Analysis of the international E-road network

Irene Valladares Duque
(Dated: July 8, 2024)

Abstract: This study examines the international E-road network, highlighting its low connectivity and minimal clustering due to its spatial layout. The network's degree distribution follows an exponential pattern, similar to a random graph, with comparable clustering coefficients and average nearest neighbor degrees obtained from a random configuration model. A Susceptible-Infectious-Susceptible (SIS) simulation identified a phase transition point, with the experimentally determined critical parameter being about 1.6 times higher than the theoretical prediction. This discrepancy may be due to the network's finite size and the limitations of the heterogeneous mean field approximation. The phase transition indicates that the network is resilient to infections, as controlling the order parameter can mitigate the spread of diseases.

I. DEFINITIONS

We introduce some basic definitions and notation which will be used throughout this report.

The list of degrees per node is called K. Each node has a degree, called k_i. The total number of links is called E. The list of nearest neighbours per node is called FN. It has been computed as an array of 2*E length where each node has assigned a space k_i where the nearest neighbours nodes are stores. Nodes start at 0 for easier array indexing.

The degree distribution, P(k), measures how many nodes have a certain degree. Its cumulative distribution measures how many nodes have the probability of having a degree lesser than a certain one.

The adyacency matrix has matrix elements given by coefficients a_{ij}. The following parameters of the network can expressed in its terms.

The average nearest neighbour degree measures the tendency of nodes of a certain degree to link with peers in terms of degree. It has been computed using the following equation:

$$\overline{k}_{nn}(k) = \sum_{k'} k' P(k' \,|\, k) = \frac{1}{N_k} \sum_{i \in \mathcal{V}(k)} \frac{1}{k_i} \sum_j a_{ij} k_j \tag{1}$$

where $P(k' \,|\, k)$ is the probability of a node having degree k if its neighbour has degree k', $\mathcal{V}(k)$ is the degree distribution of degree k, and N_k is the total number of nodes in degree class k. This quantity needs to be normalized by $\frac{<k>}{<k^2>}$.

The clustering coefficient per degree measures the tendency of the network to form triangles with nodes which belong to a certain degree. Low clustering means a sparse red not knitted together, high clustering means global interconection. This quantity can be computed using the following equation:

$$\overline{c}(k) = \frac{1}{N_k} \sum_{i \in \mathrm{V}(k)} c_i = \frac{1}{k(k-1)N_k} \sum_{i \in \mathrm{V}(k)} 2T_i \tag{2}$$

where T_i is the number of triangles formed with node 'i' of degree k. The clustering coefficient gives the ratio between the number of triangles formed and the highest number of possible triangles which could be formed.

Finally, we briefly explain the configuration model and the SIS simulation.

The configuration model, CM, makes a rewiring of the links keeping the same degree list as in the original network. It creates a random graph. The algorithm has been implemented as follows: A list of nodes is created where each node appears as many times as its degree. Then, this list is shuffled randomly to ensure random connections between nodes. Finally, nodes are paired to form edges. If a self-loop is encountered, the algorithm reestarts until the nodes are correctly shuffled.

The SIS algorithm models a process where each node has probability of being infected λ weighted by the number of infected nearest neighbours and probability δ of being cured, in each step. The time evolution is simulated using Guillespie algorithm and the criticality of the infection process is studied.

II. THE NETWORK

The data set containing the network's edge list can be found in this webpage. This network data represents the International E-road network, which connects countries all across Europe, from Portugal to Russia. These roads are labelled as 'E'. In Spain, our euroroad is called 'E1'. A map is provided in Figure 1.

As for the network, each node corresponds to a city and each link represents a motorway connecting two cities. The network is undirected and unweighted. In Figure is a

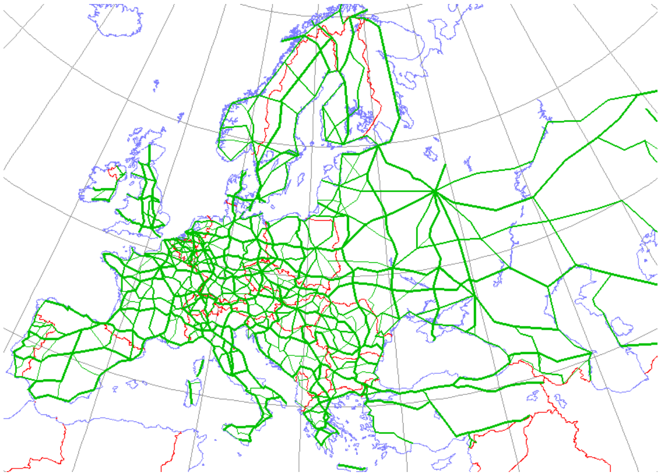

FIG. 1. Geographical map of the E-road system

representation of the network obtained using NetworksX. The sparsity of this network may be attributed to those cities in islands, such as the UK, Cyprus or Corsica.

