Analysis of Professional Trajectories using Disconnected Self-Organizing Maps

Etienne Côme, Marie Cottrell, Patrice Gaubert

HAL Id: hal-01169515

https://hal.archives-ouvertes.fr/hal-01169515

Submitted on 29 Jun 2015

Analysis of Professional Trajectories using Disconnected Self-Organizing Maps

Etienne Côme[a], Marie Cottrell[b], Patrice Gaubert[c]

[a]*Université Paris-Est, IFSTTAR, GRETTIA, F-93166 Noisy-Le-Grand, France*
[b]*Laboratoire SAMM, Université Paris 1 Panthéon-Sorbonne*
90 rue de Tolbiac, F-75634 Paris Cedex 13, France
[c]*ERUDITE, Université Paris Est Créteil,*
61, avenue du Géneral De Gaulle, 94010 Créteil, France

Abstract

In this paper we address an important economic question. Is there, as mainstream economic theory asserts it, an homogeneous labor market with mechanisms which govern supply and demand for work, producing an equilibrium with its remarkable properties? Using the Panel Study of Income Dynamics (PSID) collected on the period 1984-2003, we study the situations of American workers with respect to employment. The data include all heads of household (men or women) as well as the partners who are on the labor market, working or not. They are extracted from the complete survey and we compute a few relevant features which characterize the worker's situations.

To perform this analysis, we suggest using a Self-Organizing Map (SOM, Kohonen algorithm) with specific structure based on planar graphs, with disconnected components (called D-SOM), especially interesting for clustering. We compare the results to those obtained with a classical SOM grid and a star-shaped map (called SOS). Each component of D-SOM takes the form of a string and corresponds to an organized cluster.

From this clustering, we study the trajectories of the individuals among the classes by using the transition probability matrices for each period and the corresponding stationary distributions.

As a matter of fact, we find clear evidence of heterogeneous parts, each one with high homogeneity, representing situations well identified in terms of activity and wage levels and in degree of stability in the workplace. These results and their interpretation in economic terms contribute to the debate about flexibility which is commonly seen as a way to obtain a better level of equilibrium on the labor market.

Keywords: Kohonen algorithm, planar graphs, labor market, Markov chains

1. Introduction

The aim of this study is to identify and to analyze the succession of situations occupied by workers on a modern labor market (1984-2003), it is an extended version of [1].

Basically the dominant economic theory (neo-classical sense) is based on the concept of the labor market where supplies (individuals) and demands (firms) meet. An equilibrium price (salary) makes the adequacy of supply and demand [2]. This theory defines mechanisms explaining labor supply by the wage level and predicting the stability of the relation between a firm and a worker and its evolution over time (a career). Unfortunately, these mechanisms are not observed in most real situations.

This is the pure neo-classical model. To get closer to the real economy, many developments have been made in the representation of the behavior of economic agents, in particular with the theory of job search taking into account different types of imperfections (incomplete information, the presence of institutions, regulation of relations between firms and employees...), see for example chapter 39 in [3]. But the basic hypothesis is unchanged: a single market whose functioning is flawed with the constraints and inefficiencies caused by the actual conditions. The result is closer to the real economy, but the deeper understanding of the phenomena that affect this sector of the economy in the last 30 years and their dynamics in the changing conditions of this period cannot be properly identified.

Instead of completing the pure neo-classical market model by a set of constraints more or less complex, the idea of this work is to show that the market is not homogeneous, it is the assembly of parts whose main characteristics are very different: that is the meaning of the assertion that the labor market is heterogeneous as mentioned in [4] and [5]. To find evidence of this heterogeneity, we construct a classification of the labor market observed over twenty years: this should identify the specific characteristics of each component and, secondly, permit to observe the situation of employees, over time, in these specific markets.

Our contribution consists in the identification and characterization of each class essentially using the variables used for classification (as well as some qualitative variables, subject to a satisfactory quality of this information.) A result to be expected is the dynamic view of trajectories of employees between these classes which can be obtained by observing the transition matrices between classes.

Let us identify the diversity of situations in terms of activity. A "situation" is defined by a combination of quantitative variables as shown in these two examples:

- full time job for the whole year, high wages, seniority in the same job;

- precarious conditions, wages lower than the average, part-time jobs, short-term contracts, on-call jobs, holding of a second job.

On the basis of individual characteristics, we construct a classification of situations observed every 2 years on a specific labor market, the US labor market over two consecutive periods of nine (1984-1992) and eleven (1993-2003) years respectively. So for each individual, we can observe the sequence of the classes he belongs to. That is what we call *professional trajectories*.

We need to study the temporal changes to answer some important questions linked to the evolution of the macroeconomic environment. Recall that in 1992, the end of Reaganomics and the beginning of Clinton period lead to a global reduction of unemployment during the rest of the decade. And this change leads to ask several questions. For instance what are the real changes at the individual level after 1992? More generally, what is the impact of a reduction of unemployment associated with "new forms of employment" . And also are there different conclusions if we observe the specific situation of women on the labor market?

This article follows another paper [6] but contains necessary material (and possibly redundant) to be self-contained. It is organized as follows: Section 2 presents the data and the notations used throughout the paper. The methodology and the global architecture of the proposed procedure are described in Section 3. Section 4 discusses how to choose the more efficient topology for the map. In sections 5 and 6, the classes are analyzed from an economic point of view. Finally section 7 presents the transitions from one class to another, according to the period and gender. Section 8 is devoted to a discussion of recent articles and to a conclusion which summarizes the main results.

2. The Data: first period (1984, 86, 88, 90, 92) and second period (93, 95, 97, 99, 2001, 03)

We use the PSID (Panel Study of Income Dynamics)[1], dividing the observations in two periods in order to meet two objectives: on one hand to observe a number of workers large enough to obtain statistical indicators representative of the whole population and on the other hand to keep only individuals present all along each period to identify trajectories.

We create a sample for each period (1984-1992, 1993-2003). By looking at descriptive statistics for the quantitative variables for each period, we can assume that both periods have the same rough characteristics. So we can make the classifications with all the observations together.

In the PSID, we select households for which the head (man or woman) is present in the household every year of the period and we do it separately for each period. The administrative rule is that if there is a male in the household, he is the head, if not the head is a woman. Fortunately quite the same variables concerning the activity on the labor market are available for the wife/partner of the head, if there is one. Retrieving this information, we constitute a set of individuals (3965

---

[1]Available online at `http://psidonline.isr.umich.edu/`

in period 1, and 3607 in period 2) observed every two years in each period, with a proportion of women close to that observed in the whole population.

An observation consists of a couple (year, individual). Each one is described by 8 quantitative variables and 2 qualitative variables. See Table 1 for the list of variables and their meaning.

Name	Description	Min-Max	Type
nbhtrav	Number of worked hours per week	0-112	Quant
nbstrav	Number of worked weeks	0-52	Quant
nbschom	Number of unemployed weeks	0-52	Quant
nbsret	Number of weeks out of the labor market	0-52	Quant
salhor	Hourly wages in real dollars	0-83.85	Quant
nbex	Number of extra jobs	0-5	Quant
hortex	Number of hours worked in extra jobs	0-1664	Quant
anctrav	Seniority in current job in months	0-780	Quant
gender	Gender	2 modalities	Qual
age	Age group (<30, 30-45, >45)	3 modalities	Qual

Table 1: Variable names, description and type, for PSID dataset.

The pre-processing consist of removing observations with clearly inconsistent values such as a number of week per year greater than 52. After this filtering 41467 observations constitute our working database. Observed current wages are converted in real dollars using the Price Index of PIB in 1992 (first period) or 2003 (second period). Eventually, the 8 quantitative variables were centered and reduced to standardize the order of magnitude. We can compute the correlation matrix of these variables, displayed in Table 2.

	nbhtrav	nbstrav	nbschom	nbsret	salhor	nbex	hortex	anctrav
nbhtrav	1	0.72	-0.04	-0.14	0.36	0.05	0.01	0.23
nbstrav	0.72	1	-0.23	-0.30	0.38	0.06	0.01	0.30
nbschom	-0.04	-0.23	1	0.02	-0.09	-0.01	-0.01	-0.11
nbsret	-0.14	-0.30	0.02	1	-0.10	-0.04	-0.04	-0.12
salhor	0.36	0.38	-0.09	-0.10	1	0.07	0.05	0.31
nbex	0.05	0.06	-0.01	-0.04	0.07	1	0.72	0.00
hortex	0.01	0.01	-0.01	-0.04	0.05	0.72	1	-0.01
anctrav	0.23	0.30	-0.11	-0.12	0.31	0.00	-0.01	1

Table 2: Correlation matrix of the quantitative variables.

