Jointly published by Elsevier Science Ltd, Oxford and Akadémiai Kiadó, Budapest

Scientometrics, Vol. 37, No. 2 (1996) 223-244

REASSESSMENT OF CO-CITATION METHODS FOR SCIENCE INDICATORS: EFFECT OF METHODS IMPROVING RECALL RATES

M. ZITT[1,2] E. BASSECOULARD[1]

[1]*LERECO, INRA, BP 1627, F-44316 Nantes Cedex 03 (France)*
[2]*Observatoire des Sciences et des Techniques (OST), 93 rue de Vaugirard, F-75006 Paris, (France)*

(Received June 21, 1996)

Although co-citation techniques are very powerful structuring tools, the use of science policy indicators based on co-citation has often been criticized, especially on ISI research fronts. A major issue is the small fraction of literature retrieved, i.e. the "recall rate" problem. Our investigations indicate that at the level of micro/meso studies high recall rates can be achieved by (a) the use of appropriate clustering techniques limiting singletons and (b) the enrichment of co-cited cores by medium-cited items. This combination of appropriate clustering and extension of recall proves to be efficient, provided that careful trade-offs are sought between the extension and relevance of recall. It leads to a reassessment of the performance of the co-citation approach for structuring scientific fields and providing related indicators not limited to the 'leading edge'. It also opens new opportunities for comparison/combination with other relational methods such as co-word analysis.

Introduction

The positioning of actors in a scientific field by bibliometric methods implies a relevant corpus, a proper identification of the actors, efficient structuring methods and indicators with controlled properties. Many structuring methods rely on co-item approaches, particularly co-citation and co-word. This paper deals with co-citation methods independently pioneered by *Small*[1] and *Marshakova*[2] in the field of science (*White & McCain*[3] have noted a precursor in the humanities field). Variants of co-citation at the author level were introduced at Drexel University by *White & Griffith*[4] and further developed by *McCain*[5] and *Penan*[6] (*McCain*[7] also experimented on journal co-citation analysis). We will focus here on the document co-citation approach associated with *Small & Griffith*[8] and their co-workers at the ISI.

The essential notion is that the reference list provided by a citing author in an article creates a basic linkage between the references. The aggregation of this linkage on a set of citing authors, properly normalized, measures the association between the two cited articles, which then constitutes a basis for network design and clustering.

0138-9130/96/US $ 15.00

Conversely, clustered cores of cited papers define groups of citing papers (a 'research front') contributing to the associations in each particular core. The interpretation of these association networks is a matter of debate. Document co-citation has been applied at the ISI and by other research teams to detect homogeneous areas in research networks, thereby providing a direct illustration of Price's research fronts, as in the ISI Atlas of Science.

Co-citation analysis is considered to be an efficient means of describing mainstreams, or at least their leading edge, in academic science, but its use in science policy remains controversial because of technical and sociological objections which are well documented in the literature (*King*[9], *Hicks*[10]). A typical problem is the low "recall rate": only a fraction of the literature on a subject can be classified through co-citation. This is hardly a drawback if co-citation is viewed only as a tool for detecting the leading edge of academic activity but can be a real defect if science policy users intend to perform relevant divisions of the entire scientific domain, e.g. for better trend analysis or evaluation of actor participation. For this purpose, other structuring methods such as co-word analysis may be more efficient (*Braam* et al.[11]), despite specific problems of language treatment. Co-citation methods would be more suitable for these purposes if their recall rates were improved in a reliable manner.

We argue that such improvement is possible in the case of micro- or meso-analyses. Appropriate variants abiding by the co-citation principle can be introduced into the methods to retrieve a larger part of the literature, thus providing rather robust research front classification and associated indicators.

This paper is divided into four sections. The first section briefly reviews the limitations of co-citation techniques and introduces the recall rate problem. The second section shows the effects of clustering methods on cluster distributions and consequently on recall rates. The third section describes a second means of increasing recall rates by an extension of cores by medium-cited papers, resulting in the assignment to research fronts of previously dropped citing documents. In the fourth section, a few indicators obtained before and after extension are compared. A basic description of data and procedures is given in the Annex.

Context

Some Limitations of Co-Citation Studies

ISI databases. Contrary to co-word analyses, large-scale co-citation studies have to rely in practice on ISI sources, particularly the *Science Citation Index (SCI)*. Therefore, the well-known technical limitations and biases of this source are reflected in co-citation analyses.

A first technical problem results from ISI codification of cited references: name of author(s), page and volume number, cited journal (*Leydesdorff*[12], *Moed & Vriens*[13]). These errors can be limited by using a proper identification-key for the cited documents and, of course, by checking and cleaning the data. Another basic and frequently considered problem of ISI citation indexes is the possible bias in journal coverage. For science studies, the journal set of the *SCI*, usually considered as adequate for "big science" fields, may be less suitable for investigations of applied or technological research.

Homogeneity, immediacy, stability. Co-citation clustering suffers from the general problems of clustering methods, such as the significance of output (classifiability) or the distribution of cluster size (see below), as well as specific problems:

a) the selection of structuring documents and the consequences involved. Except for very small files, co-citation analysis implies a drastic preliminary selection of "structuring papers," i.e. cited papers to be submitted to classification. By and large, the criterion for selection is citation score. The threshold, dictated by the computing constraints involved in further clustering, is higher in practice than that desirable in a Bradfordian rationale. Another point is that the well-known variability of citation practices among fields, or inside a field between theoretical and experimental papers, may cause some distortions if uniform citation and co-citation thresholds are used (*Sullivan* et al.[14]; *Hicks*[15]). Although this can be a drawback, especially for large-scale studies, experimental records (references to widespread methods, data sources, etc.) are also likely to create spurious links. Several improvements have been introduced by specialists (first of all, those at the ISI) to overcome some of these limitations: fractional citation counting, normalized co-citation (the Jaccard index and more recently Salton's cosine formula), which has been widely adopted, variable co-citation level clustering with limitation of cluster size, and iterative clustering of clusters (*Small* et al.[16])...

b) a new research topic emerges as a co-citation cluster only when it has begun to attract a fairly large audience (*Sullivan* et al., op.cit.). This can be a drawback for addressing new works. For co-citation studies, *Healey & Rothman*[17] suggested that "some appropriate weighting to citations less than three years old" could partly account for the time lag. Other citationist methods can be mobilized to provide snapshots of early stages in the evolution of a specialty: *Glänzel & Czerwon*[18] stressed the advantages of bibliographic coupling* in this respect.

* In the same line, *Stern*[33] proposed a comparison between author-cocitation and bibliographic coupling mapping (in social sciences). Co-citation maps represent a situation in which matters are more settled.

c) fluctuations are also important for interpretation. Shifts in dominant foci of interest are traced by variations from year to year in co-citation clusters. However, changes are sometimes "disconcertingly large" (*Rip*[19]), which may be due in part to fluctuations around thresholds. The representativeness of co-citation fronts (citing literature assigned to co-citation clusters) is also affected by year-to-year variations (*Sullivan* et al., op.cit.). Time averaging, e.g. when operating on a multiple-year citing file (*Zitt & Bassecoulard*[21]), can deal partly with these fluctuations.

