ANALYSIS OF THE EUROPEAN INTERNATIONAL RAILWAY NETWORK AND PASSENGER TRANSFERS

L. Calzada-Infante[a], B. Adenso-Díaz[b*], S. García[c]

a. Business School, University of Leon, Spain,
email: lcali@unileon.es

b. Engineering School, University of Oviedo at Gijón, Spain,
email: adenso@uniovi.es

c. Engineering School, University of Oviedo at Gijón, Spain,
email: sgarcia@uniovi.es

* Corresponding author. B. Adenso-Díaz, Edificio Energía, Engineering School, University of Oviedo, 33204-Gijon Spain. adenso@uniovi.es

Publicado en: *Chaos, Solitons & Fractals*, Volume 141, December 2020, 110357

ABSTRACT.

Railway transportation is one of the most popular and greenest transportation modes for passengers. Its importance has increased in some regions owing to the deployment of high-speed infrastructures. In Europe, it is a competitive transportation mode for short and medium distance journeys, rivalling airway mode.
For topological analysis of transportation networks, Complex Networks Analysis (CNA) appears as a powerful methodology that although used in various circumstances to describe national railway networks, it has not been used thus far at a continental level. In this paper, two topological characterisations of the European International Railway Network are performed using CNA. The first analyses the direct connection among cities in the international railway service using the most commonly used metrics. The results are compared with those of the Chinese Railway Network, of similar size, observing their differences regarding assortativity. The second analysis incorporates passenger transfers between services, discussing how connectivity is improved when timetables are synchronised, remarking the importance of such coordination. Centrality metrics are defined for the assessment of the connectivity of the network. For the sake of realism, thresholds for the maximum reasonable distance to be travelled by train are introduced in the definition of the centrality metrics.

KEYWORDS: Complex Network Analysis; Railways; European transportation network

1. INTRODUCTION

Transport infrastructure plays a key role in the economic growth of a country. Both concepts, transportation and economic development, have an endogenous relationship as good transportation infrastructures induce spatial redistribution of economic activities, with wealthy economic areas requiring constant improvement of transportation conditions (Hong *et al.* 2011). A positive correlation between both variables is found in several studies (Berechman *et al.* 2006). Therefore, the accessibility of the cities is a key factor for their competitive position. This fact especially affects various businesses, significantly the tourism market (Borodako and Rudnicki, 2014).

With regard to railway transportation, Zhang *et al.* (2015) view it as a critical infrastructure, whose services and facilities are essential for the basic operations of society. Heinold and Meisel (2018) highlight its importance from the environmental perspective, given that intermodal road transportation reduces greenhouse emissions when compared to the road-only alternative.

In railway transportation of passengers, it is possible to classify trains based on the regions served and the achievable speed. Specifically high-speed trains can be observed as competitors and complement to air transportation (Adler *et al.* 2002). The viability of investing in these trains is justified in respect of the maximisation of social welfare (Adler *et al.* 2002). This evaluation assessed the costs of upgrading to high-speed infrastructure and the effect on high-speed rail, regional low-cost carriers and airline transport operator's actions using a game theory approach. Slight changes in trains classified as "high hierarchy" as with international trains, are influential on the lower ones, in particular regional trains (Landex, 2012). This interrelation is analysed in the Danish network by altering the queuing time of a railway line, observing its effects in small regions affecting the underuse capacity of the transportation system (Hansen *et al.* 2006).

As a consequence of the relevance of the transportation system for the economy, society and the environment, studies have been conducted using different methodologies. One is the Complex Network Analysis (CNA). This technique represents by links the interaction of the entities of the system (the nodes in the network model, usually cities or stations when dealing with transportation studies); a topological analysis of the network structure is performed, identifying vulnerable nodes, connectivity problems or clusters of explicitly related entities. It has been applied to multiple means of transport in various countries, namely air transportation (Barrat *et al.* 2012), underground (Lu 2018; Sun *et al.* 2017), bus (Yue *et al.* 2019), railways (Wang *et al.* 2020), or maritime (Hu 2019).

To our knowledge, no topological characterisation of international railway networks at a whole continental level (and specifically of the European International Railway Network) has been conducted, although railway structures have been analysed in various countries regarding its topological characterisation. The main contribution of this paper is to perform a topological characterisation of the complete European International Railway network, not just evaluating all the connections, but considering reasonable journeys that international railway passengers require, taking distances into account. These analyses describe the network properties of the international services, enabling comparison with an important international case, the Chinese one, ascertaining the main differences in their structure. Additionally, accounting for current timetable services, structure alterations when passenger transfers are assessed.

With that purpose, Section 2 gathers all the applications of CNA on the description of Railway Networks. Section 3 details the CNA methodology, introducing the main concepts used in our analysis. In Section 4 two networks are evaluated, the first represents the connections among stations using direct trains only (which is then compared with the Chinese Railway Network), while the second assesses how the connectivity of the network is improved when considering passenger transfers with a short waiting time. In both networks, centrality metrics identifies key stations constraining the feasible distance that passengers can travel. The last section concludes.

2. PREVIOUS RAILWAY NETWORKS CNA

The results using CNA tools are multiple in the case of railway transportation, with many recent studies at country level in various continents. Tsiotas (2017) evaluates the contribution of the Greek interregional railway network to the regional development based on socioeconomic information. He finds that non-metropolitan areas are focussed on primary productivity sector and on providing transportation services to tourism. In the Spanish case, Roanes-Lozano *et al.* (2009) evaluate the flexibility reduction of the country's railway network by cutting the deficit lines from 1956-2006, providing as a result the current radial Spanish Railway network. Yue *et al.* (2019) analyse whether intercity transportation supports the development and integration of urban agglomerations by a method called transportation cluster detection. This methodology is an iterative process that adopts hierarchical clustering, applying as criterion the proximity index based on k-shortest path. In each iteration, the quality of the cluster is controlled by geo-modularity.

For its part, Bhatia *et al.* (2015) evaluate the performance of recovering strategies in the Indian Railway Network with a CNA topological characterisation. At a multi-country level, Kurant and Thiran (2006) extract the network of traffic flow and the infrastructure network of three different railway systems in Centre and Eastern Europe from published timetables. They use these data to compare different approaches to construct networks using CNA topological characterisation.

CNA allows for understanding the structure of the complex systems based on their topological characterisation. This can be achieved in multiple ways, and depending on the definition of the network and its topology, different properties can be observed. One of the most favoured is the small-world property. It can be found in many railway networks, such as Boston and Viena Railway Networks as published by Seaton and Hackett (2004) using a bipartite model; in China (Li and Cai 2007; Wang *et al.* 2020); in Central Japan (Majima *et al.* 2007); or India (Sen *et al.* 2003).

Small-world is also observed in Pakistan (Mohmand and Wang, 2014), in the Chinese Railway Network modelled by the P-space representation (Cao *et al.* 2019). In the case of applying L-space graphs to the Greek railway, its topology appeared to share lattice-like characteristics (Tsiotas 2017).

A primary tool in CNA is the degree distribution that results as useful to observe the hub stations of the networks. Fitting this distribution can provide various results. For instance, But and Prokhorchenko (2013) find a power-law distribution over the Ukraine Freight Railway Transportation network, as Li and Cai (2007) find it in the Chinese passengers railway network. However, Sen *et al.* (2003) find an exponentially decaying distribution in the fitting for the Indian Railway network.

The relative position of the stations is achieved in CNA by using centrality metrics being the most common Degree, Betweenness and Closeness centrality. For instance, But and Prokhorchenko (2013) apply centrality metrics to a direct assortative freight railway transportation network of Ukraine, and use the closeness to detect the stations clusters; while Zhang *et al.* (2015) apply centrality metrics to the Chinese network to evaluate the importance of train stations using a modified gravity model based on population and gross domestic products of each city to weight the railway lines.

To better understand the connectivity among stations, the nodes can be grouped attending to different criteria. For example, Bhatia *et al.* (2015) use the Lovain community detection algorithm for the Indian Railway Network to detect those nodes that were connected among them more than in a random network. Regarding the Ukraine Railway Network, But and Prokhorchenko (2013) discovered three clusters based on the distribution of closeness centrality metric.

