1. **Materials and Methods**
   1. *Materials*

Information regarding occupational trajectories of directors of regulatory agencies come from several sources:

1. The Annual Social Information Report (Relação Anual de Informações Sociais - RAIS), which is an official registry of all formal workers in Brazil, maintained by the Ministry of Labor and Employment (all formal companies must fill the RAIS on a yearly basis). The RAIS contains social characteristics such as the employer, gender and age, occupation according to the Brazilian Classification of Occupations (CBO), compensation and worked hours, educational level and the employer’s size and location, among other information;
2. Superior Electoral Court (TSE) from Brazil, which allow us to track political affiliation of regulators;
3. Less structured information such as CVs, Agencies Websites, Press, Universities Websites and others to track information such as academic degrees, for instance.

We considered 117 directors and tracked their occupations 5 years earlier and 5 years after boarding. The positions assumed in each *t* were defined according to 7 dimensions, namely:

1. Male (H) or Female (M);

2 – Politically affiliated (S) or not (N) (according to TSE);

3 – Professor or scholar (S) or not (N);

4 – Whether he or she is an employee of the agency (S) or not (N);

5 – Whether he or she is a public servant (except the previous case) (S) or not (N);

6 – Whether he or she is from the regulated sector (S) or not (N);

7 – Whether he or she is a consultant (S) or not (N).

For instance, the string 1M2S3N4N5S6N7N means that individual is a woman who is politically affiliated and a public servant. Categories 4-7 are, in practice, mutually excludent.

Stability of the strings is also noticeable. In fact, when career changes happen, almost all of them take place after boarding. As a result, although there are 128 possible states (27), in practice there are only 40 different states. Moreover, the 10 most common sequences correspond to 40% of the total.

**Table 1 – 10 most common sequences**

|  |  |  |  |
| --- | --- | --- | --- |
| 5 years before | 5 years after | Freq | Percent |
| Male, public servant | Male, consultant | 11 | 9.4 |
| Male, public servant, politically affiliated | Male, public servant, politically affiliated | 6 | 5.1 |
| Male, public servant | Male, public servant | 5 | 4.3 |
| Male, public servant, politically affiliated | Male, public servant, politically affiliated | 5 | 4.3 |
| Male, from the regulated sector | Male, from the regulated sector | 4 | 3.4 |
| Male, public servant | Male, from the regulated sector | 4 | 3.4 |
| Female, public servant | Female, public servant | 4 | 3.4 |
| Male, professor or scholar, public servant | Male, professor or scholar, public servant | 3 | 2.6 |
| Male, professor or scholar, public servant | Male, professor or scholar, consultant | 3 | 2.6 |
| Male, from the regulated sector | Male, consultant | 2 | 1.7 |

Source: own elaboration

* 1. *Methods*

Careers are treated as a sequence of job positions over time (Spilerman, 1977). In recent years, one strand in the literature has applied longitudinal data techniques to compare career sequences to map patterns. One of these techniques is the Optimal Matching Analysis(OMA), introduced in career analysis by Abbott and Hrycak (1990) and revised by Abbott and Tsay (2000), Aisenbrey and Fasang (2010), and Dlouhy and Biemann (2015).

Optimal Matching Analysis defines the distance between sequences as the number of changes needed to transform one sequence into another. The lower this "cost", the more similar these sequences are. The operations allowed to transform one sequence into another are the substitution, insertion and elimination (indel operations, or indel operations) of a given state. The output of the comparison between the strings is a symmetric matrix that displays the distance from one sequence to all others. Finally, one uses this matrix to cluster sequences that are more or less similar, even though they are not necessarily the same.

Thus, besides the coding of sequences and the time frame, there are two critical decisions in applying OMA: the deletion/insertion and replacement costs between the states, when applicable; and the criterion for grouping the sequences. The TraMineR package from R 3.0.2 statistical software was used for the sequence analysis, as described by Gabadinho et al. (2011). The TraMineR algorithm is essentially that of Needleman and Wunsch, with standard optimizations (Gabadinho et al., 2011).

The transition costs between states were based on transition probabilities. This choice has been a growing trend in the literature (Aisenbrey and Fasang, 2010; Dlouhy and Biemann, 2015). Mathematically, the transition cost from state *i* to state *j* (*i ≠ j*) is equal to 2 – *p(i|j) – p (j|i)*, where *p(i|j)* is the transition rate between states *i* and *j* in the sample. The rationale behind this approach is that the transitions observed more frequently are less costly than less frequent transitions. By definition, the probability of a transition from one state to itself is equal to one, which makes the transition cost zero.

The clustering method was Ward’s hierarchical cluster, a standard in the literature. The choice of the number of clusters involved the analysis of some measures available in the R cluster package of and visual dendrogram inspection. No definitive criterion to choose the number of clusters exists; some methods and indicators aid researchers in this decision, but they often do not point towards a single solution. In the end, the choice of this number is somehow subjective. Herein, we chose a four clusters solution based on three indicators, the dendrogram and the analytical power of such a solution compared to alternatives.