Recall Rates

Tentative definitions. Last but not least, low recall rates, both on cited and citing sides, create difficulties in interpretation. In most published results, fewer than 50% of source papers cite at least one paper in co-citation clusters. The notion of the recall rate is a matter of concern at the successive stages of the process (see also *Braam* et al., op.cit.).

At the first stage, the *clustering technique* is the most usual means of building co-citation fronts. "Singleton" clusters (isolates) are discarded from co-citation cores in order to comply with the co-citation rationale of association. The first "recall rate" is *the proportion of cited papers clustered in the co-cited cores to all cited papers initially selected as structuring ones*. As we shall see below, the values of "CiteD Recall rate" (CDR) depend heavily on clustering methods. This definition addresses the central principle of the co-citation process: both terms of the ratio are unambiguous. Yet this ratio does not account for applications focused on the citing side, the "living" science.

At the second stage, *citing papers are assigned to cores* according to various decision rules. Generally speaking, a "citing recall rate" compares the assigned citing literature (numerator) to "all relevant citing literature" on the same subject (denominator), although this can be accomplished in many ways. First, a citing recall rate is only defined within a given process of assignment of citing papers, which governs the numerator. Then, in practice, there is no standard rule to define the relevant literature (the denominator). For instance, with science policy applications in mind, the "relevant citing literature" may be an expert-selected ad hoc sample, typically including non-ISI sources, while for practical reasons the numerator is *SCI*-based. Yet it must be stressed that such a citing recall rate mixes two phenomena: the intrinsic recall rate on one side and the representativeness of the literature used in front-building (in the *SCI*) on the other side. Losses due to possibly poor coverage should not be attributed to the co-citation method itself. In some extreme cases, neither the numerator nor the denominator is clear-cut, e.g. in the determination of a recall rate at the field level for ISI-built fronts*.

Inside a closed dataset the definition of the citing recall rate becomes unambiguous: a good example is a micro-study *starting from a definite citing literature* (e.g. journal-based), upon which a co-citation study is conducted after the usual cleaning operations** (our examples below). Another example of a closed set is the ISI research fronts *for all science*. In such cases, an unambiguous "Internal citinG Recall rate" (IGR) can be defined.

Recall rates in practice. As mentioned above, co-citation studies generally rely on the skew distribution of citations to select top "structuring" documents. As a result, medium-cited articles are discarded. Hence, a large number of references and consequently a high proportion of citing papers are left out.

The distributional characteristics of a field relevant for co-citation analysis may be summarized in terms of a bivariate structure (citations, references) outlined in *Zitt & Bassecoulard*[21]. The effect of these characteristics on IGR is clear when simple selection and assignment rules are used. For example, when (a) an integer citation threshold, say Y, is used to select cited paper candidates for building cores, and (b) the assignment of a citing paper to a cluster requires at least X references to cited papers in this cluster. In this case, the upper bound of IGR*** is directly determined by straightforward distributional constraints, as is the effect of substitutions between X and Y levels to reach a given recall level. Only the upper bound, and not the final value of IGR, is determined since clustering and assignment operations would cause further losses****.

As methodological choices are not always clearly mentioned in co-citation studies reported in the literature, it may be difficult to compare recall rates. If we examine

* Assigning research fronts to subfields is far from easy (see *Coward* et al.[34]). Any paper in any discipline can contribute to the creation of a front in a given discipline in "ready-made" ISI research fronts. Comparisons between this global building and local views are stimulating (*Sullivan* et al., op.cit.).

** The precise definition of citing literature may depend on these cleaning operations, e.g. discarding citing papers with very few references.

*** A paper is assigned to a cluster at threshold X when it has X references or more to papers cited Y times or more belonging to this cluster. In this case, "MaxIGR" is the frequency Z of citing documents with at least X references to papers cited Y times or more. The underlying density function is the frequency of citing documents with their Xth reference cited Y times. Other assignment methods exist.

**** The first loss, concerning the citing of papers referring only to singletons, depends directly on CDR. The second loss, due to the scattering of references of a document to cited papers in several clusters, depends on the level of cutting (the loss tends to increase with the fragmentation of the universe into a large number of clusters). This point is beyond the scope of our discussion.

some examples for which recall rates are available, we can see that internal recall rates are dispersed, depending on the datasets and methodologies used (citation countings, co-citation measurements, clustering methods), and generally poor. For instance, with single-linkage clustering on SCISEARCH (*Small*16), cited recall rates vary from 33% (fractional citation counting, cosine fixed co-citation threshold) to 66% (integer citation counting, cosine variable co-citation threshold). Moreover, using single-linkage clustering and a loose assignment rule (X=1, see above), *Milman & Gavriliva*[22] reported a 9% citing recall rate in their study on chemical engineering (integer citation counting, integer fixed co-citation threshold), whereas *Braam* et al., op.cit., achieved 21% in their chemoreception study (integer citation counting, cosine fixed co-citation threshold). As indicated below, much higher recall rates were achieved in the first tests on our file for Astronomy and Astrophysics 1989-1992 (File A2, see Annex).

Interpretation of co-citation clusters

Basically, co-citation fronts share a common focus on earlier highly-cited literature. As underlying citation behaviors are complex (see a review of problems in citation analysis by *McRoberts & McRoberts*[23]), the interpretation of co-citation cores (cited papers) is a subject of debate. Should they be considered as "shared legitimatory repertoires" (*Rip*, op.cit.) rather than central concepts or methodologies in a field? On the citing side, the original interpretation is that co-citation fronts describe the "leading edge" of scientific activity (e.g. for *Franklin*[24] such techniques are designed to be selective). Enhanced recall rates might challenge this view, suggesting that co-citation can provide partitions of a large part of *SCI* literature in a given field.

Nonetheless, results depend heavily on a set of explicit technical choices (counting methods, proximity measures, thresholds, clustering, etc.) that must be kept in mind when interpreting clusters and maps. As noted by *Hicks* (op.cit.), the "black-box effect" can be highly dangerous.

Clustering methods and cited recall rates (CDR)

The Choice of Clustering Methods

Many methods are available at the clustering stage (see for instance *Hartigan*[25]), which have been widely discussed and compared in the specialized literature, though with many difficulties, as stressed for example by the *SAS Institute*.[26] More generally, the classifiability problem is currently drawing attention to the distributional properties associated with classifications. The problem of choosing clustering criteria is often encountered in bibliometrics. For instance, *Leydesdorff*[27] recommended the Ward criterion against single linkage in a context of journal classification on citation transactions. *Todorov & Vlachy*[28] used average linkage in another context. Complete linkage may also be efficient in information-retrieval applications.

