

This is a pre print version of the following article:

La citazione bibliografica nell'epoca della sua riproducibilità tecnica. Bibliometria e analisi delle citazioni dallo Science Citation Index alla Cybermetrica [Dissertazione] / De Bellis, Nicola. - (2005).

Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

16/08/2024 11:37

(Article begins on next page)

ASSEGNATO IL PREMIO “BIBLIOTECHE OGGI”**Data:** Thursday, 06 April @ 08:38:34 CEST**Argomento:** Recensioni e segnalazioni

Biblioteche oggi News 4/2006

A conclusione del Convegno “La biblioteca su misura”, Venerdì 10 marzo 2006, ha avuto luogo presso il Palazzo delle Stelline la proclamazione del vincitore del Premio “Biblioteche oggi”, bandito dall’Editrice Bibliografica in occasione del ventesimo anniversario della rivista al fine di promuovere lo studio nell’ambito della biblioteconomia e delle scienze dell’informazione.

Pubblichiamo di seguito il testo letto in occasione della proclamazione del vincitore.

La Giuria del Premio “Biblioteche oggi”, composta da Massimo Belotti, Luigi Crocetti, Luca Ferrieri, Carlo Revelli, Michele Santoro, Giuliano Vignini, dopo aver fissato i criteri di valutazione ed aver esaminato i lavori pervenuti, è giunta alle seguenti conclusioni.

Sono pervenuti 16 lavori, che nel loro insieme hanno evidenziato un tendenziale interesse per gli aspetti innovativi che negli ultimi anni hanno riguardato la vita delle biblioteche. Ciò è dimostrato dalla presenza di un gruppo consistente di opere, volte a sviluppare i temi legati alla gestione dei servizi, esaminati alla luce delle opportunità offerte dalle tecnologie dell’informazione, specchio tanto del dibattito in corso nella letteratura professionale, quanto delle esperienze in atto nelle biblioteche. Inoltre, è stato rilevato come molti lavori si siano orientati su studi di casi o iniziative e locali: questa scelta, se da un lato ha permesso di introdurre elementi di concretezza come frutto di specifiche esperienze sul campo, in qualche caso può aver limitato gli aspetti legati alla ricerca e all’approfondimento critico.

Pur nell’apprezzamento generale sui lavori pervenuti, in qualche caso assolutamente pregevoli, la Giuria ha rilevato in molti lavori (trattandosi perlopiù di tesi di laurea e di diploma) la mancanza di un opportuno aggiornamento bibliografico e disciplinare.

Ciò premesso è stata individuata una rosa di lavori particolarmente meritevoli.

Sulla base di un’approfondita analisi di questi lavori, la Giuria – riunita a Milano nella sessione conclusiva il 30 gennaio 2006 – ha proceduto alla designazione del vincitore e all’assegnazione di alcune menzioni speciali.

Si è così deciso di riconoscere una particolare menzione a:

Chiara Faggiolo, autrice di Esperienze e prospettive dei servizi di reference digitale nelle biblioteche di pubblica lettura, Tesi di laurea in Conservazione dei beni culturali, Università degli studi di Parma, con la seguente motivazione: “Per l’accuratezza dell’indagine e l’efficace documentazione di un tema di particolare attualità”.

Rossella Naldi, autrice di Midrash e teorie moderne della lettura: la pluralità del significato, Tesi di laurea in didattica dell’italiano, Università degli studi di Milano, con la seguente motivazione: “Per l’originalità e il contributo dato ai temi della teoria della lettura”.

Roberto Raieli, autore di La problematica del Visual Retrieval, Tesi di diploma per bibliotecari presso la Scuola per archivisti e bibliotecari dell’Università “La Sapienza” di Roma, con la seguente motivazione: “Per l’approfondita trattazione di un argomento di rilevante interesse nell’odierno contesto bibliotecario, specialmente trattandosi di un campo di ricerca largamente sperimentale”.

La Giuria ha quindi deciso all’unanimità di dichiarare vincitore del Premio “Biblioteche oggi”

Nicola De Bellis, autore del saggio La citazione bibliografica nell’epoca della sua riproducibilità tecnica: bibliometria e analisi delle citazioni dallo Science Citation Index alla Cybermetrica, con la seguente

motivazione:

“Ampio e dettagliato studio su un tema non molto presente nella letteratura professionale italiana. A una approfondita conoscenza dello sviluppo storico dell'argomento affrontato, l'autore unisce una sicurezza nello stile espositivo, a partire dall'introduzione che imposta bene il discorso successivo. Nel lavoro sono presi in considerazione i vari aspetti della bibliometria, anche al di là dello specifico aspetto della misurazione delle citazioni, mentre si aprono collegamenti interessanti con altre problematiche, quali il significato della peer review, la libertà di espressione, l'attenuazione del concetto di autore. L'opera inoltre prende in esame le problematiche più recenti, quali l'open access e i problemi di citazione del web, compreso il confronto tra citazione e download”.

Il vincitore, Nicola De Bellis, è bibliotecario presso la Biblioteca Universitaria – Area medica, Università degli studi di Modena e Reggio Emilia.

Questo Articolo proviene da Biblioteche Ravenna

<http://193.207.49.2/virtual/biblio/>

L'URL per questa storia è:

<http://193.207.49.2/virtual/biblio//article.php?sid=2839>

La citazione bibliografica nell'epoca della sua riproducibilità tecnica

- Bibliometria e analisi delle citazioni dallo Science
Citation Index alla Cybermetrica -

Nicola De Bellis (*)

[Ultima revisione: 31/05/2005]

(*) BU Area Medica, Università degli Studi di Modena

Tel.: 0039 - 059 4223140 (debellis.nicola@unimore.it)

“Veniamo ora alla citazione degli autori, che negli altri libri si trovano e nel vostro invece mancano. Il rimedio è facile: voi non dovete fare altro che pigliare un libro che li citi tutti dall'A alla Z, come dite voi, e riprodurre questo elenco alfabetico nel vostro libro. Anche se si vedrà chiaramente l'inganno per la poca utilità che potevate avere a servirvene, non importa nulla: chissà che non vi sia qualcuno tanto ingenuo che creda davvero che voi li abbiate consultati tutti nella vostra storia così semplice e alla buona; e quando quel lungo catalogo d'autori non giovasse ad altro, gioverà per lo meno a dar subito di primo colpo autorità al libro. E poi, o chi volete che, non avendoci interesse, si metta a verificare se ve ne siete o non ve ne siete servito?” (Miguel De Cervantes, Don Chisciotte della Mancia, Prologo).

Indice

<i>Indice</i>	2
<i>Nota</i>	5
<i>Ringraziamenti</i>	5
<i>Introduzione</i>	6
<i>Capitolo 1 - Contare la scienza: storia breve degli studi bibliometrici</i>	22
1.1. La bibliografia statistica	24
1.2. Le leggi "troppo universali" della bibliometria	27
1.2.1. La legge di Lotka.....	28
1.2.2. La legge di Bradford e la legge di concentrazione di Garfield	31
1.2.3. La legge di Zipf.....	34
1.2.4. L'unificazione delle leggi della bibliometria	35
1.3. Dalla bibliografia statistica alla biblio(sciento)-metria	38
<i>Capitolo 2 - Le basi teoriche della scientometria</i>	47
2.1. Price e la scienza della scienza	49
2.2. Henry Small, Robert Merton e le funzioni normative della citazione bibliografica .	55
2.2.1. Il documento-simbolo: Henry Small	57
2.2.2. La citazione come ricompensa sociale: Robert Merton.....	61
<i>Capitolo 3 - Ricerche "senza soggetto": lo Science Citation Index come strumento di information retrieval</i>	66
3.1. Ideazione e realizzazione dello Science Citation Index	68
3.2. La ricerca bibliografica nello Science Citation Index	77

3.3. L'enciclopedia come impresa commerciale: ISI Web of Knowledge	85
--	----

Capitolo 4 - Mappe e paradigmi: la citazione bibliografica al servizio della storia e della sociologia della scienza 92

4.1. La storiografia algoritmica e la mappa delle influenze: HistCite.....	94
--	----

4.2. La struttura disciplinare della scienza: l'analisi delle co-citazioni	102
--	-----

Capitolo 5 - Impact Factor e valutazione degli scienziati: la citazione bibliografica al servizio della politica della scienza..... 112

5.1. La scorciatoia: l'Impact Factor delle riviste.....	116
---	-----

5.2. La via lunga: costruzione e applicazione degli indicatori scientometrici.....	123
--	-----

5.2.1. Il conteggio delle citazioni al livello dell'individuo.....	128
--	-----

5.2.2. Il conteggio delle citazioni al livello dell'istituzione	131
---	-----

5.3. Le citazioni dei brevetti tra scienza, tecnologia e giurisprudenza	133
---	-----

Capitolo 6 - Misurare la comunicazione scientifica nel XXI secolo: dalla bibliometria alla webmetrica..... 142

6.1. La citazione negli e-journals e negli archivi aperti	143
---	-----

6.1.1. Citazioni e Open Access.....	144
-------------------------------------	-----

6.1.2. CiteBase, OACI.....	149
----------------------------	-----

6.1.3. CiteSeer: Dallo Science Citation Index al Web Citation Index	151
---	-----

6.2. La citazione come hyperlink e le metriche del Web	154
--	-----

6.2.1. Webmetrica	156
-------------------------	-----

6.2.2. Analisi delle reti di hyperlink (Hyperlink Network Analysis).....	164
--	-----

6.2.3. Analisi delle reti complesse (Complex Network Analysis)	167
--	-----

Capitolo 7 - La citazione come artificio retorico e le critiche al modello normativo 172

7.1. Motivi per citare.....	173
-----------------------------	-----

7.2. Vita e morte dell'autore scientifico.....	185
--	-----

<i>Conclusioni</i>	<i>191</i>
<i>Bibliografia</i>	<i>199</i>

Nota

La traduzione italiana dei passi di opere straniere riportati nel testo, salvo diversa indicazione, è mia.

Le indicazioni bibliografiche in nota a piè di pagina contengono l'informazione completa relativa ai testi citati fatta eccezione per gli articoli estratti dagli *Essays of an Information Scientist* di Eugene Garfield, per i quali l'URL di riferimento viene fornito solo nella bibliografia finale.

Tutti i link alle pagine web e agli articoli in formato elettronico sono stati verificati l'ultima volta il 25/05/2005.

I nomi dei database e dei software ISI citati sono marchi dei rispettivi produttori soggetti a copyright.

Ringraziamenti

Si ringraziano i colleghi delle biblioteche italiane che hanno contribuito, mediante il servizio di document delivery, a recuperare in modo rapido ed efficiente il materiale necessario alla stesura del lavoro. Un ringraziamento particolare è rivolto a: Sonia Campese (Biblioteca DEIS, Università di Bologna), Vincenza Catalano (Dipartimento di Storia Antica, Università di Bologna), Andrea Lodi (BU Area Giuridica, Università di Modena), Stefania Olimpieri (BSI, Università di Modena), Carmela Palazzi (BU Area Medica, Università di Modena), Anita Ungari (Biblioteca Dipartimento di Informatica, Università di Bari).

Si ringrazia inoltre il collega, amico e "maestro" Viktor Taransky per la costanza e lo spirito critico con cui ha seguito le diverse fasi di compilazione del testo. A lui è dedicato il premio di cui l'opera è stata insignita il 10 marzo 2006.

Introduzione

Un dizionario lessicale può essere usato, alla fine di un itinerario di ricerca, per fare un gioco: si prende la parola o le parole che condensano meglio di altre l'argomento trattato e si vede se, e in che misura, per una qualche astuzia della lingua, le conclusioni raggiunte hanno una correlazione magari inconsapevole o imprevedibile con le accezioni elencate.

Alla voce 'citazione' un dizionario della lingua italiana elenca di solito quattro definizioni principali:¹ 1) un passo o parole di altri riportati in un discorso o in uno scritto («una citazione dalla *Terra desolata* di Eliot...»); 2) un riferimento bibliografico preciso ad una fonte, ad un testo («le citazioni a piè di pagina o alla fine del libro, dell'articolo...»); 3) la menzione di una persona per meriti particolari («guadagnarsi una citazione al merito ...»); 4) un documento con il quale si “cita in giudizio”, ordinando a qualcuno di comparire davanti all'autorità giudiziaria («ricevere una citazione dal tribunale ...»).

I vari significati hanno una matrice etimologica comune nel latino 'citare', forma iterativa di 'ciēre', che significa mettere in movimento, muovere fortemente, eccitare, ma anche chiamare a comparire, invitare, invocare, convocare (una persona, un testo, una divinità).

L'ultima accezione allude ad un rapporto di potere tra un individuo ed un'istituzione designata a giudicarne i comportamenti. Alla base della convocazione c'è il sospetto della violazione di una o più norme giuridiche. Anche il terzo significato implica un giudizio di valore su un individuo da parte di un'entità collettiva, stavolta però connotato in senso nettamente positivo. Si cita al merito il militare che ha dato prova di coraggio e fedeltà e la citazione, come nel caso precedente, può avere effetti immediati sulla sua situazione esistenziale (una promozione, un incarico più prestigioso). La

prima e la seconda accezione designano l'atto e l'oggetto di un prelievo e di un innesto confinati nella dimensione linguistica. Ciò che si mette in movimento sono frammenti di un testo o di un discorso, nel primo caso trapiantati in un altro testo o discorso, nel secondo sostituiti da una registrazione convenzionale degli estremi bibliografici della fonte da cui provengono. Le due operazioni possono coesistere: si cita testualmente la fonte e se ne danno subito dopo le coordinate bibliografiche. Basta sfogliare una qualunque rivista scientifica per rendersi conto però che, in questa forma peculiare di comunicazione scritta, la segnalazione bibliografica abbreviata dei documenti citati domina il palcoscenico, anzi il backstage delle note relegate a fondo pagina. Della citazione nell'altro senso non c'è traccia, sacrificata ai requisiti di sobrietà espositiva dei manuali di stile, ma l'incontro del lettore con la fonte di cui si danno le coordinate, in linea di principio, è solo differito ("leggi questo articolo e troverai la frase che giustifica le mie affermazioni").

Il dizionario avverte a questo punto che, per estensione, si parla di citazione anche a proposito della ripresa e del riecheggiamento, in un'opera letteraria, musicale o cinematografica, di motivi, luoghi, situazioni di opere precedenti. Un'estensione dettata dal buon senso, dalla familiarità con le sfide più o meno amatoriali a indovinare somiglianze e allusioni, che sottintende il carattere sociale intrinseco dei prodotti culturali: l'individualità di una creazione artistica, scientifica, tecnologica si precisa sempre sullo sfondo di modelli, stereotipi, topoi già esistenti, dei quali la critica e la storiografia cercano di stabilire modalità e forme di sopravvivenza in luoghi e tempi diversi. La musica classica, per esempio, è stata per secoli terreno di sperimentazione su idee, temi, frammenti, impasti sonori dall'origine più svariata e la pratica della citazione musicale ha raggiunto, in compositori come Schumann, una stabilità tale da erigerla a strumento consapevole di dichiarazione poetica.² Ancora di

¹ Vedi, ad esempio, la voce *Citazione* in: *De Mauro: Dizionario della lingua italiana su cd-rom*, versione 1.0.3.5 (Torino: Paravia, 2000).

² A. Malvano, *Voci da lontano. Schumann e l'arte della citazione* (Torino: EDT, 2003).

più, il jazz ha costruito la propria identità attorno alla dinamica ripetizione-variazione e alla celebrazione gioiosa dell'influenza, mentre la musica leggera ha esasperato a tal punto il riciclo di combinazioni melodiche formate da sequenze omologate di accordi e giri armonici da rendere spesso difficoltoso, alla luce delle norme giuridiche di tutela della proprietà intellettuale, l'accertamento dell'originalità dei singoli pezzi.³

Al di là della trasmissione di idee e motivi quale parte integrante del processo culturale, il culto per il richiamo, il rimando, l'allusione al già visto e al già sentito, è un elemento ricorrente nel panorama artistico contemporaneo.

Considerato uno dei sintomi più evidenti della sensibilità postmoderna, il gusto per la citazione è divenuto il segno di una vocazione comunicativa che punta a restituire, attraverso la ripetizione e la coesistenza, negli stessi testi, di elementi del passato e del presente, un'immagine sincronica di ogni storia e di ogni possibile memoria: non più quindi citazione come sperimentazione di nuove forme poetiche e narrative, ma citazione come rinuncia all'originalità, condanna dell'autore e del lettore al gioco ironico e intertestuale, alla ripetizione e riattualizzazione ossessiva di codici e materiali del passato nella consapevolezza di non poter dire altro se non ciò che è stato già detto.⁴

Lo scopo del presente lavoro è indagare la storia, le premesse teoriche e gli sviluppi più recenti degli studi bibliometrici fondati sull'analisi e trattamento quantitativo delle citazioni bibliografiche. Nelle pagine che seguono pertanto si parlerà quasi esclusivamente della citazione nel secondo senso, cioè come riferimento bibliografico preciso ad una fonte, ma non si potrà fare a meno di osservare che è impossibile seguire le peripezie della citazione bibliografica

³ M. Bovi, *Anche Mozart copiava. Cover, somiglianze, plagie e cloni* (Milano: Auditorium, 2004).

⁴ La citazione è uno dei temi centrali della riflessione estetica contemporanea, in quanto rappresenta il crocevia per alcune questioni fondamentali, quali il concetto di intertestualità, il rapporto dell'autore con la tradizione letteraria, il ruolo del lettore/spettatore/critico. Un'introduzione chiara a tali problematiche si trova in: U. Eco, *Sulla letteratura* (Milano: Bompiani, 2002), in particolare pp. 128-146 e 227-252; V. Fortunati, "Citazione" *Quaderni di Filologia Germanica della Facoltà di Lettere e Filosofia dell'Università di Bologna* 1988, 4: 7-13; V. Fortunati, "Intertestualità e citazione fra Modernismo e Postmodernismo" *Leitmotiv* 2002, 2: 87-96 <<http://www.ledonline.it/leitmotiv/allegati/leitmotiv020208.pdf>>. Sul ruolo e la tipologia delle

all'interno della letteratura bibliometrica senza un continuo rimando alle altre dimensioni implicite nelle scarse definizioni del dizionario: quella giuridico-normativa (citare rispettando delle regole di condotta o violandole), quella dei meriti socialmente riconosciuti (citare per ricompensare innalzando lo status di colui che viene citato) e quella del rapporto tra i testi, i loro autori, i lettori, la tradizione in cui si collocano (la citazione segno di un'influenza intellettuale o della sua dissimulazione).

Citare un documento in bibliografia o in nota a margine di un testo è un'operazione tutto sommato ordinaria e di immediata comprensione. In un saggio scientifico, a differenza che in un romanzo o in un articolo di quotidiano, si cita e si deve citare - nella peggiore delle ipotesi si deve fingere di farlo - per pagare un debito intellettuale. Quando assolve alla sua funzione più ovvia, la citazione bibliografica rivela un'identità al di sopra di ogni sospetto: è la moneta corrente nel commercio della comunicazione scientifica ufficiale. Moneta di piccolo taglio (costa poco citare), ma dal potere d'acquisto simbolico non indifferente. Attraverso la citazione infatti l'autore convoca a rapporto gli altri autori che, prima di lui, si sono occupati dello stesso o di argomenti simili oppure hanno fornito metodi e concetti basilari per l'indagine avviata. La loro presenza è una componente essenziale del dispositivo retorico del testo. Eccoli in fila, chiamati a rinforzare, a sfidare - in una battaglia vinta prima ancora di combatterla - o semplicemente ad ampliare il punto di vista dello scrittore, limitato dal tempo, dalla quantità di letteratura accumulata, da una giusta dose di modestia istituzionalizzata. Il dialogo è scandito da titoli di documenti, evidenziati con caratteri tipografici particolari, relegati nelle note a piè di pagina o alla fine del testo, talvolta introdotti da locuzioni dal significato ambiguo (*cfr.*, *vedi*, *vedi anche*), talvolta supportati da un testo di commento che rammenta il motivo e le circostanze della convocazione. Le regole del dialogo non sono scritte da nessuna parte: proutari e manuali di

stile guidano i neofiti nella compilazione degli apparati bibliografici,⁵ ma è impossibile usarli per ricostruire le pratiche citazionali ed i significati che l'autore ha attribuito ai riferimenti bibliografici.

Nelle discipline storico-filologiche l'accuratezza dei riferimenti bibliografici e la padronanza di uno stile di citazione appropriato certificano l'acquisizione di una professionalità e legittimano l'appartenenza dell'autore ad una corporazione di specialisti. Le note a piè di pagina, per molto tempo habitat naturale delle citazioni bibliografiche e filtro imposto dallo scrittore per l'interpretazione del loro significato, sono per lo storico pressapoco l'equivalente delle informazioni sui dati sperimentali nel campo delle scienze naturali e, al pari di quelli, hanno una funzione inscindibilmente cognitiva e sociale:⁶ presentano il fondamento empirico delle storie raccontate e delle argomentazioni proposte; al tempo stesso conferiscono autorità e rispettabilità al testo in quanto dimostrano che, mediante l'esame delle fonti primarie (i dati da interpretare) e secondarie (le interpretazioni altrui da confermare o smentire), le regole del gioco sono state rispettate,⁷ il candidato è adatto a riprodurre, nell'ambito di cariche accademiche, associazioni professionali, comitati di *referee*, congressi e così via, discorsi e comportamenti su cui si fonda la sopravvivenza istituzionale della disciplina.

Quanto il rispetto delle regole sia importante per l'identità di una disciplina emerge con chiarezza negli episodi di violazione conclamata. Esemplare in questo senso la reazione di uno dei capostipiti della filologia classica tedesca,

postmoderno (Roma: Bulzoni, 1996), in particolare pp. 87-103.

⁵ Ad esempio: U. Eco, *Come si fa una tesi di laurea. Le materie umanistiche*, 12 ed. (Milano: Bompiani, 2001); C. Revelli, *Citazione bibliografica* (Roma: Associazione Italiana Biblioteche, 2002).

⁶ Sulla storia delle note a piè di pagina i testi di riferimento restano: A. Grafton, *La nota a piè di pagina: una storia curiosa* (Milano: Sylvestre Bonnard, 2000); A. Grafton, "The footnote from De Thou to Ranke" *History and Theory* 1994, Theme Issue 33(4): 53-76.

⁷ A. Momigliano, "Le regole del gioco nello studio della storia antica," in *Sui fondamenti della storia antica* (Torino: Einaudi, 1984): 477-86. Peter Rieß, che ha tracciato le linee generali di una teoria sistematica della nota a piè di pagina (*footnotology*), scrive: «The way in which the footnote thrives and flourishes in the reality of scholarly literature shows that footnotes are here to stay. From this we can deduce that the significance of the footnote lies in the footnote itself. Scholarly publications need footnotes in order to acquire authority and respectability»: P. Riess, *Towards a Theory of the Footnote* (Berlin: De Gruyter, 1984), p. 22.

Ulrich von Wilamowitz, alla pubblicazione de *La nascita della tragedia* (1872) di Friedrich Nietzsche, all'epoca docente di filologia classica presso l'Università di Basilea.⁸ In un testo senza note a piè di pagina e citazioni dal greco, pieno in compenso di concetti metafisici e frammenti di Schopenhauer, con dieci capitoli dedicati all'imminente rinascita dello spirito dionisiaco della tragedia greca nelle opere di Richard Wagner, Wilamowitz leggeva la negazione dei principi basilari di scientificità e delle regole di condotta della ricerca storica, che doveva svolgersi, all'opposto di quanto faceva Nietzsche, in modo da studiare i fenomeni del passato alla luce dei presupposti e delle concezioni del periodo in cui si erano manifestati. Per questo motivo egli si augurava l'esclusione del concorrente dall'insegnamento accademico, cosa che puntualmente avvenne sette anni dopo. Negli anni seguenti, filologi tedeschi, inglesi, americani e italiani avrebbero affollato le note a piè di pagina dei propri saggi con riferimenti eruditi e spesso del tutto accessori a Wilamowitz, pagando debiti intellettuali equivalenti alla riscossione di altrettanti crediti in termini di approvazione e riconoscimento professionale da parte della comunità dei classicisti.

Non sempre il trucco è facile da scoprire, solo chi ha frequentato gli stessi archivi e maneggiato le stesse fonti può accorgersene. E c'è comunque la possibilità di barare. Leopold von Ranke, uno dei padri fondatori della storiografia moderna, professava il culto per la ricerca d'archivio e l'esame diretto delle fonti primarie, ma il manoscritto della sua *Storia dei popoli romanici e germanici* (1824) desta il sospetto che egli avesse aggiunto le citazioni bibliografiche tutte alla fine, e che spulciando «tra i libri e le note, tra gli excerpta e i sommari usò la saliera per aggiungere un ingrediente a uno stufato già bell'e fatto».⁹

⁸ La vicenda è ricostruita, con un'attenzione specifica per il ruolo delle note a piè di pagina, in: S. Nimis, "Fussnoten: das Fundament der Wissenschaft" *Arethusa* 1984, 17: 105-34.

⁹A. Grafton, *La nota a piè di pagina...cit.*, p. 64.

La nota bibliografica ha un ruolo, benché negativo, anche quando chi si proclama estraneo a tutte le scuole le fa pesare l'appartenenza ad un modello storiografico malato di erudizione: «le note *up to date* - scrive Georges Dumézil - allineano a volte dieci, venti riferimenti e anche più, a guisa di code bibliografiche che, non essendo nè classificatorie nè critiche, non garantiscono nemmeno che l'autore abbia fatto ricorso a tutto ciò che cita, e che ingombrano la metà inferiore delle pagine, simili ai grandi scarichi che rendono poco piacevoli i dintorni di certe città».¹⁰

Nelle scienze naturali, dove l'unità di analisi non è un testo ma un set di variabili collegate a fenomeni osservabili o, in linea di principio, riproducibili da chiunque in determinate condizioni sperimentali, l'enfasi sulla correttezza delle pratiche citazionali non è affatto inferiore. La capacità di citare le fonti giuste al posto giusto contribuisce a rinforzare la credibilità dell'autore come esperto di una materia, mentre l'incapacità di farlo può determinarne l'insuccesso professionale.¹¹

C'è stato però un momento nella storia intellettuale del '900 in cui questo ruolo strategico della citazione bibliografica, fino ad allora confinato nella ristretta cerchia degli addetti ai lavori, ha ricevuto un'attenzione inedita dal mondo dei profani, da amministratori, politici, sociologi e storici della scienza, e la citazione ha cominciato a vivere una vita propria sganciata da quella dei documenti che costituivano la sua naturale dimora. Ciò è accaduto quando, all'interno della tradizione di studi sull'*information retrieval*, ha preso consistenza il potenziale conoscitivo degli indici di citazioni, o almeno del più importante mai realizzato, lo *Science Citation Index*, ideato agli inizi degli anni '60 da Eugene Garfield (1925-), fondatore dell'Institute for Scientific Information (ISI) di Philadelphia.

¹⁰ G. Dumezil, *Feste romane* (Genova: Il Melangolo, 1989), p. 13.

¹¹ C. Berkencotter & T.N. Huckin, "You are what you cite: Novelty and intertextuality in a biologist's experimental article," in *Professional Communication: The Social Perspective*, edited by N.R. Blyler, et al. (Newbury Park, CA: Sage, 1993): 107-27. Questo punto sarà approfondito nel capitolo 7.

Un indice di citazioni (*citation index*)* è un oggetto concettualmente semplice. Gli articoli di riviste scientifiche contengono di solito, nelle note collocate a piè di pagina o alla fine del testo, una lista di riferimenti bibliografici di documenti che hanno una qualche correlazione con le tesi esposte nel corpo dell'articolo. L'indice non fa altro che elencare (su carta) o registrare (nei record di un database elettronico) le pubblicazioni citate nelle bibliografie di un gruppo selezionato di documenti, identificando per ciascuna di esse la fonte della citazione, cioè il documento o i documenti nei quali compare tra i riferimenti bibliografici. Confidando nella naturale tendenza degli scienziati che condividono lo stesso dominio disciplinare a stabilire reti comunicative intessute di riferimenti (scritti) ad un comune sostrato di concetti, metodi e strumenti, Garfield pensò che la via più semplice per potenziare il recupero dell'informazione scientifica fosse quella di raccogliere e rendere consultabili, in un unico repertorio, l'insieme dei collegamenti "spontanei" che si creano tra le pubblicazioni ogniqualevolta una di esse ne cita un'altra. La rete delle citazioni stabilisce connessioni tra documenti che non necessitano del lavoro di intermediazione di un bibliotecario per essere riconosciute: identificate, ripulite e immesse in un database, per una banale applicazione della proprietà transitiva, le citazioni consentono di risalire da un documento rilevante per me a tutti i documenti di autori che, al pari di me, hanno giudicato rilevante lo stesso contributo e, presumibilmente per lo stesso o analogo motivo, lo hanno citato. Nasceva così un apparato bibliografico radicalmente diverso dai precedenti, una bibliografia "senza soggetto", in due

* Si preferisce tradurre con la forma plurale "indice di citazioni" anziché con "indice di citazione", spesso adoperata nei contributi in lingua italiana sull'argomento, per due motivi: 1) sottolineare la composizione originaria di questa tipologia di pubblicazioni (indice nel senso di elenco alfabetico di titoli citati nella letteratura); 2) evitare confusione con il concetto di "indice di citazione" inteso come "indicatore di citazione", cioè rappresentazione semplificata di un aspetto di un fenomeno inattuabile nella sua complessità. Un'ulteriore precisazione potrebbe essere fatta a proposito della differenza tra "indice" (nel secondo senso) e "indicatore", ma su tale distinzione, della quale non si terrà conto nel resto del lavoro, si rinvia a: H. Zeisel, "Difficulties in indicator construction: Notes and queries," in *Toward a Metric of Science: the Advent of Science Indicators*, edited by Y. Elkana, et al. (New York: John Wiley and Sons, 1979): 253-58, in particolare pp. 253-254.

sensi: non ha bisogno di *subject headings* e non richiede la mediazione concettuale dei bibliotecari, sono gli scienziati stessi a scriverla senza saperlo. Le conseguenze di questa intuizione sono andate ben oltre il dominio della ricerca bibliografica.

