Groupe de Recherche en
Économie Théorique et Appliquée

Dominance relations and universities ranking

Nicolas CARAYOL

Université de Bordeaux
GREThA UMR CNRS 5113

Agenor LAHATTE

Observatoire des Sciences Techniques

Cahiers du GREThA
n° 2009-02

GRETHA UMR CNRS 5113
Université Montesquieu Bordeaux IV
Avenue Léon Duguit - 33608 PESSAC - FRANCE
Tel : +33 (0)5.56.84.25.75 - Fax : +33 (0)5.56.84.86.47 - www.gretha.fr

Relations de dominance et classement des universités

Résumé

Cet article propose une théorie permettant d'établir des relations de dominance entre universités sur la base de leur production scientifique et du nombre de citations reçues dans une certaine fenêtre temporelle. Cette théorie est appliqué au classement des Universités Françaises.

Mots-clés : Classement ; Relations de dominance ; Citations

Dominance relations and universities ranking

Abstract

This paper proposes a theory for establishing dominance relations between universities on the basis of their scientific production and the number of citations their publications received in given time window. We apply this theory to the ranking of French Universities.

Keywords: Ranking; dominance relations; citations.

JEL : D63 ; I23

Reference to this paper: Nicolas CARAYOL, Agenor LAHATTE "Dominance relations and universities ranking", *Working Papers of GREThA*, n° 2009-02 http://ideas.repec.org/p/grt/wpegrt/2009-02.html.

GRETHA UMR CNRS 5113
Université Montesquieu Bordeaux IV
Avenue Léon Duguit - 33608 PESSAC - FRANCE
Tel : +33 (0)5.56.84.25.75 - Fax : +33 (0)5.56.84.86.47 - www.gretha.fr

1 Introduction

In a context of rising competition for funds and students, the ranking of universities and departments has recently been the subject of a strong interest from the public and the managers of these institutions. Such rankings are also increasingly viewed as policy leverages. For instance, in the European context, universities rankings are often expected to increase the competition between universities across countries and to contribute to creating an integrated research and higher education system. Nevertheless, the available rankings seem to rely more on *ad hoc* techniques rather than on clear theoretical grounds. As a consequence, such rankings provide only weakly reliable information and are of a limited interest for the public, for the managers of these institutions as well as for policy makers.

This article provides a simple rationale to establish one-to-one *dominance relations*[1] between the scientific production of any given set of institutions, from which the ranking of these institutions can be established. The core of the proposed methodology is to provide a satisfactory treatment of the visibility of institutions in different segments of "scientific credit". Dominance relations between universities are established by broad scientific domains.

For that purpose, we introduce the notion of valuation function which gives a "value" to any article on the basis of the number of citations it received in a given time window. Depending of what the term value may refer to, one can retain different assumptions on the properties of the valuation function. Our goal is indeed not to propose any specific functional form for the valuation function, but to show how classes of valuation functions do transform into simple *dominance relations* of various strength between institutions. Moreover, the notion of upward dominance is introduced which allows us to focus only on high segments of scientific credit, that is to not consider publications that received less than a given number of citations. We show that all dominance relations have the transitivity property and that a hierarchy between the dominance relations do exist: a dominance relation of an university on an other university implies all weaker relations. Therefore, the more complete the dominance relation, the weaker its strength.

The main goal of this article is to show and discuss to what extent rankings can be inferred from such dominance relations. We distinguish complete dominance ranking which can deduced from dominance relations from pseudo dominance ranking which use the dominance relations to build indexes which in turn can ground rankings. The absence of any dominance relation is also used to define classes of universities among which "best(s) in class" institution(s) can be identified (they may constitute good models to compete with for the other institutions in the class).

This methodology is applied and calibrated on the set of French universities, relying upon their publications and the citation scores of these articles. Data are constituted of the detailed information on the publication of all French Universities as they appear in the ISI-WOS.[2] These data

[1]The term "dominance" refers to the stochastic dominance relations between distributions first used by von Neuman and Morgenstern (1944) in their major contribution to the theory of choice under uncertainty. Nevertheless, our notion is different as it shall become explicit below.

[2]These data are collected in house by the Observatiore des Sciences et Techniques through a very detailed and precise tehcniques that involve directly the intitutions themselves for the selection of the appropriate list of addresses mentioned in publications.

are particularly precise thanks to an interactive data collection and cleaning process involving the universities themselves so as to get rid of the variety of signing patterns among the research staff of these institutions.

The article is organized as follows. The basic notations on the scientific production of universities and its value are introduced in the second section. The dominance relations and their conditions of relevance are proposed in the third section. The fourth section presents how such dominance relations can be used to produce universities ranking and reference classes. In the fifth section, we derive a generic transitivity property of the dominance relations and show how they are linked to each other. The sixth section is dedicated to the calibration of the theory relying upon data on the French universities. The last section concludes.

2 Articles, citations and value

Let us define Θ_i^k, the structured set of institution i's publications in domain k as follows: $\Theta_i^k = \left\{ f_i^k(s) \middle| \forall s \in \mathbb{N} \right\}$, with $f_i^k(s)$ the publication performance with "visibility" s. In this paper we consider that, within a given domain, visibility is measured by the number of citations received in a given time window after publication.[3] The precise possible computations of scientific performance $f_i^k(s)$ are presented in detail in Section 6. It is not useful to go into such details at this stage. It is only necessary to know that $f_i^k(s)$ is continuous by construction, non negative and becomes null after some value of s which will differ across institutions and fields. The total number of i's publications with visibility less than or equal to x is simply given by $F_i^k(x) = \int_0^x f_i^k(s)\, ds$.

Let's now define the function $v^k(s)$ as the valuation function which gives the "value" of any article given its visibility s. The function $v^k(s)$ is domain dependent and built as follows:

$$v^k(s) = \omega_k \cdot v(s), \tag{1}$$

with ω_k a field dependent normalization constant that traduces the various average levels of visibility of the domains. In the next section, several dominance relations are defined for different assumptions one can make on the function $v(s)$. Before doing so, let us define the value of the whole publication performance of institution i in domain k as follows:

$$V^k\left(\Theta_i^k\right) = \int_0^\infty v^k(s) f_i^k(s)\, ds. \tag{2}$$

Notice that the function $V^k(\cdot)$ has the Von Neuman - Morgenstern expected utility form. We write $V\left(\Theta_i^k\right) = V^k\left(\Theta_i^k\right)/\omega_k$ the non-normalized value of the scientific production in a given domain. Since it lighten slightly the exposure, it will be used for the within domains comparisons of institutions whereas the normalized value is preferably used in the interdisciplinary approach.

