Europe Road Network

Sylvain Sylvestre Icaro Hamers

February 1, 2026

Abstract

This study focuses on the analysis of the structure and key properties found in the European road network. The network is constructed from a 2011 dataset taken from the Netzschleuder repository. Using metrics like density, clustering coefficient, shortest path length, diameter, and degree distributions we find that the European network is sparse, weakly clustered and heavily constrained to geography. European cities are pitted against each other through multiple node importance measures, including degree, closeness, betweenness, eigenvector, Katz, PageRank, and subgraph centrality from which Moscow, Paris and Warsaw are highlighted as the key cities that connect the European vehicle transport. The network is compared to Erdős–Rényi and Barabási–Albert null models with results confirming that these characteristics are not due to random chance. Regions are determined based on clustering with the Infomap and Louvain algorithms where we find that modularity based models outperform random walk algorithms. Finally, a graph is generated to visually showcase all the most interesting characteristics of the European road network.

1 Introduction

This project focuses on the network formed by the connections of roads between cities found in Europe and Central Asia. We aim to study the characteristics of this network through multiple metrics in order to find key characteristics of these European cities. The data was taken from the website *Netzschleuder* [2]. This dataset includes the cities and roads from islands found in Europe, however this study will only focus on the largest component which corresponds to the cities found in the continental mainland. It is also noticeable that this dataset was taken in 2011 and doesn't include subaquatic tunnels which nowadays would connect some of the countries together.

2 Methods

All the programming was carried out through the *Google Colab* environment using **Python** language and mostly using the *NetworkX* library [1].

2.1 Network properties

In this section we will analyze the main characteristics of the network including the number of nodes and edges, density, average clustering coefficient, average shortest–path length and effective diameter, the probability density function and finally the complementary cumulative probability function. All were computed through *NetworkX*'s built-in functions with the exception of the probability distributions which were computed by analysing the degree of all the nodes and counting the amount of times each degree appears. The complementary cumulative distribution follows formula 1.

$$\mathrm{CCDF} = 1 - \mathrm{CDF} \tag{1}$$

Where CDF stands for cumulative distribution function. This distribution is compared to a Poisson distribution, an exponential distribution and a power law distribution in order to see what sort of behaviour the network has.

2.2 Centrality structure

In this section we analyze the following centrality properties of the network:

- **Degree:** How many direct connections a node has.
- **Closeness:** How close a node is to all others (shortest paths).
- Betweenness: How often a node lies on shortest paths.
- **Eigenvector:** Importance via connections to important nodes.
- **Katz:** Eigenvector-like, but includes distant nodes with attenuation.
- **PageRank:** Random-walk importance.
- **Subgraph:** Nodes embedded in dense local structure.

All these values are computed through *NetworkX*'s built-in functions and then uploaded to a *Pandas* [3] dataframe in order to rank the cities with the highest values of these properties.

2.3 Null model comparison

In order to be sure that our network has any information of value, it is important to make sure it doesn't have the properties of a random network. Therefore, the objective of this section is to compare the network properties of two random networks to the network of European roads.

The two random networks chosen are the Erdős–Rényi network and the Barabási–Albert network. We won't go in depth in the explications of these networks but these are the basic principles:

- **Erdős–Rényi network:** The initial values of this network are the number of nodes and edges. Then these edges are randomly distributed among the nodes in the network in an uniform manner.
- **Barabási–Albert:** stochastic (random) network model based on preferential attachment, so it tends to have a few nodes with a lot of edges. The initial parameters are the number of nodes and the average degree.

2.4 Regions and visualization

In this last section we are going to search for clusters in the network, these are cities that are greatly connected to each other. Considering the network being studied, we are expecting these clusters to be countries and regions. The two clustering algorithms being tested are *Infomap* which is based on random walk and *Louvain* which is based on modularity. Both are evaluated through coverage and performance scores.

