Citation and hyperlink networks

Andrea Scharnhorst[1,*] and Mike Thelwall[2]

[1]Networked Research and Digital Information, Netherlands Institute for Scientific Information Services, Royal Netherlands Academy of Arts and Sciences, Joan Muyskenweg 25, 1090 HC Amsterdam, The Netherlands
[2]School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1SB, UK

Keywords: Citation, co-authorship networks, complex networks, hyperlink networks.

FROM the very beginning the principle of citations indexing[1] incorporated the idea of visualizing scientific information in the form of a graph. One prominent example is the concept of historiograph proposed by Garfield himself[2].

In a historiograph, an evolutionary tree is constructed with a key (or parent) paper at the top followed by subsequent temporal layers of citing papers. Co-citation graphs set a mark in the visualization of scientific specialties and research fronts[3–5]. Accompanied by co-author graphs[6], citation graphs[7] and co-word graphs[8] all these graphs have been used to visualize and analyze the growth of specialties, the structure inside scientific communities and the flow of information in science. Recently, a new branch in information science emerged which devotes itself to the visualization of knowledge domains[9]. In this area tools have been developed to automatically create graphs, to explore different visualization approaches and to navigate through massive quantities of scientific information[10–12]. Citation landscapes built in analogy to fitness landscapes[13] make the occupation and evaluation of different scientific specialities visible and allow both for navigation through research fronts and identification of possible innovative areas[14].

The visualization of emergent structures is not the only useful way to analyse collections of interconnected scientific documents, however. It is also possible to explore network structures by means of social network theory and measures developed in statistical physics. Complex network theory, a branch of statistical physics, mainly concentrates on analysing degree distributions (e.g. the numbers of citations documents in a collection receive), clustering coefficients and abstract theoretical mathematical models to explain the empirical findings. Social network analysis, in contrast, concentrates more on the interpretation of the social natures of the units and of the links between them[15]. In this paper we will address both issues.

Citation networks and co-authorship networks

Complex network theory is a field in statistical physics that analyses the statistical properties of huge networks across nature and society[16–18].

The analysis of citation and co-author networks from the point of view of statistical physics reveals some basic structural regularities. It helps to generalize the known skewed distributions of citations and collaborations (few articles are highly cited, most have few citations; few authors collaborate extensively, most have few collaborators), raises specific expectations how the science system's publication patterns may behave and points to the existence of underlying cognitive and social structures. It suggests a dynamic interplay of behavioural rules obeyed by a huge system of interacting people, which results in the observed macro-level regularities. Different modelling approaches have been used to create heuristic explanations for the observed macro-level properties[19–21]. One behavioural rule is the principle of preferential attachment[16] which reformulates the 'success breeds success' principle introduced by Price 30 years ago to explain characteristics in the 'network of papers'[22]. What most of the models are

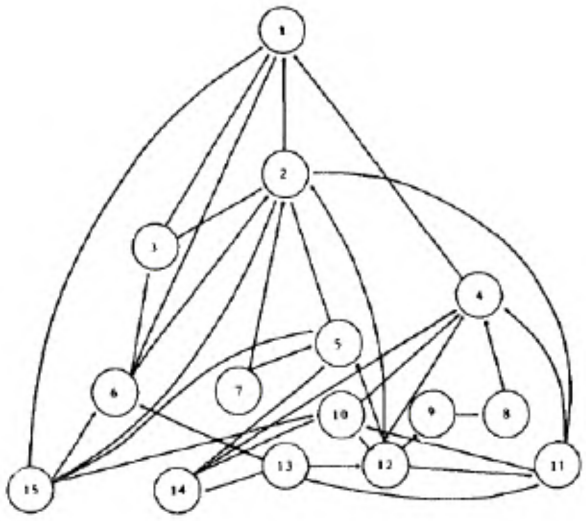

Figure 1. A diagram of citation relations in the field 'Staining Nuclid Acid' drawn from a bibliography by Gordon Allen – prototype of a historiograph (Reproduced from Garfield, E., Historiographs, librarianship and the history of science, Reprinted in *Essays of an Information Scientist*, 1974–1976, vol. 2, pp. 136–150).