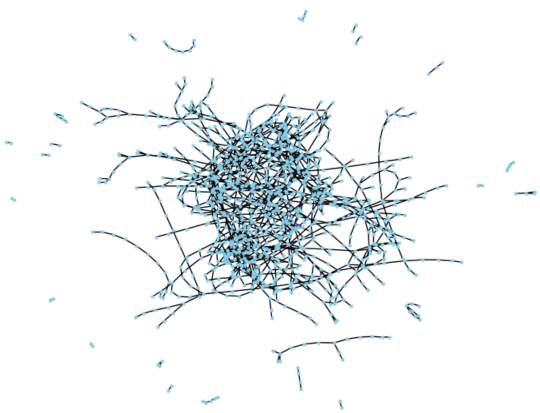

FIG. 2. Network under study. Each node is plotted in blue and each edge in black.

III. TOPOLOGICAL PROPERTIES OF THE NETWORK

During this project, the preferred way of working with the network has been through the edge list as a list of tuples. Firstly, the total number of nodes and links, the degree list and the list of nearest neighbours has been computed. Results are shown in Table I.

This network has low connectivity, as the maximum degree is only 10 and the average degree is close to 2. To see how many nodes have each degree we plot the degree distribution, P(k), and its cumulative distribution, CCCP(k). Results are given in Figure 10.

Number of nodes, N:	1174	**Number of links, E:**	1417
Average degree, $< K >$:	2.41	**Maximum degree, k_m:**	10

TABLE I. Statistics of the network

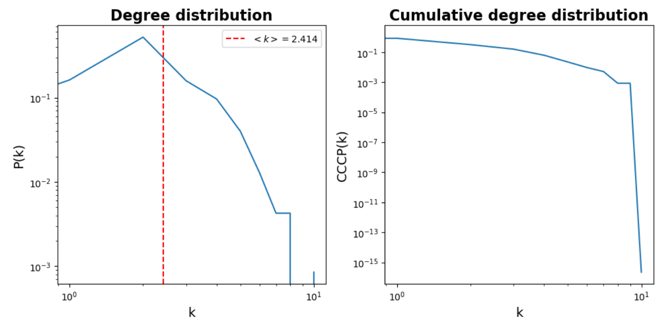

FIG. 3. Normalized degree distribution and its cumulative degree distribution. Taking a city at random, we may expect it to be connected to 2 other cities on average.

The degree distribution follows approximately an exponential distribution. As a consequence, this network does not present largely connected nodes.

Afterwards, a study of its clustering distribution and its average neighbour degree per degree were made. They are represented in Figure 4.

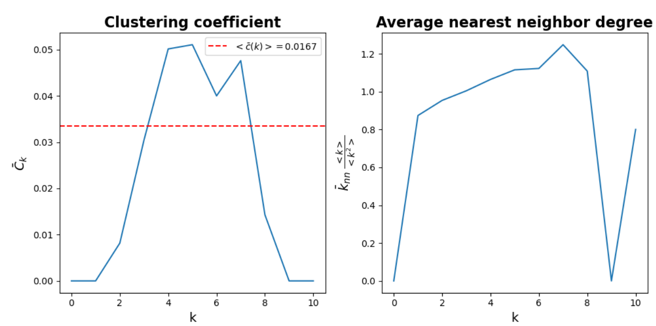

FIG. 4. Clustering coefficient per node degree and average neares neighbour degree per node degree.

Triangles:	12
$\bar{C}$**:**	0.017

TABLE II. Clustering statistics: total number of unique triangles and average clustering coefficient of the network.

As expected, there is low clustering in this network. This is due to the configuration of the roadmaps, as the intention is to connect cities which are far away. In addition, the average nearest neighbour degree is around 1, which coincides with the information obtained from the degree distribution.

IV. CONFIGURATION MODEL

Lastly, the network's conections were changed according to the configuration model algorithm explained explained on Section I. The clustering per degree and average nearest neighbour degree were computed and compared with those of the original network. The results are presented in Figure 5.

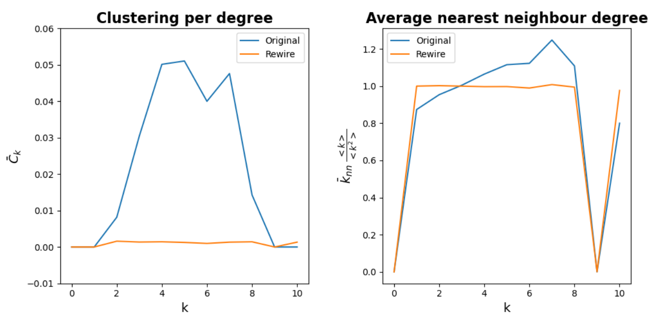

FIG. 5. Clustering per degree and average nearest neighbour degree per node degree comparing the original network and the rewired CM network.