Finally, using *NetworkX*'s graphing tool, we will show the regions detected through a graph of the network. Of course the clustering algorithms can't classify the regions so these regions were classified by myself, it is not fully rigorous but it gives a good idea of where we are in the map. Alongside the regions, we determined that the most interesting property to showcase was the node betweenness since it gives insight on which cities have the most traffic when trying to move around Europe. In the graph it is represented as the size of the nodes. Lastly, the names of the cities are displayed on each node. It is important to note that it is impossible to show in detail the graph in the paper, so it is recommended to download the full image found in the Github URL so that the reader can zoom in and see the names of all the cities.

3 Results

3.1 Network properties

The basic properties of this network are:

- **Nodes:** 1039
- **Edges:** 1305
- **Density:** 0.0024
- **Average degree:** 2.51
- **Average Clustering Coefficient:** 0.019
- **Average shortest path length:** 18.40
- **Diameter:** 62

We graph the probability distribution function (Figure 1) and the complementary cumulative distribution function (Figure 2).

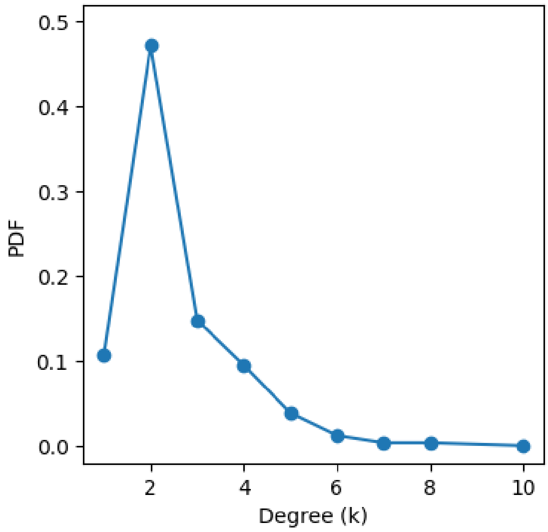

Figure 1: Probability distribution function (PDF) of the Europe roads network.

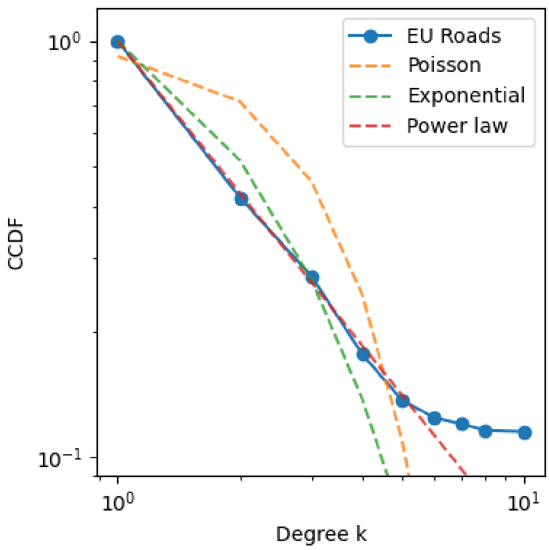

Figure 2: Complementary cumulative distribution function (CCDF) of the Europe roads network compared to a Poisson distribution, an exponential distribution and a power-law distribution. The scale of the axes are logarithmic.

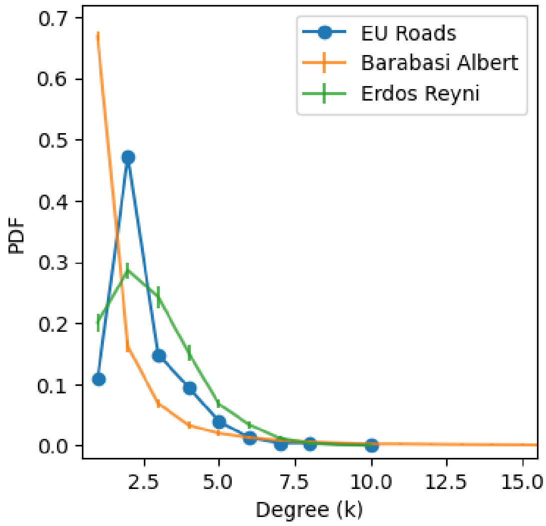

Figure 3: Probability distribution functions of the Europe roads network, the Erdős–Rényi network and the Barabási–Albert network.

