French Universities – A Melting Pot or a Hotbed of Social Segregation? A Measure of Polarisation within the French University System (2007-2015)

Romain Avouac* and Hugo Harari-Kermadec**

Abstract – Despite changes recently introduced within higher education (cluster-building policies, the influence of university rankings, etc.) which may have fueled fears of a disparity between a pocket of world-class universities and a vast group of second-tier universities, relatively few quantitative studies exist to examine this matter. Using data from the *Système d'information sur le suivi de l'étudiant* (SISE), an information system for monitoring students university enrolments in France, we provide an exhaustive overview of the university landscape taking into account the capital held by various student populations. We then apply measures of segregation and polarisation to analyse the change in heterogeneity, which increased between 2007 and 2015. Lastly, we link this polarisation to the measures implemented at domestic and international level (*Initiatives d'Excellence* in France, and university rankings globally) which shape the foundations for globalisation among universities since the mid-2000s.

*ENS Paris-Saclay and ENSAE (romain.avouac@ensae.fr); **IDHES, ENS Paris-Saclay (hugo.harari@ens-paris-saclay.fr)

The last decade has been characterised by growing globalisation in higher education, symbolised in the French public debate by the emergence of the Academic Ranking of World Universities (ARWU), or Shanghai Ranking. As part of this trend, exposure of select institutions on the international stage has been pursued as a priority by public policy through competitive procedures for the allocation of major funds, beginning with the *Initiatives d'excellence* (IDEX). A large number of institutions then took part in cluster-forming projects in an attempt to reach "critical mass" – the level required for inclusion on the international ranking tables or to rise up those tables. This raises the question of the "performativity" (Espeland & Sauder, 2007; Paradeise & Thoenig, 2015) of these measures: do international rankings actually influence the reality they are supposed to describe?

There is still no guarantee that excellence, within the global arena, of a few universities which top the rankings in France would systematically transfer to excellence within a system of higher education made generally accessible. Amsler & Bolsmann (2012) argue that, in the United Kingdom, international rankings promote a more blinkered view of education, focusing public debate on elite institutions while consigning democratisation concerns to the back-burner. Policies and the allocation of public funds would therefore tend to concentrate on a select few institutions, which risks contributing to the spread of two-tier education systems in which a small number of world-class universities receive a significant proportion of resources and a vast group of second-tier universities share what resources remain (Van Parijs, 2009). Merton (1968) refers to the "Matthew effect" to describe the cumulative effect of academic reputation within research circles.

Universities have been the main place of mass access to higher education in France in recent decades. The proportion of children of blue-collar workers who graduated from university rose from 6% in 1984 to 24% in 2009 (Peugny, 2013). As a result, we now see different types of students, from different backgrounds, rubbing shoulders at universities (Brinbaum *et al.*, 2018). Although the university sector is stratified (Frouillou, 2017; Convert 2006), we might expect that it allows students from working class backgrounds to continue accumulating academic capital but also to partially compensate for the deficit of social capital from the initial socialisation, thanks to the peers they meet at university (Truong, 2015). On the

other hand, with prestigious higher education institutions already being broadly off limits to the working classes (Ichou & Vallet, 2013), a polarised university system would, in keeping with the educational experience, lead to heavily segregated access to adult life for young people and therefore reinforce the determinism inherent in academic trajectories and subsequent professional trajectories.

But are we really witnessing polarisation within the French university system, in terms of social composition? To provide answers to this question, we study the changes in the social composition of student populations, using data available between 2007 and 2015 from the comprehensive database of enrolments in French universities (SISE[1]).

In a first step, we provide a comprehensive statistical view of the French university landscape by looking at the social and academic characteristics of the student populations at various institutions. To do this, we use a principal component analysis (PCA) in order to summarise the information in the initial data. The hierarchical ascendant classification method (HAC) then allows us to build a university typology, which we use to analyse the level of heterogeneity within a cross section of institutions. We highlight a socially heterogeneous and hierarchical university landscape, corroborating the results that Brusadelli & Lebaron (2012) obtain using a similar methodology.

In a second stage, we develop an analysis that aims to quantify the changes in the heterogeneity of the social composition of higher education institutions between 2007 and 2015. We first introduce a measure of polarisation drawn from the econometric literature (Esteban & Ray, 1994), which indicates an increase in social polarisation during the period studied. We compare these results to the ones obtained using a normalised entropy index typically used in the literature on academic segregation.

Lastly, we link social heterogeneity, which is characteristic of French higher education, to national measures (IDEX, university groups) and international measures (university rankings) which provide the foundations for globalisation at university level. These measures target institutions with the most privileged student populations. Additional (public or private) funding allocated *via* excellence-related

---

1. *Developed by the SIES (*Sous-direction des Systèmes d'Information et des Études Statistiques*), the statistical information and studies department of the French Ministry of Higher Education and Research.*

measures is therefore channelled towards those students who enjoy the most capital, of all varieties. This finding raises the question of resource redistribution *via* higher education policy.

1. A Socially Heterogeneous University Landscape

While the massification of higher education cannot be denied, the democratisation of access to all higher education institutions raises questions. To address the gap between these two dynamics, we propose using a concept of social polarisation. This requires us to first define the social composition of universities.

For each student enrolled at a French university, the comprehensive databases of enrolments at French universities (SISE for enrolments) available between 2007 and 2015 include the institution at which the student is enrolled, socio-demographic information, and data concerning the student's previous and current education (Box 1). These data allow us to characterise the composition of the student population at each institution. For example, for each university, the proportion of students for whom the reference parent is a senior executive is calculated. The tables that include qualitative variables describing students therefore lead to aggregate tables that include quantitative variables describing universities.

1.1. Principal Component Analysis as a Summary Tool

To depict the distribution of the social and academic characteristics of the student populations, we apply a principal component analysis (PCA) to the transformed data. This method allows downscaling the dimension of the problem, which is initially equal to the number of variables included to define the social composition of the universities' student populations. This approach has notably been used by authors seeking to produce socio-economic status indices based on many variables that imperfectly describe social background (Vyas & Kumaranayake, 2006). More recently, Rocher (2016) notes that "classifications of professions and social categories are limited when it comes to describing the social tonality of groups of pupils (classes, institutions, academies)" (translated from Rocher, 2016, p. 16) and therefore also uses factorial methods to construct a "student social position index".

Once aggregated, several variables at student level can be used to describe the social composition of universities. The socio-professional categories (CS) of students' parents are a significant marker of a student's social background and the capital, of multiple varieties, that the parents can pass on to the student, even though these variables do not fully quantify the influence of background upon individual academic trajectories (Boutchénik *et al.*, 2015). Although it is common practice to use the CS of an individual's father to summarise their social background, the predominant role played by the mother's level of education in the level attained by the child (Place & Vincent, 2009) provides an argument for both CS to be included. Homogeneity is a concern in our case, as we are restricted by the fact that SISE does not record the CS of the child's second reference parent until 2009. We therefore use the CS of the child's first reference parent to define the social background of students (see Appendix 1 for the classification system selected).

A more detailed description of the composition of universities may result from the inclusion of variables relating to the students' education. In particular, the baccalaureate is a significant social indicator that plays a key role in higher educational trajectories (Duru-Bellat & Kieffer,

---

Box 1 – **The SISE Database**

Since its creation in 1995 by the SIES (see footnote 1), the SISE database has been collecting data on students enrolled at the main universities. A major advantage of this database is its comprehensive nature: SISE records *all* of the (main) enrolments at the institutions surveyed. In 2007, SISE held data on 1.4 million of the 2.2 million students enrolled at a higher education institution, which equates to a coverage rate of around 63%. All university components are considered, including institutes of technology, affiliated engineering schools, remote learning, etc. For each student included in the database, information is available concerning their previous education (type of baccalaureate, year of initial enrolment in higher education, etc.) and current education (description of the institution at which the student has enrolled, the qualification undertaken, the level attained in higher education, etc.), as is socio-demographic information (gender, socio-professional category – CS hereafter – of the reference parent, nationality, etc.) and geographical information (the student's country and municipality of residence, and of their parents, etc.).

After the reference parents' CS have been coded and harmonised (see Appendix 1), the university tables on which our work is based include 81 institutions over the whole period studied, and 18 variables created from the initial individual tables relating to the students.

2008). The student's status as a (higher education) grant holder and the level of the grant also appear to be relevant. However, these variables are only available between 2007 and 2011, which is why they have not been retained in order to maintain a homogeneous approach over the period studied. Further analyses (not provided here) for this shorter period show that their inclusion does not alter the results obtained.

Lastly, we include the student's gender and the cycle of enrolment (Bachelor's degree/Master's degree/doctorate) as supplementary variables in the PCA – i.e. they play no part in the definition of the axes. While there are clear differences in the students' higher educational academic trajectories (Rosenwald, 2006), we do not believe the gender distribution among those students to be an inherent determining factor in the social composition of universities. Similarly, although the level of higher education attained depends on social background (Brinbaum *et al.*, 2018), its inclusion as an active variable could raise the issue of endogeneity given that we are interested in the effect of social background on higher educational trajectories.

1.2. The First Principal Component as a General Index of Social Composition

Ultimately, 13 active variables are used for the PCA. We present only the PCA based on the 2007 data. The axes of this PCA are the reference we use to construct the social composition index and ensure its comparability over time (Box 2).

The "scree plot" of eigenvalues can be used to evaluate the quality of the PCA's reduction of the initial data (see Appendix, Figure A3-I). The application of the "elbow criterion"[2] leads to the selection of the first two principal components (PC) to conduct the analysis. The Kaiser criterion[3] leads to the selection of the third PC. However, unlike the first two axes, the third axis appears to be difficult to interpret and provides only very minimal additional information with respect to the mean of the axes. That is why we limit the analysis to the first factorial plane, i.e. the plane of the first two PCs. The latter represents 64% of the total inertia. The representations obtained for the first two dimensions capture a large proportion of the information in the initial cluster and lend themselves well to interpretation.

