks:** At  $h = 0.3$ , the system undergoes a collapse, with #total packets exploding to 5986 (vs packets  $\approx 50$  of the other two cases). It can be interpreted through [4], which warns that if  $h$  is too small, packets are diverted to less-loaded nodes "regardless of the path length". Unlike Scale-Free networks, Random networks are homogeneous; they lack hubs that necessitate avoidance and at  $h = 0.3$ , the cost function is dominated by congestion, so one may hypothesize that this causes packets to lose their directionality, thus remaining stuck in the network.

Table 1.1: Microscopic Statistics Summary

Network	$h$	Packets	WAQ	CCM	Corr
Random	1	54	0.980	0.055	0.401
Random	0.8	46	0.430	0.028	0.374
Random	0.3	5986	32.217	0.016	0.580
SF	1	457	32.317	0.135	0.258
SF	0.8	20	0.542	0.052	0.277
SF	0.3	35	0.595	0.032	0.297
Tree	1	8215	2011.217	0.512	0.798
Tree	0.8	8402	1816.435	0.452	0.776
Tree	0.3	8500	1836.696	0.452	0.647

Note: WAQ = Weighted Avg Queue, CCM = Congestion Center Mass, Corr = Spearman Correlation.

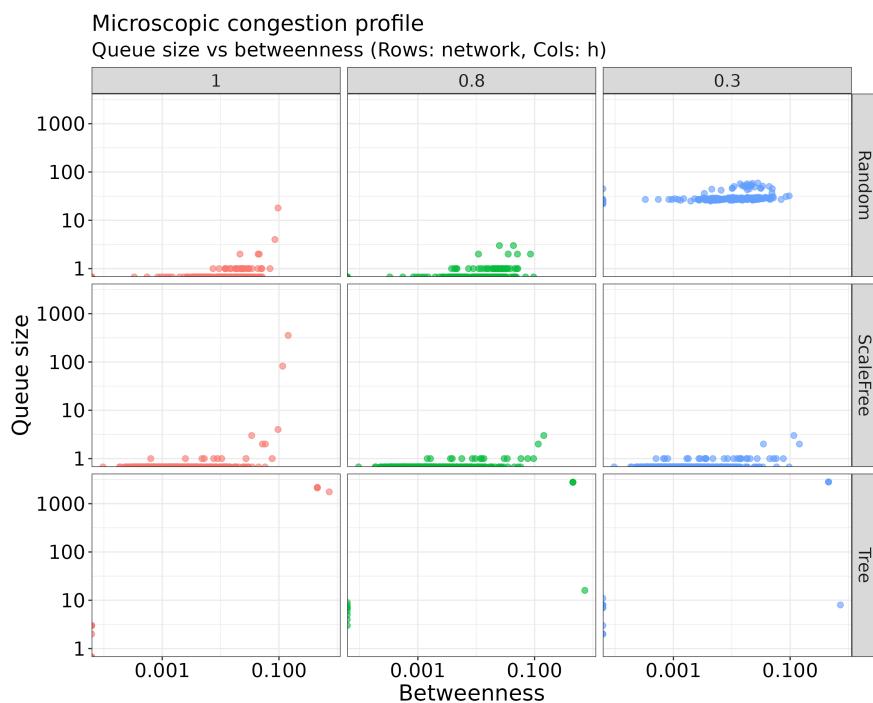


Figure 1.6: Queue size of each node at stationarity as a function of their betweenness, at  $p=10$

## 2 | European transportation networks

---

The objective of this project is to reconstruct and analyze the topology of the European rail network using high-fidelity geospatial data, the **EuroGlobalMap (EGM) 2019** dataset [5], a seamless pan-European topographic database produced by EuroGeographics.

### 2.1 | Data Description and Pre-processing

---

The raw data was provided in shapefile format, which describes geometric features rather than explicit topological graphs. Consequently, the primary challenge was to reconstruct raw spatial primitives into a graph  $G = (V, E)$ , where  $V$  represents valid stations and  $E$  represents tracks connecting them.

From the EGM specification [6], the `RailrdC` file (railway nodes) contains various facility types, identified by two attributes: *Transportation Use Category* (TUC), *Transportation Facility Type* (TFC). The TFC attribute characterizes also the `RailrdL` (railway lines) dataset, along with another main feature as the *Existence Category* (EXS). There are many other attributes however in `RailrdL`, which identify a variety of properties, such as *Railroad Power Source*, *Gauge Width*, *Location Level*, ... but they will not be considered in this analysis.