Another interesting property in CNA studies dealing with transportation is the assortativity of the network that assesses whether high connected nodes tend to be connected among them. In the case of the railway system of Singapore, the topological centralities were evaluated from a dynamical perspective during weekdays and weekends and a soft disassortative network appears in both periods (Sho *et al.* 2010).

3. METHODOLOGY

3.1 Data collection

For this analysis, the European International Railway Network (EIRN), a weighted directed network, $G(N,E)$, representing the flow of trains along the European railway transportation infrastructure was built. The dataset of all the International Railway Services in Europe used in this paper is collected (Potter 2018), providing all the services available on May 2018. The information has been transcribed, recording the timetables and frequency of each service.

The nodes N represent all the European cities connected by international train services, and each directed link in E represents the existence of at least one feasible journey between two stations without any transfers (that is, P-space representation, see Lin and Ban 2003). The links are weighted based on the frequency of services during a week. A squared matrix, whose size is the number of cities, collects the frequency of the services for the weighted evaluation, while a second matrix records the timetable of each train for each visited city. All data were processed using igraph (Csardi and Nepusz 2006) and poweRlaw (Gillespie 2014) packages of the open source statistical software R.

3.2. Metrics for topological characterisation

The topological characterisation of EIRN can be made at different levels, from a general description of the network to the evaluation of the relative position of each node in the network. Here we introduce all the metrics discussed later in the manuscript for a clearer understanding of the results.

Different metrics can be used over the weighted or the corresponding unweighted network, considering the *geodesic paths* among the nodes (i.e., the shortest path between each pair of nodes). Lin and Ban (2013) present some of these metrics in the context of transportation and city planning.

From a general perspective, some metrics such as diameter, radius and average path length provide useful information. The *diameter* measures the maximum distance of the geodesic paths (i.e., maximum number of trains necessary to use between any two cities), while the *radius* measures the minimum distance of the maximum geodesic paths from any node, thus identifying the most central node; the *average path length* assesses the mean of all the geodesic paths in the network.

Other metrics do not depend on the geodesic paths, as occurs with the concept of degree distribution. The *degree* of the node i, k_i, represents the number of cities that passengers can travel to directly, leaving from city i. As EIRN is a directed network, we can distinguish the in-degree, k_i^{in}, and out-degree, k_i^{out}, the in-degree being the number of cities from which it is possible to travel directly to city i, and the out-degree, the number of cities that passengers could travel to directly from city i.

The distribution of the degree is one of the main tools to identify the network topology and the organisation of the network attending to the connectivity of the stations. The main degree distributions are Poisson, scale free or exponential. Poisson distributions are found in random networks, while power-law distributions are common in networks with preferential attachment like scale-free networks. In the context of transportation, scale-free networks denote the existence of highly connected hubs serving as bridges for a majority of poorly connected nodes.

While degree distribution shows how the hubs are distributed in the network and how the number of connections increase, the *average degree of the nearest neighbours* allows to observe whether these hubs are connected to high connected nodes, defining an assortative network or disassortative in the other way round.

Additionally, when the network is weighted, one of the most used metrics is the *strength*, that is, the sum of all the weights of the edges linked to a node. As in the case of the degree, the strength distribution can be analysed, fitting it to different distributions by visualising the probability of a node to have a certain strength.

Apart from evaluating the direct connections of the nodes and the average degree of their neighbours, it is interesting to know how their neighbours are connected among them. For this purpose, the *clustering coefficient* measures the density of the connections among the neighbours of a node. Here, the weighed and unweighted clustering coefficient of Barrat *et al.* (2012) is applied to an undirected subgraph to avoid flow imbalance. To evaluate EIRN as an undirected network, all the directed links are transformed into undirected to avoid duplicity. In this case, the resulted weight of an arrow is the mean of the directed links that connect each pair of nodes. The purpose of this metric is to analyse the triplets: if the cluster coefficient of the weighted network is larger than for the unweighted version, which means that the more weighted links are likely to form triplets.

It is relevant in this context that the concept of *Small World Network* (SWN) is where any two nodes are expected to be not too far. These structures are specifically defined by having an average shortest path length between any two nodes that grows more slowly than the size of the network ($L \propto \log N$) and in addition, its clustering coefficient is higher than expected by random choice. In these structures it is natural that the presence of hubs serves as bridges among many nodes, and although SWN does not imply a power-law degree distribution, it holds that scale-free networks show the *ultra SWN* property ($L \propto \log \log N$).

To analyse groups of stations, the *community* structure of the network was defined using the algorithm from (Blondel *et al.* 2008). As usual, the group of stations that belong to the same community have more links among them than those that would be expected in a random network of the same size.

To evaluate the relative position of a node in the network, centrality measures such as *Closeness and Betweenness centrality*, are useful. Closeness centrality for node v is calculated by [1], being $d(v,i)$ the distance of the geodesic path between the nodes v and i. This metric is normalised by n-1, being n the number of nodes in the evaluated network. The purpose of this centrality metric is to evaluate how far the G (N,E) evaluated city is from the other cities. In the case of Betweenness centrality, the purpose of this metric is to define how likely the evaluated city is in the shortest path between pairs of stations. It is calculated by [2], being $\sigma(j,k,i)$ the number of shortest paths from j to i passing through k. High values show the criticality of that node from a security point of view. Again, this metric is normalised by $(n-2)(n-1)/2$. In each evaluated network, both centrality metrics are calculated as the average of the centrality values of all the nodes.

$$C(v) = 1/\sum_{i\neq v} d(v,i) \qquad [1]$$

$$B(v) = \sum_{i\neq v\neq j} \sigma(j,v,i) \,/ \sum_{i\neq j} \sigma(j,*,i) \qquad [2]$$

However, in the specific case of international railway transportation, nodes requiring a longer journey are not likely to be connected by train, as cheap and efficient transportation alternatives exist (namely by air). Therefore, in the calculation of centrality metrics, instead of considering the geodesic paths among every pair of cities in the network, the most accurate way would be to consider only those paths among cities that are likely to reached by train.

In this paper, a new approach is proposed in the evaluation of railway centrality metrics, introducing multiple distance thresholds representing various scenarios, with the real network that railway passengers require.. For each city i, these distance-based metrics are evaluated on the subgraph S_i (V_i,L_i) $\subseteq$ G (N,E) composed of all the stations located closer than a threshold distance H. That is, $V_i = \{j \in N \colon \|g(j) - g(i)\| < H\}$, being $g(\cdot)$ the geographic coordinates of a city. Considering the size of the subgraph associated to the value H, the metrics are normalised as mentioned. Therefore, the closeness Centrality is calculated for each station by [3], while the Betweenness is calculated by [4]

$$C(v) = \frac{1}{\sum_{i \neq v} d(v,i)} \quad \forall v \in N, \forall i \in S_v(V_v, L_v), \tag{3}$$

$$B(v) = \frac{\sum_{i \neq v \neq j} \sigma(j,v,i)}{\sum_{i \neq j} \sigma(j,*,i)} \quad \forall v \in N, \forall i,j \in S_v(V_v, L_v). \tag{4}$$

4. RESULTS

4.1 Topological description of EIRN

EIRN is composed of 412 nodes and 7732 edges in a single component and, as expected in a physical network, its density is low (4.56%).

Starting by considering this network as ***unweighted*** (i.e., the frequency of trains between two stations is set to one, and therefore the geodesic distance represents the number of trains required to travel between two cities), the network diameter is 7, while the radius is 4 and its average path length is 2.89. Considering all the geodesic paths, the most probable shortest path between two cities is 3. This means that the most probable number of transfers while on an international journey between any two stations in Europe would be 2, with a maximum of 6 transfers in the worst scenario. The average path length is not within the range between the radius and the diameter due to the topology of the network, which is not a tree.

It is important to highlight that the average path length of a random unweighted network with the same density of nodes is 2.36, which is lower than in EIRN. This fact is a consequence of the geographical and logistics limitations of EIRN. For example, it is not feasible to have an edge between Denmark and Spain without defining an international route through Germany and France at least.