The graph of variables (Figure I) allows us to analyse the pattern of correlations between active variables and to interpret the axes of the PCA. The variables which contribute most strongly to the first (horizontal) axis represent the proportion of children of senior executives on the left, and, on the opposite side, the percentages of students holding a

---

2. The PCA arranges the dimensions in decreasing order of inertia. The inertia tends to decrease rapidly for the first few dimensions, then more slowly and evenly for the subsequent, less informative dimensions. The "elbow criterion" involves selection of the dimensions before the break in the inertia gradient.
3. The Kaiser criterion involves selecting total inertia divided by the number of dimensions as the threshold. The dimensions for which inertia is above the mean are then selected and the others omitted.


Box 2 – **Comparability Over Time of the Universities' Social Composition Index**

The first axis of the PCA enables us to arrange institutions by social composition: from populations with a privileged background (on the left of the factorial plane) to populations with a working class background (on the right). A large proportion of the information in the initial cluster is also included in this axis, which makes it more relevant. We therefore select the distribution of the institutions' coordinates on this axis as a general index of the universities' social composition.

Later on in this article, we analyse the change in the polarisation of the distribution of social compositions between 2007 and 2015. This index must be comparable over time to ensure that this analysis is relevant. Although the PCA's pattern of correlations shows a high level of stability over the years – the graph of correlations of the PCA created in 2007 is very similar to those obtained when we create an equivalent PCA based on data from subsequent years – the positions of the variables vary slightly compared to the first two principal components as does the quality of the representation of the first axis. It is therefore not immediately possible to compare the changes in university coordinates over time. To compensate for this, the axes of the PCA in 2007, the only year in which the PCA is carried out, are fixed. The data from subsequent years are then projected onto the 2007 factorial plane as additional individual data. The institutions' coordinates on the first axis correspond to the value of their social composition index over time. This procedure ensures that the axes of the PCA measure the same thing each year and guarantees the comparability of successive index values over time.

A limitation to this analysis is the idiosyncratic role it gives to the reference year – 2007 in this case. Data from subsequent years are analysed with respect to the intrinsic structure of the 2007 data. To ensure that the results obtained are not overwhelmingly dependent upon this selection, we have reproduced the analysis developed in the remainder of this article by making each year of the period studied the respective reference year. The observed upward trend in polarisation is consistent between 2009 and 2015 regardless of the reference year, with greater variability during the first two years of that period.

Figure I – **PCA variables in 2007 (dimensions 1 and 2)**
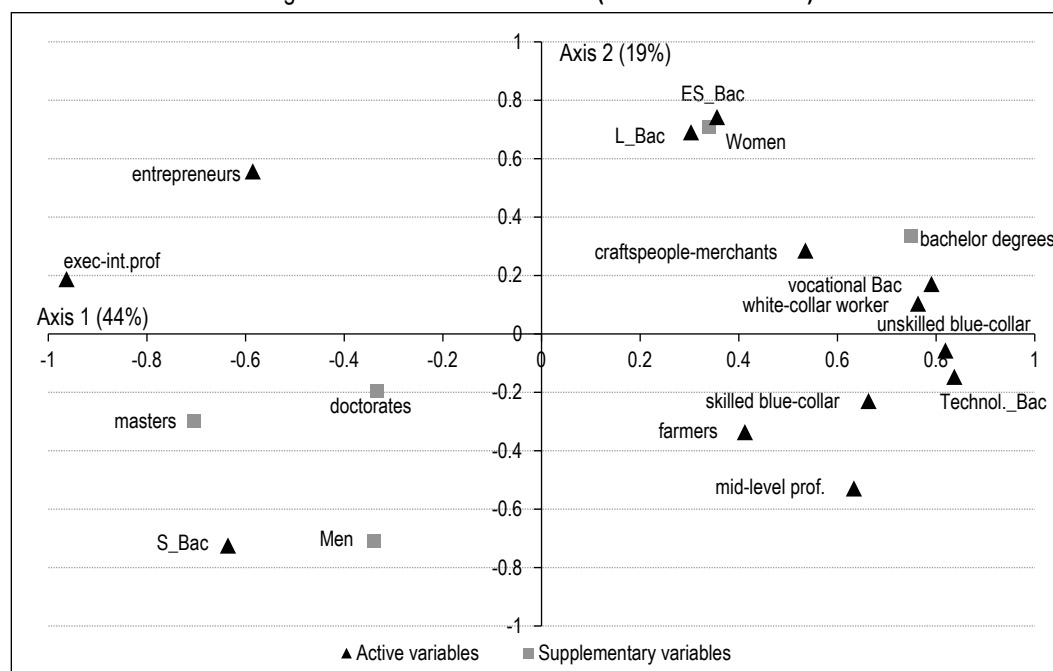
Reading note: The [exec-int.prof.] variable (which, for each institution, gives the percentage of students for whom the reference parent belongs to the CS 'Executives and intellectual professions' – see Appendix) is highly and negatively correlated with the first axis (-0.96) but its correlation with the second axis is very low (+0.19).
Sources and Coverage: MESRI-SIES, SISE; French universities and elite institutions under the authority of the Minister for Higher Education.

vocational or technological baccalaureate, as well as the proportions of children of unskilled blue-collar workers, white-collar workers, and, to a lesser extent, skilled blue-collar workers. This close correlation between the variables on the right of the axis confirms the observation that "students from working class backgrounds […] [are] over-represented in these vocational and technological streams" (translated from Duru-Bellat & Kieffer, 2008, p. 126). Accordingly, axis 1 seems to specifically contrast the most privileged children of workers with the most disadvantaged, which confirms the ongoing substantial differences between these social backgrounds in terms of their trajectories in higher education (Albouy & Tavan, 2007). With a moderate correlation with respect to both axes, children of self-employed parents are somewhat poorly represented in the factorial plane – even if their positions on axis 1 appear to be sociologically relevant. The only variable with a surprising position is the share of students whose reference parent has a mid-level profession: the level of correlation with axis 1 is slightly lower level than for skilled blue-collar workers. This may be due to this category being highly heterogeneous. Ultimately, this very clear distinction between social backgrounds and their characteristics on axis 1 enables us to interpret this as ordering the universities by their respective populations' social composition: those from a privileged background on the left,

and those from a working class background on the right. Lastly, it should be noted that this opposition appears to be highly discriminant within the French university landscape: axis 1 alone accounts for 44% of the information in the initial cluster. This axis therefore appears to be a reliable and relevant measure of university composition in a variety of ways.

1.3. Significant Inequality between Institutions

Although the second (vertical) axis is less straightforward to interpret, it provides useful insights for the analysis. The variables which contribute most significantly to this axis are the proportion of students who hold a Science Baccalaureate (S) at the lower end of the axis, and the proportions of students who hold either a Baccalaureate in Economics and Social Sciences (ES) or a Literary Baccalaureate (L) at the upper end of the axis. For a given social composition, the second axis therefore seems to oppose primarily scientific universities and those that primarily teach humanities. This interpretation is confirmed by the graph of individuals (Figure II) in which primarily humanities-focused universities appear at the top, while primarily science-focused universities, including affiliated engineering schools, appear at the bottom.

The positions of the additional variables on the factorial plane provide additional information

Figure II – **PCA individuals in 2007 (dimensions 1 and 2)**

Sources and Coverage: See Fig. I.

about the arrangement of the correlations. As these variables play no role in the definition of the PCA's axes, their position and any interpretation we may draw from that appear all the more relevant. A first notable point is the close proximity between girls and ES/L Baccalaureate holders, as is the proximity between boys and S Baccalaureate holders, in proportional terms. This proximity confirms the observation made by Rosenwald (2006), among others, that although girls tend to be more successful in their studies no matter their level and the baccalaureate stream pursued, they remain over-represented in predominantly literary streams. In addition, the contrast between the variables indicating the rates of enrolment in Bachelor's degrees (on the right) and those indicating the rates of enrolment in Master's degrees (on the left) is also revealing. The reproduction quality of the variable indicating the rates of doctoral students is poor, which hinders any interpretation. Not all universities have the same ratio of undergraduates to postgraduates, and this ratio is shaped by social factors. Although these variables relate to institutions here and not directly to students, this finding seems to corroborate the notion that not only are children from working class backgrounds less likely to enter higher education in general, but those who do are less likely to attain a Master's degree (Peugny, 2013; Selz & Vallet, 2006).

The positions on the social composition axis are relevant from a sociological perspective. The institutions with the most privileged student compositions tend to be the institutions that can select their students and in some cases even charge high enrolment fees, e.g. the "elite" institutions (Paris Dauphine University, IEP Paris, Observatoire de Paris) and the university-affiliated engineering schools (INP Toulouse, Polytech Grenoble, etc.). The over-representation of Parisian universities is clear given that they are consistently positioned on the left of the axis. In contrast, universities in the French overseas departments are the first to appear from the right on the graph, followed by medium-sized provincial cities, generally in the north (Valenciennes, Artois, Upper Alsace) and south-east (Perpignan, Avignon, Nîmes, Toulon) of France. The positions of the universities on the social composition axis therefore generally match the distribution of living standards within the various regions (as described by Auzet *et al.*, 2007, for example). This corroborates studies demonstrating that much of the inequality between universities can be traced back directly to unequal social distribution within French territories (Nicourd *et al.*, 2011).

Lastly, we observe a high degree of social heterogeneity among the institutions given the distribution of their positions along axis 1 of the

PCA (see Figure II). As no interpretation can be drawn directly from the scale of the PCA, reverting to the initial variables can better illustrate this heterogeneity. If we focus on the endpoints of the social composition axis, we can see, for example, that the proportion of children of senior executives is 66% at Paris Dauphine University and the IEP Paris, which reflects a very strong over-representation compared to the socio-professional structure of the parents' generation (Marchand, 2010). At the other end of the axis, children of senior executives account for only around 20% of students at universities in French overseas departments,[4] and 25% at universities in the north and south-east of mainland France, these universities appearing on the right of the factorial plane. Conversely, the proportion of children of unskilled blue-collar workers is 7% in the universities of Avignon and Nîmes, while these children are virtually absent from a number of institutions in Paris. Similarly, the proportion of students whose parents are skilled blue-collar workers exceeds 20% in the universities of Northern France compared to 2% in the elite institutions in Paris (Paris Dauphine University, IEP Paris, Observatoire de Paris) and 3 to 5% in inner-city Parisian universities (Panthéon-Assas, Paris 7, Paris 5, Paris 4, Panthéon-Sorbonne, etc.).