We will focus here on two particular algorithms, single and average linkage. Single linkage has been associated with co-citation techniques since the beginning, and average linkage provides an efficient trade-off for similarity analyses.

A major reason for single-linkage use is its ability to process very large universes. The drawbacks are well-known: single-linkage yields a very skew size-distribution of clusters, with generally one big cluster built by chain effects on the one hand and a high number of isolates on the other*. In practice, a strong correction of chain effects is needed, using either heavy methods (density estimates) or simpler means, e.g. various ways of limiting cluster size (ISI method). Whether the "bias-free" advantage of single-linkage holds up in such cases is questionable.

Although not unbiased, group-average linkage appears in many cases to be a low-risk strategy, with minimization of the local discrepancy between initial similarities and final ultrametrics**, good general performances against error perturbations (Milligan29) and reasonable computing constraints.

Effects on Cluster Distribution in a Document Co-citation Context

In our experiments on co-citation universes, the two methods give very different distribution of cluster sizes. The following figures illustrate these differences. The source of data was our file A1 1986-1989 (see Annex). Figure 1 addresses the critical issue of singletons, displaying the proportion of items classified in singletons at all levels of the classification tree for average linkage versus single linkage (both non-corrected). After the initial stage (leaves of the tree, upper right corner), average linkage always gave a lower proportion of isolates than single linkage. At a given cutting level, items without a strong bond (possibly unique) were left out by single linkage but aggregated by average linkage, with a collection of weaker connections.

---

* This skewness may be related to the fact that, for random proximities, the probability of two clusters merging at one step depends on the number of available targets for association (i.e. cluster sizes) at the previous step.

** This is stated by *De Virville* in *Benzecri*[35]. In an empirical comparison of criteria on controlled datasets, *Hennequet*[36] from our team obtained a good general rating for average linkage. In some cases, this method does not recover elongated clusters and, according to *Edelbrock*, tends to produce small "non-conformist" clusters of outliers (*Willet*[37])..

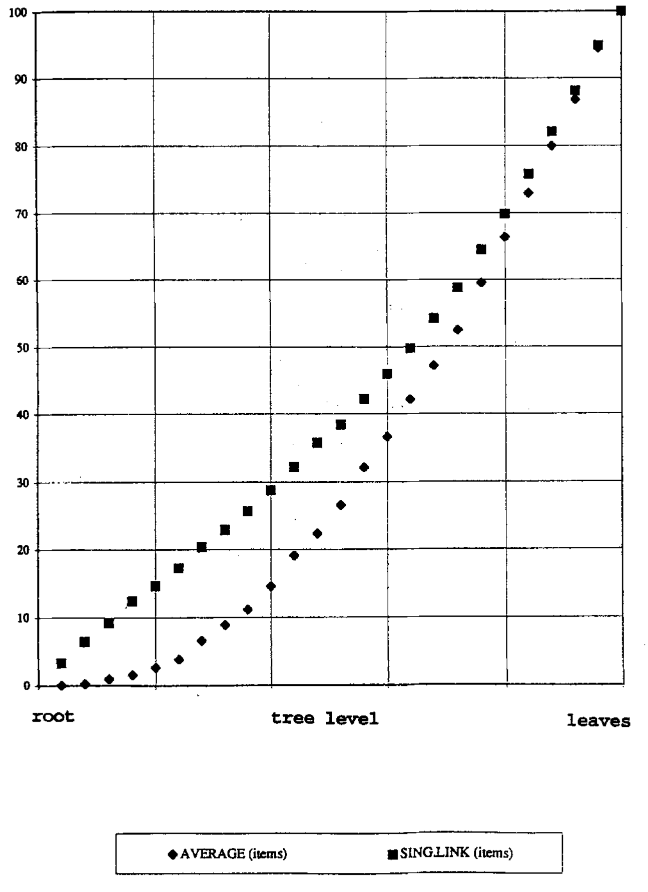

Fig. 1. Percentage of documents in singletons

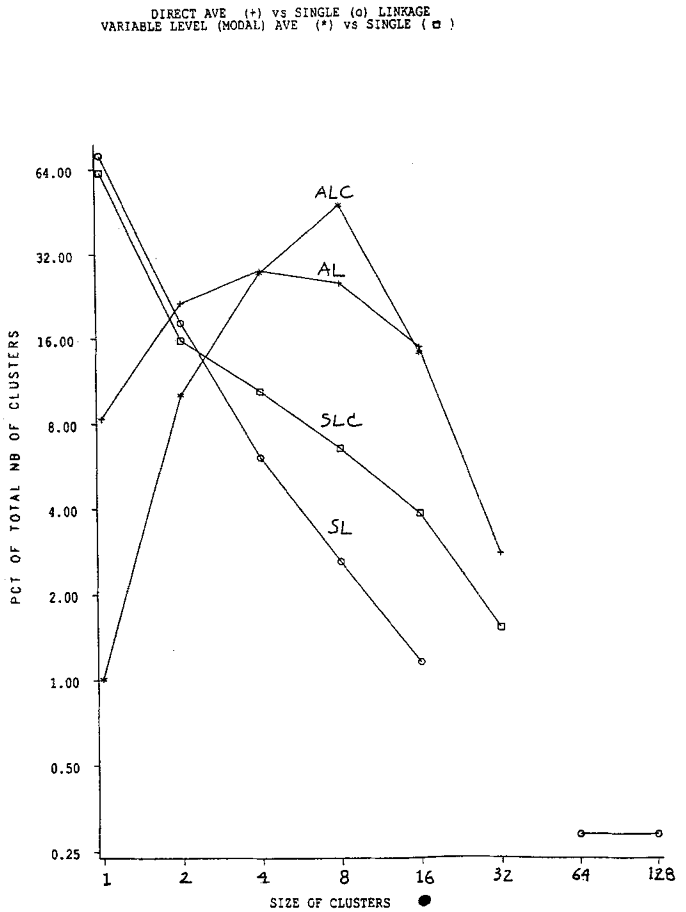

Fig. 2. Distribution of clusters (file A1, log-log plot)

Let us examine the implications in a real situation. Both methods were applied on the same corpus, first without correction and then corrected by a partial application of a modal cutting technique. We chose the cutting levels so as to obtain a reasonable number of exploitable clusters, e.g. about 100 non-singleton clusters* for about 1,000 highly-cited papers and 9,000 citing papers. Figure 2 shows the size distribution in the four experiments on a log-log plot. The difference between the hyperbolic aspect of

* Average linkage non-corrected (ALNC), 108 clusters, 99>1; single linkage non-corrected (SLNC) 347 clusters, 99>1; average linkage corrected 99 clusters, 98>1; single linkage corrected 261 clusters, 99>1

single-linkage distributions (circles and squares) and the two-tailed distributions of average-linkage distribution (log-normal-shaped but with lower kurtosis) clearly appears. The correction process adopted here reduced the proportion of singletons/doubletons sharply for average linkage (stars) and slightly for single linkage (squares), as well as the size of the largest clusters (both linkages). With or without correction, CDR* was greatly enhanced by average linkage.