Una volta compresa l'utilità di identificare, raccogliere e rendere ricercabili, attraverso un sistema ordinato di registrazioni, tutti i riferimenti bibliografici contenuti in una vasta mole di letteratura scientifica corrente, si è cominciato a riflettere sull'opportunità di contare e analizzare le citazioni ricevute da un documento o da un autore al fine di quantificarne il successo cognitivo, l'impatto sulla comunità scientifica. Il numero di citazioni sedimentate su un documento è stato acclamato come un indicatore imparziale della qualità del lavoro del suo autore, un parametro da utilizzare, all'occorrenza, come antidoto ai criteri di finanziamento e promozione accademica governati da logiche occulte e favoritismi personali. Di riflesso, un documento immesso nel circuito della comunicazione scientifica si è trasformato nel personaggio in cerca d'autore di una storia parallela a quella dei suoi usi culturali e istituzionali: la storia, scritta negli indici di citazioni, di tutte le volte in cui qualcuno si sarebbe ricordato di lui per citarlo in una nota bibliografica oppure, per dimenticanza, ignoranza o malafede, avrebbe evitato di farlo. In che modo ciò sia stato possibile partendo da un oggetto banale come una citazione, dipende da un'ambiguità costitutiva degli apparati bibliografici sulla quale hanno di recente richiamato l'attenzione Paul Wouters e Blaise Cronin. Per renderla esplicita occorre evidenziare una differenza che le consuetudini linguistiche tendono ad appiattare, quella tra “riferimento bibliografico” (*reference, Referenz*) e “citazione” (*citation, Zitat*):¹² se un documento **A** cita, nelle note a piè di pagina o nella bibliografia finale, un documento **B**, di cui riconosce in maniera diretta o indiretta la correlazione con le tesi esposte,

¹² P. Wouters, *The Citation Culture*. PhD Thesis: University of Amsterdam, 1999) <<http://www.garfield.library.upenn.edu/wouters/wouters.pdf>>, pp. 10-12; B. Cronin, “Semiotics and evaluative bibliometrics” *Journal of Documentation* 2000, 56(4): 440-53.

allora **A** contiene un “riferimento bibliografico” (*bibliographic reference*) a **B**; cambiando prospettiva, dal punto di vista di **B**, la stessa operazione implica una “citazione” (*citation*) del documento **B** da parte di **A**. Quindi il riferimento bibliografico è un attributo del testo citante, un segno che rimanda al documento citato nel tentativo di comunicare al lettore la familiarità con un testo e, di conseguenza, un contesto (di teorie, concetti, abitudini sociali). La citazione al contrario è un attributo del testo citato, un segno costruito invertendo in maniera speculare il riferimento bibliografico per mezzo di un artificio editoriale. Il riferimento bibliografico è orientato al passato, guarda all'indietro, mentre la citazione implica un movimento in avanti, verso il futuro.¹³ Nel seguito, per comodità espositiva, si continuerà a parlare indifferentemente di riferimenti bibliografici e citazioni, delegando al contesto la funzione di chiarire il senso dell'espressione, ma questa sfumatura semiotica resta un punto fermo della discussione. La sua conseguenza più importante riguarda proprio la citazione. Slegata dal contesto di provenienza e riprodotta, approfondita, tipograficamente rimodellata in un indice cartaceo o elettronico, essa perde la connotazione strettamente bibliografica che aveva nella fonte e diventa il simbolo di qualcosa di immateriale: di una presunta affinità di contenuto, di un'associazione di idee, di un'"influenza" intellettuale. Ripetendo la stessa operazione di sdoppiamento su un numero via via crescente di fonti, si crea allora uno strumento completamente diverso da una bibliografia, con il quale, oltre alle ricerche bibliografiche, si possono fare almeno altre due operazioni "meta-scientifiche": analizzare le relazioni quantitative tra documenti citati e documenti citanti al fine di misurare il tasso di ricezione dei primi e disegnare, sulla base della frequenza con cui documenti affini sono accoppiati nelle bibliografie di un certo numero di articoli, il perimetro bibliometrico delle discipline e dei fronti attivi di ricerca scientifica.

¹³ L. Egghe & R. Rousseau, *Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science* (Amsterdam: Elsevier, 1990), p. 204.

Le misure ricavate da un indice di citazioni erano, agli occhi dei primi bibliometri, misure silenziose, "discrete" (*unobtrusive*) della letteratura scientifica, perché effettuate senza interferire con il comportamento normale dei soggetti interessati. Soprattutto, erano misure che non avevano alcuna correlazione con la torre di Babele degli stili di pensiero e dei linguaggi disciplinari: nello specchio levigato dello *Science Citation Index*, il contenuto della letteratura scientifica non ha alcuna importanza, quello che conta sono le relazioni formali, di natura bibliografica, tra le unità che la compongono.¹⁴ Il sogno di politici e manager della scienza era finalmente realizzabile: non solo misurare l'input della scienza (i soldi spesi, il numero di ricercatori reclutati e così via), nè limitarsi ad una misura grezza dell'output (il numero di pubblicazioni prodotte), ma, ciò che più conta ai fini di una ripartizione strategica delle risorse, identificare i settori trainanti della ricerca, i gruppi più forti, gli scienziati più meritevoli, le aree di sviluppo e quelle di stagnazione. Gli indici di citazioni hanno incrociato così il proprio cammino con quello della tradizione di studi bibliometrici e l'incontro, documentato nel capitolo 1, è stato decisivo per entrambi, una specie di matrimonio annunciato: la bibliometria offriva ai *citation indexes* una serie di leggi (Lotka, Bradford, Zipf, Garfield) utili a legittimare, sul piano teorico, i criteri operativi di selezione dei materiali da includere nel repertorio; gli indici di citazioni, d'altro canto, dotavano la bibliometria di una fonte ineguagliabile di dati empirici da impiegare per la costruzione di indicatori di performance scientifica.

Oggi l'idea che un indice sia capace di fornire informazioni diverse da quelle per le quali era stato progettato è abbastanza familiare. Non è un segreto che, manovrando opportunamente le stringhe di parole chiave, si può usare il motore di ricerca Google per accedere ad informazioni riservate e potenzialmente lesive di interessi altrui (file di password, informazioni

¹⁴ P. Wouters, *The Citation Culture....cit.*, cap. 1, pp. 1-16.

bancarie, etc.).¹⁵ Lo stesso Google peraltro ha reso popolare la nozione di documenti (pagine web) recuperati e ordinati secondo un criterio che rispecchia sia il numero di link/citazioni ricevute che la qualità dei siti da cui tali citazioni provengono, misurata secondo il medesimo parametro: le pagine più linkate sono considerate le più rilevanti, e tra queste vengono prima le pagine linkate da siti a loro volta molto linkati. Google, e dopo di lui strumenti più ambiziosi di navigazione della letteratura specializzata come Scopus di Elsevier e Google Scholar, sono l'espressione evidente di come, immettendo la citazione bibliografica in una routine di conteggio e valutazione quantitativa, ad un certo punto non sia più possibile separare l'aspetto "neutrale" della ricerca bibliografica da quello del valore intrinseco assegnato ai documenti (e agli autori) per mezzo di un sistema di relazioni bibliografiche indipendenti dal giudizio di colui che effettua la ricerca. Uno studio sugli indici di citazioni per il recupero dell'informazione bibliografica deve essere perciò anche, necessariamente, uno studio sulla loro originaria vocazione extra-bibliografica.

In linea con tali premesse, i capitoli 3, 4, 5 e 6 mostreranno in che modo, estrapolato dal contesto di appartenenza e ingabbiato in un apparato percorribile nelle due direzioni possibili - dal citato al citante e, viceversa, dal citante al citato -, il riferimento bibliografico abbia perso le caratteristiche di individualità che lo legavano unicamente ad un codice di buona condotta professionale diventando di volta in volta: un canale di recupero dell'informazione alternativo rispetto a quelli tradizionali (capitolo 3); una base empirica per le scienze sociali (capitolo 4); un punto d'appoggio per le procedure di valutazione nel quadro del sistema istituzionale di ricompense e riconoscimento professionale degli scienziati (capitolo 5); un modello analitico per lo studio quantitativo della comunicazione scientifica mediata da

¹⁵ *Google: A hacker's best friend*, 2005 <<http://www.awarenetwork.org/home/paris2k/P2KCEGOOGLE.pdf>>.

Internet e dal World Wide Web, con i link tra pagine e siti web equiparati ai rinvii bibliografici tra documenti (capitolo 6).

Tutte queste operazioni poggiano su alcune premesse, esplicitate nel capitolo 2, relative alla stabilità delle informazioni veicolate dalle citazioni bibliografiche e all'uniformità degli stili di citazione degli scienziati. Nel capitolo 7 si illustrerà l'incompatibilità di tali premesse non solo con una parte consistente della riflessione epistemologica contemporanea, ma anche con gli esiti di alcune indagini empiriche sulla vita di laboratorio e sulle strategie comunicative messe in atto dagli scienziati nel passaggio dalla registrazione dei dati sperimentali alla loro organizzazione retorico-discorsiva sotto forma di articoli destinati alla pubblicazione.

La diffusione commerciale dei *citation indexes* ha ingigantito il significato delle citazioni bibliografiche per la vita e la carriera degli scienziati. Le citazioni sono oramai parte integrante del sistema di supporti informativi adoperati dai ricercatori ed il conteggio delle citazioni è uno degli strumenti utilizzati per stimare l'impatto e la visibilità di singoli o gruppi di scienziati. Le citazioni, in breve, sono diventate uno degli ingranaggi dell'industria scientifica e culturale. Produttori (autori), consumatori (altri scienziati), intermediari della comunicazione scientifica (bibliotecari, informatici, agenti dell'industria dell'informazione) hanno perciò maturato una consapevolezza diversa del loro valore e significato.¹⁶ Nelle università e nei college americani si usano le citazioni per classificare e attribuire un punteggio a tutto ciò che minaccia di destabilizzare la regola aurea della meritocrazia, in una specie di *rating game* giocato talvolta al massacro e senza una cognizione critica dei limiti e del valore da attribuire ai conteggi.¹⁷ Sull'esempio americano, diverse

¹⁶ «The commercialization of the citation means that those who generate citations (publishing authors), those who make use of them (other scientists), those who process and package them (the information industry), and those who mediate in their delivery (librarians and information scientists) need to have a sound grasp of what citation entails and signifies»: B. Cronin, *The Citation Process: the Role and Significance of Citations in Scientific Communications* (London: Taylor Graham, 1984), p. 24.

¹⁷ Sugli usi patologici delle citazioni in ambito accademico statunitense si è soffermato Jonathan R. Cole in: J.R. Cole, "A short history of the use of citations as a measure of the impact of scientific and scholarly work,"

nazioni europee hanno cominciato ad istituzionalizzare l'impiego degli indicatori bibliometrici basati sulle citazioni nella valutazione delle prestazioni intellettuali. Anche in Italia i *citation indexes* sono entrati nel dominio del controllo di qualità della produzione scientifica, come testimoniano il programma VPS (Valutazione Pubblicazioni Scientifiche), sviluppato sul modello del britannico *Research Assessment Exercise*, e l'iniziativa della CRUI (Conferenza dei Rettori delle Università Italiane) di condurre un'analisi delle citazioni delle pubblicazioni scientifiche nazionali.¹⁸ Pur non esistendo una tradizione italiana di studi bibliometrici, il problema dell'uso valutativo delle banche dati è emerso all'interno del dibattito sull'obiettività delle procedure di valutazione e sui meccanismi di partecipazione sociale della scienza.¹⁹ Un'attenzione specifica per l'applicazione delle tecniche bibliometriche è evidente inoltre nell'attività del Consiglio Nazionale delle Ricerche (CNR)²⁰ e nel contributo italiano all'ultima edizione dello *Handbook of Quantitative Science and Technology Research* (2004).²¹

in *The Web of Knowledge: a Festschrift in Honor of Eugene Garfield* (Medford, NJ: Information Today Inc, 2000): 281-300.

¹⁸ Sul VPS: Osservatorio per la Valutazione del Sistema Universitario, *Proposta per un programma di valutazione della produzione scientifica nelle università. Programma VPS (DOC 3/99)* (Roma: 1999). Dal 2000 al 2002 un gruppo di lavoro della CRUI ha condotto, sulle banche dati dell'ISI, uno studio della qualità della produzione scientifica nazionale: E. Brenno; G.A. Fava; V. Guardabasso, *La ricerca scientifica nelle università italiane. Una analisi delle citazioni della banca dati ISI* (Roma: CRUI, 2002).

¹⁹ Il dibattito era già attivo alla metà degli anni '90, soprattutto in ambito sanitario: A. Dracos & G. Cognetti, "La letteratura scientifica: indicatori bibliometrici e bibliografici quali criteri integrativi per la valutazione obiettiva di un'attività di ricerca" *Annali dell'Istituto Superiore di Sanità* 1995, 31(3): 381-90. La sua ripresa in anni più recenti è documentata in: A. Valente & D. Luzi (Ed.), *Partecipare la scienza* (Roma: Biblink, 2005); A.M. Tammaro, "Qualità della comunicazione scientifica" *Biblioteche Oggi* 2001(7): 104-07; A.M. Tammaro, "Qualità della comunicazione scientifica 2" *Biblioteche Oggi* 2001(8): 74-78; V. Comba, *La valutazione delle pubblicazioni: dalla letteratura a stampa agli Open Archives*, 2002

<<http://eprints.rclis.org/archive/00000095/01/valutazione.pdf>>; D. Ugolini & C. Casilli, "The visibility of italian journals" *Scientometrics* 2003, 56(3): 345-55.

²⁰ Ad esempio: F. Tuzi, "The scientific specialization of the italian regions" *Scientometrics* 2005, 62(1): 87-111. Sugli indicatori bibliometrici utilizzati per la valutazione degli istituti del CNR vedi: L. Padrielli; M. Perotti; F. Tuzi, "Criteri ed indicatori per la valutazione degli istituti del CNR," in *Partecipare la scienza*, edited by A. Valente, et al. (Roma: Biblink, 2005): 113-32.

²¹ H.F. Moed; W. Glänzel; U. Schmoch (Ed.), *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems* (Dordrecht: Kluwer Academic, 2004). I contributi italiani sono quelli di: A. Bonaccorsi; C. Darano, *Econometric approaches to the analysis of productivity of R&D systems*, pp. 51-74; F. Naldi; D. Luzi; A. Valente; I. Vannini Parenti, *Scientific and technological performance by gender*, pp. 299-314; S. Breschi; F. Lissoni, *Knowledge networks from patent data*, pp. 613-644; S. Brusoni; A. Geuna, *Specialization and integration*, pp. 733-759.

La rincorsa al modello anglosassone non è stata indolore: approssimazione, malintesi, negligenze clamorose, come l'uso indebito dell'*Impact Factor* per valutare i curricula individuali discusso nel paragrafo 5.1, hanno accompagnato questo processo di assimilazione, oscurando la natura complessa degli indici bibliometrici. Al di là degli effetti negativi, le distorsioni hanno comunque avuto il merito di richiamare l'attenzione sul problema della verifica sociale di un qualcosa che fino ad alcuni anni fa era ritenuto il non-misurabile per eccellenza: la produttività intellettuale dell'uomo di scienza, non del genio rivoluzionario beninteso, ma dello scienziato "normale" (nel senso di Kuhn) che contende ad altri scienziati "normali" lo stesso titolo accademico o lo stesso finanziamento.

È sottinteso che l'intera discussione si limita alle due aree in cui gli indici di citazioni hanno trovato un uso generalizzato sia nella ricerca bibliografica che nella valutazione della ricerca: le scienze naturali e le scienze della vita.

Scienze sociali e scienze umane, per la specificità dell'oggetto d'indagine, dei metodi applicati e delle forme di comunicazione adottate, sono ancora in gran parte estranee alle lusinghe della bibliometria.²²

Il volume e la varietà tematica della letteratura bibliometrica, testimoniati dalla comparsa di bibliografie specializzate già all'inizio degli anni '80,²³ annientano qualunque pretesa di esaustività. I materiali qui discussi rappresentano perciò una selezione delle fonti attuata nell'ottica di individuare

²² È significativo che, incaricato di elaborare una metodologia per l'accertamento della qualità nel campo delle scienze umane e storico-sociali, uno dei leader della bibliometria europea non faccia alcun uso degli indici di citazioni: H.F. Moed, "Towards research performance in the humanities" *Library Trends* 2002, 50(3): 498-520. Sulla diversità di pratiche citazionali tra campo scientifico e campo umanistico, alcune osservazioni interessanti si trovano in: C. Frost, "The use of citations in literary research: preliminary classifications of citation functions" *Library Quarterly* 1979, 49(4): 399-414.

²³ R. Hjerpe, *A Bibliography of Bibliometrics and Citation Indexing & Analysis* (Stockholm: Royal Institute of Technology, 1980); R. Hahn, *A bibliography of quantitative studies on science and its history* (Berkeley: office for History of Science and Technology, 1980); R.C. Stowe, *Annotated bibliography of publications dealing with qualitative and quantitative indicators of the quality of science* (Cambridge, Massachusetts: Harvard University, 1986); M.K. Sellen, *Bibliometrics: an Annotated Bibliography, 1970-1990* (New York: G.K. Hall, 1993). Una bibliografia online più aggiornata si trova in: J. Furner, *Scholarly Communication and Bibliometrics*, 2001 <<http://polaris.gseis.ucla.edu/jfurner/03-04/208/208res.html>>. Per un elenco delle bibliografie e delle rassegne pubblicate fino al 2001 vedi: W.W. Hood & C.S. Wilson, "The literature of Bibliometrics, Scientometrics, and Informetrics" *Scientometrics* 2001, 52(2): 291-314.

i testi più significativi per la comprensione del ruolo della citazione bibliografica negli studi quantitativi della scienza. Davanti a pubblicazioni di contenuto informativo affine, il criterio seguito per la compilazione delle note e nella bibliografia finale è semplice: hanno avuto la precedenza quelli accessibili gratuitamente online. Sotto questo profilo Eugene Garfield ha prestato un servizio importante alla comunità degli storici pubblicando sul Web, oltre a numerosi saggi e articoli di autori che si sono occupati di tematiche riconducibili agli indici di citazioni, il testo completo dei propri scritti:²⁴ monografie, tesi di dottorato, relazioni di convegni, interviste, lettere, progetti e tutti i commenti introduttivi ai fascicoli settimanali dei *Current Contents* (dal 1962 al 1995) che formano il nucleo centrale dei 15 volumi di *Essays of an Information Scientist*. Questi materiali documentano in maniera approfondita la storia e la difesa appassionata di un'idea, l'indicizzazione delle citazioni, da parte di un ricercatore/imprenditore attivo sul fronte della scienza dell'informazione dalla metà degli anni '50 in poi. Lo fanno ovviamente dal punto di vista di chi è stato in prima persona al centro dei fatti che racconta, quindi la cautela da osservare è la stessa necessaria al cospetto di tutti gli archivi personali, che hanno un'implicita vocazione a scriversi da soli la propria storia o, quantomeno, a ritagliarne la base documentaria.

²⁴ <<http://www.garfield.library.upenn.edu/>>.

Capitolo 1

Contare la scienza: storia breve degli studi bibliometrici

‘Bibliometria’, ‘scientometria’, ‘informetrica’, ‘webmetrica’, ‘netmetrica’, ‘cybermetrica’: il vocabolario della metrologia della comunicazione scientifica è ricco di termini che alludono ad aree di ricerca affini e per molti versi indistinguibili. Una facile tentazione potrebbe essere quella di stabilire una successione cronologica diretta che dalla bibliometria arrivi alla cybermetrica passando attraverso la scientometria e l’informetrica, ma la realtà è più complessa, le differenze più sfumate.

L’obiettivo di ognuna è analizzare, quantificare, misurare dei processi comunicativi, cambia invece l’ordine dei fattori e il perimetro di ciò che si vuole misurare.

La bibliometria sottolinea l’aspetto materiale dell’impresa: contare libri, articoli, pubblicazioni, citazioni, in generale qualunque elemento statisticamente significativo di un discorso scritto, a prescindere dal campo di applicazione o dalla disciplina.

La scientometria pone invece l’accento sulla misurazione del progresso scientifico e tecnologico o, in termini equivalenti, sulla valutazione quantitativa e comparativa dell’attività, della produttività e del contributo dei ricercatori allo progresso delle conoscenze. Nei limiti in cui tale accertamento viene compiuto attraverso calcoli sulle pubblicazioni e sulle citazioni, le due aree della scientometria e della bibliometria diventano praticamente indistinguibili.

L’informetrica rivendica l’appartenenza disciplinare delle tecniche bibliometriche e scientometriche alla scienza dell’informazione e, in maniera implicita, afferma il superamento della biblio(-grafia, -metria, -teca)

nell'universo dei media digitali. Secondo una delle definizioni più accreditate, l'informetrica è lo studio degli aspetti quantitativi dell'informazione in qualunque forma (non solo informazione scritta o di tipo bibliografico) ed in qualunque gruppo sociale (non solo scienziati).²⁵ In poche parole, è l'insieme degli insiemi di tutte le altre metriche poiché tutte contano un qualche tipo d'informazione.

La webmetrica (netmetrica, cybermetrica) estende l'uso di metodi e concetti informetrici ai documenti in formato digitale disseminati nella rete Internet. Nella misura in cui tali documenti sono registrati da qualche parte, non importa se in modo permanente o per un tempo limitato, la disciplina che se ne occupa è un sottoinsieme della bibliometria. I due pionieri della webmetrica, Bjerneborn e Ingwersen, propongono tuttavia di distinguere tra lo studio quantitativo dei vari aspetti legati alla costruzione e all'uso delle risorse Web (webmetrica in senso stretto) e quello più generale di tutte le applicazioni di Internet (cybermetrica).²⁶

Una storia esaustiva di queste discipline è lontana dall'essere scritta. Nel caso della cybermetrica semplicemente non esiste una storia, ma un fronte di ricerca in via di costruzione. Alcuni episodi nondimeno sono noti da tempo. Passandoli in rassegna si cercherà di evidenziarne il significato per la storia degli indici (e degli studi) citazionali.

²⁵ B.C. Brookes, "Biblio-, sciento-, infor-metrics??? What are we talking about," in *Informetrics 89/90: Selection of papers submitted for the Second International Conference on Bibliometrics, Scientometrics and Informetrics*, edited by L. Egghe, et al. (Amsterdam: Elsevier Science, 1990): 31-43; L. Egghe & R. Rousseau, *Introduction to Informetrics...cit.*, pp. 1-4; J. Tague-Sutcliffe, "An introduction to informetrics" *Information Processing & Management* 1992, 28(1): 1-3; I.N. Sengupta, "Bibliometrics, Informetrics, Scientometrics and Librametrics: An Overview" *Libri: International Library Review* 1992, 42(2): 75-98.

²⁶ L. Björneborn & P. Ingwersen, "Toward a basic framework for webometrics" *Journal of the American Society for Information Science and Technology* 2004, 55(14): 1216-27.

1.1. La bibliografia statistica

In quanto beni patrimoniali, i libri, come le persone fisiche e le terre, sono stati oggetto sin dai tempi più antichi di una forma embrionale di statistica descrittiva fondata sul conteggio e sulla classificazione a fini burocratici, fiscali, amministrativi.²⁷ Nel XVII secolo, la popolarità dei giochi d'azzardo e gli interessi legati alle assicurazioni sulla vita incentivarono lo studio di metodi matematici che, partendo da un insieme di dati iniziali, permettessero di effettuare predizioni e inferenze quantitative sull'evoluzione del loro comportamento. La teoria matematica della probabilità, nata nel '600 dalla corrispondenza tra Pascal e Fermat, divenne nel secolo seguente il nucleo matematico della statistica inferenziale, il cui metodo d'indagine fu esteso ben presto a qualunque scienza basata sull'osservazione. Alla fine dell'800, si cominciò così ad applicare la statistica e la teoria della probabilità anche ai libri e ad altri prodotti della comunicazione scritta per trarre conclusioni generali sul progresso di una disciplina scientifica o, più semplicemente, per ricavare indicazioni utili alla politica degli acquisti in biblioteca.²⁸ Il prodotto di questa attività si chiamava all'inizio "bibliografia statistica" (*statistical bibliography*) ed era possibile grazie all'ampia disponibilità di bibliografie disciplinari e apparati di *abstracting* e *indexing* redatti, dal XVIII secolo in poi, sotto la spinta della settorializzazione della ricerca scientifica.

In un trattato del 1896, Frank Campbell utilizzò la statistica per rappresentare la distribuzione degli argomenti nelle pubblicazioni elencate dalle principali

²⁷ R.N. Broadus, "Early approaches to bibliometrics" *Journal of the American Society for Information Science* 1987, 38(2): 127-29.

²⁸ Sulla storia della bibliometria vedi: D.H. Hertz, "Bibliometrics, history of the development of ideas," in *Encyclopedia of Library and Information Science*, 42, Supplement 7 (New York: M. Dekker, 1987): 144-211; F. Osareh, "Bibliometrics, citation analysis and co-citation analysis: A review of literature I" *Libri: International Library Review* 1996, 46: 149-58; F. Osareh, "Bibliometrics, citation analysis and co-citation analysis: A review of literature II" *Libri: International Library Review* 1996, 46: 217-25.

bibliografie internazionali.²⁹ F. J. Cole e Nellie Eales, nel 1917, applicarono l'analisi statistica alla letteratura sull'anatomia comparata accumulata lungo un arco di tre secoli, dal 1543 al 1860, producendo una rappresentazione grafica della crescita documentaria e correlando le fasi di sviluppo o di recessione con fattori umani, economici, sociali.³⁰ Nel 1922 E. Wyndham Hulme, bibliotecario presso il *Patent Office* di Londra, estrasse da 13 annate dello *International Catalogue of Scientific Literature* i dati relativi alla distribuzione degli scienziati per area disciplinare e censì il numero di riviste indicizzate dai servizi bibliografici nazionali. Hulme era convinto che il processo di civilizzazione andasse di pari passo con la specializzazione e che, classificando tutti i libri del mondo secondo criteri universali e ordinandoli cronologicamente all'interno di ciascuna classe, si poteva costruire un'immagine bibliografica attendibile dei progressi della mente umana.³¹ Circa vent'anni dopo, in una tesi di dottorato, Charles Gosnell applicò la statistica a tre bibliografie di testi universitari con l'obiettivo di esprimere in forma matematica il tasso di obsolescenza della letteratura. Un'operazione che partiva da un assunto analogo a quello di Hulme: il numero di libri, quindi anche la velocità con cui invecchiano nelle diverse aree del sapere, rispecchia l'evoluzione generale della cultura.³²

Nel 1927 usciva uno dei primi saggi di analisi delle citazioni, un modello per tante indagini successive: allo scopo di fornire un sostegno concreto alla politica degli acquisti librari di una piccola biblioteca di un college del sud della California, i coniugi Gross esaminarono 3633 riferimenti bibliografici di

²⁹ F. Campbell, *The Theory of the National and International Bibliography: with Special Reference to the Introduction of System in the Record of Modern Literature* (London: Library Bureau, 1896).

³⁰ F.J. Cole & N.B. Eales, "The history of comparative anatomy. Part I: A statistical analysis of the literature" *Science Progress* 1917, 11: 578-96.

³¹ «These changes in the activities of the human mind would be indicated on any bibliographical chart, if bibliography were organized on the basis of a universal and uniform classification of recorded knowledge; for if civilization is but the product of the human mind operating upon a shifting platform of its environment, we may claim for bibliography that it is not only a pillar in the structure of the edifice, but that it can function as a measure of the varying forces to which this structure is continuously subjected»: E.W. Hulme, *Statistical Bibliography in Relation to the Growth of Modern Civilization* (London: Grafton, 1923), p. 43.

un'annata del *Journal of the American Chemical Society* classificando i periodici in funzione del numero di citazioni ricevute.³³ L'equivalenza tra qualità dei periodici e numero di citazioni era così stabilita, e non sorprende che Eugene Garfield abbia richiamato esplicitamente il lavoro dei Gross in fase di progettazione dello *Science Citation Index*.³⁴

Con l'obiettivo di stabilire le interrelazioni tra i diversi settori della psicologia, nel 1936 Hulsey Cason e Marcella Lubotsky analizzarono i riferimenti bibliografici di 28 riviste di psicologia, rappresentando graficamente in una tabella la percentuale dei riferimenti incrociati.³⁵ Circa dieci anni dopo, questo contributo trovò un ideale prolungamento nel lavoro di Herman Henkle sulla letteratura periodica nel settore biochimico. Anche qui si traeva beneficio da una forma rudimentale di analisi delle citazioni mediante la compilazione di una lista di riviste di chimica ordinate in base alla frequenza con cui erano citate.³⁶

Alla fine degli anni '40, in una dissertazione per la Graduate Library School dell'Università di Chicago, Herman Fussler si servì dei riferimenti bibliografici di tre annate dei *Physical Review* e del *Journal of the American Chemical Society* per estrarre una lista di *core journals* utilizzati dai chimici e dai fisici statunitensi. Come già avevano fatto i Gross, egli mostrò che il ciclo vitale di tali riferimenti bibliografici era breve: la metà di quelli di chimica puntava a documenti pubblicati negli ultimi otto anni, la metà di quelli di fisica a documenti pubblicati negli ultimi cinque anni.³⁷

³² La tesi di dottorato di Gosnell risale al 1943. L'anno seguente i risultati salienti della tesi furono pubblicati in: C.F. Gosnell, "Obsolescence of books in college libraries" *College & Research Libraries* 1944, 5: 115-25.

