



Research Division
Federal Reserve Bank of St. Louis
Working Paper Series



Academic Rankings with RePEc

Christian Zimmermann

Working Paper 2012-023A
<http://research.stlouisfed.org/wp/2012/2012-023.pdf>

July 2012

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

The views expressed are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Federal Reserve Bank of St. Louis Working Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to Federal Reserve Bank of St. Louis Working Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors.

Academic Rankings with RePEc*

Christian Zimmermann
Federal Reserve Bank of St. Louis

July 20, 2012

Abstract

This document describes the data collection and use of data for the computation of rankings within RePEc (Research Papers in Economics). This encompasses the determination of impact factors for journals and working paper series, as well as the ranking of authors, institutions, and geographic regions. The various ranking methods are also compared, using a snapshot of the data.

*This paper benefited from discussions and electronic correspondence with Kit Baum, Oded Galor, Bill Goffe, N. Gregory Mankiw, and Ekkehard Schlicht. The data used in these rankings would not exist without the major contributions of José Manuel Barrueco Cruz, Kit Baum, Sune Karlsson, Thomas Krichel, Ivan Kurmanov and all the other volunteers working on RePEc. This version updates the previous ones with several criteria that have been added since the last version, as well as updates for the tables. The views expressed are those of individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Contents

1	Introduction	4
2	Data Gathering	5
2.1	Bibliographic Data	5
2.2	Author Data	5
2.3	Institutional Data	6
2.4	Citation Data	7
2.5	Abstract Views and Downloads Data	8
2.6	Further Refinements of the Data	9
2.7	Discussion of Coverage	10
3	Computation of Impact Factors and Ranking of Series or Journals	10
3.1	Simple Impact Factors	10
3.2	Recursive Impact Factors	11
3.3	Discounted Impact Factors	12
3.4	Recursive Discounted Impact Factors	12
3.5	H-Index	12
3.6	Abstract Views	12
3.7	Downloads	12
3.8	Aggregation	12
3.9	Discussion	13
4	Ranking of Works	13
5	Rankings of Authors	14
5.1	Criteria Based on the Number of Works	14
5.2	Criteria Based on Citation Counts	15
5.3	Criteria Based on Journal Page Counts	16
5.4	Criteria Based on Popularity on Reporting RePEc Services	17
5.5	Criteria Based on Co-Authorship Networks	18
5.6	Aggregation of Criteria	19
5.6.1	Harmonic Mean of Ranks	19
5.6.2	Arithmetic Mean of Ranks	19
5.6.3	Geometric Mean of Ranks	19
5.6.4	Lexicographic Ordering of Ranks	20
5.6.5	Graphicolexic Ordering of Ranks	20
5.6.6	Sum of Percent of Best in Criterion	20
5.6.7	Exclusion of Extremes	21
5.6.8	Discussion and Aggregation Choice	21
6	Ranking of Institutions	21
7	Ranking of Geographic Regions	23

8	Other Rankings	23
8.1	Ranking within Geographic Regions	23
8.2	Ranking of Female Economists	24
8.3	Ranking of Young Economists	25
8.4	Ranking of Deceased Economists	25
8.5	Ranking of Top-Level Institutions	25
8.6	Ranking within Fields	26
9	A Glimpse at Results	26
9.1	Impact Factors	27
9.2	Works	27
9.3	Authors	27
9.4	Institutions	29
10	Comparison with Other Ranking Methodologies	29
10.1	What RePEc Can Do and Others Not	29
10.2	What RePEc Cannot Do	30
11	Conclusions	30
12	References	31

List of Tables

1	Rank correlations of series	32
2	Rank correlations of series (top 200 series in each pannel)	33
3	Average impact factors	34
4	Rank correlations of scores for top items by criteria	34
5	Rank correlations across criteria for authors, full sample	35
6	Rank correlations across criteria for authors, top 1000 authors	36
7	Rank correlations across aggregate criteria for authors, full sample	37
8	Rank correlations across aggregate criteria for authors, top 1000 authors	37
9	Average correlations across criteria for authors	38
10	Rank correlations across criteria for institutions, full sample	39
11	Rank correlations across criteria for institutions, top 250 institutions	40
12	Rank correlations across aggregate criteria for institutions, full sample	41
13	Rank correlations across aggregate criteria for institutions, top 250 institutions	41
14	Average correlations across criteria for institutions	42

1 Introduction

RePEc has become an important bibliographic service for economics and related fields. A considerable amount of data has been collected regarding who authored which paper, where it was published, who reads it and where it is cited. One way to use this wealth of data is to compute rankings of individuals, journals (and series), institutions, and even countries. Along with the growth of the underlying data, these rankings, even though they are still experimental, have grown in importance in the profession. Indeed, there is evidence that they are used more and more for evaluation purposes (promotion and tenure decisions) and even hiring. Also, country-specific rankings have been used in various professional publications and even the popular press.

It is therefore time for the methodology behind these rankings to be explained. While a criterion such as the number of citations may appear to be simple, it is necessary to understand how it is computed. Indeed, for ranking purposes in RePEc, self-citations are not counted, but citations to other versions of an articles are counted. It is also important to understand how the citations are extracted, i.e., what citations can be considered in the statistics.

Compared with other ranking exercises, the present one also includes some criteria that are unique, such as those based on readership, those based on the number of authors citing, and those based on centrality among co-authors. It is also rare to find the same source being used both to establish impact factors of publications and rankings of authors or institutions. Finally, no other effort has included working papers, which have now become a very important way to disseminate research in economics, if not the most important.

The RePEc project would never have been possible without the efforts of the many volunteers that have participated in one way or another: the maintainers of the so-called RePEc archives who contribute the basic bibliographic data and all those who have contributed through their programming skills, making available hardware and/or bandwidth, giving advice or simply spreading the word about RePEc. RePEc is committed to honor the work of these volunteers by making sure their work will never be subject to fees, both for publishers and users, and will remain in the public domain.

The rest of the paper is structured in the following way. Section 2 describes how the various components of the data used in the rankings are gathered. Section 3 details the construction of the impact factors. Section 4 describes how articles and working papers can be ranked. The various criteria used to rank authors are introduced in Section 5, which also discusses the various ways these criteria can be aggregated and justifies the choices made for the “official” rankings. Sections 6, 7, and 8 present the procedures to rank, respectively, institutions, geographic regions, and finally other rankings. Section 9 takes a snapshot of the data and documents the concordance of the the various rank criteria. Section 10 discusses how RePEc rankings differ from other rankings. Section 11 concludes.

2 Data Gathering

This section describes how all the data are gathered to obtain the sources underlying for the rankings. All data come from RePEc and other projects related to RePEc. These data are continuously updated, and the rankings are refreshed on a monthly basis.

2.1 Bibliographic Data

The source of all the bibliographic data is RePEc. RePEc (Research Papers in Economics, <http://repec.org/>) was founded in June 1997 under the leadership of Thomas Krichel as a followup project to NetEc, founded in 1993. Under very little central management, publishers (commercial or academic) contribute the bibliographic data (called metadata) themselves using a common format. These data are provided through the servers of the publishers, which anybody can access and use. Thus RePEc is just a scheme to organize metadata and make it available in the public domain.

At the time of this writing, almost 1500 archives were contributing metadata to RePEc, thus covering: 3400+ series with 460,000 working papers, 1500 journals with 720,000 articles, 15,500 book chapters, 12,000 books, and 2,700 software components, for a total of over 1,200,000 items. Almost 1,100,000 of them are available for download in full text.

So-called RePEc services are then allowed to use these data to freely provide public access to them. Several websites directly display the data collected through RePEc, the most popular being IDEAS (<http://ideas.repec.org/>), EconPapers (<http://econpapers.repec.org/>), Inomics (<http://inomics.com/>), and finally Socionet (<http://socionet.ru/>).¹ An email notification service for new on-line working papers is also available (NEP, <http://nep.repec.org/>). Finally, data gathered by RePEc are relayed through the Open Archives Initiative and therefore made available even more widely, but to services that do not specialize in economics, such as Google Scholar and Oyster.

2.2 Author Data

For any ranking, one needs to collect information about the publications of an author. One great difficulty is the many ways an author's name may be indexed. For example, John Maynard Keynes may be listed in the bibliographic metadata as

1. John Maynard Keynes
2. John M. Keynes

¹NetEc, with its child projects WoPEc and BibEc, had display RePEc data at one time. NetEc closed, as it was not worth the maintenance effort given that competitors within RePEc were offering a superior product, according to its maintainer. Econlit also uses RePEc data for working papers through an exchange of services agreement with RePEc.

3. John Keynes
4. J. M. Keynes
5. J. Keynes
6. Keynes, John Maynard
7. Keynes, John M.
8. Keynes, John
9. Keynes, J. M.
10. Keynes, J.

and one can imagine many other ways, including misspellings. Variations are even more numerous if nicknames, titles or suffixes (Jr., Sr., III) are used or if accents are used. In addition, several people may have the same name, especially if the first name is abbreviated. Thus, an automated attribution of works to authors is bound to have a high level of errors. Human intervention is necessary here.

The best people to perform this intervention are the authors themselves. To do this, they register with the RePEc Author Service at <http://authors.repec.org/>. In doing so, they provide contact details, their affiliations (see next section), and their name variations expected in the metadata. The search engine then suggests to them works from the RePEc metadata that match the name variations, works that the author then can add to their profile.

One may ask why authors would go through that trouble. There are several incentives (Krichel and Zimmermann 2009). First, without being registered in the RePEc Author Service, an author is not ranked and his research output does not count toward the ranking of the institutions he is affiliated with. Second, when registered, an author obtains notification of new citations that are found within RePEc, a compilation of all citations, as well as a detailed ranking analysis every month.

At the time of this writing, over 32,000 authors were registered, claiming over 750,000 works as theirs, somewhat less than half of all the works listed in RePEc once the double-counts (works claimed by several authors) are taken into account.

The RePEc Author Service is based at the Economic Research Division of the Federal Reserve Bank of St. Louis and is monitored by the author of this paper. It runs on open source software written by Ivan Kurmanov and financed by a grant from the Ford Foundation, with extension funding provided by the Federal Reserve Bank of St. Louis.

2.3 Institutional Data

Institutional data are based on the institutional records collected since 1995 in EDIRC (Economics Departments, Institutes, and Research Centers in the

World, <http://edirc.repec.org/>). This website collects links to academic institutions and government agencies that would principally employ economists. The data are quite accurate; for example it lists within a university all relevant departments (economics, finance, agricultural economics, business schools, and sometimes public policy and similar departments), research centers, institutes, formal research groups, and some chairs are listed as long as economists form a substantial part of the staff or economic issues are prominent in the mission of the group. A second condition is that this listed entity have its own website. It does not need to have its own server (virtual or not), but it needs to have a web page that is more substantial than just a listing of classes: there should be at least a listing of faculty by name.