Though our results still need to be checked on other data (very similar results have already been obtained in co-word analyses), these empirical findings confirm that the choice of the similarity index and the aggregation criterion greatly affects the distribution of the resulting research fronts and thus possible interpretations concerning the underlying structure of science networks. However, it is clear that more work is needed to disentangle method artefacts and the underlying structure.

From a practical point of view, efficient clustering of cited papers can bring CDR almost to its maximum (99% in our experiments on file A2). Such coverage, already valid for discussions with experts, is also reflected on the citing side: 80% of the citing papers are recalled with the most usual assignment rule (X=1; a paper is assigned to a cluster when it cites at least one of the cluster's core-papers). To be strictly co-citationist, a more demanding rule is required, namely that a paper cite (associate) at least two references of the core (X=2). Although this is practically impossible with the usual CDR reported in co-citation studies, a high CDR makes it feasible. In our case, 53% of citing papers were retrieved with the latter rule, which can thus be expected to give better assignments.

These results depend in part on the nice distributional properties of the datasets under consideration, which would appear to be a "best case" situation. An ongoing study of a quite scattered field (ecology) gives some insight into what can be considered as a nearly "worst case" situation. Initial tests showed an equally high CDR, but an IGR of about 50% with X=1 and 25% with X=2. The additional process proposed below offers an opportunity to obtain all the benefits from a high CDR, either by maximizing citing recall rates or by using the efficient X=2 rule while keeping the citing recall rate at a fairly good level.

* CDR is one of the factors commanding the differences between internal citing recall rates (maxIGR – IGR). The higher the CDR, the lower the difference between the upper bound and the actual IGR, all things being equal. The other factor governing this difference is the distribution of references among clusters.

Extended recall: principle, effects and validation

Given the possible computer limitations on the total number of items submitted to classification, it would seem highly desirable in most cases to increase the recall rate by complementary operations, within the existing structure of co-cocited fronts. Several approaches are possible:

- Leiden's group proposes a mix of methods, using word similarities to enrich co-citation research fronts. The simultaneous use of such complementary techniques is very effective (*Braam* et al, op.cit.) but of course mixes different logics.
- Within the citation rationale, bibliographic coupling, pioneered by *Kessler*[30] and now being considered again (*Glänzel* et al., op.cit.), could be used as well as a complementary process (based on shared references between citing documents) to associate documents with research fronts.

We propose a different system still based on the co-citation principle.

Principle: Two-Stage Process Using "Cluster-Document" Co-citation

The starting stage is a fairly classical co-citation procedure. The selection of structuring documents is based on citation Y rates (say an integer threshold Y=y1). A few variants (variable level using partly "modal" cutting, average linkage, time-averaging, etc. See op.cit.[21]) are introduced for robustness.

As documents that do not cite selected highly-cited papers (y1 threshold) are ignored in conventional assignment procedures, the second stage consisted in a search for the closest neighboring core to any medium-cited paper (between y2 and y1 times, y2 <y1). The algorithm remains "co-citationist" in a broad sense, without a lexical or bibliographic procedure, and involves the calculation of a "Cluster-document" proximity (see the Annex). This process is close to reindexing used for other purposes, e.g. for the creation of macro-aggregates (*Turner* et al.[31] in a co-word context; *Small*[16,32] in co-citation). The extension procedure does not alter the basic front structure. The number and definition of the fronts remain unchanged, but their content is extended*. Otherwise, we observed that the size ratio after/before extension was fairly regular except for doubletons/singletons exhibiting higher but more irregular values. Figure 3 describes the extension process for a core.

---

* The status of singletons is modified since they generally gather other cited papers; however the robustness of the complementary assignment is in this case very questionable, as it is for doubletons, so that it is wise to keep them apart.

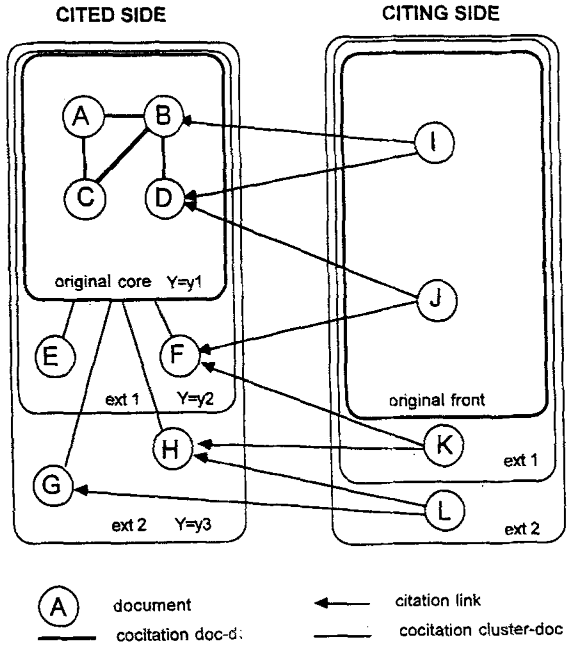

Fig. 3. Extending a research front

A, B, C, D are structuring documents (cited > y > 1 times) in the initial core of a research front. E and F are attached to the core by a moderate extension (y2 threshold), G and H by a stronger extension (y3 threshold). I, J, K, L are citing documents. With the X = 1 rule, I and J are recalled in the original front., K in the moderately extended front, L in the strongly extended front. With the more demanding X = 2 rule, only I is recalled in the original front, J is recalled in the moderately extended front, K and L in the strongly extended front.

Table 1 (cited side) shows an example of core extension for a cluster of medium size (9 core documents, SuperNovae Type IA), lowering the threshold of citation frequency y1 = 21 down to a threshold y2 = 9. Core documents (primary core = 1 in the table) were ordered using tree morphology (see Annex), whereas additional documents were ordered using co-citation similarity with the core. A few documents considered as important by experts were missing in the primary core. They were retrieved by the complementary process (e.g. FILIPPENKO on row 27), as were many recent documents close to the core. Technically, the algorithm uses SQL-type languages to handle sparse bibliometric tables as lists.

As soon as the original cores of research fronts built in the first stage are completed by secondary papers, the assignment of citing papers to enriched fronts can be carried out by the usual assignment rules. The citing recall rate is strongly enhanced by this process.

The Effects of Extension: Results and Validation

From a technical point of view, the enrichment of existing cores by complementary cited items can be achieved to almost any extent, the problem being the trade-off between quality and quantity of recall. Another way of achieving high recall rates on the citing side is to choose the lowest possible X threshold, i.e. the number of references to a research front required to assign a citing paper: X=1 gives a loose assignment, but is useful in an information-retrieval context (this is the implicit rule for *SCI*-search on-line). Full compliance with the co-citation principle requires X=2. We stressed the substituability of the rules governing X and Y in an earlier paper, and compared the effects on the citing side* of the extension of cores at cited level Y, on the one hand, and of the assignment rules X of citing items on the other [op.cit.1995].