We observe that variables *Number of worked hours per week* (nbhtrav), *Number of worked weeks* (nbstrav), *Hourly wages in dollars* (salhor) and *Seniority in current work in months* (anctrav) are strongly positively correlated, and that they are opposite to *Number of unemployed weeks* (nbschom) and *Number of weeks out of the labor market* (nbsret). The variables related to *extrajobs* are not correlated with the others.

3. SOM, Disconnected Self-Organizing Maps (D-SOM), Self Organizing Star (SOS)

3.1. The Kohonen algorithm (SOM)

In its classical presentation [7, 8], the SOM algorithm is an iterative algorithm, which iterates the two following steps over training patterns $\mathbf{x}_j$ for computing the set of code-vectors $\mathbf{m}_i, i \in \{1, \ldots, K\}$ which define the map:

- *Competitive step,* this step aims at finding the best matching unit (BMU) for sample $\mathbf{x}_j$:

$$c = \arg \min_{i \in \{1,\dots,K\}} ||\mathbf{x}_j - \mathbf{m}_i||. \qquad (1)$$

- *Cooperative step,* this step aims at moving the code-vectors of the BMU and of its neighbors (on the map), closer to the training pattern:

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + \alpha(t)h_{ci}(t)\left[\mathbf{x}_j - \mathbf{m}_i(t)\right], \qquad (2)$$

with t the time step, $\alpha(t)$ the learning rate of the algorithm and $h_{ci}(t)$ the neighborhood function between units c and i at time t.

Several neighborhood functions are commonly used such as $h_{ci}(t) = \mathbf{1}_{(d_{ci} < \sigma_t)}$ or $h_{ci}(t) = \exp(-d_{ci}^2/2\sigma_t^2)$. All of them depend on a radius σ_t which is classically decreasing during the learning process. These neighborhood functions also depend on d_{ci}: the distance between units c and i, which is determined by the map topology.

During the cooperative step, the coordinates of the best matching code-vector $\mathbf{m}_c$ are updated in order to move it closer to the training pattern. Other code-vectors $\mathbf{m}_i$ are also moved towards the training pattern according to their distance d_{ci} from the BMU defined by the lattice. Closer code-vectors from the BMU are more affected than the others. The two steps are iterated over the dataset during several epochs until convergence. Thanks to the cooperative steps, self-organization is reached at the end of the algorithm.

One can see that the distance between units in the map plays a key role in the self-organization property of the algorithm; modifications of this distance will have an impact on the results of the algorithm. Therefore, the lattice structure can be a way to incorporate prior information concerning the dataset topology into the dimensionality reduction process. In the context of classical SOM, one assumes that the dataset topology can effectively be represented by a grid or a straight line but other hypotheses can be interesting and can advantageously be investigated.

It has been already noted that graph theory can be used to define this distance [9, 10]. In this case, map units are the nodes of a graph, and distances between them are defined as the minimum number of edges needed to reach one node starting from the other, that is the so-called *shortest path distance*. We therefore propose to modify the SOM algorithm only by taking as input an adjacency matrix [2] (see [11]), which specifies the graph topology that the user desires. All undirected graphs can theoretically be used, but a specific class is of interest: the class of planar graph, because such graphs can easily be represented in a 2-dimensional setting allowing us to supply the SOM with visualization tools, as seen in the next section.

3.2. New topologies

An interesting choice is defined by a map which is composed of several disconnected one-dimensional strings. Each string will contain data which are similar at a rough level and that are displayed in an ordered disposition.

This topology has a special interest: when the map consists of not connected parts, the "cooperation" step of the algorithm only concerns the units which belong to the same component as the winning unit. The competition step is not modified, so that the algorithm meets a double goal :

1. to group the observations into macro-classes corresponding to the different disconnected components of the graph;
2. to organize the units inside the macro-classes.

Figure 1 shows an example of a disconnected neighborhood structure that we define here and of a classical grid neighborhood. In the disconnected case, for example, $d_{13,21} = +\infty$ and $d_{26,31} = 5$, while in the classical grid case, $d_{13,21} = 1$ and $d_{26,37} = 4$.

---

[2]For two units i and j, entry (i,j) of the adjacency matrix is 1 if there exists an edge between i and j, and 0 if not.
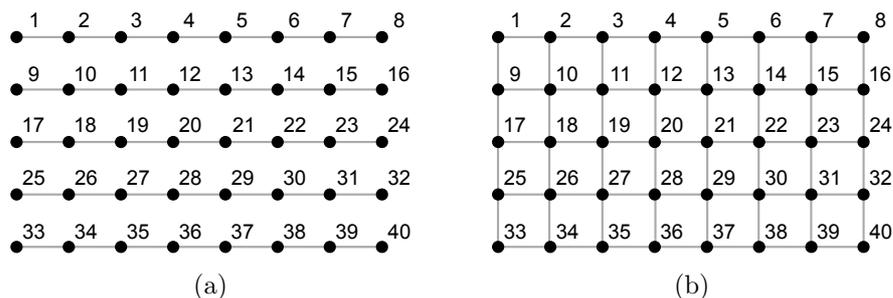
Figure 1: (a) Two-dimensional representation of a disconnected map with 5 strings of 8 units and (b) representation of classical (5 by 8) grid map.

In conclusion, by imposing a limitation of the cooperation which only acts inside the macro-classes and by keeping a competition between all units, this algorithm allows us to obtain a given number of macro-classes which are themselves self-organized. We call this topology the D-SOM.

Another interesting choice is a star-shaped graph as shown in Figure 2. This graph has a clear natural center from which different arms or rays grow. Such graph can easily be interpreted by users if different rays correspond to different well identified classes while the center gather the "normal" patterns. The organization takes place on each ray and distances to the center describe the characteristics of the patterns in an ordered way. These graphs can be characterized by their number of rays and by the length of the rays. SOM using such a topology will be called Self-Organizing Stars (SOS) (as defined in Come et al. (2010) [6]).

Figure 2 shows an example of star-shaped neighborhood structure. In this case, for example $d_{3,20} = 5$.
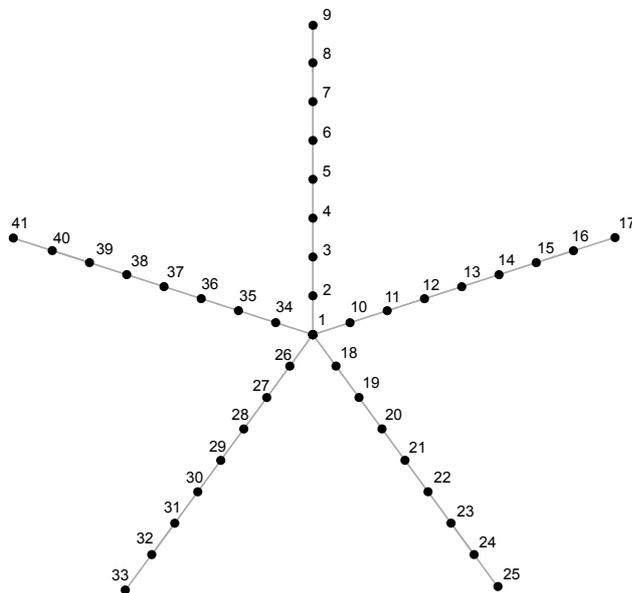
Figure 2: Two-dimensional representation of a star-shaped map with 5 rays of 8 units.

There exist other methods to obtain well-separated classes, see [12] for example. But our approach is different since we do not look for building an adjacency matrix between the code-vectors by repeating many runs of the SOM algorithm. Contrarily, we impose an a priori adjacency matrix which defines star-shaped classes or non-connected classes.