As observed previously in many railway networks, it is expected that the degree distribution of EIRN (Figure 1) follows a power-law $P(>k) = k^{-\gamma}$. Using the method described in Clauset *et al.* (2009) the optimal lower cut-off in this case is 73 with γ=4.11. Applying a goodness-of-fit test via bootstrapping the power-law distribution cannot be ruled out. Being the value of γ larger than 3, the diameter of the EIRN, whose value is 7 as mentioned above, was expected to be close to ln N (Cohen and Havlin 2010).

--Figure 1--

--Figure 2--

EIRN can be split into 9 communities following the algorithm of Blondel *et al.* (2008) as shown in Figure 2, where a clear geographical pattern can be observed. By definition, in each community the connections are denser than in cities not in the community (as well as in a random network).

The average in-degree or out-degree of EIRN is 18.77, being the average number of stations that can be reached directly, having their origin or destination in a certain city. Paris is the city with maximum in-degree and out-degree of the network (with 113 and 112 European cities reachable respectively).

As observed in Figure 3, there is a strong positive correlation (0.98) between the in-degrees and the out-degrees. The cities reach and are reached by a similar number of cities, and there is balance traffic flow among them. Notable exceptions are Nice with an out-degree of 18 and in-degree of 45, and Haute Picardie with an out-degree of 23 and in-degree 34.

--Figure 3--

Another point of view to analyse degree is its correlation with the average degree of the nearest neighbours. Figure 4 shows a positive correlation between both variables (Pearson value of 0.29±0.09), making EIRN an assortative network, therefore, the higher the degree of a node, the higher the mean degree of its neighbours. This means that the most connected cities tend to be connected to well connected cities.

--Figure 4--

Apart from evaluating the raw topology of the connections, the intensity of the connections can be assessed considering EIRN as a ***weighted*** network. The weekly frequency (w_{ij}) of the international trains between city i and j can be used to rate the connectivity among the stations as in (Wang *et al.* 2020), and its reciprocal represents how close they are, being used to calculate the network geodesic paths among cities. In other transportation networks such as air traffic, it is common to use the number of available seats in flights among the connected airports instead of the frequency of the services (Barrat *et al.* 2012).

As mentioned, one of the basic centrality metrics in weighted networks is the *strength* of a node, defined by $S_i(w) = \sum_j w_{ij}$. The strength distribution is important to characterise the topology of the network. Figure 5 shows how the cumulative distributions of the strength fit to a power-law with an optimal lower bound of 1,267 and $\gamma = 2.73$. This distribution cannot be ruled out after applying the Kolmogorov-Smirnoff test.

---Figure 5---

In the weighted EIRN, the γ value for strength distribution is lower than for the degree distribution. The difference between both distributions is due to different scales in both metrics. It is important to highlight that the nodes with the maximum degree and strength are not necessarily the same. In EIRN, the node with the maximum degree is Paris, while the node with the maximum strength is Frankfurt. This means that in a period of one week, more trains pass through Frankfurt than Paris, although the latter connects directly with more stations.

It is expected that strength increases with degree with a non-lineal dependency. Barrat *et al.* (2012) affirm that the average strength of the nodes with degree k follow an exponential relationship $S(k) \sim k^{\beta}$, and only in the case that the weight of a node is independent of its degree then β=1. EIRN follows a power law behaviour with an exponent equal to $\beta = 1.20 \pm 0.07$. Consequently, the value of the degree of vertices grows more slowly than their strength.

The nonlinear dependency between the weight and the degree of the end-points of the link is given by $W_{ij} \sim (k_i k_j)^{\alpha}$ with $\alpha = 0.07 \pm 0.02$. The value is this low due to the high number of connections with daily frequency (i.e, with $w_{ij} = 7$). Although α value is lower than in other networks like the World-wide Airport Network (Barrat *et al.* 2012) whose α is 0.5, the correlation in our case is still positive (p-value $\approx$ 0).

The average clustering coefficient is the mean clustering of all the nodes of the network. In this case it is 0.81 (0.77 for the unweighted EIRN). This means that the density of the connections among the neighbours of a city tend to be higher than expected (the random generated network has a clustering of 0.09). As mentioned above, this relationship among the clustering coefficients implies that the links with high frequency tend to form triplets.

--Figure 6--

The correlation between the clustering and degree is negative (see Figure 6). Studying the unweighted clustering, in the nonlinear dependency $C(k) \sim k^{-\eta}$, it is η= 0.27 ± 0.03. Comparing the weighted and unweighted clustering, the influence of the weight for the hubs' cities can be observed. One city has a higher clustering coefficient when it has a higher number of connections with cities that are connected to them by a direct train.

Distanced-based centrality measures

To know how well the network connects the European cities, centrality metrics are an interesting tool to assess it. As previously mentioned, Closeness centrality measures how close the nodes are in the graph based on their geodesic paths, while Betweenness centrality measures the ratio of times that a station appears in the intermediate geodesic paths. In the first case, high centrality implies a good connection between any two cities, while the second case shows the importance of a city as an intermediate station.

Note that when travelling by train, journeys lasting longer than a specified time are deemed unappealing as other more efficient travel modes exist (namely air). Due to that and for the sake of realism, the various distance-based metrics applied here must consider only feasible international railway trips under a certain threshold, assuming unreachable cities to be further than that distance limit.

Table 1 shows the results of these distance-based metrics for different thresholds. Naturally the results indicate that the size of the network increases with the threshold, and therefore the mean degree increases, the centrality measures reduce as more cities become involved, and the radius and diameter increase. For thresholds over 2000 km the distribution of the geodesic path is close to the whole network (threshold=∞). It is interesting to observe how the increment of the diameter and the radius is not proportional; indeed, its ratio is decreasing from 2.7 for the 500 km threshold, up to 1.75 in the case of the whole network. This relationship is not observed when comparing the diameter to the average path length.

---Table 1---

Moreover, most of the distances of the geodesic paths are in a small range of values. This fact influences the closeness centrality directly as it depends on the distance of the geodesic paths. Closeness is quite high for short distances of around 500 km, but for further than 1000 km it drops sharply and maintains similar values that are achieved for the whole network.

4.2 Comparison between EIRN and China railway network

To compare EIRN with another similar network, the China Railway Network (CRN) case could be of interest. The area of Europe is 10.18 million km^2, while China is 9.597 million km^2, although with a much greater population and therefore with many very populated cities, which must affect the transportation networks' characteristics.

CRN was topologically described by Wang *et al.* (2020) and a summary of its comparison with EIRN appears in Table 2, including some z-scores as defined by Zanin et al. (2018) for the sake of a more accurate analysis. In the case of CRN, it is represented as a weighted directed network, which has 255 nodes and 1,091 edges.

---Table 2---

The density of EIRN (4.56%) is higher than CRN (3.37%) in spite of the European network having more stations. Baring this, the average path length of EIRN is slightly higher than in CRN. In EIRN case, when the average path length is compared with random networks of the same size, the average path length is higher than for random networks (z-score 371.87 vs. -20.17 for CRN).

In both railway networks there is the same difference of scale between the degree and strength distribution by observing the γ values of the power law distributions. The γ of the strength distribution (1.73 for CRN and 2.73 for EIRN) are lower than that of the degree distribution (1.83 and 4.11 resp.). In the case of China, its maximum degree is 109, while its maximum strength is 968 (Beijing). Evaluating the relationship between in-degree and out-degree, both EIRN and CRN have a positive correlation ($\approx$1) with almost the same form. This is a signal that the traffic flow is balanced in both ways.

Regarding the correlation between degree and strength, in CRN the strength grows faster than for EIRN (1.30 vs. 1.20) following the non-linear dependency $S(k)\sim k^{\beta}$. Alternatively, the non-linear dependence of the clustering and degree in CRN is very similar to EIRN (0.36 vs. 0.27), but in the Chinese case the clustering decreases faster than in the European one.

EIRN and CRN differ regarding the correlation between average nearest neighbours and degree. While EIRN is assortative (+0.29), CRN is disassortative (-0.17). Moreover, the average clustering coefficient of CRN is 0.49, therefore the connections among the neighbours are denser in the case of EIRN (z-score 714.90) than in CRN (121.83).