Overall recall effect. Table 2 summarizes the results for the citing side expressed by the number of distinct papers recalled. Favorable characteristics of the field and clustering method can provide a very good citing recall rate, even when the X=2 rule is used**. The third and fifth columns from the left refer to the upper bounds of citing recall rates (MaxIGR). The following figures appear under the "observed recall" headings: (a) observed number of citing papers assigned; (b) citing papers assigned/maximum recall (percentages in boldface); (c) citing papers assigned/total number of citing documents (percentages in italics).

With respect to MaxIGR, losses in actual recall rates were of course greater when X>1 because of the distribution of references among clusters. Before extension, the upper bound was almost reached with the laxist assignment rule, but 20% of the recallable papers were left out when a truly co-citation rule (X=2) was applied.

* On the cited side, items are only known according to their *SCI* description as cited references (typically first author, journal, date, volume, page), without title. Another process is needed to retrieve their complete description from databases (ISI or others), unless they already belong to the citing set.

** Using a method similar to the early version of co-citation clustering, *Sullivan* and co-workers obtained fairly high recall rates (77.4% with X=1 and 47.8% with X=2). However, as their corpus was very small and highly coherent (bibliography on weak interactions, 124 items), the results can hardly be compared.

Table 1
Example of extended recall for cluster No. 212: super-novae type IA

DOCUMENT RANK	PRIMARY CORE	PUBLICATION YEAR-AUTHOR	CITATION FREQUENCY	CITED JOURNAL (SCI TRANSCRIPTION)
1	1	84-IBEN_I	53	ASTROPHYS_J_SUPPL_S
2	1	84-WEBBINK_RF	25	ASTROPHYS_J
3	1	86-WOOSLEY_SE	72	ANNU_REV_ASTRON_ASTR
4	1	84-NOMOTO_K	54	ASTROPHYS_J
5	1	85-IBEN_I	36	ASTROPHYS_J_SUPPL_S
6	1	86-THIELEMANN_FK	25	ASTRON_ASTROPHYS
7	1	83-BRANCH_D	26	ASTROPHYS_J
8	1	70-PACZYNSKI_B	23	ACTA_ASTRON
9	1	87-VANDENBERGH_S	35	ASTROPHYS_J
10	0	85-BRANCH_D	19	ASTROPHYS_J
11	0	88-IBEN_I	10	ASTROPHYS_J
12	0	84-SUTHERLAND_PG	18	ASTROPHYS_J
13	0	91-KHOKHLOV_AM	11	A_A
14	0	86-WOOSLEY_SE	18	ASTROPHYS_J
15	0	90-WHEELER_JC	19	REPT_PROGR_PHYS
16	0	82-NOMOTO_K	14	AP_J
17	0	69-ARNETT_WD	10	ASTROPHYS_SAPCE_SCI
18	0	85-ARNETT_WD	15	NATURE
19	0	69-COLGATE_SA	12	ASTROPHYS_J
20	0	87-IBEN_I	11	ASTROPHYS_J
21	0	87-PHILLIPS_MM	13	PUBL_ASTRON_SOC_PAC
22	0	90-BENZ_W	16	ASTROPHYS_J
23	0	88-CAPPELLARO_E	14	ASTRON_ASTROPHYS
24	0	87-IBEN_I	13	ASTROPHYS_J
25	0	90-LEIBUNDGUT_B	11	ASTRON_ASTROPHYS
26	0	90-BARBON_R	9	ASTRON_ASTROPHYS
27	0	88-FILIPPENKO_AV	13	AJ
28	0	82-FULLER_G	10	AP_J_S
29	0	90-CAPACCIOLI_M	12	ASTROPHYS_J
30	0	90-MILLER_DL	16	ASTRON_J
31	0	86-KRAANKORTEWEG_RC	20	ASTRON_ASTROPHYS_SUP
32	0	89-EVANS_R	15	ASTROPHYS_J
33	0	84-HESSER_JE	9	ASTROPHYS_J
34	0	78-ABT_HA	11	AP_J_SUPPL
35	0	71-PACZYNSKI_B	17	ACTA_ASTRON
36	0	79-BARBON_R	10	ASTRON_ASTROPHYS
37	0	84-NOMOTO_K	12	ASTROPHYS_J
38	0	82-IBEN_I	19	AP_J
39	0	86-RITTER_H	9	ASTRON_ASTROPHYS
40	0	87-PORTER_AC	10	ASTRON_J
41	0	85-HUGUES_JP	9	ASTROPHYS_J
42	0	90-CANAL_R	10	ANNU_REV_ASTRON_ASTR
43	0	78-HILLS_JG	9	MON_NOT_R_ASTRON_SOC

Table 2
Complementary assignment: recall rates

COMPLEMENTARY ASSIGNMENT	THRESHOLD (CITED)	ASSIGNMENT RULE X=1 PER CLUSTER		ASSIGNMENT RULE X=2 PER CLUSTER	
	Y	max. recalls	obs. recalls	max. recalls	obs. recalls
NO	21	8,685	(F) 8,608 **99.1** *80.4*	7,087	(A) 5,702 **80.5** *53.3*
	14	9,553	9,553 **100.0** *89.3*	8,491	(B) 7,465 **87.9** *69.7*
YES	9	10,070	10,044 **99.7** *93.8*	9,471	(C) 8,760 **92.5** *81.8*
YES	5	10,450	10,392 **99.4** *97.1*	10,168	(D) 9,637 **94.8** *90.0*
YES	2	10,705	10,573 **98.8** *98.8*	10,607	(E) 10,181 **96.0** *95.1*

On the cited side, we observed that the size ratio after/before extension was fairly regular except for doubletons exhibiting higher-than-average but more irregular values.

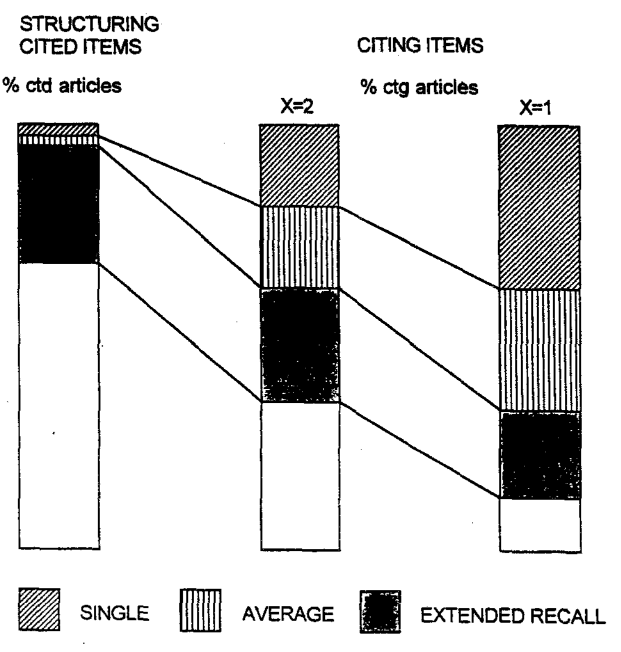

Fig. 4. Enhancement process – orders of magnitude

Figure 4 shows, in order of magnitude, the effects of the clustering technique and of the recall extension in a fictitious intermediary situation, between a nearly best case (the file A2) and a worst case in very scattered fields.

