Scientometrics (2018) 114:757–778
https://doi.org/10.1007/s11192-017-2610-9

Pennants for Garfield: bibliometrics and document retrieval

Howard D. White[1]

Received: 22 November 2017 / Published online: 14 December 2017

Abstract Eugene Garfield's name, like that of any prolific author, can designate both an oeuvre and a person. That duality is explored here with pennant diagrams, a decade-old technique that can structure information about both oeuvres and persons in one scatterplot. Such diagrams are not readily made now, but may have a place in recommender systems of the future. This paper recapitulates the basics of creating and understanding them. In pennants, every term in a bibliometric distribution is weighted with a version of the TF * IDF formula from information retrieval. The distributions are generated by a seed term, such as a cited author's name or a subject phrase, and consist of terms that co-occur with the seed in a database. TF * IDF orders the terms by relevance and specificity with respect to the seed—an outcome interpretable in light of relevance theory from linguistic pragmatics. Garfield's name appears illustratively as a seed in one pennant and as a co-cited author in five others. Another example shows works by him and others that co-occur with the phrase "Citation Analysis" in Scisearch. Pennants are richly suggestive about authors, and here they are linked to a fruitful idea of Garfield's that appeared in his first paper.

Keywords TF * IDF · Co-citation · Relevance theory · Specificity · Processing effort

Introduction

As is fitting, my tribute to Eugene Garfield makes use of his citation indexes, technological innovations that have permanently altered science and scholarship for the better. A while back I used his name to show how "ego-centered" analysis can be conducted in them—that is, analysis centered on individual authors (White 2000). Here, I feature his name

✉ Howard D. White
whitehd@drexel.edu

1 College of Computing and Informatics, Drexel University, Philadelphia, PA 19104, USA

again in a variety of pennant diagrams, introduced in White (2007a, b). These diagrams are a technique for visualizing bibliometric distributions of the sort his indexes can provide.

Pennants depict parts of literatures in a structured way. They are generated by seed terms, which are a form of query. Their purpose is to relate the seeds informatively to other data-points such as papers, authors, journals, or subject indicators. The data-point positions have regular patterns of interpretation, which Garfield's name and his citation counts in various contexts can illustrate. The present article recapitulates these interpretations, while also assembling examples of pennants and related constructs that have appeared in the past decade.

Key background

Pennants involve ideas that are rudimentary in information science, but that are not usually brought together.

Visualizing IPP's

A fundamental notion in bibliometrics is the "information production process" or IPP, as expounded by Egghe (2005: 7–8). In general, an IPP is a collection of sources that produce varying numbers of items by which the sources can be ranked in terms of yield. A basic bibliomentric distribution consists of the sources that co-occur with a seed term—that is, sources that can be ranked by how many items each contributes to the literature designated by the seed. The classic example is a collection of journals (sources) that can be ranked by the number of articles (items) each yields to a subject literature designated by a seed term. For example, given the set of article records that the term "Lubrication" retrieves from a large database, we can ask how frequently tokens of "Lubrication" (a constant) co-occur with tokens of each journal title (a variable) across the set. These co-occurrence counts can be ranked high to low, showing each journal's yield to the "Lubrication" literature. (On types and tokens, see Furner 2016.)

Another basic, rankable distribution in bibliometrics is simply the size of each source—that is, how many items appear in it overall. This is an occurrence count, as opposed to a co-occurrence count. Continuing the example, how many articles does each journal contain whether or not they are indexed by the seed term "Lubrication"?

Simultaneous co-occurrence and occurrence counts for the same terms, such as the "Lubrication" journal titles, form two dimensions on which the terms can be plotted. For pennant diagrams, both counts must first be weighted by a well-known formula from information retrieval. When this is done, the scatterplots of points take a roughly triangular shape, like that of a pennant flag, with the seed at the right tip. Different sectors of the pennant then predict that data points in them will tend to have certain qualitative characteristics relative to the seed. The predictions are rough in nature and can of course be falsified. However, they are also frequently verified, as may be noted in examples to come.

To illustrate the kinds of counts that will be used here, suppose we want to plot another IPP—the authors with whom Garfield is co-cited. This requires a standard citation database whose bibliographic records include each article's cited references (CR's). The seed "Garfield E" as a cited author will form the set of all articles that cite him at least once. We want to retrieve the names of other authors co-cited with him. What is counted in this case are all the other cited-author tokens (a variable) that co-occur with the Garfield seed token (a constant) at least once in the CR's across the set of records. (Each pairing of the seed with another author is counted only once per record, no matter how many *works* by the seed or the other author appear in the record's CR's.) At the same time, suppose we can obtain the second distribution—that is, counts of the times the various authors are cited across all records in the collection, whether or not Garfield is co-cited with them. The latter variable, of course, is simply each author's total citation count. Given these data, a pennant relating Garfield to his co-citees can be created. An example is interpreted in a later section.

Dialog's Rank command

The old Dialog search system, terminated by ProQuest in 2013, had a command called Rank that could swiftly display both co-occurrence and occurrence counts of the sort just described for any set of bibliographic records retrieved by a seed. I know of no system that can do this now. In Dialog, moreover, Rank could operate on sets retrieved from Scisearch, Social Scisearch, and Arts and Humanities Search, the same databases that are in the Web of Science. Pennants originated in this capability. Except where noted, the pennant data in the present article are from Dialog. They are thus not new, but they serve for purposes of exemplification.

A few years before White (2007a, b), it struck me that the co-occurrence and occurrence counts produced by Rank resembled the two kinds of data used in TF * IDF weighting of terms in information retrieval (IR). Although TF * IDF weights are customarily used on token counts of ordinary *words* to rank *documents* for relevance to a *query*, they could also be used to weight *bibliometric terms*—e.g., author names, journal titles, descriptors—by their token counts so as to rank them by their relevance to a *seed term*.

Weighting

A convenient formula for converting raw counts to TF and IDF weights appears in Manning and Schütze (1999: 543) and Manning et al. (2008: 109):

$$\text{Relevance to the seed} = (1 + \log\text{TF}) * (\log(\text{N/DF})).$$

Manning and Schütze (1999: 542) define term frequency (TF) as the number of occurrences of a query term in a document. They define document frequency (DF) as the number of documents in the collection (N) in which a query term occurs, showing whether it is relatively rare or common in the collection. The base-10 logs in the formula are damping factors.

In the IR paradigm, term frequency weights (TF's) are taken to measure the *relevance* (construed as the topical closeness) of each document to the query. Expressing TF's in logs gives pennants their tapering structure on the horizontal axis. In a hypothetical pennant, for example, items judged relevant to the seed *least frequently* might number in the high tens, while items judged relevant *somewhat more frequently* are numbered in the low tens, and items judged relevant *most frequently* are numbered in units.

However, the TF-relevance of each term is qualified by its inverse document frequency weight (IDF = N/DF, with collection size N, if unknown, set arbitrarily at a relatively large number, e.g., several million). This measures the term's *specificity* to the query (Sparck Jones 1972). As an inverse measure, IDF's are higher for terms with relatively low DF's in the collection—rare terms—which signals that the documents containing them are presumably more specific to the query. Terms with relatively high DF's—common terms—are presumably less specific, less discriminating, and so they are given lower IDF weights.

No author *name* is more specific than any other, of course. It is *title-words in their oeuvres* that vary in specificity with respect to *title-words in the seed authors' oeuvre*. Abbreviated strings denoting publications, such as "Garfield (1972), V178, P471, Science" or the even briefer "Garfield E, 1972," must also be translated into title-words for specificity to be compared.

Other standard versions of TF * IDF adjust TF weights by the length of the documents in which the query terms occur. Here, the set of records retrieved by Rank can be considered one long document that does not need adjustment for bibliometric visualization. Accordingly:

- The co-occurrence counts produced by Rank are the TF's—the frequencies of terms of interest in that long document.
- The occurrence counts produced by Rank are the DF's—the frequencies of those terms in records in the collection.

DeltaGraph software is useful for converting raw counts into logged weights and plotting the results as pennants. Since the data points tend to be very numerous, those below some minimum TF are usually not plotted. Even so, point labels in the remainder frequently overlap, a flaw needing correction when it occurs.

Note that, in pennants, the two factors of the TF * IDF formula are not multiplied. They are the separate variables of a scatterplot: logged TF weights position terms on the horizontal axis, and logged IDF weights position them on the vertical axis. But ranking bibliometric terms by their single TF * IDF weights is of course possible. Three such rankings are discussed by White (2017a) in the context of recommender systems for digital libraries.