Entities not based in universities can also be listed. The obvious ones are central banks and government agencies directly applying economic policy, say ministries of finance, treasury, labor, and industry, but also statistical agencies and various research agencies. The same applies to international organizations. Finally, independent research institutes and think tanks are also listed, but not most commercial institutions (banks, consultants). The only exceptions are those that have a RePEc archive or that provide substantial research for free through their website. Associations and societies are also listed.

All in all, over 12,700 institutions are listed, almost 6,000 of which are associated with an author registered in the RePEc Author Service (not counting those without claimed works). If they are specialized in a particular field, they are categorized, and almost all governmental agencies are categorized. Institutions are also categorized by countries or, in the case of the United States, by state. When authors register with the RePEc Author Service, they have the opportunity to specify with which institutions they are affiliated with among those listed in EDIRC (except associations and societies), but they can also suggest new entities. If they do not fit within the criteria of EDIRC, they are still kept in their list of affiliations without a link to an institution in EDIRC.

EDIRC is housed and managed at the Economic Research Division of the Federal Reserve Bank of St. Louis by the author of this paper.

2.4 Citation Data

Citation counts are often considered to be the most useful metric of the impact of a piece of research. Finding citations is, however, not a trivial matter. It can be performed either manually at great cost or automatically which is a process that needs considerable fine tuning and many exception rules.

All citation data for RePEc ranking purposes are provided by the CitEc project, <http://citec.repec.org/>, managed by José Manuel Barrueco Cruz, librarian at the University of Valencia. CitEc runs on hardware provided by the Valencian Economic Research Institute.

CitEc downloads all papers in pdf format it can find, typically those that are not hidden behind a password or some IP protection. Those pdf files are then successively converted to PostScript and text. The text is then parsed to recognize the references, which are then paired with items listed in RePEc

with a fuzzy matching algorithm on titles and authors. To prevent erroneous attributions, the level of confidence for a match needs to be set quite high. For somewhat lower levels of confidence, registered authors have the option to check and add appropriate citations.

At the time of this writing, over 360,000 documents have been processed, extracting over eight million references, over three million of which refer to over 400,000 items listed in RePEc. Given that only freely available documents can be analyzed, a large part of those documents are working papers. This has advantages and disadvantages. Working papers are typically more recent than published articles, thus it allows a much more up-to-date analysis than with articles alone. However, citations in published articles are considered to be much more valuable than in working papers (erroneously, as discussed further in a subsequent section). This is partially corrected in three ways: 1) publishers directly provide information to CitEc about references in their articles, either because their content is gated or because they want to increase the quality of matches; 2) for authors who have both the working paper and published article version of an item in their profile, the references found in one version can be attributed to the other; 3) on an experimental basis, authors can add references to the database, something authors are quite keen to do to increase their citation count. The system requests that all references of a paper be added, so as to provide a positive externality for others as well.

2.5 Abstract Views and Downloads Data

Another measure of the impact of research is how often it has been “looked at.” Abstract views statistics assess the attractiveness of the title, the authors or the general topic. In addition, downloads statistics indicate how much abstracts have contributed to the attractiveness of the downloaded document.

Keeping track of abstract views is not difficult using the logs of a web server. The only drawback is that abstracts displayed during uses of the search engine cannot be counted. Downloads are more difficult, given that they typically link to external servers. Thus some mechanism needs to be put in place to keep track of downloads.

The decentralized nature of RePEc complicates the compilation of these statistics. The participating services first need to keep appropriate logs and second need to make them available in an appropriate format. The LogEc project, <http://logec.repec.org/>, managed by Sune Karlsson at Örebro University, tries collect this information. The following RePEc services provide information for downloads and abstract views: EconPapers, IDEAS, NEP, EconomistsOnline, and Socionet. The defunct NetEc also used to provide data. Other services that use RePEc data, in whole or part, unfortunately do not provide statistics. Among them are EconStor, Inomics, Econlit, Oyster, and any service making use of the RePEc data made available through the Open Archive Initiative (Google Scholar, for example).

Quite obviously, these statistics are subject to manipulation, as one could repeatedly download a paper to increase its count. For this reason, various infor-

mation about the abstract viewer or downloader are recorded to prevent repeat counts. This is mainly performed through the use of the IP address, taking also into account IP clusters. Also, and this is mostly relevant for abstract views, visits by search engine robots need to be discarded as they do not represent human readership. Some robots identify themselves, and they can easily be taken care of. Others do not obey standard protocols and need to be recognized as robots. Various identification mechanisms are used to filter these additional robots from the data. Complete details on how all this performed cannot be given here. But overall, about 80% of abstract views are thus discarded—less for downloads.

Whether is it an over-count or under-count of the true count is unknown. Some robots may slip through. Some downloads are discarded as repeated despite originating from different users because they came from the same IP clusters. This happens in particular with institutions using a single cache or proxy server. We hope, however, that the statistics are sufficiently high for such accidents to even out relatively smoothly across all documents and no bias is introduced.

In addition, various checks and balances are implemented to recognize abnormal behavior, mostly from authors trying to manipulate the statistics. Obviously, these safeguards are not revealed here, but let it be known that a human eye has a final look at the server logs in these cases and that several authors have been caught.

Despite all these adjustments, LogEc records over two million abstract views and half a million downloads a month; in other words, every document's abstract is viewed once or twice time a month, and every item available on-line is downloaded once every second month, on average for reporting RePEc services.

2.6 Further Refinements of the Data

As the works covered in RePEc contain both publications and pre-publications, there is an issue with several versions of the same work being listed. In particular, a working paper may appear in several series. Thus, for any measure that considers the numbers of works someone has authored, one should count distinct works. For technical reason, the matching of different versions is done only for works that are listed in a registered author's profile. The basis is a very similar title and the author's recognition of authorship. Manual adjustments are done when titles differ, upon request to the RePEc team.

Note that such works may have been cited in their different versions. A citation to any version is counted toward all versions. The same applies to references. Any author statistics involving a count of works or citations aggregates the data from the different versions. Such matching is not performed for abstract views and download counts.

2.7 Discussion of Coverage

Quite obviously, only journals and working paper series that are listed in RePEc can be classified, and only authors that registered themselves can be included. Thus, here are omissions. This is obviously avoidable, but the structure of RePEc puts the burden of indexing on the publishers. Unlisted authors can easily correct this by registering themselves. Missing journals and working papers series can get indexed by their publishers and they will be fully considered.

Being listed is not sufficient. The listing needs to be maintained, i.e., new items added as they are published. Some publishers are better at this task than others, be it with regard to timeliness, completeness (missing items), coverage (years covered), or data quality (syntax errors, confusing author names). Again, it is up to the publishers to do their work. And registered authors also need to maintain their profile with any additions.

Deceased authors are kept in the database, but their affiliations are removed, the logic being that they cannot contribute to the academic life of their employer anymore. The RePEc Author Service maintenance team tries to keep their profiles current. Authors whose email addresses no longer function are considered to have either moved or died. Hence, their affiliations are discarded from consideration.

Note that while some journals present in other studies are not classified here, our rankings also cover working paper series that are typically neglected by other studies. There is also a limited number of chapters and books. It turns out that some working paper series have very high impact factors, while many journals have low impact factors. It is thus wrong to believe that research is valued only when it is published in a journal. There are also software components. They are either stand-alone program code or material necessary to replicate some study. Citations to them are currently not considered, as it is often difficult to disentangle them with citations to the original works. More on this later, in the discussion of impact factors.

3 Computation of Impact Factors and Ranking of Series or Journals

Many ranking exercises for institutions or authors rely heavily on impact factors calculated elsewhere, and these impact factors are usually the most controversial issue with these rankings. Here we take a different approach in that the impact factors are determined with the RePEc data. We compute four sets of impact factors.

3.1 Simple Impact Factors

The computation of this simple impact factor is rather straightforward. Just find all citations to items in that particular series or journal, count those citations and divide by the number of items in the series or journal. Several adjustments

are performed to the number of citations: 1) self-citations within the series or journal are discarded, to prevent self-inflation. Self-citations by authors are still counted, though. 2) Considering that a work may have appeared in different series, all versions of the cited and citing work are considered, but only one is counted. This matters: For example, an article may be cited, while its working paper version is not, but the working paper series is still credited with this citation.

3.2 Recursive Impact Factors

Recursive impact factors are computed in the same way as the simple impact factors, except that every citation carries some weight. That weight is the recursive impact factor. It is thus the fix point of a function that could be specified in the following way:

$$R_I = \frac{1}{\sum_{\forall I} \sum_{i \in I} R_I} \frac{\sum_{c_J \in I} R_J}{\sum_{i \in I} 1} \quad \forall I,$$

where R_I is the recursive impact factor of series or journal I , which has items i . c_J represents all citations from journal J . To guarantee that a fix point exists, the weights are normalized such that the average item (article or working papers) has a recursive impact factor of one. Also, when there are several versions of a citing item, the one with the highest impact factor is considered.

These factors are computed by iteration. In the first pass, simple impact factors are used, and then in each pass the recursive impact factors from the previous iteration are taken. This does, however, never converge completely, as new items and citations are continuously added to the database. The results are relatively stable, though. Concretely, the weights are recomputed every day for all series and journals that are refreshed on IDEAS, that is, those that have had any amendments in the bibliographic data and those that have not been refreshed for thirty days.

The recursive impact factor computed here is similar to the Google PageRank (Brin and Page 1998), which ranks web pages higher if they are linked to many others, even more so if it is by web sites that have a high PageRank. The difference is that Google computes a different factor for every page, whereas we compute one for every journal or paper series. The idea of the PageRank is to determine the probability that a web surfer clicking randomly would end up at that page. In our case, this would be the probability, or rather something proportional to it, that a reader randomly following references in articles and papers would end up with a particular journal or working paper series.²

The recursive impact factor also bears some similarity with the Article Influence, which is a journal's Eigenfactor divided by the proportion of articles

²Strictly speaking this would only be true if we did not account for different versions of the same item. Also, the reader would need to follow all citations, as the impact factor is not divided by the number of cited items. Some versions of PageRank do this, however.

from that journal³. We have, however, not checked how close to each other the two are.

