Entropy Analysis of a Railway Network’s Complexity

Duarte Valério 1,†, António M. Lopes 2,*‡ and José A. Tenreiro Machado 3,†

1 IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal; duarte.valerio@tecnico.ulisboa.pt
2 UISPA - LAETA/INEGI, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
3 Department of Electrical Engineering, Institute of Engineering, Polytechnic of Porto, R. Dr. António Bernardino de Almeida, 431, 4249-015 Porto, Portugal; jtm@isep.ipp.pt
* Correspondence: aml@fe.up.pt; Tel.: +351-91-349-9471
† These authors contributed equally to this work.

Academic Editors: Kevin H. Knuth and Adom Giffin
Received: 12 August 2016; Accepted: 27 October 2016; Published: 31 October 2016;

Abstract: Railway networks are among the many physical systems that reveal a fractal structure. This paper studies the Portuguese railway system, and analyzes how it evolved over time, namely what concerns the structure of its different levels, and its distribution over the territory. Different mathematical tools are adopted, such as fractal dimension, entropy and state space portrait. The results are consistent with the historical evolution of the network.

Keywords: railway network; entropy; fractal dimension; fractional calculus

1. Introduction

It is known that many physical systems can be described by fractals. Railway networks (RN) are one such type of system. Suburban railway networks of several cities (Paris, Moscow, Rhine conurbation, Seoul, Beijing, and Mexico City) were studied in this respect [1–4], but published literature analyses these networks only at the time of writing. The sole exception is [3], which covers a few decades.

Biological systems with fractal structure include the circulatory and respiratory systems in humans [5], or the circulation of sap in trees [6,7]. Among these, commuting resembles the respiratory system most, where flow takes place in both directions in the same pathway. The knowledge of a system’s fractal structure is important to determine its dynamic behavior, which often turns out to include phenomena of fractional or anomalous diffusion [8,9].

This paper studies the Portuguese railway system, for the entire territory located in Europe, and analyzes its evolution over time, since the first line was opened in 1856. Loosely speaking, we study how evenly the RNs spread over the Portuguese territory over one and a half centuries. Our purpose is to study the RN dynamics and to compare what is qualitatively known from history with our quantitative results. For that purpose, we adopt, as mathematical tools, fractal dimension, entropy, and state space portrait (SSP). The first tool, as mentioned, has been used for suburban RNs; the last two are, to the best of our knowledge, new for this application. They have, however, been applied in [10–12]. The analysis presented is descriptive, as is the case with [1–4], and serves to confirm the qualitative history of the RN. It is aimed at an academic study in the area of geography. Once the dynamic behavior of the RN is known, understanding its fractal geometry will be of importance for planning extensions or closures of lines, making the method relevant for a wider audience.

The evolution of the RN was crucial for the economic development of the country, as alternative means of transportation available until the second quarter of the 20th century were clearly inferior: navigable rivers in Portugal are not long and allow access to only a limited part of the territory,
and road transport only becomes a serious alternative with the development of efficient internal combustion engines. While limited by geographical reasons (especially by the mountainous orography of the North of Portugal, which anyway also hinders road construction and completely precludes building canals), the RN was faster and cheaper, and allowed an increased regional specialization of the economy [13]. Even though modern economic growth failed to take hold of Portugal before the end of the Second World War, the social and economic changes induced by the introduction of railway transportation were very significant [14], and it is therefore relevant to study the evolution of its structure.

The paper is organized as follows. Section 2 briefly describes the Portuguese railway system. Sections 3 and 4 study the time evolution of the RN by means of fractal geometry and entropy versus time, and SSP of entropy, respectively. Section 5 discusses the results obtained. Finally, Section 6 draws the main conclusions, perspectives, and future work, and foresees possible manners of incorporating these results in planning.

2. The Portuguese Railway System

The development of the RN has been closely related to the Portuguese history over the last century and a half. Figure 1 is a timeline of the main socio-economic events.

![Figure 1. Timeline of Portuguese history over the last 165 years; see [15] for more details.](image)

The RN began in 1856 when the first stations of a line that would connect Lisbon to Porto and to Madrid, Spain were opened. Lines were built both by public and private initiatives. In fact, the original aim was to link Lisbon to Madrid, with a branch to reach Porto. Shortly thereafter, this plan became a line to connect Lisbon and Porto, and the connection to Spain became secondary. Connections to France and the rest of Europe always had to overcome the difficulty of a change of gauge at the Spanish–French border.