An insight

The revelatory moment in applying TF * IDF weights to bibliomentric data came when I happened to use the seed *Moby Dick* as a cited work (CW) to retrieve the hundreds of works co-cited with it in Dialog's Arts & Humanities Search (White 2007a). CW's in this database are not papers but books and journals. When I looked at the long DeltaGraph column of works ranked by their TF * IDF weights, it became quite clear what such weighting does. *Moby Dick* itself had the top weight. Immediately under it were dozens of other works from Melville studies: some of his other novels, biographical and critical studies of him, critical studies of *Moby Dick*, books on whaling, and so on. The middle section of the same distribution contained famous American novels with which *Moby Dick* is frequently discussed (e.g., *The Scarlet Letter* by Melville's friend Hawthorne) and critical studies of nineteenth century American literature, which consider Melville and Hawthorne among others. The bottom section contained journals such as *American Literature* and *The New England Quarterly*, synoptic treatises such as Frye's *The Anatomy of Criticism*, and world classics such as the *Iliad* and *Paradise Lost*.

The TF factor was co-citation evidence that all of these works are at least somehow relevant to *Moby Dick*, but the IDF specificity factor had automatically ordered them by the *obviousness* of that relevance. There was a top group of works whose title-relevance was easy to see, a middle group whose title-relevance required more conjecture, and a bottom group whose relevance was not at all evident from their titles. Promoting and demoting terms in this way is just what the IDF factor is supposed to do, but in word-oriented IR its operation is taken for granted and almost never illustrated in any detail. The titles co-cited with *Moby Dick* brought the outcomes of TF * IDF weighting sharply and transparently into focus.[1]

This notion of *relative ease of seeing relevance* played into my longstanding interest in giving bibliometrics a more psychological cast. Although bibliometricians usually treat their data as depersonalized, TF's and DF's reflect human relevance judgments. They thus record actual cognitive associations, some of which were apparently easy to make, in the sense that relatively many persons made them; others of which were apparently harder to make, in the sense that fewer persons made them. In other words, human relevance judgments in the aggregate seem conditioned by an effort factor.

Relevance theory

During that time I had also been studying Sperber and Wilson's (1986, 1995) relevance theory from linguistic pragmatics as a possible improvement on views of relevance in information science (cf. Harter 1992). Relevance theory (RT) grounds an account of human communication (how individuals determine what spoken and written inputs mean) in an account of human cognition (how the mind efficiently processes information). S & W hold that the relevance of any input varies directly with its cognitive effects and inversely with the effort the input costs to process. The input may strengthen an assumption, eliminate it, or combine with it to yield a new conclusion—one not possible with either the input or the assumption alone. Cognitive effects are, in short, *the consequences inferred from an input*: the greater they are, the greater the input's relevance. But processing inputs also requires cognitive effort: the less effort an input costs, the more relevant it is.

Relevance in RT thus has degrees. However, individuals cannot introspect its degrees on a ratio-level scale. As in information science, it is measurable only ordinally as *greater* or *lesser*.

In the *Moby Dick* study, judgments to cite a work at all (the DF's) or to cite it with *Moby Dick* (the TF's) clearly involved cognitive effects in authors (see White 2011). At the same time, some judgments seemed easier to make than others (the IDF's) because of their relative obviousness at title level. This led me to conclude that TF and IDF in information science could be regarded as a special case of RT's terse formulation of relevance to an individual:

$$\text{Relevance of an input} = \text{effects/effort}$$
$$\text{Relevance of an input} = \quad \text{TF} * \text{IDF}$$

Multiplying TF's by IDF's, an inverse factor, is like dividing cognitive effects by processing effort. But, again, the precise-looking result is a crude estimate of greater or lesser relevance, not a true measure of it.

This line of thought underlies the papers in which I adapt RT to data from information science, where both relevance and least-effort behavior are central ideas (White 2007a, b, 2009, 2010a, 2011, 2014, 2017b). In a sense, the effects/effort ratio explains psychologically why TF * IDF weighting has been used so widely in document retrieval.

[1] I recently learned that CiteSeer uses TF * IDF to recommend publications that are bibliographically coupled with (not co-cited with) a seed document. The weight, briefly explained in Lawrence et al. (1999), is called CCIDF, standing for "the common citations between any pair of documents weighted by the inverse frequency of citation." But CiteSeer simply lists the titles so retrieved as "Active Documents" and does not explain what that means.

TF predicts effects (how well a document matches the intended sense of one's search terms), while IDF predicts effort (document matches on *specific* terms are easier to process than matches on less specific terms).

Pennant predictions

Pennants are modeled on the effects/effort ratio. The farther right items are on a pennant's horizontal axis, the greater their predicted cognitive effects in relation to the seed at the tip. Prediction here is based on the TF's of past relevance judgments by citers. On the basis of these past judgments, for example, a new reader may find that works by authors pulled close to a seed:

- Strengthen a claim by the seed author, or
- Undermine a claim by the seed author, or
- Lead to valuable new conclusions if read with the seed author's works.

The bulleted points are tailored versions of RT's three kinds of cognitive effects. If none of them happens as a consequence of reading a recommended work, its predicted relevance fails. (It could still be relevant to the reader on other grounds.)

The higher on a pennant's vertical axis items are, the less effort they are predicted to cost when considered in relation to the seed. Expressing IDF's in logs on this axis spreads items over at least two orders of magnitude, giving pennants their vertical structure: items whose relevance to the seed is specific and easy to see are pushed upward, while items whose relevance is increasingly less specific and less easy to see are pushed downward.

In pennants and elsehere, author names can mean both oeuvres and persons (White 2016a), and interpretations of pennants will often involve the names in both senses. Thus the string "Garfield E" can denote both a multi-volume body of writings and a man known for (among other things) founding the Institute for Scientific Information. Taking pairs of authors as *persons* allows consideration of their social relationships, such as mentor and student, boss and employee, colleague and colleague, foe and foe. Taking pairs of authors as *oeuvres* allows discussions of the joint verbal contexts their writings create. (Sandstrom and White (2007) presents an extensive dual analysis of this kind on the anthropologist Marvin Harris, although it uses graphics other than pennants.)

Recall that, in pennants displaying authors, the DF's are simply total citation counts. If these counts are taken to reflect oeuvres, a high count (and low IDF) suggests an oeuvre that has been used in many contexts, with a corresponding wealth of implications. Such an oeuvre may well be hard to relate to the seed at title level, owing to little or no commonality of language. This is especially likely for high-count oeuvres not drawn markedly toward the seed—that is, oeuvres whose low TF's keep them in the pennant's lower left area. In contrast, an oeuvre with a low count (and high IDF) often indicates title language that is relatively easy to relate to the seed's title language.

If the DF counts are taken to reflect persons, a count markedly lower than the seed's tends to indicate an author who is:

- Younger than the seed author.
- Junior to the seed author in years since doctorate or stage of career.
- Less famous, less established than the seed (e.g., a seed author's doctoral student).
- Identified with research topics similar to those of the seed.
- In the same research specialty and discipline as the seed.
- Identified with relatively few and relatively short publications (articles).

In contrast, a count markedly higher than the seed's tends to indicate an author who is:

- Older than the seed author.
- Senior to the seed author in years since doctorate or stage of career.
- More famous, more established, than the seed (e.g., for most information scientists, Garfield).
- Identified with a broader range of topics than the seed.
- In a different specialty and sometimes different discipline than the seed's.
- Identified with relatively many publications, often including long ones (books).

These examples deal with the two extremes to show the logic of IDF when applied to bibliometric data. When pennants involve numerous points, the greatest number of items may fall between the two extremes, as will be evident below (see White 2007a: 550–557). To clarify positioning on the vertical axis, I adopted the convention of subjectively dividing pennants into sectors A, B, and C in White (2007a, b). Generalizations could thus be made about what is predicted by placement in a given sector. The same convention is seen here. In pennants of authors as persons, the top or A sector has been identified with the seed's "juniors," the middle or B sector with the seed's "peers," and the bottom or C sector with the seed's "seniors." If authors are oeuvres, the predicted relation to the seed is always from more specific at top to broader and vaguer at bottom.