It is also possible to use U-matrix visualization as in [13] or [14], or direct clustering of code-

5

vectors as in [15] to define the macro-classes. In the following we achieve an Hierarchical Ascending Classification of the code-vectors in the classical grid case, to group the classes into a previously fixed number of macro-classes, see [8].

Both kinds of topology (D-SOM, SOS) are well adapted to the analysis of labor market segmentation, since one looks for a segmentation into macro-classes well discriminated, split into organized classes. In a general case, the question of the choice of the number of macro-classes is guided by *a priori* argument if it exists. In our case, we choose 5 macro-classes which is the best choice to get contrasting and well identified situations. In fact, in the literature, the authors generally choose 4 segments, (for example, see two recent studies on real markets, the French and the German labor markets, [4], [5]). But we have to add a fifth segment specific to the US economy, which corresponds to the regular practice of two or more jobs (it is a fairly common practice in the US economy, while it is rare in France and Germany). This is the reason for choosing five macro-classes and no other numbers that the economic model probably could not explain.

Let us now describe the results that we get using these three topologies for the PSID data.

4. Comparison of the maps, choice of a topology

We use the Kohonen algorithm for three different topologies : the classical one on a (5 by 8) grid, a D-SOM with 5 strings of 8 units, a SOS with 5 rays of 8 units. The total numbers of classes are almost the same (40, 40 and 41 units). The data include 41467 couples (year, individual) represented by a 8-vector composed of the 8 quantitative variables.

The number of initial classes (micro-level) is determined by the number of available observations: descriptive statistics constructed at this level must be calculated with a sufficient number of observations in each of these classes. With approximately 40,000 observations, dividing each macro-class in 8 units gives 40 classes with 1000 observations by class on average.

Figures 3 (a), (b) and (c) show the code-vectors for each map: the 8 components of each code-vector are displayed according to the order defined in Table 1. They are well organized. The SOM map is organized in all directions, while the others are organized inside each string.

From a quantitative point of view, a *first indicator* to compare the three maps is the *quantization error*, which is a measure of quality of the clustering. We consider the within sum of squares as:

$$SC_{within} = \sum_x \|x - \mathbf{m}_{c(x)}\|^2, \tag{3}$$

where

$$c(x) = \arg \min_{i \in \{1,...,K\}} \|\mathbf{x} - \mathbf{m}_i\|. \tag{4}$$

This is simply the sum of the squared distances between each pattern x and the code-vector of its BMU, in the pattern space.

Then we define the total sum of squares as

$$SC_{total} = \sum_x \|x - \bar{x}\|^2. \tag{5}$$

So we can define the *relative quantization error* as:

$$RQE = \frac{SC_{within}}{SC_{total}}. \tag{6}$$

The smaller the relative quantization error, the better the classification.

A *second indicator* is the ratio between the sum of squares extended to neighbor code-vectors and the total sum of squares. If we note (as in [16])

$$SC_{extended} = \sum_x \sum_{k \in \mathcal{V}(c(x))} \frac{1}{|\mathcal{V}(c(x))|} \|x - m_k\|^2, \tag{7}$$

6
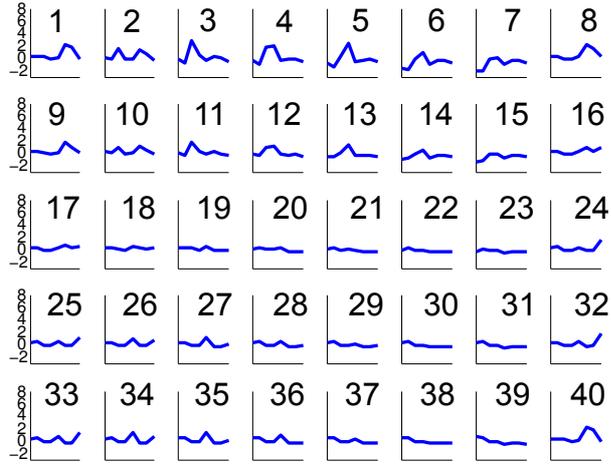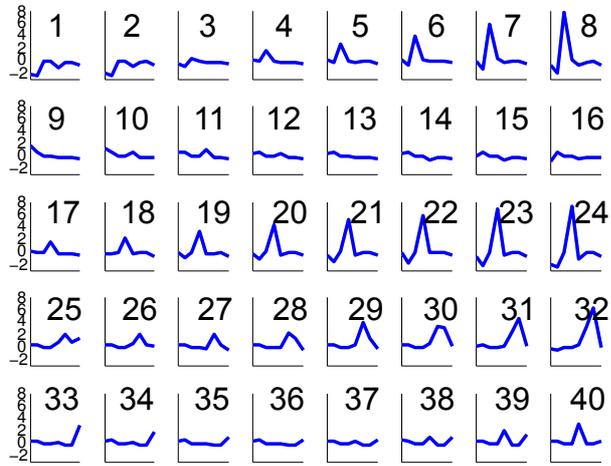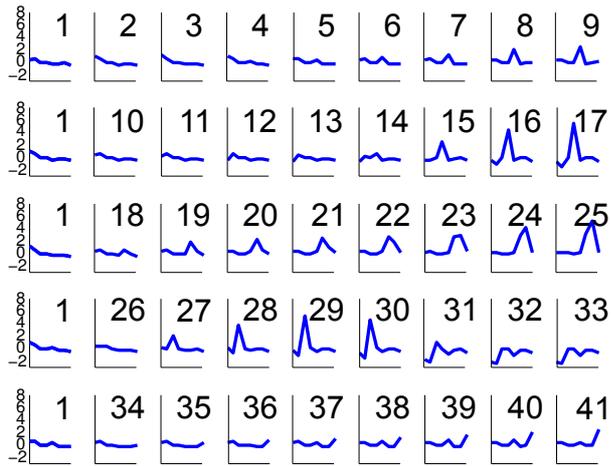
Figure 3: (a) SOM map (classical grid neighborhood) codebook representation, (b) D-SOM map codebook representation, (c) SOS map codebook representation. For each codebook its coordinates in the feature space are depicted with features ordered as in Table 1. For each subplot, in abscissa we find the features number and in ordinate the standardized values of the features for the codebook.

where $\mathcal{V}(c(x))$ is the set of neighbors of $c(x)$, as defined by the adjacency matrix of the graph, we can compute the *relative extended quantization error*:

$$RQE_{ext} = \frac{SC_{extended}}{SC_{total}}. \tag{8}$$

A small value of RQE_{ext} indicates a good organization, since it implies that neighbor code-vectors on the map are close in the pattern space.

For comparison, the three possible solutions namely SOM, D-SOM and SOS were fitted to the PSID data with the same procedure and parameters (linear decreasing learning rate and Gaussian neighborhood function one run per method), the corresponding results are given in Table 3.

Table 3 shows that D-SOM gets better quantization than the others at the unit level. This is because the map is less constrained, the adaptation algorithm finds a better minimum for RQE. As to the relative extended quantization error RQE_{ext}, the results of D-SOM and SOS are close and better than the SOM case results. Furthermore, D-SOM and SOS allow us to get well-contrasted and easy-to-interpret macro-classes. As our main goal is to build robust macro-classes, we decide to consider only these two topologies.

	RQE	RQE_{ext}
SOM	22.23%	40.37%
D-SOM	12.79%	22.01%
SOS	16.79%	22.36%

Table 3: Relative quantization error and relative extended quantization error for the three topologies in %.

At the unit level the D-SOM topology achieves therefore the best result on this particular dataset and such a topology seems therefore better fitted to such a case.