3 Dominance relations

We now introduce several dominance relations: strong dominance, dominance and weak dominance. Each dominance relation holds for different assumptions on the function $v(s)$. Lemmas establish the

[3] Alternatively, one could also consider the journals Impact Factor as an appropriate (though more indirect) visibility measure. More complex measurement techniques could also take into account where do citations come from, for instance as suggested by Palacios-Huerta and Volij (2004).

necessary and sufficient conditions for each dominance relation to hold. Next upward dominance relations are introduced. Upward dominance relations only depart from the former dominance relations in that they only take into account the scientific production among the most visible articles in the field.

3.1 Strong dominance

We now define the notion of strong dominance over the set of institutions I. This dominance relation requires only the function $v(s)$ to be non negative, that is no article will contribute negatively to the scientific performance of any institution. This assumption is likely to be consistent for any different precise understanding of the value notion.

Definition 1 *The scientific production of institution i in field k strongly dominates the one of institution j, noted $i \blacktriangleright_k j$, if, for any non negative function $v(\cdot)$, $\int_0^\infty v(s) f_i^k(s)\, ds \geq \int_0^\infty v(s) f_j^k(s)\, ds$.*

Lemma 1 *$i \blacktriangleright_k j$ if and only if $\forall x \in [0, \infty[\,, f_i^k(x) \geq f_j^k(x)$.*

Proof. *See Appendix A.*

That is, a university strongly dominates another one in field k, if it's publication performance is greater in all possible citation classes.

3.2 Dominance

We now turn to the notion of dominance. The dominance relation requires the function $v(s)$ to be again non negative and now also non decreasing, that is articles which receive more citations have a higher value (within a given domain). This assumption is again likely to be consistent for different value notion. More cited papers have been more noticed: they are likely to have contributed more to the advancement of knowledge. They also contribute more to the visibility of the institution.

Definition 2 *The scientific production of institution i in field k dominates the one of institution j, noted $i \rhd_k j$, if, for any non decreasing function $v(\cdot)$, $\int_0^\infty v(s) f_i^k(s)\, ds \geq \int_0^\infty v(s) f_j^k(s)\, ds$.*

Lemma 2 *$i \rhd_k j$ if and only if $\forall x \in [0, \infty[\,, \int_x^\infty f_i^k(s)\, ds \geq \int_x^\infty f_j^k(s)\, ds$.*

Proof. *See Appendix A.*

3.3 Weak dominance

The notion of weak dominance requires the function $v(s)$ to be in addition to the above mentioned properties to be also weakly convex. That assumption implies that the value function gives proportional or more than proportional weight to highly cited papers. If $v(s)$ accounts for the social value of knowledge, this assumption would be barely acceptable since the arrival of citations is known to be a cumulative process. More cited papers are also more widely known and are thus more likely to be cited. More, citation fads tend to appear so that the number of citations would increase more than proportionally with real social value[4]. Alternatively, if $v(s)$ accounts for the contribution of the paper to the visibility of the parent institution, then the weak convexity assumption is more likely to be consistent since universities CEOs and their trustees are usually more willing the institution to be performing in the higher segments of visibility whereas weakly cited papers tend to be less than proportionally considered.

Definition 3 *The scientific production of institution i in field k weakly dominates the one of institution j, noted $i \trianglerighteq_k j$, if, for any non decreasing and weakly convex function $v(\cdot)$, $\int_0^\infty v(s) f_i^k(s)\, ds \geq \int_0^\infty v(s) f_j^k(s)\, ds$.*

Lemma 3 *$i \trianglerighteq_k j$ if and only if $\int_0^\infty s f_i^k(s)\, ds \geq \int_0^\infty s f_j^k(s)\, ds$.*

Proof. *See Appendix A.*

3.4 Upward dominance relations

We now consider the possibility that the dominance relations introduced so far are to be established only for high citation classes, that is only considering a certain share of the most cited articles in the field. To do so, we first need to define s_k^ϕ as the smallest visibility a paper may exhibit among the ϕ% most visible papers in the field k (in the world). We can then define, for each type of dominance, its corresponding upward dominance notion and introduce the necessary and sufficient conditions for such dominance relation to hold. This is the purpose of the following three definitions and lemmas. The proofs of the lemmas are obviously extended from the proofs of the three first lemmas (see appendix) and are thus omitted.

Definition 4 *The scientific production of institution i in field k upward strongly dominates at order ϕ the one of institution j, noted $i \blacktriangleright_k^\phi j$, if, for any positive function $v(\cdot)$, $\int_{s_k^\phi}^\infty v(s) f_i^k(s)\, ds \geq \int_{s_k^\phi}^\infty v(s) f_j^k(s)\, ds$.*

Lemma 4 *$i \triangleright_k^\phi j$ if and only if $\forall x \in \left[s_k^\phi, \infty\right[, f_i^k(x) \geq f_j^k(x)$.*

Proof. *The proof derives trivially from Lemma 1.*

Definition 5 *The scientific production of institution i in field k upward dominates at order ϕ the one of institution j, noted $i \triangleright_k^\phi j$, if, for any positive and non decreasing function $v(\cdot)$, $\int_{s_k^\phi}^\infty v(s) f_i^k(s)\, ds \geq \int_{s_k^\phi}^\infty v(s) f_j^k(s)\, ds$.*

Lemma 5 *$i \triangleright_k^\phi j$ if and only if $\forall x \in \left[s_k^\phi, \infty\right[, \int_x^\infty f_i^k(s)\, ds \geq \int_x^\infty f_j^k(s)\, ds$.*

[4]The distribution of paper according to the number of citations received is known to be highly skew: few researchers publish many articles and many researchers each publishing only few papers. The shape of the distribution can be well approximated by an inverse power distribution (Power Law) given by the following function: $f(n) = an^{-k}$ with $f(n)$ the number of authors having published n papers, a and k being the parameters of the law. When $k = 2$, this expression is identical to the one initially proposed by Lotka (1926). Many empirical studies confirmed of the relevance of such distribution for different scientific domains: *e.g.* Murphy (1973) for Humanities, Radhakrishnan and Kernizan (1979) in Computer sciences, Cox and Chung (1991) in Economics, etc.