We observe that although both networks have similar density, the diameter, average path length and clustering are lower in the Chinese case (see z-scores in Table 2). There are notable differences between both networks, such as their hub distributions, the EIRN tendency to connect hubs with high connected stations (which is not observed in the Chinese case), and how CRN clustering decreases faster with the degree (η=0.36) than for the EIRN. The Chinese and European demography could explain these differences.

4.3 EIRN with passenger transfers (EIRN-t)

So far, as carried out by most previous researches, we have considered the network where links represent cities connected by direct trains, with no transfer of passengers among trains allowed. However, in the real world it is possible to travel by train between two cities, although no direct train exists between both stations as long as the timetables allow the passengers to transfer to an intermediate station.

In this section, a new EIRN is evaluated that considers that two cities are connected if the passengers can transfer to an intermediate station, when the waiting time is less than the time limit (one hour in our case). Therefore, a link exists if there is a direct train between two stations or if two stations are connected by two trains, the time differential between the departing and arrival time at an intermediate station being less than one hour. Note that although other authors have already considered transfers when analysing railway structures (Sen *et al.* 2003) consideration was not given to the real possibility of a transfer by checking the timetables as done here, but only the existence of intermediate stations connecting various lines.

The new network, called EIRN-t, is composed of the same nodes and links as EIRN, but with some new 270,891 links (an additional 350%) owing to the transfers. This significant increment of connections is due to good timetable synchronisation on many routes.

In the case of EIRN-t the density is 20.6% (more than 5 times the EIRN density). The diameter for the unweighted network has decreased to 5 and its radius to 3, while in the original EIRN 6 transfers, as a maximum, were required to be achieved in any city of the network. As a consequence of the high density, the average path length for the unweighted network has decreased to 2.0, and the most probable shortest path is 2 (see Table 2).

Regarding the degree, its cumulative distribution can be observed in Figure 7, applying the Kolmogorov-Smirnoff test to power-law, log-normal and exponential distribution. In the case of power-law distribution, the lower bound is 340 and $\gamma = 6.10$, and the goodness-of-fit via bootstrapping does not rule out the power-law distribution. This significant increment of the γ value indicates that there are fewer hubs with a high number of connections, and a low probability of having super hubs.

---Figure 7---

Studying the weight distribution, the power-law distribution cannot be ruled out by the method described in Clauset *et al.* (2009). Figure 7 shows the power-law, log-normal and exponential distributions fitting to the cumulative distribution of the strength using the Kolmogorov-Smirnoff test. The power-law distribution has a lower bound of 351.6 and $\gamma = 3.06$.

Because of the transfers, the correlation between the in- and out- degree has decreased to 0.96. In this case, the maximum in-degree is 127 in Frankfurt and the maximum out-degree is 412 in Zurich. Moreover, the strength grows faster with the degree than for the original EIRN with $\beta = 1.42 \pm 0.06$, while the correlation between the frequency weight and the degree of the end-points of the links is higher with $\alpha = 0.43 \pm 0.02$. The correlation between the unweighted average nearest neighbours and the degree is higher with a value of 0.38, which means that EIRN-t is more assertive than EIRN and therefore high connected nodes tend to be connected among them.

Baring this, the unweighted average clustering is 0.74 (z-score 564.67, see Table 2), lower than for EIRN (714.90), and much higher than for a random network. The same happens for the weighted average clustering. This means that there are fewer triplets. Indeed, the unweighted clustering of EIRN-t grows more slowly than for the original EIRN, $\eta = 0.17 \pm 0.02$.

---Table 3--

Comparing the results of the distanced-based metrics in EIRN-t and EIRN in Table 1 and 3, the behaviour of closeness and betweenness centrality is fairly similar, although EIRN-t is much denser than EIRN. An important difference among both networks is the higher values of closeness centrality for short distances (under 1000 km). The low values of betweenness centrality occur as mentioned due to the definition of the network based on route corridors and the existence of links between all the cities in a corridor. As the degree of a node increases, owing to the transfers, the importance of the intermediate stations reduces and the betweenness is even smaller than in the EIRN case.

5. CONCLUSIONS

We have investigated the topological properties of the European International Railway Network, without (EIRN) and with (EIRN-t) possible transfers assuming a maximum waiting time in the intermediate station of one hour. The EIRN behaves as a small world network, although the average path length is higher than the corresponding to a random network since the international routes are geographically constrained. This confirms the hypothesis of Sen *et al.* (2003) who stated that in their opinion SWNs should be expected when analysing the railway structure of any country.

Moreover, the EIRN has been compared with the China Railway Network (CRN), observing their similarities and their differences. In both cases, clustering is higher than for random networks, and both networks have a balanced traffic flow in both ways. In CRN, the strength of the cities grows faster with the degree than in the EIRN case, while the clustering coefficient decreases faster. Among all these similarities, the main difference is how EIRN cities with high degree tend to be connected each other, while in CRN they do not.

Comparing EIRN and EIRN-t, the connectivity when there are transfers is much higher due to its high density, its high mean degree, and the strength growing faster with the degree. It is remarkable how in EIRN-t the number of potential cities to travel to and from well connected cities explodes when passengers transfer because of the higher assortativity, and how the average path length its higher compared with random networks of the same size. This last feature is also observed in EIRN, given that the train routes are defined by corridors.

Although the connectivity is higher in EIRN-t, its unweighted clustering coefficient is lower, whilst the weighted clustering decrease slower with the degree. This means that the number of triplets is lower (i.e., the number of groups of three cities connected among them is smaller).

For the sake of a more accurate evaluation of the centrality metrics, a new approach limiting the reachability to a maximum geographic distance was defined. Varying the threshold of the distance that passengers are willing to travel by railway, the relative position of the cities can be evaluated considering their closest cities. It is significant how the Closeness and Betweenness centrality metrics do not differ between EIRN and EIRN-greatly. This means that the relative positions of the cities in the network on average do not change after considering transfers.

In this paper we have seen the potential to use passenger transfer to change the centrality of the cities. However, this is only possible when the timetables are synchronised to prevent a lengthy passenger wait at the station. As a future research, it would be interesting to see how the connectivity of the European cities could be improved by a better redesign of the current railway timetables that optimise the synchronisation.

Acknowledgments. This research was carried out with the financial support of the Spanish Ministry of Science and the European Regional Development Fund (ERDF), grant DPI2017-85343-P.

REFERENCES

Adler, N., Friedman, L., & Sinuany-Stern, Z. (2002). Review of ranking methods in the data envelopment analysis context. *European Journal of Operational Research*, *140*(2), 249–265. doi:10.1016/S0377-2217(02)00068-1

Barrat, A., Barthélemy, M., & Vespignani, A. (2012). The Architecture of Complex Weighted Networks: Measurements and Models, 67–92. doi:10.1142/9789812771681_0005

Berechman, J., Ozmen, D., & Ozbay, K. (2006). Empirical analysis of transportation investment and economic development at state, county and municipality levels. *Transportation*, *33*, 537–551. doi:10.1007/s11116-006-8472-2

Bhatia, U., Kumar, D., Kodra, E., & Ganguly, A. R. (2015). Network science based quantification of resilience demonstrated on the indian railways network. *PLoS ONE*, *10*(11), 1–17. doi:10.1371/journal.pone.0141890

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, P10008. doi:10.1088/1742-5468/2008/10/P10008

Borodako, K., & Rudnicki, M. (2014). Transport Accessibility in Business Travel - a Case Study of Central and East European Cities. *International Journal of Tourism Research*, *16*, 137–145. doi:10.1002/jtr

But, T., & Prokhorchenko, A. (2013). Investigation into Train Flow System on Ukraine's Railways with Methods of Complex Network Analysis. *American Journal of Industrial Engineering*, *1*(3), 41–45. doi:10.12691/ajie-1-3-1

Cao, W., Feng, X., Jia, J., & Zhang, H. (2019). Characterizing the Structure of the Railway Network in China: A Complex Weighted Network Approach. *Journal of Advanced Transportation*, *2019*, 1–10. doi:10.1155/2019/3928260

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. *SIAM Review*, *51*(4), 661–703. doi:10.1137/070710111

Cohen, R., & Havlin, S. (2010). *Complex networks: Structure, robustness and function*. *Complex Networks: Structure, Robustness and Function*. Cambridge (United Kingdom). doi:10.1017/CBO9780511780356

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, *1695*.