1.4. A Typology of French Higher Education Institutions

The fact that institutions are highly scattered on the first factorial plane suggests that a typology can be based on the PCA's results. To achieve this, we proceed with a hierarchical ascendant classification (HAC) based on the coordinates of the institutions on the principal components of the PCA.[5]

We retain the first two dimensions of the PCA to construct the typology. Although the first axis appears to be a relevant measure of the social composition of the universities, it does not fully summarise the social heterogeneity seen in the factorial plane. The variable indicating the proportions of Science Baccalaureate holders, for example, appears to be significantly correlated with both axes 1 and 2, which underlines the Science Baccalaureate's role as a social indicator. So, while the vertical axis measures a university's primary discipline first and foremost, it also includes some of the social heterogeneity of the institutions that the horizontal axis fails to register. The dynamics of social differentiation at play in higher education are therefore described more comprehensively by constructing the typology from the first two

dimensions. In addition, we show that the classifications obtained are sociologically relevant (cf. Box 2).

Beyond their descriptive capacity, the typologies that we present will allow us to subsequently apply the polarisation measure of Esteban & Ray (1994) as part of a dynamic analysis. This measure is based on the existence of well-defined groups in the population studied. For the analysis of the changes in polarisation to be justified, the typologies produced must retain their interpretation over the period studied while highlighting a deformation of the groups, i.e. potential changes in the group of certain institutions over the period studied, which are expressed through a shift on the factorial plane. To meet both conditions, we perform the HAC using data from all years simultaneously. In this way, the groups produced keep, necessarily, the same interpretation over the years, and changes of group of certain institutions are observed over the period studied.

The literature offers many criteria to guide the decision on how many groups to select. In practice, this selection is the result of a trade-off between parsimony and homogeneity within the groups, "the principal criterion being that the typology finally selected should be coherent and informative with respect to the research being conducted" (translated from Robette, 2011, p. 19). Following this logic, we choose to focus the analysis on the four-group typology derived from the HAC (Figure III). Broadly speaking, the left side of the plane groups together institutions with a privileged composition – institutions based in Paris for the most part – and universities in large provincial towns and cities. The right side of the plane is populated by institutions with populations from a background ranging from intermediate to working class, most of these institutions being located in average-sized cities and in French overseas departments. Among institutions that have a privileged composition, the classification also draws a distinction between those for which the main field of study is literature or economics and social sciences (at the top of the plane) and those for which it is mainly scientific (at the bottom). Heterogeneity along axis 2 is much more pronounced on the left of the plane, among institutions with a privileged composition, whereas institutions

---

4. We checked that restricting the coverage to metropolitan institutions only did not alter the main results.
5. The factorial and classification methods complement one another. Directly using the principal components to produce the classification rather than the initial data reduces the influence of statistical noise in those data (Husson et al., 2010).

Figure III – **Representation of the four-group, HAC-derived typology**
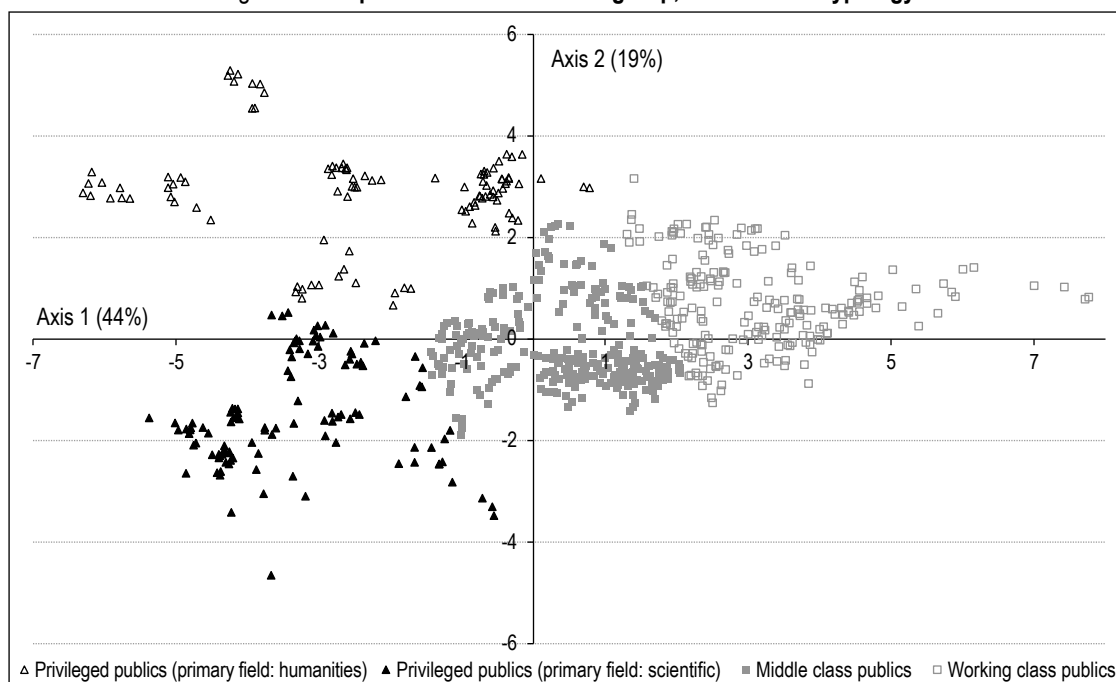
Sources and Coverage: See Fig. I.

with an intermediate or working class student population, on the right of the plane, display significant homogeneity along axis 2. This makes it possible to distinguish, among institutions with a privileged social composition, those that are primarily scientific from the others, while institutions with an intermediate or working class student population do not seem to be marked in terms of field of study. In this respect, it seems relevant to separate these different groups because they may be characterised by distinct polarisation logics. To limit the sensitivity of the results to the selected typology, the following analyses have all been reproduced using the two- and three-group typologies (see Appendix, Figures A3-II and A3-III). Similar results are obtained in all cases.

The relevance of the typology is confirmed by the statistical description of the various groups (see Appendix 2). In terms of the social composition, there is a clear social gradient for each of the determinant variables of axis 1 among the three groups of institutions (privileged, intermediate and working class populations). The proportion of children of senior executives is thus 52-54% in the two privileged population groups, compared to 38% in the intermediate population group and 28% in the working class population group. Conversely, the proportion of children of skilled blue-collar workers does not exceed 6% in institutions with a privileged composition, compared with an average of 13% for institutions with a working class population.

This social gradient can also be seen very clearly when we focus on the distribution of the various levels of study (Bachelor's/Master's/Doctorate) within the groups.

With the exception of vocational and technological baccalaureates, which are mainly social indicators and therefore behave in a similar manner to the variables described above, the analysis of the distribution of the various types of baccalaureate shows the relevance of separating the privileged groups by field of study. The proportion of Science Baccalaureate holders varies by 48 percentage points between both groups with a privileged composition. However, it should be noted that the proportion of Science Baccalaureate holders in the group of institutions with a working class social composition is almost the same as in the group of institutions with a privileged and primarily humanities-focused composition. This observation confirms that the Science Baccalaureate also plays a significant role of social marker, which, among other things, justifies the decision to base the typology on the factorial plane, rather than just axis 1, in order to capture all social differentiation processes. By contrast, ES and L Baccalaureate holders are highly under-represented in the group of socially privileged compositions in which science is the primary field of study and, conversely, over-represented in the institutions in which literature is the primary field of study. Nevertheless, as the heterogeneity of the groups in relation to these types of baccalaureate appears to be less

pronounced, so too appears the social role of these types of baccalaureate, which corroborates the position of the associated variables on the factorial plane. Lastly, the variables indicating the proportion of women and men are in very close proximity to those indicating the proportions of S and ES/L Baccalaureate holders, which is in keeping with the significant differentiation of disciplines by gender already observed.

2. Growing Social Polarisation among French Higher Education Institutions

The first factorial plane obtained from the PCA highlights, statically, a significant level of social heterogeneity at institution level. The literature on recent developments in higher educational policy, including the increased use of quantitative performance indicators, suggests that a shift in polarisation within the higher education system can be observed (Van Parijs, 2009). It is also possible that the democratisation of access to university is accompanied by a segregative dimension, similar to that observed in access to the Baccalaureate (Merle, 2002). Based on these hypotheses, we propose a dynamic analysis using two different indices: a polarisation index derived from the econometric literature, and a multi-group segregation index, typical in the literature on academic segregation.
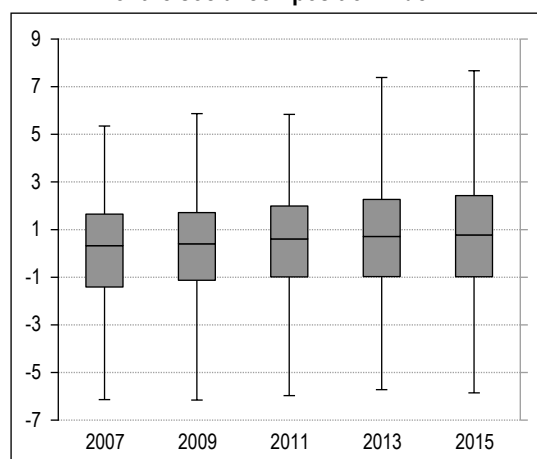
First, we examine the changes in the social composition index from the PCA, between 2007 and 2015 (Figure IV). The advantage of this analysis is that it does not depend on any typology. The standard deviation of the distribution, a measure of the degree of inequality of the distribution of social compositions (McKenzie, 2005), rises by 15% between 2007 and 2015. The interquartile range increases by

11%. Finally, the maximum range increases by 18% between 2007 and 2015. This is the gap between the institutions that are farthest apart – Paris-Dauphine University and the University of French Polynesia throughout the period studied – on the social composition axis.