Proof. *The proof derives trivially from Lemma 2.*

Definition 6 *The scientific production of institution i in field k upward weakly dominates at order ϕ the one of institution j, noted $i \trianglerighteq_k^\phi j$, if, for any positive, non decreasing and weakly convex function $v(\cdot)$, $\int_{s_k^\phi}^\infty v(s) f_i^k(s)\, ds \geq \int_{s_k^\phi}^\infty v(s) f_j^k(s)\, ds$.*

Lemma 6 $i \trianglerighteq_k^\phi j$ *if and only if* $\int_{s_k^\phi}^\infty s f_i^k(s)\, ds \geq \int_{s_k^\phi}^\infty s f_j^k(s)\, ds$.

Proof. *The proof derives trivially from Lemma 3.*

It should be noticed that the upward dominance relations are obviously generalization of the dominance relations introduced in Definitions 1, 2 and 3 which are equivalent to upward dominance relations at order 1.

4 The properties of dominance relations

This section presents the properties of the dominance relations. It is first shown that all dominance relations share the transitivity property. Next, a theorem which summarizes the causal relations between dominance relations is introduced.

4.1 Transitivity

Lemma 7 *All the dominance relations introduced are transitive, that is, with symbol $\succ$ accounting for any one of the dominance relations introduced ($\blacktriangleright_k^\phi, \triangleright_k^\phi$ or $\trianglerighteq_k^\phi$, $\forall \phi \in]0,1]$), if $i \succ j$ and $j \succ h$, then $i \succ h$.*

Proof. *The proof derives trivially from the Definitions 1 to 6.*

4.2 Relations between dominance relations

The definition below establishes that a dominance relation is stronger than another one if a dominance of the former type between two institutions necessarily implies a dominance of the latter between these two institutions. .

Definition 7 *A dominance relation $\succ$ is stronger than dominance relation $\succ'$, noted $\succ \gg \succ'$, if, $\forall i, j$, $i \succ j$ implies $i \succ' j$.*

The theorem introduces causal relations between the dominance relations.

Theorem 8 $\forall \phi, \phi' \in [0,1]$ *such that* $\phi \geq \phi'$, $\blacktriangleright_k^\phi \gg \triangleright_k^\phi \gg \trianglerighteq_k^\phi$, $\blacktriangleright_k^\phi \gg \blacktriangleright_k^{\phi'}$, $\triangleright_k^\phi \gg \triangleright_k^{\phi'}$ *and* $\trianglerighteq_k^\phi \gg \trianglerighteq_k^{\phi'}$

Proof. *The proof derives straightforwardly from the Definitions 4, 5, 6, and 7.*

Thus the weaker a dominance relation the more dominance relations it is possible to establish between the institutions of any given set I.

5 Rankings and reference classes

We now turn to the use of the dominance relations for the implementation of ranking procedures and reference classes. For that purpose, we first define dominance networks which constitute both a robust manner to write down formally the structuration of dominance relations between institutions of a given set and a convenient way to picture it. Next two types of dominance ranking are introduced (complete dominance ranking and pseudo dominance ranking). Lastly, it is shown how the dominance relations can be used to implement reference classes.

5.1 Directed dominance networks

We can define a directed network of dominance relations between institutions. Let us consider $\succ$, which could be any one of the dominance relations examined above ($\blacktriangleright_k^\phi, \rhd_k^\phi$ or $\unrhd_k^\phi$, with $\forall \phi \in]0,1]$). Let's build the (directed) dominance network $\vec{g}$ associated to dominance relation $\succ$ and the institutions set I by writing a link from institution i to institution j if i dominates j. That is formally: $\forall i, j \in I, ij \in \vec{g}$ if $i \succ j$.

In this network, transitive dominance triplets (without ex aequos) are uninformative since we know that the transitivity property holds. Therefore, for picturing purposes, it is convenient to define the adjusted dominance network $\vec{g}'$ which is derived from $\vec{g}$ by deleting such triplets. Formally, we begin to build $\vec{g}'$ by assigning a link from i to j if $ij \in \vec{g}$. But $\forall i, j, k \in I$, if $ij, ik, jk \in \vec{g}$ and $kj \notin \vec{g}$ then $ik \notin \vec{g}'$.

5.2 Complete dominance rankings

A ranking over a set I, written R_I is a vector of dimension $\#I$ with unitary element $r_i \in \mathbb{N}$ giving the rank of institution i. A ranking, conventionally written $R_I^\succ$, is said to be based on a dominance relation $\succ$ if it is possible to attribute a rank $r_i, \forall i \in I$, from the series of inequalities $r_i < r_j$ established each time $i \succ j$.[5] Nevertheless not all dominance relations can constitute the basis of a ranking. We shall show that such ranking over set I can only be constructed on the basis of an I-complete dominance relation defined below. Such a ranking is thus called a complete dominance ranking over I.

Definition 8 *A dominance relation $\succ$ is said to be I-complete if $\forall i, j \in I, i \succ j$ or $j \succ i$.*

When a dominance relation is complete over a set I, then one can always establish (at least) such a dominance relation between any two institutions of I. Then, ranking these institutions becomes an easy and unambiguous task on the basis of that dominance relation.