Gillespie, C. S. (2014). Fitting heavy tailed distributions: the poweRlaw package. *Journal of Statistical Software*, *64*(2), 1–16. doi:10.18637/jss.v064.i02

Hansen, S., Landex, A., & Kaas, A. H. (2006). The network effects of railway investments. *WIT Transactions on the Built Environment*, *88*, 45–54. doi:10.2495/CR060051

Heinold, A., & Meisel, F. (2018). Emission rates of intermodal rail/road and road-only transportation in Europe: A comprehensive simulation study. *Transportation Research Part D: Transport and Environment*, *65*(September), 421–437. doi:10.1016/j.trd.2018.09.003

Hong, J., Chu, Z., & Wang, Q. (2011). Transport infrastructure and regional economic growth: Evidence from China. *Transportation*, *38*(5), 737–752. doi:10.1007/s11116-011-9349-6

Hu, Z.-H. (2019). Vietnam's Connectivity and Embeddedness in the Maritime Silk Road and Global Maritime Network. *IEEE Access*, *7*, 79592–79601. doi:10.1109/ACCESS.2019.2923528

Kurant, M., & Thiran, P. (2006). Extraction and analysis of traffic topologies of transportation networks. *Physical Review E*, *74* (036114).

Landex, A. (2012). Network effects in railways. *WIT Transactions on the Built Environment*, *127*, 391–402. doi:10.2495/CR120331

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

Lin, J., & Ban, Y. (2013). Complex network topology of transportation systems. *Transport Reviews: A Transnational Transdisciplinary Journal*, *33*(6), 658–685. doi: 10.1080/01441647.2013.848955

Lu, Q. C. (2018). Modeling network resilience of rail transit under operational incidents. *Transportation Research Part A*, *117*, 227–237. doi:10.1016/j.tra.2018.08.015

Majima, T., Katuhara, M. & Takadama, K. (2007). Analysis on transport network of railway, subway and waterbus in Japan. *In: Namatame A., Kurihara S., Nakashima H. (eds) Emergent Intelligence of Networked Agents. Studies in Computational Intelligence, vol 56. Springer, Berlin, Heidelberg.*

Mohmand, Y.T. & Wang, A. (2014). Complex Network Analysis of Pakistan Railways. *Discrete Dynamics in Nature and Society*, *2014*, Article ID 126261. doi: 10.1155/2014/126261.

Potter, J. (2018). *European Rail Timetable Limited*. (C. Woodcock, Ed.). Northamptonshire, United Kingdom: CPI Group (UK) Ltd, Croydon, CR0 4YY.

Roanes-Lozano, E., Laita, L. M., Roanes-Macías, E., Wester, M. J., Ruiz-Lozano, J. L., & Roncero, C. (2009). Evolution of railway network flexibility: The Spanish broad gauge case. *Mathematics and Computers in Simulation*. doi:10.1016/j.matcom.2008.11.007

Seaton, K. A., & Hackett, L. M. (2004). Stations, trains and small-world networks. *Physica A: Statistical Mechanics and its Applications*, *339*, 635–644. doi:10.1016/j.physa.2004.03.019

Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P. A., Mukherjee, G., & Manna, S. S. (2003). Small-world properties of the Indian railway network. *Physical Review E*, *67*(036106). doi:10.1103/PhysRevE.67.036106

Soh, H., Lim, S., Zhang, T., Fu, X., Lee, G. K. K., Hung, T. G. G., *et al.* (2010). Weighted complex network analysis of travel routes on the Singapore public transportation system. *Physica A: Statistical Mechanics and its Applications*, *389*(24), 5852–5863. doi:10.1016/j.physa.2010.08.015

Sun, S., Li, H., & Xu, X. (2017). A key station identification method for urban rail transit: A case study of Beijing subway. *Promet - Traffic & Transportation*, *29*(3), 267–273. doi:10.7307/ptt.v29i3.2133

Tsiotas, D. (2017). Links between network topology and socioeconomic framework of railway transport: Evidence from Greece. *Journal of Engineering Science and Technology Review*, *10*(3), 175–187. doi:10.25103/jestr.103.23

Wang, W., Cai, K., Du, W., Wu, X., Tong, L. (Carol), Zhu, X., & Cao, X. (2020). Analysis of the Chinese railway system as a complex network. *Chaos, Solitons & Fractals*, *130*, 109408. doi:10.1016/j.chaos.2019.109408

Yue, H., Guan, Q., Pan, Y., Chen, L., Lv, J., & Yao, Y. (2019). Detecting clusters over intercity transportation networks using K-shortest paths and hierarchical clustering: a case study of mainland China. *International Journal of Geographical Information Science*, *33*(5), 1082–1105. doi:10.1080/13658816.2019.1566551

Zanin, M., Sun, X., & Wandelt, S. (2018). Studying the topology of transportation systems through complex networks: Handle with care. *Journal of Advanced Transportation*, 18, 3156137. doi: 10.155/2018/3156137

Zhang, Z., Li, X., & Li, H. (2015). A quantitative approach for assessing the critical nodal and linear elements of a railway infrastructure. *International Journal of Critical Infrastructure Protection*, *8*, 3–15. doi:10.1016/j.ijcip.2014.11.001