Our initial analysis shows an increase in inequality in terms of social composition over the course of the period studied. However, this analysis is limited by its inability to determine where in the distribution the changes take place. A way around this limit is to analyse the change in the positions of the various institutions on the factorial plane, according to their group. To do this, we plot the evolution of each group's barycentres – i.e. the average points of each group on both PCA dimensions, weighted by the number of students enrolled in the various institutions (Figure V). Among the institutions with a privileged social composition, the group in which science is the primary field of study appears to be closing, while the humanities-focused group reveals a mild social accessibility. Similarly, the group of intermediate social compositions is marginally accessible. Lastly, the group of institutions with a working class composition consistently shifts to the right. Ultimately, it therefore seems that the deepening inequality observed can be largely explained by the dynamics of extremes: the inaccessibility of the group of scientific institutions with a privileged student population versus the accessibility of those institutions with a population of students from a working class background.

To explore this in greater depth, we now use econometric indices of segregation and polarisation for measuring the heterogeneity of a population split into groups.

2.1. A Decrease in the Level of Segregation...

First, we use the normalised entropy index, which is conventionally used in the literature on academic segregation and its evolution. A common feature of segregation indices is that they are based on the division of a population into several categories, following which the population is distributed into different units. The observed distribution of the categories within the different units is then compared to the overall distribution of these categories in the population, in order to quantify the deviation from what would be a homogeneous distribution of the categories. Many indices have been developed and applied in the literature, reflecting the many different situations

Figure IV – **Change in the dispersion of the social composition index**

Notes: The unit is that of the PCA's axes.
Sources and Coverage: See Fig. I.

Figure V – **Change in the barycentres in the four-group, HAC-derived typology between 2007 and 2015**
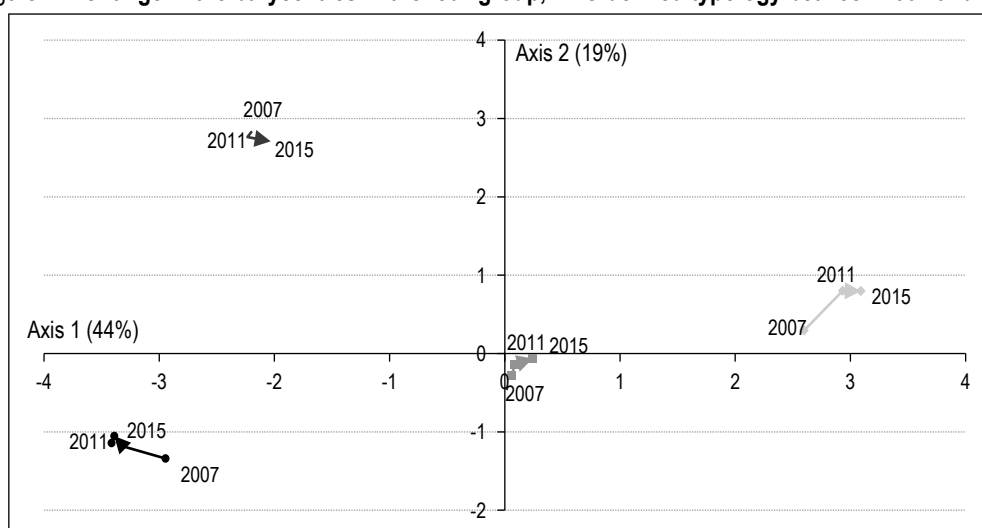
Notes: The arrows indicate the trajectory of the barycentres on the factorial design between 2007 and 2015.
Sources and Coverage: See Fig. I.

that these indices may encounter – binary or multi-category distribution, a variable used to characterise the categories, selection of units, etc. Among the available indices, we select the normalised entropy index H (see a formalised presentation in the Online Appendix – link at the end of the article), which is relevant to study multi-category segregation. This index has a set of properties that are desirable for that type of index, namely statistical properties that enable comparisons over time. It also has the property of additive decomposability across both categories and units (see Reardon & Firebaugh, 2002 for a detailed presentation of these properties, as well as a formalised comparison with several other common multi-category indices). This index has been widely used in the literature on multi-category segregation in schools because of these advantages (see, for example, Fack *et al.*, 2014; Givord *et al.*, 2016 for applications to secondary education in France).

Here, students are categorised according to their social background, as measured by the reference parent's CS, and distributed into higher education institutions. We then measure the distribution of the CS of the students' reference parents across the four groups derived from the HAC. The segregation index shows a relatively steady total decline of around 10% from 2010 onwards (Figure VI).

2.2. ... and an Increase in Polarisation

We now test the hypothesis that the social compositions of French higher education institutions are polarised. The economic literature traditionally addresses this type of question using measures of inequality such as the Gini

coefficient to determine whether the concentration of a distribution (usually income inequality) increases over time. However, as noted by Esteban & Ray (1994), in the case of a shift from unimodal to bimodal distribution over time – i.e. a polarisation of two groups based on local averages – classical measures of inequality founded on the Pigou-Dalton transfer principle would indicate that inequality decreases. Nevertheless, such a situation could indeed characterise a polarising effect (the gap increases in favour of those institutions that have a better position at the outset). Polarisation measures make it possible to overcome this limitation of measures of inequality.

Figure VI – **Change in the normalised entropy index, broken down according to the four-group typology**
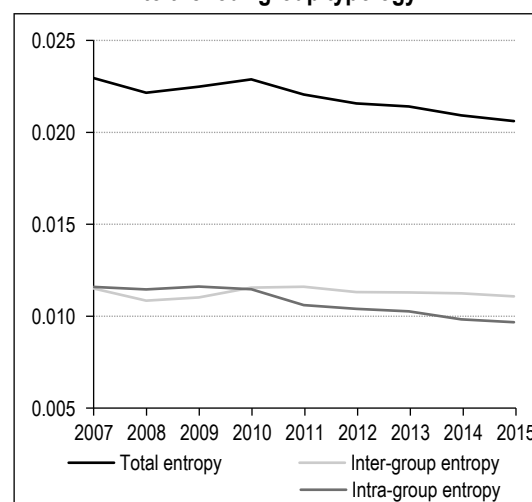
Reading note: In 2007, the normalised entropy index is 0.023, the sum of the inter-group component (0.011) and the intra-group component (0.012).
Sources and Coverage: See Fig. I.

We therefore seek to understand a situation in which the distribution of the social composition and academic capital of the student populations within French universities tends to evolve from a unimodal situation to a bimodal or multi-modal one, i.e. a situation in which the distribution for each mode tends to become more concentrated and/or the modes move further apart. The econometric polarisation framework that we use, which is proposed by Esteban & Ray (1994), is presented in more detail in the Online Appendix. Here, the groups are identified *a priori via* the typology derived from the HAC. We therefore apply the measure developed in the initial article (for an application of this measure to income data in France, see Échevin & Parent, 2002).

There are a number of reasons to apply the polarisation framework to the subject of our study. First, the polarisation's axioms appear to be met. Each of the typologies derived from the HAC produces a small number of groups (between two and four groups). These groups are inherently homogeneous when considered individually, but highly heterogeneous when considered collectively, as shown by the descriptive statistics (see Appendix 2, Table A2 for the four-group configuration). One condition is that each group must be "significant in size". Although this condition is somewhat vague, it seems to be true here, since the relative sizes of the groups are similar in each of the configurations.

Additionally, the creation of pressure groups, which include university presidents, at the end of the period studied suggests that the subjective aspects of polarisation considered by Esteban & Ray (1994) are also present. In 2015, 21 French university presidents collectively signed an open letter, condemning "the obsession with international rankings [which prevails] in budgetary support decision-making, the notion of scientific excellence and critical mass [which] are becoming key, to the detriment of university's other goals, which include success for all and lifelong learning".[6] This open letter, which will lead to the *Alliance des Universités de Recherche et de Formation* (AUREF, an alliance of universities), only includes signatories from the intermediate and working class groups in our typology. At the opposite extreme of the principal plane are the institutions presided by signatories of the 2017 open letter distinguishing "a few research universities that intend to be competitive at global level",[7] and the members of the *Coordination des universités de recherche intensive françaises* (CURIF, a network French research-intensive universities, created in 2008

and since restructured as Udice). These examples illustrate both the sense of identification that can exist between institutions with similar social compositions, and the alienation they can feel towards institutions that are distant from them in the distribution, in the words of Esteban & Ray (1994).

The econometric approach proposed by Esteban & Ray (1994) to study the change in the degree of social polarisation between institutions (see Online Appendix) leads to the reverse outcome to that which emerged from the segregation approach. There is indeed an increase in social polarisation among higher education institutions between 2005 and 2017 (Figure VII). The change in the degree of polarisation is of the order of 20% for our main four-group typology. This result is robust if we retain a 2- or 3-group typology, and it is also robust to the choice of the polarisation sensitivity parameter α, i.e. the assumption made about the degree of homogeneity of the groups in a static approach (see Online Appendix).