Lemma 9 *A (complete) dominance ranking $R_I^\succ$ can be constructed over institutions set I on the basis of dominance relation $\succ$ if and only if $\succ$ is an I-complete dominance relation.*

Proof. *If $\succ$ is an I-complete dominance relation then $\forall i, j \in I$, it can always be inferred that $r_i < r_j$ or (and, when i and j are ex aequos) $r_j < r_i$ and thus a ranking can be established. If a*

[5] If $i \succ j$ and $j \succ i$ then i and j are *ex aequos*, that is $r_i = r_j$. The rank of *ex aequos* is set to the median of the possible ranks, conventionally choosing the minimal value when the median is ambiguous.

ranking can be established on institutions set I the it can be inferred that $r_i < r_j$ or (and, when i and j when ex aequos) $r_j < r_i$, which is equivalent to setting that $i \succ j$ or (and) $j \succ i$, and thus the relation $\succ$ is an I-complete dominance relation.$\square$

Lemma 10 *The dominance network $\vec{g}$ associated with any I-complete dominance relation $\succ$ is one-sided complete, that is $\forall i, j \in I$, with $i \neq j$, then $ij \in \vec{g}$ or $ji \in \vec{g}$.*

Proof. *Straightforward and thus omitted.*

5.3 Pseudo dominance rankings

A more indirect procedure can be established to build a ranking from a dominance relation. It relies on a score computed thanks to that dominance relation. In the pseudo dominance ranking defined below, the proposed score is the number of institutions an institution dominates.

Definition 9 *A pseudo dominance ranking $R_I'^{\succ}$ over a given set of institutions I is established on the basis of the number of dominance relations of type $\succ$ which emanate from each institution over the remaining other institutions in I. The score of institution i is $n_i = \# \{j \in I \,| i \neq j, i \succ j\}$.*

Pseudo dominance rankings are less reliable than the complete dominance rankings but have the considerable advantage that they can always be produced on the basis of any given dominance relation.

5.4 Reference classes

The fact that a dominance relation can not be established between two institutions tells that the two institutions are, to some extent, similar and thus comparable. Building on this idea, we introduce the notion of reference class: the reference class of institution i and associated with dominance relation $\succ$, is the set of institutions noted $c_i^{\succ} \subseteq I$. The definition follows.

Definition 10 $\forall k \in I$, $k \in c_i^{\succ} \subseteq I$ *if $i \not\succ k$ and $k \not\succ i$.*

Notice that then all institutions belong to their own reference class and that the relation is reciprocal (that is, if $k \in c_i^{\succ}$ then $i \in c_k^{\succ}$).

Theorem 11 *For any dominance relations $\succ$ and $\succ'$, if $\succ \gg \succ'$ then $\forall i$, $c_i^{\succ} \subseteq c_i^{\succ'}$.*

Proof. *The proof derives straightforwardly from Theorem 8 and Definition 10.*

Moreover, pseudo ranking $R_I'^{\succ}$ can be used to find reference institutions within reference classes: universities managers are interested by the universities they would like to look like among the ones their university does not differ too much.

6 Calibration exercises on French universities

Several solutions have been introduced so far to rank research institutions and produce reference classes according to several dominance relations. The purpose of this section is to test and calibrate the various possibilities thanks to publications data of French universities. The data are first presented and the calibration exercise follows.

6.1 The Data

The data come from the French national program for recording the publications of French universities (IPERU) operated by the French national institute dedicated to the production of information on sciences and technologies (OST, Observatoire des Sciences et Techniques). Due to the complexity of the French research system, there is a great variety of patterns of referencing the employing institution which in turn leads to great difficulties in recording the publication output of universities. To overcome that difficulty, the program relies upon the validation, by the research support services of the universities, of the correct list of signing patterns their scholars and researchers use. The initial list of candidate references was established by OST. It was constituted of references observed in the publications filtered out on a precise geographic basis (post codes provided by the institutions). This work is done on a yearly basis and concerns the 106 French universities and Grandes Ecoles which are associated to the French ministry of research and higher Education and which are not totally specialized in the social sciences and/or humanities.[6] Out of the 106, 5 (mainly small universities and schools) decided not to participate in the project which leaves us with 101 institutions covered by the study.

The publication and citation data come from the SCI-expanded database (Thomson-Reuters) which constitutes a reference product in scientometrics studies. It is recording the standard scientific information of all articles published in a maintained list of about 6,500 journals. These journals are selected on the basis of their impact factor, their regularity and the respect of some editorial criteria (like peer review, rules for referencing authors and cited articles). The journals are associated to detailed scientific domains (potentially to several ones) which can be aggregated in nine large domains. The eight first, exposed in Table 1, correspond to clear disciplinary lines of inquiry, whereas the ninth, labeled Multidisciplinary Sciences, corresponds mainly to those well established journal in which articles from different disciplines can be published. So as to correct for the bias that would constitute the elimination of a significant share of the best articles of all disciplines, the articles published in PNAS (Proceedings of the National Academy of Sciences USA), Science and Nature have been reallocated to their parent discipline thanks to a lexicographic work.

Since the cleaned publication data are available only from 2002, and that the citations data available only until year 2006, a two-years publications period (2002 and 2003) and a three-years citations window (2002, 2003 and 2004 for year 2002 and 2003, 2004 and 2005 for year 2003) were selected.

6.2 Computing scientific production

This subsection exposes how the publication performance is computed from the SCI-expanded database on publications and citations. Several measures could be used.[7] In this article we use the so called

[6] The study thus deos not concern the schools which are associatd to other ministeries (Defense, Industry and Agriculture mainly).

[7] We have also tested the number of papers in which each institution was mentioned. This measure meets the critics that articles usually reference several institutions which reflect either the copublication between scholars affiliated to different institutions or the affiliation of scholars to different institutions, and that journals can also be associated to several domains. Since in addition to those crittics, it does not change the results, we do not presents the results on this measure.

fractional count measure. An article a, referencing at least one address associated to institution i, brings a score of:

$$p_{i,a}^{k} = \frac{\#\left\{i \text{ mentioned in } a\right\}}{\#\left\{j\right| j \text{ mentioned in } a\}} \times \frac{1\left\{j(a) \text{ associated to domain } k\right\}}{\#\left\{k\right| j(a) \text{ associated to } k\}}. \tag{3}$$

The function $1\{.\}$ is the indicator function which is equal to 1 if the condition into brackets is equal to one and zero otherwise. The expression $\#\{.\}$ denotes the cardinal of the set defined into brackets. The first ratio of the right hand side of 3 is the weight associated to the institution i. The second ratio filters and weights for the domain k.