FIGURES

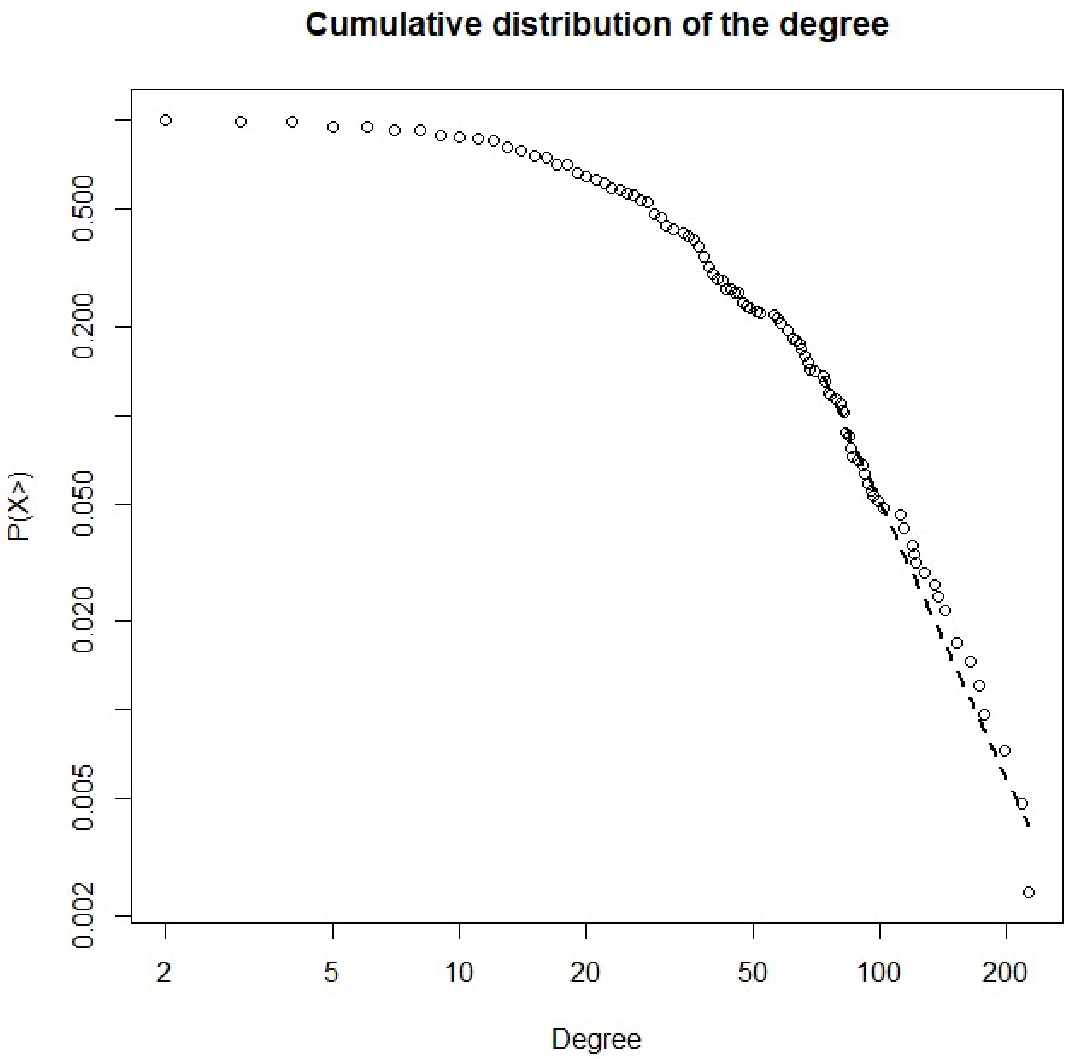

Figure 1 - Degree distribution of EIRN fitted to a power-law

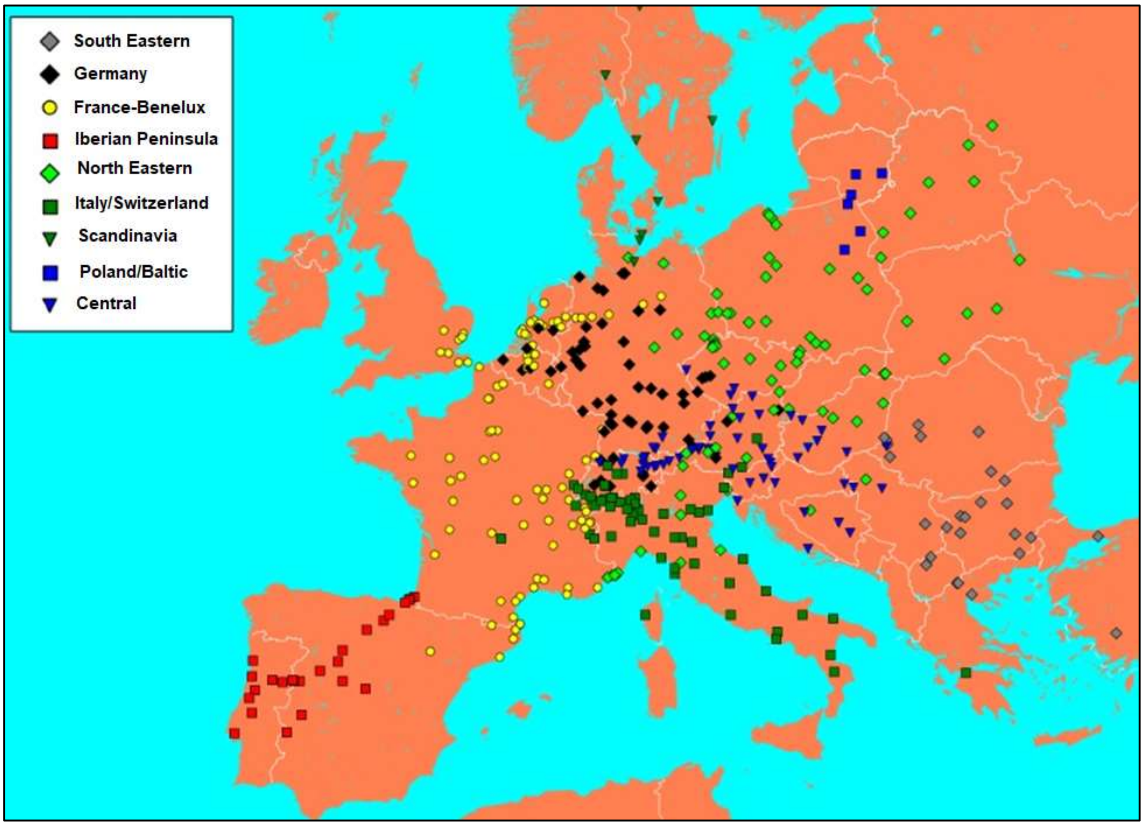

Figure 2 – Division of the stations by communities