³³ P.L.K. Gross & E.M. Gross, "College libraries and chemical education" *Science* 1927, 66(1713): 385-89.

³⁴ E. Garfield, "Citation Indexes for science: a new dimension in documentation through association of ideas," in *Essays of an Information Scientist*, 6 (Philadelphia: ISI Press, 1983): 468-71, p. 469.

³⁵ H. Cason & M. Lubotsky, "The influence and dependence of psychological journals on each other" *Psychological Bulletin* 1936, 33: 95-103.

³⁶ H.H. Henkle, "The periodical literature of biochemistry" *Bulletin of the Medical Library Association* 1938, 27: 139-47.

³⁷ H.H. Fussler, "Characteristics of the research literature used by chemists and physicists in the United States" *Library Quarterly* 1949, 19: 19-35.

Nel frattempo si manifestavano i segni di una nuova consapevolezza terminologica e disciplinare.

Paul Otlet aveva dedicato una sezione del suo *Traité de documentation* (1934) all'argomento *Le livre et la mesure - Bibliometrie*. S. R. Ranganathan, in occasione della conferenza ASLIB di Leamington Spa (1948), auspicò l'uso di metodi matematici per promuovere, sul modello delle metrologie scientifiche esistenti (*Biometry, Econometry, Psychometry* etc.), una *Librametry* intesa ad ottimizzare i flussi informativi nei servizi di biblioteca,³⁸ ed egli stesso applicò metodi bibliometrici nella gestione della biblioteca dell'Università di Madras.³⁹ Alan Pritchard, in un articolo del 1969, propose di sostituire il termine 'bibliografia statistica' (*statistical bibliography*) con 'bibliometria' (*bibliometrics*) per designare l'applicazione della matematica e di metodi statistici ai libri e alle altre forme di comunicazione scritta.⁴⁰ Robert Fairthorne, scrivendo nello stesso anno di Pritchard, generalizzò lo scopo della bibliometria nel trattamento quantitativo delle proprietà di qualunque discorso scritto e delle azioni ad esso associate.⁴¹ Tre anni dopo, lo stesso Pritchard specificò che l'oggetto della misura bibliometrica è il processo di trasferimento dell'informazione allo scopo di analizzarlo e controllarlo.⁴²

1.2. Le leggi "troppo universali" della bibliometria

Tra gli anni '20 e '30 del '900, hanno visto la luce tre studi fondamentali per la storia della bibliometria: il lavoro di Lotka sulla distribuzione della

³⁸ S.R. Ranganathan, *Aslib Proceedings* 1949, 1(2): 102-03.

³⁹ S.R. Ranganathan, "Librametry and its scope," in *7th Documentation Research Training Centre. Annual Seminars I*. (Bangalore: DTRC, 1969).

⁴⁰ «BIBLIOMETRICS, i.e. the application of mathematics and statistical methods to books and other media of communication»: A. Pritchard, "Statistical bibliography or bibliometrics?" *Journal of Documentation* 1969, 25(4): 348-49, p. 349.

⁴¹ R.A. Fairthorne, "Empirical hyperbolic distributions (Bradford-Zipf-Mandelbrot) for bibliometric description and prediction" *Journal of Documentation* 1969, 25: 319-43.

⁴² A. Pritchard, "Bibliometrics and information transfer" *Research in Librarianship* 1972, 4: 37-46.

produttività scientifica in relazione al numero degli autori;⁴³ quello di Bradford sulla distribuzione degli articoli nelle riviste;⁴⁴ quello di Zipf sulla distribuzione delle parole in un testo.⁴⁵

Queste pubblicazioni descrivono, sotto forma di semplici equazioni, delle regolarità, impropriamente chiamate leggi, caratterizzate da una strana similitudine nel meccanismo probabilistico degli eventi descritti. Un meccanismo nel quale, a differenza di quanto accade per la maggior parte delle grandezze naturali, non esiste un valore medio più probabile degli altri attorno a cui si distribiscono in maniera ordinata tutti i restanti valori (distribuzione normale o gaussiana), ma molti valori piccoli coesistono con pochi valori grandi e la distribuzione complessiva è fortemente asimmetrica (distribuzione iperbolica o regolata da una legge di potenza). La prima situazione è tipica di variabili derivanti dall'aggregazione di molti comportamenti casuali: il valore centrale (la media) è la norma, mentre i valori che si discostano dalla media possono essere considerati deviazioni dalla norma, "errori" casuali. La seconda situazione si verifica invece quando la casualità dei comportamenti che determinano le variabili è compatibile con l'esistenza di valori "straordinari", al di fuori della norma.

1.2.1. La legge di Lotka

Alfred Lotka (1880-1949) è noto innanzitutto per aver dato vita, contemporaneamente a Vito Volterra e in modo indipendente, ad un modello matematico di dinamica delle popolazioni animali utile a spiegare le

⁴³ A.J. Lotka, "Statistics - The frequency distribution of scientific productivity" *Journal of Washington Academy of Sciences* 1926, 16(12): 317-25.

⁴⁴ Formulata in: S.C. Bradford, "Sources of information on specific subjects" *Engineering* 1934, 137: 85-86, la legge cominciò a ricevere attenzione solo alla fine degli anni '40, in particolare dopo la pubblicazione di: S.C. Bradford, *Documentation* (Washington, D.C.: Public Affairs Press, 1950).

⁴⁵ G.K. Zipf, *The Psycho-Biology of Language: An Introduction to Dynamic Philology* (London: Routledge, 1936).

fluttuazioni nel rapporto tra predatori e prede. Rivolgendosi al settore della documentazione, Lotka cercò di determinare in che misura scienziati di diverso calibro, impegnati nella lotta per la vita sul fronte della comunicazione scientifica, contribuiscono al progresso delle conoscenze. Egli contò un certo numero di autori e relative pubblicazioni schedate nei *Chemical Abstracts* (dal 1907 al 1916) e nelle *Geschichtstafeln der Physik* di Auerbach (dalle origini al 1900) e generalizzò le frequenze osservate in una legge empirica secondo la quale il numero di autori che producono n contributi in un'area di ricerca è all'incirca pari a $1/n^2$ del numero di autori che producono un solo contributo. In simboli:

$$p(n) = K/n^2$$

dove p è il numero di autori che producono n contributi e K una costante caratteristica dell'area disciplinare. Quindi, se il 60% degli autori di un'area disciplinare produce un solo contributo, il 15% produrrà due pubblicazioni ($1/2^2 * 60$), il 7% tre pubblicazioni ($1/3^2 * 60$) e così via. In poche parole, il dono della produttività scientifica non è equamente distribuito tra colleghi: esistono pochi autori molto produttivi ed un numero molto più alto di autori con una bassa produttività. «Distribuzioni di questo tipo - ammetteva Lotka - «hanno un vasto raggio di applicabilità ad una varietà di fenomeni, e la mera forma di tale distribuzione getta poca o nessuna luce sulle relazioni fisiche sottostanti».⁴⁶

Il lavoro di Lotka fu ignorato per diverso tempo. A partire dagli anni '70, in compenso, è andato incontro ad una serie di applicazioni, reinterpretazioni, smentite e presunte riconferme.⁴⁷ Ad esempio ne è stata mostrata

⁴⁶ A.J. Lotka, "Statistics..." cit., p. 323.

⁴⁷ Il primo numero della rivista *Scientometrics*, pubblicato nel 1978, conteneva una bibliografia di lavori su Lotka di 437 titoli: J. Vlachy, "Frequency distributions of scientific performance: A bibliography of Lotka's law and related phenomena" *Scientometrics* 1978, 1: 109-30.

l'incongruenza rispetto alla distribuzione della produttività degli autori attivi nei settori delle scienze umane, della biblioteconomia, dell'*information science*, della medicina legale e della *computer science*,⁴⁸ mentre uno studio recente sullo sviluppo di software *open source* da parte dei programmatori disseminati nel cyberspazio ne ha reclamato la sostanziale validità.⁴⁹ Come spesso accade quando un dispositivo funziona senza sapere bene perché funziona, molti autori hanno citato la letteratura precedente lasciando credere che qualcuno avesse già dimostrato la validità della legge di Lotka in determinati campi scientifici, ma non era così. Coloro che hanno provato a testarne la validità invece si sono spesso concentrati su serie temporali e gruppi di autori diversi da quelli presi in esame da Lotka, producendo risultati difficilmente comparabili con gli originali. E l'enigma arriva fino alla fonte: applicando un test statistico ai dati adoperati dallo stesso Lotka, si trova che la curva da egli disegnata per descrivere le frequenze osservate si adatta solo ad una parte dei valori registrati.⁵⁰ Quando poi il numero di autori degli articoli cresce oltre un certo limite, una situazione comune in settori come la fisica della alte energie dal secondo dopoguerra in poi, l'equazione originaria non è più adeguata ai dati empirici, ed è necessario ricorrere ad una formulazione matematica più generale per conservarne la validità.⁵¹

⁴⁸ L.J. Murphy, "Lotka's law in the humanities?" *Journal of the American Society for Information Science* 1973, 24: 461-62; H. Voos, "Lotka and information science" *Journal of the American Society for Information Science* 1974, 25(4): 270-72; A.E. Schorr, "Lotka's law and map librarianship" *Journal of the American Society for Information Science* 1975, 26: 189-90; A.E. Schorr, "Lotka's law and library science" *RQ. Reference Quarterly* 1974, 14(1): 32-33; A.E. Schorr, "Lotka's law and the history of legal medicine" *Research in Librarianship* 1975, 30: 205-09; T. Radhakrishnan & R. Kernizan, "Lotka's law and computer science literature" *Journal of the American Society for Information Science* 1979, 30(1): 51-54.

⁴⁹ G.B. Newby; J. Greenberg; P. Jones, "Open source software development and Lotka's law: bibliometric patterns in programming" *Journal of the American Society for Information Science and Technology* 2003, 54(2): 169-78.

⁵⁰ W.G. Potter, "Lotka's law revisited" *Library Trends* 1981, 30(1): 21-39. Il test applicato in questo caso è il Kolmogorov-Smirnov.

⁵¹ L. Egghe, "Extension of the general "success breeds success" principle to the case that items can have multiple sources," in *Proceedings of the 5th Biennial Conference of the International Society for Scientometrics and Informetrics*, edited by M. Koenig, et al. (Medford, N.J.: Learned Information, 1995): 147-56; H. Kretschmer & R. Rousseau, "Author inflation leads to a breakdown of Lotka's law" *Journal of the American Society for Information Science and Technology* 2001, 52(8): 610-14.

1.2.2. La legge di Bradford e la legge di concentrazione di Garfield

Samuel Bradford (1878-1948), bibliotecario presso lo Science Museum di Londra, condivideva con Paul Otlet e Henry-Marie LaFontaine il miraggio di un sistema di classificazione in grado di realizzare il controllo universale della documentazione tecnico-scientifica. L'ostacolo principale alla sua costruzione era, secondo lui, l'incapacità dei servizi di *abstracting* e *indexing* dell'epoca di ricondurre la letteratura disponibile ad una griglia omogenea di voci di soggetto. La legge di Bradford nacque dall'insoddisfazione per questo sistema e dal tentativo di approfondirne le cause. Bradford si ispirò ad essa allorché, progettando di trasformare la biblioteca dello Science Museum in un centro nazionale di supporto per la documentazione scientifica, gettò le basi del moderno concetto di document delivery.

La legge fu elaborata a partire da un'analisi statistica di due bibliografie di geofisica: la *Current Bibliography of Applied Geophysics* e la *Quarterly Bibliography of Lubrication*. Essa afferma che, classificando un insieme di riviste in ordine di produttività decrescente in rapporto al numero di articoli su un dato argomento, si otterrà una sequenza formata da un nucleo di periodici fondamentali, responsabili della maggior parte degli articoli sull'argomento in questione, e diversi altri gruppi contenenti lo stesso numero di articoli del nucleo, ma disseminati in una quantità sempre maggiore di riviste. Il numero di periodici nel nucleo e nei successivi gruppi è espresso dalla proporzione:

$$1:n:n^2 \dots$$

La correlazione traduce in termini matematici un dato abbastanza intuitivo e ben noto ai bibliotecari addetti agli acquisti di materiale periodico: dove esiste un'area d'indagine specialistica, un numero limitato di riviste produce il 90% circa della letteratura essenziale alla sua sopravvivenza, mentre, se si vuole

ottenere il 100% del controllo bibliografico, bisogna aggiungere al nucleo iniziale un numero molto maggiore (che cresce in modo esponenziale) di riviste di aree disciplinari più o meno contigue. «Per la stessa ragione - sosteneva Bradford - le biblioteche specializzate non possono raccogliere la letteratura completa del loro settore [...] Un terzo dei materiali di una biblioteca specializzata è di solito correlato in maniera precisa con il loro scopo, i restanti due terzi comprendono invece letteratura su argomenti di confine e meno pertinenti».⁵²

Consegnata alla comunità degli studiosi, anche la legge di Bradford ha subito, in virtù della sua stessa generalità, analisi, estensioni, contestazioni, comparazioni e fusioni con altre leggi. È una legge ambigua soprattutto perché ne sono state date interpretazioni matematicamente diverse, e la sua corrispondenza con i dati empirici dipende in larga misura dalla scelta e dal trattamento del campione iniziale.⁵³

Una prima conferma arrivò negli anni '50. Conteggiando i prestiti di materiale periodico effettuati nel 1956 dalla biblioteca dello Science Museum, Donald Urquhart trovò che meno del 10% delle riviste erano sufficienti a rendere conto dell'80% circa del materiale richiesto e formulò una legge in base alla quale il numero di prestiti interbibliotecari di un periodico è una misura attendibile del suo utilizzo globale.⁵⁴

Garfield, estrapolando i risultati di uno studio condotto su un campione di riviste dello *Science Citation Index* (1969) e del *Current Abstracts of Chemistry and Index Chemicus* (1974), ha ampliato il raggio di applicazione della legge di Bradford e ne ha esteso la validità alla scienza nella sua totalità, concepita come un insieme di aree disciplinari che si intersecano e sovrappongono in diversi punti. La legge di Bradford rappresentava la

⁵² S.C. Bradford, "Sources of information..." cit., p. 86.

⁵³ E.A. Wilkinson, "The ambiguity of Bradford's law" *Journal of Documentation* 1972, 28(2): 122-30. Per un inquadramento teorico ed una difesa della validità ed estendibilità della legge al di là dei confini bibliografici vedi: B.C. Brookes, "Theory of the Bradford law" *Journal of Documentation* 1977, 33(3): 173-250.

letteratura scientifica di una disciplina come una cometa: il nucleo contiene le poche riviste fondamentali, la coda comprende le riviste che sporadicamente pubblicano articoli importanti nello stesso settore. La legge di concentrazione di Garfield afferma che la coda della letteratura di una disciplina è fatta, in gran parte, dai nuclei documentari delle altre discipline. «Talmente ampia infatti è la sovrapposizione tra discipline, che la letteratura del nucleo di tutte le discipline scientifiche coinvolge un gruppo di non più di 1000 riviste, e può arrivare a comprenderne anche solo 500».⁵⁵ Contro la percezione comune dello scienziato sommerso da una marea di informazione disseminata in un numero sempre crescente di riviste scientifiche, la legge di Bradford e la legge di concentrazione di Garfield rassicuravano del fatto che «un numero limitato di riviste produce la maggior parte dei risultati scientifici degni di nota».⁵⁶ In tempi più recenti, l'ISI ha richiamato questa legge per legittimare la validità dei propri criteri di selezione della letteratura periodica, in virtù dei quali solo il 10-12% dei circa 2000 periodici candidati annualmente all'inclusione nel repertorio supera l'esame di ammissione.⁵⁷ Confidando nella validità della stessa legge, Garfield ha potuto così sostenere che, dei 4500 periodici coperti nel 1996 dallo *Science Citation Index* e dal *Social Science Citation Index*,

⁵⁴ S.J. Bensman, "Urquhart's and Garfield's Laws: The British controversy over their validity" *Journal of the American Society for Information Science and Technology* 2001, 52(9): 714-24.

⁵⁵ E. Garfield, *Citation Indexing: its Theory and Application in Science, Technology, and Humanities* (Philadelphia: ISI Press, 1983), p. 23. Preannunciata in: E. Garfield, "The mystery of the transposed journal lists wherein Bradford's law of scattering is generalized according to Garfield's law of concentration," in *Essays of an Information Scientist 1962-1973*, 1 (Philadelphia: ISI Press, 1977): 222-23, la legge di concentrazione fu successivamente discussa in: E. Garfield, "Bradford's Law and related statistical patterns," in *Essays of an Information Scientist 1979-1980*, 4 (Philadelphia: ISI Press, 1981): 476-83. Le edizioni originali dei due articoli citati risalgono rispettivamente al 1971 e al 1980.

⁵⁶ «It has been five years since I reported on the "journal myth" – that is, the popular perception that researchers are being overwhelmed by an ever-increasing flood of scientific journals. Experience has taught me it is essential that a simple message be repeated regularly: A small number of journals accounts for the bulk of significant scientific results»: E. Garfield, "The significant scientific literature appears in a small core of journals" *The Scientist* 1996, 10(17): 13-16

<[http://www.garfield.library.upenn.edu/commentaries/tsv10\(17\)p13y090296.html](http://www.garfield.library.upenn.edu/commentaries/tsv10(17)p13y090296.html)>, p. 13. L'articolo precedente cui Garfield si riferisce: E. Garfield, "In truth, the 'flood' of scientific literature is only a myth" *The Scientist* 1991, 5(17) <[http://www.garfield.library.upenn.edu/commentaries/tsv05\(17\)y19910902.html](http://www.garfield.library.upenn.edu/commentaries/tsv05(17)y19910902.html)>.

⁵⁷ Thomson ISI, *The ISI® database: The journal selection process*, 2005 <<http://www.isinet.com/essays/selectionofmaterialforcoverage/199701.html>>.

«500 rendono conto del 50% di ciò che viene pubblicato e del 75% di ciò che viene citato».⁵⁸

1.2.3. La legge di Zipf

George K. Zipf (1902-1950), docente di linguistica ad Harvard, era convinto che fosse possibile «studiare il linguaggio come un fenomeno naturale [...] alla maniera delle scienze esatte, attraverso l'applicazione diretta di principi statistici ai fenomeni linguistici oggettivi».⁵⁹ Poco importa che egli avesse poca familiarità con la matematica ed intendesse giustificare tale impresa mediante una qualche forza mistica operante in natura: la storia della scienza mostra tanti esempi di programmi di ricerca e risultati scientifici validi nati da (o innestati in) un complesso discutibile di credenze metafisiche.

Zipf osservò l'andamento iperbolico di molti fenomeni del linguaggio, giungendo alla conclusione che la chiave per comprenderne gli aspetti essenziali sia una tendenza statisticamente definita a mantenere l'equilibrio tra dimensioni e frequenza degli elementi che lo compongono. L'equilibrio è un caso particolare del principio che egli poneva all'origine di tutte le azioni umane e collettive: il principio del minimo sforzo.⁶⁰ La *legge di Zipf*, ispirata dal lavoro dello stenografo francese Estoup, afferma che, in un testo relativamente lungo, classificando le parole in ordine decrescente di frequenza, la posizione occupata da una parola nella classifica moltiplicata per la sua frequenza di apparizione è uguale ad una costante. In simboli:

$$r * f = k$$

⁵⁸ E. Garfield, "How can impact factors be improved?" *BMJ* 1996, 313(7054): 411-3
<[http://www.garfield.library.upenn.edu/papers/bmj313\(7054\)p411y1996.html](http://www.garfield.library.upenn.edu/papers/bmj313(7054)p411y1996.html)>, p. 413.

⁵⁹ G.K. Zipf, *The Psycho-Biology of Language: An Introduction to Dinamic Philology* (London: Routledge, 1936), prefazione, p. v.

⁶⁰ G.K. Zipf, *Human Behavior and the Principle of least Effort. An Introduction to Human Ecology* (Reading, Mass: Addison Wesley Press, 1949).

dove r è la posizione della parola nella classifica, f la sua frequenza di apparizione, k la costante.

Analizzando un indice delle parole dell'*Ulisse* di Joyce, Zipf ha scoperto, ad esempio, che la decima parola in ordine di frequenza era menzionata 2653 volte, la centesima 265 volte, la duecentesima 133 volte e così via. Esiste, nella distribuzione delle parole di un testo, la stessa asimmetria che caratterizza gli autori (Lotka) e le riviste (Bradford): poche parole molto frequenti ed un numero di gran lunga più alto di parole meno frequenti. La legge di Zipf ha subito diversi tentativi di generalizzazione e fondazione matematica rigorosa, tra cui quelli di Benoît Mandelbrot, che impiegò la teoria dell'informazione di Shannon per spiegare la relazione ordine/frequenza delle parole. La maggior parte delle reinterpretazioni è di tipo probabilistico, punta cioè a calcolare, date certe condizioni iniziali, la probabilità con cui una parola comparirà con una determinata frequenza in un corpus arbitrario. Resta però un enigma di fondo: sebbene diversi modelli matematici possano adattarsi alla distribuzione di Zipf, «siamo ancora lontani dal capire perché esista e perché sia così diffusa». ⁶¹

1.2.4. L'unificazione delle leggi della bibliometria

Un articolo molto citato ha più probabilità di ricevere altre citazioni rispetto ad un articolo mai o poco citato. Un autore che ha scritto molti articoli ha più probabilità di scriverne altri rispetto ad uno che non ha mai scritto niente o ha scritto poco. Una rivista molto consultata ha più probabilità di essere consultata di un'altra mai sfogliata. Un miliardario ha più probabilità di aumentare il suo capitale di quante ne abbia un impiegato. È abbastanza

⁶¹ R.E. Wyllys, "Empirical and theoretical bases of Zipf's law" *Library Trends* 1981, 30(1): 53-64, p. 62.

comune, lavorando su dati bibliometrici o socio-economici, imbattersi nella legge secondo la quale chi ha già avuto avrà ancora, il "successo genera successo" (*success breeds success*). Ma per dare un fondamento teorico al trattamento quantitativo dell'informazione non basta scoprire che esistono delle regolarità negli aspetti formali della documentazione. Il fatto che eventi simili si ripetano un certo numero di volte non garantisce che continueranno a verificarsi in futuro nello stesso modo: una legge scientifica non nasce dalla generalizzazione induttiva di casi particolari.⁶² Servono ipotesi capaci di mettere ordine nella complessità dei fenomeni, di spiegarne le cause e predirne il comportamento futuro al di là delle circostanze particolari in cui sono stati descritti la prima volta.

Il passo decisivo per le sorti della bibliometria è stato compiuto quando, a partire dalla fine degli anni '60, si è cercato di dimostrare che le distribuzioni di Lotka, Bradford, Zipf sono matematicamente equivalenti e discendono da uno stesso principio universale. Negli scritti di Fairthorne, Price, Brookes, Bookstein, Egghe, il principio dei vantaggi cumulativi, l'idea che "il successo genera successo" si traduce in un set manipolabile di funzioni matematiche capaci di descrivere un vasto assortimento di fenomeni bibliografici e sociali.⁶³ Le equazioni di Lotka, Bradford, Zipf sono divenute così varianti di una medesima distribuzione probabilistica, descritta da una legge di potenza, nella quale una variabile indipendente che aumenta in progressione geometrica genera una variabile dipendente che cresce in progressione aritmetica.

⁶² R. Carnap, *I fondamenti filosofici della fisica: Introduzione alla filosofia della scienza* (Milano: Il Saggiatore, 1971); K.R. Popper, *Logica della scoperta scientifica: Il carattere autocorrettivo della scienza* (Torino: Einaudi, 1970).

⁶³ R.A. Fairthorne, "Empirical hyperbolic distributions (Bradford-Zipf-Mandelbrot) for bibliometric description and prediction" *Journal of Documentation* 1969, 25: 319-43; D.J.d.S. Price, "A general theory of bibliometric and other cumulative advantage processes" *Journal of the American Society for Information Science* 1976, 27: 292-306; B.C. Brookes, "Theory of the Bradford law" *Journal of Documentation* 1977, 33(3): 173-250; A. Bookstein, "Informetric distributions, Part I: Unified Overview" *Journal of the American Society for Information Science* 1990, 41(5): 368-75; A. Bookstein, "Informetric distributions, Part II: Resilience to ambiguity" *Journal of the American Society for Information Science* 1990, 41(5): 376-86; L. Egghe, "Consequences of Lotka's law for the law of Bradford" *Journal of Documentation* 1985, 41: 173-89; L. Egghe & R. Rousseau, "Generalized success-breeds-success principle leading to time-dependent informetric distributions" *Journal of the American Society for Information Science* 1995, 46(6): 426-45.

Oggi le leggi di potenza sono al centro dei modelli matematici di analisi dei processi informativi,⁶⁴ e più avanti se ne constaterà il revival nella nuova scienza delle reti (capitolo 6). La loro ubiquità nondimeno ha sempre destato dei sospetti. La distribuzione dei redditi, delle popolazioni nei centri abitati, delle frequenze di accesso alle pagine web, delle note in una performance musicale, delle specie biologiche, del diametro dei crateri lunari si adattano a questo tipo di distribuzione al pari di molte altre serie di eventi sociali o naturali. L'onnipresenza ha qualcosa di mistico, al punto che è facile imbattersi in formule che accostano le distribuzioni iperboliche alla sezione aurea, alla geometria della piramide di Cheope e alla simmetria della molecola dell'acqua considerandole tutte manifestazioni convergenti di una legge superiore che regola l'armonia dell'universo.⁶⁵

Le leggi bibliometriche tuttavia descrivono un insieme di regolarità osservabili in un campione limitato di letteratura e permettono, entro certi limiti, di predire l'andamento futuro di una data variabile, ma non spiegano in alcun modo perché tali regolarità si verificano, quali sono i meccanismi e le cause sottostanti.⁶⁶ «I matematici - commentava lo statistico Gustav Herdan - credono nella legge di Zipf perché pensano che i linguisti l'abbiano dimostrata come legge linguistica, e i linguisti, dal canto loro, ci credono perché pensano che i matematici l'abbiano dimostrata come legge matematica».⁶⁷ Potere predittivo e potere esplicativo, in fondo, non sono complementari: per secoli l'astronomia ha effettuato previsioni molto accurate senza porsi il problema di spiegare la struttura fisica e il funzionamento meccanico dell'universo, ma limitandosi a "salvare i fenomeni", cioè a predire le posizioni apparenti dei

⁶⁴ L. Egghe, *Power Laws in the Information Production Process: Lotkian Informetrics* (Amsterdam: Elsevier Academic Press, 2005).

⁶⁵ Per esempio: C. Lange, *L'angolo aureo*, 2005 <<http://www.sectioaurea.com/sectioaurea/angoloaureo.htm>>.

⁶⁶ A. Rapoport, "Rank-size relations," in *International Encyclopedia of Statistics*, edited by W. Kruskal, et al. (New York: The Free Press, 1978): 847-54; D.O. O'Connor & H. Voos, "Empirical laws, theory construction and bibliometrics" *Library Trends* 1981, 30(1): 9-20. In Italia l'eco delle critiche al potere esplicativo delle leggi bibliometriche si trova in: A. Serrai, *Dai "loci communes" alla bibliometria* (Roma: Bulzoni, 1984), pp. 199-218.

⁶⁷ Citato in: R.E. Wyllys, "Empirical..." cit., p. 53.

pianeti nella sfera celeste mediante combinazioni geometriche di moti circolari uniformi. Le leggi di Lotka, Bradford, Zipf non spiegano perché un autore scrive più di un altro, perché una rivista pubblica più contributi fondamentali in una disciplina di un'altra, perché certe parole di un testo sono più frequenti di altre, ma aiutano i bibliotecari a progettare regole di catalogazione adeguate alle tipologie di *authorship* (Lotka), a scegliere i periodici cruciali per una disciplina (Bradford), a progettare sistemi più efficienti di *information retrieval* (Zipf). Il loro significato per la storia degli indici di citazioni non si colloca però a questo livello. A prescindere dal valore esplicativo e predittivo, le leggi bibliometriche invocano un principio che si rivelerà strategico per gli usi extra-bibliografici delle citazioni: si tratti del numero di pubblicazioni per scienziato, del numero di articoli per rivista o del numero di parole significative per testo scritto, appare probabilisticamente fondata la tesi che, nell'organizzazione bibliografica della conoscenza umana, pochi elementi sono sufficienti a rendere conto di un complesso costituito da molti elementi.