The same type of quality measures can be computed at the macro-classes level, as a *third indicator*. To this end we define:

$$SC_{macro} = \sum_x \|x - \mathbf{M}_{b(x)}\|^2, \tag{9}$$

where $b(x)$ give the macro-class number of x such as $b(x) = s, s \in \{1, \ldots, S\}$ with S the number of macro classes in the map and $\mathbf{M}_s$ is the empirical mean of macro-class s members. Normalizing this quantity by the total sum of square as previously give a normalized quality measure:

$$RQE_{macro} = \frac{SC_{macro}}{SC_{total}}. \tag{10}$$

This quantity can be easily computed for the D-SOM topology and also for the SOS topology if each ray of the star is associated to a macro-class plus one macro-class for the center of the star. Since the classical SOM topology based on rectangular grid did not define such macro-classes, one has to use an additional step to build them using K-means [15] or Hierarchical Ascending Clustering (HAC) [8]. We used the later here to build the macro-classes in the classical SOM map. Using such an approach, the evolution of RQE_{macro} with respect to the number of macro-classes for the SOM map can be computed and is depicted in Figure 4. So we are able to do the comparison of the *relative quantization errors at the macro-class level* for the three maps with 5 macro-classes. The results are presented in Table 4.

	RQE_{macro}
SOM+HAC	60.4%
D-SOM	47.5%
SOS	55.7%

Table 4: Relative quantization error at the macro-classes level for the three topologies in %.
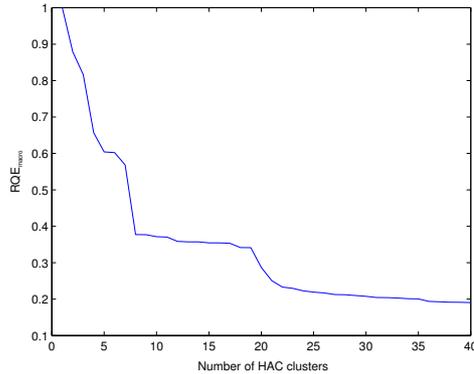
Figure 4: Evolution of RQE_{macro} with respect to the number of macro-classes for SOM+HAC.

This quality measure leads to the same conclusion as for the previous, D-SOM performs better on the PSID dataset and the differences with the two other approaches is clearer. As expected the results are poorer than at the unit level since the description of the data is coarser. All these elements lead us to choose the D-SOM topology with five macro-classes as a reference to study the PSID dataset in the rest of the paper.

Even if D-SOM seems to be the best choice, we can go deeper in the comparison between SOS and D-SOM in analyzing the classifications *from an economic point of view*. The best one is the one offering the clearest interpretation of a specific labor market during almost twenty years, with the influence of changing economic policies and changing economic environment. The latent global interpretation is that this labor market is not a homogeneous market, but a set of sub-markets well differentiated in terms of level of activity, wages and seniority; the links between them, for instance the worker's ability to move from one segment to another one, are important points to understand the economic system.

4.1. Analysis of the macro-classes resulting from D-SOM algorithm

This topology is well adapted, applied to the whole sample, if the basic hypothesis "the labor market is constituted of 5 segments well differentiated independently from the period of observation" is true. If this was false, we could see that some macro-classes correspond to the first period and the others to the second one: this is not what we have found. See Table 5. All macro-classes are almost equally divided between the two periods.

	Class 1	Class 2	Class 3	Class 4	Class 5	Sample
Whole sample	6051	15597	1664	5043	13112	41467
Period 1 (%)	61.4	47.3	58.4	47.0	41.1	47.8
Period 2 (%)	38.6	52.7	41.6	53.0	58.9	52.2

Table 5: Share of each period in each class.

The proportions are approximately the same for both periods in each class, as we see in Table 5.

So we may summarize the main results of the classification in Table 6, where each column shows a macro-class: the eight variables used by the algorithm describe the similarities and differences between the segments from a quantitative point of view (means).

- **Class 1**: low activity measured by hours of labor by week and number of weeks of labor per year and significant periods of unemployment (close to 7 weeks);

- **Class 2**: high activity but for a wage lower than the whole average and a short seniority; this may be identified as a set of jobs with a great flexibility;

9

	Class 1	Class 2	Class 3	Class 4	Class 5	Sample
	6051	15597	1664	5043	13112	41467
nbhtrav	12.34	**42.42**	28.33	40.24	41.90	37.04
nbstrav	9.63	**49.18**	21.67	46.04	48.01	41.55
nbschom	**6.87**	0.11	0.73	0.47	0.10	1.16
nbsret	0.42	0.12	**30.09**	0.36	0.14	1.40
salhor	3.37	11.92	7.57	14.39	**18.61**	12.91
nbex	0.02	0.00	0.04	**1.20**	0.00	0.15
hortex	4.70	0.05	6.22	**458.44**	0.98	57.02
anctrav	11.19	35.16	17.47	93.77	**202.68**	91.05

Table 6: Mean values for each variable by the 5 D-SOM macro-classes and for the whole sample; the figures in bold are the maximum values for each variable, the figures below class names are the class sizes.

- **Class 3**: mainly defined by a part-time activity over the year with more than half the year out of the labor market and a very short seniority [3];

- **Class 4**: constituted with people having more than one job at the same time; the obtained wage is then greater than the average;

- **Class 5**: constituted of the "good jobs", with high activity, wages approximately 50% greater than the average and a great stability with quite 17 years in the same place.

4.2. Analysis of the macro-classes obtained using SOS algorithm

Comparing D-SOM and SOS algorithms, a first difference comes from the structure of the map. The SOS structure is very convenient if the phenomenon under study contains a reference situation or "*normal*" situation (the center of the star) and several shifts or gaps of one or several variables which are represented on the different rays of the star. It could be a good representation if the labor market was conceived as a pure mechanism that could be observed in the real economic system (what is not true due to institutional and social constraints).

From the computational point of view, the results are very similar to that obtained with D-SOM, although there is a sixth macro-class (the center of the star).

We calculate the arithmetic means of the eight variables and can emphasize the contrasts between the macro-classes, (the five rays and the center). In Table 7, let C1-SOS, C2-SOS, ..., C6-SOS be the classes obtained using SOS algorithm. Class C6-SOS is the center of the star.

	C1-SOS	C2-SOS	C3-SOS	C4-SOS	C5-SOS	C6-SOS	Sample
	10177	7210	4913	6146	12447	574	41467
nbhtrav	**48.12**	31.31	40.32	13.04	41.56	43.08	37.04
nbstrav	**48.64**	42.78	46.03	10.69	48.19	48.66	41.55
nbschom	0.10	0.21	0.47	**6.89**	0.08	0.02	1.16
nbsret	0.15	**6.99**	0.35	0.43	0.13	0.13	1.40
salhor	**19.96**	7.35	14.27	3.36	14.71	9.45	12.91
nbex	0.01	0.01	**1.21**	0.02	0.01	0.00	0.15
hortex	1.29	1.44	**466.91**	4.81	1.39	0.00	57.02
anctrav	52.34	24.11	86.04	11.08	**205.39**	37.92	91.05

Table 7: Characteristics of the 6 macro-classes; the figures in bold are the maximum values for each variable, the figures below class names are the class sizes.

Three macro-classes are very close to their equivalent in the D-SOM classification (C2-SOS, C3-SOS and C4-SOS respectively for Class 3, Class 4 and Class 1 in the D-SOM classification):

---

[3]Presumably, unemployment and other situations out of activity are not correctly declared by the individuals: in this macro-class there is an average of 7 weeks unemployed, less than one week temporarily out of the market and close to 10 weeks at work, letting 34 weeks of a year unexplained.

out of the market temporarily, more than one job at the same time, low activity with some period of complete unemployment. The last two (the sixth being treated separately) show situations with a great level of activity and wages that are very high (class C1-SOS) or slightly above the global mean (class C5-SOS).

We encounter some problems in interpretation: in opposition to the first classification, the class C1-SOS associates high wages and short seniority, while C5-SOS associates a relatively moderate wages with more than 17 years of seniority. As a matter of fact this opposition was observed inside macro-class Class 5 of good jobs (D-SOM classification) between the units: at one end of the string are found the jobs with high wages and low seniority and at the other end the exact opposite situation. This has been interpreted as a trade-off between these two positive characteristics: you have to move on the market to obtain better careers with significant higher wages.