3.3 Discounted Impact Factors

This factor is similar to the simple impact factor, with one important difference: Each citation counts for the inverse of the age in years (plus one) of the citing paper. Thus, if an article is cited in a paper dated in 2009 and we are in 2012, this citations would count for 0.25.

Such a factor gives an edge to what is cited now, and therefore highlights the publications series that are hot now. It does, however, not mean that its most recent publications are well cited, only that some of them, possibly old, are well cited now.

3.4 Recursive Discounted Impact Factors

This factor is the recursive version of the discounted impact factor. It thus uses its own factors as weights, multiplied by the age factor. This highlights publications series currently well cited in series that are currently well cited.

3.5 H-Index

This statistic is typically used for authors, and hence it is more thoroughly discussed when we go through author rankings. A journal would have H-index of h if h articles have at least h citations. Quite obviously, this favors older journals or series that have a good and numerous stock of articles that attract citations. It takes quite a few years for young publications series to rank well.

3.6 Abstract Views

This criterion simply extracts the abstract views statistics from the LogEc project, using the numbers for the past twelve months.

3.7 Downloads

As for the previous criterion, download numbers are used for the past twelve months.

3.8 Aggregation

With six criteria, rankings are obviously going to differ, and every editor or publisher is going to find a favorite. There is nothing wrong with that, but one may want to have a more authoritative ranking. We suggest that aggregating these rankings may do the trick. For reasons explained below on the aggregation of author rankings, the harmonic mean of ranks of all six criteria is used.

³See <http://www.eigenfactor.org/> for details.

3.9 Discussion

Some other published impact factors differentiate by type of article, for example, by giving different weights to full articles, notes and book reviews. One may also want to discard corrigenda. The metadata do not contain the type of the article, and the title in the vast majority of cases does not allow one to infer the type. We thus abstract from these considerations.

Also, some journal issues are different. For example, the *American Economic Review* has one issue a year with non-refereed short articles, the Papers and Proceedings of the annual meeting of the American Economic Association. These short papers are less likely to be cited and add to the article count, thereby diluting the impact factor of the regular article. One could isolate these special issues, but the task then becomes subjective as other journals are subject to the same issues at varying degrees. We want to stay objective in our ranking and thus do not adjust. In this particular example, the American Economic Association does not want this distinction to be made anyway.

There are also some small sample issues. Some working paper series especially have few items and may as a results have unexpectedly high or low impact factors, high if just one item is often cited. The current solution is not to rank series or journals with fewer than 50 items. The impact factors are, however, used as is.

Finally, there is a problem when a journal changes publishers. Technically, it is now a different journal in RePEc, as its metadata are supplied from a different source. Publishers have the opportunity to record in the journal metadata what the predecessor of the successor was, but few do (or are aware they can). When recognized, this is adjusted by hand in adding pairs, or even triplets, into an exception file. Then statistics are aggregated among them.

4 Ranking of Works

There are six different ways to rank works (working papers, articles, chapters, books). One is to simply count the number of citations it has gathered, again adjusting for different versions of the same item. The second is to discount each citation by its age. The remaining four are to weigh those citations by the impact factors of the citing series or journals.

Thus, if one were to add up all citations to articles in a particular journal, then divide the result by the number of articles, one would obtain the simple impact factor (except that self-citations within the journal need to be excluded). Or if one were to add up the scores of all articles in a journal, with scores using the recursive impact factors and excluding self-citations, one would obtain the recursive impact factor. Doing this with simple impact factors would result in the factors of the first pass in the recursive impact factor computation.

RePEc publicizes rankings for the top 1% items for each ranking method. In addition, items published five years ago or more recently that are among the top 2% are also listed. As there are several criteria, one could also think of

aggregating them. Because of the large amount of data, the required computational time and the fact that these rankings are updated daily as abstract pages are refreshed, this has not been implemented.

5 Rankings of Authors

Every person registered in the RePEc Author Service with works listed in the profile is ranked. There are many ways to rank authors and this section discusses those used in the RePEc rankings. The strategy to aggregate the various rankings is then discussed.

5.1 Criteria Based on the Number of Works

The simplest of all ways to rank authors is by the number of works they have authored. However, as working papers are also considered, the same work may appear several times, in different versions. These duplicates cannot therefore be considered. A ranking including the duplicates is provided, but it is not used in the calculation of the aggregate rankings.

The number of distinct works thus serves as basis for the following criteria. They are a combination of simple counts and counts with weights from the simple or recursive impact factors with those counts either divided by the number of authors or not. Thus, the following criteria are used (with their respective labels in bold face):

1. **NbWorks**: Simple count;
2. **DNbWorks**: Count divided by number of authors on each work;
3. **ScWorks**: Count with simple impact factor weights;
4. **AScWorks**: Count with simple impact factor weights divided by number of authors on each work;
5. **WScWorks**: Count with recursive impact factor weights;
6. **AWScWorks**: Count with recursive impact factor weights divided by number of authors on each work.

The first two criteria merely indicate how prolific an author is. The four others measure one characteristic of the quality of one's work: where it was published. It is an imperfect measure, given one may simply ride on the coat tails of other papers published in the same series or journal that have been frequently cited. But such count based solely on the impact factors are the ones most frequently used, as they do not necessitate the compilation of citations if one simply takes the impact factors from somewhere else.

Note that the discounted impact factors and recursive discounted impact factors are not used here. They could also be considered, but this would put too much weight on criteria based on the number of works in the overall rankings.

5.2 Criteria Based on Citation Counts

Here, we have criteria similar to those based on the work counts, but we count citations. Self-citations are eliminated to avoid artificial and in some cases malicious inflation of citation scores. We may apply to each citation weights by any of the four impact factors, or no weight. And all these criteria may be divided by the number of authors or not.

In addition, we provide the h-index introduced by Hirsch (2005). His definition: *A scientist has index h if h of his/her N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each.* Thus, this author would have at least h^2 citations (at least h papers with at least h citations each). Such a criterion puts more emphasis on an important body of work, instead of a few very highly cited papers, by giving higher score to those who have many cited papers. This index was developed for physics, where scientists write a lot of papers and also cite rather generously. Some physicists have h above 100, but in economics it is very rare to have an h above 20, mainly due to the fact that economists write fewer, but more involved papers.

A variation of the h-index is provided, the so-called Wu-index following Wu (2008): *A scientist has index w if w of his/her N_p papers have at least $10w$ citations each, and the other $(N_p - w)$ papers have no more than $10w$ citations each.*

Finally, two criteria count the number of registered authors citing a particular author: first a simple count, second a count considering the rank of the citing author, giving more points for highly ranked citers. This can measure how widely an author is cited. For example, this penalizes those that cite each other repeatedly (“citing clubs”). Note that each co-author counts for these criteria if she has some self-citations. This is the only case where a self-citation may count. It is possible to compute these criteria thanks to the very nature of the RePEc data with author profiles. We are not aware of any other ranking using such criteria.

Thus, we have the following criteria based on citations:

1. **NbCites**: Simple citation count;
2. **ANbCites**: Citation count divided by number of authors on each work;
3. **ScCites**: Citation count with simple impact factor weights;
4. **AScCites**: Citation count with simple impact factor weights divided by number of authors on each work;
5. **WScCites**: Citation count with recursive impact factor weights;
6. **AWScCites**: Citation count with recursive impact factor weights divided by number of authors on each work;
7. **DCites**: Citation count discounted by age;

8. **ADCites**: Citation count discounted by age and divided by number of authors on each work;
9. **DScCites**: Citation count with discounted impact factor weights;
10. **ADScCites**: Citation count with discounted impact factor weights divided by number of authors on each work;
11. **WDSccites**: Citation count with recursive discounted impact factor weights;
12. **AWDSccites**: Citation count with recursive discounted factor weights divided by number of authors on each work;
13. **HIndex**: h-index;
14. **WIndex**: Wu-index;
15. **NCAuthors**: Count of citing registered authors;
16. **RCAuthors**: Rank weighted count of citing registered authors.

Due to scheduling differences between the upload of new citations and the ranking computations, the new citations are included for a minority of the authors in current ranking, but they are for all authors in the next issue of the rankings. And again, all self-citations by the author are of course excluded.

5.3 Criteria Based on Journal Page Counts

The following criteria concern only journal articles. Whether one publishes a note, which is shorter, or a full-length article is an indication how editors feel about the contribution of an article. Also, some argue that editors allow particularly good pieces to run longer, while less important works are cut. Thus the page count can be an indication of the worth of one's publication record. Again, the page count can be weighted or not and divided by the number of authors or not.

1. **NbPages**: Simple page count;
2. **ScPages**: Page count divided by number of authors on each work;
3. **WSCPages**: Page count with simple impact factor weights;
4. **ANbPages**: Page count with simple impact factor weights divided by number of authors on each work;
5. **AScPages**: Page count with recursive impact factor weights;
6. **AWScPages**: Page count with recursive impact factor weights divided by number of authors on each work;

Thus publishing a long article in an obscure journal is valued highly with the two first criteria, but barely factors in with the four others. Note that these are criteria that, in contrast to the others, pertain to a subset of all documents (articles). Also, these criteria can sometimes be somewhat misleading. For example, if a journal does not provide page numbers, either because they are missing in the metadata or because the article is on-line only and not in a paginated format, the number of pages defaults to one. This is justified by the fact that in some cases only the number of the starting page is provided, with it indistinguishable from a one-page article. In addition, these criteria do not take into account the size of the pages. Some journals publish in A4 or Letter format, whereas most have smaller formats. Font size may vary as well, thus actual content of a page could be quite different from one journal to another. No such adjustments are performed as there is no way to systematically verify those parameters and how they may change through the years, except through intensive manual labor that would count the average number of words per page or something of that order.

Note also that the discounted impact factors are not considered. Adding them would be giving more weight to publications in journals. Given that many journals have impact factors lower than working paper series, there is no particular reason to privilege journals. Let the market decide what the better publication outlet is.

5.4 Criteria Based on Popularity on Reporting RePEc Services

Here, we measure how many times document abstracts have been viewed and how often they have been downloaded. As described in the section on LogEc, these statistics pertain to the subset of RePEc services that report such statistics. Furthermore, as all the metadata collected by RePEc are in the public domain, one cannot track how much it is used. But looking at the collected subset can still give good indications. Note that these statistics are checked for multiple views or downloads, and robot and web spider activity is excluded, as described above.

Again, we provide statistics with the criteria either divided by the number of authors or not. Thus the following four criteria are available in the category:

1. **AbsViews**: Total abstract views in the past 12 months;
2. **AAbsViews**: Total abstract views per author in the past 12 months;
3. **Downloads**: Total downloads in the past 12 months;
4. **ADownloads**: Total downloads per author in the past 12 months.