Segregation and polarisation therefore have opposite dynamics, but the fact that the two indices cannot be interpreted directly makes it difficult to explain these contrasting

6. *https://blogs.mediapart.fr/edition/les-invites-de-mediapart/article/290515/quel-avenir-pour-l-enseignement-superieur-et-la-recherche-francais*
7. *https://lemonde.fr/idees/article/2017/05/30/enseignement-superieur-rapprocher-les-meilleurs-organismes-de-recherche-des-meilleurs-etablissements_5135898_3232.html*

Figure VII – **Change in the polarisation index of Esteban & Ray with the four-group typology**
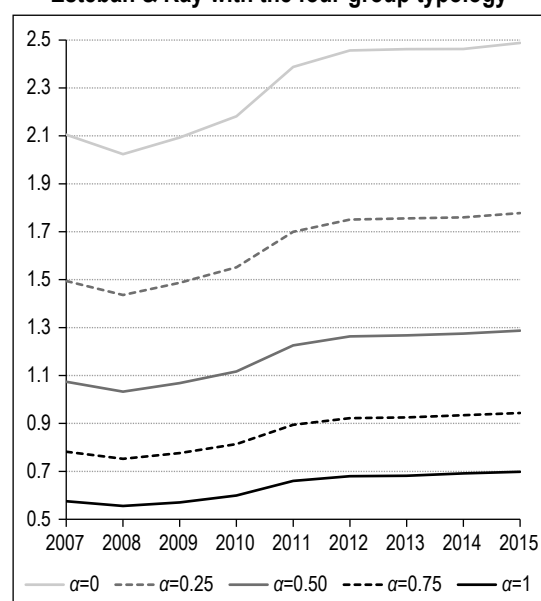
Reading note: In 2007, the polarisation index of Esteban & Ray is equal to 2.1 when the polarisation sensitivity parameter is set to zero.
Sources and Coverage: See Fig. I.

developments. Two explanations emerge when we revisit the underlying raw data used to calculate these indices, however.

A first possibility is that this divergence is partly due to the different concepts that these two indices aim to measure. The decomposition of the entropy index according to the HAC's groups into an inter-group component and an intra-group component shows that virtually the entire decline in total entropy over the period can be explained by the decline in the intra-group component, while the inter-group component remains unchanged (cf. Figure VII). This observation means that most of the decline in the level of segregation observed over the period is explained not by a convergence of the different groups of institutions but by an homogenisation of these groups in terms of their social composition. So although different institutions in the same group have increasingly similar student populations, in terms of the reference parent's CS, the gap between each individual group pairing remains consistent. This is precisely one of the dynamics that the polarisation index aims to measure and this is what distinguishes it from the conventional inequality and segregation indices. This kind of situation, in which the centres of the groups studied do not move and instead the institutions move closer to the centre within each group, tends to produce a more marked multi-modal distribution, which corresponds to the first polarisation axiom (Esteban & Ray, 1994). Segregation and polarisation develop together with heterogeneity between group centres, but this development is inverse when intra-group heterogeneity is introduced. When intra-group heterogeneity decreases, segregation decreases but polarisation increases.

Another explanation, which is also backed up by the data, is that these two analyses do not use the same definition of student background. Only the reference parent's CS is used to calculate the entropy index. The general index of social composition produced by the PCA – from which the polarisation index is calculated – is based on a wider definition of institutional social composition, which includes academic capital. In particular, the graph of the PCA variables (cf. Figure I) shows that the vocational baccalaureate is a social indicator with an effect similar to that of being a child of an unskilled white-collar or blue-collar worker. However, a precise analysis of the change in the socio-academic profile of students over the period reveals a considerable increase in the number of vocational baccalaureate holders in higher education. This number increased by a factor of 2.6 between 2007 and

2015, whereas the development for other types of baccalaureate remains stable or marginal (10%). If we break this change down according to the HAC-derived typology, we see that most of this increase is attributable to institutions belonging to the working class social composition group. At the outset, these institutions already had a significantly higher percentage of vocational baccalaureate holders among their student populations (Figure VIII). It therefore seems that the particularly pronounced change in the barycentre of the group of institutions with a working class social composition (cf. Figure V) can be largely explained by the fact that, over the period 2007-2015, there was a very high level of access to higher education for holders of a vocational baccalaureate, and enrolments of these students were very unequally distributed among institutions. The polarisation analysis identifies this trend due to the general index of socio-academic capital produced by the PCA. However, a segregation analysis based solely on the reference parent's CS would not be able to detect such a trend. The endogeneity of the polarisation index is what allows the various trends to be identified, whereas the segregation analysis requires to select beforehand the relevant categories (in this case, the percentage of vocational baccalaureate holders).

Figure VIII – **Change in the proportion of vocational baccalaureate holders among students in the four groups derived from the HAC, between 2007 and 2015**
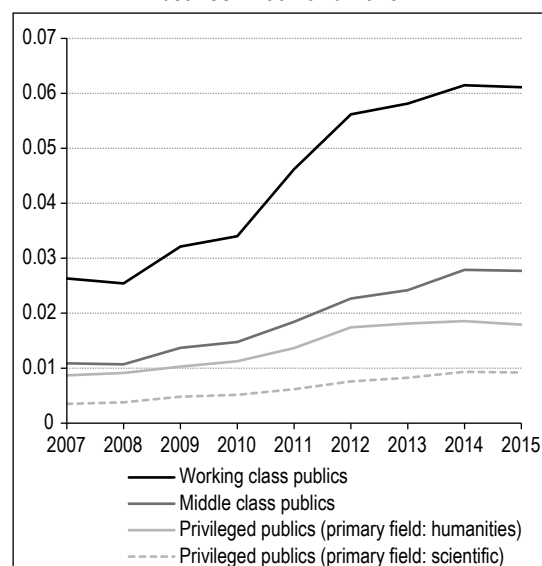
Reading note: The proportion of vocational baccalaureate holders among students in the group of institutions with a working class social composition increases from 2.6% to 6.1% between 2007 and 2015.
Notes: The number of vocational baccalaureate holders among students enrolled in higher education increases from 10,917 to 28,772 between 2007 and 2015.
Sources and Coverage: See Fig. I.

The conclusion of the dynamic analysis therefore depends directly on the operationalisation of the concept used to measure the shift in the distribution of social compositions. The conclusion of the analysis of a polarisation index is that of an increase, insofar as this distribution appears to be increasingly marked by the existence of groups of institutions that are clearly identified by the socio-academic capital held by their students. However, the conclusion drawn from the analysis of a segregation index is that of a decrease, although this is essentially the product of the homogenisation of the social compositions within the groups derived from the typology. These different analyses generally lead to the conclusion that French universities, which were highly socially heterogeneous at the outset, do not exhibit any overall homogenisation that would reflect a converging composition of student populations in the institutions during the period studied.

3. "Policies of Excellence" in Higher Education: A Reverse Redistribution?

Our results indicate that higher education institutions are statically characterised by a high degree of social heterogeneity. This heterogeneity also appears to increase between 2007 and 2015. However, identifying the causes of such a process, which is probably multifaceted, is complicated.

For example, we have shown that the (static) heterogeneity of the social compositions somewhat mirrored the unequal distribution of those compositions in various regions. It is therefore possible that greater social inequality between regions will systematically influence the social composition of the universities affected. However, this explanation assumes that students have a low level of mobility in connection with their family home. In fact, the propensity for student mobility cannot be neglected and is inherently dependent on social background. In 2007, in groups of institutions with an intermediate or working class social composition, approximately 25% of students are enrolled in an institution for which the academy is located in a place other than the place of residence of the students' parents. In institutions that accept more privileged students, this rises to 40% in the primarily science-focused group and 50% in the primarily humanities-focused group. This mobility makes it difficult to conclude that changes in regional inequality are systematically transmitted to the academic sphere. Furthermore, the distribution of social compositions within French regions appears to be relatively

consistent over the period studied. To arrive at this observation, we take the normalised entropy index – which is a useful way of quantifying changes in the degree of separation between several groups in a region – and apply it to the distributions of socio-professional categories by academy, calculated using data taken from recent censuses. We therefore see a 2% increase in the index between 1999 and 2010, followed by a 3.8% decrease between 2010 and 2015. These changes do not seem to be sufficient to explain the observed trends of social recruitment within higher education.

Increased polarisation is one of the theoretical predictions to emerge from the literature which analyses the performativity of the measures that have partly shaped recent policy in higher education: *Initiatives d'excellence* (IDEXs), university rankings to justify groupings of universities and institutions (COMUEs), etc. (Espeland & Sauder, 2007; Van Parijs, 2009; Halffman & Leydesdorff, 2010; Brusadelli & Lebaron, 2012; Paradeise & Thoenig, 2015). However, it would appear difficult to empirically establish the causal link between these two phenomena, especially given that a reverse causal link is possible. For example, a privileged social composition could "attract" IDEXs if the average level of students in these institutions were higher. In this section, we focus on the link between the social composition of institutions and the fact of being distinguished by the aforementioned measures. We particularly concentrate on IDEXs, which, due to the potential financial benefits they bring, appear to be a major component of higher education policy during the period studied. Considering the social heterogeneity highlighted in the previous analyses, this relationship seriously raises the question of fairness and the redistribution of resources within higher education.

The aim of the *Programme d'investissements d'avenir* (PIA – "Investments for the future" programme), which was introduced by the Amending Finance Law of 9 March 2010, is to support innovative projects in a number of sectors deemed "national priorities". Academia is high up on this list of priorities. 7.7 billion euros of funding were granted to IDEXs, which comfortably makes it the largest budget item.[8] In 2018, after the first two waves of investments for the future (PIA1 and PIA2), ten institutions had been awarded the IDEX label – three of which are in a

---

8. *See the 2015* Rapport relatif à la mise en œuvre et au suivi des investissements d'avenir de 2015 *(https://www.performance-publique.budget. gouv.fr/sites/performance_publique/files/farandole/ressources/2015/pap/ pdf/jaunes/jaune2015_investissements_avenir.pdf).*

probatory period: the Paris-Saclay University, PSL Research University and University of Paris projects. In addition, the PIA2 programmes also recognised *Initiatives-Science–Innovation–Territoires–Économie* (I-SITEs) among the IDEXs. The innovative potential of I-SITEs is recognised but they are not intended to become world-class universities, and consequently receive substantially less funding than IDEXs.

On the factorial plane, we show the universities covered that belong to a group – generally a COMUE – assigned the IDEX label (Figure IX). To do so, we had to construct the University of Clermont Auvergne (resulting from the merger of Blaise Pascal University and the University of Auvergne in 2017) on an *ex ante* basis using the 2015 data from the SISE database, in order to make our scope comparable to the universities benefiting from the IDEX label.