Other studies

Since the original pennant studies, different types of seeds and associated items have appeared in further publications, as Table 1 reveals. Thus, the predictions of the three sectors need to be adapted to whatever is being analyzed in the pennant. Readers may consult these replications of the pennant idea to judge how well the RT-influenced predictions are borne out. A table in the "Appendix'' describes papers that do not contain pennants but that explicitly or implicitly relate RT to bibliometric distributions, adding examples of relative ease of processing.

The most recent pennant in Table 1—Akbulut (2016a: 39)—supports interpretations with a particularly artful graphic design. Its seed is Maron and Kuhns (1960), the pioneering article on probabilistic indexing, and it uses color-coded backgrounds to differentiate the topical implications and processing costs of nine groups of co-cited papers within the sectors. Although the original thesis is in Turkish, an abbreviated version, Akbulut (2016b), is in English.

New pennants

The latest pennant here dates from 2013, the last year Dialog's Rank command was available. I will comment mainly on implications of pennant *content* and *shape*. The latter is affected by where the seed appears with respect to the other points on the vertical axis—the inverse scale of citedness. Examples of pennants more or less symmetrical in shape are Cronin's in White (2010a) and my own in White (2007a). Here, the asymmetry of several pennants is pronounced. I chose them, in part, for what they suggest about the seed author's stage of career.

Table 1 Works with pennants published during 2007–2016

Articles	Retrieved items and seed	Source
White (2007a)	Books and journals co-cited with Moby Dick	Arts & Humanities Search
	Authors co-cited with White HD	Social Scisearch
	Papers co-cited with Harter (1992)	Social Scisearch
White (2007b)	Index terms co-occuring with "Dark Tower"	MLA Bibliography
	Journals co-occurring with "Lubrication"	Scisearch
Schneider et al. (2007) and Larsen (2008)	Authors co-cited with Ingwersen P	Social Scisearch
White (2009)	Authors co-cited with Strindberg A	Arts & Humanities Search
	Authors co-cited with Persson O	Social Scisearch
White (2010a)	Authors co-cited with Cronin B	Social Scisearch
	Descriptors co-occurring with Börner K	INSPEC
Holmberg (2012)	Authors co-cited with Rothstein B	Scopus
	Authors co-cited with Hjørland B	Scopus
	Author affiliations co-cited with Rothstein's	Scopus
White and Mayr (2013)	Descriptors co-occurring with "Migration"	Social Sciences Abstracts
Tonta and Çelik (2013)	Papers co-cited with Arf C (1941)	Web of Science
White (2014)	Authors co-cited with Mann T	AHCI
	Authors co-cited with Bates MJ	AHCI
	Authors co-cited with Bal M	AHCI
	Authors co-cited with Milton A	AHCI
	Authors co-cited with Holberg L	AHCI
	Authors co-cited with Kraus K	AHCI
Akbulut (2016a)	Papers co-cited with Maron and Kuhns (1960)	Web of Science

AuthorWeb

Figures 1, 2 and 3 were created with AuthorWeb, a now-discontinued system for instantly visualizing the 24 authors most frequently co-cited with a seed author as a PFNET, a Kohonen map, or a pennant. As explained more fully in White (2014), the AuthorWeb software operated on all the records—some 1.26 million—in Arts & Humanities Citation Index (AHCI) for the period 1988–1997. This database was granted to Drexel University for research purposes by Garfield's company, the Institute for Scientific Information (ISI).

AuthorWeb is good for illustrating features of pennants, because relatively few authors are mapped and their labels are non-overlapping. AuthorWeb pennants are also directly comparable, coming from a database with fixed characteristics. The seed authors in Figs. 1, 2 and 3 were three friends—Garfield himself, Derek de Solla Price, and Belver C. Griffith. The fact that these pennants were made in a 10-year subset of a humanities-oriented

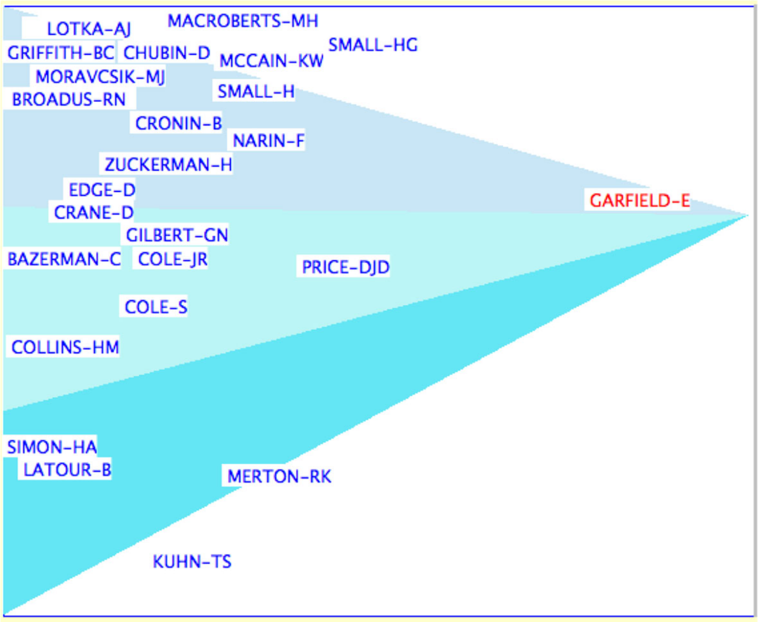

Fig. 1 AuthorWeb pennant for Eugene Garfield

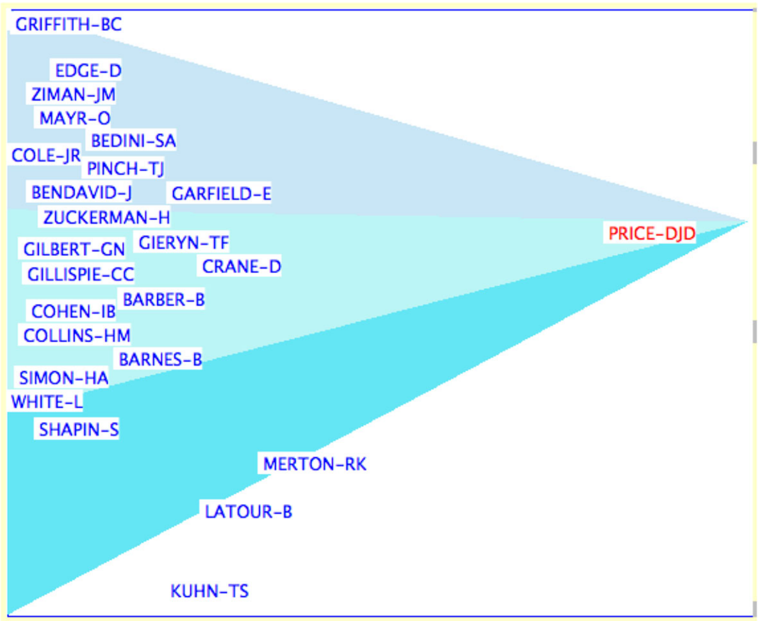

Fig. 2 AuthorWeb pennant for Derek J. de Solla Price

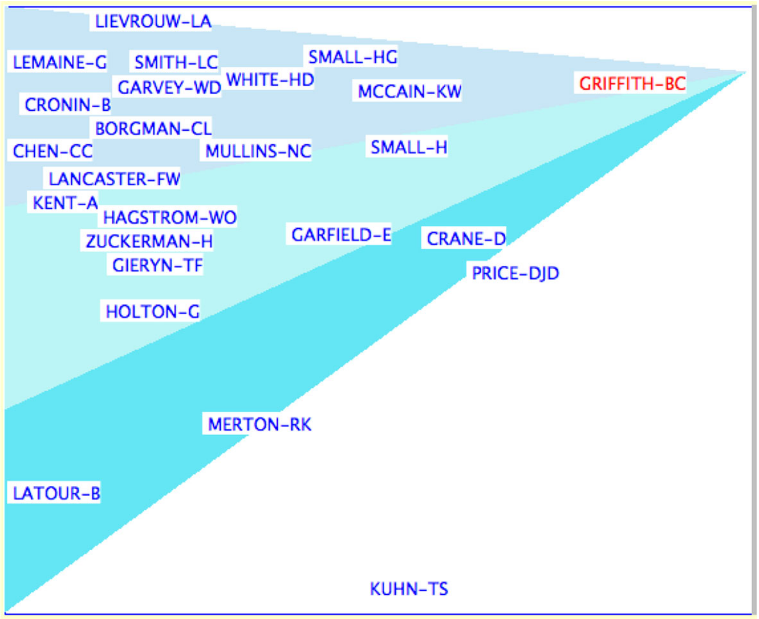

Fig. 3 AuthorWeb pennant for Belver C. Griffith

database makes them less representative of the seeds' true citation records than pennants from the full Social Sciences Citation Index would be, because SSCI covers the information science literature much better than AHCI. But they are reasonable portrayals of the three seeds from the standpoint of the humanities—including the fields of history and science studies—during 1988–1997.