4.3. Crossing the classifications

Finally if we cross both classifications at the individual level, we can show more precisely what the problem is. See Table 8:

	C1-SOS	C2-SOS	C3-SOS	C4-SOS	C5-SOS	C6-SOS	Total
Class 1	49	79	2	5900	21	0	6051
Class 2	7874	5519	3	216	1416	569	15597
Class 3	35	1608	2	11	3	5	1664
Class 4	13	3	4905	5	117	0	5043
Class 5	2206	1	1	14	10890	0	13112
Total	10177	7210	4913	6146	12447	574	41467

Table 8: Cross-tabulation of the SOS classification (in column) and DSOM classification (in row).

This table presents how a macro-class obtained using one algorithm (D-SOM if one reads the table row by row) is split in several parts, each one corresponding to a macro-class produced by the other one (SOS) and whose frequencies are the figures read on the same line. If both algorithms produce very similar classification each class obtained with the first algorithm is mainly observed in another class from the second one: most of the observations figuring as the total of a row are concentrated in one cell. The number identifying a class being attributed randomly by the computer, most of the time it will not be situated on the diagonal. This is the case, in table 8, for classes D-SOM 1, 3 and 4:

- the 6051 observations of D-SOM class 1 are mainly (5900) in SOS class 4 and residual numbers in the five other ones (6 classes are obtained with SOS);

- on the 1664 observations grouped in the D-SOM class 3, 1608 belong to the SOS class 2;

- it is quite the same with the 5043 observations of the D-SOM class 4: 4905 are in the SOS class 3.

The main difference between the results of the algorithms in question here is to be found in the contents of the classes 2 and 5 (D-SOM):

- the main part (50.5%) of D-SOM class 2 is also in the SOS class 1 with the same characteristics (high activity, wages higher than the average of the class);

- 35.4 % are in restricted activity (in hours per week or in number of weeks);

- 9.1 % of D-SOM class 2 belong to the SOS class 5 (longer seniority);

- the D-SOM class 5 is split in two parts: 16.8 % belonging to the SOS class 1 (those with higher wages and short seniority) and the main part (83.8 %) found in the SOS class 5 (very high seniority and wages above the average).

This difference has a major importance because the clear differentiation between a class with precarious jobs, weak wages and short seniority and what we called the class of good jobs is not visible in the classification produced by SOS. Whereas, the D-SOM algorithm led to a satisfactory interpretation of a characteristic of the class of good jobs: faster career progression is achieved by a change of job and therefore lower seniority where wages are higher. The relationship between two key variables to build the classification is not visible in the result obtained with SOS and is what ultimately determines the choice of D-SOM. So this implies that the main problem observed is due to this exchange between the two classes of high activity: individuals with relatively low wages combined with high seniority have been transferred to the other ray.

5. Visualization of the 5 macro-classes in D-SOM

The comparison between the three topologies lead us to only consider the D-SOM map and we will concentrate the rest of the analysis on it. One can find the description of the 5 classes in Section 4.1.

The results of D-SOM can be displayed in several ways, and we present some of them in this section.

Figure 5 shows the 41467 couples (year, individual) classified into 40 classes, and 5 disconnected macro-classes (the rows of the figure). In each class, we draw the stacking of individual lines obtained by connecting individual values (standardized) for the eight variables used for classification. It is a visual tool used to verify the homogeneity of the classes. As each class is very homogeneous, this figure 5 is very close to the figure 3-c, which represents the code-vectors.
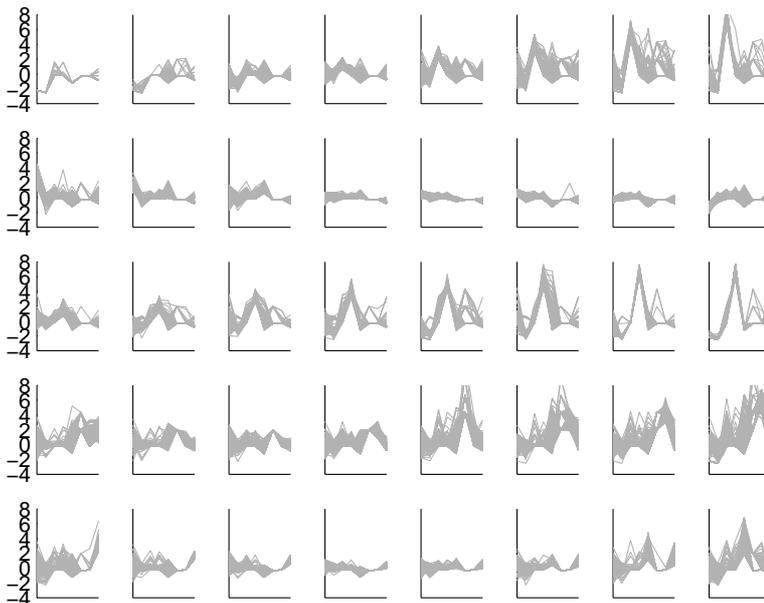
Figure 5: Contents of the classes in the D-SOM map. For each subplot, in abscissa we find the features number and in ordinate the standardized values of the features, all members of the classes are represented by a line.

Figure 6 contains five subplots which present the evolution of the code-vectors along a macro-class from unit one to unit eight. All the variables are centered and reduced and are drawn on the same scale $[-4, 8]$. The major characteristic of each class can be seen by observing the noticeable variation of one (or two) of the variables used for classification, increasing or decreasing when going from the first to the last unit.

More precisely, Figure 6 shows that:

- Class 1 has a specific growth of the number of weeks of unemployment;

- Class 2 shows very weak salary and seniority while the number of hours of work is decreasing from unit 1 to unit 8;

- Class 3 is characterized by the number of weeks out of the market;

- Class 4 shows the importance of the two variables measuring the practice of two jobs or more;

- Class 5 combines a very high seniority and wages very superior to the sample average, but from the first to the last unit of this class, they change in exactly opposite directions.

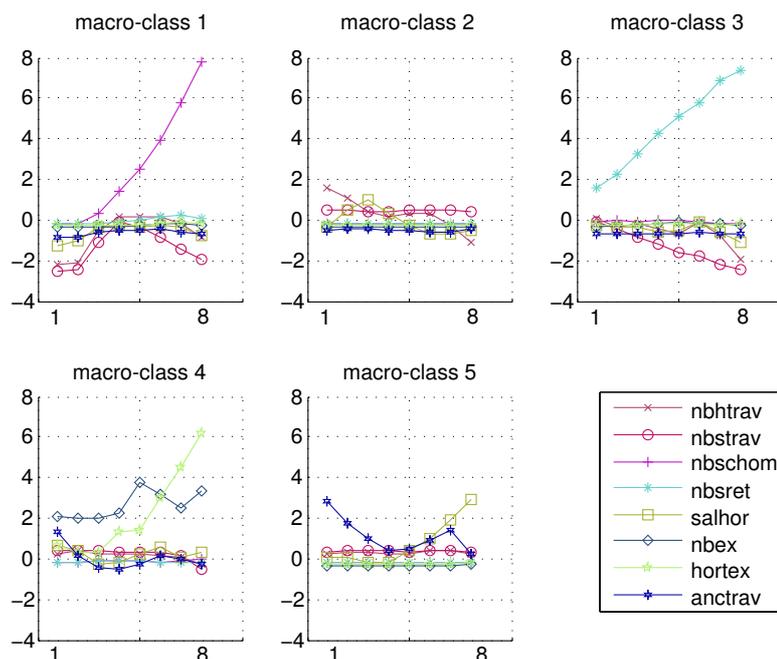Figure 6: Multivariate profiles of the different macro-classes. For each subplot, in abscissa we find the unit number (inside the macro-class) and in ordinate the standardized values of the features for the codebook of each unit.

Figure 7 presents the 8 variables on the whole D-SOM map, with 5 macro-classes of 8 units each one. It is a representation which is the dual one of Figure 6: for each variable used in the classification, we see the extent to which a class is strongly influenced by it. For each variable, the 5 macro-classes are represented as stacked rows, showing the mean values computed at the unit level, using a color code from black for lowest value until white for the highest. It is a visual tool to compare the five classes through the importance of each variable.

All these representations confirm the descriptions of the 5 macro-classes we did in previous section.