Several observations can be made on the basis of this figure. First, universities belonging to an IDEX institution tend to be positioned in the left quadrant of the factorial plane, and these therefore correspond to universities with socially privileged student populations. In reality, our representation certainly underestimates this fact. In practice, these universities generally belong to a COMUE along with prestigious higher education and research institutions (the *Grandes Écoles*). If we had the data required to fully reconstruct the field components of these groups, it is very likely that they would shift further to

the left on the plane. This is particularly true for all the universities in the right quadrant: the University of Evry-Val-d'Essonne (UEVE, on the right of the plane) is part of the University of Paris-Saclay, together with Versailles Saint-Quentin-en-Yvelines University, Paris-Sud University and the *Grandes Écoles* (ENS Paris-Saclay, AgroParisTech, CentraleSupélec, etc.). Lyon 2 and Lyon 3 belong to the University of Lyon COMUE, together with Lyon 1, the *Grandes Écoles* (ENS Lyon, Institute of Political Studies of Lyon, EM Lyon Business School, etc.) and others.

We also note that the hierarchy of the IDEXs has a somewhat natural order along the social composition axis of the PCA. If we consider the HAC-derived classifications, the vast majority of IDEXs belong to the groups of institutions with socially privileged student populations, and the I-SITEs belong to the groups with an intermediate social composition, but no institution belonging to the group with a working class social composition has the IDEX label – apart from the University of Evry-Val-d'Essonne (UEVE), but this is more a matter of statistical artefact for the aforementioned reasons.

The labels of excellence and associated exceptional funding are thus awarded to institutions with relatively high levels of the most privileged students, reproducing within the university system itself a well-established reverse redistribution of public resources between this university system

Figure IX – **Projection on the PCA factorial plane of the institutions granted an IDEX in 2015**
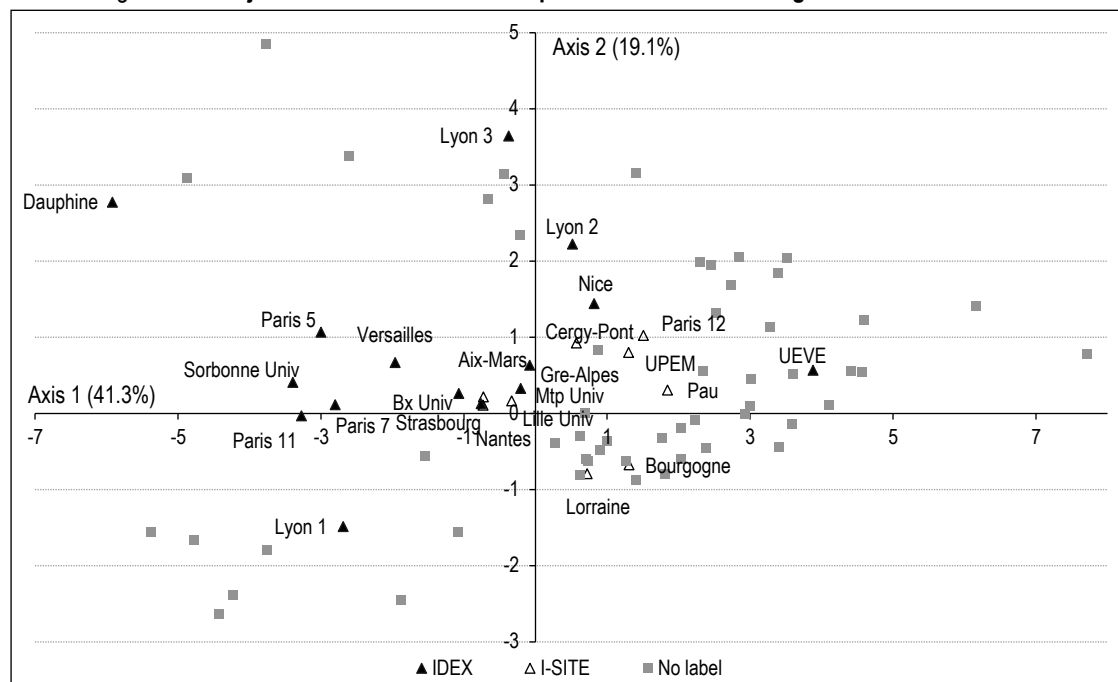
Sources and Coverage: See Fig. I.

Figure X – **Projection on the PCA factorial plane of the institutions appearing on the Shanghai Ranking (ARWU) in 2015**
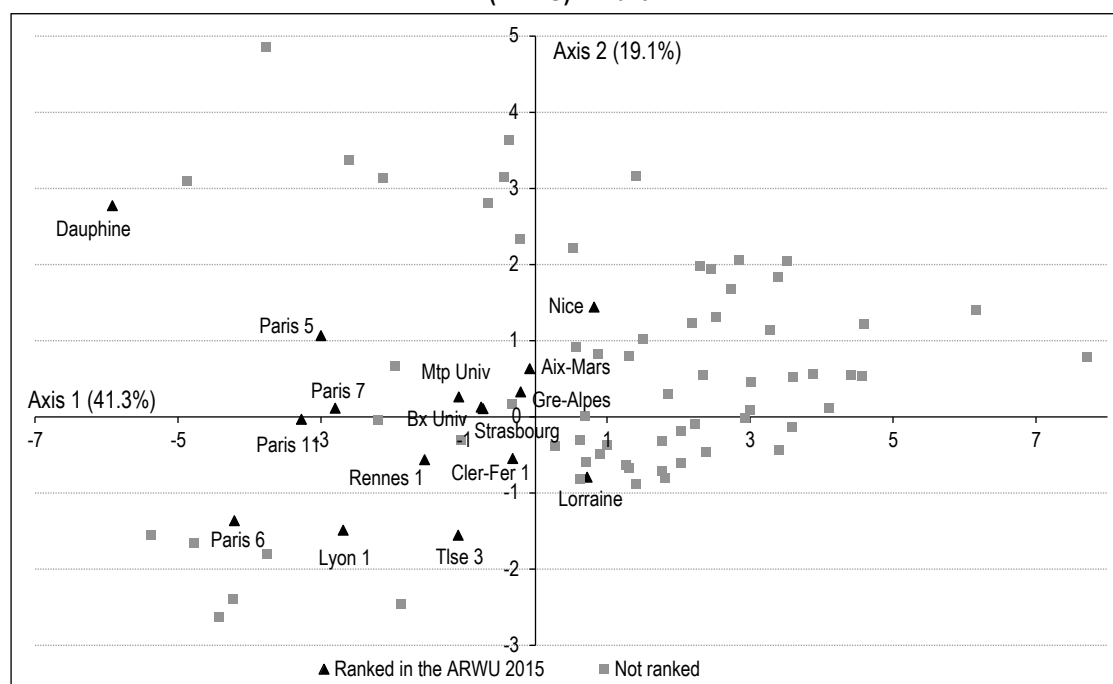
Sources and Coverage: See Fig. I.

and the Preparatory classes to the *"Grandes Écoles"* (Flacher *et al.*, 2018). On the other hand, the competitive procedure between institutions to obtain these labels produces a selection which strongly reflects the social composition of their student populations, even though it is based solely on an evaluation of the institutional and scientific projects of those institutions.

A similar observation can be made concerning the main university rankings, the increasing application of which to justify higher education policy was noted as early as 2007 (Hazelkorn, 2007). Rankings are widely used in calls for projects in connection with IDEXs, in particular. Of the twenty successful institutions in the competitive procedure for IDEX and I-SITE initiatives, fourteen cite the opportunity to improve their position in the Shanghai Ranking as justification for their project (Charpin *et al.*, 2017). Unsurprisingly, the projection on the factorial plane of institutions appearing in the 2015 Shanghai Ranking largely coincides with that for the IDEXs, in particular those institutions granted an IDEX (Figure X). These rankings, which are only very imperfect measures of university excellence due to significant methodological shortcomings (Billaut *et al.*, 2010), therefore also appear to distinguish institutions with a privileged social composition for the most part. We may then question the extent to which these various measures encourage genuine competition and allocate more resources to the most worthy institutions, or, on the contrary,

whether they in fact support institutions whose role in the massification of higher education is merely peripheral.

* *
*

Using data from comprehensive student enrolment databases, we show that the French university system is characterised by a high level of social heterogeneity. Although higher education is pursuing greater accessibility for young working class people, a process that we have been witnessing for a number of decades, this accessibility does not result in greater social diversity in universities. The indicators that we use suggest that the level of social heterogeneity did not decrease significantly between 2007 and 2015. We also show that the *Initiatives d'excellence* (IDEXs), which are a major component of recent higher education policy, give priority to institutions with a privileged social composition. Given the high level of academic massification which is borne unequally by the institutions, this observation justifies a thorough reconsideration of the distribution of funding in higher education.

Other recent developments suggest that this polarisation may continue in the years to come. After some uncertainty concerning the statutes of the university groups (associations, clusters, communities, etc.), the drafting of exceptional

statutes for IDEXs in the second half of 2018 could result in the existence of two distinct types of universities in France. In connection with this legal development, the fact that universities with a status of *Grand établissement* can introduce much higher enrolment fees than other institutions may increase social polarisation, both because it acts as a barrier to entry to institutions with a socially privileged student population and because it widens the gap in resources between the respective institutions in both groups.