Then the vector of structured publication outcome of institution i in domain k, noted $\Theta_i^k = \left\{ f_i^k(s) \middle| \forall s \in \mathbb{N} \right\}$ is computed by summing over all articles a and filtering out the article scores in the sum on the right side of equation 3 on the basis of the number of citations the associated articles received. It gives:

$$f_i(s) = \sum_a p_{i,a}^k \times 1\left\{a \text{ received } s \text{ citations}\right\}.$$

6.3 Tests and calibration

Out of all the dominance relations introduced so far, we are concerned with the appropriate and robust manners to rank universities and to build reference classes. More specific questions can be addressed such as: are dominance relations complete over the set of French institutions? We will see that we have mainly negative results in this respect: complete ranking can often not be produced and, when it can be produced, it is only on very small share of the scientific production. Then, we turn toward the pseudo dominance ranking exercise and show how it can be implemented. We discuss the sensibility of the technique with the associated types of dominance relations.

6.3.1 The completeness of dominance relations

We now introduce the notion of rate of completeness of dominance relations over institutions set I, defined as follows:

$$C_{I,\succ} = \frac{\#\left\{(i,j) \in I^2 \left| i > j, i \succ j \text{ or } j \succ i \right.\right\}}{\#\left\{(i,j) \in I^2 \left| i > j \right.\right\}}, \tag{4}$$

which simple indicates the share of pairs of (distinct) institutions for which one can establish (at least) one dominance relation of type $\succ$. The rates of completeness of dominance relations $\blacktriangleright, \rhd, \unrhd$ $, \blacktriangleright^{.5}, \rhd^{.5}, \unrhd^{.5}, \blacktriangleright^{.1}, \rhd^{.1}, \unrhd^{.1}$ are given in Table 2. Though the results show that completeness varies across domains and types of dominance. Engineering sciences exhibit significantly lower rates than other disciplines across all dominance relations. Other disciplines exhibit high rates mostly between 60 and 90%.

It is to be noticed that none of the dominance relations exposed in Table 2 are I-complete (the rates are below the unity) and so no complete dominance ranking can be constructed on the basis of such dominance relations. Incomplete dominance relations can be pictured using the dominance networks notion. Adjusted dominance networks of French universities in Physics are pictured in Figure 1.

A different manner to handle the issue is to find, for each type of dominance (strong dominance, dominance, weak dominance), the largest ϕ such that such dominance relation is I-complete. Such largest ϕ are given in Table 3. The results indicate that complete dominance relations can be established only when one restricts to articles which are among less the .01% most cited ones (in the world) for all fields but Chemistry, Physics and Science of the Universe which have their largest ϕs always less than .03%. These results hold because, for all disciplines, it is quite difficult establish dominance relation among the most performing institutions.

6.3.2 Pseudo rankings and reference classes

Unlike the complete dominance ranking, pseudo dominance rankings can be built for any dominance relation. We provide in Table 4 an example of the top ranking (sixth first) in the field of physics for different types of dominance relations.

We also measure of the degree of dissimilarity between pseudo rankings based on different dominance relations with the following index:

$$D_{\succ,\succ'} = \frac{\sum_{i \in I} |r_i - r'_i|}{n(n-1)}, \tag{5}$$

where r_i is the rank of institution i according to the pseudo ranking based on dominance relation $\succ$, and r'_i is its rank according to dominance relation $\succ'$. If the two ranking are maximally dissimilar, that index is equal to the unity whereas it is equal to zero if they are identical.

We compare pseudo rankings operated with strong domination, domination and weak domination computed for three given shares of the most visible papers in the different fields k. For all domains, we find that the largest degree of dissimilarity is between the pseudo rankings build upon strong dominance and weak dominance relations. Dissimilarity is minimal between rankings built upon dominance and weak dominance relations. Lastly, dissimilarity decreases when one restricts to smaller shares of the most cited papers. Again leaving aside Engineering sciences, the dissimilarity between dominance and weak dominance rankings is never greater than 4% and is most of the time lower than 3%, which corresponds to a quite reduced volatility across ranking procedures.

References classes are to be built upon the impossibility to establish dominance relations between pairs of institutions. The nice property of reference classes (synthesized in Theorem 11) is that the weaker the associated dominance relation, the more the reference class is reduced to a smaller set of institutions. Therefore it is possible to adjust the size of the reference class by strengthening or weakening the associated dominance relation (see Table 6).

7 Conclusion

This article introduces a new economic theory for establishing binary dominance relations by disciplines between the scientific production of any set of institutions. As illustrated with the scientific production of French Universities, such a theory can ground several ranking procedures and building reference classes. Therefore, it may become a tool in the benchmarking of universities. Nevertheless, further developments are still needed towards developing an interdisciplinary approach, assessing statistical significance of dominance relations as well as further testing on different sets of institutions.

8 References

Atkinson A.B., 1987, On the measurement of poverty, Econometrica, 55, 749-764.

Beilock R. P., Polopolus L. C., Correal M., 1986, Ranking of Agricultural Economics Departments by Citations. American Journal of Agricultural Economics 68, 595-604.

Cox R.A., Chung K.H., 1991, Patterns of research output and author concentration in the economic literature, Review of Economics and Statistics 73 (4), 740-747.

Lotka, A. J. 1926, Statistics - the frequency distribution of scientific productivity. Journal of the Washington Academy of Science 16, 317- 25.

Lubrano M., Protopopescu C., 2004, Density inference for ranking european research systems in the field of economics, Journal of Econometrics 123, 345-369.

Meyer, J., 1977, Second degree stochastic dominance with respect to a function. International Economic Review 18, 477-487.

Murphy L., 1973, Lotka's Law in the humanities, Journal of the American Society for Information Science 24, 461-462.

Palacios-Huerta I., Volij O., 2004, The measurement of intellectual influence. Econometrica 72, 963-977.

Radhakrishnan T. et Kernizan R., 1979, Lotka's law and computer science literature, Journal of the American Society for Information Science, 30, p. 51-54.

von Neumann, J., Morgenstern atld O., 1944, Theory of Games and Economic Behavior. Princeton: Princeton University Press.

9 Appendix A.

9.1 Proof of Lemma 1

The *if and only if* statement shall be proven by demonstrating that the causality holds both ways.

$i)$ We first demonstrate the left-to-right implication: $i \blacktriangleright_k j \rightarrow \forall x \in [0, \infty[\,, f_i^k(x) \geq f_j^k(x)$.