The configuration model eliminates any correlations present in the original network, as it randomizes the nearest neighbour connections. Therefore, the clustering coefficient is lower than the original and the average nearest neighbour degree becomes regular, as they are chosen at random without a preference. This rewiring could join together by a direct road cities such as Madrid with Berlin, which is unrealistic.

V. RESULS OF SIS SIMULATION

According to the heterogeneous mean field aproximation, the critical value for λ order parameter is given by:

$$\lambda_c = \frac{< k >}{< k^2 >} \tag{3}$$

which in this network is λ_c =0,33.

To check whether this network will undergo a phase transition depending on the order parameter, a time evolution simulation has been made using Guillespie algorithm. The results are presented in Figure 6.

In Figure 6 presents the number of infected nodes per time starting from the same random initial configuration, where roughly a half of the total nodes are infected. For larger values than λ_c the pandemic evolves very fast and

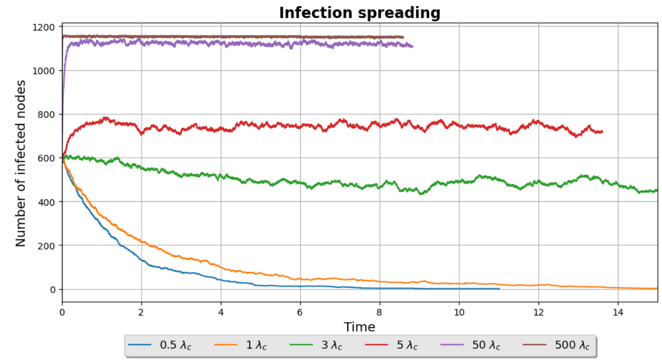

FIG. 6. Time evolution of the infection spreading for different order parameter values.

there appears a stationary plateau of infected nodes, which corresponds to the total number of nodes for $\lambda = 500\lambda_c$. For parameter values under the critical value it can be seen the pandemic stops because all nodes are cured.

In the range of $\{\lambda_c, 2\,\lambda_c\}$ the behaviour changes. At the critical value is expected that the number of infected nodes will wildly fluctuate for a long time, without curing all the nodes. However, for this theoretical value of the order parameter, the pandemic ends very quickly, as can be seen in Figure 7.

A better candidate for the critical point appears to be around 1.5 or 1.6 λ_c, as they do not present a steady plateau of infected nodes. To obtain the real critical value for this network, a plot has been made opposing the averaged plateau of infected nodes with their corresponding value for the order parameter. Results are presented in Figure 8.

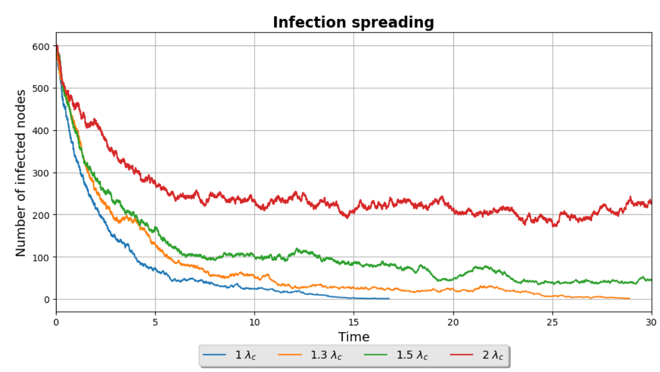

FIG. 7. Time evolution of the infection spreading for different order parameter values around the theoretical critical value. A plateau of infected nodes is obtained for $2\,\lambda_c$.

As expected, the critical value should be around $1.6\,\lambda_c$. In Figure 8 the point corresponding to this value appears to be displaced from its expected place following

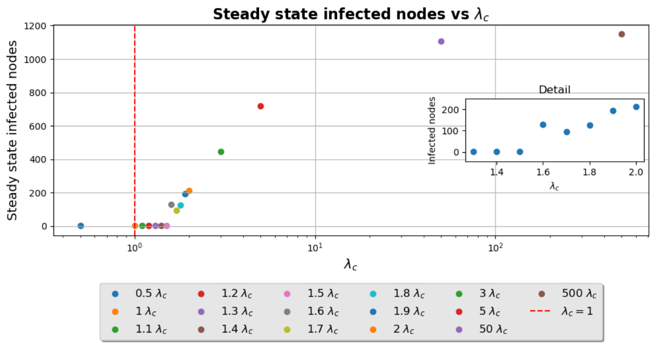

FIG. 8. Averaged final number of infected nodes per value of the order parameter. The red line represents the theoretical value of the order parameter. A detail around the range of $\{\lambda_c, 2\,\lambda_c\}$ is being presented inside the plot as well.

the curve. This may be due to statistical fluctuations. To test them, the algorithm was repeated 200 times starting from the same initial configuration and letting Guillespie algorithm run a large, fixed number of steps. The results are shown in Figure 9 along the results for the rewired CM network.