1.3. Dalla bibliografia statistica alla biblio(sciento)-metria

La bibliografia statistica è solo una delle tradizioni intellettuali che si incrociano nello studio quantitativo della scienza. Le rudimentali statistiche assemblate o inventate, a partire dalla seconda metà dell'800, da Charles Babbage, Julian Huxley e John D. Bernal a supporto delle loro rivendicazioni politiche, rientrano di diritto nella catena degli antecedenti della scientometria, così come vi rientrano, sotto una veste più colta, la sociologia positivista di August Comte, Herbert Spencer e William Ogburn, le indagini sul genio scientifico di Alphonse de Candolle e Francis Galton, gli studi di Georges

Sarton e Pitrim Sorokin sulla distribuzione del genio scientifico nelle varie epoche.⁶⁸

La scienza, in effetti, si può misurare in diversi modi, applicando metodi statistici non solo alle pubblicazioni, ma a qualunque tipo di dati quantitativi, compresi gli aspetti economici, sociali, biografici. Prevale tuttavia tra gli scientometri la convinzione che, se uno scienziato ha qualcosa di importante da dire, pubblica o cerca di pubblicare le sue scoperte su una rivista internazionale. L'articolo scientifico *peer reviewed* è considerato l'apice del lavoro creativo degli scienziati: ne legittima le rivendicazioni conoscitive e preannuncia la ricompensa istituzionale dei loro sforzi. Ecco perché, come dimostra l'orientamento editoriale della rivista *Scientometrics*, una buona parte della ricerca scientometrica è di natura bibliometrica, si fonda cioè sull'applicazione della matematica e di metodi statistici ai prodotti della comunicazione scritta ed in particolar modo agli articoli di riviste. Per questo motivo si parlerà indifferentemente, d'ora in avanti, di bibliometria e di scientometria con l'intento di limitare la discussione unicamente a questa tradizione di studi.

L'idea della bibliografia statistica di misurare la documentazione scientifica rimase nell'ombra per diversi anni. Il suo recupero, con la conseguente riscoperta della letteratura precedente sull'argomento, avvenne solo dopo la seconda guerra mondiale.⁶⁹

«Possiamo affermare - scriveva Derek John de Solla Price nel 1963 - che l'80-90% degli scienziati mai vissuti è viva adesso».⁷⁰ La scienza appariva in

⁶⁸ A. Thackray, "Measurement in the Historiography of Science," in *Toward a Metric of Science: the Advent of Science Indicators*, edited by Y. Elkana, et al. (New York: John Wiley and Sons, 1979): 11-30.

⁶⁹ Per le brevi note sulla storia della scientometria presentate in questo paragrafo mi sono servito in particolare di: L. Leydesdorff, *The Evaluation of Research and the Scientometric Research Program: Historical Evolution and Redefinitions of the Relationship*, 2004
<<http://users.fmg.uva.nl/lleydesdorff/sss04/>>; B. Godin, *Outline for a History of Science Measurement*, 2000
<http://www.csiic.ca/PDF/Godin_1.pdf>.

⁷⁰ D.J.d.S. Price, *Little Science, Big Science...and Beyond* (New York: Columbia University Press, 1986), p. 1. L'edizione originale risale al 1963 ed è il risultato delle Pegram Lectures tenute da Price a Brookhaven l'anno precedente. Esiste una traduzione italiana di questa prima edizione: D.J.d.S. Price, *Sociologia della*

quegli anni un'impresa essenzialmente cumulativa, collaborativa e soprattutto contemporanea, in cui l'essenziale era accaduto da Einstein in poi, e l'intero processo era stato graduale ma inesorabile. La fiducia positivista nella capacità della scienza di conoscere e trasformare il mondo ne legittimava la percezione come forma di conoscenza radicalmente diversa dall'arte, dalla letteratura, dalla storiografia. Il fatalismo insito in questa percezione rendeva per altri versi l'attività scientifica un oggetto dal comportamento prevedibile e in linea di principio controllabile, al punto da ritenere che «se Michelangelo o Beethoven non fossero esistiti, i loro capolavori sarebbero stati sostituiti da qualcosa di affatto diverso. Se Copernico e Fermi non fossero mai esistiti, essenzialmente gli stessi contributi sarebbero arrivati da altre persone. C'è, in effetti, solo un mondo da scoprire ...». ⁷¹ La lunga lista di scoperte multiple indipendenti, cioè scoperte simili fatte, simultaneamente o in tempi diversi, da scienziati che lavoravano ognuno per proprio conto senza conoscere l'operato dei colleghi o predecessori, era una conferma dell'inevitabilità del progresso scientifico: date certe condizioni ambientali, alcune acquisizioni teoriche sembravano destinate a verificarsi prima o poi, a prescindere dal fatto che il merito andasse ascrivito ad un individuo particolarmente geniale o ad un gruppo di persone meno geniali ma funzionalmente equivalenti al genio individuale. ⁷²

La guerra aveva mostrato che scienza e tecnologia potevano essere controllate, manipolate e indirizzate verso obiettivi specifici. Inoltre era sempre più chiaro che la crescita economica ininterrotta dei paesi occidentali non poteva essere spiegata solo con i fattori economici tradizionali della terra, del lavoro, del capitale. L'impatto militare della scienza e della tecnologia lo

creatività scientifica (Milano: Bompiani, 1967). I passi citati (e tradotti) nel seguito si riferiscono tuttavia all'edizione del 1986.

⁷¹ D.J.d.S. Price, *Little Science...cit.*, p. 62.

⁷² La teoria sociologica del genio scientifico come equivalente funzionale di un gran numero di altri scienziati fu sviluppata da Robert Merton. Vedi: R.K. Merton, "Scoperte singole e multiple nella scienza (1961)," in *La sociologia della scienza. Indagini teoriche ed empiriche*, edited by N.W. Storer (Milano: Franco Angeli, 1981): 436-66.

confermava: il valore aggiunto andava cercato nei processi di sviluppo e innovazione azionati dalle scienze di base.

La preoccupazione iniziale della classe politica non si rivolse tuttavia all'output, ai risultati effettivamente conseguiti, bensì all'input, alle risorse umane e finanziarie investite, come si evince dagli standard di monitoraggio statistico delle attività di ricerca e sviluppo sponsorizzati, a partire dal 1963, dall'OECD (Organization for Economic Cooperation and Development). Era implicita in compilazioni come il *Frascati Manual* l'idea, già espressa con chiarezza nel famoso report di Vannevar Bush *The Endless Frontier*, che l'autonomia istituzionale della scienza andasse rispettata in virtù della sua capacità di promuovere automaticamente il progresso economico-sociale, e che il governo dovesse limitarsi a finanziare la ricerca di base, senza preoccuparsi delle misure di impatto e produttività. Il resto lo avrebbero fatto gli scienziati da soli, affidando il controllo di qualità al vecchio e collaudato sistema di *peer reviewing*. Il *Frascati Manual* e gli affini *Oslo manual* e *Canberra Manual* disdegnavano gli indicatori bibliometrici con la motivazione che ciò di cui essi dovevano essere una rappresentazione semplificata, cioè le complesse relazioni tra scienza, tecnologia, economia e società, era troppo complesso da rappresentare, data anche la carenza di linee guida internazionali al riguardo.⁷³

Dopo il lancio dello Sputnik da parte dell'Unione Sovietica nel 1957, si ebbe un drastico cambiamento di rotta. Era chiaro che il modello liberale statunitense non aveva saputo adeguarsi al nuovo scenario con la stessa rapidità di quello *non-liberal* del blocco contrapposto. L'Unione Sovietica infatti padroneggiava da tempo un apparato burocratico di controllo e orientamento della ricerca scientifica. La scientometria, del resto, è in gran parte un'invenzione sovietica. Lo statistico russo Nalimov aveva elaborato, già verso la fine degli anni '50, dei modelli matematici di rappresentazione della

⁷³ E. Spinak, "Indicadores cientometricos" *Ciência da informação* 1998, 27(2): 141-48.

crescita scientifica mondiale e subito dopo aveva assistito al decollo dello *Science Citation Index*, suggerendo per questo tipo di ricerche quantitative il termine *naukometriya* (scientometria) nel 1966 e nel 1969.⁷⁴

Il primato russo è abbastanza comprensibile: in regime di pianificazione economica e finanziaria, misurare la ricerca scientifica e tecnologica in termini di documentazione prodotta e “usata” era un modo per consolidarne l’assetto burocratico e il controllo centralizzato, un strategia di *science management* più sofisticata di quella amministrativa. Non a caso Nalimov vide nello *Science Citation Index* una possibilità concreta di realizzazione del progetto scientometrico, salutando l’avvento di una nuova epoca della storia della scienza e paragonando l’introduzione del repertorio di Garfield alla comparsa delle prime cronache che avevano reso possibile la storia documentata delle società umane.⁷⁵

La risposta degli Stati Uniti al lancio dello Sputnik e alle minacce di aggressione militare si concretizzò, sul piano istituzionale, nella creazione della NASA (National Aeronautics and Space Administration), dell'ARPA (Advanced Research Project Agency) e dell'OECD. Quest'ultima, nata nel 1961 dalle ceneri del Piano Marshall, aveva il compito di coordinare le politiche nazionali in materia di scienza e tecnologia dei paesi industrializzati, ma la sua principale preoccupazione era e restava di tipo economico. La

⁷⁴ Un resoconto dettagliato della scientometria russa negli anni '60-70 si trova in: L.G. Gurjeva, *Early Soviet Scientometrics and Scientometricians*. Thesis for the degree of MSc in Science Dynamics: Universiteit van Amsterdam, 1992). Sul contributo di Nalimov, in particolare: Y.V. Granovsky, “Is it possible to measure science? V. V. Nalimov’s research in scientometrics” *Scientometrics* 2001, 52(2): 127-50.

⁷⁵ «...with the advent of SCI the documented history of science began in the same way as with the advent of chronicles the documented history of society began»: V.V. Nalimov & B.M. Mulchenko, *Measurement of Science. Study of the Development of Science as an Information Process* (Washington, DC: Foreign Technology Division, US Air Force System Command, 1971)

<<http://garfield.library.upenn.edu/nalimovmeasurementofscience/book.pdf>>, p. 96. Lo stesso Garfield, in occasione del 50° anniversario del VINITI (Vsesojuznyi Institut Naucnoj I Tehnic'eskoj Informac'ii, l'agenzia russa per l'informazione tecnico-scientifica), ha ricordato i molteplici legami tra l'ISI di Philadelphia ed il VINITI, elogiando in particolare l'approccio multidisciplinare all'indicizzazione della letteratura scientifica introdotto dall'agenzia sovietica, poi divenuto un modello di riferimento per la filosofia dell'ISI: E. Garfield, “The ISI-VINITI Connection - Remarks on the 50th anniversary of VINITI,” in *Proceedings of the 6th International Conference of Information Society, Intelligent Information Processing and Information Technologies. October 16-18, 2002* (Moscow, 2002): 409-11

valutazione della ricerca restava confinata nel recinto accademico del sistema disciplinare e della comunicazione specialistica.

Intanto la scienza aveva oltrepassato le mura dell'università intrecciando il suo destino con quello del capitalismo industriale e delle politiche economiche nazionali. Il nuovo assetto industriale dell'impresa scientifica e il conseguente incremento esponenziale della letteratura evidenziavano la necessità di nuovi strumenti di controllo. Il progresso andava sostenuto e, se possibile, accelerato con adeguati strumenti di recupero e disseminazione dell'informazione.

Davanti al volume di documentazione disponibile, alla iper-specializzazione delle aree di ricerca e ai limiti dei sistemi manuali di indicizzazione della letteratura, produttori e consumatori di scienza trovavano infatti sempre più difficile reperire informazioni pertinenti al proprio dominio di indagine. Il rischio di duplicare programmi di ricerca, che la storia delle scoperte multiple rendeva tangibile, comportava potenziale sperpero di denaro. La dimensione manageriale assunta dall'impresa scientifica imponeva, al contrario, scelte amministrative votate alla razionalizzazione degli impegni di spesa.

È proprio qui che comincia la strana storia dello *Science Citation Index* di Eugene Garfield. Dopo il 1961, con l'intermediazione del genetista premio Nobel Joshua Lederberg, il progetto di un indice interdisciplinare della letteratura scientifica guadagnò credibilità politica e sostegno finanziario. Né Garfield né Lederberg tuttavia potevano prevedere il destino cui il repertorio sarebbe andato incontro. Le citazioni bibliografiche legavano documenti e autori secondo una logica interna alle dinamiche intellettuali di produzione della conoscenza, quindi, oltre a facilitare la ricerca della letteratura, offrivano una misura imparziale dell'impatto esercitato dai documenti (e dagli autori) citati sui documenti (e gli autori) citanti. E per come la scienza veniva percepita in quel periodo, l'impatto implicava sempre una spinta in avanti, mai un arresto improvviso o un'esitazione.

L'uso degli indicatori bibliometrici nella valutazione delle attività di ricerca si affermò, a partire dagli anni '60-'70, dapprima negli Stati Uniti, poi in Europa, e subì sin dall'inizio il fascino discreto degli indici di citazioni.

Nel 1973 vide la luce il primo di una serie di *Science Indicators Reports* a cura del National Science Board, con l'obiettivo di descrivere quantitativamente la condizione della ricerca scientifica negli USA. I *Reports* annoveravano il calcolo delle citazioni tra gli indicatori qualitativi e facevano uso dei dati ricavati dallo *Science Citation Index* di Garfield. La loro risonanza, nel bene e nel male, fu tale da spingere intellettuali di diversa provenienza ed estrazione (storici, sociologi, politici ed economisti) ad interrogarsi sul significato e sulla portata degli indicatori scientometrici in una conferenza, tenuta a Standord nel giugno 1974, dalla quale nacque il volume *Toward a Metric of Science*, una delle fonti principali di ispirazione dei bibliometri successivi.⁷⁶

Analoghi repertori con inclinazione verso i *citation indexes* apparvero negli anni seguenti al di fuori degli Stati Uniti, in Canada, in Australia, in Europa (ad esempio l'*Indicateurs de sciences et de technologies* francese, il *Wetenschaps- en Technologie-Indicatoren* olandese, il *Vlaams Indicatorenboek* belga, l'*European Report on S&T Indicators*). Governi e istituzioni cominciarono a finanziare progetti di ricerca mirati a perfezionare gli strumenti di gestione e diffusione dell'informazione scientifica, promuovendo al contempo lo studio degli indicatori bibliometrici come supporto alle decisioni in merito alla politica della scienza ed ai finanziamenti della ricerca.

Nel 1976 fu pubblicato, sotto gli auspici della National Science Foundation, il lavoro di Francis Narin *Evaluative Bibliometrics*, che identificava nell'impatto a livello internazionale ricavabile dall'analisi delle citazioni il parametro

⁷⁶ Questo impatto è documentato da: Y. Elkana; J. Lederberg; R.K. Merton et al. (Ed.), *Toward a Metric of Science: the Advent of Science Indicators* (New York: John Wiley and Sons, 1979).

cruciale per l'accertamento della qualità scientifica.⁷⁷ Negli stessi anni Henry Small avviò il progetto di "cartografare" la struttura disciplinare della scienza classificandone la letteratura attraverso la tecnica delle co-citazioni.

Gli studi americani si focalizzarono all'inizio soprattutto sull'organizzazione della letteratura scientifica, mentre in alcune nazioni europee si cominciò a sperimentare l'applicazione degli indicatori bibliometrici alla valutazione delle performance di università, istituti e gruppi di ricerca. Sull'esempio di Ben Martin e John Irvine e del loro studio dedicato alla misurazione dell'efficienza di quattro osservatori radioastronomici,⁷⁸ si distinse in questo lavoro pionieristico l'Università di Leida, il cui modello di analisi fu generalizzato, nei primi anni '90, dal programma britannico *Research Assessment Exercises* per il finanziamento della ricerca universitaria.⁷⁹

Nel frattempo, dalla metà degli anni '70 in poi, grazie anche al rapido sviluppo delle tecnologie informatiche e alla disponibilità crescente di database bibliografici in formato elettronico, gli studi quantitativi della scienza si sono evoluti in una disciplina a sè stante, con tanto di manualistica e dizionari specializzati (il primo *Handbook of Quantitative Studies of Science and Technology* è del 1988), riviste (*Scientometrics* dal 1978, *Research Evaluation* dal 1991, ma anche il *Journal of the American Society of Information Science* e il *Journal of Information Science*), conferenze tematiche (dal 1983), società scientifiche (l'*International Society for Scientometrics and Informetrics* dal 1993), cattedre universitarie.

Il paradosso di questa disciplina, ciò che la fa sembrare qualcosa di meno e qualcosa di più delle altre forme istituzionalizzate di conoscenza, è la sua originaria interdisciplinarietà. Al di fuori della comune premessa circa l'esistenza di un qualcosa dai contorni ben definiti come "la scienza" e circa la

⁷⁷ F. Narin, *Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity* (Cherry Hill, NJ: Computer Horizons, 1976).

⁷⁸ B.R. Martin & J. Irvine, "Assessing basic research. Some partial indicators of scientific progress in radio astronomy" *Research Policy* 1983, 12: 61-90.

misurabilità di vari aspetti di questo qualcosa (premesse non banali dal punto di vista epistemologico), non esiste infatti accordo su una teoria della scienza che faccia da guida nella costruzione degli indicatori nè sulla metodologia di misura da adottare. La procedura normalmente seguita perciò è di tipo empirico: si raccolgono e si elaborano i dati senza partire da una cornice teorica predefinita, ma con una serie di tacite idee su come vanno le cose (*how the things works*), poi si cerca un modello esplicativo dei risultati.⁸⁰

Le attuali ricerche bibliometriche rivelano una tale diversità di provenienza, scopo, impostazione metodologica da rendere praticamente impossibile la loro collocazione in in una cornice teoretica unitaria. Schematizzando si possono distinguere una bibliometria per i bibliometri, fatta prevalentemente di riflessioni sullo statuto metodologico della disciplina, una bibliometria per le discipline scientifiche e storico-sociali, orientata alle diverse modalità di trattamento quantitativo dell'informazione specifica di ciascuna disciplina, ed una bibliometria per la politica ed il management della scienza, di gran lunga l'area più delicata, dove i calcoli sono finalizzati alla valutazione comparativa di scienziati, istituzioni, nazioni.⁸¹

⁷⁹ H.F. Moed; W.J.M. Burger; J.G. Frankfort et al., "The use of bibliometric data for the measurement of university research performance" *Research Policy* 1985, 14: 131-49.

⁸⁰ A.F.J. van Raan, "Measuring science. Capita selecta of current main issues," in *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems*, edited by H.F. Moed, et al. (Dordrecht: Kluwer Academic, 2004): 19-50.

⁸¹ Ho trovato questa schematizzazione in: W. Glänzel, *Bibliometrics as a Research Field. A Course on Theory and Application of Bibliometric Indicators*, 2003
<http://www.norslis.net/2004/Bib_Module_KUL.pdf>.

Capitolo 2

Le basi teoriche della scientometria

Dopo l'introduzione dello *Science Citation Index*, che forniva gran parte dei dati grezzi necessari per la costruzione di modelli bibliometrici, e la pubblicazione dei volumi *Science since Babylon* (1961) e *Little Science Big Science* (1963) di Derek John de Solla Price, la nuova area d'indagine ebbe a disposizione un vasto arsenale empirico e concettuale. Nel 1962 usciva inoltre *La struttura delle rivoluzioni scientifiche* di Thomas Kuhn (1922-1996), un libro completamente diverso dai precedenti, ma fondamentale per il futuro degli studi scientometrici perché insegnò a guardare alla scienza come ad un fenomeno storicamente circoscritto, non riducibile al solo aspetto metodologico e linguistico, governato da dinamiche sociali interne alle singole discipline.

Sul piano teorico, i contributi di Derek John de Solla Price (1922-1983), al pari delle analisi sociologiche di Robert K. Merton (1910-2003), furono decisivi per le sorti della scientometria. Entrambi, come in seguito Garfield, riconobbero tuttavia nell'opera dello scienziato irlandese John D. Bernal (1901-1971) alcune delle premesse teoriche della nuova area d'indagine. Bernal era un cristallografo di fama internazionale il cui impegno politico e culturale, manifestato soprattutto attraverso l'interesse per i problemi della comunicazione e della documentazione scientifica, ne oscurarono ben presto la fama di scienziato. La scienza era per lui un'attività sociale intimamente connessa all'intera gamma delle altre attività umane, e tale correlazione discendeva dall'estensione del materialismo dialettico di Engels allo studio della natura e della società. A differenza dei marxisti ortodossi, però, egli non riduceva la scienza a riflesso meccanico di un contesto sociale, anzi la

considerava il fondamento di ogni conoscenza. Il suo *The Social Function of Science* (1939) divenne, nel periodo bellico e post-bellico, la bibbia della *red science*, la sinistra scientifica britannica e internazionale. Bernal vi prefigurava una rivoluzione delle strutture della comunicazione scientifica resa necessaria dall'ingovernabilità della mole di documentazione prodotta quotidianamente dai ricercatori. La rivoluzione doveva attuarsi in due passaggi: primo, destabilizzare il monopolio sulla letteratura periodica esercitato dalle società scientifiche, con i loro interessi particolari (in un appunto manoscritto si legge: «communication chaotic, scientific periodicals to be abolished»);⁸² secondo, creare un sistema centralizzato di raccolta, archiviazione, coordinamento e distribuzione dell'informazione scientifica nel quale «tutta l'informazione rilevante deve essere disponibile per ciascun ricercatore in ampiezza proporzionale al suo grado di rilevanza [...] non solo l'informazione deve essere disponibile, ma deve essere messa a disposizione del ricercatore senza che egli debba fare dei passi speciali per entrarne in possesso».⁸³

La prima previsione non si sarebbe mai attuata. La seconda divenne la missione commerciale dei *citation indexes* dell'ISI.⁸⁴

⁸² Citato in: D. Muddiman, "Red information scientist: the information career of J.D. Bernal" *Journal of Documentation* 2003, 59(4): 387-409, p. 391.

⁸³ J.D. Bernal, *The Social Function of Science* (Cambridge, MA: M.I.T. Press, 1967), p. 294.

⁸⁴ Un'esplicita connessione tra lo scopo dei *citation indexes* e le proposte di Bernal si trova in: E. Garfield, "Citation indexes - new paths to scientific knowledge" *The Chemical Bulletin* 1956, 43(4): 11 <<http://www.garfield.library.upenn.edu/papers/31.html>>. Nel 1964 Bernal divenne consulente dell'ISI. L'anno seguente scrisse una acuta recensione dello *Science Citation Index* ristampata da Garfield in: J.D. Bernal, "Review of the *Science Citation Index*," in *Essays of an Information Scientist 1981-1982*, 5 (Philadelphia: ISI Press, 1983): 521-23. Garfield ha commemorato il maestro in: E. Garfield, "J.D. Bernal—The sage of Cambridge. 4S Award memorializes his contributions to the social studies of science," in *Essays of an Information Scientist 1981-1982*, 5 (Philadelphia: ISI Press, 1983): 511-20.

2.1. Price e la scienza della scienza

Derek John de Solla Price, docente alla Yale University dal 1959, era uno storico britannico della scienza e della tecnologia con una solida formazione fisico-matematica. Le sue idee sulla natura e l'evoluzione della conoscenza scientifica, derivate in gran parte dalla sociologia di Merton, da una rilettura delle opere di Bernal e da un'attenta riflessione sui dati ricavati dagli indici di Garfield, poggiano su una premessa fondamentale, una specie di "riduzionismo bibliometrico" che contraddistingue l'intera tradizione di studi quantitativi:⁸⁵ la scienza si identifica con la letteratura scientifica pubblicata; uno scienziato non si riconosce dal fatto che ha ricevuto una certa formazione o lavora in un'istituzione etichettata come "scientifica", bensì dal fatto che, almeno una volta nella sua vita, ha prodotto una pubblicazione letta e approvata dalla comunità dei colleghi.

Al VI Congresso Internazionale di Storia della Scienza (1950), Price formulò con chiarezza la tesi che l'enumerazione, classificazione e rappresentazione sotto forma di serie temporali degli articoli di riviste fornisce un indicatore efficace del livello di attività e produzione scientifica.⁸⁶ Dall'osservazione e dal confronto di tali serie temporali egli progettava di scoprire le leggi quantitative di crescita della conoscenza. Gli storici si erano fino ad allora limitati ad investigare gli aspetti interni, qualitativi dell'attività di ricerca, avevano studiato i contenuti (le scoperte scientifiche e i loro usi, le biografie dei singoli personaggi) in relazione ai contesti sociali, economici, politici. Era arrivato invece il momento di riformulare su basi solide i problemi di organizzazione, amministrazione e controllo politico della *Big Science*. Per farlo bisognava applicare alla scienza le stesse metodologie d'indagine che

⁸⁵ La presente discussione delle teorie di Price deve molto a: X. Polanco, "Aux sources de la scientométrie" *Solaris* 1995, 2 <<http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d02/2polanco1.html>>.

essa applicava alla natura, in primo luogo la statistica, in modo da scoprire le leggi fondamentali che ne regolano la crescita e l'andamento complessivo. Il modello di riferimento era quello della termodinamica dei gas, dove non interessano la posizione e la velocità della singola molecola (il singolo scienziato, la singola scoperta) in un dato istante, ma la media dei valori riferiti al gas nel suo insieme (le statistiche su scienziati, articoli, citazioni). Per trattare quantitativamente la struttura e lo sviluppo della scienza nel suo complesso era tuttavia sufficiente una statistica elementare, meno raffinata di quella della termodinamica («statistically, in a not very mathematical fashion»)⁸⁷.

Dopo aver classificato e conteggiato un numero enorme di articoli tratti dalle *Philosophical Transactions of the Royal Society of London*, dai *Physics Abstracts* e da una compilazione di articoli sulla teoria matematica delle matrici e dei determinanti, Price enunciò quella che riteneva la legge fondamentale di ogni analisi della scienza: qualunque indicatore numerico dei vari settori e aspetti della scienza moderna (dalla metà del '600 in poi) si prenda in considerazione - dal numero di riviste scientifiche, di articoli o di abstract pubblicati, al numero di scienziati e ingegneri vissuti - il suo modo normale di crescita è esponenziale, cioè si moltiplica, in eguali periodi di tempo, di un fattore costante, a prescindere da eventi storici e politici potenzialmente nocivi per il progresso come le guerre.⁸⁸ La crescita esponenziale, ribadita negli stessi anni dal russo Nalimov, non procede con lo

⁸⁶ D.J.d.S. Price, "Quantitative measures of the development of science" *Archives Internationales d'Histoire des Sciences* 1951, 14: 85-93.

⁸⁷ «Why should we not turn the tools of science on science itself? Why not measure and generalize, make hypotheses, and derive conclusions? [...] My approach will be to deal statistically, in a not very mathematical fashion, with general problems of the shape and size of science and the ground rules governing growth and behavior of science-in-the-large [...] The method to be used is similar to that of thermodynamics, in which is discussed the behavior of a gas under various conditions of temperature and pressure. One does not fix one's gaze on a specific molecule called George, traveling at a specific velocity and being in a specific place at some given instant; one considers only an average of the total assemblage in which some molecules are faster than others, and in which they are spaced out randomly and moving in different directions»: D.J.d.S. Price, *Little Science, Big Science...*cit., preface, pp. xv-xvi.

stesso ritmo in tutti i settori. Inoltre non è e non può essere infinita. Arriva un punto di saturazione nel quale la curva esponenziale si trasforma in curva logistica (a forma di S) e il processo descritto dalla variabile sotto osservazione subisce una completa riorganizzazione oppure va incontro a violente fluttuazioni.⁸⁹

Passando dalle serie temporali al modo in cui le unità sono distribuite in un dato momento storico, Price si convinse che le leggi di Lotka, Bradford, Zipf governano l'universo della comunicazione scientifica ed elaborò una versione modificata della legge di Lotka: il numero di scienziati produttivi in un dominio, cioè di autori responsabili all'incirca della metà degli articoli pubblicati, è proporzionale alla radice quadrata del numero totale di scienziati.⁹⁰ Il tasso di crescita della conoscenza scientifica non dipende dal numero di persone reclutate nella carriera scientifica in un dato periodo, ma dal numero, che cresce molto più lentamente del primo, di ricercatori capaci di produrre lavori importanti.⁹¹ Le leggi iperboliche spiegano come mai il fronte attivo della ricerca scientifica corrisponda, in ogni periodo storico, al lavoro di poche centinaia di scienziati e al contenuto di poche migliaia di riviste fondamentali, rispetto alle quali la maggior parte dei periodici disponibili sul mercato è puro "rumore di sfondo".⁹²

Price considerava l'asimmetria delle leggi bibliometriche espressione di un fenomeno sociale e culturale di portata più ampia, già descritto da Robert

⁸⁸ «if any sufficiently large segment of science is measured in any reasonable way, the normal mode of growth is exponential [...] I have no hesitation in suggesting it as the fundamental law of any analysis of science»: D.J.d.S. Price, *Little Science, Big Science...*cit., p. 4.

⁸⁹ : «From this we are led to suggest a second basic law of the analysis of science: all the apparently exponential laws of growth must ultimately be logistic, and this implies a period of crisis extending on either side of the date of midpoint for about a generation. The outcome of the battle at the point of no return is complete reorganization or violent fluctuation or death of the variable»: Ibid., p. 25.

⁹⁰ «This immediately gives an objective method for separating the major from the minor contributors. One may set a limit and say that half the work is done by those with more than 10 papers to their credit, or that the number of high producers seems to be the same order of magnitude as the square root of the total number of authors. [...] the second way, suggesting that the number of men goes up as the square of the number of good ones, seems consistent with the previous findings that the number of scientists doubles every 10 years, but the number of noteworthy scientists only every 20 years»: Ibid., p. 42.

⁹¹ Una critica alla legge di Price si trova in: S. Cole & G.S. Meyer, "Little Science Big Science revisited" *Scientometrics* 1985, 7(3-6): 443-58.

⁹² D.J.d.S. Price, "Networks of scientific papers" *Science* 1965, 149: 510-15, p. 515.