It was chosen to focus only on population-related infrastructures, being fully aware that this is not the only relevant aspect. In the event of an attack, for example, stations would not be the only possible targets. This was simply a deliberate choice. Thus, data was filtered accordingly.

About nodes, only those corresponding to the following attributes were retained:

- **TFC 15, 31, 32, 34:** Railway Station, Joint R.S., Halt, Terminal.
- **TUC: 26, 45:** Passenger and General respectively.

About lines instead:

- **EXS 28:** Operational (excluding "Under Construction" or "Abandoned");
- **TUC: 26, 45:** same as nodes.

Also, the EGM is indeed a seamless database, in fact all features have valid inputs for all attributes, eventually values that mark the inappropriateness of the value; one can therefore easily identify them and filter them out (*Null/No value*, *Unknown*, *Unpopulated*, as well as specific integers: -32768, -29999, -29997, -29998, 997, 998).

## 2.2 | Topology Reconstruction

---

The aim of this section is to build the network topology. [6] defines, in fact, connectivity at the level of geometric primitives, where edges meet, only, at connected nodes; a node, in RailrdC, is any point where a line starts, ends, or connects to another line; all lines are connected to each other by nodes. However, shapefiles lack explicit mapping from "Source ID" to "Target ID" nodes.

First, to prevent artificial fragmentation of the graph, `hclust` (Agglomerative hierarchical clustering) was applied on the station coordinates, since [6] states that connected points within 200 meters should be combined; the centroid of each cluster was assigned as the unique coordinate for that station node.

Then, since line segments in shapefiles differ from graph edges, the start and end coordinates of every valid line segment in `RailrdL` were extracted, and unique coordinate pairs were identified to form the set of potential topological nodes. A "Snapping" algorithm was in fact implemented to classify these nodes:

1. **Stations:** If a topological node is located within a tolerance of 50m from a filtered `RailrdC` point, it is labeled as a *Station*. The declared tolerance is of 30m; 50m were used for safety.
2. **Junctions:** Any topological node not matching a station is labeled as a *Junction*. These are in fact required to track geometry but they are not relevant for passenger boarding.

To get to study the basic statistics of this network in terms of passenger flow, a *logical network*  $G_{log}$  had to be derived. It was done via a "graph traversal" algorithm that initiates a "walker" at every `Station` node, it traverses the physical edges of those nodes passing through `Junctions` without stopping, stopping instead as soon as it encounters another `Station` node.  $G_{log}$  thus results in an undirected graph where nodes are strictly `Station` and edges represent (actual) `Station-Station` connections.

## 2.3 | Analysis and Results

---

Let's start by analyzing the global properties of the network from Table 2.1. One can see that the reconstructed "logical" network of the European Union consists of  $N = 8133$  stations and  $E = 71739$  edges. The global average degree is  $\langle k \rangle \approx 17.6$ , but this aggregate figure masks a profound structural heterogeneity between national sub-networks. Two distinct topological regimes can in fact be observed:

- **High connectivity** (e.g., France  $\langle k \rangle \approx 86.6$ , Spain  $\langle k \rangle \approx 32.4$ ): in these countries the "logical" abstraction results in extremely high connectivity, likely due to "intercity" or "high-speed" lines. These lines cross vast areas, allowing direct connections between stations that would otherwise remain unlinked. However, they are dedicated lines, not accessible to all trains, and therefore an artefact has been introduced. At the same time, excluding them would still yield an inaccurate representation of connectivity. It was therefore decided to proceed under the assumption that all lines are equally accessible.
- **Low(er) connectivity** (e.g., Germany  $\langle k \rangle \approx 9.2$ , Poland  $\langle k \rangle \approx 11.0$ ): an average

degree closer to likely physical track constraints, indicating a higher density of valid stations and a more "granular" stopping pattern.

One can also see that the network displays an anomalously high Assortativity Coefficient ( $r \approx 0.97$ ), which conveys structural segregation, of "modules" with internally consistent topologies that do not mix; apparently, cross-border integration is too weak to blend these distinct topological characteristics. These "modules" can be the "high-degree, low-degree" clusters that integrate minimally, islands and detached regions. The Giant Component comprises infact 76.6% of the nodes, mostly because neither the "Channel line" nor the "Øresund Line" are accounted for, then there are minor islands, the Kaliningrad Oblast and potential data discontinuities at specific borders.