Statistics are computed for the past 12 months. On the one hand, including a longer period allows the smoothing out of inherent short-term variability—for example, new papers announced through NEP get a large one-time boost, and

authors may not yet have claimed them in their profile. On the other hand, the period considered should not be too long. First, this allows one to take into account what is popular now; second, it corrects for bias stemming from items having been listed for a long time, while even older material may have been added only recently.

Note that counting abstract views and downloads starts as soon as the research item (article, paper, etc.) is added to RePEc, and these numbers are aggregated for registered authors. Thus, when an author creates a profile, the statistics for his/her papers are added also for the period where he/she was not yet registered.

For computational reasons, the criteria with statistics per author are computed with a one-month delay.

5.5 Criteria Based on Co-Authorship Networks

These two criteria have been recently included and exploit the new CollEc project, <http://collec.repec.org/>, run by Thomas Krichel of Long Island University and hosted by the Economics Department at Washington University. CollEc looks at all registered authors and computes a network of all of them, using their ties through co-authorship. Several disconnected networks emerge from this analysis, one of them encompassing the majority of the authors (as of this writing, 23994 of 32664). Within this network, the shortest path between any two authors is computed. With this, we can compute two criteria:

1. **Close**: The average number of degrees of separation through co-authorship with all other registered authors.
2. **Betweenn**⁴: The frequency the author appears on the shortest path through co-authorship between any two other registered authors.

The first measures the average number of hops through the co-authorship network that are necessary to reach all member authors. This is similar to the Erdős number mathematicians use to relate themselves to the most prolific of them, Paul Erdős. In economics, there is no such standout author, and we average over all. This is the only criterion where a smaller number is better.

The second looks at how likely a shortest path (from the set of all shortest paths) is likely to run through a particular author. Because some authors that are in the network are on the end-points of shortest paths (“dangling nodes”), the numbers of those that can be ranked with betweenness is smaller than for closeness, currently 16791. In both cases, authors with null scores are ranked just behind the last author with a score.

Why are these criteria used for rankings? For one, they measure how involved and networked authors may be. Co-authorship is penalized by other criteria, so this may be some redemption for authors who have helped many

⁴This is not a misspelling. “Between” is a reserved word in relational database management systems and can thus not be used as a variable name.

others with their work. But because these criteria do not merely measure the number of co-authors but also the centrality in the co-authorship network, it is more relevant to measuring the pre-eminence of an author. In addition, it encourages authors to get their co-authors to sign up with RePEc.

5.6 Aggregation of Criteria

Quite obviously, with so many criteria, it is difficult to agree on who the best economists are, especially as the rankings certainly do not correlate perfectly.⁵ Some way to aggregate the rankings is required and unfortunately different ways of doing so give different results. In fact, they emphasize different aspects that all have some relevance. We discuss here some of them and then discuss our choice.

5.6.1 Harmonic Mean of Ranks

The harmonic mean is defined as

$$M_{-1} = N \frac{1}{\sum_{i=1}^N \frac{1}{r_i}},$$

where r_i is the ranking of an author in criterion i . In such a mean, very good rankings have a lot of weight; for example, the first rank counts twice as much as the second one. But a one rank difference carries very little weight for higher numbers. This aggregation method therefore rewards those who are particularly good in some category, but perhaps rewards too much. For this reason, the harmonic mean is dampened somewhat by adding a constant (currently one) to each rank and then subtracting it from the mean.

5.6.2 Arithmetic Mean of Ranks

This is the easiest and most frequently used way to aggregate criteria and create indices. It is defined as

$$M_1 = \frac{1}{N} \sum_{i=1}^N r_i.$$

Doing poorly on one criterion penalizes an author particularly hard. Doing particularly well on one criterion to compensate is much more difficult. Thus, the arithmetic mean rewards those who rank consistently across criteria.

5.6.3 Geometric Mean of Ranks

The geometric mean is defined as

⁵We will discuss these correlations a few pages down.

$$M_0 = \left(\prod_{i=1}^N r_i \right)^{\frac{1}{N}},$$

where \prod symbolizes the product. The geometric mean penalizes poor rankings and emphasizes good rankings. To see this, notice that the geometric mean is the exponential of the arithmetic mean, and thus it dramatizes the features of the latter. Or put in another way, given a generalized mean with exponent p defined as

$$M_p = \left(\frac{1}{n} \sum x_i^p \right)^{\frac{1}{p}},$$

the geometric mean corresponds to $p = 0$, which is between the arithmetic mean ($p = 1$) and the harmonic mean ($p = -1$).

5.6.4 Lexicographic Ordering of Ranks

Ranking extremely well for a particular criterion is the most rewarding with this aggregation method. For an author, all ranks are ordered from best to worst, then all authors are ranked in the following way: first all those with their best rank being a first rank, the tie breaker being their second best rank, then third best. Once all authors ranked first for any criterion are exhausted, those with rank two as their best rank are taken, etc. This is akin to the ordering of words in the dictionary, hence it is named “lexicographic.” This concept is also used in economics to describe some preference classes in utility theory.

5.6.5 Graphicolexic Ordering of Ranks

This method takes the lexicographic method, but turns it on its head, hence its newly coined name: authors are ranked by their worst rank under any criterions, then their second worst rank to break ties, etc. This rewards authors that do not have a slip-up according to some criterion.

5.6.6 Sum of Percent of Best in Criterion

All the aggregation methods above consider only how someone is ranked according to the various criteria, but not far apart the ranks are apart from each other for each criterion. For example, barely being first is valued in the same way as when there is a large gap between the first-ranked and the second-ranked. One way to take the latter into account is to attribute 100% to the first ranked, and then proportionally percentages to the lower ranked authors. All these scores are then added. This aggregation method benefits the most those who have criteria where they are significantly better than others, especially for criteria where the dispersion of scores is larger.

5.6.7 Exclusion of Extremes

The truncated mean excludes the x largest and smallest values. This reduces the impact of outliers. In particular, if one thinks that the particular aggregation mean one has chosen is too much influenced by such outliers, using truncation can make the mean more credible. There is no particular guideline to choose what the value of x should be. An alternative is the Winsorized mean, where the truncated criteria are set to the rank of the largest respectively lowest remaining ranks.

5.6.8 Discussion and Aggregation Choice

We have identified 35 different criteria for ranking authors and could have easily added more. In addition, we presented six aggregation methods, which can even be varied with the number of extremes to exclude and some other degrees of freedom. Each of the criteria can be multiplied by some weight. This is a dismaying array of possibilities, but we need to make choices. Those choices are easier if the criteria or aggregation methods lead to similar results. To some extent they do, as we see in a subsequent section, but there are noticeable differences. We still need to make a choice, take a stand.

Everyone would probably favor a combination of criteria and aggregation method that would favor oneself. We need to find something that is credible, in the sense that a person outside the profession would find it agreeable. We want to highlight the particular achievement, say that an author is particularly successful in downloads despite not having published much (yet), or that an author elicited many citations despite not being prolific. The harmonic mean achieves this, but needs to be tempered somewhat, and we thus add a constant of one to each rank. Also we include all criteria but two, the simple number of works NbWorks (which does not distinguish distinct works, as multiple versions of the same work inflate this count) and the Wu-index (as it leads to a large number of ties and in particular a lot of null scores), in the aggregation. For each author, we further truncate by dropping the best and worst ranking. Thus, in summary: we consider for each author 31 rankings from a pool of 33, having dropped for all authors NbWorks and the Wu-index from the 35 presneted rankings, with aggregation through an adjusted harmonic mean.

These choices can, and should, be argued and we leave the reader the opportunity to try other ways to rank on the website.⁶

6 Ranking of Institutions

When registering, each author has the opportunity to affiliate himself with some institution(s). For those that are listed in EDIRC, the affiliation is recorded with an identifier that can be used to aggregate all authors from that institution. This allows subsequently also to rank institutions.

⁶<http://ideas.repec.org/cgi-bin/newrank.cgi>

A few rules apply. Only institutions listed in EDIRC are ranked. An author can affiliate himself with several institutions and all receive credit for that author. If an institution is a sub-entity of another institution also listed in EDIRC, the latter also receives credit. The ranking score of sub-entities is computed, but these institutions do not increment ranking counters. This allows to ask what if this sub-entity were a stand-alone one. For each criterion, the institution's score is just the sum of the scores of each affiliated author. The only exceptions are the h-index and the Wu-index, see below.

Quite obviously, institutions with many authors are advantaged. Clearly, taking an average score within an institution would make little sense, as author registration is not mandatory, and potentially lower-ranked authors may be discouraged to register. On the contrary, adding up all authors' scores gives the right incentive: everyone should register, including students who already have authored something in RePEc.

One controversial aspect, though, is how to treat authors with multiple affiliations. Until the December 2008 ranking, each affiliation counted equally and fully, which counted some authors multiple times, and some institutions with numerous "courtesy" appointments would rank much higher than expected. Since the January 2009 ranking, the rules for multiple appointments have changed in the following way. For each affiliation i , the number of registered authors is counted; call it N_i . Then, the weight of that institution is

$$w_i = \frac{1}{2} \frac{\frac{\sum_j N_j}{N_i}}{\sum_k \frac{\sum_j N_j}{N_k}} = \frac{\frac{1}{2N_i}}{\sum_j \frac{1}{N_j}}.$$

Note that these weight add up to 0.5. The remaining 0.5 is attributed to the affiliations whose website domain most closely matches the email address or personal website of the author (ties are split equally). If it is impossible to identify a principal affiliation, for example for authors without institutional homepages and with email accounts at Gmail or alumni accounts, all weights are doubled. For affiliations that are not listed in EDIRC, and thus that do not have a well-defined N_i , by default the number of authors divided by the number of institutions in EDIRC with authors is taken.

Of course, authors may disagree with the weights. Since February 2012, authors can specify the weights themselves. In fact, any change in affiliation now requires the author to set these weights, with the hope that system-set weights will gradually fade out. In the end, these weights are supposed to better take into account courtesy appointments by giving them less weight and attribute authors to the location where they mostly work.

Finally, we need to explain how the h-index is computed in the case of institutions. Remember that for authors h is defined as the number of works with at least h citations. For institutions, we follow Schubert (2007) and define the institutional h as the number of authors affiliated to that institution with an h-index of at least h . As the h can only be an integer and the support of its distribution is even smaller than for authors, there are numerous ties.