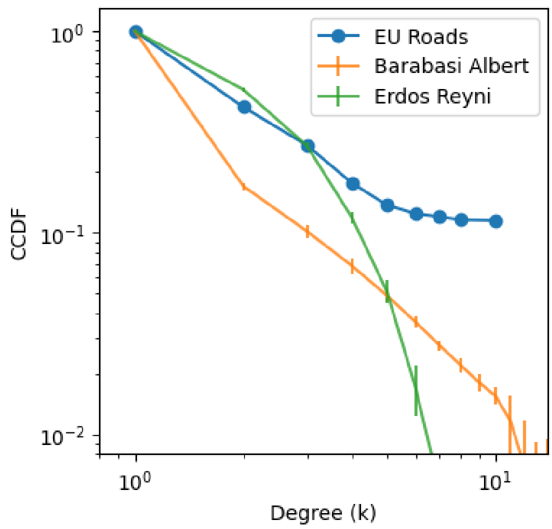

Figure 4: Complementary cumulative distribution functions of the Europe roads network, the Erdős–Rényi network and the Barabási–Albert network. The scales of the axes are logarithmic.

3.2 Centrality structure

The top 25 cities in centrality parameters are represented in Table 1.

3.3 Null model comparison

Erdős–Rényi:

- **Density:** 0.002972 ± 0.000050
- **Average Clustering Coefficient:** 0.0018 ± 0.0012
- **Average shortest path length:** 7.259 ± 0.098
- **Diameter:** 17.300 ± 1.3454

Barabási–Albert:

- **Nº Edges:** 1038.0 ± 0.0
- **Density:** 0.0019 ± 0.0000
- **Average Clustering Coefficient:** 0.0 ± 0.0
- **Average shortest path length:** 6.8061 ± 0.3188
- **Diameter:** 17.600 ± 1.0198

The probability distribution functions and complementary cumulative distribution functions of the random networks are represented in figures 3 and 4. It is important to note that the Barabási–Albert networks had nodes with up to 60 degrees but in order to appreciate the graphs we decided to only show nodes up to 15 degrees.

3.4 Regions and visualization

The Infomap clustering algorithm found only 2 communities, on the other hand the Louvain algorithm found 25 communities.

The Coverage scores are:

- **Infomap:** 0.9985
- **Louvain:** 0.9149

The performance scores are:

- **Infomap:** 0.2414
- **Louvain:** 0.9573

Finally, the network is visualized in Figure 5.

4 Discussion

4.1 Network properties

In this study we focused on the largest component of the network, which has 1039 nodes and 1305 edges. Comparing these values to the original network found in *Netzschleuder* [2] (Nodes: 1174, Edges: 1417) we see that 135 cities were disconnected from the main group, these correspond to cities found in islands. With these stats we can estimate that by focusing on the largest component we are omitting around 11.5% of Europe's cities, this includes entire countries like The United Kingdom, Ireland and Cyprus. It can also be used as a rough estimation of the percentage of Europe's island surface.