Several extensions of the study presented here can be considered. Looking further back in time, we could test the presence of a period of social accessibility within the university system with less or even no polarisation. However, gradually expanding the scope of the SISE university database risks making this type of study less feasible. We could include institutions under the authority of other ministries, which could produce more useful results and would test the hypothesis of a convergence of universities and prestigious higher education institutions. This would strengthen our interpretation regarding IDEXs. Finally, a multi-scale analysis is also possible: within the institutions themselves, can we observe an internal social polarisation between Bachelor's and Master's degrees for example, between disciplines or between different campuses at the same university? What about during the Bachelor's degree cycle, when students from the Preparatory classes to Higher Schools enter the third year of a Bachelor's degree or in the case of dual-honours degrees and other special arrangements? We could then break down the respective roles of the institution, campus, discipline, cycle, etc. in the social mix or polarisation experienced by young students. □

Link to Online Appendix:
https://www.insee.fr/en/statistiques/fichier/5391226/ES_Avouac-Harari_Annexe-en-ligne_Online-Appendix.pdf

---

BIBLIOGRAPHY

Albouy, V. & Tavan, C. (2007). Accès à l'enseignement supérieur en France: une démocratisation réelle mais de faible ampleur. *Économie et Statistique*, 410, 3–22. https://doi.org/10.3406/estat.2007.7053

Amsler, S.-S. & Bolsmann, C.-H. (2012). University Ranking as Social Exclusion. *British Journal of Sociology of Education*, 33, 283–301. https://doi.org/10.1080/01425692.2011.649835

Auzet, L. & Février, M. (2007). Niveaux de vie et pauvreté en France : les départements du Nord et du Sud sont les plus touchés par la pauvreté et les inégalités. *Insee Première* N° 1162.
https://www.epsilon.insee.fr/jspui/bitstream/1/91/1/ip1162.pdf

Billaut, J. C., Bouyssou, D. & Vincke, P. (2010). Should you believe in the Shanghai ranking? *Scientometrics*, 84, 237–263. https://akjournals.com/view/journals/11192/84/1/article-p237.xml

Boutchénik, B., Coron, C., Grobon, S., Goffette, C. & Vallet, L.-A. (2015). Quantifier l'influence totale de la famille d'origine sur le devenir scolaire et professionnel des individus. *Économie et Statistique*, 477, 5–23. https://doi.org/10.3406/estat.2015.10532

Brinbaum, Y., Hugrée, C. & Poullaouec, T. (2018). 50 % à la licence... mais comment? Les jeunes de familles populaires à l'université en France. *Économie et Statistique,* 499, 79–105.
https://doi.org/10.24187/ecostat.2018.499s.1941

Brusadelli, N. & Lebaron, F. (2012). Les indicateurs de « performance » universitaire. *Savoir/Agir*, 4, 97–104. https://doi-org.ezproxy.universite-paris-saclay.fr/10.3917/sava.022.0097

Charpin, J.-M., Filliatreau, G., Aimé, P., Dulbecco, P. & Foucault, M. (2017). La prise en compte des classements internationaux dans les politiques de site. Rapport de l'IGF et de l'IGAENR.
https://www.enseignementsup-recherche.gouv.fr/cid119176/la-prise-en-compte-des-classements-internationaux-dans-les-politiques-de-site.html

Convert, B. (2006). *Les impasses de la démocratisation scolaire*. Paris: Raisons d'Agir.

Duru-Bellat, M. & Kieffer, A. (2008). Du baccalauréat à l'enseignement supérieur en France: déplacement et recomposition des inégalités. *Population*, 63, 123–157. https://doi.org/10.3917/popu.801.0123

Échevin, D. & Parent, A. (2002). Les indicateurs de polarisation et leur application à la France. *Économie & prévision*, 155, 13–30. https://doi.org/10.3917/ecop.155.0013

Espeland, W.-N. & Sauder, M. (2007). Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology,* 113, 1–40. https://doi.org/10.1086/517897

Esteban, J.-M. & Ray, D. (1994). On the Measurement of Polarization. *Econometrica,* 62, 819–851.
https://doi.org/10.2307/2951734

Fack, G., Grenet, J. & Benhenda, A. (2014). L'impact des procédures de sectorisation et d'affectation sur la mixité sociale et scolaire dans les lycées d'Île-de-France. Institut des Politiques Publiques, Rapport N° 3. https://hal.archives-ouvertes.fr/hal-02464102/

Flacher, D., Harari-Kermadec, H. & Moulin, L. (2018). Régime par répartition dans l'enseignement supérieur : fondements théoriques et estimations empiriques. *Économie et institutions* [en ligne], 27. https://doi.org/10.4000/ei.6233

Frouillou, L. (2017). Des publics hétérogènes ? Tableau des étudiants en Géographie des universités françaises. *Carnets de géographes*, 10, 1–17. https://doi.org/10.4000/cdg.1129

Givord, P., Guillerm, M., Monso, O. & Murat, F. (2016). La ségrégation sociale entre les collèges : Quelles différences entre public et privé, aux niveaux national, académique et local ? *Éducation et Formations*, Massification scolaire et mixité sociale, 53–76. https://hal.archives-ouvertes.fr/halshs-01447198/

Halffman, W. & Leydesdorff, L. (2010). Is Inequality Among Universities Increasing? Gini Coefficients and the Elusive Rise of Elite Universities. *Minerva,* 48, 55–72. https://arxiv.org/abs/1001.2921

Hazelkorn, E. (2007). L'impact du classement des établissements sur la prise de décision dans l'enseignement supérieur. *Politiques et gestion de l'enseignement supérieur*, 19, 95–122. https://doi.org/10.1787/hemp-v19-art12-fr

Husson, F., Josse, J. & Pagès, J. (2010). Principal component methods – hierarchical clustering - partitional clustering: why would we need to choose for visualizing data? Agrocampus, Rapport technique. http://www.sthda.com/english/upload/hcpc_husson_josse.pdf

Ichou, M. & Vallet, L.-A. (2013). Academic achievement, tracking decisions, and their relative contribution to educational inequalities: Change over four decades in France. *In:* M. Jackson (Ed.). *Determined to succeed? Performance versus choice in educational attainment,* pp. 116–148. Stanford (Cal.): Stanford University Press.

Jauneau, Y. (2009). Les employés et ouvriers non qualifiés. *Insee Première* N° 1250. http://www.epsilon.insee.fr/jspui/bitstream/1/672/1/ip1250.pdf

Marchand, O. (2010). 50 ans de mutations de l'emploi. *Insee Première* N° 1312. http://www.epsilon.insee.fr/jspui/bitstream/1/583/1/ip1312.pdf

McKenzie, D.-J. (2005). Measuring Inequality with Asset Indicators. *Journal of Population Economics*, 18, 229–260. https://doi.org/10.1007/s00148-005-0224-7

Merle, P. (2002). Démocratisation ou accroissement des inégalités scolaires? L'exemple de l'évolution de la durée des études en France (1988-1998). *Population*, 57(4-5), 633–659. https://doi.org/10.2307/1534799

Merton, R.-K. (1968). The Matthew Effect in Science. *Science*, 159, 56–63. https://doi.org/10.1126/science.159.3810.56

Nicourd, S., Samuel, O. & Vilter, S. (2011). Les inégalités territoriales à l'université: effets sur les parcours des étudiants d'origine populaire. *Revue française de pédagogie*, 176, 27–40. https://doi.org/10.4000/rfp.3153

Paradeise, C. & Thoenig, J. C. (2015). *In Search of Academic Quality*. London: Palgrave MacMillan.

Peugny, C. (2013). *Le destin au berceau. Inégalités et reproduction sociale*. Paris: Seuil.

Place, D. & Vincent, B. (2009). L'influence des caractéristiques sociodémographiques sur les diplômes et les compétences. *Économie et Statistique*, 424-425, 125–147. https://doi.org/10.3406/estat.2009.8034

Reardon, S.-F. & Firebaugh, G. (2002). Measures of Multigroup Segregation. *Sociological Methodology*, 32, 33–67. https://doi.org/10.1111/1467-9531.00110

Robette, N. (2011). Explorer et décrire les parcours de vie : les typologies de trajectoires. Paris: CEPED.

Rocher, T. (2016). Construction d'un indice de position sociale des élèves. *Éducation et Formations*, 90, 5–27. https://hal.archives-ouvertes.fr/hal-01350095/

Rosenwald, F. (2006). Filles et garçons dans le système éducatif depuis vingt ans. Insee, *Données sociales : La société française*, pp. 87–94. https://www.insee.fr/fr/statistiques/fichier/1371989/donsoc06i.pdf

Selz, M. & Vallet, L.-A. (2006). La démocratisation de l'enseignement et son paradoxe apparent. Insee, *Données sociales: La société française*, 101–107. https://www.insee.fr/fr/statistiques/fichier/1371955/donsoc06k.pdf

Truong, F. (2015). Comment les « nouveaux étudiants » s'ajustent aux études supérieures par le bas. *Diversité*, 181, 68–73. https://www.academia.edu/16262361/Comment_les_nouveaux_%C3%A9tudiants_s_ajustent_aux_%C3%A9tudes_sup%C3%A9rieures_par_le_bas

Van Parijs, P. (2009). European Higher Education Under the Spell of University Rankings. *Ethical Perspectives*, 16, 189–206. https://doi.org/10.2143/EP.16.2.2041651

Vyas, S. & Kumaranayake, L. (2006). Constructing Socio-Economics Status Indices: How to Use Principal Components Analysis. *Health Policy and Planning*, 21, 459–468. https://doi.org/10.1093/heapol/czl029

CODING OF THE REFERENCE PARENTS' SOCIO-PROFESSIONAL CATEGORY (CS) AND HARMONISATION

Selection of the classification system for coding the socio-professional category of the reference parent

Information concerning the CS of the student's reference parent is coded in SISE using a classification system similar to the two-digit socio-professional classification system used by INSEE (PCS). The size of our sample, whereby the analysis relates to 81 higher education institutions rather than students, makes it impossible to directly include all of these conditions into the PCA as variables. However, the single-digit PCS classification system does not appear to be entirely satisfactory either: certain categories present significant heterogeneity, particularly in terms of social and cultural capital, and this could conceal clear differences in higher educational trajectories. Therefore, we use a refined version of this system, using divisions frequently used by sociologists for social stratification purposes. In particular, we separate entrepreneurs with ten or more employees (CS 23) from craftspeople/traders (CS 21 and 22). We also draw a distinction between skilled blue-collar workers (CS 61 in SISE) and unskilled blue-collar workers (CS 66 and 69 in SISE). Similarly, there is no doubt that it would have been useful to distinguish the skill level of white-collar workers, but this requires classification systems that are more detailed (Jauneau, 2009).