To break these ties, we adapt Ruane and Tol (2008). They augment h by a rational number between zero and one measuring the distance to the next h-index considering how many citations are required to reach it. In our case, we measure a similar distance, but we consider how many authors with appropriate h-indices are necessary to reach the next step. Note that for multiple affiliations, it is impossible to use the weights w_i discussed above. The h of member authors is fully counted toward each institution.

A final note regarding institutions: due to the nature of criteria, the measure of centrality in the co-authorship network, Close and Betweenness, cannot be computed.

7 Ranking of Geographic Regions

To rank geographic regions (countries, U.S. states), the same logic is used as for ranking institutions. All authors affiliated with institutions in a particular region are added to the pool of that region. However, authors with multiple affiliations have their scores split among all regions according to the weights discussed in the previous section.

For authors with affiliations not listed in EDIRC, the geographic location of their affiliation is guessed from the address of its web page. If it still cannot be found, then the home page of the author and then the email address are used. Obviously, this can still fail, as addresses with .com, .net, .org or .info are not geographically informative. But at least we tried.⁷

Once all these attributions are made, we simply add up the scores, properly weighted. The only exceptions are, again, the h-index, where the same scheme as for institutions is used, and the closeness and betweenness indicators in the co-authorship network. Note that we do not calculate scores for the United States as a whole, as it would obviously be number one in every aspect. Rankings for every state are given, though.

8 Other Rankings

A wealth of data is available, and this allows us to establish various other rankings. A few examples are below, and more will be added once sufficient critical mass is present to display somewhat credible results.

8.1 Ranking within Geographic Regions

Once authors have been attributed to a particular region, it is easy to rank them within that region as well. The same applies to institutions within that region. Publishing rankings with very few entities or authors do not make much

⁷Some errors are unavoidable. For example, at the time of this writing, the Pacific island nations of Niue and Nauru are ranked thanks to two authors using courtesy domains from these micro-nations.

sense, though. For this reason, a minimum of five authors or five institutions need to be present. In some regions, there is little hope for authors to be listed, whatever their prestige, due to lack of participation by others in RePEc, or in small countries, due to the lack of economists. Therefore, rankings for regional conglomerates are presented as well, say the Mountain states in the United States, Central America and the Caribbean, or Africa.

Again, we need to mention authors with multiple affiliations here. If those span several geographic regions, their score is multiplied by the appropriate weight w_i as computed above.⁸

A ranking that uses a straight excerpt from the world rankings is also provided for information (take the world ranking, and pick those from the specific region in the same order). But this ranking can differ significantly from the regional ranking for several reasons: first and as mentioned, authors with multiple affiliations across regions can only count part of their score toward a regional ranking; second, aggregate rankings are computed afresh within the region. This means that an author who far ahead in the world ranking under some criteria (say, because of very high citation counts) is still ahead under the regional ranking, but not by much. This can matter for the aggregation of ranks.

The same rules apply for ranking institutions within regions, where author scores (multiplied by relevant weights) are added. And in a similar way, regional rankings may differ from a regional extraction from the world rankings.

8.2 Ranking of Female Economists

Women are, unfortunately, quite underrepresented in the economics profession. It appears, from a limited investigation, that they are further underrepresented within RePEc. One can still try to make a meaningful ranking with data collected within RePEc. Unfortunately, an author registering with RePEc does not declare his or her gender. This needs to be inferred from the first and middle names using a name data bank. There are, however, several difficulties: some names may be used for both females and males, and this may vary by culture. Also, given the international nature of RePEc, there is an incredible diversity in first names.

The following rules are applied for gender attribution: if there is more than 90% confidence the gender is correct, it is so attributed. The ambiguous ones and the unrecognized ones are then manually entered in exception tables—one for names that were not in the original tables, the other for case by case attributions.⁹ In the end, only 0.4% are left without a gender. Close to 18% are identified as female.

The ranking of female economists is performed solely among female economists, that is, without considering the gender wide ranking: females are ranked within their group according to each criterion and then the rankings are aggregated.

⁸Before January 2009, the weight was one, which led to the perversion of rankings in some countries where it is a habit to provide courtesy appointments to foreign scholars. The new weighing scheme now ranks true residents on top.

⁹Thanks to many authors for putting a picture of themselves on their web page!

This makes it possible that the order of female economists among themselves may be different from the classification of female economists among all economists, as it happens for the regional rankings described above.

8.3 Ranking of Young Economists

It takes a long time for economists to make it into the top ranks; thus, it is of interest to compute rankings limited to young economists so that they have a chance of getting some visibility. However, the RePEc Author Service does not collect data about birth date or graduation dates. As a proxy for age or professional experience, one can use the date of the first publication, whatever its form. It is commonplace to publish at least a working paper within a year of graduation, if not before finishing studies.

There is a small percentage of records in RePEc that do not carry dates. There is nothing that can be done about that, but we can just hope that those items are not the first works of some authors. For all others, the selection criterion is that the first work be within 5, 10, 15, or 20 years of the current year, counting whole years. As obviously young economists have fewer papers and citations, the rankings are much less stable once you go past the top ones, especially for the youngest. For this reason, rankings are limited to the top 200.

8.4 Ranking of Deceased Economists

Unfortunately, economists eventually die. When we learn about a death, the deceased author is flagged and the profile continues to be maintained, as some works may still be added (posthumous publications as well as late additions to the database). By principle, deceased authors do not have affiliations. They are ranked along with the others.

As by now they are about 150 of them, it becomes possible to rank deceased authors. It does not serve any particular purpose, except that it can be done.

8.5 Ranking of Top-Level Institutions

Institution rankings are performed at the department, school, institute, or center level—that is, whichever unit has a substantial number of economists. But some are not affiliated in such units, say, a political science department. In a few cases they are senior administrators of the university. In addition, many universities have economists dispersed in several independent units that are listed in EDIRC; for example, the departments of economics, agricultural economics, public policy, and finance. The strength of a university with many such units may not be properly reflected in ranking based on units. The same applies to the Federal Reserve System whose constituents are treated as separate.

For this purpose, a separate ranking is created for “top-level” institutions. For example, anyone affiliated with any unit at Harvard University is counted toward the university. The aggregation is based on EDIRC records, and for those without an EDIRC affiliation the domain of the institutional webpage

is used (authors can submit free text affiliations, where an institutional URL is required). The Federal Reserve Banks are aggregated as well. From there, ranking is performed in the same way as other institutional rankings.

8.6 Ranking within Fields

When registering, authors do not declare a field of research. It is therefore difficult to classify them within each field, although one could try to infer it from the JEL codes attached to their papers. However, as it is customary to put several JEL codes on each paper, and only about 20% of all papers have such a code, inferred field attributions would not be reliable. However, we can attribute authors to fields by using data collected with the NEP project, <http://nep.repec.org/>.

NEP disseminates new working papers by email. At the time of writing, there are 91 field-specific NEP reports, each managed by an editor who selects from all new papers the ones fitting within her field. We use these assignments to classify authors. Thus an author who had 75% of his papers in NEP announced in field A would get 75% of his score attributed towards his ranking in that field. To be ranked, a minimum threshold of 5 papers or 25% is required. As a paper can be announced in several NEP fields, an author may have attributions adding to more than 100%.

To rank institutions within fields, author scores are added for those affiliated, using the appropriate field and affiliation weights. No minimum threshold is used, the rationale being that institutions are expected to have much more diverse expertise than individuals.

In addition, one can also use the field code in EDIRC for institutions. For example, institutions working in agricultural economics or finance are well identified. Also, certain institution types are well documented: central banks, think tanks, international organizations. For others, patterns in their names (or their English translation) are used. This is the case for economics departments and business schools. For all of them, separate rankings are released, including for U.S. Economics departments. Note that for economics departments, an effort is made to remove mis-fits and add those missed by the automatic categorization.

Note that, as for regional rankings, ranking points are computed within the set of admissible authors or institutions and thus can differ from an excerpt of the world rankings, as for other rankings of subsets described above.

9 A Glimpse at Results

We do not want to give detailed rankings here; they are constantly updated and available at <http://ideas.repec.org/top/>. In the following, we present a comparison of the various criteria and aggregation methods using a snapshot of the data on July 9, 2012, with 32,731 authors registered affiliated with 5,825 institutions.

9.1 Impact Factors

How do the impact factors compare? Table 1 provides a summary with rank correlations. All of them are very high. This is quite natural as series with many citations ought also to be cited by series with high impact factors. Overall, it does not seem to matter which criterion is used when it comes to ranking series or journals.

Looking only at the top 100 series (Table 2) correlations are reduced: the disparities between the top and worst series do not count anymore. This is reinforced when one does not filter out the series with few items, which introduce considerable noise. This is the reason they are not ranked on the web pages.

Of particular interest here is to compare the impact of journal articles relative to working papers. Table 3 shows that there is no clear winner, which could surprise many. We have to keep in mind that some journals have very low impact factors, while some working paper series have impact factors superior to most journals. Note also, as explained in the previous sections, that if the article version of a paper is cited, it counts toward both. So these numbers do not reflect where the citing author found the reference.

9.2 Works

How do the various rankings compare? Taking all articles and papers that are ranked in the top 500 in any of the six categories on February 23, 2009, and narrowing them down to those listed in all six categories, we obtain a sample of 416 items. The fact that 83% of the top 500 according to one criterion are listed in all other criteria is already an indication of high correlation. Within this (rather small) sample, the rank correlations are still fairly high, averaging 0.647 (Table 4). Rank correlations over the whole sample would be much larger, as demonstrated in other contexts below, but much more difficult to compute, for technical reasons.

9.3 Authors

We have 34 different ways to rank authors¹⁰; thus if we want to compare how differently they perform, we need to look at 1122 correlations ($34^2 - 34$). Table 5 reports them. While all these numbers can be overwhelming, the following can be extracted: The average correlation stands at 0.822 and varies between 0.561 and 0.997. The table groups the criteria in categories (number of works, citations, derived from citations, article pages, visibility on RePEc); not surprisingly, correlations within these categories tend to be higher than within other categories. It is more interesting to see where criteria seem to differ most: article pages versus co-authorship centrality on RePEc, with an average correlation of 0.646. This does not mean that they are orthogonal, though; 0.646 is still a significant correlation. But it is revealing that publishing in journals, or even in

¹⁰We do not consider the Wu-Index.

good journals more specifically, has *relatively* little to do with how much people are connected to each other.