Merton, attraverso il quale il "successo genera successo", la ricchezza reale o simbolica tende spontaneamente ad accrescersi concentrandosi in poche mani. Egli cercò pertanto di derivare le leggi di Lotka, Bradford, Zipf da una teoria probabilistica generale dei vantaggi cumulativi (*cumulative advantage processes*).⁹³ Il suo modello di sviluppo scientifico puntava gran parte della posta sugli indici di citazioni di Garfield, nella convinzione che, oltre ad essere un indicatore affidabile dell'uso della letteratura precedente da parte degli scienziati, la rete delle citazioni rifletta la struttura socio-cognitiva di un campo di ricerca. Per mezzo di strumenti come la teoria dei grafi e l'algebra delle matrici, questa struttura può essere evidenziata e rappresentata graficamente, in modo da ottenere una mappa topografica della letteratura scientifica corrente: «una volta stabilita tale topografia, si potrebbero forse indicare la sovrapposizione e l'importanza relativa delle riviste, dei paesi, degli autori o di contributi individuali dalla posizione che occupano nella mappa e dal loro grado di centralità strategica». ⁹⁴ Un progetto del genere sarebbe stato avviato concretamente, negli anni 70, da Henry Small e Belver Griffith impiegando l'analisi delle co-citazioni.

Da una riflessione sulle prime analisi quantitative dello *Science Citation Index* disponibili agli inizi degli anni '60, Price ricavò alcune conclusioni che Garfield e i teorici successivi dell'analisi citazionale avrebbero apprezzato. Innanzitutto osservò che, in un dato anno, la metà dei riferimenti bibliografici di un corpo selezionato di letteratura punta ad un numero limitato di pubblicazioni, mentre l'altra metà si distribuisce in maniera più o meno irregolare su una quantità molto maggiore di letteratura degli anni precedenti. Quindi «ciascun gruppo di nuovi articoli è legato saldamente ad una piccola,

⁹³ D.J.d.S. Price, "A general theory of bibliometric and other cumulative advantage processes" *Journal of the American Society for Information Science* 1976, 27: 292-306.

⁹⁴ «With such a topography established, one could perhaps indicate the overlap and relative importance of journals and, indeed, of countries, authors, or individual papers by the place they occupied within the map, and by their degree of strategic centralness within a given strip»: D.J.d.S. Price, "Networks of scientific papers" *Science* 1965, 149: 510-15, p. 515. L'articolo è ristampato in: D.J.d.S. Price, *Little Science, Big Science...cit.*, pp. 103-118.

selezionata quota della letteratura scientifica esistente, ma è connesso in maniera piuttosto debole e casuale ad una parte molto più grande». ⁹⁵ Grazie allo *Science Citation Index* diventava allora ipotizzabile l'identificazione automatica dei "classici" e dei "superclassici" in un'area disciplinare e addirittura la realizzazione di un fantomatico *Journal of Really Important Papers*. ⁹⁶ In secondo luogo, circa il 35% delle pubblicazioni esistenti in un anno non viene mai citato ed un altro 49% è citato solo una volta, mentre solo l'1% è citato 5, 6 o più volte: la citazione è una merce rara, a conferma ulteriore del suo valore. Una terza osservazione importante investe la dimensione del tempo: il 30% delle citazioni «sono riferimenti altamente selettivi alla letteratura recente [...] la metà della quale è formata da articoli con età compresa tra 1 e 6 anni». ⁹⁷ Questo carattere "epidermico", la tendenza a citare in prevalenza documentazione recente (*immediacy factor*), differenzia la scienza da altre discipline intellettuali, ne giustifica il carattere cumulativo e conferma l'esistenza, in una disciplina scientifica, di un fronte di ricerca attivo, nel quale gruppi ristretti di studiosi, che Price chiamava "collegi invisibili" (*invisible colleges*), collaborano a definire i problemi da risolvere e le soluzioni da adottare. ⁹⁸ La formazione di tali gruppi è una conseguenza diretta della profonda asimmetria regnante nell'universo sociale e cognitivo della scienza: esistono pochi autori molto produttivi e molto citati, a fronte di

⁹⁵ Ibidem, p. 512.

⁹⁶ «further work in this area might well lead to the discovery that classic papers could be rapidly identified, and that perhaps even the "superclassics" would prove so distinctive that they could be picked automatically by means of citation-index-production procedures and published as a single U. S. (or World) Journal of Really Important Papers»: Ibid., p. 512.

⁹⁷ «30% are highly selective references to recent literature [...] half of the 30 percent being papers between 1 and 6 years old»: Ibid., p. 513.

⁹⁸ D.J.d.S. Price, *Little Science, Big Science...* cit., capitoli 3 (pp. 56-81) e 6 (pp. 119-134). Il termine e il concetto di "collegio invisibile" (*invisible college*) risalgono a Robert Boyle (1627-1691), il quale lo utilizzò per riferirsi al gruppo di studiosi inglesi dalla cui collaborazione sarebbe nata la Royal Society of London. Negli anni '60 del '900, sotto la spinta dell'interesse per lo studio sociale della scienza promosso dai lavori pionieristici di Merton, Kuhn, Price, il concetto fu ripreso e approfondito nel tentativo di comprendere il modo in cui gli scienziati comunicano all'interno di circuiti informali. Un modello speculativo di tali processi si trova in: D. Crane, *Invisible Colleges: Diffusion of Knowledge in Scientific Communication* (Chicago: University of Chicago Press, 1972). La letteratura sull'argomento si è moltiplicata negli anni '70 come dimostra la bibliografia di oltre trecento titoli pubblicata in: D.E. Chubin, *Sociology of Sciences: An Annotated Bibliography on Invisible Colleges, 1972-1981* (New York: Garland, 1983).

una massa di autori che scrivono poco e sono citati ancora meno. Sembra allora naturale che i primi tendano a stabilire rapporti più o meno ufficiali di collaborazione. Un "collegio invisibile" esiste proprio in virtù delle relazioni formali e informali che l'élite degli scienziati che contano («everybody who is really somebody in a field») intrattiene, al di là di confini nazionali e istituzionali, in una fase evoluta dello sviluppo di una disciplina, ad esempio collaborando attivamente ad un progetto oppure scambiandosi documenti, suggerimenti, informazioni sui più recenti progressi del fronte di ricerca. I membri del gruppo risolvono così almeno due problemi in un colpo solo: evitare una crisi comunicativa limitando il numero di pubblicazioni necessarie per tenersi reciprocamente al corrente degli sviluppi nell'area d'indagine condivisa e conferire, a ciascun membro del gruppo, il prestigio derivante dalla stima e dal riconoscimento dei pari.

Price fissò convenzionalmente a 5 anni il limite di età media dei documenti citati in un articolo situato sul fronte di ricerca e sostenne che gli articoli con un indice d'immediatezza più basso, tipici delle "scienze molli" e delle discipline umanistiche, trovano la loro collocazione ideale in un sistema informativo formato da una biblioteca e da un sistema di classificazione delle conoscenze (*archival library*), mentre gli articoli con alto indice d'immediatezza necessitano di un sistema di *current awareness* più rapido ed efficiente di quelli basati sull'indicizzazione per soggetto. In questo secondo caso, aggiunse, «sembra ovvio che l'indicizzazione delle citazioni e l'ingegneria sociale dei collegi invisibili producono risultati di gran lunga superiori alle possibilità di una normale biblioteca o centro informativo». ⁹⁹

⁹⁹ «For such people it seems obvious that citation indexing and the social engineering of invisible colleges produce results vastly superior to any possibilities of normal library or information organization»: D.J.d.S. Price, *Little Science, Big Science...*cit., p. 179. Lo stesso concetto si trova ribadito in altra sede: «The research front builds on recent work, and the network becomes very tight. To cope with this, the scientist (particularly, I presume, in physics and molecular biology) needs an alerting service that will keep him posted, probably by citation indexing, on the works of his peers and colleagues»: D.J.d.S. Price, "Networks of scientific papers"...cit., p. 515.

È stato obiettato che il modello termodinamico di Price, dominato da un positivismo ingenuo, oscura la struttura cognitiva complessa dell'attività scientifica nonché le differenze tra discipline ed aree di ricerca,¹⁰⁰ e che le leggi su cui si basa sono prive di un fondamento empirico solido.¹⁰¹

Nonostante gli sforzi per dotarla di una formulazione matematica rigorosa, per esempio, la sua legge della radice quadrata non sembra trovare conferma nelle distribuzioni empiriche.¹⁰² A prescindere dalla loro validità, le critiche non oscurano il fatto che la metodologia e i principi teorici professati da Price hanno contribuito in maniera decisiva al decollo degli indici di citazioni nell'analisi sociologica e nella valutazione politica della scienza.

2.2. Henry Small, Robert Merton e le funzioni normative della citazione bibliografica

L'articolo di ricerca sperimentale è nato come genere letterario nell'Europa della seconda metà del '600. In quel periodo le prime accademie e società scientifiche si preoccuparono di offrire ai ricercatori uno strumento di comunicazione alternativo e più agile rispetto al libro-trattato della tradizione scolastica. Pubblicazioni periodiche come le *Philosophical Transactions of the Royal Society* e il *Journal des Sçavan*, a differenza delle attuali riviste, servivano non tanto alla diffusione di contributi originali quanto al monitoraggio e alla divulgazione, in forma di compendio, del gran numero di comunicazioni personali, lettere e studi individuali prodotti nelle diverse

¹⁰⁰ G. Holton, "Can science be measured?," in *Toward a Metric of Science: the Advent of Science Indicators*, edited by Y. Elkana, et al. (New York: John Wiley and Sons, 1979): 39-68.

¹⁰¹ J. Tague-Sutcliffe; J. Beheshti; L. Reespotter, "The law of exponential growth. Evidence, implications and forecasts" *Library Trends* 1981, 30(1): 125-45. Una rassegna aggiornata sul modello esponenziale di Price si trova in: A. Fernandez-Cano; M. Torralbo; M. Vallejo, "Reconsidering Price's model of scientific growth: An overview" *Scientometrics* 2004, 61(3): 301-21.

¹⁰² W. Glänzel & A. Schubert, "Price distribution. An exact formulation of Price's square root law" *Scientometrics* 1985, 7(3-6): 211-19; P.T. Nicholls, "Price's square root law: empirical validity and relation to Lotka's law" *Information Processing & Management* 1988, 24(4): 469-77.

branche della scienza naturale.¹⁰³ Era uno scopo sociale, pratico prima ancora che conoscitivo: mettere al corrente uno scienziato su "cosa" stavano facendo i suoi colleghi, "chi" e "dove" si stava occupando di un certo oggetto d'indagine. In quelle riviste si citavano i contributi di altri autori, ma i riferimenti, spesso privi di correlazione con le singole affermazioni, erano generici e disseminati nel testo sotto forma di: "Mr. Newton", "Mr. Leibnitz", "Mr. Huygens".

Fu solo a partire dalla seconda metà del XIX secolo che l'articolo scientifico subì una serie di trasformazioni interne al termine delle quali, nei primi decenni del secolo successivo, acquisì la forma attuale diventando il mezzo più importante di tutela della proprietà intellettuale sulle idee e le teorie scientifiche. Una delle trasformazioni più importanti riguarda proprio l'organizzazione ed il trattamento dei dati bibliografici. La loro compilazione divenne gradualmente più accurata, il formato si standardizzò, la collocazione fisica si allineò con quella concettuale (la fonte giusta citata al posto giusto) adattandosi alla duplice funzione che gli veniva attribuita dalla comunità dei pari: innanzitutto consentire ad un pubblico selezionato di risalire alla fonte di una certa affermazione ed eventualmente verificarla; contestualmente ricompensare, riconoscere, a qualunque livello (metodi, concetti, risultati sperimentali), il lavoro dei predecessori esercitando una forma peculiare di giustizia distributiva.

Su queste due funzioni, trasmettere conoscenza e ricompensare, poggia ogni giustificazione ideologica dell'uso di un *citation index* per ricercare la letteratura e per misurare, ricostruire, valutare le dinamiche della comunicazione scientifica.

¹⁰³ D.A. Kronick, *A History of Scientific and Technical Periodicals* (New York: Scarecrow Press, 1962), cap. 5, pp. 77-117. Sulla nascita del genere del saggio sperimentale nei secoli XVII e XVIII vedi: C. Bazerman, *Le origini della scrittura scientifica: come è nata e come funziona l'argomentazione del saggio sperimentale* (Ancona; Bologna: Transeuropa, 1991).

2.2.1. Il documento-simbolo: Henry Small

Henry Small, collega e collaboratore di Garfield all'ISI, ha difeso il contenuto cognitivo delle citazioni affermando che il riferimento bibliografico, quando non risulti da una frode o da un uso deviato, incarna un'idea discussa dall'autore, è il simbolo relativamente stabile di un concetto scientifico (*concept symbol*).¹⁰⁴

Small ha derivato la sua concezione del documento-simbolo dall'antropologo britannico Edmund Leach (1910-1989) che, in *Culture and Communication* (1976), applicava l'analisi strutturalista ad alcune tematiche fondamentali dell'antropologia sociale illustrando una tesi oggi ritenuta scontata: al pari del linguaggio, la cultura è un sistema di segni da decodificare, ed ogni fenomeno culturale è, in maniera più o meno consapevole, un processo di comunicazione.

Leach, sulla scorta di Chomsky, era convinto che fosse possibile analizzare le dimensioni non verbali della cultura - la moda, l'abbigliamento, la cucina, la musica, gli atteggiamenti - allo scopo di coglierne la struttura grammaticale profonda. Semplicando la terminologia ereditata dalla tradizione semiotica che da Peirce, de Saussure, Cassirer arriva fino a Hjelmslev, Morris, Jakobson e Barthes, egli sosteneva che la comunicazione umana si realizza attraverso azioni espressive che operano come segnali, segni e simboli.¹⁰⁵ Nei segni l'oggetto, l'indice che funziona come segno di un certo significato è contiguo a ciò che viene significato (relazione metonimica): la corona è un segno di regalità nel contesto delle tradizioni politiche europee; la sequenza di lettere m-e-l-a è un segno per un particolare frutto nel contesto delle convenzioni della lingua italiana. Nei simboli invece non esiste alcun rapporto di contiguità

¹⁰⁴ H.G. Small, "Cited documents as concept symbols" *Social Studies of Science* 1978, 8(3): 327-40.

¹⁰⁵ E.R. Leach, *Cultura e comunicazione: la logica della connessione simbolica : un'introduzione all'uso dell'analisi strutturale nell'antropologia sociale* (Milano: Franco Angeli, 1981), in particolare pp. 24-32.

tra l'oggetto e ciò che esso significa, anzi le due entità appartengono a contesti culturali diversi e la relazione stabilita tra di essi è puramente convenzionale (relazione metaforica): la corona usata come marchio di fabbrica di una birra è un simbolo, non un segno della birra; il serpente nel Giardino dell'Eden è un simbolo, non un segno del peccato. Esiste però una differenza sostanziale tra i simboli occasionali, per esempio i simboli dei sogni e delle immagini poetiche, «che non comunicano alcuna informazione "pubblica", finché essi non siano arricchiti da un commento»,¹⁰⁶ ed i simboli standardizzati, che comunicano informazione di pubblico dominio e tendono pertanto a veicolare significati costanti.

Small ha trasferito questo schema nel regno delle citazioni bibliografiche. Una citazione è al tempo stesso un segno ed un simbolo: è un segno che sta per un oggetto fisico, cioè il documento citato in carne e ossa, col quale condivide alcune caratteristiche (relazione metonimica); è un simbolo del concetto o dei concetti espressi nel testo del documento (relazione metaforica).

Negli articoli scientifici molte citazioni sono simboli individuali che l'autore usa per comunicare le proprie idee, mentre «altre citazioni si riferiscono a documenti il cui significato può essere condiviso da una comunità o da un gruppo di scienziati e tali documenti tendono ad essere citati frequentemente: nella terminologia di Leach sono simboli standardizzati».¹⁰⁷

In alcune circostanze l'accordo degli scienziati sul significato da attribuire ad un concetto è talmente forte che si cita sempre lo stesso documento per richiamarsi ad esso. Uno degli articoli più citati della storia della scienza, il lavoro di Lowry e dei suoi collaboratori sul metodo di misurazione delle proteine che sfrutta il fenolo come reagente,¹⁰⁸ è un esempio classico di documento al quale la comunità dei citanti attribuisce sempre lo stesso

¹⁰⁶ Ibid., p. 31.

¹⁰⁷ «Other citation are to documents whose significant content may be shared by a community or group of scientists, and such documents are likely to be frequently cited: in Leach's terminology they are 'standard symbols'»: H.G. Small, "Cited documents..." cit., p. 329.

significato, cioè, per l'appunto, "metodo Lowry per la determinazione delle proteine".

I simboli standardizzati sono evidentemente «il prodotto di un dialogo e di un processo di selezione da parte di molti individui lungo un periodo di tempo» ed è ipotizzabile che ogni scienziato «porti con sé un repertorio di tali concetti collettivi e dei corrispondenti documenti-simbolo». ¹⁰⁹ Citando un documento lo scienziato lo etichetta, gli attribuisce un significato specifico, che il lettore può desumere dall'analisi del contesto della citazione. Il concetto cui si riferisce la citazione non è necessariamente un'entità astratta, può trattarsi di «risultati sperimentali, metodologie, tipologie di dati, nozioni metafisiche, formulazioni teoretiche o equazioni». ¹¹⁰ Non è detto inoltre che il citante attribuisca al documento citato o ad una sua parte lo stesso significato che l'autore ha inteso conferirgli. ¹¹¹ Esiste, nella scienza come nella letteratura, un certo grado di autonomia del prodotto intellettuale allorché viene immesso nel circuito della comunicazione, ma il caso più comune resta, per Small e per tutta la tradizione bibliometrica, quello del consenso normativo. ¹¹² In fin dei conti, si potrebbe aggiungere, chi sta citando è uno scienziato e le operazioni compiute dalla mente di uno scienziato sono di gran lunga più prevedibili di quelle della mente di un artista o di un comune mortale. Per secoli gli intellettuali hanno cercato una "lingua perfetta" capace di istituire un rapporto

¹⁰⁸ O.H. Lowry; N.J. Rosebrough; A.L. Farr et al., "Protein measurement with the folin phenol reagent" *Journal of Biological Chemistry* 1951, 193: 265-75.

¹⁰⁹ «In the case of 'standard symbols', the 'idea' is the product of a dialogue and selection process on the part of many individuals over a period of time [...] One of the hypotheses to be explored in this paper is that a scientist carries with him a repertoire of such collective concepts and their correspondent document-symbols»: H.G. Small, "Cited documents..." cit., p. 329.

¹¹⁰ «I regard ideas or concepts as residing in the mind and expressed in language. I do not, however, restrict these terms to abstract or theoretical formulations: I include experimental findings, methodologies, types of data, metaphysical notions, theoretical statements or equations...»: Loc. cit.

¹¹¹ «The citing author may transform the work into something unrecognizable to the cited author. It may be the case that prior literature is necessarily in a constant state of reinterpretation, adapting to changes in knowledge within the field»: H.G. Small, "On the shoulders of Robert Merton: towards a normative theory of citation" *Scientometrics* 2004, 60(1): 71-79, p. 75.

¹¹² «For an individual document the idea for which the author is being credited can be congruent or at odds with the author's original message or with consensual usage in the field. When the constructed meaning is coincident with the author's original message as well as common usage, we may say that there is strong normative compliance. This would, I suspect, be the most common case»: Ibid., p. 76.

in scala uno a uno con il complesso dei concetti astratti necessari a descrivere il mondo esterno. Se una lingua del genere esistesse, la traduzione, persino la traduzione automatica, non sarebbe un grosso problema al pari della comunicazione tra gli esseri umani.¹¹³ Gli scienziati non hanno il privilegio della lingua perfetta ma, in misura maggiore rispetto ad altre categorie di individui creativi, tendono a trovare un accordo sulle regole di generazione e manipolazione delle proposizioni linguistiche riferite agli stati del mondo. Se si accetta questa premessa e l'equivalenza concetti-citazioni viene generalizzata per un gran numero di documenti citati nella letteratura specializzata, allora lo *Science Citation Index* può essere considerato un indice di concetti scientifici equiparabile ad un thesauro disciplinare. In tal caso è lecito interrogarlo per una ricerca tematica della letteratura e per ricavare informazioni utili a valutare in termini di impatto cognitivo il numero di citazioni ricevute da un documento.

Un indice di questo tipo, come Garfield riconosce esplicitamente,¹¹⁴ aveva in realtà un'ambizione ancora più grande. Attraverso l'equivalenza stabilita da Small tra citazioni e concetti, l'unificazione bibliografica della letteratura scientifica aspirava al ruolo di una delle possibili soluzioni concrete al progetto che, dalla fine degli anni '20, aveva impegnato gli empiristi logici del Circolo di Vienna nella costruzione di un linguaggio unificato della scienza: ridurre i concetti delle varie discipline scientifiche alle registrazioni (protocolli) dei dati immediati ricavati dall'esperienza sensibile, espresse nel linguaggio più aderente alla descrizione degli stati del mondo, ovvero il linguaggio della fisica; ridurre, mediante la logica simbolica, il senso di ogni concetto, indipendentemente dal settore disciplinare di appartenenza, ai

¹¹³ U. Eco, *Dire quasi la stessa cosa. Esperienze di traduzione* (Milano: Bompiani, 2003). pp. 345-364.

¹¹⁴ E. Garfield, "Citation consciousness. Interview with Eugene Garfield, Chairman Emeritus of I.S.I., Philadelphia, U.S.A." *Password* 2002(June): 22-25
<<http://www.garfield.library.upenn.edu/papers/passwordinterview062002.pdf>>.

concetti di livello più basso riferibili ai dati empirici.¹¹⁵ «Ciò che Neurath chiama "integrazione enciclopedica delle asserzioni scientifiche" - scriveva Garfield nel 1959 - [è] ciò che io chiamo "Indice Unificato della Scienza"». ¹¹⁶

2.2.2. La citazione come ricompensa sociale: Robert Merton

La scientometria ha potuto contare sin dall'inizio su un'alleanza strategica con l'indirizzo metodologico prevalente nella sociologia della scienza del XX secolo. Già nel 1942, in un periodo nel quale l'autonomia intellettuale degli scienziati era messa in pericolo dalla guerra, Robert Merton formulava un insieme di criteri per descrivere l'*ethos* della scienza, ossia il complesso di norme, espresse sotto forma di prescrizioni, divieti, valori, che si suppone governino l'operato degli scienziati.¹¹⁷ Le norme principali sono quattro: universalismo, comunismo, disinteresse, scetticismo organizzato.

L'universalismo implica che «ogni pretesa di verità, qualsiasi sia la sua fonte, deve essere soggetta a criteri impersonali prestabiliti»: ¹¹⁸ la scienza, non diversamente dal mito e dalla religione, aspira a conoscenze universali, ma si tratta di un'universalità che ambisce ad essere "qualitativamente" superiore a quella del mito e della religione per il modo (si suppone razionale) in cui viene attinta. Il comunismo, antitesi della segretezza tipica della pseudoscienza di maghi e alchimisti, concerne l'aspettativa che i risultati della ricerca scientifica siano resi pubblicamente accessibili e manipolabili, fermo

¹¹⁵ Le teorie neoempiriste sull'unificazione della scienza culminarono, alla fine degli anni '30, nel progetto di una *International Encyclopaedia of Unified Science*, il cui primo volume vide la luce, sotto la direzione di Otto Neurath (1882-1960), a Chicago nel 1938. Vedi: H. Hahn; O. Neurath; R. Carnap, *La concezione scientifica del mondo: il Circolo di Vienna* (Roma; Bari: Laterza, 1979).

¹¹⁶ «The primary purpose of this paper is to discuss a plan for accomplishing what Neurath calls "an encyclopedic integration of scientific statements", what I call a "Unified Index to Science."»: E. Garfield, "A unified index to science," in *Proceedings of the International Conference on Scientific Information*, 1 (Washington, D.C.: National Academy of Sciences - National Research Council, 1959): 461-74, p. 461.

¹¹⁷ L'articolo, apparso originariamente nel *Journal of Legal and Political Sociology*, è stato ristampato in traduzione italiana in: R.K. Merton, "La struttura normativa della scienza (1942)," in *La sociologia della scienza. Indagini teoriche ed empiriche*, edited by N.W. Storer (Milano: Franco Angeli, 1981): 349-62.

restando il diritto per lo scienziato di essere riconosciuto e ricompensato dai colleghi per la novità del suo contributo: la scienza è un'impresa cumulativa e collaborativa, pertanto ogni restrizione privatistica sui prodotti ufficiali della sua attività rischia di rallentare il progresso (un argomento tornato in prima fila nell'attuale dibattito sull'*open access*). Il disinteresse è l'assenza di motivazioni economiche o personali nel perseguimento dei fini conoscitivi. Anche un artista, in teoria, opera in maniera disinteressata, ma nel caso dello scienziato c'è qualcosa in più di un atteggiamento psicologico: il disinteresse si materializza in un insieme di strutture istituzionali che formano un sistema neutrale di sorveglianza sui comportamenti dei singoli ricercatori. Lo scetticismo sistematico è la «sospensione del giudizio e l'esame distaccato delle credenze secondo criteri logici ed empirici»,¹¹⁹ in altri termini la rinuncia ad affermazioni che non siano verificate o verificabili.

Per Merton il successo straordinario dimostrato dalla scienza nella scoperta delle leggi naturali si spiega ammettendo un'omogeneità ed una continuità di fondo tra i diversi livelli in cui si esplica l'attività di ricerca. La pratica scientifica, quello che gli scienziati fanno concretamente nel laboratorio, produce dei risultati, che si esprimono sotto forma di scritti destinati alla pubblicazione, tipicamente articoli di riviste. Gli scienziati hanno urgenza di pubblicare per poter reclamare la priorità sulle idee comunicate: diversamente da quanto accade nel mondo degli interessi economici, la proprietà del bene intellettuale (la scoperta scientifica) è tanto più al sicuro quanto più esso viene ceduto alla collettività sotto forma di un documento letto e approvato dal gruppo dei pari.¹²⁰ Gli articoli pubblicati sono letti e recepiti dalla comunità degli esperti di una disciplina, i quali ne riconoscono il valore citandoli nei loro lavori. La citazione bibliografica si configura quindi come un atomo di

¹¹⁸ Ibid., p. 352.

¹¹⁹ Ibid., p. 359.

¹²⁰ R.K. Merton, "La priorità nella scoperta scientifica (1957)," in *La sociologia della scienza. Indagini teoriche ed empiriche*, edited by N.W. Storer (Milano: Franco Angeli, 1981): 371-414.

peer recognition.¹²¹ A questo punto, il sistema di gerarchie e ricompense della scienza premia i ricercatori che hanno prodotto i risultati più originali sia garantendo loro un tasso di citazioni bibliografiche nettamente più alto della media, sia dispensando cattedre e titoli onorifici. Il credito simbolico fluisce così spontaneamente dai documenti agli autori, che lo convertono in ruoli istituzionali e posizioni di potere, e dagli autori alle istituzioni, alle nazioni, alle riviste.

Merton naturalmente non era così ingenuo da pensare che gli scienziati seguano le norme di condotta come un breviario e che l'aspetto immacolato degli articoli scientifici rispecchi l'andamento effettivo della ricerca, con le sue tipiche esitazioni, deviazioni, strategie di accomodamento. La continuità tra vita di laboratorio, resoconti ufficiali e *reward system* è comunque garantita dal fatto che gli scienziati, o almeno la maggior parte di essi, operano in accordo con le prescrizioni dell'universalismo, del comunismo, del disinteresse e dello scetticismo organizzato. Non è necessario che essi credano in quelle norme, è sufficiente che si comportino "come se" ci credessero per garantire comunque il successo dell'impresa.

Il sistema funziona bene anche per il modo in cui organizza la divisione del lavoro. La scienza, al pari di altri universi sociali, è un'istituzione stratificata, in cui il potere e le risorse sono concentrati nelle mani di una minoranza relativamente ristretta di individui. A differenza di altre organizzazioni sociali, però, le norme di Merton garantiscono che in campo scientifico le disuguaglianze si realizzano secondo criteri rigorosamente universalistici, in

¹²¹ «A composite cognitive and moral framework calls for the systematic use of references and citations. As with all normative constraints in society, the depth and consequential force of the moral obligation to acknowledge one's sources become most evident when the norm is violated (and the violation is publicly visible). The failure to cite the original text that one has quoted at length or drawn upon becomes socially defined as theft, as intellectual larceny or, as it is better known since at least the seventeenth century, as plagiarism. Plagiarism involves expropriating the one kind of private property that even the dedicated abolitionist of private productive property, Karl Marx, passionately regarded as inalienable (as witness his preface to the first edition of *Capital* and his further thunderings on the subject throughout that revolutionary work)»: R.K. Merton, "The Matthew effect in science, 2: cumulative advantage and the symbolism of intellectual property" *ISIS* 1988, 79(299): 606-23, p. 622.

modo da premiare gli autori dei contributi più significativi.¹²² Mantenere l'equilibrio non è faticoso anche perché, in perfetta armonia con le leggi bibliometriche, il numero di documenti e autori realmente significativi è basso.