Let us consider that $i \blacktriangleright_k j$ holds and let us further assume that there exists an $x_0 \in [0, \infty[$ such that $f_i^k(x_0) < f_j^k(x_0)$. Given the latter statement, one can always find a non negative function $v(.)$ such that $V\left(\Theta_i^k\right) < V\left(\Theta_j^k\right)$. For instance, if $v(.)$ is such that $v(x_0) > 0$ and $v(x) = 0$ otherwise, then obviously $f_i^k(x_0) < f_j^k(x_0)$ implies that $\int_0^\infty v(s) f_i^k(s)\, ds < \int_0^\infty v(s) f_i^k(s)\, ds$. We thus obtain a contradiction with the initial statement $i \blacktriangleright_k j$. Thus the inequality $f_i^k(x) \geq f_j^k(x)$ is always verified when i strongly dominates j.$\square$

$ii)$ The right-to-left implication, $\forall x \in [0, \infty[\,, f_i^k(x) \geq f_j^k(x) \rightarrow i \blacktriangleright_k j$, is immediate.

Indeed when $\forall x \in [0, \infty[\,, f_i^k(x) \geq f_j^k(x)$, we can multiply both sides by any non negative function $v(.)$ and the inequality still holds for all $x \in [0, \infty[$. We can also integrate both sides of the inequality and then we have $\int_0^\infty v(s) f_i^k(s)\, ds \geq \int_0^\infty v(s) f_j^k(s)\, ds$, that is i strongly dominates j.$\square$

9.2 Proof of Lemma 2

The *if and only if* statement shall again be proven by demonstrating that the causality holds both ways.

$i)$ Consider first the implication: $i \rhd_k j \rightarrow \forall x \in [0, \infty[\,, \int_x^\infty f_i^k(s)\, ds \geq \int_x^\infty f_j^k(s)\, ds$.

Assume that $i \rhd_k j$ and that there exists an $x_0 \in [0, \infty[$ such that $\int_{x_0}^{\infty} f_i^k(s)\, ds < \int_{x_0}^{\infty} f_j^k(s)\, ds$. For any (non negative and non decreasing) function $v(.)$ such that $v(s) = v_0 > 0$ if $s \geq x_0$ and $v(s) = 0$ otherwise, we can write $\int_{x_0}^{\infty} v_0 f_i^k(s)\, ds < \int_{x_0}^{\infty} v_0 f_j^k(s)\, ds$, which implies $\int_0^{\infty} v(s) f_i^k(s)\, ds < \int_0^{\infty} v(s) f_j^k(s)\, ds$ since $v(s) = 0$ when $s < x_0$. This leads to a contradiction with the (assumed) inequality $\int_x^{\infty} v(s) f_i^k(s)\, ds \geq \int_x^{\infty} v(s) f_j^k(s)\, ds$. Therefore the inequality $\int_{x_0}^{\infty} f_i^k(s)\, ds \geq \int_{x_0}^{\infty} f_j^k(s)\, ds$ is satisfied for all $x \in [0, \infty[$ when $i \rhd_k j$. □

ii) We now demonstrate the right-to-left implication: $\forall x \in [0, \infty[\,, \int_x^{\infty} f_i^k(s)\, ds \geq \int_x^{\infty} f_j^k(s)\, ds \rightarrow i \rhd_k j$.

First assume that $\forall x \in [0, \infty[\,, \int_x^{\infty} f_i^k(s)\, ds \geq \int_x^{\infty} f_j^k(s)\, ds$. Let us further assume that institution i does not dominates institution j. Thus, there exists an $x_0 \in [0, \infty[$ and a non negative and non decreasing function $v(.)$, such that $\int_0^{\infty} v(s) f_i^k(s)\, ds < \int_0^{\infty} v(s) f_j^k(s)\, ds \iff \int_0^{x_0} v(s) f_i^k(s)\, ds + \int_{x_0}^{\infty} v(s) f_i^k(s)\, ds < \int_0^{x_0} v(s) f_j^k(s)\, ds + \int_{x_0}^{\infty} v(s) f_j^k(s)\, ds$. For instance assuming that $v(.)$ is such that $v(s) = v_0 > 0$ if $s > x_0$ and $v(s) = 0$ otherwise, yields to the inequality $\int_{x_0}^{\infty} f_i^k(s)\, ds < \int_{x_0}^{\infty} f_j^k(s)\, ds$, which is in conflict with the anthill assumed statement. Therefore the dominance relation $i \rhd_k j$ is always satisfied when $\int_x^{\infty} f_i^k(s)\, ds \geq \int_x^{\infty} f_j^k(s)\, ds$ for all $x \in [0, \infty[$.□

9.3 Proof of Lemma 3.

The *if and only if* statement shall again be proven by demonstrating that the causality holds both ways.

i) We begin by the left-to-right implication: $i \unrhd_k j \rightarrow \int_0^{\infty} s f_i^k(s)\, ds \geq \int_0^{\infty} s f_j^k(s)\, ds$.

We assume weak dominance of i over j and inequality $\int_0^{\infty} s f_i^k(s)\, ds < \int_0^{\infty} s f_j^k(s)\, ds$. The last inequality implies is equivalent to $\int_0^{\infty} v(s) f_i^k(s)\, ds < \int_0^{\infty} v(s) f_j^k(s)\, ds$ when the function v is such that $v(x) = x$. That inequality contradicts the initial statement. Accordingly, if the institution i weakly dominates the institution j in the field k, the inequality $\int_0^{\infty} s f_i^k(s)\, ds \geq \int_0^{\infty} s f_j^k(s)\, ds$ must be true.□

ii) Consider now the right-to-left implication: $\int_0^{\infty} s f_i^k(s)\, ds \geq \int_0^{\infty} s f_j^k(s)\, ds \rightarrow i \unrhd_k j$.

Let us assume that the weak dominance relation is not verified: $\int_0^{\infty} v(s) f_i^k(s)\, ds < \int_0^{\infty} v(s) f_j^k(s)\, ds$. Then taking a function $v(.)$ such that $v(x) = x \; \forall \; x \in [0, \infty[$ yields to an obvious contradiction with the inequality $\int_0^{\infty} s f_i^k(s)\, ds \geq \int_0^{\infty} s f_j^k(s)\, ds$. Thus $i \unrhd_k j$ if $\int_0^{\infty} s f_i^k(s)\, ds \geq \int_0^{\infty} s f_j^k(s)\, ds$.