Table 1: Top 25 cities in centrality parameters

Rank	Degree	Closeness	Betweenness	Eigenvector	Katz	PageRank	Subgraph
1	Moscow	Warsaw	Brest	Paris	Moscow	Moscow	Moscow
2	Paris	Brest	Moscow	Metz	Paris	Berlin	Paris
3	Munich	Minsk	Saint Petersburg	Reims	Munich	Budapest	Munich
4	Berlin	Lviv	Le Mans	Brussels	Liège	Liège	Liège
5	Budapest	Lublin	Rennes	Liège	Berlin	Munich	Vienna
6	Liège	Kaunas	Minsk	Le Mans	Budapest	Paris	Bratislava
7	Metz	Rennes	Warsaw	Orléans	Metz	Larissa	Metz
8	Warsaw	Smolensk	Smolensk	Luxembourg	Vienna	İzmir	Berlin
9	Prague	Piotrków Trybunalski	Vyborg	Charleville-Mézières	Bratislava	Malmö	Budapest
10	Vienna	Moscow	Vaalimaa	Lyon	Warsaw	Bratislava	Brno
11	Bratislava	Kiev	Kotka	Rouen	Prague	Vienna	Prague
12	Cologne	Gomel	Helsinki	Geneva	Salzburg	Prague	Warsaw
13	Nuremberg	Riga	Paris	Calais	Brno	Innsbruck	Salzburg
14	Bucharest	Rivne	Jyväskylä	Lausanne	Nuremberg	Warsaw	Zagreb
15	Kiev	Radom	Kemi	Tours	Lyon	Lyon	Graz
16	Zagreb	Babruysk	Oulu	St. Vith	Zagreb	Lamia	Milan
17	Salzburg	Pskov	Tornio	Bordeaux	Riga	Riga	Kiev
18	Oradea	Poznań	Kiev	Toulouse	Cologne	Metz	Larissa
19	Riga	Mukachevo	Luleå	Leuven	Kiev	Milan	Bordeaux
20	Graz	Vilnius	Haparanda	Le Havre	Graz	Bucharest	Toulouse
21	Larissa	Saint Petersburg	Umeå	Chaumont	Frankfurt am Main	Cologne	Nuremberg
22	Frankfurt am Main	Le Mans	Niš	Saarbrücken	Bucharest	Dushanbe	Orléans
23	Innsbruck	Kovel	Örnsköldsvik	Charleroi	Milan	Graz	Linz
24	Milan	Elbląg	Sundsvall	Chambéry	Oradea	Oradea	Lyon
25	Lyon	Katowice	Lviv	Eupen	Larissa	Bukhara	Reims