*For correspondence. (e-mail: andrea.scharnhorst@vks.knaw.nl)

missing is the incorporation of social science theoretical approaches, an omission avoided by Price when he based his principle on Merton's theory on scientific reward and the Matthew effect in science[23,24]. What remains as a task is to understand the network-creating mechanisms as part of a specific communication culture in science, which distinguishes the science system from other systems inside society and at the same time links the science system to them[25]. In other words, citations theories have to be taken into account in the verification of network models and the interpretation of indicators based on them. We return to this point later in the paper.

Hyperlink networks

Complex network theory has taken bibliometric citation/co-author networks as examples of social networks, and the same holds for networks based on linked web pages. Indeed the development of complex networks was very much influenced by the availability of data on the web.

Without the availability of digitized data none of the examples actually used in complex network theory could have been exploited. Furthermore, the Internet and Web provided objects for investigation. In a recent review, Dorogovtsev and Mendes[26] took the Internet and the World Wide Web as THE examples for complex networks. They write about the starting point of the rapid growth in this field: 'The first experimental data, mostly for the simplest structural characteristics of the communication networks, were obtained in 1997–1999. … After these findings, physicists started intensive study of networks in various areas, from communications to biology and public relations'. The key literature of this beginning, according to them, was the following, all of which have the Web as their main topic.

- Huberman, B. A., Pirolli, P. L. T., Pitkow, J. E. and Lukose, R. J., Strong regularities in World Wide Web surfing. *Science,* 1998, **280**, 95.
- Albert, R., Jeong, H. and Barabási, Albert-László. The diameter of the world-wide web. *Nature,* 1999, **401**, 131, cond-mat/9907038.
- Faloutsos, M., Faloutsos, P. and Faloutsos, C., On power–law relationships of the internet topology. *Comput. Commun. Rev.*, 1999, **29**, 251.
- Huberman, B. A. and Adamic, L. A., Growth dynamics of the world-wide web. *Nature*, 1999, **401**, 131, cond-mat/9901071.
- Kumar, R., Raghavan, P., Rajagopalan, S. and Tomkins, A., Extracting large-scale knowledge bases from the web. *The VLDB J.*, http://www.Vldb.Org/Conf/1999/P60.pdf.
- Kleinberg, J. M., Kumar, R., Raghavan, P., Rajagopalan, S. and Tomkins, A., The web as a graph: measurements, models, and methods. 1999; http://www.almaden.ibm.com/cs/k53/cocoon.ps

In parallel with developments from statistical physics and computer science, information scientists have taken inspiration from statistical network analysis[27] or have bypassed these theories and moved directly from citation analysis to hyperlink analysis to understand better the mechanism of science and to get alternative instruments for research evaluation[28–32]. Hyperlink analysis covers now links between academic units on different levels of aggregation. For example, Heimeriks *et al.*[33] explored the use of hyperlink networks on the level of countries (links between academic institutions in different European countries), universities across countries and in single countries, departments, and individual researchers.

One early research hope for the web was that different areas in society would be present and interact online, allowing investigations into the interfaces between the different sections of society, such as universities and industry[34]. This led to investigations into co-words appearance on web pages (for instance 'University' and 'Government'), revealing similarities and dissimilarities between different countries (country domains)[35]. This attempt to identify connectivity patterns between countries other than by direct links shows that information on the web about topics and institutions may be found outside its hyperlink structure, such as in the content of the web pages.

The context in which specific hyperlinks appear can give insights into the functionalities of hyperlinking that are unavailable from large sample studies. This kind of individualized hyperlink analysis proposed by virtual ethnography[36,37], combined with interviews and hyperlink studies on a meso level of single specialties[38] could enrich the debate about the meaning of hyperlink networks for webometric purposes.

The topological structure of hyperlink networks revealed in large-scale analysis can reveal which mixtures of features of an academic institution or a scientific field is visible in networks. For instance, it has been shown that for universities, the importance of geography and the ability to use the web technology are visible in hyperlink networks[33,39]. On the other hand, small-scale analyses of institutional websites have revealed that collaboration structures are visible in hyperlinks[40], and website harvesting around specific groups in scientific specialties can detect a complex 'websphere' reflecting informal and formal communication and reputation patterns visible in the composition hyperlinks between websites[41].