Muovendosi in questa direzione, Jonathan e Stephen Cole hanno confutato la tesi, sostenuta tra gli altri dal filosofo spagnolo Ortega y Gasset e nota perciò come "ipotesi di Ortega", secondo cui la crescita della conoscenza scientifica sia da ascrivere al lavoro umile e sommerso di una miriade di ricercatori mediocri che con le loro piccole scoperte hanno preparato il terreno al grande uomo di genio. L'analisi delle citazioni mostra invece che la maggior parte delle fonti usate negli articoli fondamentali per un settore di ricerca è prodotta solo da un piccolo numero di ricercatori.¹²³

Il sistema però non è infallibile, anzi scricchiola in diversi punti. Ad esempio quando, per malafede, per un errore di giudizio o semplicemente per il limitato numero di crediti disponibili, non premia equamente tutti coloro che lo meriterebbero, ma privilegia alcuni a scapito di altri. In certi casi, descritti da Merton come esemplificazioni dell'“effetto S. Matteo”, accade che, come nel vangelo omonimo, «a chiunque ha, sarà dato e sarà nell'abbondanza; ma a chi non ha, sarà tolto anche quello che ha» (Matteo 25,29). Gli scienziati ricompensati una volta entrano di fatto in un'élite destinata sia ad un accesso privilegiato ai mezzi di produzione scientifica (fondi di ricerca, collaborazioni internazionali, ruoli di primo piano in accademie e società scientifiche) sia ad una maggiore visibilità rispetto alla massa “anonima” dei colleghi. Dal punto di vista delle pratiche citazionali, l'“effetto S. Matteo” implica la tendenza degli autori a citare solo o preferibilmente lavori di scienziati dalla fama

¹²² «...the social processes that produce stratification in science may approximate the ideal of a meritocracy based on the application of universalistic principles, and may do this far more than most, perhaps more than all, institutions in our society»: J.R. Cole & S. Cole, *Social Stratification in Science* (Chicago: University of Chicago Press, 1973), p. 89.

¹²³ J.R. Cole & S. Cole, “The Ortega hypothesis” *Science* 1972, 178(4059): 368-75.

consolidata, trascurandone altri potenzialmente altrettanto (o forse più) pertinenti.

Merton, che negli anni '60 aveva celebrato i *citation indexes* come un nuovo strumento di analisi sociologica, non ha lavorato in prima persona nel campo della teoria e dell'analisi delle citazioni, anzi ha sottolineato il carattere rudimentale delle misure d'impatto e qualità scientifica ricavate dagli indicatori bibliometrici ed ha insistito su fenomeni che provocano la perdita di molti dati significativi, uno su tutti l'"obliterazione per incorporazione", cioè l'esistenza di documenti talmente importanti per una disciplina da non essere più citati in quanto inglobati nel corpus delle conoscenze acquisite.¹²⁴ Le sue norme sono apparse da subito troppo generiche per caratterizzare in maniera univoca l'attività scientifica. Eppure i bibliometri le hanno prese sul serio, ricavandone una cornice intellettuale di riferimento per la comprensione dei comportamenti citazionali, i quali appaiono razionali, quindi prevedibili, in quanto regolati da un'etica professionale, in primo luogo dalla prescrizione che impone il riconoscimento dei meriti individuali all'interno dell'impresa collettiva della scienza.

¹²⁴ R. K. Merton (1983) "Foreword" in E. Garfield, *Citation Indexing - its Theory and Application...*cit., pp. v-ix.

Capitolo 3

Ricerche "senza soggetto": lo Science Citation Index come strumento di information retrieval

La genesi dello *Science Citation Index* nasconde un paradosso: proiettato nel futuro, progettato per promuovere la ricerca e il potenziamento dell'informazione scientifica, l'indice di Garfield è, dal punto di vista tecnico, il discendente diretto di un modello culturale - il sistema giuridico anglosassone - saldamente ancorato al passato, al formalismo del "ragionare per precedenti" e alla dottrina conservatrice dello *stare decisis*.¹²⁵ Malgrado tavole di "casi precedenti" annesse alle rassegne giudiziarie si trovino già nel '700 ed un *citation index* a sè stante di sentenze californiane abbia visto la luce nel 1860,¹²⁶ l'applicazione più conosciuta di questo modello è lo *Shepard's Citation*, un apparato di consultazione giuridica usato dagli avvocati americani sin dal 1875.¹²⁷ Partendo da un caso discusso in uno dei tribunali di 48 stati americani o nei tribunali federali, identificato univocamente con un codice formato dall'indicazione del volume e del numero di pagina del documento originario in cui era riportato, l'indice elencava sotto di esso tutte le pubblicazioni successive che lo avevano discusso unitamente a tutte le istanze dei tribunali che vi si erano richiamati per confermare, mettere in discussione o capovolgere la sentenza originaria.

¹²⁵ S.R. Heifetz, "Blue in the face: the Bluebook, the Bar exam, and the paradox of our legal culture" *Rutgers Law Review* 1999, 51: 695-711.

¹²⁶ F.R. Shapiro, "Origins of bibliometrics, citation indexing, and citation analysis: the neglected legal literature" *Journal of the American Society for Information Science* 1992, 43(5): 337-39.

¹²⁷ Per una storia dei *legal citation indexes* nel contesto del sistema giuridico anglosassone, compresi gli antecedenti e contemporanei allo *Shepard's*, vedi: P. Ogden, "'Mastering the lawless science of our law": a story of legal citation indexes" *Law Library Journal* 1993, 85: 1-47

Le radici giuridiche dei *citation indexes* affondano ancora più lontano. Manoscritti ebraici contenenti indici di citazioni datano dal XII secolo e, a partire dal 1546, l'edizione a stampa del *Talmud* incorporava un indice dei luoghi dell'opera di Maimonide (1135-1204) in cui erano citate le leggi del corpus. In epoca medievale i cristiani compilavano prevalentemente indici per soggetto, gli ebrei indici di citazioni. Una possibile spiegazione risiede nel fatto che, mentre per i cristiani «il termine della legge è Cristo» (Lettera ai Romani 10,4), nella tradizione talmudica la religione ha un fondamento legalistico ed il testo originario in cui la legge è contenuta risulta da una stratificazione di testi diversi (la Bibbia, il Talmud, i codici). Un indice di citazioni serviva perciò a vari scopi: a preparare le omelie, a seguire lo sviluppo di una norma di condotta, a trovare le autorità in grado di legittimare un'argomentazione.¹²⁸

La componente giuridica si è conservata anche quando i nodi da connettere sono diventati articoli scientifici anziché testi sacri, ma con un significato diverso: lo scienziato moderno cita per pagare un debito in modo che tutti si accorgano che ha anche un credito (cognitivo, sociale) da riscuotere; così facendo legittima il diritto di proprietà intellettuale su un'idea nuova o presunta tale.

¹²⁸ B.H. Weinberg, "The earliest hebrew citation indexes" *Journal of the American Society for Information Science* 1997, 48: 318-30; B.H. Weinberg, "Indexes and religion: reflections on research in the history of indexes" *Indexer* 1999, 21(3): 111-18. In questo secondo articolo ho trovato il riferimento all'Epistola ai Romani e la tesi degli "strati" testuali che differenziano la tradizione talmudica da quella cristiana.

3.1. Ideazione e realizzazione dello Science Citation Index

Nel 1951 Eugene Garfield aderì al *Johns Hopkins University's Welch Medical Library Indexing Project*, sponsorizzato dall'Army Medical Library, la futura National Library of Medicine. Scopo principale del progetto, imperniato sulla *Current List of Medical Literature* (antenato dell'*Index Medicus*), era capire se, e in che modo, i computer potevano essere usati per rendere più efficiente l'indicizzazione e la ricerca della letteratura biomedica. L'orientamento prevalente era quello di cercare un sistema per trasformare semplici *tables of contents* delle riviste in registrazioni analitiche con tanto di *subject headings* desunti da un thesauro disciplinare (il *Medical Subject Headings*). Garfield, responsabile della standardizzazione della nomenclatura chimica, si rese conto dei limiti principali di tale orientamento: innanzitutto il ritardo di pubblicazione imputabile al fattore umano dell'indicizzazione semantica manuale; in secondo luogo l'arbitrarietà nella scelta dei *subject headings*, troppo settoriali e condizionati dalla scarsa aderenza alla dinamica attiva della ricerca scientifica. Il dibattito preliminare inoltre aveva evidenziato la scarsa utilità di un indice costruito solo con i termini desunti dal titolo e dall'abstract di un documento, spesso scarsamente rappresentativi del suo contenuto, unita alla difficoltà, con i mezzi tecnici allora disponibili, di automatizzare la procedura d'indicizzazione dei *full text*.

Il ritardo congenito dei repertori bibliografici disciplinari poteva essere risolto recuperando, potenziando ed estendendo al di là dell'ambito biomedico l'idea, peraltro già ventilata nell'ambito del progetto, di un sistema semi-automatico di indicizzazione basato sulla fotocopiazione dei sommari delle riviste.

Nacquero così i *Current Contents (CC)*, che resero accessibili ad un pubblico sempre più ampio, con periodicità settimanale, gli spogli delle riviste di biblioteconomia e scienze dell'informazione (1953), di management e scienze sociali (1955), di farmacia e scienze biomediche (1957), di scienze fisiche

(1960). Il punto di forza del sistema era la velocità di aggiornamento del repertorio, spesso, come accade oggi per gli articoli disponibili sul web *ahead of print*, in anticipo sulla distribuzione dei fascicoli a stampa. La Garfield Associates, che nel 1960 divenne Institute for Scientific Information (ISI), inserì i *Current Contents* in una strategia culturale e commerciale di portata ancora più ampia, che comprendeva sia un servizio di document delivery a pagamento unico nel suo genere, in quanto dedito alla fornitura non di copie ma di articoli originali “strappati” dai molteplici esemplari dei fascicoli di riviste pervenute all’ISI,¹²⁹ sia la redazione e l’aggiornamento di un indirizzario degli autori utile a facilitare lo scambio di informazioni e reprint tra membri dello stesso “collegio invisibile”. La popolarità dei *Current Contents* presso università e istituti di ricerca crebbe negli anni seguenti al punto da prefigurare una situazione di reciproca dipendenza tra repertorio bibliografico, industria editoriale e comunità scientifica che avrebbe caratterizzato anche lo *Science Citation Index* e prodotti derivati: l’inclusione nei *Current Contents* poteva alterare sensibilmente la situazione finanziaria di una rivista, i cui standard editoriali dovevano allinearsi con gli standard di qualità dell’ISI; dal canto loro, gli scienziati cominciavano a selezionare le riviste cui inviare i propri lavori in funzione del trattamento ad esse riservato dai *Current Contents*.¹³⁰

Il complesso di esperienze, intuizioni e incontri casuali appena descritti condusse, in poco tempo, alla maturazione dell’idea e dei presupposti tecnici di un *citation index* per le scienze.

Lavorando come volontario ai *Chemical Abstracts*, Garfield si era reso conto dell’utilità di includere nelle schede di spoglio i riferimenti bibliografici

¹²⁹ Il nome del servizio era per l’appunto OATS, cioè *Original Article Tear Sheet Service*: E. Garfield, “The mysteries of ISI’s International OATS Library Service Revealed (Again!),” in *Essays of an Information Scientist 1974-1976*, 2 (Philadelphia: ISI Press, 1977): 55-57.

¹³⁰ «A listing in CC can dramatically alter a journal’s financial and editorial condition. In many instances editors have informed us that scientists in their country were more willing to publish in a journal listed in CC. The effect of such a listing is obvious from reprint requests which are received due to listings in CC»: E.

collegati ai documenti indicizzati. Dietro suggerimento del letterato e storico della scienza Chauncey D. Leake, egli aveva concentrato in particolare l'attenzione su una tipologia di articolo scientifico, la rassegna (*review article*), cioè l'articolo che, senza apportare contributi originali, sintetizza lo stato dell'arte su un particolare argomento. In una rassegna l'uso delle citazioni è particolarmente disciplinato: l'autore aggiunge un riferimento bibliografico quasi ad ogni frase del testo ed il significato che attribuisce a quel riferimento è esplicito, privo di ambiguità, anzi è proprio quello comunicato dalla frase che lo introduce. Di conseguenza la frase stessa può essere considerata un'insolita ma autorevole stringa d'indicizzazione del lavoro citato («an unusually definitive indexing statement of the cited work»)¹³¹.

Intanto, nel marzo 1953, l'ex presidente della *Frank Shepard Company* William C. Adair aveva inviato allo staff del *Welch Project* una lettera in cui proponeva una soluzione collaudata al problema del controllo bibliografico: shepardizzare la letteratura biomedica, cioè sbrogliarne la matassa seguendo il filo dei nessi citazionali. La proposta fu ignorata, ma permise a Garfield di apprendere dell'esistenza dello *Shepard's Citation*. La struttura di questo repertorio mostrava che, dal punto di vista tecnico, l'impresa di un'organizzazione della letteratura in funzione della ricerca di singole unità informative era fattibile al di fuori del paradigma consolidato dei *subject headings*. I contatti con Adair proseguirono e il disegno di un indice di citazioni per le scienze, da compilare in modo automatico con le schede perforate dell'IBM, acquisì contorni sempre più precisi.

Nonostante i limiti tecnologici dei calcolatori disponibili, lo stato dell'arte in materia di indicizzazione automatica lasciava presagire delle possibilità. Alla fine degli anni '40, il padre gesuita Roberto Busa aveva iniziato a produrre con i calcolatori dell'IBM l'indice completo delle concordanze dell'opera di S.

Garfield, "Current Contents - Ninth Anniversary," in *Essays of an Information Scientist 1962-1973*, 1 (Philadelphia: ISI Press, 1977): 12-15, p. 14.

¹³¹ E. Garfield, *Citation Indexing: its Theory and Application...cit.*, p. 7.

Tommaso d'Aquino e cominciavano a diffondersi, nella critica testuale, i primi metodi di collazione automatica. All'incirca nello stesso periodo, gli esperti di linguistica computazionale sperimentavano la costruzione di sistemi efficaci di traduzione automatica. Pesava sui loro sforzi il monito di Yehoshua Bar-Hillel: una traduzione completamente automatizzata di un testo può essere ottenuta solo sacrificando l'accuratezza del risultato; per tradurre bene e in fretta un computer non deve solo incorporare un dizionario bensì un'intera enciclopedia di concetti e delle loro relazioni.¹³²

Garfield, che nel frattempo aveva conseguito una laurea in *Library Science* e un Ph.D. in linguistica strutturale, doveva risolvere un problema molto simile a quello della traduzione, cioè estrarre in modo automatico, da un corpus di letteratura specializzata, un apparato di termini sufficientemente rappresentativi da utilizzare in un sistema efficiente di recupero dell'informazione. La difficoltà principale era data dall'incapacità della macchina di formulare giudizi attendibili sul peso relativo delle parole che compongono un testo e di stabilire relazioni tra concetti che andassero oltre il contenuto esplicito dei documenti. Un'analisi lessicale dei singoli vocaboli, così come un'analisi sintattica delle relazioni grammaticali tra le parole, non erano di grande aiuto, un indicizzatore umano sapeva fare molto di più. Ma a questo punto, anziché rincorrere vanamente la chimera di un sistema automatico d'indicizzazione per parola in grado di gareggiare con il buonsenso degli esseri umani, Garfield considerò l'opportunità di cambiare direzione trasformando radicalmente l'unità di analisi: non più le parole, di per sé ambigue per la dipendenza del loro significato dal contesto, ma i riferimenti bibliografici, la cui stabilità semantica è garantita, nell'ambito di una

¹³² Y. Bar-Hillel, "The present state of research on mechanical translation" *American Documentation* 1951, 2(4): 229-37. Il riflesso di questo dibattito sul progetto di Garfield è documentato in: E. Garfield, "From computational linguistics to algorithmic historiography" *Lazerow Lecture held in conjunction with panel on "Knowledge and Language: Building Large-Scale Knowledge Bases for Intelligent Applications"*, University of Pittsburgh, September 2001 <<http://www.garfield.library.upenn.edu/papers/pittsburgh92001.pdf>>.

disciplina scientifica, dall'accordo degli autori che tendono ad usarli nei loro articoli mettendoli in relazione quasi sempre con gli stessi concetti.

In un articolo apparso su *Science* nel 1955,¹³³ Garfield esponeva finalmente il suo progetto di un sistema bibliografico semi-automatizzato di indicizzazione delle citazioni della letteratura scientifica costruito sul modello dello *Shepard's*. Un sistema destinato ad «eliminare la citazione acritica di dati fraudolenti, incompleti o obsoleti consentendo allo studioso diligente di essere al corrente delle critiche indirizzate ai lavori precedenti».¹³⁴

Per rendere l'oggetto più gradito ai bibliotecari, l'autore ne rintracciava gli ideali precursori nelle compilazioni di nomenclatura chimica (il *Beilstein*) e in alcuni cataloghi di grandi biblioteche (la Bibliothèque National, il British Museum, la Library of Congress).¹³⁵ L'idea era estremamente semplice, a dispetto delle difficoltà tecniche che ne condizionarono la realizzazione.

Partendo da un nucleo preselezionato di riviste scientifiche, l'indice avrebbe elencato, in corrispondenza di ciascun articolo pubblicato lungo un numero predefinito di anni, tutti i riferimenti bibliografici in esso contenuti.

L'originalità dell'operazione consisteva nel fatto che, una volta compilata, la lista iniziale poteva essere letta o interrogata al rovescio, permettendo così di risalire dal testo (quindi dall'autore, dal gruppo di ricerca, dal dipartimento) citato all'insieme dei testi citanti nonché dal numero e dalla distribuzione

¹³³ E. Garfield, "Citation Indexes for science: a new dimension in documentation through association of ideas" *Science* 1955, 122(3159): 108-11. Ristampato in: E. Garfield, "Citation Indexes for science: a new dimension in documentation through association of ideas," in *Essays of an Information Scientist*, 6 (Philadelphia: ISI Press, 1983): 468-71. Contemporaneamente veniva pubblicato il lavoro di Adair imperniato sulla stessa proposta: W.C. Adair, "Citation indexes for scientific literature?" *American Documentation* 1955, 6: 31-32 <<http://www.garfield.library.upenn.edu/papers/adaircitationindexesforscientificliterature1955.html>>.

¹³⁴ E. Garfield, "Citation Indexes...cit.", p. 468.

¹³⁵ «There are analogies in bibliographic operations. For example, in cataloguing books for booksellers' or library catalogs, an attempt is made to find references to each book in one or more authoritative bibliographic sources, such as the catalogs of the British Museum (BM), Bibliothèque Nationale (BN), or the Library of Congress (LC). The "authority" card used in cataloguing sometimes looks like a Shepard entry»: Ibid., p. 469. Più avanti il richiamo al *Beilstein*: «In a certain sense a citation index is not very different from a compendium like *Beilstein*, which gives a rather complete record of a compound, compiled by a similar method»: Ibid., p. 471. L'*Handbuch der Organische Chemie* di Friedrich Beilstein (prima edizione 1881) e l'*Handbuch der Anorganische Chemie* di Leopold Gmelin (prima edizione 1819) sono assimilabili ad un indice citazionale nella misura in cui associano ad ogni composto chimico (organico e inorganico) il riferimento ai testi della letteratura chimica da cui sono ricavati i dati.

nello spazio e nel tempo dei testi citanti ad una serie di considerazioni extra-bibliografiche circa l'impatto cognitivo del testo citato.

Una bozza di *citation index* del Vecchio Testamento ed un primo esperimento di indicizzazione delle citazioni di 5000 brevetti chimico-farmaceutici videro la luce verso la metà degli anni '50, ma erano solo prove tecniche. La produzione su larga scala richiedeva soldi e la giusta dose di complicità da parte di editori, scienziati politicamente attrezzati e politici sensibili al destino della scienza.

I genetisti Gordon Allen e Joshua Lederberg incoraggiarono il progetto e si adoperarono per garantirgli l'appoggio finanziario. L'occasione giusta si presentò quando la sezione di studi genetici del National Institutes of Health dovette fronteggiare la richiesta, da parte dell'amministrazione, di documentare l'impatto esercitato dalle proprie ricerche. Un *citation index* sembrava pensato apposta per soddisfare questa esigenza e fu proprio Lederberg, che nell'ottobre 1961 era entrato in una commissione presidenziale di consulenza in materia di informazione scientifica (Weinberg Committee), ad esaltarne le doti nelle sedi opportune. Egli ha rivestito un ruolo cruciale nel conferire all'indice di Garfield lo status di oggetto "politicamente" desiderabile, presentandolo come il giusto rimedio alla «enorme crescita della letteratura» e al «pericolo che la scienza si frammenti in un ammasso di scoperte ripetitive o, peggio, in specialità in conflitto tra loro».¹³⁶ Sotto la spinta del dibattito sul futuro dell'informazione scientifica negli Stati Uniti, la costruzione dell'indice, prima ancora che un'impresa tecnica e commerciale, divenne una necessità politica, un antidoto alla crisi che minacciava l'identità stessa della scienza.

Il programma di riforma del sistema dell'informazione scientifica che Lederberg e Garfield concepivano in quegli anni nella corrispondenza privata era ancora più radicale e prospettava un'autentica rivoluzione dell'assetto

¹³⁶ Citato in: P. Wouters, *The Citation Culture...cit.*, p. 63.

esistente, ispirata alle tesi di John D. Bernal: sottrarre agli editori commerciali il controllo delle riviste fondamentali; debellare l'anarchia causata dalla proliferazione dei periodici istituendo un sistema centralizzato di archiviazione e disseminazione selettiva di tutta la letteratura scientifica; sostituire i periodici correnti con una serie di *abstract journals* e pubblicare i risultati più rilevanti della ricerca in un *Daily Newspaper of Science* con annesso indice di citazioni giornaliero.¹³⁷ L'indice di citazioni, in virtù della struttura interdisciplinare e della capacità di connettere concetti scientifici al di là di barriere tematiche e temporali, era il fulcro del sistema, la stanza di compensazione dei crediti che i ricercatori si dispensavano a vicenda ed una bussola per non perdersi nell'enciclopedia unificata delle scienze. A patto però che il comportamento citazionale degli scienziati si adeguasse alle regole di disciplina e prevedibilità imposte dal gioco. Regola numero uno: citare solo o in prevalenza la letteratura davvero rilevante per l'argomento.

Niente di tutto ciò accadde. Le pubblicazioni scientifiche restarono saldamente nelle mani degli editori commerciali e gli scienziati continuarono a citare in maniera indisciplinata, ma la conquista del mercato da parte degli indici di citazioni era cominciata e, a modo suo, lo *Science Citation Index* avrebbe messo ordine nel bazar dell'editoria scientifica concedendo visibilità citazionale solo ad un numero limitato di riviste.

Nel 1963, grazie ad un finanziamento annuale di 50 mila dollari ottenuto tre anni prima dal National Institute of Health, fu pubblicato il *Genetics Citation Index*. Qui prese forma, una volta per tutte, l'impostazione multidisciplinare del repertorio. Gli articoli di interesse genetico avevano un difetto, che in ultima analisi si rivelò un pregio: non erano contenuti in un numero ristretto di riviste di taglio specialistico, ma disseminati un pò ovunque (si pensi all'articolo di Watson e Crick uscito su *Nature* nel 1959). Una delle soluzioni adottate fu allora quella di costruire un database omnicomprensivo basato

¹³⁷ Il carteggio è documentato in: P. Wouters, *The Citation Culture...cit.*, cap. 3, pp. 59-78.

sullo spoglio delle 600 riviste coperte dai *Current Contents* e sviluppare un insieme di criteri per estrarre in modo automatico, dal database iniziale, tutto il materiale virtualmente rilevante per la genetica. Il risultato intermedio si rivelò migliore di quello finale: un indice multidisciplinare apriva orizzonti più ampi di quelli della ricerca bibliografica e lasciava intravedere nuove aree di applicazione, quali lo studio dei confini tra settori di ricerca e discipline, la mappatura della rete di interrelazioni tra scienziati e gruppi di ricerca, la valutazione dell'impatto del lavoro individuale sulla comunità dei ricercatori. L'accoglienza iniziale fu tiepida, soprattutto da parte dei bibliotecari,¹³⁸ ed anche tra gli scienziati non mancarono commenti negativi circa l'utilità effettiva dell'indice ed i pericoli insiti nel possibile abuso politico dei dati bibliografici.¹³⁹

La National Science Foundation non diede altri soldi all'ISI cosicché il lavoro poteva proseguire solo su base commerciale. Nel 1964, grazie ai proventi derivati dalla vendita dei *Current Contents* e al supporto finanziario di un gruppo di investitori di Wall Street, iniziò finalmente la pubblicazione dello *Science Citation Index* affiancato, alcuni anni dopo, dal *Social Sciences Citation Index* (1972) e dallo *Arts & Humanities Citation Index* (1978). Nella fase operativa, l'indicizzazione su larga scala imponeva due scelte iniziali: la copertura disciplinare, cioè quante e quali riviste indicizzare, e il sistema di produzione e organizzazione dell'indice. Entrambe richiedevano una soluzione che rispettasse il giusto equilibrio tra costi ed efficacia. La selezione delle riviste si svolgeva in base a criteri rigorosi. Il principio ispiratore era la legge di Bradford, almeno nell'interpretazione datane da Garfield: la letteratura scientifica che conta è concentrata in un numero relativamente basso di periodici, quindi sarebbe inutile, oltre che finanziariamente dannoso, pretendere la copertura totale. Il problema non era tanto identificare i periodici fondamentali di una disciplina, cosa che un

¹³⁸ J. Martyn, "An examination of citation indexes" *Aslib Proceedings* 1965, 17(6): 184-96.

qualunque specialista sarebbe stato in grado di fare in poco tempo, ma definire criteri qualitativi e quantitativi per espandere la copertura al di là del nucleo di riviste di "ovvia" importanza. I criteri qualitativi, gli unici utilizzabili nel caso di riviste di recente fondazione, erano abbastanza prevedibili: il giudizio di un comitato di esperti, la reputazione dell'editore, la distribuzione geografica dei membri del comitato di redazione, il curriculum degli autori e la qualità della loro precedente produzione scientifica, il rispetto di alcuni standard editoriali di base, come la puntualità nell'uscita dei fascicoli, l'accuratezza e completezza dei riferimenti bibliografici e l'inclusione di indici e abstract in lingua inglese. Requisiti essenziali, sebbene non classificabili come qualitativi, erano anche la fotoreproducibilità degli indici, un formato dei dati bibliografici predisposto alla codifica e all'elaborazione computerizzata e la completezza delle informazioni relative a nomi e indirizzi degli autori.¹⁴⁰ I criteri quantitativi invece non potevano che fondarsi su un dispositivo già azionato, ovvero il conteggio delle citazioni, che Garfield pensò di rendere più affidabile introducendo un parametro destinato a scatenare polemiche e incomprensioni, l'*impact factor* delle riviste, cioè il numero medio di citazioni ricevute dagli articoli di una rivista in un periodo di tempo predefinito. All'inizio la pressione e la benevolenza degli editori potevano condizionare in modo decisivo la scelta di un titolo,¹⁴¹ ma, una volta messa in moto, la

¹³⁹ H.B. Steinbach, "The quest for certainty: Science citation index" *Science* 1964, 145: 142-43.

¹⁴⁰ «Since we are producing contents page services and citation indexes, it is important that the journal contain the basic features needed to process it for these types of reference tools. Therefore we give careful consideration to the contents page, including its photoreproducibility, the substance of the article titles presented there, the completeness of the information, the ease with which it can be scanned, the language of the contents page, etc. The journal should also include bibliographic references if it is also to be processed for a citation index. The presentation of these references should be in a format that can easily be processed and which provides complete information. Beyond these basic elements, we are concerned with author address availability and presentation since this information is used in the *Current Bibliographic Directory of the Arts & Sciences* as well as CC and the citation indexes. The accuracy of these addresses determines the ease with which authors can communicate and request reprints»: E. Garfield, "How do we select journals for *Current Contents?*," in *Essays of an Information Scientist 1979-1980*, 4 (Philadelphia: ISI Press, 1981): 309-12, p. 310.

¹⁴¹ «...in our early years, the friendliness of publishers often influenced our selection of additional journals. In the early 1960s, we introduced the SCI and began to analyze the citations journals gave to and received from other journals. For the first time, we could quantify the relative impact of journals and identify additional ones for coverage in CC»: E. Garfield, "How do we select journals for *Current Contents?*," in *Essays of an Information Scientist 1979-1980*, 4 (Philadelphia: ISI Press, 1981): 309-12, pp. 309-10.

macchina citazionale si autoalimentava da sola e, grazie al calcolo dell'*impact factor*, poteva procedere alla selezione dei periodici sulla base di un presunto criterio oggettivo: «è difficile - scrive Garfield - immaginare una rivista importante che sfugga a questa rete di citazioni». ¹⁴²

Sul fronte della produzione materiale, avendo eliminato lo sforzo intellettuale associato alle operazioni di indicizzazione per soggetto, le fasi di lavorazione dello *Science Citation Index* erano poco più articolate di quelle di una catena di montaggio, con un faticoso lavoro di pre-editing mirato a: 1) selezionare le unità da processare per ogni fascicolo di rivista; 2) standardizzare le informazioni destinate a confluire nell'indice; 3) riformattare le citazioni disseminate nel testo e nelle note a piè di pagina o redatte con formati arcaici.

3.2. La ricerca bibliografica nello Science Citation Index

Nella versione cartacea, la cui struttura è rimasta sostanzialmente invariata in ambiente digitale, lo *Science Citation Index* combina tre indici diversi, pubblicati con periodicità trimestrale: l'indice delle fonti (*Source Index*), l'indice delle citazioni propriamente detto (*Citation Index*) e l'indice dei soggetti (*Permuterm Subject Index*). Indici separati sono previsti anche per le organizzazioni di appartenenza degli autori (*Corporate Index*) e i brevetti (*Patent Citation Index*). ¹⁴³

Il *Source Index*, ordinato alfabeticamente per nome del primo autore, comprende le registrazioni bibliografiche complete di tutti gli articoli pubblicati nei fascicoli delle riviste selezionate per l'indicizzazione. Il *Citation Index* elenca, in ordine alfabetico per nome del primo autore, tutti i documenti, editi e inediti, citati negli articoli delle fonti, riportando in forma

¹⁴² E. Garfield, *Citation Indexing: its Theory and Application...cit.*, p. 25.

abbreviata, in corrispondenza di ognuno, gli estremi bibliografici essenziali degli articoli dell'annata corrente del *Source Index* in cui tali documenti sono citati. Mentre il *Source Index* è limitato dal punto di vista tipologico (contiene solo gli articoli delle riviste "elette" dall'ISI) e cronologico (la data di pubblicazione degli articoli è quella dell'anno in cui è effettuato lo spoglio), il *Citation Index* non conosce discriminazioni di tipologia o di data: non solo libri e articoli di riviste, ma anche rapporti di ricerca, tesi, relazioni di convegni, materiali non librari, comunicazioni personali; non solo scienziati contemporanei e operanti nei fronti di ricerca attivi ma anche, se qualcuno li ha citati, Leonardo da Vinci, Copernico, Galileo Galilei. Un'edizione annuale dello *Science Citation Index* contiene quindi sia una percentuale rilevante di articoli significativi pubblicati in quel dato anno (*Source Index*), sia una grossa percentuale di documenti significativi pubblicati negli anni precedenti (*Citation Index*).