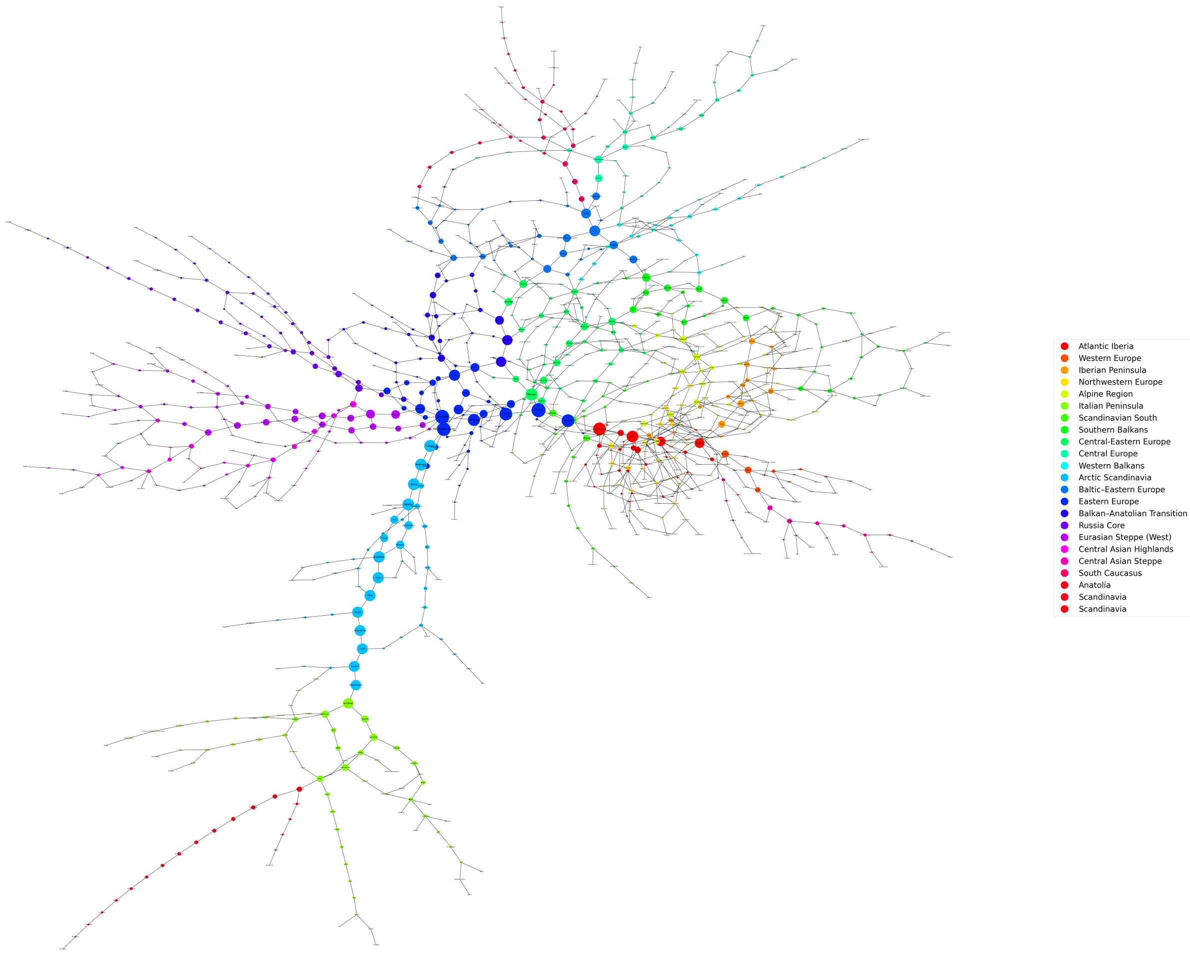

Figure 5: Europe roads network. Colour coded to different regions and cities scaled based on node betweenness.

The density ($\simeq$ 0.0024) is extremely low as expected from this network since, as an example, an edge between Madrid and Moscow is geographically impossible. The average degree ($\simeq$ 2.5) is also makes conceptual sense since each city having around 2-3 connections is reasonable. From the low average clustering coefficient ($\simeq$ 0.019) we can gather that cities connected to a main city are not often connected to each other, meaning that in order to travel from one city to another, even if they are close to each other, it is very likely that we will need to pass through a city with lots of connections. This ties in with the city betweenness that we see in figure 5 where capital cities tend to have a lot of traffic due to their connectivity to the rest of the country.

The average shortest path length ($\simeq$ 18.40) shows us that in order to get from one random place in Europe to another you need to pass through 18-19 cities on average. So if you are doing a car trip around Europe, you have a lot of places to visit before getting to your destination. The shortest trip among the most distant places goes through 62 cities, which correspond to Rennesøy (Norway) - Gytheio (Greece), a beautiful and calm trip which passes through Helsinki, Saint Petersburg, Kiev and Bucharest among others.

4.2 Centrality structure

Looking at table 1 we can see that Moscow is an extremely important city in Europe, being almost the top city in five centrality parameters. This result is expected since this network includes countries in Centre Asia like Kazakhstan and Uzbekistan which are only connected to Europe through Russia and Turkey. A surprising result is how Liège ranks higher than it's country's capital Brussels in most parameters. Other notable cities are Paris and Warsaw which appear as often as Moscow (6 times) and Kiev, Liège, Lyon and Metz which all appear 5 times in the table. There seems to be a preference for East Europe probably due to the density of cities found in this region. We can also see many Belgian cities in the Eigenvector ranking, meaning that Belgium is well connected to the most "important" cities in the network, this may be one of the reasons why this country is considered to be the "centre" of Europe even though geographically it's clearly eastern.