10 Appendix B. Tables and figures

k	Domain
1	Fundamental biology
2	Medicine
3	Applied biology/ecology
4	Chemistry
5	Physics
6	Science of the universe
7	Engineering sciences
8	Mathematics

Table 1. *The domains.*

		dominance relations								
domains		$\blacktriangleright$	$\triangleright$	$\trianglerighteq$	$\blacktriangleright^{.5}$	$\triangleright^{.5}$	$\trianglerighteq^{.5}$	$\blacktriangleright^{.1}$	$\triangleright^{.1}$	$\trianglerighteq^{.1}$
1	Fund. bio	0.54	0.79	0.85	0.56	0.81	0.86	0.74	0.89	0.91
2	Medicine	0.60	0.82	0.88	0.62	0.84	0.88	0.77	0.90	0.92
3	Ap. bio./ecol.	0.63	0.81	0.87	0.65	0.83	0.87	0.75	0.89	0.91
4	Chemistry	0.41	0.71	0.78	0.43	0.73	0.79	0.54	0.80	0.84
5	Physics	0.43	0.72	0.79	0.46	0.73	0.80	0.60	0.82	0.85
6	Sc. universe	0.55	0.77	0.84	0.58	0.80	0.85	0.73	0.89	0.91
7	Engineering	0.31	0.60	0.70	0.31	0.60	0.70	0.35	0.67	0.74
8	Maths	0.53	0.72	0.80	0.53	0.72	0.80	0.59	0.78	0.83

Table 2. *The rate of completness of a series of dominance relations over the set of 101 French higher Education and research institutions.*

	1	2	3	4	5	6	7	8
$\blacktriangleright$	.009	.004	.005	.022	.016	.019	.002	.009
$\rhd$	.009	.004	.008	.024	.016	.026	.002	.009
$\unrhd$	.009	.004	.008	.024	.052	.026	.007	.009

Table 3. *The largest ϕ (when it exists) such that such dominance relation is I-complete over the set of 101 French higher Education and research institutions.*

physics	dominance relations																	
	$\blacktriangleright$		$\triangleright$		$\trianglerighteq$		$\blacktriangleright^{.5}$		$\triangleright^{.5}$		$\trianglerighteq^{.5}$		$\blacktriangleright^{.1}$		$\triangleright^{.1}$		$\trianglerighteq^{.1}$	
universities	r_i	n_i	r_i	n_i	r_i	n_i	r_i	n_i	r_i	n_i	r_i	n_i	r_i	n_i	r_i	n_i	r_i	n_i
U. Paris 6	1	78	1	99	1	99	1	78	1	99	1	99	1	78	1	99	1	99
U. Paris Sud 11	2	70	2	98	2	98	2	70	2	98	2	98	2	70	2	98	2	98
ENS Paris	4	60	3	95	3	97	4	60	3	96	3	97	4	60	3	97	2	98
U. Strasbourg 1	8	52	4	89	4	93	8	52	4	89	4	93	8	52	4	91	4	94
U. Grenoble 1	3	63	5	87	5	90	3	63	5	87	5	90	3	63	5	87	5	90
INPG	6	54	6	77	6	85	6	54	6	77	6	85	6	54	6	77	6	85

Table 4. *Top-ranked French higher Education and research institutions according to pseudo dominance ranking build upon three dominance relations.*

		$\blacktriangleright$	$\triangleright$	$\trianglerighteq$		$\blacktriangleright^{.5}$	$\triangleright^{.5}$	$\trianglerighteq^{.5}$		$\blacktriangleright^{.1}$	$\triangleright^{.1}$	$\trianglerighteq^{.1}$
Fund. bio	$\blacktriangleright$				$\blacktriangleright^{.5}$				$\blacktriangleright^{.1}$			
	$\triangleright$	.041			$\triangleright^{.5}$	.044			$\triangleright^{.1}$	.037		
	$\trianglerighteq$	.054	.027		$\trianglerighteq^{.5}$	.054	.022		$\trianglerighteq^{.1}$	.038	.014	
Medicine	$\blacktriangleright$				$\blacktriangleright^{.5}$				$\blacktriangleright^{.1}$			
	$\triangleright$	.045			$\triangleright^{.5}$	.043			$\triangleright^{.1}$	.028		
	$\trianglerighteq$	.065	.029		$\trianglerighteq^{.5}$	.056	.021		$\trianglerighteq^{.1}$	.036	.014	
App. bio.	$\blacktriangleright$				$\blacktriangleright^{.5}$				$\blacktriangleright^{.1}$			
	$\triangleright$	.049			$\triangleright^{.5}$	.050			$\triangleright^{.1}$	.041		
	$\trianglerighteq$	.071	.034		$\trianglerighteq^{.5}$	.063	.025		$\trianglerighteq^{.1}$	.052	.018	
Chemistry	$\blacktriangleright$				$\blacktriangleright^{.5}$				$\blacktriangleright^{.1}$			
	$\triangleright$	.072			$\triangleright^{.5}$	.067			$\triangleright^{.1}$	.067		
	$\trianglerighteq$	.096	.033		$\trianglerighteq^{.5}$	.083	.025		$\trianglerighteq^{.1}$	.080	.020	
Physics	$\blacktriangleright$				$\blacktriangleright^{.5}$				$\blacktriangleright^{.1}$			
	$\triangleright$	.050			$\triangleright^{.5}$	.058			$\triangleright^{.1}$	.061		
	$\trianglerighteq$	.075	.036		$\trianglerighteq^{.5}$	.075	.028		$\trianglerighteq^{.1}$	.074	.021	
Sc. univ.	$\blacktriangleright$				$\blacktriangleright^{.5}$				$\blacktriangleright^{.1}$			
	$\triangleright$	.047			$\triangleright^{.5}$	.058			$\triangleright^{.1}$	.043		
	$\trianglerighteq$	.072	.035		$\trianglerighteq^{.5}$	.076	.025		$\trianglerighteq^{.1}$	.052	.012	
Engin.	$\blacktriangleright$				$\blacktriangleright^{.5}$				$\blacktriangleright^{.1}$			
	$\triangleright$	.079			$\triangleright^{.5}$	.079			$\triangleright^{.1}$	.098		
	$\trianglerighteq$	.107	.045		$\trianglerighteq^{.5}$	.107	.045		$\trianglerighteq^{.1}$	.111	.032	
Maths	$\blacktriangleright$				$\blacktriangleright^{.5}$				$\blacktriangleright^{.1}$			
	$\triangleright$	.063			$\triangleright^{.5}$	.063			$\triangleright^{.1}$	.059		
	$\trianglerighteq$	.082	.036		$\trianglerighteq^{.5}$	.082	.036		$\trianglerighteq^{.1}$	.076	.030	