Speaking of significance of correlations, there is a statistic that allows one to measure how independent the criteria are from each other, χ^2 . Here,

$$\chi_{n-1}^2 = (n-1)((p-1)\bar{r} + 1),$$

where $p = 34$ is the number of criteria¹¹, $\bar{r} = 0.822$ is the average correlation, $n = 32731$ is the number of authors, and $\chi_{32730}^2 = 855104$. To be significant at 5%, the statistic would need to be below 32310. Therefore, we easily reject the null hypothesis that the criteria are independent.

Looking at only the 1000 top authors (Table 6, considering the 1000 authors with the most listed works), the correlations are smaller, between 0.010 and 0.998 and averaging 0.636, but they follow the same patterns as above. The lowest correlation by criterion category is visibility on RePEc and co-authorship centrality, at 0.361. While this seems a small number, one should take into account that this is within a subsample of authors that are jointly different from the rest of the sample (they all have a lot of publications). Again, if we apply the χ^2 statistics, we find 19968, which is very far from the 5% threshold of 927.

One should expect that correlations are higher when we consider the aggregate ranking criteria than for the individual ranking criteria. This turns out to be wrong; see Table 7. They average 0.771, with a minimum of -0.112 and a maximum of 1. This is because the “percent” aggregator is not well defined for the Close criterion, where a smaller number is better.¹² Excluding the best and worst criterion for each author makes a significant impact on the overall picture, however, as the Close criterion is then excluded for most in the “percent” aggregations but has little impact otherwise. Indeed, experience shows that the exclusion of extremes can alter the rankings at the very top for a few authors with a large variance in the rankings across criteria, but it does relatively little for others. The only exception in the “percent” aggregation, where a strong lead in a category can cause a drastic reduction in ranking when it is excluded, for example. It is also remarkable that harmonic, arithmetic, and geometric aggregation methods are all very close to each other.

As for individual criteria, correlations are lower when looking at the top 1000 authors, fluctuating between 0.242 and 1 for an average 0.690; see Table 8. The patterns across aggregation methods are similar to the full sample. For additional statistics for other subsamples, see Table 9. Interestingly, in some small subsamples for lower-ranked authors, some correlations between individual criteria can get negative. Lower ranks are characterized by many ties (one or two citations, publications in series with a zero impact factors), and very little can mean large changes in rankings. But mean correlations are still high, despite these “accidents.”

¹¹We do not consider the Wu-Index.

¹²For ranking computations, the opposite of Close is used.

9.4 Institutions

The concordance of rankings across institutions is higher than that of authors for individual criteria and for aggregate criteria¹³; see Tables 10 to 14. Looking at the individual correlations, the patterns are also somewhat different compared with authors. For example, the h-index and citing authors rankings typically correlate less, while page counts and RePEc visibility correlate more. And of course, the “percent” aggregation correlates much more.

10 Comparison with Other Ranking Methodologies

The goal of this section is not to compare how the impact factors or rankings obtained by RePEc differ from other exercises.¹⁴ It is rather to highlight some of the conceptual differences: what RePEc may miss and what others may miss.

10.1 What RePEc Can Do and Others Not

The rankings described above make use of the many facets of the data collected within the RePEc project. Some of them are quite unique, which certainly gives these rankings some added value when compared with existing rankings:

1. **Timeliness:** The data in RePEc are constantly updated and the results are continuously refreshed on its websites. For example, a working paper or article is typically listed within 24 hours of the publisher indexing it, its citation analysis is released within a month, and its downloads are continuously monitored.
2. **Current affiliations:** Rankings of institutions reflect the current affiliations of authors and can take the move of an author from one affiliation to the other into account within a month. Other counts typically take into account only the affiliation at the time of publication.
3. **Pre-publications:** Established citation aggregators typically consider only citations in journals to journal articles. Even the set of journals is often severely limited. There are no such restrictions in RePEc. In fact, working papers are a very important means of dissemination in economics (and RePEc may have contributed to this) that should not be neglected. Note that analyzing working papers also significantly contributes to the timeliness of rankings.
4. **Certainty about authorship:** Given that authors acknowledge what works they have authored when they maintain their RePEc profiles, one

¹³Keep in mind that the Close and Between criteria are not computed for institutions, and the Close criterion is the one that draws down severely the aggregate correlations for authors.

¹⁴For a list of such ranking exercises, as indexed on RePEc, see <http://ideas.repec.org/k/ranking.html>

big issue in ranking authors is resolved: name ambiguities. Indeed, many publications provide only the initial of the first name. Also, there are homonyms in the profession. The use of RePEc data leave no doubt.

5. **New ranking criteria:** Thanks to the fact that authors build profiles in RePEc, it is possible to reliably count how many different authors cite a particular author. We do not know of the use of the NCAuthors, RCAuthors, Close, and Betweenn criteria elsewhere. The same applies to the h-index for journals, series, and institutions.

10.2 What RePEc Cannot Do

There is very little human intervention in anything that RePEc does. Thus various aspects of other ranking analyses cannot be performed here:

1. **Errors:** Citation analysis is very much based on automatic reference extraction from texts and pattern matching of titles. Errors can obviously happen, and probably more so than with analysis by humans. The most important case is when a list of other working papers in a particular series is printed on the last page of a paper, and this list is interpreted as the continuation of the citations. This is adjusted when reported, but affected authors have little incentive to report this. Authors can now remove citations that are not accurate, though.
2. **Adjustments:** Any criteria based on page counts can be adjusted by the size of the page or its average word count in order to truly reflect the length of the article. RePEc does not do this, as it is completely automated.
3. **Stable impact factors:** Due to the constant adjustments in RePEc, impact factors change frequently, within bounds. But this makes the use of such factors difficult for third parties.
4. **Comprehensiveness:** Some important publications are still missing in RePEc, but RePEc has no staff to index them. Also, not all authors are registered with RePEc, and some do little to maintain the accuracy of their records.

11 Conclusions

In this paper, we hope to have demonstrated that the ranking exercises performed in RePEc are based on a sound methodology and can be useful. It should also be clear that they are a work in progress, as the data are not yet as comprehensive as they could be, both in terms of listed publications and, especially, registered authors. The citation database is the component that is the most experimental¹⁵ at this point, as reference extraction and matching

¹⁵And it is mentioned everywhere in they rankings that they are experimental because of this. One metric that ca be used to remove this label is when the number of items with references exceeds the number of items that are cited.

is difficult and error prone. As more publishers and more authors join in the RePEc project, as we perfect the analysis of the data, our confidence in the rankings will rise, and we hope the RePEc rankings will be regarded as a useful tool in the profession.

12 References

Sergey Brin and Larry Page, 1998. “The Anatomy of a Large-Scale Hypertextual Web Search Engine”, *Computer Networks* vol. 30(1-7), pages 107–117.

Jorge E. Hirsch, 2005. “An index to quantify an individual’s scientific out”, *Proceedings of the National Academy of Science*, vol. 102, 16569, accessible at <http://arxiv.org/abs/physics/0508025>.

Thomas Krichel and Christian Zimmermann, 2009. “The Economics of Open Source Bibliographic Data Provision”, *Economic Analysis and Policy*, vol. 39(1), pages 143-152, March.

Frances Ruane and Richard S. J. Tol, 2008. “Rational (Successive) H-Indices: An Application to Economics in the Republic of Ireland”, *Scientometrics*, vol. 75(2), pages 395–405 May.

Andr as Schubert, 2007. “Successive h-indices”, *Scientometrics* vol. 70(1), pages 201–205.

Qian Wu, 2008. “The w-index: A significant improvement of the h-index”, manuscript, accessible at <http://arxiv.org/abs/0805.4650>.

Table 1: Rank correlations of series

Impact factor	All						
	All series						
Simple factors	1	.956	.985	.951	.03	.456	.381
Recursive factors	.956	1	.95	.993	.022	.506	.424
Discounted factors	.985	.95	1	.96	.031	.475	.419
Recursive discounted factors	.951	.993	.96	1	.023	.508	.44
H-index	.03	.022	.031	.023	1	-.03	-.024
Abstract views	.456	.506	.475	.508	-.03	1	.904
Downloads	.381	.424	.419	.44	-.024	.904	1
	Journals						
Simple factors	1	.968	.99	.966	-.014	.55	.557
Recursive factors	.968	1	.954	.994	-.017	.533	.535
Discounted factors	.99	.954	1	.965	-.013	.531	.547
Recursive discounted factors	.966	.994	.965	1	-.016	.515	.524
H-index	-.014	-.017	-.013	-.016	1	-.037	-.03
Abstract views	.55	.533	.531	.515	-.037	1	.929
Downloads	.557	.535	.547	.524	-.03	.929	1
	Working paper series						
Simple factors	1	.954	.98	.948	.027	.508	.369
Recursive factors	.954	1	.947	.992	.021	.572	.424
Discounted factors	.98	.947	1	.961	.028	.543	.433
Recursive discounted factors	.948	.992	.961	1	.023	.585	.457
H-index	.027	.021	.028	.023	1	-.024	-.015
Abstract views	.508	.572	.543	.585	-.024	1	.887
Downloads	.369	.424	.433	.457	-.015	.887	1
Impact factor	w/ ≥ 50 items						
	All series						
Simple factors	1	.963	.985	.957	-.056	.367	.341
Recursive factors	.963	1	.951	.991	-.065	.359	.335
Discounted factors	.985	.951	1	.964	-.051	.366	.364
Recursive discounted factors	.957	.991	.964	1	-.062	.352	.345
H-index	-.056	-.065	-.051	-.062	1	-.031	-.023
Abstract views	.367	.359	.366	.352	-.031	1	.903
Downloads	.341	.335	.364	.345	-.023	.903	1
	Journals						
Simple factors	1	.957	.988	.956	-.06	.477	.504
Recursive factors	.957	1	.94	.992	-.067	.448	.471
Discounted factors	.988	.94	1	.954	-.061	.45	.489
Recursive discounted factors	.956	.992	.954	1	-.069	.422	.455
H-index	-.06	-.067	-.061	-.069	1	-.039	-.033
Abstract views	.477	.448	.45	.422	-.039	1	.917
Downloads	.504	.471	.489	.455	-.033	.917	1
	Working paper series						
Simple factors	1	.963	.977	.952	-.057	.372	.258
Recursive factors	.963	1	.949	.988	-.065	.379	.272
Discounted factors	.977	.949	1	.966	-.048	.398	.322
Recursive discounted factors	.952	.988	.966	1	-.057	.394	.312
H-index	-.057	-.065	-.048	-.057	1	-.017	-.010
Abstract views	.372	.379	.398	.394	-.017	1	.905
Downloads	.258	.272	.322	.312	-.010	.905	1

Table 2: Rank correlations of series (top 200 series in each panel)