Note that AuthorWeb allowed one to pass directly from authors' names in a pennant to the journal articles in which these authors were co-cited. One could then pass directly to the *works* of theirs that were co-cited. This should be borne in mind in subsequent discussions of what authors' names connote. Such names are simply the top level of a system for exploring and retrieving specific *documents*. AuthorWeb pennants, in other words, were part of a VIRI—a visual information retrieval interface. With author pennants, a reasonable question is "How can I possibly know what all these names imply?" The answer is: you don't need to. Domain knowledge of authors is of course desirable; it lets one use familiar names as seeds and recognize others that appear in the pennant. However, a proper system would permit immediate passage to more detailed bibliographic records and even to passages in full texts, where the actual implications of authors' names are made apparent. The pennants discussed here prefigure ways to exploit data in recommender systems of the future (Carevic and Mayr 2014; White 2017a).

Garfield, Price, and Griffith are all in each other's pennant, and are thus among the names that bind these visualizations together. The other authors common to all three pennants are Diana Crane, Thomas Kuhn, Bruno Latour, Robert Merton, and Harriet Zuckerman—each with considerable bearing on scientometrics, but none of them information scientists or heavily quantitative. Latour is present mainly because of what he wrote about the functions of scientific papers, but his subversive views on science in general made him very popular among humanists, and during 1987–1998 his citation count was accordingly high.

Garfield

The three oeuvres predicted to be most relevant to Garfield's in Fig. 1 are Henry Small's, Derek Price's, and Robert Merton's, which makes good sense. Small's point would be even closer to Garfield's if the counts for the two cited forms of his name (Small HG and Small H) were combined. The oeuvres of MacRoberts, McCain, and Narin are also predicted to be highly relevant to Garfield's. Co-citees that are unique to Garfield in AuthorWeb are Bazerman, Broadus, Chubin, Stephen Cole, Lotka, MacRoberts, Moravcsik, and Narin. All are easily relatable to the analysis of scientific literatures or, in the case of Bazerman, the study of science writing.

If the authors are considered as persons, the pennant is dense with social relations—for example, Small was Garfield's director of research at ISI, Garfield's intellectual opponents included MacRoberts, Edge, and Collins; Garfield's intellectual advocates included Merton, Zuckerman (Merton's wife) and the Cole brothers (Merton's students and then colleagues).

The sectors in these pennants were programmed simply to be about equal in size, but they do capture IDF differences reasonably well. Note that, from Narin upward, everyone in Garfield's pennant is in some way a citation analyst. Garfield's name itself is pulled slightly upward into the upper sector with these others. This is *specificity of specialty* emerging from blind IDF weights, although one must look at subject indicators from the underlying oeuvres to see it. (In other words, citers have assembled works from these oeuvres on the basis of topic-relatedness that is invisible here.) Below Narin, the authors are less specific—that is, broader and more diffuse in their implications. In the humanities they are not known principally as citation analysts; rather, they are historians (Price, Kuhn), rhetoricians (Gilbert, Bazerman, Latour), and sociologists (Zuckerman, the two Coles, Collins). Note as well that, insofar as the pennant represents bibliometrics, it is bibliometrics of a relatively non-mathematical kind, as one would expect in a database of humanities literature. With respect to eminence in the humanities, everyone in the pennant from Gilbert down is more highly cited than Garfield. Moreover, Price, Merton, Kuhn, and Simon are older, as their points below his suggest.

Price

In Fig. 2, the co-citee predicted to be most relevant in reading Price is Merton, with Garfield, Crane, Latour, and Kuhn close behind. (Crane, for example, developed Price's notion of modern invisible colleges sociologically.) In the humanities Garfield is in a slightly "junior" position with respect to Price, which probably would have made both of them smile. Griffith was Price's explicit disciple in scientometics, and so his "highly specific" position in Price's pennant seems appropriate. Griffith's point has almost exactly the same position in Garfield's pennant—that is, it connotes scientometrics as a specialty—but all of the others with this specialty in Garfield have disappeared in Price, to be replaced by non-scientometricians.

Price is, in fact, the most humanities-oriented of the three AuthorWeb seeds. Certainly some of his co-citees suggest his work in bibliometrics and citation analysis, either wholly (e.g., Griffith) or in part (e.g., Jonathan R. Cole), but the majority are not information scientists at all. They are sociologists, philosophers, and historians who speialize in science and technology studies (STS). As many readers will know, Price made large, non-scientometric contributions in STS, and the co-citations he accrued for it drive out most of the information scientists. Even so, the majority of Price's co-citees from STS are ranged down the left side of his pennant, indicating that, in AuthorWeb, their relevance to him is not sharply differentiated.

Griffith

A relatively low citation count pushes Griffith's name into the top sector of the pennant in Fig. 3. Some implications of a low count were bulleted earlier—the author is younger, less established, more specialized, and so on. Griffith was in fact younger and less famous than demigods like Garfield, Price, Merton, and Kuhn. But in this case it also means that the humanities were not his home turf. He did research on the co-citation mapping of scientific specialties, communication in science, and bibliometrics, and the group of papers concerned with these topics in AuthorWeb was not large. Most of the swarm of co-citees drawn toward him are thus from his own topical areas in information science, or from closely related areas of STS (Lemaine, Gieryn). Six of his co-citees are also his co-authors (Garvey, Mullins, Small, White, McCain, and Price himself). Small would be Griffith's leading co-citee if, again, his name forms were combined. As it is, Griffith's foremost co-citee is Price. The reverse is not true; Price's pennant shows Griffith as a distant junior.

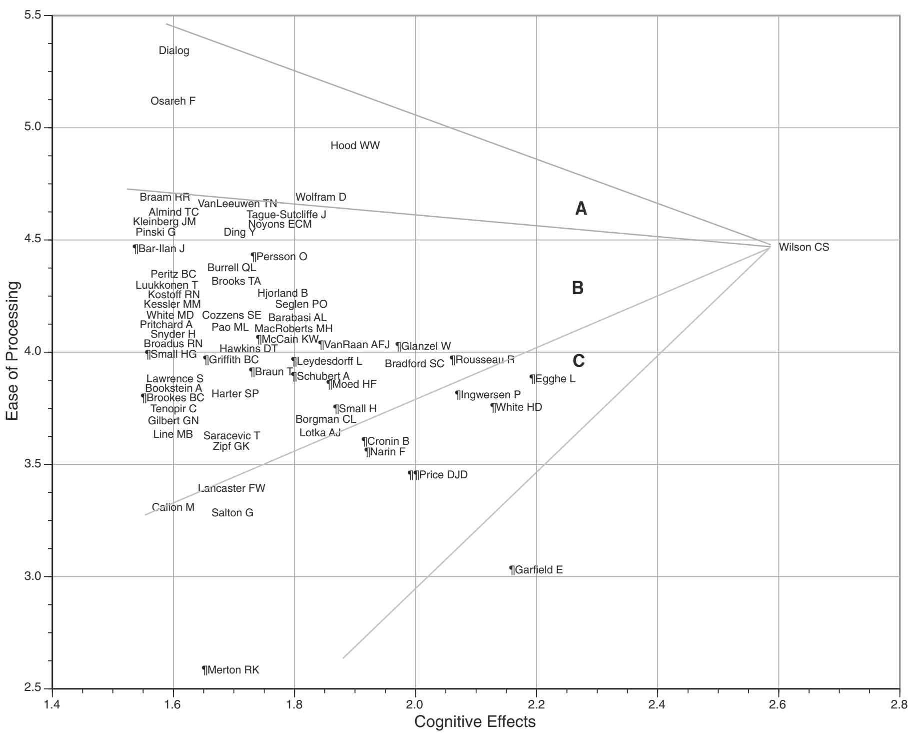

Fig. 4 Social Scisearch pennant for Connie S. Wilson and her co-citees

Wilson

Figure 4 displays the authors with whom Concepión (Connie) S. Wilson was co-cited in at least four papers in Social Scisearch on Dialog as of mid-2007. The top-ranked names in the underlying distribution were briefly discussed in White (2010a). Wilson's co-citees are largely information scientists—in particular, citation analysts and bibliometricians, but also incuding some retrievalists (e.g., Lancaster, Salton) and some researchers widely recognized in information science but outside it (e.g., Callon, Kleinberg, Barabási). Wilson's work is thus perceived by citers as quite specialized in its implications. To simplify, her name connotes *informetrics*, the quantitative side of information science, owing to her extensive work in defining it as a field (e.g., Wilson 1999).