Changes to the database coverage

The coding of the reference parent's CS in SISE forces us to omit individuals from our analysis. First, the categories of retirees (CS 71 to 76 in SISE) are combined: for example, CS 76 corresponds to retired white-collar and blue-collar workers. These categories therefore cannot be broken down using the CS classification system that we have selected. That is why we do not include any students for whom the reference parent is a retiree. In addition, we also omit individuals for whom the reference parent's code is "unemployed having never worked" (too few observations to create a variable associated with this condition) or "other not engaged in reference activity" (imprecise and potentially heterogeneous category), or for whom the reference parent's CS is not provided.

To ensure that our analyses can be compared over time, we must limit the database coverage. First, certain university components – such as the *Instituts Universitaires de Formation des Maîtres* (IUFM, institutes for teacher training) in Martinique, Guadeloupe and French Guiana, Centre Universitaire de Mayotte as well as the Paris-Est centre for research and higher education (PRES) – either emerge or change significantly during the period studied, so we omit them to ensure that our coverage remains consistent. Similarly, between 2007 and 2008, certain engineering schools affiliated with an institution become non-university engineering schools in the coverage. These are therefore no longer present in the SISE university tables.

Several universities merge into a single entity during the period studied. We are choosing to reconstruct these mergers on an *ex-ante* basis to ensure that our results can be compared over time. For example, the University of Bordeaux was created in 2014 following the merger of the Bordeaux I, Bordeaux II and Bordeaux IV universities. We therefore reconstruct this consolidated institution from 2007 onwards by combining the student populations from these three separate universities during each year between 2007 and 2013. This reconstruction is not neutral from a statistical perspective. Mergers generally have an "averaging" effect on the PCA factorial plane, with the consolidated institution tending to be positioned in the centre of the plane. This is because mergers effectively aggregate institutions that previously might have occupied very distinct positions on the factorial plane. However, due to the number and size of institutions resulting from mergers, deleting them would produce a significant representativeness bias.

APPENDIX 2

FOUR-GROUP TYPOLOGY

Table A2 – **Description of the HAC-derived groups in 2007**

%

Group	Privileged populations (primarily science)	Privileged populations (primarily humanities)	Intermediate populations	Working class populations	Overall
Male	50	37	43	41	43
Female	50	63	57	59	57
S Baccalaureate	77	29	48	31	48
ES Baccalaureate	10	34	24	26	23
L Baccalaureate	5	28	15	22	16
Voc. baccalaureate	0	1	1	3	1
Tech. baccalaureate	8	8	12	19	12
Farmers	2	1	3	3	2
Craftspeople/traders	6	6	6	7	6
Entrepreneurs	3	4	2	2	2
Executives and intellectual professions	52	54	38	28	41
Mid-level professionals	17	13	19	20	18
White-collar workers	13	14	17	22	17
Skilled blue-collar workers	6	6	11	13	10
Unskilled blue-collar workers	2	2	3	5	3
Bachelor's	53	61	67	79	66
Master's	42	34	30	20	31
Doctorate	5	5	3	2	3
Student total	136,656	122,464	581,775	115,510	956,405

Reading note: In 2007, the primarily scientific institutions with a privileged student population consist of 50% male students, 77% Science Baccalaureate holders, 2% students for whom the reference parent is a farmer and 53% students enrolled on an undergraduate course of study.
Sources and Coverage: MESRI-SIES, SISE; French universities and elite institutions under the authority of the Minister for Higher Education.
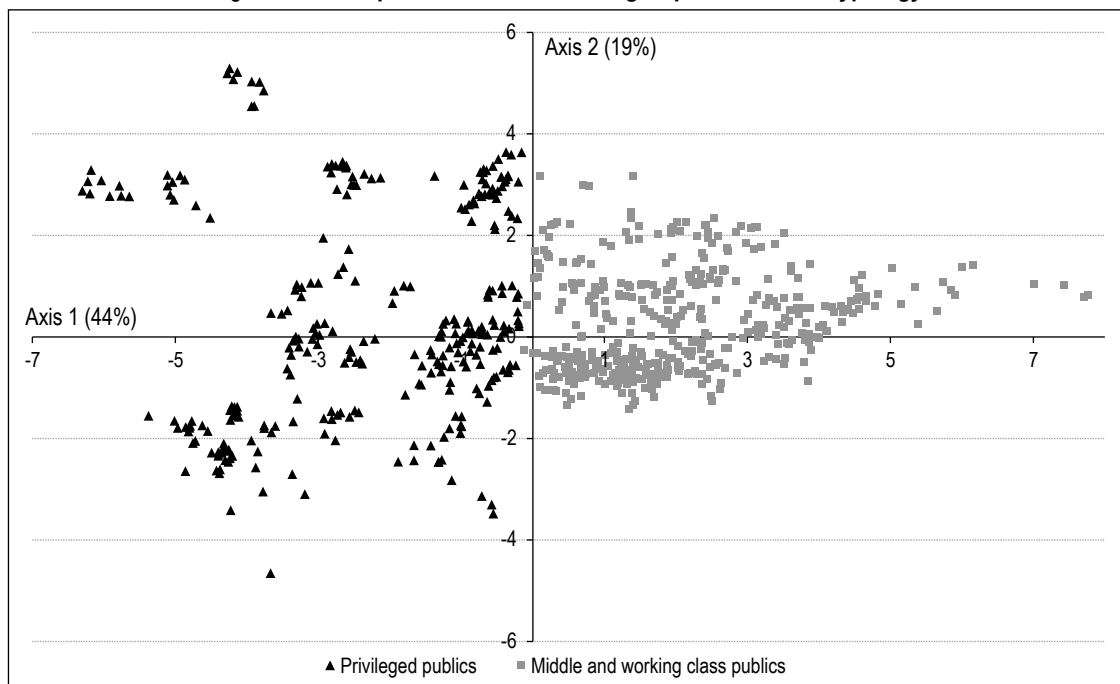
PCA AND TYPOLOGY

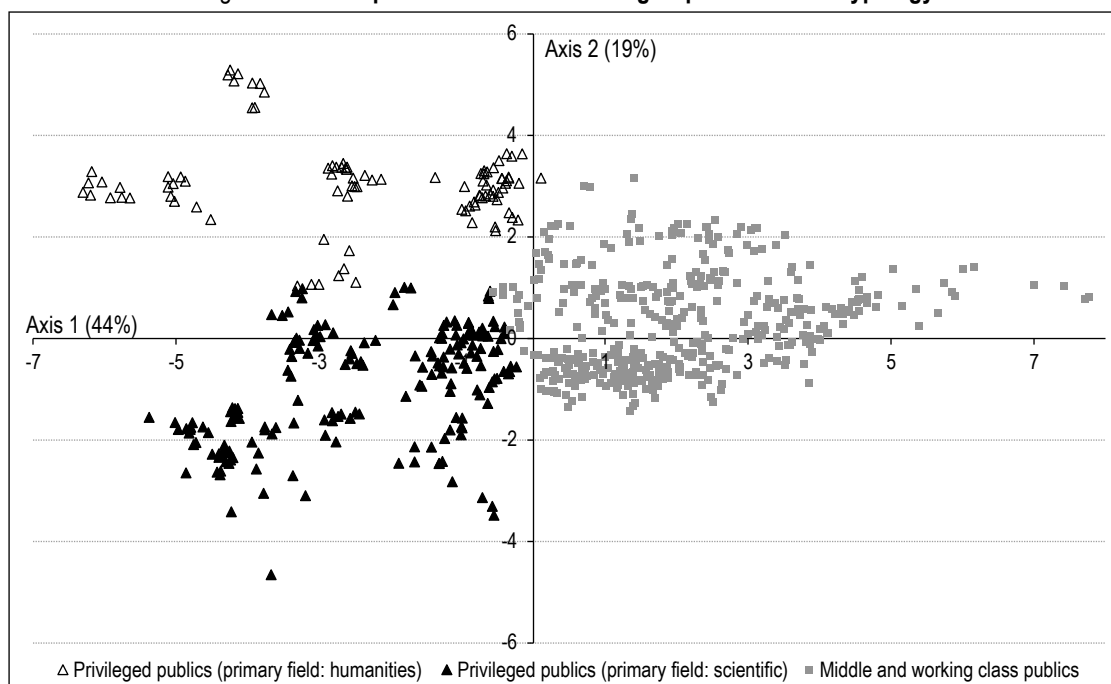Figure A3-I – **Variance explained by the various dimensions of the PCA in 2007**
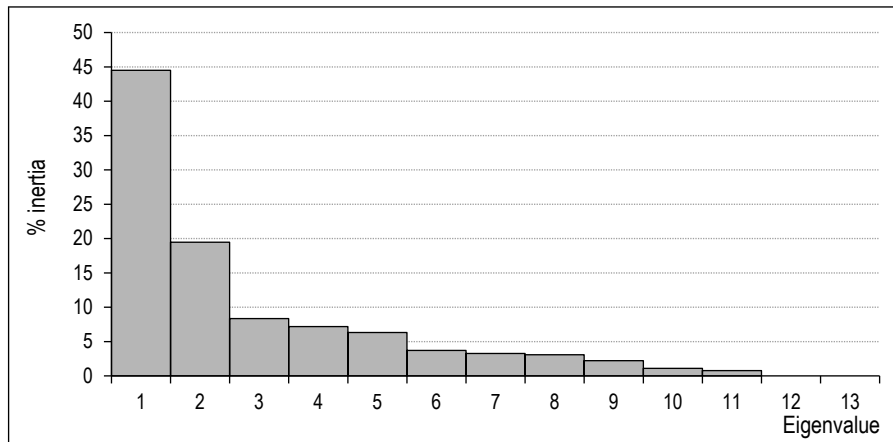
Reading note: The first dimension of the PCA reproduces 44% of the initial cluster's inertia.
Sources and Coverage: See Fig. I.

Figure A3-II – **Representation of the two-group HAC-derived typology**

Sources and Coverage: See Fig. I.

Figure A3-III – **Representation of the three-group HAC-derived typology**

△ Privileged publics (primary field: humanities)　▲ Privileged publics (primary field: scientific)　■ Middle and working class publics

Sources and Coverage: See Fig. I.