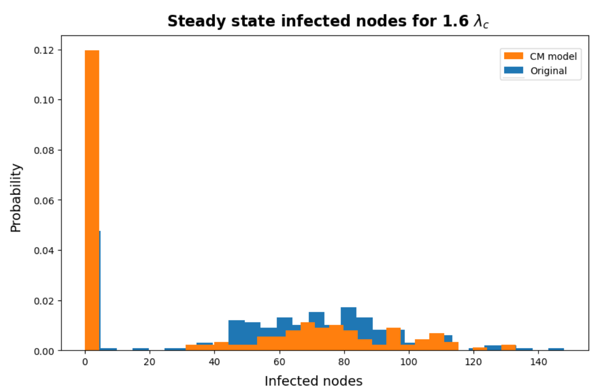

FIG. 9. Distribution of infected nodes for $1.6\,\lambda$ after large number of repetitions for the real network and the CM rewiring. Data are normalized.

The results show that the final number of infected nodes have more probability to be zero, therefore, the pandemic ends (if the number of infected nodes reach 0, the infected arrays storage 1 as their last value). However, there is a probability that the pandemic arrives at a plateau of a hundred of infected nodes. Endlessly fluctuating behaviour is characteristic of criticality, but as the system is finite, it may explain why sometimes fluctuations are suppresed and all nodes can be cured.

Finally, to test whether this experimental critical value is characteristic of this network or it may appear in a random graph with the same degree distribution, a CM rewiring has been applied and the SIS algorithm has been used to simulate its infection dynamics. Following the same steps and averages as in the normal network, the final infected nodes have been plotted against their parameter value. Results are shown in Figure 10.

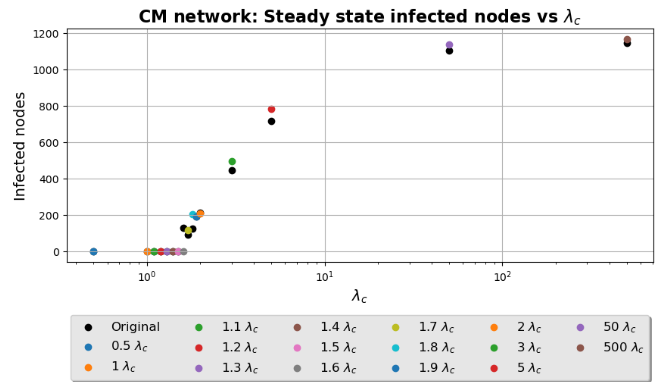

FIG. 10. Averaged final number of infected nodes per value of the order parameter in the rewired CM network. Dots in black represent the curve depicted in Figure 7.

A similar critical curve has been obtained for the normal network and the rewired. This suggest the critical value only depends on degree distribution and not on the nearest neighbour arrangement. However, a significant difference can be found in the fluctuations around the critical point, shown in Figure 9. There is a higher probability for the network to become uninfected, even though there still appear fluctuations on the final number of infected nodes. This may be due to the rewiring flattening the average nearest neighbour degree, which would make easier for the pandemic to spread uniformingly.

VI. CONCUSIONS

In this report we have carried out an analysis of the international E-road network. Its topological features have been extracted and compared with those of a random configuration model network. This network has low connectivity, as the average degree is small, and has virtually no clustering. These features are given by the spatial arrangement of motorways across Europe. In addition, its degree distribution follows an exponential distribution, so its statistics are very similar to those of a random graph. Indeed, the rewiring using the configuration model provide results of a similar order of magnitude in the clustering coefficient and average nearest neighbour degree, although lower.

Afterwards, a SIS simulation was made and plots were made to locate the exact parameter in which a phase transition occurs. The heterogeneous mean field aproximation predicts a critical value for the order parameter which is much smaller than the obtained experimentally, which is around 1.6 times higher. The reason of this discordance between theoretical and experimental results may lay in the finite size of the system as well as on the non satisfaction of the coarse graining hypothesis, in which the heterogeneous mean field aproximation bases its results. Further studies should be made to better locate the exact value for the critical order parameter, perhaps analyzing the fluctuations on infected nodes around the critical point or developing an analytical framework which would better describe its dynamics.

The phase transition makes this network resilient to infections, as the control of the order parameter allows slowing down the pandemic state.