6. Crossing the classification obtained by D-SOM with exogeneous qualitative variables

6.1. Two periods

We have already indicated that the two periods were very similar (see Table 5), but some changes must be reported (see Table 9).

- Significant reduction in the proportion of fully unemployed or employed a small part of the year (class 1), 18.75 to 10.78%

- Increase of the same order from the good jobs (Class 5) 27.18 to 35.69%.
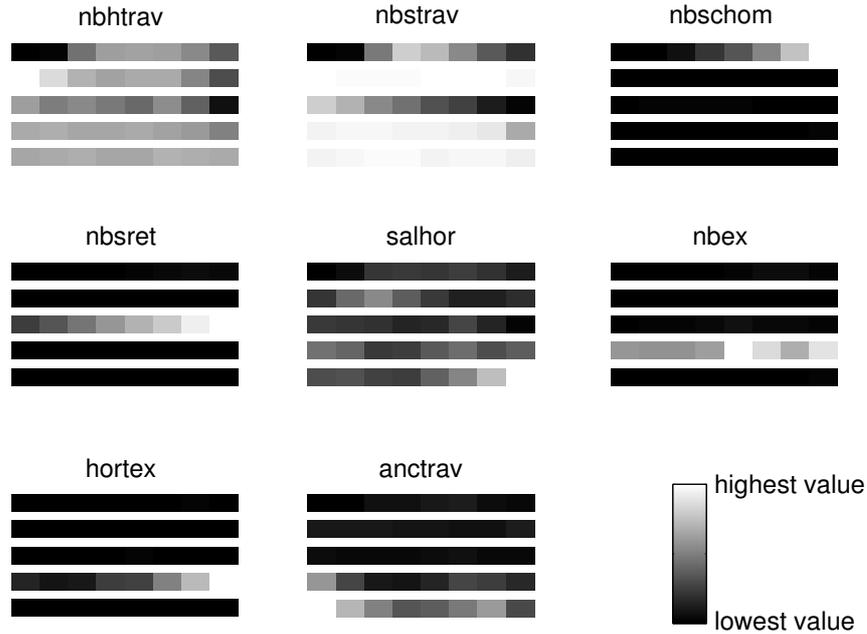
13

Figure 7: 8 variables on the whole D-SOM map, with 5 macro-classes of 8 units each one

Classes	1984-92	1993-2003
Class 1	18.75	10.78
Class 2	37.22	37.98
Class 3	4.90	3.20
Class 4	11.95	12.35
Class 5	27.18	35.69
Sample	100	100

Table 9: Distribution of the 5 classes for both periods (in %).

This is consistent with the overall finding made by all observers of the American macro-economy by comparing the late 80s and the 90s: the increased flexibility of the labor market results in a reduction in the overall unemployment and in an increase in various forms of employment.

According to Table 10 other categories vary slightly from one period to the other (the number of hours worked per week and number of weeks worked in the year, for example) for three active classes (Classes 2, 4 and 5). You still have to observe the growth of real wages for the class of good jobs (from 15.02 dollars per labor hour to 21.11 \$ / h), while for the class of precarious jobs hourly wage ranges from 10.74 to 12.99 \$ / h.

The resulting classes can be better defined by comparing the distribution of the sample according to the terms of age and gender.

6.2. By gender

According to table 11 for the entire observation period women are relatively more likely to be unemployed (more than 1 / 5) than men (1 / 13) and in return men are more often in precarious employment or perform two jobs simultaneously. In the second period, the percentage of women in stable jobs is growing.

6.3. Age groups

The analysis of the results with respect to the age groups are given in Table 12. Most of the under than 30 are concentrated in two classes, Class 2 (48.37%) and Class 1 (19.34%), while the middle-aged (30 to 45) individuals are more present in the class of precarious activities (Class 2)

	Class 1	Class 2	Class 3	Class 4	Class 5
Period 1					
1984-92	3717	7378	971	2370	5389
Period 2					
1993-2003	2334	8219	693	2673	7723
nbhtrav	13.53	**41.99**	27.03	40.77	41.76
	10.44	**42.80**	30.15	39.76	42.00
nbstrav	10.20	**48.69**	18.57	47.27	47.21
	8.73	**49.62**	26.01	44.95	48.57
nbschom	**7.65**	0.14	0.88	0.42	0.12
	5.61	0.08	0.52	0.50	0.09
nbsret	0.35	0.06	**31.73**	0.15	0.02
	0.52	0.17	**27.79**	0.53	0.22
salhor	3.44	10.74	6.43	11.75	**15.02**
	3.24	12.99	9.18	16.73	**21.11**
nbex	0.02	0.00	0.03	**1.13**	0.00
	0.02	0.00	0.06	**1.27**	0.01
hortex	4.58	0.10	4.45	**344.76**	0.22
	4.90	0.01	8.69	**559.24**	1.51
anctrav	11.97	35.06	15.33	84.87	**184.47**
	9.95	35.26	20.47	101.66	**215.39**

Table 10: Mean values for each variable by macro-class, for periods 1 and 2 (one above the other), the figures in bold are the maximum values for each row, the figures below the class names are the class sizes.

By gender	Class 1	Class 2	Class 3	Class 4	Class 5
	Low	Low skill	Temporary	Two	Good
	activity	precarious	Withdrawal	jobs	jobs
Whole Sample 1					
Men	7.64	40.71	1.64	15.29	34.70
Women	22.32	34.17	6.65	8.68	28.19
Period 1					
Men	8.80	40.71	1.50	16.15	32.77
Women	29.22	33.51	8.50	7.51	21.26
Period 2					
Men	6.56	40.72	1.77	14.54	16.13
Women	15.69	34.79	4.86	9.81	34.85

Table 11: Structure in percentage by gender of the macro-classes for the whole sample and by period.

and more than 45 predominate in Class 5. This is not surprising, but it reinforces the credibility of the obtained classification.

By age	Class 1	Class 2	Class 3	Class 4	Class 5
	Low	Low skill	Temporary	Two	Good
	activity	precarious	Withdrawal	jobs	jobs
less than 30, (8.66 %)	19.34	48.37	8.44	11.54	12.32
from 30 to 45, (60.64 %)	12.81	39.06	3.97	12.80	31.36
more than 45, (30.70 %)	16.78	31.72	2.84	11.08	37.58

Table 12: Structure by age group of the macro-classes.

Other qualitative variables such as skill and branch of work could be crossed with the obtained classification in order to better describe the heterogeneity of subsets that are the real labor market.

The quality of this information in PSID has deteriorated significantly in period 2 (inconsistencies between variables, updates not made from one period to another, notably in case of loss or change of job...). This degradation makes it very uncertain interpretation of crossings obtained with subsets of the market. We choose not to present these results.

7. Transitions between D-SOM macro-classes, empirical and limit distributions.

In order to study the trajectories followed by individuals over each period, we consider that the successive situations are observations of a Markov chain.

Let us recall some elements of finite Markov chain theory (see for example [17] or [18]).

If $\{1, \ldots, K\}$ is the set of possibles states for a discrete process $X(t)$ (here $K = 5$ and $X(t)$ is the class number of an observation at time t), the transition matrix Π is a $K \times K$-matrix, with

$$\Pi(i,j) = \mathbb{P}(X(t+1) = j \mid X(t) = i). \tag{11}$$

Note that $0 \leq \Pi(i,j) \leq 1, \forall i, \forall j$ and that $\sum_j Pi(i,j) = 1, \forall i$. Each row of the transition matrix sum to 1, and it is the probability distribution of the next state, conditionally to a starting position. One assumes that $\Pi(i,j)$ does not depend on t. Then such a discrete process $X(t)$ is a finite Markov chain, defined by its transition matrix Π.

We estimate the transitions probabilities $\Pi(i,j)$ by computing the empirical frequencies to be in Class j at the next time, belonging in Class i at present [4]. See in Table 13 the matrices for both periods.