Table 5. *The degree of similarity between pseudo rankings based on different dominance relations over the set of 101 French higher education and research institutions.*

domains		dominance relations								
		$\blacktriangleright$	$\triangleright$	$\trianglerighteq$	$\blacktriangleright^{.5}$	$\triangleright^{.5}$	$\trianglerighteq^{.5}$	$\blacktriangleright^{.1}$	$\triangleright^{.1}$	$\trianglerighteq^{.1}$
1	Fund. bio	47(21)	22 (12)	16 (9)	45 (22)	20 (12)	15 (9)	27 (23)	12 (11)	10 (9)
2	Medicine	41 (20)	19 (10)	13 (8)	39 (20)	17 (10)	13 (8)	24 (22)	11 (11)	9 (9)
3	Ap. bio./ecol.	38 (22)	20 (13)	14 (10)	36 (23)	18 (13)	14 (10)	26 (22)	12 (11)	10 (9)
4	Chemistry	60 (20)	30 (13)	23 (11)	58 (21)	28 (13)	22 (11)	47 (26)	21 (14)	17 (11)
5	Physics	58 (18)	29 (13)	22 (11)	55 (21)	27 (13)	21 (11)	41 (26)	19 (15)	16 (13)
6	Sc. universe	46 (20)	24 (13)	17 (10)	43 (21)	21 (13)	16 (11)	28 (21)	12 (9)	10 (7)
7	Engineering	70 (19)	41 (15)	31 (13)	70 (19)	41 (15)	31 (13)	66 (21)	34 (15)	27 (13)
8	Maths	48 (23)	29 (15)	21 (12)	48 (23)	29 (15)	21 (12)	42 (24)	23 (15)	18 (12)

Table 6. *The average size (std. errors) of reference classes in the different domains according to different dominance relations.*

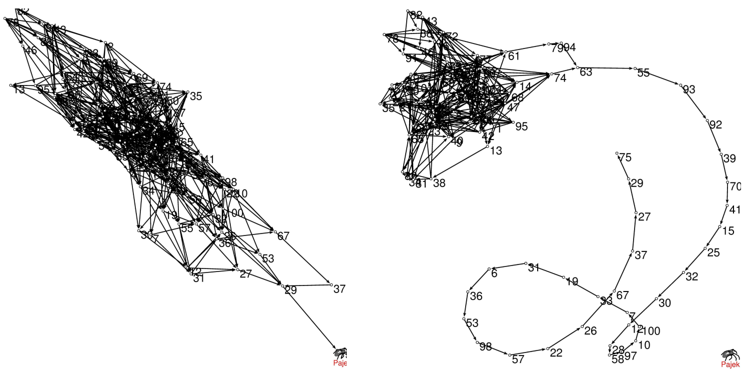

Figure 1. Adjusted dominance networks $\vec{g}'$ built for dominance relations $\rhd_5^{.5}$ (left graph) and $\rhd_5^{.05}$ (left graph) (domain 5 is Physics).

Cahiers du GREThA
Working papers of GREThA

GREThA UMR CNRS 5113

Université Montesquieu Bordeaux IV
Avenue Léon Duguit
33608 PESSAC - FRANCE
Tel : +33 (0)5.56.84.25.75
Fax : +33 (0)5.56.84.86.47

www.gretha.fr

Cahiers du GREThA (derniers numéros)

2008-17 : YILDIZOGLU Murat, *Reinforcing the patent system? Effects of patent fences and knowledge diffusion on the development of new industries, technical progress and social welfare*

2008-18 : ROUILLON Sébastien, *On the Existence of Anonymous and Balanced Mechanisms Implementing the Lindahl Allocations*

2008-19 : MALFAIT Jean-Jacques, PAJOT Guillaume, *Séquestration des flux de carbone forestier : rotations des peuplements, prise en compte des produits bois et optimisation des stocks de carbone*

2008-20 : LAYAN Jean-Bernard, LUNG Yannick, *Attractivité et agglomération de l'industrie automobile au Maroc et en Tunisie : une analyse comparative*

2008-21 : CABANNES Michel, *La place de la sphère résidentielle dans le développement territorial : Quelques éléments d'appréciations*

2008-22 : NICET-CHENAF Dalila, ROUGIER Eric, *Recent exports matter: export discoveries, FDI and Growth, an empirical assessment for MENA countries*

2008-23 : MAGDALOU Brice, MOYES Patrick, *Social Welfare, Inequality and Deprivation*

2008-24 : BERR Eric, *Le développement soutenable dans une perspective post keynésienne : retour aux sources de l'écodéveloppement*

2008-25 : BERROU Jean-Philippe, COMBARNOUS François, *Ties configuration in entrepreneurs' personal network and economic performances in African urban informal economy*

2008-26 : AMABLE Bruno, LUNG Yannick, *The European Socio-Economic Models of a Knowledge-based society. Main findings and conclusion*

2008-27 : MAROUANE Alaya, NICET-CHENAF Dalila, ROUGIER Eric, *The law of growth and attraction: an endogenous model of absorptive capacities, FDI and income for MENA countries*

2008-28 : OLTRA Vanessa, *Environmental innovation and industrial dynamics: the contributions of evolutionary economics*

2009-01 : MONTALBAN Matthieu, *L'influence de la financiarisation sur les modèles productifs dans l'industrie pharmaceutique : domination et contradictions de la conception du contrôle blockbuster*

2009-02 : Nicolas CARAYOL, Agenor LAHATTE, *Dominance relations and universities ranking*

La coordination scientifique des Cahiers du GREThA est assurée par Sylvie FERRARI et Vincent FRIGANT. La mise en page est assurée par Dominique REBOLLO.