The lexical way as validation. Within a citationist rationale, it is possible to test extension by repeating a co-citation study at lower thresholds or by bibliographic coupling. To achieve independent validation, we tried another approach: we measured the changes in cluster contents before and after extension through comparison of the title words of citing papers, for clusters with a large citing population (99 out of 126) to avoid too flimsy a lexical basis. This study led to the following conclusions (ibid.,1995): (a) as expected, the drift of vocabulary gradually increases with the force of recall, but no threshold appears that could be considered suggestive of optimum recall extension*; (b) the comparison between quantitatively equivalent combinations (F: laxist X rule on original cores; C: strict co-citationist X rule on extended cores) prove favorable to the latter. Despite the obvious limitations of the title vocabulary for describing cluster contents, these results are very promising for the extension procedure.

Extended recall: comparison of indicators before and after extension

Actor Participation in Clusters

Let us turn now to practical results and implications for the production of indicators. The strategic positioning of actors is usually described by a series of measures: areas of relative strength or weakness (activity indexes), partnerships in collaborative networks, and above all the basic factor of actor shares. We shall not attempt to compare the outcomes of the clustering methods, namely single vs average, at the cluster level since this would bring up the serious difficulty of the non-superposability of the two partitions. However, within a clustering choice (here group average), it is possible to explore the effects of variants while maintaining the fundamentals of the research front (its original core) and only enriching its contents in cited and citing documents.

Does the content of a cluster change significantly in its successive states of extension? Again we shall limit our comparison to the following states: A (original), F (relaxing X rule) and C (relaxing Y rule by the core-extension process). For the sake

* The sensitivity of very small clusters (e.g. original doubletons) to this processing suggests that they could be used as a marker for optimal rates, but this sensitivity may also be due to lexical fluctuations.

of simplicity, we have merely chosen to assign a citing paper to the cluster it cites most (ties allowed). The counts for country participation are fractional.

On the whole file, country shares look very much alike in the successive states. However, a cluster-by-cluster analysis gave different conclusions. An example will be presented for two clusters: Lithium Abundance and SuperNovae type IA in Table 3.

Table 3
Country share in given clusters

Lithium Abundance	STATE A (25 Art) %	STATE F (64 Art) %	STATE C (53 Art) %	SuperNovae Type IA	STATE A (59 Art) %	STATE F (116 Art) %	STATE C (78 Art) %
USA	65.2	52.8	56.8	USA	46.9	44.1	33.8
FRANCE	13.3	8.3	8.6	GERMANY	15.1	12.2	13.2
CANADA	6.0	8.1	5.6	ITALY	5.1	9.3	12.6
GERMANY	2.0	5.5	4.2	ISRAEL	6.1	5.7	6.8
ITALY	4.0	4.7	3.7	USSR	5.9	5.6	7.0
UNITED KINGDOM	1.3	4.4	5.3	SPAIN	2.0	4.6	4.5
POLAND	2.8	3.2	3.1	JAPAN	5.2	3.3	4.7
FINLAND	0.0	2.3	2.8	CANADA	3.4	3.0	4.4
INDIA	0.0	2.3	3.5	SWITZERLAND	0.8	2.8	4.7
TURKEY	4.0	1.6	1.4	NETHERLANDS	2.5	2.2	3.2
SWITZERLAND	0.0	1.6	2.8	FRANCE	2.3	2.0	1.7
CHILE	0.0	1.6	0.0	AUSTRALIA	2.1	1.7	0.6
TAIWAN	0.0	0.8	0.0	CHILE	0.3	1.2	1.7
SPAIN	0.0	0.8	0.7	INDIA	0.0	0.9	1.2
BULGARIA	0.0	0.8	0.0	BELGIUM	0.8	0.4	0.0
USSR	0.0	0.8	0.0	DENMARK	0.8	0.4	0.0
DENMARK	1.3	0.5	1.8	FINLAND	0.0	0.4	0.0
				UNITED KINGDOM	0.4	0.2	0.0

As expected, both types of extension increase recall rates (number of citing papers) but do not show the same effects on contents. Since co-cited cores remain unchanged in the first type of extension (relaxing X rule), a cluster could only gain new citing papers and possibly new countries. Since the cores were enriched by medium-cited papers in the second type of extension, some shifts could occur in the assignments of particular papers.

For the first cluster, Lithium Abundance, the three leading countries remained the same in all cases. However, some changes occurred in a downward direction: new countries appeared (e.g. India) or some substantially increased their participation (e.g. the U.K.). For the second cluster, SuperNovae type IA, the changes were greater: Italy, Spain or Switzerland, for instance, appeared to be more active on the field than might have been expected from the first configuration, whereas the presence of India was not detected at all. Conversely, for the U.S.A., even though its leading position was not affected, the share and activity indexes decreased strongly for both types of extension. These results indicate that such comparisons can be crucial for studies of the scientific activity of smaller countries on specific topics.

Table 4
Cluster contents: country distribution

	Nb of signif..differences (chi2) at 10% level	Rank corr. median value over clusters	Nb of clusters with rank corr. <0.9/0.8/0.7
State (F) / State (A)	24 out of 93	0.83	86/47/29 out of 112
State (C) / State (A)	13 out of 80	0.83	85/58/37 out of 106
State (F) / State (C)	19 out of 93	0.85	77/49/32 out of 109

We also used two means of comparing cluster contents on the whole file: (a) A chi-square comparison at the 10% level by grouping all countries with less than 5 participations (category "others"). Clusters with less than 3 countries were discarded. (b) a Spearman rank correlation was performed for all clusters but singletons.

The comparison of cluster contents for the three configurations is summarized in Table 4. For the first type of extension (relaxing X rule), the rank correlation was less than 0.7 for about one cluster out of four, and chi-square differences were significant. For the second type of extension, the rank correlation was less than 0.7 for one-third of the clusters, with significant chi-square differences for only 16% of the clusters. These results were due in part to the enrichment of clusters by small countries, as in the example above.

Temporal Features of Research Fronts: Trend and Immediacy

We previously suggested (op.cit.[21]) that combination of cited side features ("immediacy" in the general sense of citation age) and citing side features (growth of a cluster in a multi-year analysis) could help describing the dynamism of research fronts. Improvement in the recall rate is likely to modify these measures slightly, which are now being calculated for a larger number of documents both on the cited and citing side.