Moving to microscopic analysis, three centrality measures were computed:

**Harmonic Centrality** Given the disconnected nature of the graph, Closeness Centrality is ill-defined (Figure 2.2). Harmonic Centrality was then used (Figure 2.2), which shows a reasonable core-periphery structure.

**Betweenness Centrality** Its pattern (Figure 2.2) looks peculiar, but can be explained, starting from the linear "rows" of high-betweenness stations increasing from east to west in Romania. Those are in fact the only crossings that allow the, dense, eastern european region to connect to the rest of the continent. The same applies to the corresponding patterns in the northeast. Finally, from Figure 2.5, one can see that the other points with significant betweenness all appear to be non-peripheral nodes acting as "entry" or "exit" points to highly connected areas. All these behaviours are compatible with the definition of Betweenness Centrality.

**Katz Centrality** The pattern of Katz Centrality reflects the position of the nodes with high logical degree (Figure 2.2, Figure 2.1), which matches with the concept of centrality that Katz-centrality implements: a node is important according to the (weighted) degree of its high-order neighbors.

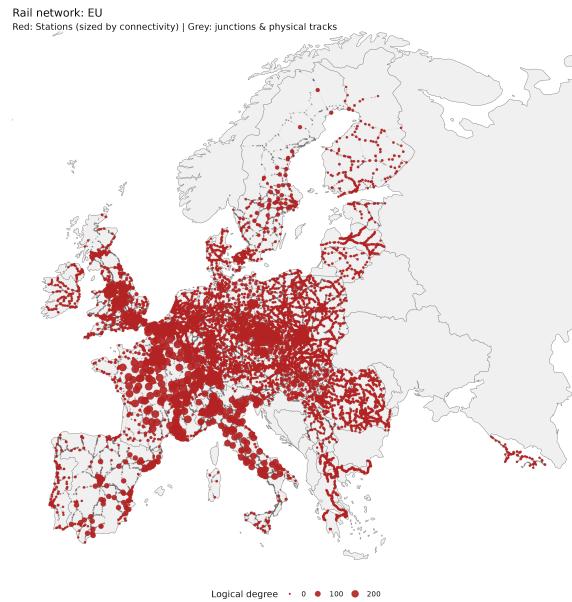


Figure 2.1: Red nodes are the Stations, with a size proportional to their "logical" degree (# Station-Station connections), gray nodes are the Junctions