The heyday of the RN was the mid-20th century; the economic and social importance of this means of transportation has since been decreasing, particularly due to an increasing importance of road transportation for both passengers and cargo. Nearly all existing lines are built with Iberian gauges (1668 mm according to present standards). Two lines with meter gauges (1000 mm) still exist; this gauge was typical in secondary lines in the north of Portugal (where the mountainous orography makes a narrower gauge cheaper to build), but these have nearly all been closed. Indeed, many lines with low traffic and no commercial interest ceased operations in 1988–1992, and others in 2008–2012. A few short lines have been built in this period, all of them for cargo, with the exception a suburban connection from Lisbon to its south suburbs on the other side of the Tagus River. Readers interested in a detailed history of the Portuguese railway system may consult [16,17].

While there is no room to show changes for every single year in which stations opened or closed, Figure 2 presents several maps, for chosen significant years unevenly spaced in time, from...
which the evolution of the RN with time is clear. Notice that, for obvious geographical reasons, all lines are on the European continent; no RNs have ever been built in the Azores and Madeira islands (Lines in Portuguese colonies during this period are not only outside the scope of this paper, but also have little interest for our purpose, since no colony ever had a proper network, but, at most, only unconnected lines).

Figure 2. Snapshots of the Portuguese railway network in years 1856, 1864, 1882, 1945, 1987 and 2012.

This paper concentrates on the railway system proper, not taking into account underground networks or trams; these are not connected to the railway system except for passenger interfaces, while Iberian and meter gauge lines were connected for both passenger and cargo. For each year, those lines that were operated publicly or privately are considered; therefore, lines under construction or lines still available, but without traffic, are disregarded. The RN is considered as a graph, where each station is a node. The geographical coordinates of train stations and the years when they were open and closed to service were found conjugating information in [17–19]. For international lines, the first station in Spanish territory was included. This graph concerns the geography of the line only, so differences in traffic flow are not taken into account, which lines are double-track, which are electrified, the maximum speeds attainable, extant signaling technology, or any other variable related to the quality of the line or the intensity of its use.

3. Fractal Dimension of the Portuguese Railway Network

A fractal is a geometrical object displaying a pattern repeated at different scales. The pattern may occur repeated exactly or approximately. While mathematical objects can repeat a pattern at all scales, infinitely large and infinitely small, actual plants reveal a fractal nature only within a range of scales. An RN is a good example, since scales larger than a continent or smaller than the gauge have no physical sense. The fractal dimension is a measure of how much the fractal fills the space as it is magnified from larger to smaller scales, for which there are several definitions, that in general do not coincide [20–22]. The definition more often found in practice is the box-counting fractal dimension, due to its easy numerical implementation. The algorithm can be described as follows:

- Repeat:
  - cover the fractal object $S$ with a grid consisting of squares (the boxes) with size $\epsilon > 0$, 

- find the number of boxes that include part of the fractal \( N_\varepsilon(S) \in \mathbb{N} \),
- decrease \( \varepsilon \).

- The fractal dimension \( b \in \mathbb{R} \) is the slope of the log-log plot of \( N_\varepsilon(S) \) vs. \( \varepsilon \), i.e.,

\[
b(S) = - \lim_{\varepsilon \to 0^+} \frac{\log N_\varepsilon(S)}{\log \varepsilon}.
\]

In other words, the fractal dimension is estimated as the exponent of a power law:

\[
y = ae^b.
\]

Here, \( a \in \mathbb{R}^+ \) is related to the length between bifurcations of the RN.

Figure 3 depicts the evolution of parameters \( a \) and \( b \) with time. Figure 4 shows how they are related to each other. From these figures, it is seen that results can be conveniently divided into five periods, discussed in Section 6.

**Figure 3.** Parameters \( a \) and \( b \) versus time; for periods A–E, see Section 6.

**Figure 4.** Locus of \( a \) versus \( b \), and periods A–E; dots get lighter colours with the passing of time.
4. Entropy and SSP Analysis of the Portuguese Railway Network

The availability of the RN was never uniform in the Portuguese territory. In this section, the uniformity, or lack thereof, of the distribution of train lines throughout the country is assessed by means of entropy and SSP.

4.1. Entropy of the Railway Network

Entropy plays an important role in the area of information theory and statistical mechanics. In the last decades, generalizations of the classic Boltzmann–Gibbs–Shannon formulation have been proposed, namely to explain the statistical nature of many complex systems that follow asymptotic power–law statistical distributions. These generalizations are often expressed in terms of a set of parameters, and, usually, do not obey the fourth Khinchin axiom [23]. Examples are the Tsallis [24], Rényi [25], Ubriaco [26], Kaniadakis [27,28], Sharma–Mittal [29,30], and Gamma [31,32] entropies, among others [33–39]. For certain particular values of the parameters, the classical entropic formulations are recovered.

The Boltzmann–Gibbs–Shannon entropy is given by:

$$S = \sum_i (-\ln p_i) p_i,$$

and represents the expected value of the information content, \(I(p_i) = -\ln p_i\), of an event with probability of occurrence \(p_i\), where \(\sum p_i = 1\).