Remarkable evidence of this connotation is that her co-citees include 20 of the 26 recipients of the Price Medal through 2017. (A ¶ appears before their names, with ¶¶ before Price himself.) Garfield was the first winner of the medal; the most recent at this writing is Judit Bar-Ilan. Speaking more generally, the many names below Wilson's on the vertical axis suggest she was principally co-cited with persons senior to her in their careers. (Her own citation record began to take off in the mid-1990s.) The lines indicating the sectors I drew arbitrarily, and readers can make their own determinations, but there are very few names above Wilson's in the A sector. Hood and Osareh were her doctoral students and then co-authors; Dialog, a corporate author, pertained to her work on online searching. Of course, many more names could be added to all sectors of the pennant if the threshold for inclusion were dropped below four.

At lower right one sees the importance of Garfield, who is virtually tied with Leo Egghe as Wilson's most frequent co-citee. Garfield's citation count is so high relative to those of other information scientists that he almost always appears in a low "infrastructural" or "foundational" position in their pennants. (See both mine and Blaise Cronin's.) But in his book, his research papers, and his very numerous informal essays, he also published on a great variety of topics (including jazz saxophone transcriptions), which makes the relation of his subject matter to the seed's difficult to predict. Egghe, by contrast, will usually deal with topics in mathematical bibliometrics, and that more restricted focus puts his point higher than Garfield's on the ease-of-processing scale.

Yue

Weiping Yue is another of Connie Wilson's doctoral students and co-authors. Her pennant in Fig. 5 is that of a young researcher at an early stage of her career: she is, so to speak, in her own A sector. Counts for her and those co-cited with her at least twice were obtained from Social Scisearch on Dialog in mid-2009.

The Yue example comes a talk on pennant methodology I gave at the Institute of Scientific and Technical Information of China in 2009; she gave the Chinese version of it. (The session was organized by the Beijing office of the successor to ISI—Thomson Reuters, now Clarivate Analytics—where Yue worked.) Her pennant is reproduced from a slide in that talk (unfortunately, the raw data file has been lost). Of primary note here is how the major figures of citation analysis are once again pulled toward the seed in foundational positions. Garfield is first among her co-citees; Henk Moed is second; Egghe, Merton and Price reappear. Yue's other co-citees represent the bibliometric side of information science quite purely. Names from the field's other large specialties, such as

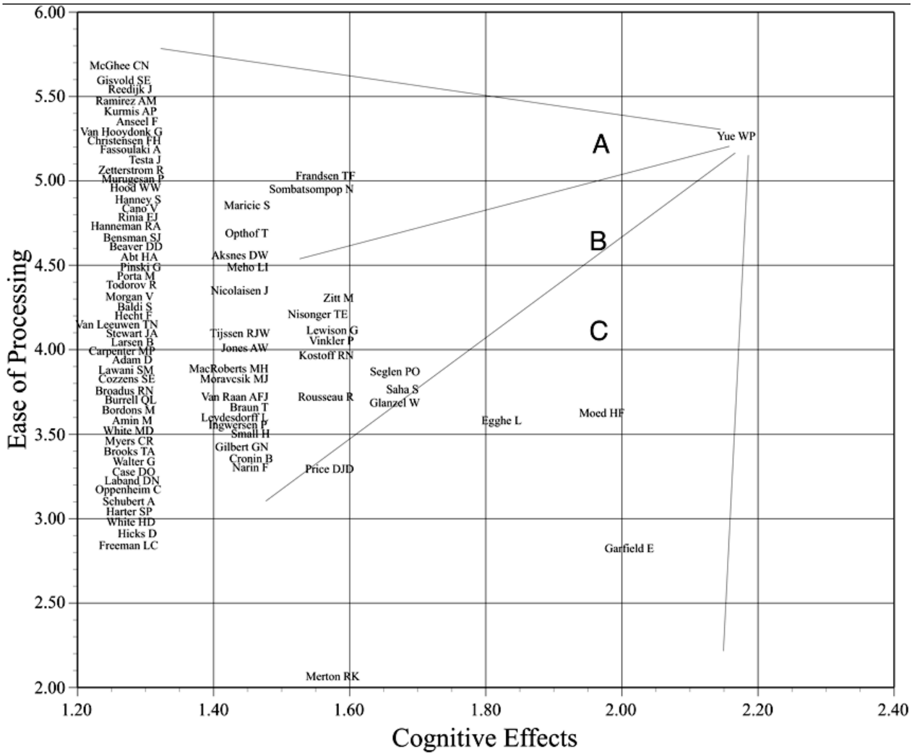

Fig. 5 Social Scisearch pennant for Weiping Yue and her co-citees

information retrieval and user behavior, were not linked to her research in any significant way as of 2009.

Another slide from the talk demonstrates the two kinds of counts underlying pennants. My input to Dialog asked for the set of documents in which Yue was cited and a ranking of her co-citees in it. Figure 6 displays the top lines of output from Dialog's Rank command on the retrieved set of 16 documents. The output has been edited to assign the counts to my relavance-theoretic and TF * IDF categories. (Dialog simply has "Items in File" for my "idf" column and "Items Ranked" for my "tf" column.) The slide helps to explain how total counts move different authors to different sectors of the pennant. The left-pointing arrows identify counts that push three famous authors into the C sector, while the right-pointing arrows identify counts that push two non-famous authors into the A sector.

Taking authors as persons, it is not hard to understand their pennant positions in terms of relative fame, stage of career, and the like. It is when authors are taken as oeuvres that matters become less clear. What needs further illustration is how title-terms in oeuvres vary in congruity to title-terms in the seed's oeuvre. By the TF factor, Merton and Price are the same as Frandsen and Sombatsompop in their relevance to Yue (four co-citations each), and Cronin and Narin are the same as Maricic and Opthof (three co-citations each). But by the IDF factor, the title-relevance of Merton, Price, Cronin, and Narin is much more difficult to see.

Some title-phrases from Wue's publications are listed in the first column of Table 2. The remaining columns list analogous terms from four of her co-citees in sector A. Yue's

```
             "Ease of       "Cognitive
             Processing"     Effects"
             in relation    in relation
              to seed        to seed
             (Inverse       (Direct
RANK No.   Measure: idf)   measure: tf) %Items Ranked  Term
--------   -------------   ------------ -------------  ----
    1              16         16             100.0%     YUE WP
    2        ───►4113         11              68.8%  ──►GARFIELD E
    3             719         10              62.5%     MOED HF
    4             720          7              43.8%     EGGHE L
    5             624          5              31.3%     GLANZEL W
    6             555          5              31.3%     SAHA S
    7             412          5              31.3%     SEGLEN PO
    8              34 ◄───     4              25.0%     FRANDSEN TF◄───
    9             329          4              25.0%     KOSTOFF RN
   10             197          4              25.0%     LEWISON G
   11       ────►11426         4              25.0%  ──►MERTON RK
   12             207          4              25.0%     NISONGER TE
   13        ───►1537          4              25.0%  ──►PRICE DJD
   14             570          4              25.0%     ROUSSEAU R
   15              34 ◄───     4              25.0%     SOMBATSOMPOP N◄───
```

Fig. 6 Partial output from Dialog Rank command for Yue and her co-citees

Table 2 Selected title terms from works by Yue and four co-citees from sector A of her pennant

Yue	Sobatsompop	Frandsen	Maricic	Opthof
Research journals	Academic journals	Journal citation identity	Croatian journals	European Heart Journal
Citation impact	Impact factor	Journal citation image	Evaluating… journals	Impact factor
Journal impact factor	Impact factors of journals	Journal impact factor	Impact factors of… journals	Citation analysis
Journal quality	Citation quality	Journal interaction	Citation histories	
Immediacy index	Citation report	Journal self-citations	Citation context	
	Science Citation Index		Science Citation Index	

terminology and theirs turn out to be close at about the same high level of specificity. They are phrases whose mutual relevance can be readily inferred.