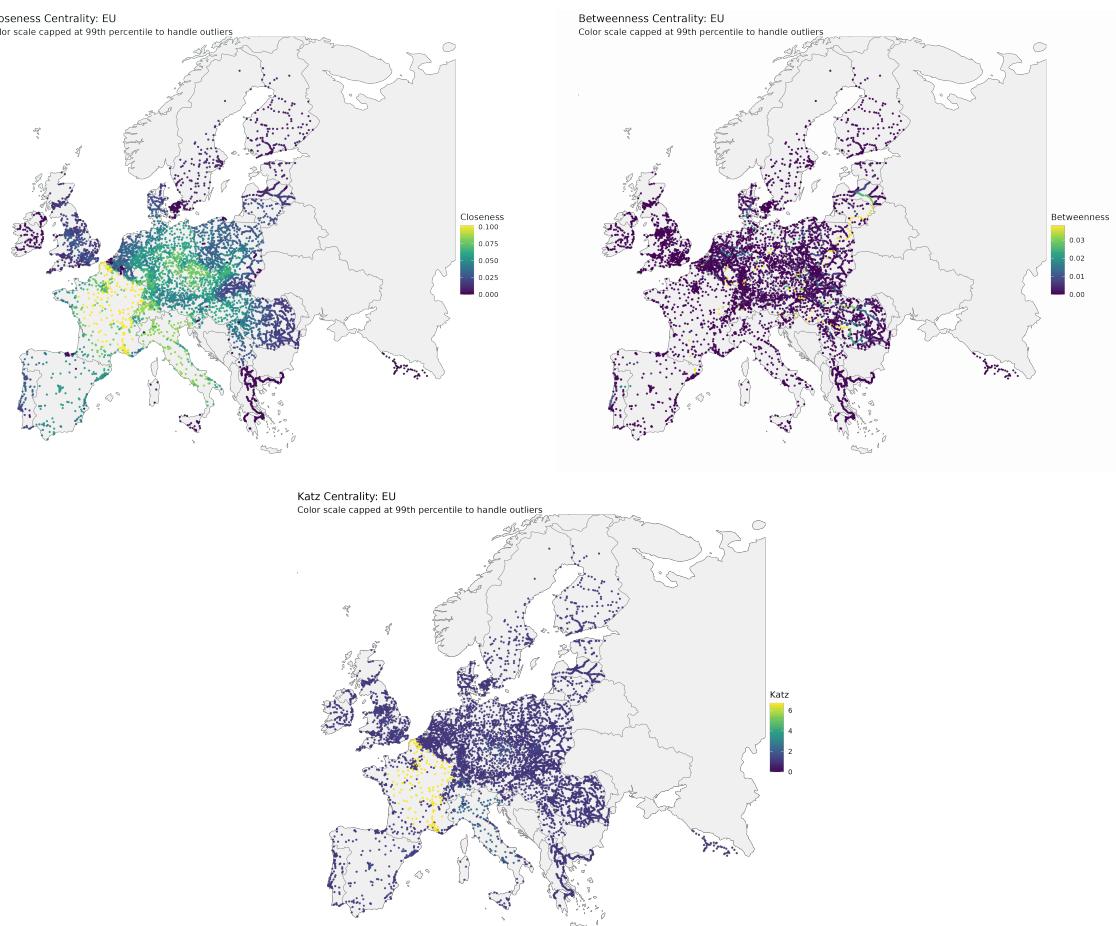


Figure 2.2: Centrality Measures Analysis. Top-Left: Harmonic Centrality showing a center-periphery pattern. Top-Right: Betweenness Centrality. Bottom: Katz Centrality, displaying a pattern that resembles the "logical" degree distribution (see Figure 2.1).

# Appendix

---

Table 2.1: Some quantities for a global and mesoscopic analysis. 'EU' stands for the whole continent. Not all the countries that from the continental-level dataset are present because individual data for those countries were not available

Country	Nodes	Edges	GCC Size	GCC (%)	$\langle k \rangle$	Glob. Clust.	Assort.
AT	258	708	255	98.84	5.488	0.853	0.890
BE	540	1037	446	82.59	3.841	0.821	0.906
CHLI	152	1567	151	99.34	20.618	0.922	0.771
CZ	207	4259	207	100.00	41.150	0.981	0.909
DE	1102	5056	1093	99.18	9.176	0.841	0.835
DK	179	239	107	59.78	2.670	0.402	0.612
EE	31	35	31	100.00	2.258	0.327	0.413
ES	200	3182	175	87.50	31.820	0.982	0.816
EU	8124	72409	6223	76.60	17.826	0.979	0.974
FI	167	222	166	99.40	2.659	0.329	0.541
FR	614	26628	614	100.00	86.736	0.989	0.824
GB	707	6673	697	98.59	18.877	0.952	0.897
GE	44	58	42	95.45	2.636	0.514	0.388
GR	361	390	317	87.81	2.161	0.228	0.657
HU	136	1147	136	100.00	16.868	0.868	0.717
IE	77	92	77	100.00	2.390	0.314	0.418

## Centrality Measures

---

### Harmonic Centrality

$$C_{\text{harmonic}}(i) = \sum_{j \neq i} \frac{1}{d(i, j)} \quad (2.1)$$

Where:

- $C_{\text{harmonic}}(i)$  is the harmonic centrality of node  $i$ .
- $d(i, j)$  is the geodesic distance (shortest path) between node  $i$  and node  $j$ .
- If  $i$  and  $j$  are not connected,  $d(i, j) = \infty$  and  $1/d(i, j) = 0$ .

### Betweenness Centrality

$$C_B(i) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}} \quad (2.2)$$

Where:

- $C_B(i)$  is the betweenness centrality of node  $i$ .
- $s$  and  $t$  are node pairs in the network distinct from  $i$ .
- $\sigma_{st}$  is the total number of shortest paths between node  $s$  and node  $t$ .
- $\sigma_{st}(i)$  is the number of those shortest paths that pass through node  $i$ .

### Katz Centrality

$$C_{\text{Katz}}(i) = \sum_{k=1}^{\infty} \sum_{j=1}^N \alpha^k (A^k)_{ji} \quad (2.3)$$

Where:

- $C_{\text{Katz}}(i)$  is the Katz centrality of node  $i$ .
- $N$  is the total number of nodes.
- $A$  is the adjacency matrix of the network ( $A_{ji} = 1$  if connected, 0 otherwise).
- $\alpha$  is the attenuation factor ( $0 < \alpha < 1/\lambda_{\max}$ , where  $\lambda_{\max}$  is the largest eigenvalue of  $A$ ).
- $(A^k)_{ji}$  represents the number of walks of length  $k$  connecting node  $j$  to node  $i$ .