The Tsallis, \(S_q\), and Rényi, \(S_r\), formulations are parametric entropies, given by:

$$S_q = \frac{1}{q-1} \left(1 - \sum_i p_i^q\right),$$

$$S_r = \frac{1}{1-r} \log \left(\sum_i p_i^r\right),$$

where \(q \in \mathbb{R}\) and \(r \in \mathbb{R}^+\setminus\{1\}\). In the limit \(\{q,r\} \to 1\), the Boltzmann–Gibbs–Shannon entropy is obtained.

Fractional calculus (FC) deals with the generalization of integrals and derivatives to an arbitrary real or complex order [40–45]. FC goes back to the beginning of the theory of calculus, but it was just in the last decades that FC was found to play an important role in modeling many physical phenomena, and emerged as an important tool in the area of dynamical systems with complex behavior. FC is able to capture long range phenomena, often overlooked by standard differential equations, being an interesting tool for analyzing phenomena that occur in many dynamical systems [5,7,40,46–56].

By adopting the tools of FC, the concepts of information content and entropy were recently revisited, leading to the information content and entropy of fractional order \(\alpha \in \mathbb{R}\) [57,58]:

$$I_{\alpha} (p_i) = D^{\alpha} I \{p_i\} = -\frac{p_i^{-\alpha}}{\Gamma(\alpha + 1)} \left[\ln p_i + \psi(1) - \psi(1 - \alpha)\right],$$

$$S_{\alpha} = \sum_i \left\{-\frac{p_i^{-\alpha}}{\Gamma(\alpha + 1)} \left[\ln p_i + \psi(1) - \psi(1 - \alpha)\right]\right\} p_i,$$

where \(D^{\alpha} \cdot\) is the derivative of order \(\alpha\) and \(\Gamma(\cdot)\) and \(\psi(\cdot)\) represent the gamma and digamma functions, respectively.
This fractional (or generalized) entropy (Equation (7)) does not obey some of the Khinchin axioms, except for $\alpha = 0$ [57]. In this case, it leads to the classical Boltzmann–Gibbs–Shannon entropy (Equation (3)).

For calculating the entropy of the RN, the probabilities are estimated according to the following procedure. First, the train lines are interpreted as a succession of evenly spaced points, with a step of approximately 1 km (chosen so that a lower value does not significantly change results), interpolated in-between the stations. Second, a rectangular grid with a 0.05° step (chosen like the step between points) and $N$ divisions is superimposed over the RN. Finally, the probabilities are computed by means of $p_i = \frac{n_i}{n}$, where $n_i$ is the number of points in each division of the grid $i$, and $n$ is the total number of points in all $N$ divisions.

Figure 5 depicts the variation of the Tsallis, Rényi and fractional entropies versus time and the parameters $\{q, r, \alpha\} \in [0.1, 0.9]$, respectively. The fractional entropy, $S_\alpha$, has a maximum for $\alpha = 0.76$. The Tsallis and Rényi entropies are both monotonic in the parameters, and $S_r$ varies little with $r$. Furthermore, $S_\alpha$ is more sensitive to changes in its parameter $\alpha$, and leads to better discrimination between entropy values. The fractional entropy was also compared with other entropic formulations [59] leading to similar conclusions.

Figure 6 depicts the time evolution of the Boltzmann–Gibbs–Shannon, $S$, and fractional, $S_\alpha (\alpha = 0.76)$, entropies for the period 1856–2012. We verify that $S_\alpha$ is able to “amplify” details improving the resolution of the entropy charts.

Figure 5. Tsallis, Rényi and fractional entropies, versus year and parameters $\{q, r, \alpha\} \in [0.1, 0.9]$, respectively: (a) $S_q$; (b) $S_r$; and (c) $S_\alpha$. 
4.2. The SSP of the Railway Network Entropy

For a \( k \)-dimensional dynamical system, the state space consists of the set of all possible states, each one corresponding to a unique point. As time evolves, sequences of points describing trajectories are obtained. The set of trajectories is the SSP. For first, second, and third-order systems, the SSP can be depicted and the system behavior can be inferred from the corresponding graphical representation [60,61].

For constructing the SSP, phase variables are adopted, namely the time evolution of the entropy in the period 1856–2012 and its \( k - 1 \) \((k \in \mathbb{N})\) time derivatives. As the order of the system differential model is unknown, successive increasing values of \( k \) were tested, and it was found that the two-dimensional space is a good compromise between assertiveness and feasibility of the SSP representation.