These different types of content-motivated hyperlink analysis produce hypotheses about the motivation and dynamics of hyperlinking behaviour that must be tested with research directly addressing hyperlink motivations. The combination of large-scale structural analysis with content studies of communication patterns on a meso level and single case studies which include the behaviour of the actors creates in theory a possibility to develop a rounded theory of linking behaviour. However, the ongoing debate around citation theories shows that the complicated issue

of collective behaviour which creates patterns of interaction and appearance in a self-organized way allows and maybe even requires for several, sometimes competing, theories of explanation[25,42].

Motivations for linking: Citations vs hyperlinks

Citer motivations

A considerable body of research in bibliometrics has been devoted to finding out why authors cite papers and there is emerging parallel research into why web authors create hyperlinks. Cronin[43] has eloquently argued for the value of citation analysis and in this section we draw out a range of issues that are particularly useful to illuminate parallel phenomena in link analysis. We make no apology for some duplication of Cronin's paper (written before ours) because of the importance of the issue of citer motivations. The theories of the influential sociologist Merton suggest that citations in papers tend to show intellectual debt[23]. A paper's references indicate the work that it is built upon and are also a formal acknowledgement of the prior contributions of other published researchers. This is an important and influential idea in bibliometrics because the logical reverse of the argument suggests that the most frequently cited papers contain the most influential research, justifying the use of citation counts as indicators of the impact of publications and authors. This can be seen in the citation-based Impact Factors of the Garfield-founded Institute for Scientific Information, as well as the use of citation counts for various research evaluation purposes[44,45] and to map the intellectual structure of science[12].

The Merton-inspired connection between citations and research impact has been subjected to scrutiny and criticism from different perspectives, although most identified problems can be avoided by using appropriate methods. For example, field differences in the extent to which citations are used means that citation counts should not be compared across significantly different fields[44,45]. One of the key criticisms, however, is more fundamental: the wide range of influences on citer motivations, Cronin's (this issue) main theme. The following list is a small sample of the range of known direct and indirect citer influences[46,47].

- Methodological research and review articles are more likely to be cited.
- Papers are more likely to be cited if their authors are known to the citing authors, live in the same country, write in the same language and/or inhabit the same physical office location.
- There seems to be a tendency to cite classic works as a touchstone or symbol for an idea or field.

In addition, there are also negative reasons for citing, such as to refute prior claims.

Even though there are many citer motivations, citation counts could reasonably be held as indicators of impact if intellectual influence was sufficiently strong to dominate or obscure the other motivations[43]. The lack of clear empirical evidence for this is perhaps the reason why it continues to be controversial. Moreover, field differences in research styles[38] suggest that intellectual influence varies across fields.

Despite the lack of clear evidence to show that citations are predominantly dispassionate indicators of intellectual influence, many studies have shown that citation counts, when appropriately calculated, give figures that show good agreement with peer judgements of research quality. This provides indirect evidence that citation counts could be 'measuring' research quality. An alternative explanation is that highly cited articles tend to get a disproportionate share of 'intellectual influence' citations[42]. Hence, *differences* in citation counts may significantly indicate differences in influence, particularly for highly cited articles, even if many citations are unreliable indicators of intellectual influence.

Linker motivations

Different kinds of web phenomena that could be loosely termed 'citation' have been investigated, as summarized below.

- Hyperlinks in web pages are a form of inter-document connection, as are citations. Rousseau[48] has termed them sitations but here they will be described as hyperlinks or links.
- Hyperlinks in online journal articles are more directly analogous to citations than general links; these will be called 'article hyperlinks'.
- An article may be mentioned in web pages that are not journal articles. Citations from grey literature (e.g. unpublished reports) are included here. This is a type of citation, but differs from traditional citation in that the source is not a refereed journal article or conference paper. This will be termed 'web citation'.
- Normal citations can also be found online through the reference sections of e-journal articles and in electronic copies of academic journal or conference papers.