4.3 Null model comparison

Comparing the random networks to the Europe roads network we see that they have similar density and that the Barabási-Albert network has a similar amount of edges. However, every other parameter is much lower in random networks than in our network. We find a lower clustering coefficient in random networks which means that our network has a low value not because of the structure of the network but rather as a result of the low number of edges. The low average shortest path length and diameter found in random networks also shows that the high values found in the European road network are not due to randomness but rather due to the intrinsic nature of geography.

We can however find in figure 3 that the probability distribution of our network has a similar shape to a Erdős–Rényi network with the biggest difference being the drop found in the CCDF for high values of the degree. This is due to the Erdős–Rényi network not being able to replicate the nature of capitals, nodes with high degree values found all around the network. The Barabási–Albert network looks nothing like the European network so it is safe to assume that European cities don't follow a "rich get richer" type of structure.

4.4 Regions and visualization

The coverage scores imply that both algorithms could find clusters with high internal edges, on the other hand, the performance scores indicate that the Louvain algorithm strongly separates internal and external edges between clusters. This can be seen in how the Infomap algorithm only finds 2 regions while the Louvain algorithm finds 25 which is a bit more than expected but reasonable considering that some countries like Italy or Norway/Sweden are very disconnected from the rest of Europe.

Due to the nature of the network, random walk based algorithms seemed like the best choice to find communities, however it seems like modularity based algorithms work much better. An interpretation of these results is that if you were to choose random roads when navigating Europe, after a few travels you would find yourself in a very distant city from which you started from. It can also be interpreted that countries are generally well connected with each other.

In figure 5 we can see how there is a clear line of high betweenness in the network which spans from Sundsvall (Sweden) to Toulouse (France). This line goes through Helsinki (Finland), Moscow (Russia), Minsk (Belarus) and Paris (France) among other cities. This result makes sense since these cities are the capitals found along the line that Europe makes, so it is likely that if you want to travel from one side of Europe to the other you will pass through one of these cities. We can also estimate these cities to be potential bottlenecks of traffic flow.

It is through this last observation that an error in the dataset was detected, the data treats Brest (Belarus) and Brest (France) as the same city, making it a teleportation device between countries. From experimentation it has been observed that it is likely an error found in the original dataset rather than a misstep handling the data, since in the original data only one "Brest" can be found so the error probably comes from the edges list given. This is a great letdown since this jump between countries greatly modifies the topology of the network, so it is likely that in a correct network the diameter and average shortest path would increase along with the decrease in value of most of the centrality parameters.

5 Conclusions

In this project we have (with a caveat) successfully studied the properties of the European roads network and found the most "important" cities. Through a comparison with null models we have confirmed that the characteristics found in the network weren't due to pure chance. Next we determined that modularity clustering algorithms were better suited to find regions in this network and finally we found the cities in Europe most likely to get high traffic, generally corresponding to capital cities due to their interconnectivity to the rest of the country.

To further explore this topic it would be interesting to analyze a modern dataset that includes the subaquatic tunnels which not only would introduce some countries to the network but also reduce the diameter and average shortest path due to it connecting what were previously distant countries like Denmark and Norway.

It would also be interesting to compare this network to other continent's road networks, Asia and Africa would probably have lower relative diameters due to their rounder geography. We could also analyze other means of transportation like the train stations network or include airports which would probably transform the network into having "small world" properties.

References

[1] NetworkX Developers. Networkx: Network analysis in python. `https://networkx.org/en/`, 2025. Accessed: YYYY-MM-DD.

[2] Netzschleuder network catalogue and repository. Euroroad network (2011). `https://networks.skewed.de/net/euroroad`, 2026. Network dataset from the Netzschleuder catalogue.

[3] Pandas Developers. pandas: Python data analysis library. `https://pandas.pydata.org/`. Accessed: 2026-01-31.

Appendix

Most of the code written was either self made or copied directly from the *Complex Network*'s provided lectures. Of course AI, mainly ChatGPT, was used as a tool to help with some of the most challenging coding sections, these sections are properly highlighted as being AI assisted. All the code written can be found in the following GitHub URL: `https://github.com/SylvainHam/Europe-Roads-Network`.