Impact factor	All						
	All series						
Simple factors	1	.096	1	.089	.999	-.065	-.108
Recursive factors	.096	1	.096	.999	.092	.246	.260
Discounted factors	1	.096	1	.089	.999	-.065	-.108
Recursive discounted factors	.089	.999	.089	1	.085	.242	.258
H-index	.999	.092	.999	.085	1	-.067	-.109
Abstract views	-.065	.246	-.065	.242	-.067	1	.920
Downloads	-.108	.260	-.108	.258	-.109	.920	1
	Journals						
Simple factors	1	1	1	1	.997	.123	.118
Recursive factors	1	1	1	1	.997	.124	.119
Discounted factors	1	1	1	1	.997	.123	.118
Recursive discounted factors	1	1	1	1	.997	.124	.119
H-index	.997	.997	.997	.997	1	.128	.121
Abstract views	.123	.124	.123	.124	.128	1	.938
Downloads	.118	.119	.118	.119	.121	.938	1
	Working paper series						
Simple factors	1	.471	1	.478	.998	.004	-.059
Recursive factors	.471	1	.471	.997	.478	.106	.030
Discounted factors	1	.471	1	.478	.998	.004	-.059
Recursive discounted factors	.478	.997	.478	1	.485	.121	.040
H-index	.998	.478	.998	.485	1	-.007	-.072
Abstract views	.004	.106	.004	.121	-.007	1	.894
Downloads	-.059	.030	-.059	.040	-.072	.894	1
Impact factor	w/ ≥ 50 items						
	All series						
Simple factors	1	.396	.998	.347	.852	.099	.069
Recursive factors	.396	1	.397	.992	.253	.288	.315
Discounted factors	.998	.397	1	.349	.843	.111	.081
Recursive discounted factors	.347	.992	.349	1	.227	.273	.306
H-index	.852	.253	.843	.227	1	-.098	-.052
Abstract views	.099	.288	.111	.273	-.098	1	.898
Downloads	.069	.315	.081	.306	-.052	.898	1
	Journals						
Simple factors	1	.858	.958	.867	.327	.277	.263
Recursive factors	.858	1	.783	.971	.207	.291	.232
Discounted factors	.958	.783	1	.841	.287	.266	.282
Recursive discounted factors	.867	.971	.841	1	.217	.293	.257
H-index	.327	.207	.287	.217	1	-.038	-.007
Abstract views	.277	.291	.266	.293	-.038	1	.91
Downloads	.263	.232	.282	.257	-.007	.91	1
	Working paper series						
Simple factors	1	.733	.897	.758	.002	.200	.145
Recursive factors	.733	1	.623	.954	-.116	.219	.151
Discounted factors	.897	.623	1	.729	.032	.273	.278
Recursive discounted factors	.758	.954	.729	1	-.099	.259	.226
H-index	.002	-.116	.032	-.099	1	-.044	-.005
Abstract views	.200	.219	.273	.259	-.044	1	.861
Downloads	.145	.151	.278	.226	-.005	.861	1

Table 3: Average impact factors

	Papers	Journals
Simple factors	3.67	2.79
Recursive factors	0.27	0.19
Discounted simple factors	0.86	0.61
Discounted recursive factors	0.31	0.19

Table 4: Rank correlations of scores for top items by criteria

	Criteria from left column					
Number of citations	1	.909	.465	.895	.831	.429
Simple factors	.909	1	.543	.783	.883	.432
Recursive factors	.465	.543	1	.416	.483	.598
Discounted citations	.895	.783	.416	1	.867	.508
Discounted simple factors	.831	.883	.483	.867	1	.564
Discounted recursive factors	.429	.432	.598	.508	.564	1

Table 7: Rank correlations across aggregate criteria for authors, full sample

exclude outliers?	harmonic		arithmetic		geometric		lexicographic		graphicolexic		percent	
	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
harmonic no	1	1	.9910	.9912	.9963	.9964	.9568	.9568	.9140	.9140	-.0872	.9129
harmonic yes	1	1	.9909	.9911	.9963	.9964	.9570	.9570	.9132	.9132	-.0871	.9122
arithmetic no	.9910	.9909	1	1	.9983	.9982	.9247	.9247	.9369	.9369	-.1038	.9358
arithmetic yes	.9912	.9911	1	1	.9984	.9983	.9250	.9250	.9352	.9352	-.1034	.9341
geometric no	.9963	.9963	.9983	.9984	1	1	.9385	.9385	.9271	.9271	-.0961	.9260
geometric yes	.9964	.9964	.9982	.9983	1	1	.9387	.9387	.9259	.9259	-.0959	.9248
lexicographic no	.9568	.9570	.9247	.9250	.9385	.9387	1	1	.8459	.8459	-.0915	.8449
lexicographic yes	.9568	.9570	.9247	.9250	.9385	.9387	1	1	.8459	.8459	-.0916	.8449
graphicolexic no	.9140	.9132	.9369	.9352	.9271	.9259	.8459	.8459	1	1	-.1112	.9999
graphicolexic yes	.9140	.9132	.9369	.9352	.9271	.9259	.8459	.8459	1	1	-.1112	.9999
percent no	-.0872	-.0871	-.1038	-.1034	-.0961	-.0959	-.0915	-.0916	-.1112	-.1112	1	-.1046
percent yes	.9129	.9122	.9358	.9341	.9260	.9248	.8449	.8449	.9999	.9999	-.1046	1

Table 8: Rank correlations across aggregate criteria for authors, top 1000 authors

exclude outliers?	harmonic		arithmetic		geometric		lexicographic		graphicolexic		percent	
	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
harmonic no	1	1	.7667	.7695	.8703	.8720	.8213	.8212	.5748	.5748	.7231	.5755
harmonic yes	1	1	.7659	.7686	.8697	.8713	.8221	.8221	.5736	.5736	.7231	.5744
arithmetic no	.7667	.7659	1	.9999	.9772	.9764	.4386	.4385	.8775	.8775	.5006	.8770
arithmetic yes	.7695	.7686	.9999	1	.9783	.9775	.4420	.4419	.8716	.8716	.5055	.8711
geometric no	.8703	.8697	.9772	.9783	1	1	.5568	.5567	.8182	.8182	.5905	.8176
geometric yes	.8720	.8713	.9764	.9775	1	1	.5591	.5590	.8154	.8154	.5927	.8148
lexicographic no	.8213	.8221	.4386	.4420	.5568	.5591	1	1	.2805	.2805	.5724	.2814
lexicographic yes	.8212	.8221	.4385	.4419	.5567	.5590	1	1	.2805	.2805	.5724	.2813
graphicolexic no	.5748	.5736	.8775	.8716	.8182	.8154	.2805	.2805	1	1	.2421	.9999
graphicolexic yes	.5748	.5736	.8775	.8716	.8182	.8154	.2805	.2805	1	1	.2421	.9999
percent no	.7231	.7231	.5006	.5055	.5905	.5927	.5724	.5724	.2421	.2421	1	.2471
percent yes	.5755	.5744	.8770	.8711	.8176	.8148	.2814	.2813	.9999	.9999	.2471	1

Table 9: Average correlations across criteria for authors

Sample	Individual criteria			Aggregate criteria		
	mean	max	min	mean	max	min
Full	.822	.997	.561	.771	1	-.112
1-250	.608	.998	-.139	.587	1	.133
1-500	.617	.998	-.086	.616	1	.088
1-750	.630	.998	-.034	.654	1	.014
1-1000	.636	.998	.010	.690	1	.242
1-2000	.645	.997	.076	.717	1	.269
1-3000	.642	.997	.076	.728	1	.236
1-4000	.646	.997	.112	.734	1	.186
1-5000	.653	.997	.151	.751	1	.218
5001-10000	.563	.996	-.144	.700	1	-.082
10001-15000	.519	.995	-.230	.640	1	-.294
15001-20000	.473	.994	-.327	.608	1	-.336
20001-25000	.431	.994	-.342	.587	1	-.311
25001-30000	.384	.994	-.393	.566	1	-.375
1001-2000	.596	.997	-.207	.694	1	.203
2001-3000	.583	.998	-.269	.717	1	.124
3001-4000	.559	.997	-.286	.683	1	-.035
4001-5000	.564	.997	-.241	.723	1	.110
5001-6000	.537	.997	-.284	.694	1	.023
6001-7000	.538	.996	-.326	.710	1	.005
7001-8000	.548	.996	-.310	.697	1	-.140
8001-9000	.511	.996	-.403	.655	1	-.212
9000-10000	.497	.996	-.319	.661	1	-.220

Table 11: Rank correlations across criteria for institutions, top 250 institutions