Period 1 1984-92	Class 1 3717	Class 2 7378	Class 3 971	Class 4 2370	Class 5 5389
Class 1	**56.51**	24.44	11.,73	4.11	3.69
Class 2	8.44	**65.46**	3.62	7.98	14.51
Class 3	30.19	**43.01**	18.00	7.19	1.61
Class 4	3.74	28.14	2.29	**50.76**	15.08
Class 5	3.75	7.92	1.21	5.92	**81.21**
Period 2 1993-2003	Class 1 2334	Class 2 8219	Class 3 693	Class 4 2673	Class 5 7723
Class 1	**57.67**	26.77	8.35	3.97	3.23
Class 2	5.08	**67.66**	2.15	9.55	15.57
Class 3	19.85	**52.73**	13.18	7.51	6.74
Class 4	2.79	29.44	1.23	**46.81**	19.74
Class 5	2.34	13.98	0.96	6.47	**76.26**

Table 13: Transition matrices, period 1 and period 2, the figures below the class names are the class sizes, other values are expressed as percentages, values in bold are maxima of the row.

The study of situations on the labor market can be realized in a dynamic sense, since the successive positions of each individual have been observed and used to construct the classification.

The most visible result is that the major part of a class has not moved between year t and year $t + 2$: this is observed since the diagonal entries are the maximum of the line and even greater than 50, except for Class 3 and Class 4 in period 2.

Some global results can be pointed:

- the transition from unemployment is mainly towards unemployment itself and towards precarious situations of Class 2 in a lower proportion;

- the transition from macro-class 2 towards "good jobs" is slightly greater than the contrary (moves from Class 5 towards Class 2);

---

[4]The data are observed each two years

- the most stable class over both periods is Class 5, the one with "goods jobs" , but it is less stable in period 2.

These findings can be confronted with the economic policies that characterize these two periods. Since the 90s a large number of publications dealing with the labor market, were devoted to the theme of flexibility as the preferred means of return to full employment and growth without inflation. This is particularly highlighted by the distinction between job stability, throughout the year, and employment flexibility: the worker is busy most of the year but through changes between successive jobs short in duration. What we see with the Table 13 is a reduction of the probability of maintaining a stable and well-paid employment in average (class 5), while the probability of remaining unemployed (class 1) or in precarious employment (class 2) has increased slightly in the period 2. One may say that the new form of employment named *flexibility* seems to have very weak effects, except for the reduction of the stability of Class 5.

It is interesting to study these transitions for women, in order to see if these global results are verified in a gender perspective. See Table 14.

Period 1	Class 1	Class 2	Class 3	Class 4	Class 5
1984-92	2058	2599	692	594	1847
Class 1	**62.80**	19.38	13.57	2.64	1.62
Class 2	11.56	**60.38**	6.43	6.86	14.77
Class 3	31.41	**40.90**	19.70	6.46	1.54
Class 4	4.55	32.71	5.64	**41.94**	15.16
Class 5	4.41	7.24	2.75	4.81	**80.00**
Period 2	Class 1	Class 2	Class 3	Class 4	Class 5
1993-2003	1270	3230	355	7963	2694
Class 1	**60.30**	23.87	9.99	3.16	2.68
Class 2	7.46	**65.83**	3.55	8.18	14.98
Class 3	20.62	**51.16**	13.91	7.97	6.33
Class 4	3.69	33.95	2.07	**41.87**	18.43
Class 5	3.11	16.11	1.36	5.88	**73.53**

Table 14: Transition matrices for women, period 1 and period 2, the figures below the class names are the class sizes, other values are expressed as percentages, values in bold are maxima.

We can observe that women stay longer in Class 2 (that is the precarious jobs class) in period 2 than in period 1. Conversely, as the mean is almost constant, it means that men more often leave precarious jobs to get Class 5 (goods jobs).

At the same time, in period 2, women are less often in Class 5 (goods jobs). When they are initially out of the market (Class 3), they stay shorter in class 3 (withdrawal) and they change more often towards Class 2 and less often towards to class 1 (unemployment).

For each period, we can compare the observed distributions of individuals across the five macroclasses (average for 4 transitions over the first period and 5 transitions over the second one) to the theoretical limit distributions, computed under the hypothesis that everything in the environment stays unchanged during the period.

The limit distribution is estimated by iterating the transition matrix. As shown by Markov chain theory ([17] or [18]), the powers of the transition matrix converge [5], to a matrix where all rows are equal to the limit distribution. So this limit distribution does not depend anymore on the starting value.

---

[5]when the Markov chain is irreducible, i.e. when all the probabilities to go from i to j, in every number of steps, are strictly positive

The empirical and theoretical distributions are displayed in Table 15. We see that there is a change between periods 1 and 2. The theoretical and observed distributions are closer in period 2 than in period 1. This indicates that the system has become more stable, i.e. the successive distributions are approximately the same during period 2.

	C1	C2	C3	C4	C5
Empirical distribution (first period)	0.19	0.37	0.05	0.12	**0.27**
Limit distribution (first period)	0.14	0.33	0.04	0.12	**0.37**
Empirical distribution (second period)	0.11	0.38	0.03	0.12	**0.36**
Limit distribution (second period)	0.08	0.34	0.02	0.13	**0.42**

Table 15: Empirical and limit distributions, period 1 and 2.

We observe that the previsions are quite good for classes 1 to 4, but that there is a large discrepancy for class 5. That fact may suggest that the hypothesis of the stationarity during each period is not totally founded. However we see that the drifts are in the same direction.

8. Discussion and conclusion

8.1. Results on SOM alternative topologies

This paper has investigated several alternative topologies to the classical sheets used with the SOM algorithm. To do so, the SOM algorithm was slightly generalized to deals with arbitrary topology defined by un-oriented graphs. In this setting, two particulars topologies where advocated since they present natural interest in terms of analysis and interpretations, one based on a star shaped graph called SOS and a second one based on distinct strings called "D-SOM". These two topologies were compared with a classical one on an economical dataset and the "D-SOM" was found to be relevant and to ease the interpretation of the produced projection and quantization. Furthermore, the two levels of analysis offered by these two topologies that we call the macro class level and the class level was also interesting for interpreting the projection in our case study. They are however remaining points to be tackled, in particular with respect to topology selection. This difficult problem, was addressed in this paper using prior knowledge from the field study and by the calculation of projection quality indicators based quantization error. But such indicators suffer from several disadvantages, they are in particular sensitive to the elasticity of the topology *i.e.* a topology with less edges will be favoured by such indicators. This may be an interesting path for further works together with the analysis of new type of topology.

8.2. Discussion about some recent publications on the issue of the functioning of the real labor market

Some recent publications deal with the labor market functioning: see for example [19], [20], [21], [22], [23], [24], [25]. The main objection we can do to these publications is methodological: they use a macroeconomic approach to explain a behavior of the individual present on a single market and following a specified dynamics and they ignore difference due to gender. An accurate treatment of this behavior cannot be obtained through macroeconomic studies, while the authors want to check a real heterogeneity in the labor market. This is the case, for instance, in studies initiated by the European Union on the quality of employment, in the spirit of the Laeken indicators ([19], [20]).

The reduction of individual information by aggregation procedures and the use of specific statistical treatments (Common Factor Model, [21]) lead to results as the so-called "trend improvement in quality of employment" like in most institutional publications of the European Union. This appears to be in complete contradiction with most of real labor markets, in other words the idea of quality has been reduced to a single dimension ("to be occupied over the whole year" for instance) instead of a multidimensional concept. In real situations, the quality of a job cannot be enhanced over all the dimensions simultaneously, as we observed for the PSID data.

The irrelevance is much worse if we refer to the works based on the notion of a representative agent [26] (chapter 2, pp. 37-90), where all individuals are supposed to behave like a typical agent. Thus excluding by construction the heterogeneity (see for example some recent works as [22], [23], [24]).

At the opposite of the last ones, interesting results are obtained in studies conducted at a microeconomic level ([25]) to highlight the question of quality of jobs and its evolution over time.