In contrast, consider the less specific, more inclusive title phrases from pre-2009 works by authors low in Yue's pennant:

- Cronin's "citation process," "academic writing," and "web-based citation analysis".
- Narin's "publication and citation analysis" and "bibliometric performance measures".
- Price's "networks of scientific papers" and "citation measures of hard science".
- Merton's "Matthew effect in science" and "sociology of science".

Price's (1970) phrase "citation measures of hard science" actually comes from the title of the chapter in which he introduced the "immediacy index" that Yue writes about, but his title is clearly broader than hers.

It happens in this case that the sector C oeuvres by Garfield, Moed, Egghe, and Glänzel include titles that *are* easy to relate to Yue's, and these authors are also her top-ranked co-citees. That improves pennant interpretability, but it is also somewhat unusual. The cause may be that Yue built her early oeuvre on papers by figures such as Garfield and Moed who directly addressed her topics, e.g., journal impact factors, in their titles. The fact remains that authors in sector C have usually written on many topics, and so their names have broader, less predictable connotations than names in sector A.

Mayr

Garfield's work and Philipp Mayr's are related in the eyes of citers, but distantly. The latter (now Mayr-Schlegel) is a mid-career researcher at GESIS, the Leibniz-Institute for the Social Sciences, whose mission includes providing digital library services to social scientists throughout Germany. Mayr publishes on techniques from various specialties that can be adapted to the digital library environment so as to provide innovative services for users. Figure 7 shows the authors who were co-cited with him at least four times in Social Scisearch on Dialog as of mid-2013. Their names connote the broad ties of the digital library movement to technical areas of document retrieval, bibliometrics, webometrics, and visualization.

Mayr's pennant shows pre-web and post-web specialties intermingling. Thus, many of the bibliometric stalwarts seen in Wilson's and Yue's pennants are found in his at lower

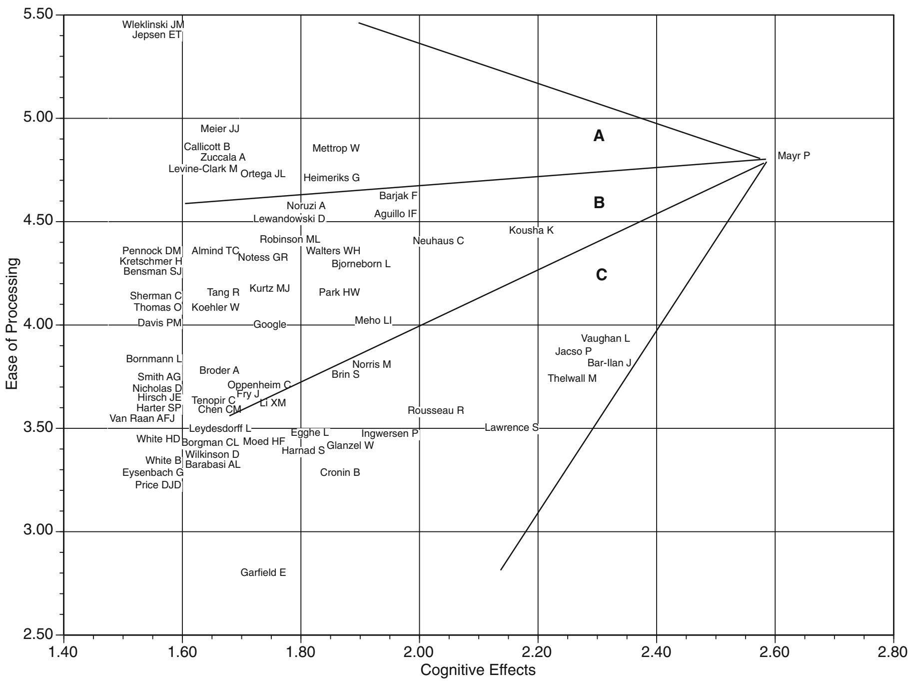

Fig. 7 Social Scisearch pennant for Philipp Mayr and his co-citees

left, with Garfield farthest down. The emergence of newer lines of research is symbolized in sector C by authors such as Steve Lawrence and Seymour Brin, and in sector B by Google as a corporate author.

More particularly, the authors closest to Mayr represent web retrieval and webometrics rather than traditional bibliometrics. His own most highly cited papers include several analyses of Google and Google Scholar. Kousha, Vaughan, Jacso, Bar-Ilan, and Thelwall are leaders in putting these and other search engines to new uses, often with evaluative citation analysis as a precedent. For example, Vaughan, Kousha, and Thelwall have tested Google Scholar against comparable tools, developing new data-gathering techniques in the process; Bar-Ilan and Jacso have explored Google Scholar's weaknesses. This makes their predicted high relevance to Mayr intelligible.

One further example with respect to authors as persons: Cronin's pennant (White 2010a) was based on Social Scisearch data from early 2008, and Thelwall is in Cronin's sector A. In Mayr's pennant, created five-and-a-half years later, he is in sector C. Such movement is of course relative to the seed's total citation count, but at this writing in 2017, Thelwall has become another Price medalist, and he would now be eminent when matched against anyone in information science. Pennants can exhibit such developments.

Citation analysis

The pennant in Fig. 8 is different from the others. It was made in DeltaGraph in 2002 from all the records of papers in Dialog's Scisearch that contained "Citation Analysis" as a natural-language phrase. These records included lists of the papers' cited references

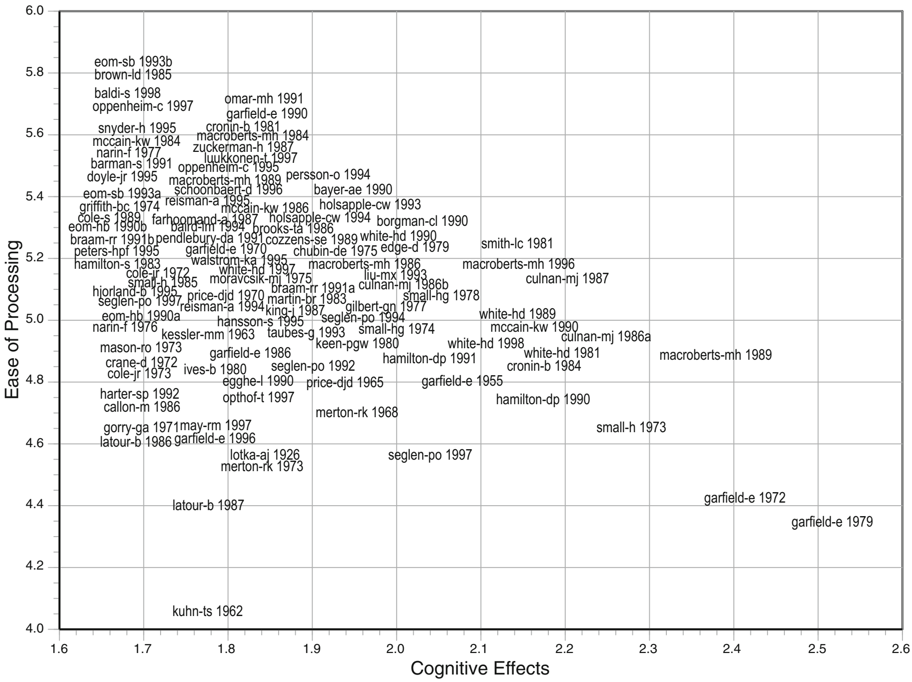

Fig. 8 Scisearch pennant of works co-occurring with the term "Citation Analysis"

(CR's). The latter can designate cited articles, books, or other publications. The Rank command was used to rank all the CR's in the set by how many tokens each yielded— for example, all tokens of the type "Garfield E, 1970, V227, P669, Nature." At the time there were exactly 250 records with at least one CR.

The pennant had to be tailored to fit a journal page. Its actual seed, "Citation Analysis," has an occurrence count much larger than any co-occurrence count with a CR, and so displaying it would make the pennant over-wide. The data points are thus almost all in the seed's A sector. I have also left out sector lines and two works whose DF counts made the pennant undisplayably deep (one was Lowry et al. 1951, the most cited of all papers). For as much legibility as possible, I show only CR's co-occurring with the seed at least five times and, like Akbulut (2016a: 39), truncate them to author and date. Nevertheless, many overlapping labels had to be teased apart, resulting in small displacements of points in a crowded display.