Note that the above list includes both hyperlinks and text citations, although in some cases a text citation will also link to the cited article.

One big difference between links and citations is that the web is not a quality-controlled refereed product: in principle information can be published online by anybody. It is logical, therefore, to restrict attention to a particular type of web page or site when investigating link creation motivations. With the increasing availability of electronic journals, a logical first question is whether e-journal citer

motivations differ from traditional citer motivations. A study of Indiana University electronic journal article authors suggested a few new motivations for citing electronic sources, such as the availability of the information online and the ability to cite digital content[49]. Although Kim's study[49] used author interviews, most other investigations have made indirect inferences about link creation motivations from link typologies created by classifying links through interpreting the text surrounding the link in the source page, and by visiting the target page.

Vaughan and Shaw[31] have introduced the possibility that web citations of journal articles be used to estimate Impact Factors for journals that are not in the ISI's database. They base their argument upon evidence that web citation counts for journal articles significantly correlate with ISI citation counts in the majority of cases. Other studies have investigated why journal articles are cited in general web pages. Articles are commonly invoked in educational contexts, such as online course reading lists. This is interesting because this kind of citation is an indicator of impact and usefulness of an article, even though it is not evidence that the article is being used for further research. This type of citation raises the possibility that web citation counts may measure a different kind of impact to traditional citations and may be useful for triangulation.

Hyperlink motivation studies have had a similar rationale to electronic and non-electronic citer motivation research: to justify the use of link counts as indicators of influence. Since early findings pointed to a very wide range of uses for hyperlinks, even those in academic web pages[50], subsequent research has broadened into general investigations into why links are created and what link counts might signify[51–55]. It seems that even in academic contexts, hyperlinks are created for a very wide range of reasons that might be loosely described as informal scholarly and educational communication. It should be noted, however, that many links appear to be merely signs, or to have no real function at all[56]. Hence interpreting general academic link counts is significantly more problematic than interpreting citation counts.

A new type of data: Usage statistics

The future of bibliometrics and network approaches may be significantly different because of the possibilities introduced by digital libraries. The major journal publishers in developed nations now maintain digital libraries of their journals, selling digital subscriptions to libraries and individuals. Additionally, some institutions and research fields have developed their own online publication archives[57]. Whilst these electronic environments can deliver citation counts, they can also produce usage statistics, allowing detailed evidence to be gathered about the number of people that have viewed each article. As an example, although in the area of preprints, one can look at the physics arXive (UK mirror uk.arxiv.org). The CiteBase function documents the number of automatically identified citations to an article as well as the number of hits to the UK server. Thus 'article views' has the potential to be an additional metric that may even replace citation counts as a more direct indicator of impact. Before this could be achieved, however, there are technical and political problems to be overcome in order to get accurate and comparable usage statistics from the major journal publishers. It is not clear that this is possible but in the meantime research has started that compares citation to usage which will reveal new insights into the value and interpretation of citation metrics[58]. It would not be difficult to incorporate these statistics into network analyses and this will be an interesting future type of research.

Conclusion

Research over the past ten years into hyperlinks and into citation and hyperlink networks is a legacy of Garfield and other key early citation analysts, whether this is explicitly acknowledged or not. This research has produced some interesting findings. Currently, however, the value of these approaches cannot compete with that of citation analysis, particularly for research evaluation. Nevertheless, they can complement citation analysis by giving a different perspective on science as a social system. For hyperlink data, its advantage is its timeliness: often research projects are flagged on the web even before any papers have been written. Nevertheless, its disadvantages are the wide range of reasons for hyperlink creation and even if the web is now becoming central to scientific communication, the same could not be claimed for hyperlinks between academic web sites. The near future promises hyperlink analysis and network theories as supporting actors to mainstream citation analysis, although this may eventually change in the future as the web and digital library technologies evolve.