As noted above, the number of cited documents can be greatly increased for each cluster. We compared the median cited dates before and after extension by relaxing rule Y in a real context (partial modal cutting). As expected, the regression line stands off the diagonal. The process clearly enriches the oldest cores (roughly before 1985) with new documents and the most recent cores with old documents. For example, our SuperNovae Type IA cluster is "younger" after recall. The median date was 1984 for the original core (9 documents cited at threshold 21) and 1986 for the enriched core (43 documents cited at threshold 9). As expected, the process reduces the overall deviation of cluster dates in the file: the differences are less marked but depend on a larger base. Some changes also took place on the citing side, though the proportion of newcomers is typically smaller than for the cited side. Without going into details, the extension may in some cases change the judgment about the significance of the average date of the citing papers in the cluster (growth indicator).

The last two points indicate that neither the distribution of countries in particular clusters nor the dynamic features were insensitive to methodological choices, even though the overall landscape was similar for these variants. Our preference for the description goes to the extended recall.

Conclusion

Our study shows that a combination of the methods available for micro/meso studies (appropriate clustering, extension of recall) can dramatically increase recall rates in the co-citation methods. First, the choice of a clustering method is fundamental. The sensitivity of cluster size distribution to methodological choices raises once again the problem of artefact effects in describing the structures of science. For example, the use of average linkage can practically solve the singleton problem.

Thus, extended assignment enriches cores in an efficient manner. It should be emphasized that this process is strongly complementary to the first one. Extension does not create new clusters but is particularly warranted since the original cores already provide a robust and extensive coverage of the field. Complementary assignment also allows better assignment rules for citing items. However, it must be carefully conducted and achieve a trade-off between the "quantity" and "quality" of recall.

The structuring and mapping of science are generally part of interactive processes involving experts. From this point of view, the methods proposed here have two distinct advantages: first, the size distribution of clusters yields a very workable disaggregation of the universe; secondly, discussion with experts about research front cores is facilitated by the extension process. The experts find a more detailed picture of their literature of reference in comparison with the original cores which are often considered as too limited and too old.

Actor distribution inside clusters is a key-point for strategic interpretations of co-citation studies. Within the performing clustering method adopted, distribution is moderately sensitive to further variants (extension of recall/relaxation of assignment rules), which emphasizes the importance of methodological choices. The temporal features of clusters, either on the cited or citing side, are likely to be more robust with this set of methods, e.g. for trend extrapolation, if the fundamental caveats concerning the chaotic nature of changes are borne in mind.

In conclusion, co-citation methods can be reassessed to serve as partitioning tools rather than as leading edge extractors, and in this view can provide reliable indicators. Good levels of recall open up new possibilities for improved comparisons or combinations and are complementary to, rather than competive with, related methods providing good coverage, e.g. co-word analysis.

References

1. H. G. SMALL, Co-citation in the scientific literature, *Journal of the American Society for Information Science,* 24 (1973) 265-269.
2. I. V. MARSHAKOVA, Document coupling system based on references taken from *Science Citation Index* (in Russian), *Nauchno - Teknicheskaya Informatsiya, Ser. 2* (1973) No. 6,3.
3. H. D. WHITE, K. W. MCCAIN, Bibliometrics, *Annual Review of Information Science and Technology,* 24 (1989) 119-186.
4. H. D. WHITE, B. C. GRIFFITH, Author Co-citation: A Literature Measure of Intellectual Structure, *Journal of the American Society for Information Science,* 32, 3 (1981) 163-172.
5. K. W. MCCAIN, The Author Cocitation Structure of Macro-Economics, *Scientometrics* 5 (1983) 277-289.
6. H. PENAN, Analyse Des Citations, Principes, Applications à la Théorie Microéconomique, in DESVALS, DOU, *La Veille Technologique*, Dunod 1992, 277-330.
7. K. W. MCCAIN, Mapping Economics through the Journal Literature: An Experiment in Journal Co-citation Analysis, *Journal of the American Society for Information Science* 42,4 (1991) 290-296.
8. H. G. SMALL, B. C. GRIFFITH, The Structure of Scientific Literature. I: Identifying and Graphing Specialties, *Science Studies*, 4, (1974) 17-40. B. C. GRIFFITH, H. G. SMALL, J. A. STONEHILL, S. DEY, The Structure of Scientific Literature. II: Toward a Macro and Microstructure for Science, *Science Studies*, 4, (1974) 339-365.
9. J. KING, A Review of Bibliometric and Other Science Indicators and their Role in Research Evaluation, *Journal of Information Science*, 13 (1987) 261-276.
10. D. HICKS, Limitations of Co-citation Analysis/Bibliometric Modelling: A Reply to Franklin, *Social Studies of Science*, 17 (1987) 295-316.
11. R. R. BRAAM, H. F. MOED, A. F. J. VAN RAAN, Mapping of Science by combined Co-citation and Word Analysis, I Structural Aspects, *Journal of the American Society of Information Science*, 42 (1991) 233-251.
12. L. LEYDESDORFF, The Generation of Aggregated Journal-journal Citation Maps on the Basis of the CD-ROM Version of the *Science Citation Index*, *Scientometrics*, 31, 1 (1994) 59-84.
13. H. F. MOED, M. VRIENS, Possible Inaccuracies Occurring in Citation Analysis, *Journal of Information Science*, 15 (1989) 95-107.
14. D. SULLIVAN, D. H. WHITE, E. J. BARBONI, Co-Citation Analyses of Science-Evaluation, *Social Studies of Science*, 7, 2 (1977) 223-240.
15. D. HICKS, Limitations and More Limitations of Co-citation Analysis as a Tool for Science Policy, *Social Studies of Science*, 18 (1988) 375-384.
16. H. G. SMALL, E. SWEENEY, Clustering the *Science Citation Index* using Co-citations. I A comparison of methods, *Scientometrics*, 7, 3-6 (1985) 391-409 H. G. SMALL, E. SWEENEY, E. GREENLEE, Clustering the *Science Citation Index* using Co-citations. II Mapping Science, *Scientometrics* , 8, 5-6 (1985) 321-340.