8.3. Conclusion: Strong results obtained with the classification

The results can be assessed in three perspectives. The first one is the question of the heterogeneity of the actual labor market. Our approach challenges the standard paradigm by providing particularly clear results on a real labor market. In fact what is called the labor market is a set of components of very different qualities in the sense that they can be completely opposed in some dimensions. This is very clear using simple statistic indicators applied to the eight quantitative variables used by the algorithm. The five main classes are well contrasted by means of two or more of these variables. The titles given to these classes are a practical expression of the obtained multidimensional differentiation . As we say before, this differentiation cannot be represented by some unique scaled variable (score). These components are the permanent structure of the market even if discrepancies may occur over the time, each component having its own dynamics.

Another perspective is that of the dynamics of individual situations as it can be represented using the panel structure of the data. This is obtained with the transition matrices built with the classes. For each individual, classification provides the class he belongs to for each observed year.

The observation of empirical probabilities of transition between classes presented in table 13 shows that the change possibility strongly depends on the present situation:

- Firstly the probability of remaining in the class of "good jobs" is 4 out of 5 chances in the first period and 3 out of 4 chances for the second one

- Secondly, when someone is in a class of low quality jobs he/she has 2 chances out of 3 to stay in it against 1 chance out of 6 or 7 to move to the class of good jobs. This is true for both periods.

This analysis is not the explanatory model of the job market: this one remains to be built. It is simply a step in this construction. Contrary to the idea often put forward, the transfer of the unemployed to new forms of precarious employment (also called "flexibility") is not clearly observed. The most visible movements between classes, considering the whole sample, do not verify this assertion and unemployment remains a relatively stable situation, in the second period as in the first one. In the second period there seems to be a much less frequent move back to unemployment after a temporary withdrawal from the labor market.

The last perspective is that of gender. The share of those who remain in a status of low activity (activity very restricted or unemployment, Class 1) is very stable and represents a mean of more than 56 %, for both genders, see Tables 13, 14.

But it is not the same for those who are in good jobs (Class 5), see Tables 13, 14. In fact, the overall results mask the real evolution of women's situation on the whole period. For the complete sample (men and women), the probability to remain in good jobs decreases slightly (from 81 % to 76 %) from the first period to the second. In fact this average is misleading because simultaneously the probability remains equal for men (about 82%) and decreases for women (from 80 % to 73 %). This indicates a deepening of the gap between men and women in terms of maintaining into the category of good jobs.

In terms of wages, the results are less clear given the arbitration, already mentioned, between job tenure and the level of wages. For high values of the seniority, the level of remuneration for men

is significantly higher than that of women (about 50 %), both in the first period and the second period. For low values of seniority (high levels of wages) the difference is in favor of men but in a much smaller proportion.

8.4. Future work

As we mention previously, unfortunately, in PSID survey, poor quality of data on qualifications and location of employment in terms of branches of industry (very imprecise classifications, lack of updated information in each survey wave) does not allow to specify the content of classes in these areas. To overcome these drawbacks, we intend to repeat this classification technique and the exploitation of results with another large database, the European Community Household Panel (SILC) built over 27 countries, and now available (on demand) on line on the Eurostat website [6].

Moreover, according to some recent studies [27], we plan to improve the theoretical study of the heterogeneous labor market, by modeling the transitions between segments over time using some econometric tools, like a dynamic logit model, with measured factors and controlling unobserved heterogeneity in a Hidden Markov Chain modelization.

References

[1] E. Côme, M. Cottrell, P. Gaubert, Analysis of Professional Trajectories using Disconnected Self-Organizing Maps, in: Proceedings of WSOM2012, Vol. 198 of Advances in Intelligent System and Computing, 2012, pp. 295–304.

[2] P. Cahuc, A. Zylberberg, Labor Economics, The MIT Press, 2004.

[3] D. Mortensen, C. Pissarides, New developments in model of search in the labor market, in: A. O., C. D. (Eds.), Handbook of Labor Economics, Vol. 3B of Lecture Notes in Computer Science, Elsevier, 1999, pp. 2567–2627.

[4] T. Le Barbanchon, F. Malherbet, An anatomy of the French labour market: Country case study on labour market segmentation, Tech. Rep. 142, ILO, Genève (2013).
URL http://digitalcommons.ilr.cornell.edu/intl/279/

[5] W. Eichhorst, V. Tobsch, Has atypical work become typical in Germany?: Country case study on labour market segmentation, Tech. Rep. 145, ILO, Genève (2013).
URL http://digitalcommons.ilr.cornell.edu/intl/276/

[6] E. Côme, M. Cottrell, M. Verleysen, J. Lacaille, Self Organizing Star (SOS) for health monitoring, in: Proceedings of the European Symposium on Artificial Neural Networks, 2010, pp. 1341–1346.

[7] T. Kohonen, Self-Organizing Maps, Information sciences, Springer, 1995.

[8] M. Cottrell, J.-C. Fort, G. Pagès, Theoretical aspects of the SOM algorithm, Neurocomputing 21 (1998) 119–138.

[9] A. Barsi, Neural Self-Organization using graphs, in: Proceedings of the Third International Conference Machine Learning and Data Mining in Pattern Recognition, Leipzing (Germany), Vol. 2734 of Lecture Notes in Artificial Intelligence, Springer, 2003, pp. 343–352.

[10] J. Pakkanen, J. Iivarinen, E. Oja, The evolving tree - analysis and applications, IEEE Transactions on Neural Networks 17 (2006) 591–603.

[11] C. Godsil, G. Royle, Algebraic Graph Theory, Springer, 2001.

[12] M. Resta, Assessing the efficiency of health care providers: a som perspective, in: Advances in Self-Organizing Maps, Vol. 6731 of Lecture Notes in Computer Science, 2011, pp. 30–39.

---

[6]http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database

[13] A. Ultsch, H. Siemon, Kohonen's self-organizing features maps for exploratory data analysis, in: Proceedings of International Neural Networks Conference, Kluwer, 1990, pp. 305–308.

[14] J. Vesanto, J. Himberg, E. Alhoniemi, J. Parhankangas, Som toolbox for matlab 5, Tech. Rep. A57, Helsinki University of Technilogy, Laboratory of Computer and Information Science (2000).

[15] J. Vesanto, E. Alhoniemi, Clustering of the self-organizing map, IEEE Transactions on Neural Network 11 (2000) 586–600.

[16] H. Ritter, T. Martinetz, K. Schulten, Neural Computation and Self-Organizing Maps, Addison-Wesley, 1992.

[17] P. Hoel, S. Port, C. Stone, Introduction to Stochastic Processes, Houghton Mifflin Company, 1987.

[18] P. Billingsley, Probability and Measure, John Wiley and Sons, 1979.

[19] F. Fremigacci, Y. L'Horty, La qualité de l'emploi en France: tendance et cycle, Tech. Rep. 51, Centre d'Etudes de l'Emploi (2005).
URL http://www.cee-recherche.fr/publications/document-de-travail/la-qualite-de-lemploi-en-france-tendance-et-cycle

[20] L. Davoine, E. C., La qualité de l'emploi en Europe : une approche dynamique, Economie et Statistique 410 (2008) 47–69.

[21] M. Forni, M. Hallin, M. Lippi, L. Reichlin, The generalized dynamic factor model consistency and rates, Journal of Econometrics 119 (2004) 231–255.

[22] K. Krishna, M. A., C. Yavas, Trade with labor market distortions, Tech. Rep. 9086, NBER (2002).
URL http://www.nber.org/papers/w9086.pdf

[23] M. Pries, Heterogeneity and labor market volatility in matching models, Review of Economic Dynamics 11 (2011) 664–678.

[24] S. Danthine, Two-sided search, heterogeneous skills and labor market performance, Tech. Rep. 1572, IZA (2005).
URL http://ftp.iza.org/dp1572.pdf

[25] EUROFOUND, Employment polarisation and job quality in the crisis: European jobs monitor 2013, Tech. rep., EUROFOUND (2013).
URL http://www.eurofound.europa.eu/pubdocs/2013/04/en/1/EF1304EN.pdf

[26] O. Blanchard, S. Fisher, Lectures on Macroeconomics, 5th Edition, MIT Press, Cambridge (MA), 92.

[27] F. Bartolucci, A. Farcomeni, A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure, Journal of American Statistical Association 104 (2009) 816–831.