1. Garfield, E., Citation indexes for science. *Science*, 1955, **122**(3159), 108–111.
2. Garfield, E., Historiographs, librarianship and the history of science. *Essays of an Inf. Sci.*, 1974, **2**, 136–150 http://www.garfield.library.upenn.edu/essays/v132p136y1974-1976.pdf.
3. Marshakova, I. V., System of document connections based on references. *Nauchno-Teknich.Inf.*, 1973, **2**, 3–8.
4. Small, H., Co-citation in the scientific literature: A new measure of the relationship between two documents. *J. Am. Soc. Inf. Sci.*, 1973, **24**(4), 265–269.
5. White, H. D. and Griffith, B. C., Author co-citation: a literature measure of intellectual structure. *J. Am. Soc. Inf. Sci.*, 1982, **32**(3), 163–172.
6. Arunachalam, S. and Doss, M. J., Mapping international collaboration in science in Asia through coauthorship analysis. *Curr. Sci.*, 2000, **79**(5), 621–628.

7. Cawkell, T., Visualizing citation connections. In *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield* (ed. Cronin, B.), Information Today Inc., Medford, NJ, 2000, pp. 177–194.
8. Leydesdorff, L., Words and co-words as indicators of intellectual organization. *Res. Policy,* 1989, **18**, 209-223.
9. Chen, C., Visualising semantic spaces and author co-citation networks in digital libraries. *Inf. Proc. Manage.*, 1999, **35**(3), 401–420.
10. Börner, K., Chen, C. M. and Boyack, K. W., Visualizing knowledge domains. *Annu. Rev. Inf. Sci. Technol.,* 2003, **37**, 179–255.
11. Garfield, E., Pudovkin, A. I. and Istomin, V. I., Mapping the output of topical searches in the *Web of Knowledge* and the case of Watson–Crick. *Inf. Technol. Libr.*, 2003, **22**(4), 183–187.
12. Small, H., Visualising science through citation mapping. *J. Am. Soc. Inf. Sci.,* 1999, **50**(9), 799–813.
13. Scharnhorst, A., Evolution in adaptive landscapes – examples of science and technology development. In *Collaboration in Science* (eds Havemann, F. and Kretschmer, H.), Gesellschaft für Wissenschaftsforschung, Berlin, 2002, pp. 118–142.
14. Davidson, G. S., Hendrickson, B., Johnson, D. K., Meyers, C. E. and Wylie, B. N., Knowledge mining with VxInsight: discovery through interaction. *J. Intell. Inf. Syst.*, 1998, **11**(3), 259–285.
15. Marion, L. S., Garfield, E., Hargens, L. L., Lievrouw, L. A., White, H. D. and Wilson, C. S., Social network analysis and citation network analysis: Complementary approaches to the study of scientific communication. ASIST 2003: Proceedings of the 66th ASIST Annual Meeting, 2003, vol. 40, pp. 486–487.
16. Barabási, A. L., *Linked: The New Science of Networks*, Perseus Publishing, Cambridge, 2002.
17. Huberman, B. A., *The Laws of the Web: Patterns in the Ecology of Information*, The MIT Press, Cambridge, 2001.
18. Scharnhorst, A., Complex networks and the web: Insights from nonlinear physics. *J. Comput. Med. Commun.*, 2003, **8**(4), http://www.ascusc.org/jcmc/vol8/issue4/scharnhorst.html
19. Newman, M. E. J., The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA*, 2001, **98**(2), 404–409.
20. Barabási, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., Vicsek, T., Evolution of the social network of scientific collaborations. *Physica*, 2002, **A311**(3-4), 590–614, http://arxiv.org/abs/cond-mat/0104162/
21. Li, M., Wu, J., Wang, D., Zhou, T., Di, Z. and Fan, Y., Evolving model of weighted networks inspired by scientific collaboration networks. arXiv:cond-mat/0501655v2, 2005, http://arxiv.org/abs/cond-mat/0501655
22. Price, D. d. S., A general theory of bibliometric and other cumulative advantage processes. *J. Am. Soc. Inf. Sci.,* 1976, **27**, 292–306.
23. Merton, R. K., *The Sociology of Science. Theoretical and Empirical Investigations*, University of Chicago Press, Chicago, 1973.
24. Bonitz, M., Bruckner, E. and Scharnhorst, A., Characteristics and impact of the Matthew effect for countries. *Scientometrics,* 1997, **40**(3), 407–422.
25. Leydesdorff, L., Theories of citation. *Scientometrics,* 1998, **43**(1), 5–25.
26. Dorogovtsev, S. N. and Mendes, J. F. F., Evolution of networks. *Adv. Phys.,* 2002, **51**, 1079–1187.
27. Björneborn, L., Small-world link structures across an academic web space: a library and information science approach, Ph D dissertation, Royal School of Library and Information Science, Copenhagen, 2004, available at the URL http://www.db.dk/lb/phd/phd-thesis.pdf.