	Nb Works	DNb Works	Sc Works	WSc Works	ANb Works	ASc Works	AWSc Works	Nb Cites	D Cites	Sc Cites	DSc Cites	WSc Cites	WDSc Cites	ANb Cites	AD Cites	ASc Cites	ADSc Cites	AWSc Cites	AWDSc Cites	H Index	NC Authors	RC Authors	Nb Pages	Sc Pages	WSc Pages	ANb Pages	ASc Pages	AWSc Pages	Abs Views	Down loads	AAbs Views	ADown loads
NbWorks	1	.984	.760	.690	.975	.742	.672	.630	.681	.528	.580	.506	.554	.624	.679	.523	.579	.501	.553	.676	.698	.684	.903	.689	.617	.877	.665	.598	.890	.866	.845	.824
DNbWorks	.984	1	.704	.628	.989	.695	.619	.580	.628	.473	.518	.451	.492	.585	.637	.476	.527	.453	.499	.674	.647	.632	.913	.650	.574	.895	.635	.563	.868	.841	.849	.822
ScWorks	.760	.704	1	.991	.722	.990	.981	.901	.909	.883	.899	.873	.890	.894	.909	.880	.906	.869	.896	.537	.934	.936	.776	.946	.934	.757	.925	.915	.752	.731	.695	.687
WScWorks	.690	.628	.991	1	.650	.980	.990	.902	.902	.907	.916	.903	.913	.891	.899	.900	.920	.895	.916	.486	.928	.933	.715	.939	.944	.697	.916	.923	.701	.681	.642	.636
ANbWorks	.975	.989	.722	.650	1	.725	.651	.592	.632	.491	.531	.470	.507	.606	.653	.502	.548	.480	.523	.663	.654	.641	.908	.669	.597	.906	.664	.594	.861	.830	.861	.829
AScWorks	.742	.695	.990	.980	.725	1	.990	.882	.881	.866	.872	.857	.864	.894	.902	.881	.899	.871	.889	.526	.913	.916	.771	.944	.931	.765	.939	.929	.731	.707	.695	.684
AWScWorks	.672	.619	.981	.990	.651	.990	1	.883	.875	.890	.888	.886	.886	.892	.892	.902	.912	.897	.909	.474	.908	.914	.709	.935	.940	.703	.930	.937	.680	.657	.641	.631
NbCites	.630	.580	.901	.902	.592	.882	.883	1	.987	.973	.972	.960	.960	.980	.976	.956	.965	.943	.953	.450	.972	.971	.672	.890	.894	.651	.863	.867	.714	.703	.657	.663
DCites	.681	.628	.909	.902	.632	.881	.875	.987	1	.949	.973	.934	.957	.959	.979	.924	.956	.909	.940	.496	.978	.974	.708	.888	.882	.681	.854	.849	.755	.750	.686	.696
ScCites	.528	.473	.883	.907	.491	.866	.890	.973	.949	1	.986	.998	.987	.950	.935	.981	.978	.980	.979	.372	.945	.951	.581	.876	.905	.564	.850	.878	.625	.614	.570	.576
DScCites	.580	.518	.899	.916	.531	.872	.888	.972	.973	.986	1	.981	.997	.938	.946	.957	.979	.952	.977	.420	.96	.964	.620	.883	.903	.597	.849	.868	.667	.661	.598	.607
WScCites	.506	.451	.873	.903	.470	.857	.886	.960	.934	.998	.981	1	.986	.937	.919	.980	.974	.982	.979	.355	.932	.939	.561	.869	.903	.546	.843	.876	.603	.593	.55	.556
WDScCites	.554	.492	.890	.913	.507	.864	.886	.960	.957	.987	.997	.986	1	.927	.931	.958	.977	.957	.980	.401	.948	.953	.598	.877	.903	.577	.844	.870	.642	.637	.575	.584
ANbCites	.624	.585	.894	.891	.606	.894	.892	.980	.959	.950	.938	.937	.927	1	.987	.970	.970	.956	.957	.453	.955	.954	.682	.896	.895	.675	.888	.888	.700	.687	.668	.673
ADCites	.679	.637	.909	.899	.653	.902	.892	.976	.979	.935	.946	.919	.931	.987	1	.946	.969	.930	.953	.502	.967	.964	.724	.901	.893	.712	.888	.880	.746	.739	.706	.715
AScCites	.523	.476	.880	.900	.502	.881	.902	.956	.924	.981	.957	.980	.958	.970	.946	1	.987	.998	.988	.374	.930	.937	.589	.883	.908	.584	.875	.900	.611	.599	.579	.584
ADScCites	.579	.527	.906	.920	.548	.899	.912	.965	.956	.978	.979	.974	.977	.970	.969	.987	1	.982	.997	.426	.955	.959	.635	.902	.918	.625	.887	.904	.659	.651	.616	.624
AWScCites	.501	.453	.869	.895	.480	.871	.897	.943	.909	.980	.952	.982	.957	.956	.930	.998	.982	1	.987	.356	.917	.925	.569	.875	.905	.564	.866	.897	.589	.578	.557	.563
AWDScCites	.553	.499	.896	.916	.523	.889	.909	.953	.940	.979	.977	.979	.980	.957	.953	.988	.997	.987	1	.405	.942	.948	.612	.894	.917	.602	.880	.903	.634	.627	.591	.599
HIndex	.676	.674	.537	.486	.663	.526	.474	.450	.496	.372	.420	.355	.401	.453	.502	.374	.426	.356	.405	1	.517	.507	.647	.510	.455	.632	.499	.446	.614	.603	.585	.577
NCAuthors	.698	.647	.934	.928	.654	.913	.908	.972	.978	.945	.960	.932	.948	.955	.967	.930	.955	.917	.942	.517	1	.999	.729	.915	.910	.704	.885	.882	.750	.742	.688	.694
RCAuthors	.684	.632	.936	.933	.641	.916	.914	.971	.974	.951	.964	.939	.953	.954	.964	.937	.959	.925	.948	.507	.999	1	.718	.919	.918	.694	.890	.890	.736	.726	.675	.681
NbPages	.903	.913	.776	.715	.908	.771	.709	.672	.708	.581	.620	.561	.598	.682	.724	.589	.635	.569	.612	.647	.729	.718	1	.791	.719	.989	.781	.713	.804	.782	.779	.762
ScPages	.689	.650	.946	.939	.669	.944	.935	.890	.888	.876	.883	.869	.877	.896	.901	.883	.902	.875	.894	.510	.915	.919	.791	1	.989	.780	.989	.980	.687	.667	.647	.641
WScPages	.617	.574	.934	.944	.597	.931	.940	.894	.882	.905	.903	.903	.903	.895	.893	.908	.918	.905	.917	.455	.910	.918	.719	.989	1	.711	.978	.989	.638	.619	.599	.594
ANbPages	.877	.895	.757	.697	.906	.765	.703	.651	.681	.564	.597	.546	.577	.675	.712	.584	.625	.564	.602	.632	.704	.694	.989	.780	.711	1	.786	.718	.777	.752	.774	.753
AScPages	.665	.635	.925	.916	.664	.939	.930	.863	.854	.850	.849	.843	.844	.888	.888	.875	.887	.866	.880	.499	.885	.890	.781	.989	.978	.786	1	.989	.659	.637	.638	.630
AWScPages	.598	.563	.915	.923	.594	.929	.937	.867	.849	.878	.868	.876	.870	.888	.88	.900	.904	.897	.903	.446	.882	.890	.713	.980	.989	.718	.989	1	.612	.591	.591	.584
AbsViews	.890	.868	.752	.701	.861	.731	.680	.714	.755	.625	.667	.603	.642	.700	.746	.611	.659	.589	.634	.614	.750	.736	.804	.687	.638	.777	.659	.612	1	.975	.967	.948
Downloads	.866	.841	.731	.681	.830	.707	.657	.703	.750	.614	.661	.593	.637	.687	.739	.599	.651	.578	.627	.603	.742	.726	.782	.667	.619	.752	.637	.591	.975	1	.934	.967
AAbsViews	.845	.849	.695	.642	.861	.695	.641	.657	.686	.570	.598	.550	.575	.668	.706	.579	.616	.557	.591	.585	.688	.675	.779	.647	.599	.774	.638	.591	.967	.934	1	.968
ADownloads	.824	.822	.687	.636	.829	.684	.631	.663	.696	.576	.607	.556	.584	.673	.715	.584	.624	.563	.599	.577	.694	.681	.762	.641	.594	.753	.630	.584	.948	.967	.968	1

Table 12: Rank correlations across aggregate criteria for institutions, full sample

exclude outliers?		harmonic		arithmetic		geometric		lexicographic		graphicolexic		percent	
		no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
harmonic	no	1	.9992	.6096	.6105	.7063	.7038	.8849	.8849	.5241	.5241	.5254	.5423
harmonic	yes	.9992	1	.6304	.6315	.7272	.7250	.8966	.8966	.5437	.5437	.5462	.5637
arithmetic	no	.6096	.6304	1	.9998	.9602	.9589	.7533	.7533	.9165	.9165	.9481	.9668
arithmetic	yes	.6105	.6315	.9998	1	.9630	.9619	.7561	.7561	.9138	.9138	.9490	.9677
geometric	no	.7063	.7272	.9602	.9630	1	.9998	.8559	.8559	.8599	.8599	.9048	.9241
geometric	yes	.7038	.7250	.9589	.9619	.9998	1	.8557	.8557	.8579	.8579	.9046	.9239
lexicographic	no	.8849	.8966	.7533	.7561	.8559	.8557	1	1	.6551	.6551	.6904	.7093
lexicographic	yes	.8849	.8966	.7533	.7561	.8559	.8557	1	1	.6551	.6551	.6904	.7093
graphicolexic	no	.5241	.5437	.9165	.9138	.8599	.8579	.6551	.6551	1	1	.8593	.8774
graphicolexic	yes	.5241	.5437	.9165	.9138	.8599	.8579	.6551	.6551	1	1	.8593	.8774
percent	no	.5254	.5462	.9481	.9490	.9048	.9046	.6904	.6904	.8593	.8593	1	.9959
percent	yes	.5423	.5637	.9668	.9677	.9241	.9239	.7093	.7093	.8774	.8774	.9959	1

Table 13: Rank correlations across aggregate criteria for institutions, top 250 institutions

exclude outliers?		harmonic		arithmetic		geometric		lexicographic		graphicolexic		percent	
		no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
harmonic	no	1	.9999	.5886	.5828	.8281	.8317	.9531	.9531	.6633	.6633	.3316	.3400
harmonic	yes	.9999	1	.5864	.5805	.8279	.8317	.9546	.9546	.6612	.6612	.3272	.3355
arithmetic	no	.5886	.5864	1	.9982	.8471	.8379	.5732	.5732	.9707	.9707	.7410	.7513
arithmetic	yes	.5828	.5805	.9982	1	.8526	.8442	.5672	.5672	.9569	.9569	.7502	.7606
geometric	no	.8281	.8279	.8471	.8526	1	.9995	.8196	.8196	.8264	.8264	.5153	.5273
geometric	yes	.8317	.8317	.8379	.8442	.9995	1	.8252	.8252	.8160	.8160	.5125	.5244
lexicographic	no	.9531	.9546	.5732	.5672	.8196	.8252	1	1	.6439	.6439	.3619	.3685
lexicographic	yes	.9531	.9546	.5732	.5672	.8196	.8252	1	1	.6439	.6439	.3619	.3685
graphicolexic	no	.6633	.6612	.9707	.9569	.8264	.8160	.6439	.6439	1	1	.7025	.7126
graphicolexic	yes	.6633	.6612	.9707	.9569	.8264	.8160	.6439	.6439	1	1	.7025	.7126
percent	no	.3316	.3272	.7410	.7502	.5153	.5125	.3619	.3619	.7025	.7025	1	.9990
percent	yes	.3400	.3355	.7513	.7606	.5273	.5244	.3685	.3685	.7126	.7126	.9990	1

Table 14: Average correlations across criteria for institutions

Sample	Individual criteria			Aggregate criteria		
	mean	max	min	mean	max	min
Full	.891	1	.592	.803	.968	.524
1-250	.781	.999	.355	.720	.971	.327
1-500	.821	.999	.479	.717	.977	.347
1-750	.842	1	.526	.713	.980	.336
1-1000	.865	1	.569	.710	.981	.316
1001-2000	.820	1	.253	.835	.988	.505
2001-3000	.835	1	.100	.909	.997	.760