That said, the pennant can be read exactly like the others. Someone with enough patience to do lookups in Google Scholar will find the specifity factor has been maintained. Titles that are higher or rightward tend to be easier to relate to "Citation Anlysis" than titles that are lower or leftward.

The two works most relevant to "Citation Analysis" are by Garfield, and he himself has noted their significance. In Garfield (2013: 5), after mentioning an ISI project in which thousands of scientists commented on their most cited papers, he continues: "Quite often authors would say that these were not necessarily their most important papers. So when I was asked to write a commentary about my 1955 paper in *Science,* I had to point out that this primordial paper on Citation Indexing was neither my most cited work nor my most significant. My 1972 paper in *Science*, on using citation analysis to evaluate journals, has attracted much more attention. My most cited work is in fact my 1979 book, *Citation Indexing*" (Garfield 1955, 1972, 1979). The pennant renders these remarks visually, and sets them in the context of numerous associated works. Visualizations of Garfield's works in context by Yongxia et al. (2009) are in the same spirit but relatively chaotic.

As if made by a recommender system for a hypothetical user, the pennant in Fig. 8 prioritizes classic articles and books from the pre-2000 citation-analytic literature. The third most relevant item, MacRoberts and MacRoberts (1989), is an attack on citation analysis that citers reflexively use to symbolize the anti-citationist position; hence its high TF count. (Garfield 1997 is a rebuttal.) The pennant also picks up significant works on *co*-citation analysis by Small, Culnan, McCain, and myself, and widens leftward into further well-known contributions to the specialty designated by the seed and to scientometrics in general. Among them are Smith's and Liu's reviews of citation analysis, Cronin's trend-setting book on it, Seglen's cautionary account of impact factors, and Hamilton's papers on uncitedness, important because they appeared in *Science.*

Perhaps the most striking outcome is that the pennant has substantial structure in common with those shown earlier in this paper, especially Garfield's in AuthorWeb. Once again we see Latour, Merton, and Kuhn in supporting roles at lower left, only this time it is individual works of theirs that are mapped and that can be bibliographically traced to full texts. Major works by Price, Cronin, Crane, Cole, and Griffith are visible, along with seven items by Garfield, the most in the map. The convergence with the humanities-oriented AuthorWeb map happened despite (1) use of a much larger *science*-oriented database, in which (2) the seed was a subject phrase rather than an author name, and (3) Dialog was asked for co-occurring cited works rather than co-cited authors.

This argues for the strength of the pennant technique. If it prioritizes works in one specialty this intelligibly, it can do the same in others.

Conclusion

Pennants like those from Fig. 4 onward are labor-intensive to create, and were so even when Dialog made obtaining data for them relatively easy. Similarly, the Akbulut (2016a) pennant, whose data were drawn from the Web of Science, is crafted at a level that is anything but casual. By contrast, a user exploring AuthorWeb could quickly and automatically generate pennants at will. Both artists' and scholars' names would work as seeds, and a co-citee's name from one might serve as the seed in another. Furthermore, as stated above, the bibliographic data underlying the names could be called up immediately.

The contrast evokes what, in talks, I have called a "poster vs. Kleenex" distinction between visualizations. In the present case, a "poster" would be a visually striking pennant like Akbulut's that might be mounted on a wall for long-term exhibition, or more likely published in an article or book. A "Kleenex" would be a quickly disposable pennant, computer-produced as part of a sequence for rapid browsing—e.g., one by AuthorWeb. Poster pennants enhance in-depth discussions of a seed's significance (such as the historical interest of the mathematician Arf in Tonta and Çelik 2013). Kleenex pennants are basically a possibility for recommender systems. If a system recommends more items than can be easily labeled in one display (as in Fig. 8), mouse-over labeling is an option. Users simply need to know what different areas of the pennant imply.

Assuming quickly disposable visualizations are a likelier prospect, one must stress that citation-based pennants can make recommendations over more variegated topical areas than standard recommender systems. In IR research, the paradigmatic test of such systems is how well a retrieved set of documents matches a set pre-scored as relevant to a query. The pre-scoring is usually done by a single judge on the basis of document-query similarity at the global level. But *global topical similarity* is not the sole basis for inferring relevance; that was explicitly why Harter (1992) rejected the IR evaluation paradigm and introduced Sperber and Wilson's (1986) more inclusive notion of relevance to information science (cf. White 2017b). Pennants can and do pick up global topical similarities by means of high TF and high IDF weights, but they can also pick up non-global kinds of relevance as perceived by multiple citers (White 2017a). For instance, Garfield's (1979) book on citation indexing is not self-evidently relevant to Maron and Kuhns's (1960) paper on probabilistic indexing, yet in Akbulut (2016a: 39) it appears with other bibliometric works in sector C of the Maron–Kuhns's pennant. This exemplifies one of the countless inferences by citers that relevance judges in IR experiments would not ordinarily make.

A related argument has long been used for marketing purposes by ISI and its successors: citation databases capture relevance relationships not captured by standard word-matching and subject indexing. The very first article on citation indexing, Garfield (1955), provides a good example: although certain papers citing Selye (1946), a famous paper in endocrinology, were very diverse in subject matter, they could still provide, in Garfield's phrase, "confirmatory evidence for some of Selye's claims." *Strengthening a claim* is not global topical relevance; it is the first of Sperber & Wilson's cognitive effects.

In that same article Garfield wrote: "If one considers the book as the macro unit of thought and the periodical article the micro unit of thought, then the citation index in some respects deals in the submicro or molecular unit of thought. It is here that most [subject] indexes are inadequate, because the scientist is quite often concerned with a particular idea rather than with a complete concept. 'Thought' indexes can be extremely useful if they are properly conceived and developed." Further on, he added: "By using authors' references in compiling the citation index, we are in reality utilizing an army of indexers, for every time an author makes a reference he is in effect indexing that work from his point of view."

Pennants are one more extension and enrichment of this argument, begun so ably by Garfield at the start of his magnificent career.

Appendix

See Table 3.

Table 3 Papers with examples that bear on pennant analysis. As stated earlier, they do not contain pennants but explicitly or implicitly relate relevance theory to bibliometric distributions, adding examples of relative ease of processing. The first two involve terms weighted by TF * IDF; the remainder involve terms that are unweighted

Articles	Retrieved items and seed	Source
White (2010a)	Descriptors co-occurring with "Information Needs"	ERIC
	Descriptors co-occurring with Börner K	INSPEC
	Authors co-cited with Cronin B	Social Scisearch
	Authors co-cited with Wilson CS	Social Scisearch
	Books and journals co-cited with Kuhn (1962)	Scisearch
White (2017a)	Papers co-cited with Bates (1989)	Social Scisearch
	Papers co-cited with Manning et al. (2008)	Social Scisearch
	Papers co-cited with Bonacich (1987)	Social Scisearch
White (2010b)	Authors cited by Ingwersen P	Social Scisearch
	Authors co-cited with Ingwersen P	Social Scisearch
	Authors cited by White HD	Social Scisearch
	Authors co-cited with White HD	Social Scisearch
White (2011)	Readings assigned in lectures of an online course	Deirdre Wilson
	Authors cited by Carston R	Social Scisearch and
	Authors co-cited with Carston R	Arts & Humanities Search
	Authors cited by Recanati F	Social Scisearch and
	Authors co-cited with Recanati F	Arts & Humanities Search
	Authors cited by Newman MEJ	Scisearch
	Authors co-cited with Newman MEJ	Scisearch
	Papers and books cited by Newman MEJ	Scisearch
White (2016a)	Authors co-cited with Smiraglia R	Social Scisearch
	Authors co-cited with Ingwersen P	Social Scisearch
	Authors cited by Cronin B	Social Scisearch
	Authors co-cited with Cronin B	Social Scisearch
White (2016b)	Journals citing Swain JM	Web of Science
	Books acquired in LC class Z711 in the 1980s	OCLC WorldCat
	Books acquired in LC class Z711 in the 2000s	OCLC WorldCat
	Journals co-occurring with "Lubrication"	SCI (WoS)
	Journals co-occurring with "Kierkegaard"	SSCI and AHCI (WoS)
	Journals co-occurring with "Fairy Tales"	SSCI and AHCI (WoS)
	Readings assigned in lectures of an online course	Deirdre Wilson

References

Akbulut, M. (2016a). Atıf klasiklerinin etkisinin ve ilgililik sıralamalarının pennant diyagramları ile analizi [The analysis of the impact of citation classics and relevance rankings using pennant diagrams]. Yayımlanmamış yüksek lisans tezi, Hacettepe Üniversitesi, Ankara [Unpublished master's thesis, Hacettepe University, Ankara]. http://www.mugeakbulut.com/yayinlar/Muge_Akbulut_YL_Tez.pdf.