28. Cronin, B., Bibliometrics and beyond: some thoughts on web-based citation analysis. *J. Inf. Sci.,* 2001, **27**(1), 1–7.
29. Ingwersen, P., The calculation of web impact factors. *J. Document.,* 1998, **54**(2), 236–243.
30. Vaughan, L. and Hysen, K., Relationship between links to journal web sites and impact factors. *ASLIB Proc.,* 2002, **54**(6), 356–361.
31. Vaughan, L. and Shaw, D., Bibliographic and web citations: What is the difference?. *J. Am. Soc. Inf. Sci. Technol.*, 2003, **54**(14), 1313–1322.
32. Vaughan, L. and Wu, G., Links to commercial websites as a source of business information. *Scientometrics*, 2004, **60**(3), 487–496.
33. Heimeriks, G., Hörlesberger, M. and van den Besselaar, P., Mapping communication and collaboration in heterogeneous research networks. *Scientometrics*, 2003, **58**(2), 391–413.
34. Leydesdorff, L. and Scharnhorst, A., Measuring the knowledge base: a program of innovation studies. Report to the Bundesministerium für Bildung und Forschung. Berlin-Brandenburgische Akademie der Wissenschaften. 2002, http://www.sciencepolicy-studies.de/dok/expertise-leydesdorff-scharnhorst.pdf
35. Leydesdorff, L. and Curran, M., Mapping university–industry–government relations on the Internet: the construction of indicators for a knowledge-based economy. *Cybermetrics*, 2000, **4**, http://www.cindoc.csic.es/cybermetrics/articles/v4i1p2.html.
36. Beaulieu, A., From brainbank to database: the informational turn in the study of the brain. *Studies History Philos. Biol. Biomed. Sci.*, 2004, **35**(2), 367–390.
37. Beaulieu, A., Sociable hyperlinks: an ethnographic approach to connectivity. In *Virtual Methods – Issues in Social Research on the Internet* (ed. Hine, C.) Berg, Oxford, 2005, pp. 183–197.
38. Fry, J. and Talja, S., *The cultural shaping of scholarly communication: Explaining e-journal use within and across academic fields.* Paper presented at the ASIST 2004: Proceedings of the 67th ASIST Annual Meeting, 2004, pp. 20–30.
39. Thelwall, M., Evidence for the existence of geographic trends in university web site interlinking. *J. Document.*, 2002, **58**(5), 563-574.
40. Vasileiadou, E. and Besselaar, P. van den, Linking shallow, linking deep – How scientific intermediaries use the web for their network of collaborators. *Cybermetrics* (in prep.), 2005.
41. Fry, J., Understanding disciplinary differential on the web. *Cybermetrics* (in prep.), 2005.
42. van Raan, A. F. J., In matters of quantitative studies of science the fault of theorists is offering too little and asking too much. *Scientometrics*, 1998, **43**(1), 129–148.
43. Cronin, B., A hundred million acts of whimsy?, *Curr. Sci.*, 2005, **89**, 1505–1509.
44. Moed, H. F., The impact-factors debate: the ISI's uses and limits. *Nature,* 2002, **415**, 731–732.
45. van Raan, A. F. J., The Pandora's box of citation analysis: Measuring scientific excellence – The last evil? In *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield* (eds Cronin, B. and Atkins, H. B.), ASIS Monograph Series, Information Today, Inc, Medford, 2000, pp. 301–319.
46. Borgman, C. L. and Furner, J., Scholarly communication and bibliometrics. *Annu. Rev. Inf. Sci. Technol.*, 2002, **36**, 3–72.
47. Cronin, B., *The Citation Process: The Role and Significance of Citations in Scientific Communication*, Taylor Graham, London, 1984.
48. Rousseau, R., Sitations: an exploratory study. *Cybermetrics,* 1997, **1**(1), http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html.
49. Kim, H. J., Motivations for hyperlinking in scholarly electronic articles: A qualitative study. *J. Am. Soc. Inf. Sci.*, 2000, **51**(10), 887–899.
50. Smith, A. G., A tale of two web spaces; comparing sites using web impact factors. *J. Document.*, 1999, **55**(5), 577–592.
51. Bar-Ilan, J., A microscopic link analysis of academic institutions within a country – the case of Israel. *Scientometrics*, 2004a, **59**(3), 391–403.
52. Bar-Ilan, J., Self-linking and self-linked rates of academic institutions on the web. *Scientometrics,* 2004b, **59**(1), 29–41.
53. Bar-Ilan, J., What do we know about links and linking? A framework for studying links in academic environments. *Inf. Proc. Manage.*, 2005, **41**(3), 973–986.