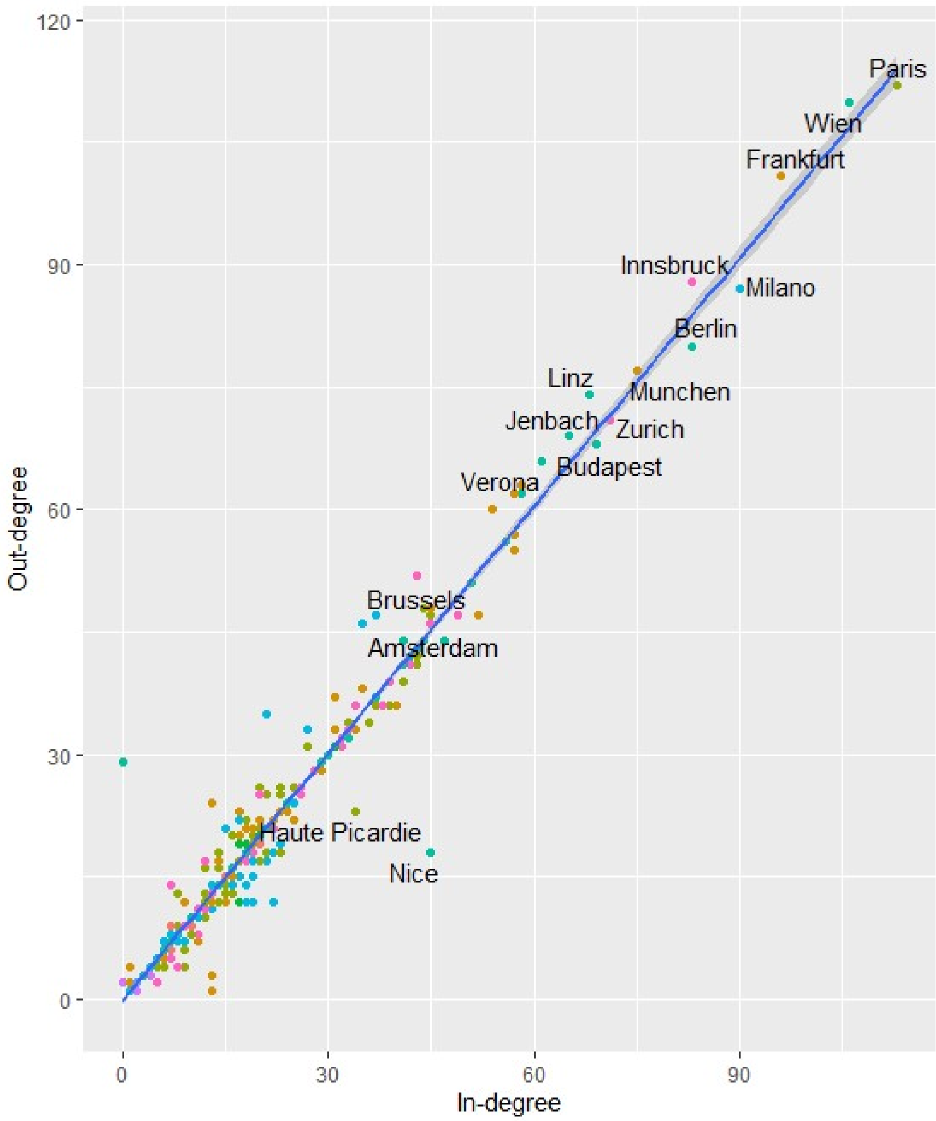

Figure 3 – Relationship between indegree and outdegree in EIRN

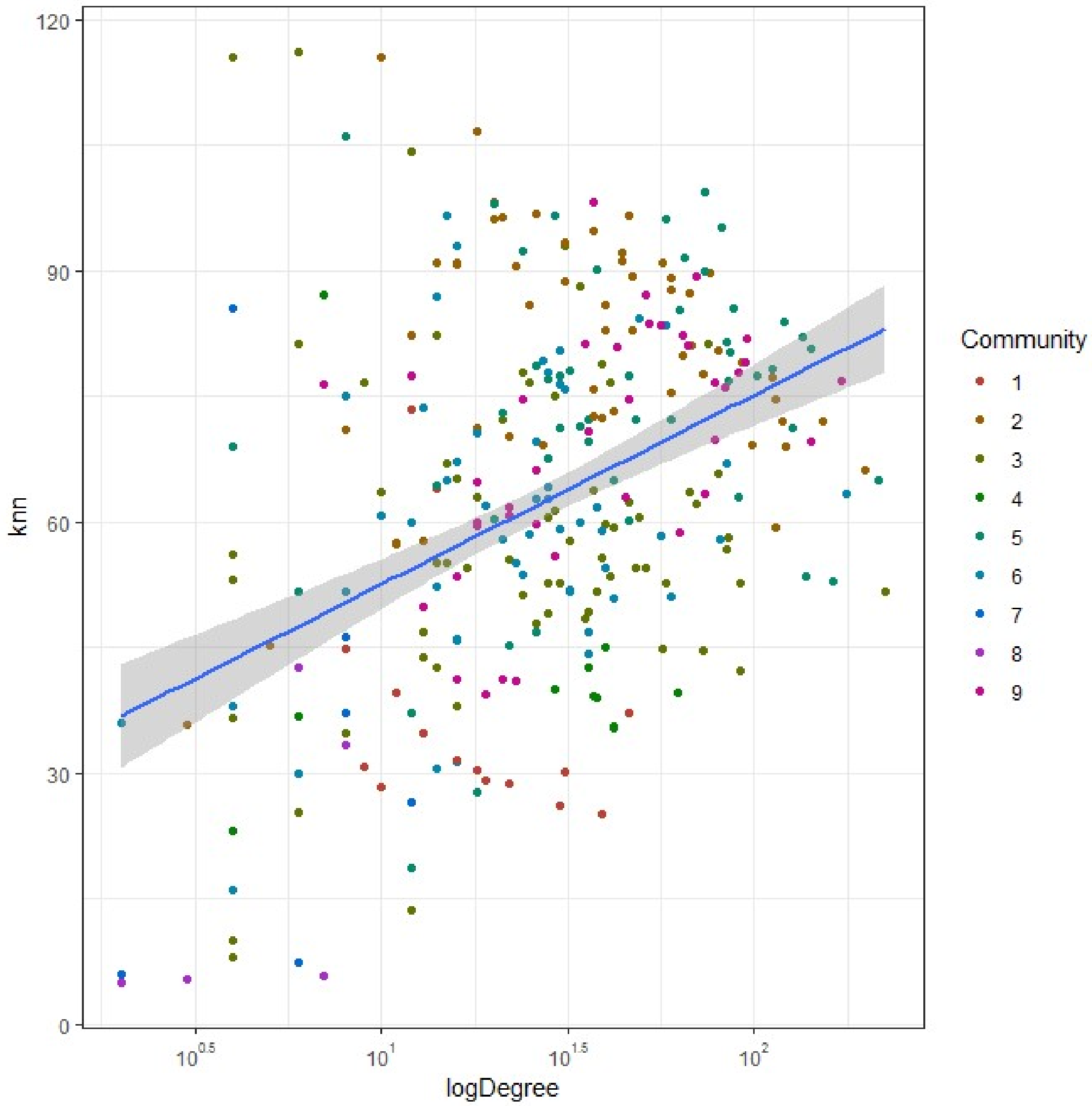

Figure 4 - Correlation between average degree of nearest neighbours (k_{nn}) and degree