L'indice dei brevetti (*Patents Index*), ordinato per numero di brevetto, elenca in una sezione separata tutti i brevetti (americani ed esteri) citati in una delle riviste coperte dallo *Science Citation Index* e fornisce per ognuno informazioni relative all'anno di emissione, al nome dell'inventore, alla nazione di provenienza.

L'indice permutato delle parole chiave (*Permuterm Subject Index*) è costruito con le parole dei titoli degli articoli schedati nel *Source Index*, ed utilizza le combinazioni di tali parole come voci di soggetto in corrispondenza delle quali si trovano elencati i nomi dei primi autori degli articoli di provenienza. Permutazione è intesa qui nel senso matematico di "tutti i possibili modi di ordinare un certo numero di oggetti": dato un titolo, l'indice permutato risulta da tutte le possibili coppie ordinate delle parole che lo compongono. Un indice di questo tipo, oltre al vantaggio di poter essere redatto interamente dal

¹⁴³ Nell'edizione cartacea il *Citation Index* precede il *Source Index*. Un'esposizione chiara e concisa della struttura dello *Science Citation Index* e degli indici di citazioni si trova in: M. Weinstock, "Citation indexes," in *Encyclopaedia of Library and Information Science*, 5 (New York: Marcel Dekker, 1971): 16-40.

calcolatore, ha la prerogativa, secondo Garfield e i partigiani dell'ISI, di fondarsi sul linguaggio naturale della scienza e della tecnologia anziché sul letto di Procuste di un sistema di classificazione o di una lista più o meno arbitraria di termini precompilati.¹⁴⁴

Nell'architettura dello *Science Citation Index*, la ricerca per parola o per soggetto serve solo a trovare un punto di partenza - un documento o un autore che abbia scritto qualcosa di attinente all'argomento indagato - utile ad azionare la leva del *citation index*. Una volta trovato il nome dell'autore, si cerca la voce corrispondente in una delle edizioni dello *Science Citation Index* e si risale, dalle segnalazioni bibliografiche abbreviate dei suoi contributi (*cited references*), a quelle degli articoli in cui sono stati citati (*citing articles*); si localizzano poi, mediante il *Source Index*, le registrazioni bibliografiche complete degli articoli citanti e, se uno di questi risulta particolarmente rilevante, si ripete la procedura utilizzando il nuovo documento o uno dei documenti citati nella sua bibliografia come chiave di ricerca (*citation cycling*). L'esito della strategia dipende in larga misura dalla rilevanza e dall'effettiva aderenza del documento iniziale (*seed document*) ai bisogni informativi del ricercatore.

Oltre alle ricerche per argomento, in un *citation index* si possono compiere diversi altri tipi di indagine, dalla più banale, come la verifica di un riferimento bibliografico incompleto, ad altre meno scontate:¹⁴⁵

¹⁴⁴ «The third section of the *SCI* is the *Permuterm Subject Index*. Briefly, it uses words from the titles of articles listed in the *Source Index* as subject headings referring to the first author in the *Source Index* entries. The *Permuterm Subject Index* is thus a so-called 'natural language' index. It uses the current language, that is the current terminology, of science and technology as its subject headings, not a terminology that has been chopped or stretched to the arbitrary demands of some Procrustean hierarchy, classification, category listing, etc. »: E. Garfield, *The Science Citation Index and ISI's Journal Citation Reports: their implications for journal editors. Presented at: Third General Assembly of the European Association of Editors of Biological Periodicals. Paris, 10—12 May 1976, 1976* <<http://www.garfield.library.upenn.edu/papers/255.html>>.

¹⁴⁵ E. Garfield, *Citation Indexing: its Theory and Application...cit.*, cap. 5, pp. 41-61. Esempi concreti di ricerche complesse condotte sulla versione elettronica dello *Science Citation Index* si trovano in: A.E. Cawkell, "Checking research progress on 'image retrieval by shape-matching' using the Web of Science" *Aslib Proceedings* 1998, 50(2): 27-31; A.E. Cawkell, "Methods of information retrieval using Web of Science: pulmonary hypertension as a subject example" *Journal of Information Science* 2000, 26(1): 66-70.

- 1) ricerche eponimiche, riferite ad argomenti etichettati con il nome di una persona (linfoma di Hodgkin, paralisi di Bell): non sempre le liste di *subject headings* registrano tali argomenti sotto voci specifiche, mentre in un *citation index* è sufficiente utilizzare come chiave di ricerca il documento "primordiale" associato al nome eponimico, a patto ovviamente che esso non sia stato obliterato diventando un dato acquisito, motivo per cui nessuno lo cita più formalmente;
- 2) ricerche metodologiche: spesso gli autori scientifici citano per identificare metodologie e tecniche d'indagine specifiche che non emergono dall'analisi delle parti significative di un documento condotta secondo i dettami dell'indicizzazione convenzionale per soggetto;
- 3) ricerche di tipo *follow-up*: frequenti in aree come la chimica o la medicina, servono a seguire gli sviluppi di un particolare evento (una reazione chimica, una terapia, una tecnica chirurgica) a partire dal documento in cui è stato descritto per la prima volta;
- 4) ricerche su questioni specifiche che necessitano di "profondità" storica, ad esempio per sapere se una data teoria o ipotesi scientifica è stata confermata o smentita nella letteratura successiva;
- 5) ricerche interdisciplinari, nelle quali un approccio per argomento richiederebbe la consultazione di diversi indici specializzati: in un indice di citazioni la pertinenza disciplinare non ha un valore assoluto, ma dipende dai percorsi di utilizzo della letteratura precedente da parte degli autori attivamente impegnati in un fronte di ricerca, quindi la ricerca per citazioni può rivelare legami intellettuali che sarebbe difficile inquadrare in un set precostituito di *subject headings*;
- 6) ricerche per similarità, familiari agli utenti di database e motori di ricerca del Web attraverso funzioni come *Related Articles* (MEDLINE) o *Pagine Simili* (Google): puntano al recupero di documenti affini ad un documento di partenza per la condivisione di uno o più riferimenti bibliografici

(*bibliographic coupling*), oppure di documenti che si riferiscono contemporaneamente, nelle loro bibliografie, a due o più documenti di partenza (*co-citation*);

6) ricerche rapide per ricostruire lo “stato dell’arte” in un settore: un indice di citazioni permette di identificare gli articoli su un certo argomento che contengono un numero di riferimenti bibliografici superiore ad una soglia prefissata e che pertanto offrono, con buona probabilità, una sintesi della letteratura precedente.

Non esiste un test definitivo che permetta di confrontare su una scala assoluta l'efficienza di una strategia di ricerca basata sugli indici di citazioni con quella di un'interrogazione convenzionale per soggetto o parola chiave. Documenti citanti e documenti citati hanno in genere affinità di contenuto linguistico,¹⁴⁶ ma non mancano evidenze contrarie: è possibile infatti che un documento sull'argomento **A** sia citato da un documento sull'argomento **B** senza che tra i due si possa stabilire una corrispondenza terminologica a livello di titoli, abstract, parole chiave.¹⁴⁷ L'uso combinato di citazioni e indici convenzionali (parole del titolo, parole chiave, soggetti) incrementa, secondo alcuni studi, le performance di un sistema di *information retrieval*.¹⁴⁸ Esperimenti condotti su MEDLINE e altri database biomedici mostrano nondimeno che i risultati ottenuti applicando alternativamente le due metodologie allo stesso set di termini di ricerca sono solo in minima parte sovrapponibili e formano

¹⁴⁶ H.P.F. Peters; R.R. Braam; A.F.J. van Raan, “Cognitive resemblance and citation relations in chemical engineering publications” *Journal of the American Society for Information Science* 1995, 46(1): 9-22.

¹⁴⁷ S.P. Harter; T.E. Nisonger; A. Weng, “Semantic relationships between cited and citing articles in library and information science journals” *Journal of the American Society for Information Science* 1993, 44(9): 543 - 52.

¹⁴⁸ G. Salton, “Automatic indexing using bibliographic citations” *Journal of Documentation* 1971, 27(2): 98-110; M.L. Pao, “Term and citation retrieval - a field study” *Information Processing & Management* 1993, 29(1): 95-112; M.L. Pao & D.B. Worthen, “Retrieval effectiveness by semantic and citation searching” *Journal of the American Society for Information Science* 1989, 40(4): 226-35; Y. He; S.C. Hui; A.C.M. Fong, “Citation-based retrieval for scholarly publications” *Intelligent Systems, IEEE* 2003, 18(2): 58-65.

piuttosto insiemi complementari.¹⁴⁹ Un rimedio per colmare questa distanza, suggerito dalle prime revisioni critiche dei *citation indexes* negli anni '70, è "contestualizzare" le citazioni in modo da associare ad ognuna di esse un coefficiente testuale che ne illustri il ruolo in un dato documento.¹⁵⁰ Un'altra ipotesi è considerare le citazioni bibliografiche alla stregua di una delle molteplici rappresentazioni cognitive del contenuto di un documento utilizzabili per ridurre l'incertezza in fase di ricerca. Un sistema ottimale di *information retrieval*, secondo i fautori di questa tesi, deve sovrapporre tante diverse rappresentazioni di un testo quante sono necessarie per soddisfare, con un alto tasso di probabilità, i bisogni informativi dei suoi potenziali destinatari.¹⁵¹

La moltiplicazione di banche dati testuali accessibili via Web ha evidenziato i limiti dei sistemi convenzionali di *information retrieval*, costruiti attorno agli indici invertiti e alle tecniche booleane d'interrogazione e basati sulla semplice analisi lessicale delle parole. Il limite principale è l'incapacità di risolvere in maniera soddisfacente il problema della rilevanza dei dati recuperati, un problema essenzialmente semantico. La rilevanza dipende infatti dal giudizio di colui che deve valutare la corrispondenza dei risultati ottenuti alle domande

¹⁴⁹ K.W. McCain, "Descriptor and citation retrieval in the medical behavioral sciences literature: Retrieval overlaps and novelty distribution" *Journal of the American Society for Information Science* 1989, 40(2): 110-14. In termini di complementarietà si esprime anche Garfield: E. Garfield, *Citation-based and descriptor-based search strategies*, 1994 <<http://www.isinet.com/essays/useofcitationdatabases/3.html/>>.

¹⁵⁰ Alcuni progetti in tale direzione sono falliti a causa dei costi eccessivi legati all'implementazione delle informazioni contestuali in un sistema di *information retrieval* su larga scala. Due di tali progetti sono descritti in: B.-A. Lipetz, "Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators" *American Documentation* 1965, 16(2): 81-90; E.B. Duncan, "Qualified citation indexing: its relevance to educational technology," in *Information Retrieval in Educational Technology. Proceedings of the First Symposium on Information Retrieval in Educational Technology, held at ETIC '81, Aberdeen, Scotland, 1st April, 1981*, edited by E.B. Duncan, et al. (Aberdeen: University of Aberdeen, 1981): 70-79.

¹⁵¹ Questa tesi deriva da un approccio cognitivo all'*information retrieval* che tiene conto, in fase di progettazione, sia della solidità del sistema e dell'efficacia degli algoritmi che degli aspetti psicologici e comportamentali dei potenziali utenti. Vedi: P. Ingwersen, *Information Retrieval Interaction* (London: Taylor Graham, 1992) <http://www.db.dk/pi/iri/files/Ingwersen_IRI.pdf>; P. Ingwersen, "Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory" *Journal of Documentation* 1996, 52(1): 3-50. Un tentativo recente di investigare, da un punto di vista cognitivo, il ruolo delle citazioni bibliografiche in un sistema automatico di indicizzazione e recupero dell'informazione si trova in: B. Larsen, *References and Citations in Automatic Indexing and Retrieval Systems - Experiments with the Boomerang Effect*. PhD dissertation (Copenhagen: Department of Information Studies, Royal School of Library and Information Science, 2004) <<http://www.db.dk/blar/dissertation>>.

formulate. Le incertezze linguistiche in fase di indicizzazione e di ricerca e l'aumento esponenziale del volume delle informazioni trattate rendono problematica la corrispondenza e conducono il più delle volte a risultati generici, che richiedono uno sforzo manuale non indifferente per separare i dati significativi da quelli inconsistenti. Si spiegano allora i tentativi di sviluppare, nella tradizione di ricerca sulle reti neurali artificiali, sistemi di apprendimento automatico capaci di riconoscere in maniera autonoma similarità, relazioni, regolarità, quindi di raggruppare i documenti in classi (*clusters*) senza ricorrere a griglie di classificazione precostituite.¹⁵² In tale contesto, richiamo e precisione non sono più gli unici parametri coi quali valutare un sistema di recupero dell'informazione: se l'utente ha già interrogato diverse banche dati ed ha estratto da ognuna di esse un certo numero di documenti rilevanti, è probabile che la sua richiesta successiva sia quella di interagire con un sistema che gli restituisca informazioni nuove,¹⁵³ non sovrapponibili con quelle già in suo possesso. Un indice di citazioni si presta più di altri strumenti a soddisfare questa esigenza proprio perché i riferimenti bibliografici funzionano come indicatori estremamente flessibili del contenuto di un documento e possono costituire il punto di innesto di reti semantiche diverse. Analogamente, quando l'indagine coinvolge settori interdisciplinari o in via di formazione, una ricerca per citazioni può rivelarsi più efficace di una ricerca per termini di un dizionario controllato, a causa del ritardo congenito dei linguaggi di indicizzazione e della maggiore aderenza dei riferimenti bibliografici alle dinamiche concrete di produzione della letteratura scientifica.

¹⁵² Un'applicazione delle reti neurali, in particolare del sistema delle mappe di Kohonen (*self-organizing maps*), al problema del recupero dell'informazione da un database di citazioni è descritta in: S.C. Hui & A.C.M. Fong, "Document retrieval from a citation database using conceptual clustering and co-word analysis" *Online Information Review* 2004, 28(1): 22-32.

¹⁵³ Vedi ad esempio: A. Fujii & T. Ishikawa, "A novelty-based evaluation method for information retrieval," in *Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC-2000)*, Athens, Greece 31.5.-2.6.2000 (Paris: European Language Resources Association, 2000): 1637-41.

Garfield ha esaltato le potenzialità dello *Science Citation Index* per la ricerca bibliografica con diversi argomenti. Il più interessante sul piano biblioteconomico asserisce che le citazioni bibliografiche, nella misura in cui identificano lavori precedenti su argomenti affini a quelli trattati nel documento che le produce, incorporano un sistema di indicizzazione della letteratura scientifica. Tale sistema è più stabile di quello basato sui *subject headings* perché le citazioni, a differenza delle voci di soggetto, non subiscono il processo di rapida obsolescenza tipico del linguaggio discorsivo della scienza e della tecnologia.¹⁵⁴ La differenza però non è solo di grado o di resistenza all'invecchiamento, ma di valore e portata cognitiva. Un indice di citazioni è in grado infatti di ripristinare legami tra documenti che sfuggono ai linguaggi di indicizzazione convenzionali: l'argomento o gli argomenti di una pubblicazione, così come si trovano formulati nelle voci di soggetto, sono senza dubbio significativi ad un macro (in una monografia) o ad un micro (in un articolo) livello di analisi, ma, ad un livello di analisi submicro o "molecolare", dove lo scienziato è alle prese con la costruzione del suo impianto teorico-discorsivo, non esistono argomenti facilmente traducibili in *subject headings*, esistono invece "idee" o "eventi" particolari che si connettono ad altre "idee" e ad altri "eventi", e questa correlazione è meglio espressa da un legame di tipo citazionale. Le citazioni rivelano pertanto legami intellettuali tra la letteratura del passato e quella del presente, tra particolari eventi scientifici descritti nei documenti citati e la sorte cui sono andati incontro nei documenti citanti (conferme, estensioni, correzioni, smentite).¹⁵⁵ La scarsa abilità degli indicizzatori di stabilire correlazioni tra

¹⁵⁴ E. Garfield, *Citation Indexing: its Theory and Application...cit.*, p. 41.

¹⁵⁵ «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 indexers are inadequate, because the scientist is quite often concerned with a particular idea rather than with a particular concept. "Thought" indexes can be extremely useful if they are properly conceived and developed»: E. Garfield, "Citation Indexes for science...cit.", p. 468. A rafforzare il concetto: «Citation indexing goes beyond the function of categorizing the literature. It explicitly reveals the intellectual relationships that exist between old and new literature. Each article is a published record of a particular event in the process of scientific development. The citation index shows the relationships between individual events

idee, concetti, argomenti viene così rimpiazzata «dalla capacità di gran lunga superiore dell'intera comunità scientifica di fare la stessa cosa».¹⁵⁶

3.3. L'enciclopedia come impresa commerciale: ISI Web of Knowledge

Oggi è quasi scontato, navigando nei portali dei grandi editori e aggregatori commerciali di database, muoversi senza soluzione di continuità da un riferimento bibliografico al testo completo dell'articolo citato. Gli archivi digitali di Elsevier, Blackwell, Wiley, Springer, EBSCO, OVID contengono un indice dei riferimenti bibliografici collegati, quando possibile, al full-text degli articoli. Analogamente i maggiori servizi di *abstracting* e *indexing* della letteratura scientifica hanno cominciato da tempo, in forme più o meno raffinate, ad arricchire i record dei loro database con un campo di *cited references*: *CINHAL - Current Index to Nursing and Allied Health Literature* ci aveva già pensato alla metà degli anni '90, *PsychINFO* ha seguito l'esempio nel 2001, fino ad arrivare agli ambiziosi progetti universalistici di Scopus e Google Scholar nel 2004.¹⁵⁷

CrossRef,¹⁵⁸ il servizio di *reference linking* nato alla fine del 1999 da un'iniziativa di dodici editori internazionali, sia commerciali che no-profit, è probabilmente la migliore esemplificazione di questa potenzialità. Grazie alla tecnologia DOI (Digital Object Identifier), esso permette all'utente che gode degli opportuni permessi di muoversi senza soluzione di continuità dalla

at different points in time. That makes a citation index particularly effective in telling us what has happened to some idea or experiment - whether it has been confirmed, extended, improved, tried, or corrected»: E.

Garfield, *Citation Indexing: its Theory and Application...cit.*, p. 41.

¹⁵⁶ Ibid., p. 9.

¹⁵⁷ P. Jacsó, "Citation-enhanced indexing/abstracting databases" *Online Information Review* 2004, 28(3): 235-38.

¹⁵⁸ <<http://www.crossref.org/crossrefsearch.html>>. Sul ruolo delle citazioni bibliografiche in CrossRef vedi: A. Brand, "CrossRef turns one" *D-Lib Magazine* 2001, 7(5) <<http://www.dlib.org/dlib/may01/brand/05brand.html>>.

citazione bibliografica al testo completo di un articolo ospitato in uno qualunque dei server consociati.

Gli standard per la definizione di linguaggi di marcatura, in particolare SGML e soprattutto XML, nonché le norme internazionali per la redazione di manoscritti digitali (ISO 12083:1994), hanno facilitato enormemente la vita della citazione bibliografica nel nuovo ambiente. In virtù di queste codifiche infatti le unità informative di un testo, nel caso specifico le singole voci di una bibliografia, possono essere marcate e rese riconoscibili da un'etichetta che ne facilita il collegamento con il full text degli articoli citati o l'inserimento automatico in un indice di un motore di ricerca appositamente istruito. In tale contesto, un aspetto importante del successo commerciale degli editori è stata la capacità di sfruttare le tecnologie e i protocolli disponibili per moltiplicare la visibilità e reperibilità dell'informazione venduta. Se una strategia di mercato è tanto più efficace quanto più, partendo da un prodotto, riesce a stimolare e soddisfare il bisogno di un altro prodotto della stessa linea, allora la citazione bibliografica non si sottrae al gioco: in un documento opportunamente formattato, essa è il canale attivo di commercializzazione di tutti i documenti ad esso collegati per mezzo di legami ipertestuali.

Sullo sfondo di questo paesaggio oramai consueto si collocano gli indici di Garfield e i capitoli della loro storia recente. Con uno scarto decisivo però rispetto agli altri archivi digitali: in un database convenzionale l'informazione si cerca punto e basta, tutto ciò che succede prima e dopo la ricerca, l'uso che è possibile fare dell'informazione trovata, è fuori dai confini logici del database; nelle banche dati ISI l'informazione si cerca e nello stesso istante in cui viene trovata ha qualcosa "in più" da dire oltre alla corrispondenza in termini di richiamo e precisione con i termini di ricerca, è un'informazione targata, con un peso specifico e un valore che derivano dal fatto di essere iscritta in un sistema di valutazione bibliometrica dell'impatto e della qualità

dei documenti (e degli autori) fondato sul conteggio e sull'analisi delle citazioni.

Nel 1992 l'ISI è stata acquisita dalla Thomson Business Information, una sezione della Thomson Corporation, che ha traghettato l'indice di Garfield nell'universo delle reti telematiche e del World Wide Web, esaltandone la natura interdisciplinare e la capacità di integrare contenuti e strumenti analitici di provenienza e finalità differente sulla base del comune denominatore rappresentato dalla citazione bibliografica. La tecnologia dell'ipertesto ha annullato di fatto la separazione fisica tra indici di citazioni e indici di parole chiave, di conseguenza lo *Science Citation Index* del XXI secolo può aspirare alla sintesi di almeno due tradizioni di acquisizione, organizzazione e manipolazione dell'informazione: quella degli indici citazionali di Garfield e della tradizione giuridica e quella dei più importanti servizi novecenteschi di *abstracting e indexing* della letteratura periodica specializzata.

L'attuale struttura e funzionalità di ricerca degli indici citazionali sono il risultato di un'evoluzione tecnica che il suo fondatore difficilmente avrebbe potuto prevedere. Il punto di partenza è rimasto lo stesso, cioè cattura e descrizione standardizzata di tutti i riferimenti bibliografici collegati agli articoli scientifici pubblicati su un certo numero di riviste scientifiche internazionali (per ora circa 8700), ma solo 40 anni dopo essere stata concepita, la rete citazionale è divenuta percorribile fino nelle maglie più profonde grazie ai protocolli di Internet e alla tecnologia dell'ipertesto. La navigazione ipertestuale ha reso tangibili e fisicamente trasparenti i legami concettuali tra documenti incorporati nelle citazioni bibliografiche. Navigando tra i link si possono ora scoprire le fonti di un dato documento o, nella direzione opposta, gli articoli che lo hanno citato, si possono identificare documenti ad esso affini grazie alla tecnica della co-citazione, si possono visualizzare le citazioni in comune tra due documenti, in modo da stabilire il loro grado di correlazione, si può cambiare il criterio di ordinamento dei

risultati in conformità ad una serie di parametri quali autore, istituzione, tipo di documento, linguaggio, anno di pubblicazione, soggetto, titolo della rivista. *ISI Web of Knowledge* è la piattaforma Web che integra le due creature di Garfield, *Web of Science* (gli indici di citazioni) e *Current Contents Connect* (gli spogli di periodici e monografie), con una serie di fonti informative di diversa provenienza e con gli strumenti avanzati di valutazione scientometrica sponsorizzati dall'ISI. Negli ultimi anni, il servizio è stato potenziato sia attraverso l'hosting di banche dati prodotte da partner commerciali, quali *INSPEC*, *CAB Abstract*, *FSTA (Food Science and Technology Abstracts)*, *PsycINFO* e *BIOSIS Preview* (di recente acquistata dalla stessa Thomson), sia mediante l'accesso a contenuti di banche dati bibliografiche e full text esterne, quali *AGRICOLA*, *PubMed*, *NCBI GenBank*, *arXiv.org*, *Popline*, la *Cochrane Library*, l'*AIAA (American Institute of Aeronautics and Astronautics) Meeting Papers*, l'*ASCE Civil Engineering Database*, il *National Technical Information Service*, *AskEric*.

Le risorse che compongono *ISI Web of Knowledge* promettono una copertura bibliografica enciclopedica: fonti specialistiche e interdisciplinari, riviste, monografie, atti di convegni, preprint, brevetti, formule chimiche, siti web selezionati, fondi di ricerca. L'informazione venduta è ad un tempo multidisciplinare e selettiva: multidisciplinare perché, come recita un saggio di presentazione dell'ISI, è il fronte stesso della ricerca scientifica che «diventa di natura sempre più interdisciplinare»; selettiva perché, con l'esplosione del Web, «la necessità di trovare informazione è stata rimpiazzata dalla necessità di trovare informazione rilevante; il bisogno della qualità ha di gran lunga oscurato quello della quantità».¹⁵⁹ Si tratta degli stessi argomenti con cui, negli anni '60, Garfield e Lederberg legittimavano l'utilità degli indici di citazioni davanti alla crisi informativa causata dalla marea di documentazione scientifica esistente.

Le principali banche dati accessibili all'interno della piattaforma possono essere raggruppate in quattro grandi aree.

1) Il portale *Web of Science* è la spina dorsale del sistema, in quanto comprende i tre indici che hanno garantito fama e ricchezze all'ISI, con il loro patrimonio di circa 8700 riviste indicizzate: *Arts & Humanities Citation Index* (A&HCI), con copertura retrospettiva dal 1975, *Social Sciences Citation Index* (SSCI), con copertura dal 1956, e *Science Citation Index Expanded* (SCIE), con copertura dal 1945. L'iniziativa *Century of Science*, lanciata nel gennaio 2005, ha dilatato ulteriormente i confini temporali del database avviando il recupero e lo spoglio di oltre 260 periodici scientifici pubblicati nella prima metà del XX secolo.¹⁶⁰ In aggiunta, il portale ospita due edizioni di indici settoriali per la chimica, l'*Index Chemicus* e il *Current Chemical Reactions*.

2) Il *Derwent Innovations Index*, nato dalla fusione del *Derwent World Patents Index* e del *Patents Citation Index*, indicizza oltre 22 milioni di brevetti rilasciati da 40 autorità internazionali dal 1963 ad oggi, con abstract riscritti dagli indicizzatori dell'ISI. Il database è corredato da *Derwent Analytics*, un software di data-mining per effettuare correlazioni e analisi di tendenze di natura economica e strategica.

3) Il database *ISI Proceedings* indicizza atti di convegni a livello internazionale sia in ambito scientifico che umanistico, con copertura retrospettiva dal 1990.

4) Il database *Current Contents Connect*, erede dei *Current Contents*, effettua lo spoglio di circa 7600 riviste, 2000 libri e oltre 4400 siti Web selezionati. Per questi ultimi, raggruppati nel database *ISI Current Web Contents*, è disponibile una funzione di ricerca nei full-text delle pagine Web indicizzate.

¹⁵⁹ H. Szigeti, *The ISI Web of Knowledgesm Platform: Current and Future Directions*, 2001 <<http://www.isinet.com/media/presentrep/essayspdf/wokplat.pdf>>, p. 4.

Gli utenti del servizio possono ora avvalersi anche della funzione *eFirst*, che indicizza gli articoli resi disponibili dagli editori prima della pubblicazione a stampa.

L'eterogeneità delle risorse è bilanciata dalla possibilità di effettuare ricerche trasversali per parola chiave mediante una tecnologia (*MuscatDiscovery*) che, intervenendo “dietro le quinte” sull’informazione strutturata (articoli, monografie) e su quella non strutturata (pagine Web, file PostScript, pdf, txt), filtra i risultati e li ordina secondo un criterio di rilevanza di tipo probabilistico.¹⁶¹

Il portale ha un'infrastruttura di linking molto solida che, oltre a sovrintendere alle operazioni di smistamento e controllo degli accessi, garantisce un sistema centralizzato di gestione delle diverse tipologie di link tra contenuti: a) link intra-documentali, ad esempio tra un articolo di rivista e l'indice del fascicolo in cui si trova oppure tra un articolo e il conteggio delle volte in cui è stato citato; b) link inter-documentali, ad esempio tra un articolo di rivista e le reazioni chimiche in esso trattate oppure tra un brevetto e gli articoli citati nella sua bibliografia; c) link esterni con il full-text degli articoli ospitati sui server dell'ISI o di altri fornitori (editori, *open archives*), con servizi di pay per view, con OPAC locali; d) link esterni di tipo *context-sensitive*, capaci di garantire, grazie al protocollo di interoperabilità OpenURL, la personalizzazione degli accessi in funzione di specifiche esigenze istituzionali (sono supportati, per esempio, i link server SFX e 1Cate).¹⁶²

¹⁶⁰ Thomson ISI, *Century of Science*TM, 2005

<<http://www.isinet.com/media/presentrep/facts/centuryofscience.pdf>>

¹⁶¹ Le due interfacce di ricerca unificata si chiamano *CrossSearch* e *Current Contents eSearch*. Nella nuova versione di *CrossSearch* è possibile valicare i confini di *Web of Knowledge* fino ad includere nei risultati della ricerca anche banche dati esterne di tipo commerciale o ad accesso gratuito.