Noise may occur in the numerical calculation of the time derivatives. Therefore, the following algorithm [62] that avoids saturating the resulting signals is considered:

\[
y'(t) = \frac{1}{8h} \left\{ 2 \left[ y(t+h) - y(t-h) \right] + y(t+2h) - y(t-2h) \right\},
\]

where \( h \) denotes the sampling period.

It must be noted that (Equation (8)) requires two additional points at the left and the right. Therefore, the SSP turns out to be limited to the period 1858–2010. Several numerical experiments demonstrated that these formulae produce good results for the signals under analysis.

Figures 7 and 8 show the two-dimensional SSP representations, with \( h = 1 \) year, for entropies \( S \) and \( S_\alpha \) (\( \alpha = 0.76 \)), respectively. As can be seen, the fractional entropy leads to a SSP representation where details can be identified easily, since the chart has more resolution. On the contrary, the main loop corresponding to years 1958–1864 saturates the classic SSP chart, making for difficult the observation of small cycles.

The results obtained from fractal dimension and entropy plus SSP analysis are consistent with each other, and are discussed in the next section.
5. Discussion

The five periods of the Portuguese RN that can be found from the results above are as follows:

*Period A: 1856–1864.* The first train line opened in 1856, but there was not yet a proper RN during the first eight years. Therefore, this initial period should be overlooked. Spurious oscillations can be
seen in the variations of the fractal parameters $a$ and $b$. The entropy grows fast, as lines start reaching all regions; this corresponds to the biggest loop in the SSP plane of $S$ and $S_a$.

Period B: 1864–1882. There is growth of the fractal dimension $b$ of the network, but, more important still, there is an increase of the length of the average connection between two bifurcations of the network, as parameter $a$ shows. Entropy grows again increasingly, even if not as much as before. In other words, more than interconnections, longer lines were being built.

Period C: 1882–1945. The fractal dimension $b$ of the network keeps growing; periods of little increase are indeed those of little growth of the network. However, the inter-connectivity of the RN now increases in such a way that the average length of the connection between two bifurcations decreases. Entropy grows, but at an increasingly lower rate until stagnation occurs. In other words, it is not necessary to travel long before finding some bifurcation.

Period D: 1945–1987. This is a period of stagnation, during which few changes in the network take place.

Period E: 1987–2012. The closing of lines at the end of the 1980s was a big change: parameter $b$ decreases sharply, there is likewise a significant drop in entropy, and the same happens again with the closing of lines in 2008–2012. The average length of the connection between two bifurcations increases as lines are closed, but goes down as new short lines are built, as can be seen from the evolution of $a$.

All mentioned periods can be easily identified in the SPP chart of the entropy. Therefore, we conclude that the methods presented here lead to identical results.

6. Conclusions

The results discussed in the section above fit well with the short sketch of the evolution of the RN outlined in Section 2 and found in the references quoted. However, and even though they do reflect the changing role of railway transportation, they do so only partially, since line closures take place well after trains cease to be relevant in a certain region. Many lines are kept open for many years with few trains, few passengers, and few tons of cargo, awaiting a political decision about what to do. Actually, the 1945–1987 period of stagnation is a period when, while the network remains the same, significant changes are actually taking place (e.g., introduction of diesel locomotives, overhead electrification, end of the system of separate company franchises running different lines).

To study these issues quantitatively, future work will include analyzing figures for traffic of both passengers and cargo, as the flow efficiency and the percentage of the usage of the RN will have to play a more quantified role. Once this is done, it might be possible to use changes in fractal dimension and entropy to help plan new branches or minimize the effects of closures. In fact, only an RN with significant values of fractal dimension and entropy will be resilient to unscheduled service interruptions. When capacity is considered, not only new lines and line closures, but also variations in the capacity of existing lines or branches may be studied in this regard. Such results will, of course, have to be integrated with the existent models, and with financial, political or social concerns. However, it is relevant that, though the fractal nature of some suburban RNs has been studied for years, it is not considered in state-of-the-art modelling of passenger flows [63] or freight flows [64–66].

Such developments will be considered in future research, in the image of the biological systems with a fractal nature.

Acknowledgments: The authors thank Nuno Valério and Maria Eugénia Mata for pointing out interesting bibliographical items. This work was partially supported by Fundação para a Ciência e a Tecnologia, through IDMEC under LAETA.

Author Contributions: Duarte Valério, António M. Lopes and José A. Tenreiro Machado conceived, designed and performed the experiments, analyzed the data and wrote the paper. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.
References

29. Sharma, B.D.; Taneja, I.J. Entropy of type ($\alpha$, $\beta$) and other generalized measures in information theory. Metrika 1975, 22, 205–215.


47. Hanel, R.; Thurner, S. A comprehensive classification of complex statistical systems and an axiomatic derivation of their entropy and distribution functions. *EPL* 2011, 93, 20006.


© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).