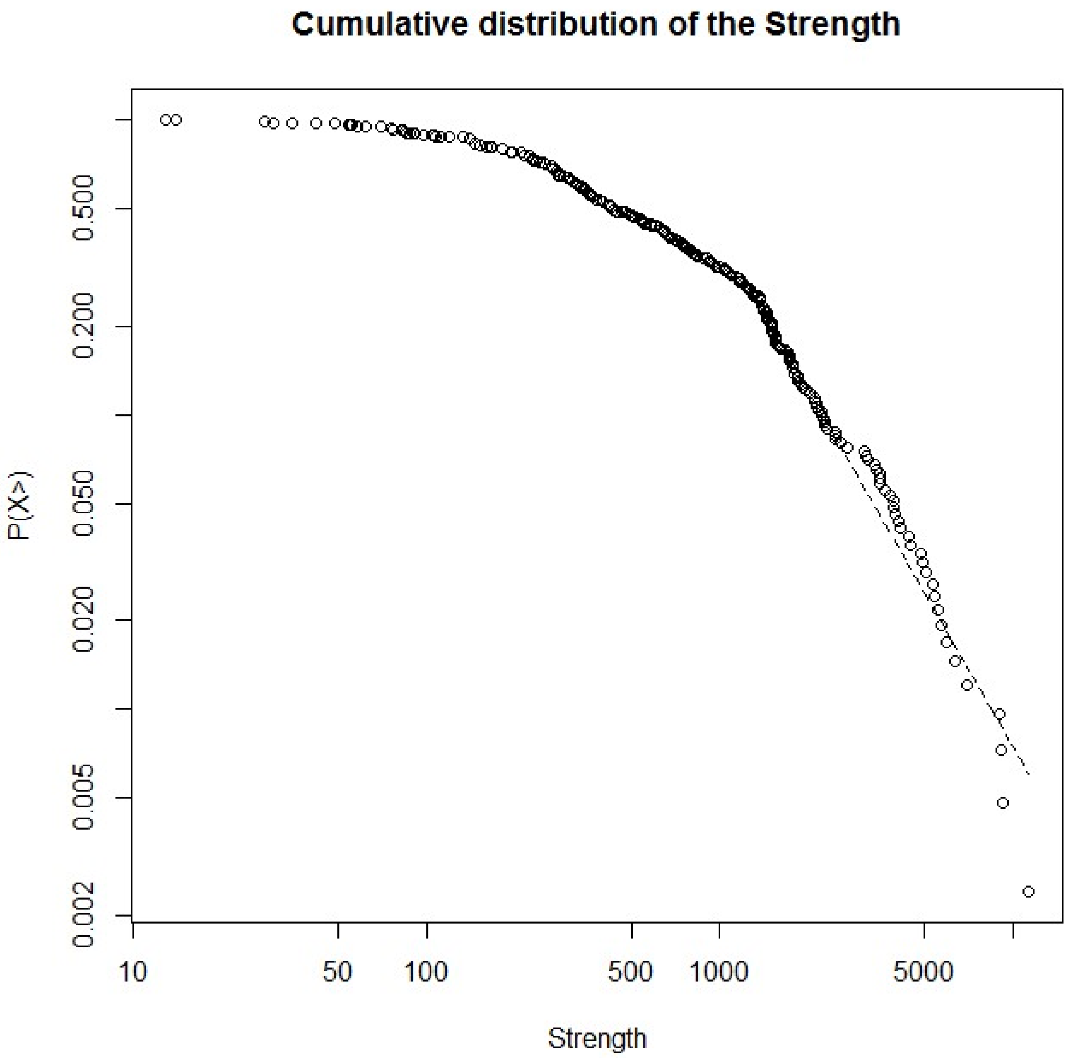

Figure 5 – Cumulative strength distributions of EIRN fitted to a power-law

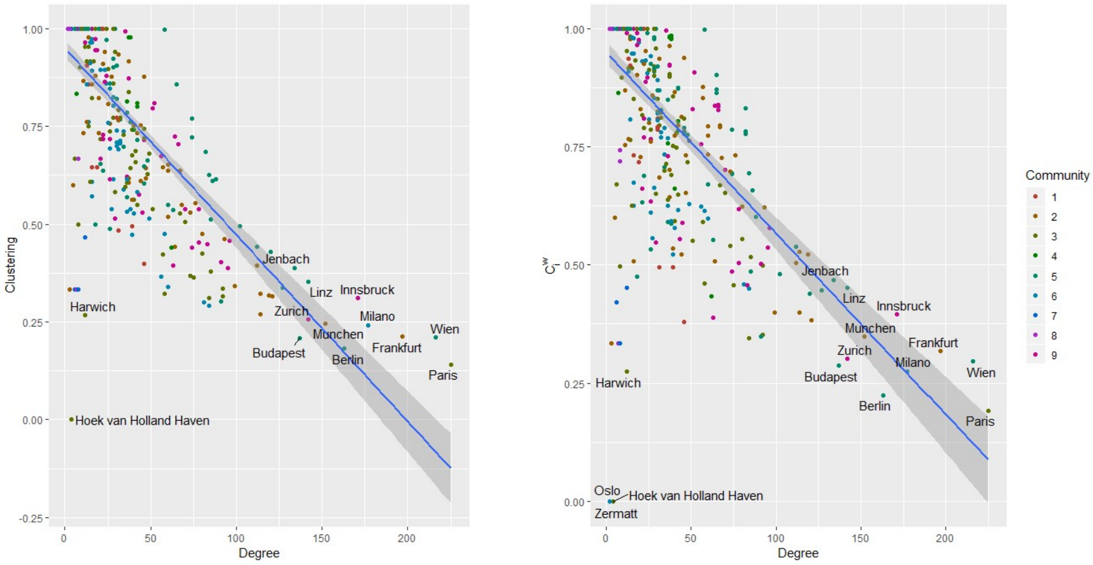

Figure 6: Correlation between degree and clustering (left) and weighted clustering (right)

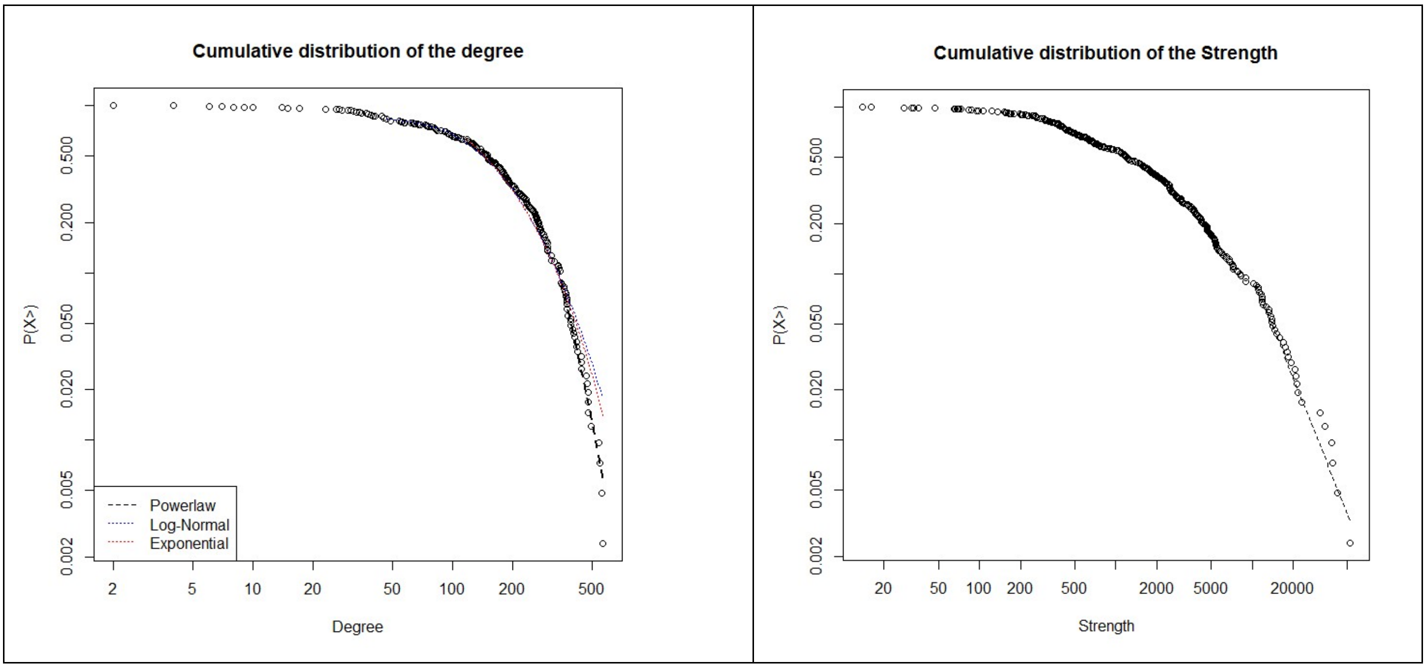

Figure 7 - Cumulative distribution of Degree and Strengths for EIRN-t

TABLES

Table 1: Influenced means of the distanced based metrics in EIRN depending on the geographical threshold.

Distance threshold (km)	Mean Degree	Mean Closeness	Mean Betweenness	Diameter	Radius	Average Path Length
500	14.27	1.34	0.03	5.16	1.88	2.02
1000	18.61	0.67	0.02	6.74	2.87	2.53
1500	19.30	0.51	0.01	6.97	3.64	2.73
2000	19.75	0.47	0.01	7.02	3.94	2.83
2500	19.87	0.47	0.01	7.02	3.99	2.86
∞	19.90	0.46	0.01	7.00	4.00	2.89

Table 2: Topological comparison between the European and the Chine networks (in brackets, z-scores)

METRIC	EUROPE (EIRN)	CHINA (CRN)	(EIRN-t)
Nodes / Edges	412 / 7,732	255 / 1,091	412 / 34,823
Density	4.56%	3.37%	20.6%
Diameter / Radius	7 [65.58] / 4	5 [-5.43] / -	5 [72.78] / 3
Average Path Length (unweighted)	2.89 [371.87]	2.66 [-20.17]	2.00 [845258.3]
Most probable shortest path	3	3	2.00
Maximum in-degree / Maximum out-degree/ Average degree	113 (Paris)/ 112 (Paris) / 18.77	- / - / 8.56	290 (Frankfurt) / 283 (Zurich)/ 84.52
In-degree Out-degree correlation	0.98	≈ 1	0.96
Average Degree Nearest Neighbours vs. Degree	(Correl. +0.29) Assortativity	(Correl. -0.17) Disassortativity	(Correl. +0.38) Assortativity
Average clustering coefficient	C=0.77 [714.90] (C^w=0.81 [614.53] for weighted)	0.49 [121.83]	C=0.74 [564.67] (C^w=0.86 [562.88] for weighted)
Degree Powerlaw	γ=4.11 (degree lower bound 73)	γ=0.82 for small degrees γ=2.09 for large degrees	γ=6.10 (degree lower bound 340)
Strength vs. degree: $S(k) \sim k^{\beta}$	β = 1.20 ± 0.07	β=1.3± 0.01	β = 1.42 ± 0.06.
Clustering vs. degree: $C \sim k^{-\eta}$	η= 0.27 ± 0.03	η= 0.36	η = 0.17 ± 0.02
Top 3 cities with highest degree	Paris (117) Wien (111) Frankfurt (102)	Beijing (109) Shanghai (85) Guangzhou (67)	Zurich (314) Frankfurt (313) Munchen (293)
Top 3 cities with highest strength	Frankfurt Koln Wien	Beijing Guangzhou Shanghai	Frankfurt Brussels Paris

Table 3: Influenced means of the distanced based metrics in EIRN-t depending on the geographical threshold.

Distance threshold (km)	Mean Degree	Mean Closeness	Mean Betweenness	Diameter	Radius	Average Path Length
500	53.35	2.34	0.03	3.64	1.77	1.55
1000	89.40	0.86	0.01	4.67	2.09	1.77
1500	96.71	0.52	0.01	4.93	2.73	1.89
2000	99.77	0.43	0.01	4.98	2.95	1.95
2500	100.88	0.42	0.01	5.00	2.99	1.98
∞	101.08	0.41	0.01	5.00	3.00	1.99