## Additional plots

---

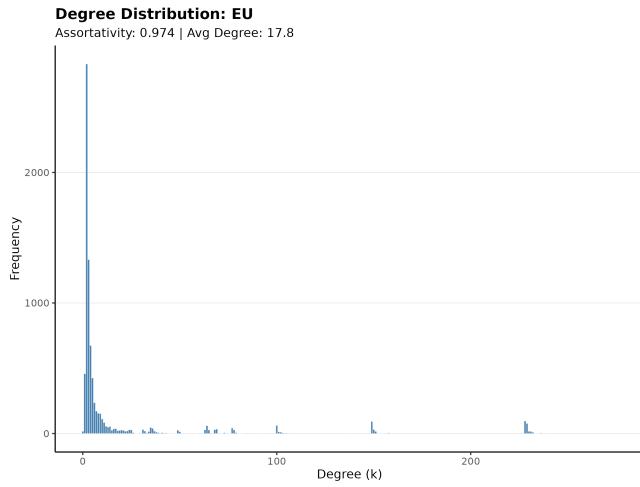


Figure 2.3: Distribution of the "logical" degree of the Stations. Clearly a heterogeneous distribution, like many empirical networks. Very high degrees may be artefacts, due to the "logical" version of the degree this plot represents.

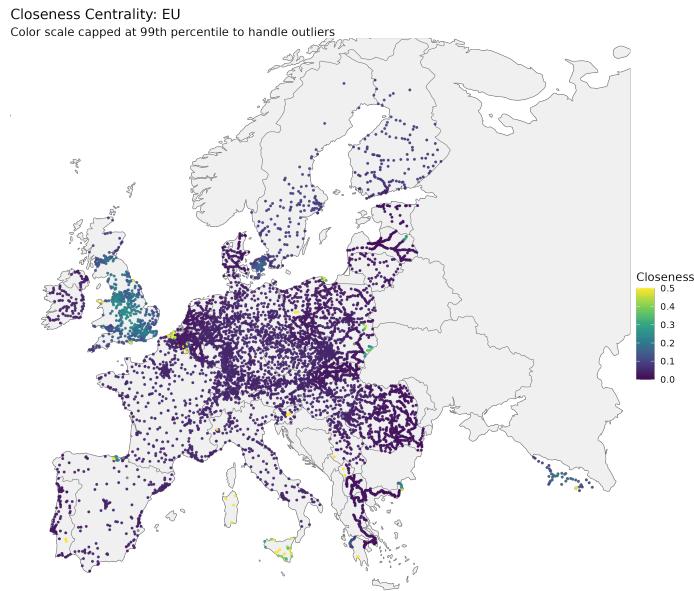


Figure 2.4: Closeness Centrality: since the network is disconnected, we see that this quantity is ill-defined: Sicily, e.g., has one among the highest centrality values.