17. P. HEALEY, H. ROTHMAN, P. K. HOCH, An Experiment in Science Mapping for Research Planning, *Research Policy*, 15 (1986) 233-251.
18. W. GLÄNZEL, H. J. CZERWON, A New Methodological Approach to Bibliographic Coupling and its Application to Research-Front and Other Core Documents, *Proceedings of the Fifth Biennial Conference of the International Society for Scientometrics and Informetrics*, River Forest, 7-10 June 1995, Learned Information Inc., Medford, New Jersey, 167-176.
19. A. RIP, Mapping of Science: Possibilities and Limitations, *Handbook of Quantitative Studies of Science and Technology*, A. F. J. VAN RAAN (Ed.), Elsevier Science Publishers, Amsterdam, 1988, 253-273.
20. M. ZITT, E. BASSECOULARD, Development of a method for detection and trend analysis of research fronts built by lexical or co-citation analysis.*Scientometrics*, 30, 1 (1994) 333-351.
21. M. ZITT, E. BASSECOULARD, Recall Rates of Co-citation Techniques: field Bibliometric Constraints and Strategies of Improvement in Micro Studies, *Proceedings of the Fifth Biennial Conference of the International Society for Scientometrics and Informetrics*, River Forest, 7-10 June 1995, Learned Information Inc., Medford, New Jersey, 167-176.
22. B. L. MILMAN, YU.A. GAVRILOVA, Citations in Chemical Engineering, *Scientometrics*, 27, 1 (1993) 53-74.
23. M. H. MCROBERTS, B. R. MCROBERTS, Problems of Citation Analysis: A Critical Review, *Journal of the American Society for Information Science*, 40, 5 (1989) 342-349.
24. J. J. FRANKLIN, Testing and Using Quantitative Methods in Science Policy Contexts: A Response to Hicks, *Social Studies of Science*, 18 (1988) 365-375.
25. J. A. HARTIGAN, *Clustering Algorithms*, John Wiley, New York, 1975.
26. SAS INSTITUTE Inc., Introduction to Clustering Procedures, in *SAS/STAT™ User's Guide, Release 6,03 Edition*. Cary, NC:SAS Institute Inc., 1988, 47-83.
27. L. LEYDESDORFF, Various Methods for the Mapping of Science, *Scientometrics*, 11, 5-6 (1987) 295-324.
28. R. TODOROV, J. VLACHY, National Patterns in Physics Output by Subfields, *Czechoslowak Journal of Physics*, B36, (1986], 163-166.
29. G. W. MILLIGAN, An Examination of the Effects of Six Types of Error Perturbation on Fifteen Clustering Algorithms, *Psychometrika*, 45, 3, (1980) 325-342.
30. M. M. KESSLER, Comparison of Results of Bibliographic Coupling and Analytic Subject Indexing, *American Documentation*, 16 (1965) 223-233.
31. W. A. TURNER, G. CHARTRON, F. LAVILLE F., B. MICHELET, Packaging Information for Peer Review: new Co-word Analysis Techniques, *Handbook of Quantitative Studies of Science and Technology*, A. F. J. VAN RAAN (Ed.), Elsevier Science Publishers, Amsterdam, 1988, 291-323.
32. H. G. SMALL, Macro-level Changes in the Structure of Co-citation Clusters: 1983-1989, *Scientometrics*, 26, 1 (1993) 5-20.
33. P. STERN, Bibliometric Maps as Valid Representations of Scientific Practices: Controversies and Closure in the Social Sciences, *Proceedings of the Fifth Biennial Conference of the International Society for Scientometrics and Informetrics*, River Forest, 7-10 June 1995, Learned Information Inc., Medford, New Jersey, 167-176.
34. R. H. COWARD, R. R. FRESNE, Taxonomies of Science and Bibliometric Indicators: Developing a Methodology for Classifying Research Fronts in a Co-citation Database, *Proceedings of the Joint EC-Leiden Conference on Science and Technology Indicators*, Leiden, 23-25 Oct.1991, DSWO Press, Leiden University.
35. J. P. BENZECRI et al., *L'Analyse des Données, 1 La Taxinomie*, Paris, Dunod, 1973, 180-181,
36. C. HENNEQUET, *Exploration et Validation de Méthodes de Classification en vue d'Analyses Bibliometriques*, DESS Report, Univ. Paul Sabatier, Toulouse/INRA LERECO, 1993.
37. P. WILLETT, Recent trends in hierarchic document clustering: a critical review, *Information Processing and Management*, 24, 5, (1988) 577-597.

Annex

·Data

Co-citation analysis was performed on a citing set of all articles from 1986 through 1992 in the Astrophysical Journal, published in the U.S.A, and Astronomy & Astrophysics, published in Europe. This citing dataset was split into two files: ASTRO1 (A1) containing the years 1986 through 1989 and ASTRO2 (A2) containing the years 1989 through 1992 (there was thus one overlapping year). Data-processing was carried out with the SINDBAD sequence of programs developed in our laboratory using the SAS^R statistical package.

A1: 9,152 citing articles; 66,913 cited articles; 210,960 citations. Integer citation threshold Y1 for cited articles: 21; 991 selected cited articles.

A2: 10,755 citing articles (10,705 with references to articles cited more than once); 81,270 cited articles (37,769 cited more than once); 258,545 citations (215,044 more than once). Integer citation threshold Y1 for cited articles: 21; 1,183 selected cited articles.

Citations were identified by the following key: first 4 letters of the first author's name/ publication year/ volume/ first page.

Clustering of highly-cited papers

Clustering procedures were reported in an earlier paper (op.cit. 1994) describing experiments on file A1. Briefly:

- the co-citation strength between cited papers is weighted

paper k with n_k cited references: c_{ki} boolean; $c_{ki} = 1$ if i cited by k.

$cooc(i,j) = \Sigma[p_k(c_{ki}*c_{kj})] / (\Sigma[p_k\, c_{ki}]\cdot\Sigma[p_k\, c_{kj}])^{1/2}$ Appropriate weights are $p_k = 1/n_k$ (in the reported experiments) and $p_k = 1/(n_k)^{1/2}$.

- the clustering method used is group-average linkage.

-cutting level is variable: a range is defined for cutting levels; inside the range, we use modal cutting which allows a cluster to collect peripheral elements.

- the position of each cited paper within a cluster is qualified by an empirical coefficient of "internality" given by the morphological characteristics of the ultrametric tree.

Enrichment of co-cited cores

After the clustering stage, a new citation threshold is established ($Y2 < Y1$). In the original set (table citing documents/cited documents), references to cited documents beneath the threshold are ignored.

Reindexing: for each citing document, references to cited documents corresponding to "cores" are replaced by a reference to this core, considered as a superdocument. References to cited documents outside cores (CDOC) are retained.

Selection: only the p clusters most referred to are retained by citing document. p=2 or p=3 seems to be a reasonable choice to avoid noise since a given document (except for review articles) is unlikely to be originally distributed among a large number of research fronts for the usual partition levels.

Document-cluster co-citation: we consider that each pair i-J is an elementary co-citation link [i for a CDOC cited c(i) times, J for a cluster cited c(J) times]. Co-citation links are summed up for all citing documents, giving c(i,J) and, after a Salton-Ochiai normalization, w(i,J).

In addition, counts are weighted as for document-document co-citation.

The CDOC i is eventually assigned to the cluster J for which w(i,J) is maximum. A multi-assignment is possible, but a single assignment is more consistent with the rationale of co-citation partitions.

Assignment of citing documents

A new fraction of the cited literature is now assigned to "cores" enriched by secondary documents. Assignment of citing documents is operated on this new basis according to the usual rules: a paper is assigned to a cluster at threshold X when X references or more to papers cited Y times or more belong to this cluster. Generally, multi-assignment is allowed but limited to 3 clusters. For the above tests, a paper was only assigned to the cluster(s) it cites most often, allowing ties.