54. Harries, G., Wilkinson, D., Price, E., Fairclough, R. and Thelwall, M., Hyperlinks as a data source for science mapping. *J. Inf. Sci.*, 2004, **30**(5), 436–447.
55. Wilkinson, D., Harries, G., Thelwall, M. and Price, E., Motivations for academic Web site interlinking: Evidence for the Web as a novel source of information on informal scholarly communication. *J. Inf. Sci.*, 2003, **29**(1), 49–56.
56. Thelwall, M., What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation. *Inf. Res.*, 2003, **8**(3), http://informationr.net/ir/8-3/paper151.html.
57. Harnard, S. and Carr, L., Integrating, navigating, and analysing open eprint archives through open citation linking (the OpCit project). *Curr. Sci.*, 2000, **79**(5), 629–638.
58. Moed, H. F., Statistical relationships between downloads and citations at the level of individual documents within a single journal. *J. Am. Soc. Inf. Sci. Technol.*, 2005, **56**(10), 1088–1097.

ACKNOWLEDGEMENTS. This work was supported by a grant from the Common Basis for Science, Technology and Innovation Indicators part of the Improving Human Research Potential specific programme of the Fifth Framework for Research ad Technological Development of the European Commission. It is part of the WISER project (web indicators for scientific, technological and innovation research) (Contract HPV2–CT–2002–00015) (www.webindicators.org).

Quotations by Eugene Garfield

'However, I would like to add this personal philosophical note. I did not start my career with the aim of becoming wealthy. I was born with the "information consciousness" gene in me. To this day, I cannot resist writing friends to tell them they have been cited in a recent obscure publication. When I heard an author recently speak on his biography of Benjamin Franklin I sent him dozens of citations to his work he had never known about. I relished the role I played in informing thousands of authors that they had published a Citation Classic. I might have remained simply a researcher and communicator but that was not to be. I tried in vain to convince several non-profit and governmental organizations to take up the challenge of citation indexing but they adamantly refused to consider the idea. So I was forced to choose the private entrepreneurial route in order to achieve my goal. Were it not for the initial financial success of another harebrained idea, Current Contents, the Science Citation Index would not have seen the light of day.'

— Garfield
Some personal recollections on the occasion of receiving the Kaula Award for 2000,
Washington DC, 3 April 2003
http://garfield.library.upenn.edu/papers/kaulaaward.html

'It would be more relevant to use the actual impact (citation frequency) of individual papers in evaluating the work of individual scientists rather than using the journal impact factor as a surrogate.'

— Garfield
Nature, 2001, **411**, 522

'If I had my way, there would be a government WPA project to go back and create a Citation Index for all the literature that was ever published.'

— Garfield
Chemical Herditage Foundation Oral History, 1997

'The problem with Current Contents is that it is so simple and utilitarian that it gives theoreticians little to talk about. Current awareness is one thing – information retrieval is something else. It is somewhat telling that when I taught at the Moore School of Engineering at Penn in the 60s, the engineers called CC an information retrieval tool.'

— Garfield
Conference on the History and Heritage of Science Information Systems, Pittsburgh, PA, 1998