Akbulut, M. (2016b). Extended abstract: The analysis of the impact of citation classics and relevance rankings using pennant diagrams. http://www.mugeakbulut.com/yayinlar/tez_extended_abstract.pdf.

Arf, C. (1941). Untersuchungen über quadratische Formen in Körpern der Charakteristik 2. Teil I. [Investigations on quadratic forms in bodies of characteristic 2. Part I.] Journal für die Reine und Angewandte Mathematik [Journal for pure and applied mathematics], 183, pp. 148–167.

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review, 13,* 407–424.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology, 92*(5), 1170–1182.

Carevic, Z., & Mayr, P. (2014). Recommender systems using pennant diagrams in digital libraries. In *Paper presented at NKOS workshop, London*, September 12, 2014. https://arxiv.org/ftp/arxiv/papers/1407/1407.7276.pdf.

Egghe, L. (2005). *Power laws in the information production process: Lotkaian informetrics*. Amsterdam: Elsevier.

Furner, J. (2016). Type-token theory and bibliometrics. In C. R. Sugimoto (Ed.), *Theories of informetrics and scholarly communication*. Berlin: Walter de Gruyter GmbH & Co KG.

Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science* (Vol. 122, pp. 108–111). http://www.garfield.library.upenn.edu/papers/science_v122v3159p108y1955.html.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science* (Vol. 178, pp. 471–479). http://www.garfield.library.upenn.edu/essays/V1p527y1962-73.pdf.

Garfield, E. (1979). *Citation indexing: Its theory and application in science, technology, and humanities*. New York: Wiley. http://www.garfield.library.upenn.edu/ci/title.pdf.

Garfield, E. (1997). Validation of citation analysis [Letter, with a rejoinder by MacRoberts]. *Journal of the American Society for Information Science, 48*(10), 962–963.

Garfield, E. (2013). A century of citation indexing: Keynote address. COLLNET 2011 In *Proceedings* (pp. 5–10). https://mafiadoc.com/collnet-2011-proceedings_59aff6e81723ddb8c56166a8.html.

Harter, S. P. (1992). Psychological relevance and information science. *Journal of the American Society for Information Science, 43*(9), 602–615.

Holmberg, J.H. (2012). Dynamisk kunskapsorganisation: teoretisk ansats och implementering. [Dynamic knowledge organization: Theoretical approach and implementation]. Bachelor Thesis, University of Borås/Swedish School of Library and Information Science.

Kuhn, T. S. (1962). *The structure of scientific revelutions*. Chicago: University of Chicago Press.

Larsen, B. (2008). Informetrics and IR. Presentation at the Nordic Research School in Information Studies (NORSLIS), Umea, Sweden. http://itlab.dbit.dk/~blar/files/Norslis_Umea-june2008_BL2.ppt.

Lawrence, S., Giles, C. L., & Bollacker, K. (1999). Digital libraries and autonomous citation indexing. *IEEE Computer, 32*(6), 67–71.

Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951). Protein measurement with the Folin phenol reagent. *Journal of Biological Chemistry, 193*(1), 265–275.

MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for information Science, 40*(5), 342–349.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *An introduction to information retrieval*. Cambridge: Cambridge University Press.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge: MIT Press.

Maron, M. E., & Kuhns, J. L. (1960). On relevance, probabilistic indexing and information retrieval. *Journal of the ACM, 7*(3), 216–244.

Price, D. J. D. (1970). Citation measures of hard science, soft science, technology, and nonscience. In C. E. Nelson & D. K. Pollock (Eds.), *Communication among scientists and engineers*. Lexington: Heath Lexington Books.

Sandstrom, P. E., & White, H. D. (2007). The impact of cultural materialism: A bibliometric analysis of the writings of Marvin Harris. In L. A. Kuznar & S. K. Sanderson (Eds.), *Studying societies and cultures: Marvin Harris's cultural materialism and its legacy* (pp. 20–55). Boulder: Paradigm Publishers.

Schneider, J.W., Larsen, B., & Ingwersen, P. (2007). Pennant diagrams, what is it [sic], what are the possibilities and are they useful? In *Presentation at the Nordic Workshop on Bibliometrics and Research Policy*, Copenhagen, September 13-14, 2007. https://pdfs.semanticscholar.org/b674/7068496b8b72a5b017281b2dce75844b1e3d.pdf.

Selye, H. (1946). The general adaptation syndrome and the diseases of adaptation. *Journal of Clinical Endocrinology, 6*(2), 117–230.

Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application to retrieval. *Journal of Documentation, 28*(1), 11–21.

Sperber, D., & Wilson, D. (1986). Relevance: Communication and cognition. Harvard University Press.

Sperber, D., & Wilson, D. (1995). Relevance: Communication and cognition. 2nd edn. with postface. Blackwell.

Tonta, Y., & Çelik, A. E. Ö. (2013). Cahit Arf: Exploring his scientific influence using social network analysis, author co-citation maps and single publication h index. *Journal of Scientometric Research*, 2(1), pp. 37–51. http://www.jscires.org/article/38.

White, H. D. (2000). Toward ego-centered citation analysis. In B. Cronin & H. B. Atkins (Eds.), *The web of knowledge: A festschrift in honor of Eugene Garfield* (pp. 475–496). Medford Township: Information Today.

White, H. D. (2007a). Combining bibliometrics, information retrieval, and relevance theory, part 1: First examples of a synthesis. *Journal of the Association for Information Science and Technology, 58*(4), 536–559.

White, H. D. (2007b). Combining bibliometrics, information retrieval, and relevance theory, part 2: Some implications for information science. *Journal of the Association for Information Science and Technology, 58*(4), 583–605.

White, H.D. (2009). Pennants for Strindberg and Persson. Celebrating scholarly communication studies: A festschrift for Olle Persson at his 60th birthday, Special volume of the e-newletter of the International Society for Scientometrics and Informetrics, 5-S, 71–83. https://www.researchgate.net/publication/229861362_Pennants_for_Strindberg_and_Persson.

White, H. D. (2010a). Some new tests of relevance theory in information science. *Scientometrics, 83*(3), 653–667.

White, H.D. (2010b). Ingwersen's image and identity compared. In Larsen, B., Schneider, J.W., Ångström, F.,Schlemmer, B. (Eds.), The Janus faced scholar: A festschrift in honour of Peter Ingwersen. Special volume of the e-zine of the International Society for Scientometrics and Informetrics, 6-S, pp. 219–227. http://vbn.aau.dk/files/90357690/JanusFacedScholer_Festschrift_PeterIngwersen_2010.pdf#page=222.

White, H. D. (2011). Relevance theory and citations. *Journal of Pragmatics, 43*(14), 3345–3361.

White, H. D. (2014). Co-cited author retrieval and relevance theory: Examples from the humanities. *Scientometrics, 102*(3), 2275–2299.

White, H. D. (2016a). Authors as persons and authors as bundles of words. In C. R. Sugimoto (Ed.), *Theories of informetrics and scholarly communication: A festschrift in honor of Blaise Cronin*. Berlin: Walter de Gruyter.

White, H. D. (2016b). Bibliometrics, librarians, and bibliograms. *Education for Information, 32*(2), 125–148.

White, H. D. (2017a). Bag of works retrieval: TF * IDF weighting of works co-cited with a seed. *International Journal on Digital Libraries* pp. 1–11. https://link.springer.com/article/10.1007/s00799-017-0217-7.

White, H. D. (2017b). Relevance theory and distributions of judgments in document retrieval. *Information Processing and Management, 53*(5), 1080–1102.

White, H.D., & Mayr, P. (2013). Pennants for descriptors. In *Paper presented at the NKOS Workshop*, Valletta, Malta, September 26, 2013. https://arxiv.org/abs/1310.3808.

Wilson, C. S. (1999). Informetrics. *Annual Review of Information Science and Technology, 34,* 107–247.

Yongxia, L., Liu, Z., & Chen, C.M. (2009). Eugene Garfield's contributions to the formation and development of citation analysis—A visual analysis of Eugene Garfield's publications in celebration of his 84th birthday. In *Proceedings of the Fifth International Conference on WIS & Tenth COLLNET Meeting, Dalian, China*, September 13-16, 2009. CD-ROM.