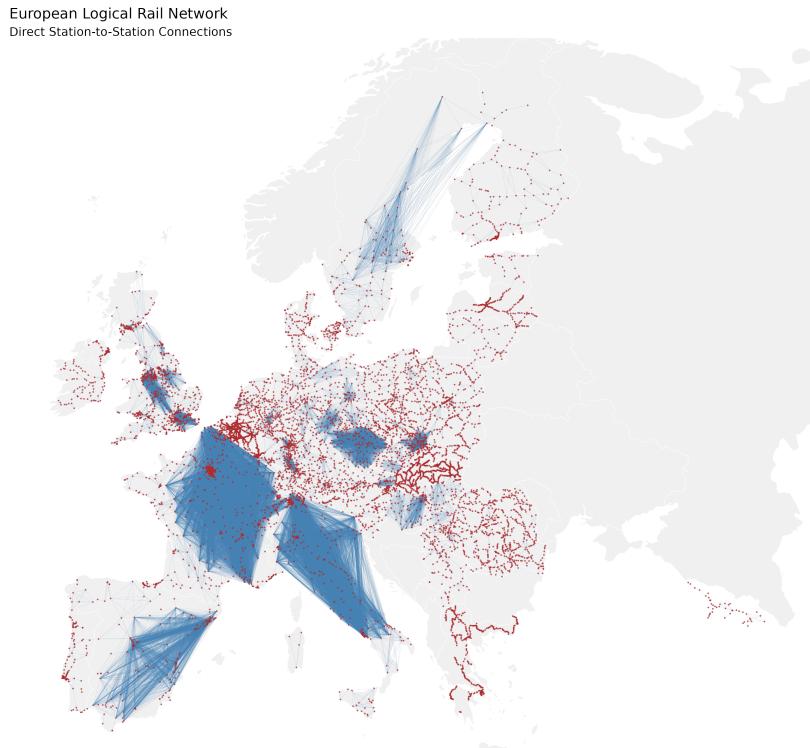


Figure 2.5: Network with Station-Station "logical" connections

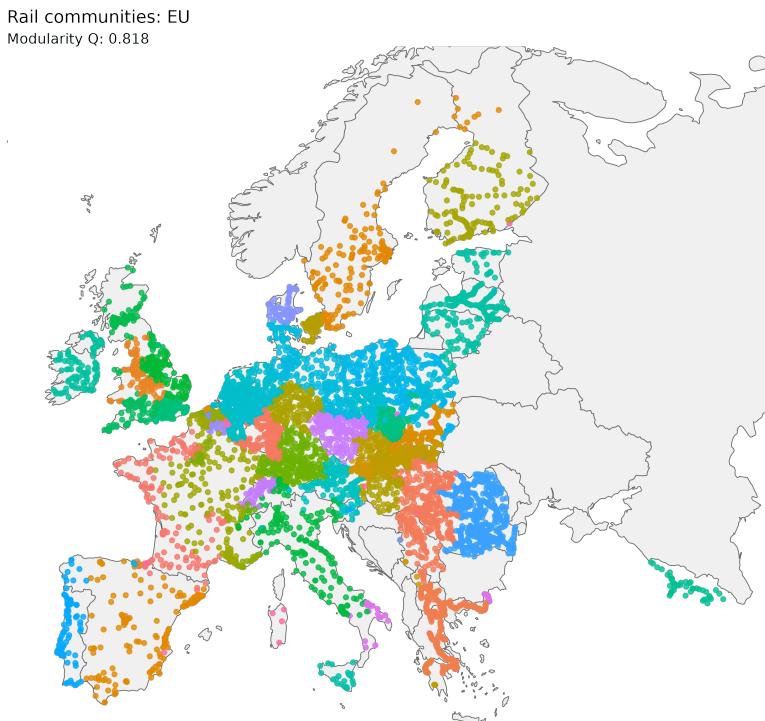


Figure 2.6: Community analysis via a spin-glass-like Pott's mode using the Louvian alforithm. No constraints where put in the number of communities; it is intended to be a basic application of such algorithm to compare the results with the borders of the european countries (since a significant portion of them follow natural boundaries)

## 3 | Bibliography

---

- [1] The assistance of AI (ChatGPT) was used to realize this project. It was used while writing the code to make specific sections more efficient—for example, by ensuring that built-in implementations were used instead of custom ones, by revising the computational logic to make it more concise, and by properly handling special cases through the necessary checks. It was also helpful for improving the plots and the management of their automated generation. Finally, it proved useful in revising some portions of the text to enhance their readability.
- [2] A. Arenas, A. Díaz-Guilera, and R. Guimerà. Communication in networks with hierarchical branching. *Phys. Rev. Lett.*, 86:3196–3199, Apr 2001. doi: 10.1103/PhysRevLett.86.3196. URL <https://link.aps.org/doi/10.1103/PhysRevLett.86.3196>.
- [3] P. Echenique, J. Gómez-Gardeñes, and Y. Moreno. Dynamics of jamming transitions in complex networks. *Europhysics Letters*, 71(2):325, jun 2005. doi: 10.1209/epl/i2005-10080-8. URL <https://doi.org/10.1209/epl/i2005-10080-8>.
- [4] Pablo Echenique, Jesús Gómez-Gardeñes, and Yamir Moreno. Improved routing strategies for internet traffic delivery. *Phys. Rev. E*, 70:056105, Nov 2004. doi: 10.1103/PhysRevE.70.056105. URL <https://link.aps.org/doi/10.1103/PhysRevE.70.056105>.
- [5] EuroGeographics. Euroglobalmap (egm) 2019. Dataset, 2019. URL <https://eurogeographics.org/maps-for-europe/open-data/>.
- [6] EuroGeographics. Euroglobalmap release 2019: Pan-european database at small scale — specification and data catalogue. Technical Specification EGM\_2019\_DataSpecification, EuroGeographics, 2019. URL <https://eurogeographics.org>.