¹⁶² «The idea behind this gateway environment – just like the idea behind the platform as a whole – is simple in theory: provide a dependable system that identifies who a user is (authentication), where that user wants to go (subscription access), and then sends them to the right place (appropriate copy)»: H. Szigeti, *The ISI Web*

Il collegamento esterno al full text degli articoli coinvolge anche editori e archivi che non collaborano direttamente con l'ISI nell'aggiornamento dei metadati necessari a localizzare in maniera stabile una risorsa nel tempo: una tecnologia di linking "algoritmico" o "ottimistico" (*RoboLinks*) permette infatti di prevenire il cambiamento inatteso di indirizzo di una risorsa e di ricostruire a distanza di tempo, con buona probabilità di successo, il link per raggiungerla.

Nel marchingegno appena descritto, gli ingranaggi vitali sono ancora e sempre quelli innescati dalla citazione bibliografica e dall'insieme degli usi extra-bibliografici che, dalla metà degli anni '50 in poi, sono lì fuori ad attenderla. In un database bibliografico tradizionale si cerca un titolo con l'idea che, una volta trovato, bisogna leggere il documento per accertarsi del suo valore. Nella rete dell'ISI il documento è già "pesato", battezzato, con tanto di indice scientometrico attaccato al collo, pertanto è inevitabile che la ricerca bibliografica per citazioni trovi il suo sbocco naturale in alcuni strumenti di analisi e valutazione bibliometrica dei dati: a) il *Journal Citation Reports*, che contiene il resoconto statistico delle relazioni tra documenti citanti e documenti citati in termini di impatto citazionale delle riviste; b) l'*Essential Science Indicators*, che classifica, in funzione del numero di citazioni ricevute, articoli, autori, istituzioni, nazioni e permette di individuare, nell'ambito di 22 aree di ricerca scientifica, sia gli *Highly Cited Papers* (gli articoli più citati negli ultimi dieci anni) che gli *Hot Papers* (gli articoli più citati negli ultimi due anni); c) *HighlyCited.com*, che pubblica e rende accessibile gratuitamente la classifica e le generalità dei 250 ricercatori più citati al mondo.

Capitolo 4

Mappe e paradigmi: la citazione bibliografica al servizio della storia e della sociologia della scienza

La scienza è un mosaico di specialità - un puzzle (*jigsaw puzzle*) preferiscono dire i bibliometri - il cui disegno generale è ignoto agli scienziati, per lo più abituati a vedere e manipolare solo le singole tessere. Il piano sul quale il mosaico si distende, al pari dei contorni delle tessere, non è fisso, ma in costante movimento: cambia in accordo con le trasformazioni interne delle pratiche disciplinari; cambia anche, inevitabilmente, a seconda della prospettiva teorica e dei metodi con cui viene indagato.

Costruire mappe dettagliate delle specialità scientifiche, dei temi ricorrenti, delle aree di ricerca attive e di quelle in via di formazione, è uno degli obiettivi degli studi quantitativi della scienza. La mappa, è noto da tempo, non riproduce fedelmente un paesaggio. È un modello, una semplificazione della realtà che, nell'atto stesso in cui sceglie cosa rappresentare e come rappresentarlo, rivela il proprio contesto sociale di produzione e diventa una tecnologia di potere nelle mani del cartografo, del committente o del fruitore.¹⁶³

Per costruire una mappa in due o tre dimensioni di un campo di ricerca la bibliometria sfrutta generalmente il principio della co-occorrenza di informazioni bibliografiche: più due elementi compaiono assieme nello stesso documento, maggiore la probabilità che essi abbiano una correlazione

¹⁶³ Lo studio delle mappe geografiche come costruzioni sociali, testi da decostruire anziché fedeli rappresentazioni di un paesaggio, è stato condotto alla fine degli anni '80 da J.B. Harley sulla scorta di Foucault, Derrida e Barthes: J.B. Harley, "Maps, knowledge and power," in *The Iconography of Landscape: essays on the Symbolic Representation, Design and Use of Past Environments*, edited by D. Cosgrove, et al. (Cambridge: Cambridge University Press, 1988): 277-312. Vedi sullo stesso punto: U. Eco,

strutturale con i concetti caratteristici del campo e che la loro posizione sulla superficie della mappa sia rappresentata da punti vicini. Una mappa di questo tipo si può fare, per esempio, estraendo da un corpo di letteratura specializzata le co-occorrenze di termini (*co-word analysis*): parole del titolo o dell'abstract, codici di classificazione oppure termini scelti convenzionalmente (da autori o bibliotecari) per rappresentare il contenuto dei documenti.¹⁶⁴ Così pure si possono rilevare le co-occorrenze degli autori dei documenti e confidare nel fatto che la frequenza di tali correlazioni rifletta la struttura sociale di un campo.¹⁶⁵ Allo stesso modo, accettando il postulato che le citazioni sono parole di un linguaggio simbolico altamente specializzato le cui regole d'uso dipendono dalle funzioni normative discusse nel paragrafo 2.2., la mappa della scienza si può fare registrando le co-occorrenze di citazioni bibliografiche. Se la citazione non è un orpello inutile o un artificio retorico del testo, ma un dispositivo non ambiguo di trasmissione di idee e concetti (funzione cognitiva), e se gli scienziati citano seguendo un codice istituzionale di buona condotta (funzione sociale), allora la ricostruzione della rete delle citazioni tra articoli prodotti in un'area disciplinare o in un periodo storico determinato rispecchia la struttura e l'evoluzione della comunicazione scientifica in quell'area disciplinare o lungo quel periodo storico.

“Dell'impossibilità di costruire la carta dell'impero 1 a 1,” in *Il secondo diario minimo* (Milano: Bompiani, 1992): 157-63.

¹⁶⁴ L'analisi delle *co-words* è stata sviluppata agli inizi degli anni '80 dal Centre de Sociologie de l'Innovation dell'École des Mines di Parigi: J.P. Courtial & M. Callon, “Is indexing trustworthy? Classification of articles through co-word analysis” *Journal of Information Science* 1984, 9(2): 47-56.

¹⁶⁵ H.P.F. Peters & A.F.J. van Raan, “Structuring scientific activities by co-author analysis: an exercise on a university faculty level” *Scientometrics* 1991, 20(1): 235-55.

4.1. La storiografia algoritmica e la mappa delle influenze: HistCite

Agli inizi degli anni '60, dietro suggerimento del genetista Gordon Allen, Garfield sviluppò la seguente idea: prendere un saggio, una ricostruzione storiografica esauriente di un episodio significativo della storia della scienza e tracciare su un foglio di carta una mappa con nodi e linee (un grafo) corrispondenti alle tappe principali (autori e documenti chiave) della storia raccontata; sovrapporre poi la mappa ad un'altra ricavata estraendo dallo *Science Citation Index* le connessioni bibliografiche tra i documenti chiave identificati. I punti coincidenti delle due mappe sovrapposte avrebbero rivelato il grado di attendibilità di una storia scritta seguendo il filo d'Arianna delle citazioni. Una storia che, con poco sforzo, anche un computer avrebbe potuto scrivere, anzi disegnare.¹⁶⁶ L'esperimento venne condotto su un evento che pochi anni prima aveva rivoluzionato diversi settori della ricerca biomedica: la scoperta del codice del DNA. Il libro di Asimov *The genetic code* (1963) fu scelto come versione discorsiva ufficiale della storia, in quanto illustrava le tappe principali che condussero alla duplicazione in laboratorio del processo di sintesi delle proteine sotto il controllo del DNA. Il *Genetics Citation Index* e la prima edizione dello *Science Citation Index* servirono per estrarre i dati citazionali relativi ai documenti chiave identificati a partire dal libro di Asimov, e riferiti al periodo che va da Mendel a Nirenberg (1962). I due diagrammi risultarono sovrapponibili in molti punti. Inoltre i legami bibliografici suggerivano nuove e significative relazioni tra nodi che non

¹⁶⁶ E. Garfield; I.H. Sher; R.J. Thorpie, *The Use of Citation Data in Writing the History of Science* (Philadelphia: Institute for Scientific Information, 1964) <<http://www.garfield.library.upenn.edu/papers/useofcitdatawritinghistofsci.pdf>>. La stessa mappa cronologica della storia della genetica è stata rielaborata con metodi sociometrici di analisi delle reti in: N.P. Hummon & P. Doreian, "Connectivity in a citation network: the development of DNA theory" *Social Networks* 1989, 11: 39-63 <<http://www.garfield.library.upenn.edu/papers/hummondoreian1989.pdf>>.

emergevano dal resoconto ufficiale, confermando l'utilità storiografica degli "storio-grafi" (*historiographs*) nell'individuazione dei rapporti genealogici tra documenti.

Il software *HistCite*, messo a punto da Garfield in collaborazione con Alexander Pudovkin e Vladimir Istomin, è la versione computerizzata del procedimento manuale di mappatura cronologica della scienza appena delineato. Partendo da una bibliografia compilata dopo una ricerca per citazioni o per parola chiave in *ISI Web of Science*, si ottiene una lista di articoli di una data area tematica, ciascuno con i relativi riferimenti bibliografici. *HistCite* rielabora questa rete bibliografica, che può essere corretta ed ampliata a piacimento dall'utente, e la trasforma in un mini *citation index*, costituito da una tabella cronologica di titoli ordinati per anno, rivista, volume, pagina. Ciascun titolo è associato ad un *Global Citation Score* (numero di volte in cui è citato in *ISI Web of Science*) e ad un *Local Citation Score* (numero di volte in cui è citato nella bibliografia iniziale), in modo da rendere subito riconoscibili i nodi attorno ai quali si addensa il maggior numero di citazioni. Il sistema estrae automaticamente ed elenca in una tabella separata anche i riferimenti esterni (*outer references*), cioè documenti di qualunque tipologia che, pur non appartenendo al set bibliografico iniziale, risultano citati dagli articoli in esso contenuti e rappresentano perciò un ulteriore nucleo di candidati al ruolo di documenti chiave (*key papers*) nell'area tematica indagata. I riferimenti esterni con un tasso di citazione superiore ad una soglia prestabilita possono essere integrati nel nucleo bibliografico iniziale e l'intero set può essere riprocessato per arricchire la complessità della trama. Al termine dell'elaborazione, uno "storio-grafo" permette di visualizzare graficamente la distribuzione spaziale dei nodi e dei link corrispondenti ai documenti più citati e ai legami bibliografici che li connettono.

Ovviamente la storiografia algoritmica non ha la pretesa di scrivere la storia da sola: per farlo servono ipotesi, intuizione, capacità interpretativa ed un arsenale metodologico collaudato. Ma prima di esercitare tali competenze lo storico deve compiere l'operazione preliminare di reperire la letteratura rilevante per l'argomento e stabilire i nessi genealogici tra le unità che la compongono. Sulla scorta del riduzionismo bibliometrico di Price e dalla fiducia mertoniana nel carattere normativo delle pratiche citazionali, Garfield può così formulare «l'assunto basilare che l'informazione bibliografica contenuta in un gruppo di articoli scientifici pubblicati è sufficiente per catturare la struttura storiografica del campo».¹⁶⁷ La storia della scienza riscritta e visualizzata attraverso i legami citazionali facilita l'identificazione dei *key papers* e dei punti di discontinuità bibliografica all'interno di un'area di ricerca permettendo di risalire, dai cambiamenti nella distribuzione delle citazioni, alle trasformazioni dei concetti basilari e del modo in cui il paradigma condiviso da una comunità di scienziati viene percepito. In un saggio del 1963, Garfield prefigurava addirittura la possibilità di determinare, mediante la manipolazione computerizzata dei nodi di uno "storio-grafo", i destini alternativi della letteratura scientifica: cosa sarebbe successo, per esempio, se i lavori di Einstein sulla relatività ristretta e generale non fossero

¹⁶⁷ « In our conception of facilitating historiography - that is, writing the history of modern science - we make the basic assumption that the bibliographic information contained in a collection of published scientific articles is sufficient for the purpose of recapturing the historiographic structure of the field»: E. Garfield; A.I. Pudovkin; V.S. Istomin, "Why do we need algorithmic historiography?" *Journal of the American Society for Information Science and Technology* 2003, 54(5): 400-12 <[http://www.garfield.library.upenn.edu/papers/jasist54\(5\)400y2003.pdf](http://www.garfield.library.upenn.edu/papers/jasist54(5)400y2003.pdf)>, p. 400. Sullo stesso argomento e con diversi esempi applicativi di "storio-grafi": E. Garfield, "Historiographs, librarianship, and the history of science," in *Essays of an Information Scientist 1974-1976*, 2 (Philadelphia: ISI Press, 1977): 136-50; E. Garfield; A.I. Pudovkin; V.S. Istomin, "Algorithmic citation-linked historiography - Mapping the literature of science" *ASIS&T 2002: Information, Connections and Community. 65th Annual Meeting of ASIST*, Philadelphia, PA, November 18-21, 2002 <<http://garfield.library.upenn.edu/papers/asis2002/asis2002presentation.html>>; E. Garfield, "Historiographic mapping of knowledge domains literature" *Arthur M. Sackler Colloquium on "Mapping Knowledge Domains"*, Beckman Center of the National Academy of Sciences, Irvine, CA, May 9-11, 2003 <<http://www.garfield.library.upenn.edu/papers/knowledgedomain2003.pdf>>.

mai stati pubblicati? Quante e quali scoperte si sarebbero comunque verificate attraverso percorsi alternativi rivelati dai legami bibliografici?¹⁶⁸

Davanti ad una costellazione così ampia di opportunità, bastava una mente entusiasta e filosoficamente attrezzata per far esplodere oltre i limiti immaginabili le ambizioni conoscitive degli indici di citazioni. Già negli anni '60, Price aveva ipotizzato che fosse possibile esplorare, attraverso le cumulazioni annuali dello *Science Citation Index*, la rete mondiale delle pubblicazioni scientifiche («the total world network of scientific papers»), allo scopo di individuare il meccanismo che distingue la crescita cumulativa della scienza dall'andamento disordinato e aleatorio della non-scienza. Garfield, richiamandosi ad un testo del romanziere H. G. Wells, ha descritto tale impresa come la creazione di un gigantesco cervello mondiale, di un'enciclopedia universale dell'informazione che non deve soltanto catalogare, ma anche correlare le idee formulate nella letteratura scientifica mediante i legami documentali prodotti spontaneamente dalla rete globale delle citazioni. La rete di rinvii bibliografici simbolizza il modo in cui la «mente collettiva» della scienza plasma e organizza la letteratura disponibile.¹⁶⁹ Ma per essere utile sul piano della ricerca storica, dell'analisi sociologica e della prassi amministrativa, essa doveva proseguire sulla strada indicata dalla storiografia algoritmica ed estendersi al complesso delle relazioni interdisciplinari latenti negli indici di citazioni, allo scopo di chiarire alcune questioni di fondo: Quali sono le unità strutturali della scienza? Sono discipline istituzionalizzate come la fisica e la chimica o si tratta di unità più piccole e dai contorni meno nitidi, come le specialità o le reti informali di collaborazione che Price chiamava "collegi invisibili"? Quali forze ne garantiscono la stabilità? Che tipo di interconnessione esiste tra le singole

¹⁶⁸ E. Garfield, "Citation indexes in sociological and historical research," in *Essays of an Information Scientist 1962-1973*, 1 (Philadelphia: ISI Press, 1977): 43-46.

¹⁶⁹ E. Garfield, "Co-citation analysis of the scientific literature: Henry Small on mapping the collective mind of science," in *Essays of an Information Scientist 1992-1993. Of Nobel Class, Women in Science, Citation Classics and other Essays*, 15 (Philadelphia: ISI Press, 1993): 293-94, p. 293.

unità? Sono integrate le une con le altre o nascono e si evolvono in maniera isolata? Come muta nel tempo la struttura della scienza ad un macro e ad un micro livello di analisi?¹⁷⁰

Le visioni di Price e Garfield erano dunque destinate a tradursi in progetti concreti e tecnicamente affidabili di catalogazione della letteratura scientifica ad uso di storici, sociologi, manager della scienza. Non basta sapere “chi” ha citato “cosa” per stabilire relazioni significative tra documenti, scienziati, istituzioni, bisogna capire “dove” sono posizionati gli attori nel territorio delle discipline scientifiche e qual’è la loro forza relativa.

Serviva perciò un metodo capace di riconoscere, nella rete globale delle citazioni, similarità, percorsi, associazioni costanti, strutture. La rete bibliografica doveva perdere ogni carattere di astrazione concettuale per dispiegarsi concretamente in uno spazio a due dimensioni, in poche parole doveva tradursi in una mappa esauriente del paesaggio della scienza, segnando il passaggio dalla storia di fatti individuali alla struttura che li sottende.

Thomas Kuhn aveva insegnato agli scientometri che una comunità scientifica riesce a far progredire un settore di ricerca solo attraverso l’adesione incondizionata e quasi dogmatica dei suoi membri ad una costellazione condivisa di pratiche, metodi, strumenti, valori, teorie, credenze. In punti diversi della sua opera, egli si era riferito a tale costellazione di elementi comuni con i termini “paradigma” e “matrice disciplinare”,¹⁷¹ ed aveva sottolineato l’importanza, tra gli elementi che compongono la serie ordinata di una matrice disciplinare, di alcuni documenti “esemplari”, i quali contengono «le concrete soluzioni-di-rompicapo che, usate come modelli o come esempi,

¹⁷⁰ H. Small; E. Sweeney; E. Grenlee, “Clustering the Science Citation Index using co-citations. 2. Mapping science” *Scientometrics* 1985, 8(5-6): 321-40, p. 322.

¹⁷¹ Il concetto di “paradigma” è discusso ampiamente, sebbene in maniera non univoca, ne *La struttura delle rivoluzioni scientifiche* (1962), mentre il *Poscritto 1969* introduce il concetto di “matrice disciplinare”: T. Kuhn, “Poscritto 1969,” in *La struttura delle rivoluzioni scientifiche* (Torino: Einaudi, 1978): 209-51. Un’ulteriore precisazione del concetto si trova in: T.S. Kuhn, “Nuove riflessioni sui paradigmi,” in *La tensione essenziale. Cambiamenti e continuità nella scienza* (Torino: Einaudi, 1985): 321-50.

possono sostituire regole esplicite come base per la soluzione dei rimanenti rompicapo della scienza normale». ¹⁷²

Lo spazio d'azione per i citazionisti si apriva esattamente qui, sul perimetro di un circolo che rischiava di diventare vizioso: un paradigma o matrice disciplinare è ciò che viene condiviso dai membri di una comunità scientifica; inversamente una comunità scientifica consiste di coloro che condividono un certo paradigma; per uscire dal circolo è necessario allora che le comunità scientifiche vengano individuate senza il precedente ricorso a paradigmi. Ciò è relativamente facile per gruppi o sottogruppi professionali già istituzionalizzati: l'area di ricerca frequentata, l'appartenenza ad associazioni professionali, i periodici scientifici letti, sono parametri sufficienti per distinguere un fisico da un chimico e da un biologo o per differenziare, a sua volta, un fisico dello stato solido da un fisico delle alte energie, un chimico organico da un chimico inorganico. È al livello immediatamente inferiore, laddove un'area di ricerca ultraspecializzata è in via di formazione attorno ad alcuni documenti e autori chiave, che sorgono problemi empirici di identificazione. Ed è a questo livello che, su indicazione di Kuhn e Price, ¹⁷³ i bibliometri cominciarono a progettare tecniche efficaci per identificare, attraverso la rete dei riferimenti bibliografici reciproci, i "collegi invisibili" di studiosi che, senza essere legati da vincoli istituzionali ufficiali, collaborano a formare il fronte di ricerca attivo in una data specialità scientifica. Ammesso che le mappe bibliografiche permettano di identificare un "collegio invisibile", resta ancora da capire come si fa a riconoscere il paradigma condiviso dai suoi membri. Kuhn non ha lasciato indicazioni concrete al riguardo, anzi l'ambiguità del concetto di paradigma è stata oggetto di

¹⁷² T. Kuhn, "Poscritto 1969," in *La struttura delle rivoluzioni scientifiche* (Torino: Einaudi, 1978): 209-51, p. 212.

¹⁷³ A questo livello, scrive Kuhn, «bisogna ricorrere alla partecipazione a congressi speciali, alla distribuzione di manoscritti o di bozze prima della pubblicazione, e soprattutto a tutte le reti di comunicazione sia formali che informali comprese quelle scoperte nella corrispondenza e nei rinvii delle citazioni»: Ibid., p. 215. In nota, Kuhn cita sia Garfield che Price. Un riferimento analogo si trova in: T.S. Kuhn, "Nuove riflessioni sui

contestazione sin dalla prima edizione de *La struttura delle rivoluzioni scientifiche*.¹⁷⁴ In una riflessione successiva però, egli aveva aperto un varco alla scienza dell'informazione sostenendo che il modo in cui uno scienziato impara a riconoscere classi distinti di problemi come similari a problemi già incontrati e risolti nella letteratura è del tutto analogo al processo mediante il quale un bambino che passeggia nel giardino zoologico con suo padre impara a distinguere, grazie agli esemplari concreti di volatili osservati e riconosciuti sotto la guida dall'adulto, le oche dalle anatre e dai cigni, ed entrambi i processi possono essere facilmente simulati da un calcolatore.¹⁷⁵ Questa indicazione è stata interpretata dai bibliometri come un invito ad approfondire il problema su un registro a loro familiare. Negli anni '70 infatti, il gruppo di ricerca formatosi all'ISI attorno ad Henry Small e Belver Griffith, ha riformulato la questione nei termini dell'analisi delle citazioni. Nonostante la riluttanza di Kuhn ad una lettura bibliometrica e linguistica del concetto di paradigma,¹⁷⁶ un documento molto citato sembrava in effetti una buona approssimazione empirica della nozione di "caso esemplare". Siccome gli scienziati che risolvono problemi inerenti all'interpretazione dei processi naturali non applicano quasi mai delle regole astratte ma imparano, attraverso esempi condivisi di pratica efficace, a percepire la similarità tra un problema nuovo e problemi già risolti dai loro colleghi e predecessori, è verosimile che uno o più documenti incorporino tali modelli di pratica condivisa e per questo

paradigmi,” in *La tensione essenziale. Cambiamenti e continuità nella scienza* (Torino: Einaudi, 1985): 321-50, p. 323-325.

¹⁷⁴ D. Shapere, “The Structure of Scientific Revolutions” *Philosophical Review* 1964, 73: 383-94; M. Mastermann, “La natura di un paradigma,” in *Critica e crescita della conoscenza*, edited by I. Lakatos, et al. (Milano: Feltrinelli, 1979): 129-63.

¹⁷⁵ T.S. Kuhn, “Nuove riflessioni sui paradigmi,” in *La tensione essenziale. Cambiamenti e continuità nella scienza* (Torino: Einaudi, 1985): 321-50, pp. 339-340.

¹⁷⁶ La reazione negativa di Kuhn è documentata da Henry Small: «Kuhn’s reaction to our initial papers on the cocitation structure of science was one of puzzlement. He wondered why we had focused on highly cited papers and authors, rather than defining the total community of researchers involved with the topic, representing the complete paradigm-sharing community. Any notion that highly cited papers might stand for exemplars or other paradigmatic constructs was not on his radar. Presumably, Kuhn saw bibliometric methods only as a means of performing a social inventory of a specialty, not as way to define the paradigm itself. Therefore, it is unlikely that Kuhn would have accepted any of our bibliometric reinterpretations of his theory»: H.G. Small, “Paradigms, citations and maps of science: a personal history” *Journal of the American Society for Information Science and Technology* 2003, 54(5): 394-99, pp. 395-396.

motivo non solo ricevano più citazioni degli altri, ma il significato di tali citazioni mostri una sorprendente regolarità d'uso. Se, come pretende Small, le citazioni bibliografiche possono essere considerate simboli o marcatori intellettuali di concetti scientifici più stabili delle parole, e se da un'analisi del contesto delle citazioni risulta che la maggior parte degli specialisti in un settore disciplinare attribuisce loro lo stesso significato, allora la co-occorrenza delle stesse citazioni in un gruppo di articoli può essere considerata un buon indizio della presenza di una struttura concettuale condivisa dai loro autori. Analogamente si può ipotizzare che la sostituzione, in una disciplina o specialità, di un gruppo di documenti citati di frequente con un gruppo diverso sia un indizio dell'occorrenza di un mutamento di paradigma o rivoluzione scientifica.

All'interno di un progetto di documentazione della storia della fisica nucleare, Small progettò di disegnare una mappa della scienza che traducesse in termini visivi il concetto kuhniano di paradigma scientifico e che, in proiezione storica, permettesse di analizzare le variazioni topografiche nello scenario dei documenti esemplari in coincidenza con le trasformazioni della struttura concettuale di una specialità scientifica.¹⁷⁷ L'analisi delle coppie di citazioni o co-citazioni (*co-citation* o *cluster analysis*) si rivelò il candidato ideale per questa duplice missione.

¹⁷⁷ Nel marzo 2003 il *Journal of the American Society for Information Science and Technology* ha dedicato un fascicolo intero alle tecniche bibliometriche di visualizzazione dei paradigmi scientifici: "Special issue: Visualizing Scientific Paradigms" *Journal of the American Society for Information Science and Technology* 2003, 54(5).

4.2. La struttura disciplinare della scienza: l'analisi delle co-citazioni

L'idea di costruire coppie bibliografiche non è nata con lo *Science Citation Index*. Nel 1963 M. M. Kessler del MIT aveva ipotizzato l'esistenza di una correlazione significativa tra gruppi di articoli scientifici che condividono almeno un riferimento bibliografico. Egli verificò tale ipotesi sugli articoli pubblicati in 35 volumi di *Physical Review*, producendo una classificazione dei documenti in funzione del loro uso alternativa a quella fondata su parole chiave e thesauri disciplinari.¹⁷⁸ Sottoposte al giudizio di un profano, le classi di documenti costruite mediante coppie bibliografiche potevano anche apparire frammentarie e prive di senso, ma la mente di uno scienziato funziona in modo diverso da quella del profano e sarebbe stata in grado, secondo Kessler, di scorgere parentele e affinità concettuali che sfuggivano a quest'ultima.¹⁷⁹

Garfield e i suoi collaboratori hanno sminuito l'utilità delle coppie bibliografiche di Kessler per l'analisi sociologica, giudicandole misure statiche, retrospettive, dipendenti solo dalle citazioni contenute "nei" documenti accoppiati.¹⁸⁰ Occorreva invece un metodo capace di rappresentare le relazioni dinamiche tra documenti in accordo con gli orientamenti e le percezioni collettive delle comunità di specialisti che scrivono e pubblicano. La proposta dell'ISI si concentrò allora sull'unità analitica inversa rispetto a quella di Kessler: non le coppie di documenti che condividono una citazione,

¹⁷⁸ M.M. Kessler, "Bibliographic coupling between scientific papers" *American Documentation* 1963, 14(January-February): 10-25. Una rassegna sul metodo di *bibliographic coupling* si trova in: B.H. Weinberg, "Bibliographic coupling: a review" *Information Storage & Retrieval* 1974, 10: 189-96.

¹⁷⁹ Sull'esistenza di interconnessioni "non ovvie" nella letteratura scientifica vedi i *case histories* presentati in: M.M. Kessler, "Bibliographic coupling extended in time: ten case histories" *Information Storage & Retrieval* 1963, 1: 169-87.

¹⁸⁰ H.G. Small & B.C. Griffith, "The structure of scientific literature. I: Identifying and graphing specialties" *Science Studies* 1974, 4: 17-40, p. 20.

ma le coppie di citazioni condivise da due o più documenti (co-citazioni). L'analisi delle co-citazioni (*cluster* o *co-citation analysis*), sviluppata contemporaneamente (e indipendentemente) da Henry Small a Philadelphia e da Irina Marschakowa a Mosca, poggia su due ipotesi: 1) se due lavori sono citati da un terzo, esiste un vincolo cognitivo tra i due; 2) la forza di questo vincolo dipende dal numero di volte in cui i due documenti sono citati assieme nella letteratura specialistica.

Un articolo che cita in bibliografia due documenti effettua un qualche tipo di associazione tra di essi. Se altri articoli operano la stessa associazione citando contemporaneamente gli stessi documenti allora è ipotizzabile che la correlazione sia qualcosa in più di un caso fortuito. Concordando su ciò che costituisce la letteratura significativa precedente, gli scienziati definiscono in qualche modo la struttura disciplinare del loro campo d'indagine. La frequenza di co-citazioni misura perciò il grado di affinità (concettuale, metodologica) tra documenti così come è percepita dal gruppo degli autori citanti.¹⁸¹ Se i documenti molto citati equivalgono a concetti chiave, metodi o esperimenti tipici di una disciplina, allora le co-citazioni possono essere usate per rintracciare le relazioni tra questi documenti/concetti chiave e, nella misura in cui tali relazioni sono condivise da una comunità di studiosi, per definire e visualizzare graficamente la struttura di una disciplina o di una specialità, le sue connessioni con altri settori e l'articolazione di un'area d'indagine in sottoaree e nuovi fronti di ricerca.¹⁸²

¹⁸¹ H.G. Small, "Co-citation in the scientific literature: a new measure of the relationship between two documents" *Journal of the American Society for Information Science* 1973, 24: 265-69.

¹⁸² In una variante del metodo, gli autori co-citati sono conteggiati al posto dei documenti, sulla base dell'assunto che l'associazione bibliografica costante di due autori sia un indicatore affidabile del modo in cui gli altri membri della comunità scientifica percepiscono la similarità di contenuto tra i documenti da essi prodotti. Sul metodo della co-citazione degli autori: H.D. White & B.C. Griffith, "Author cocitation: A literature measure on intellectual structure" *Journal of the American Society for Information Science* 1981, 32: 163-71. Un'applicazione al campo della scienza dell'informazione si trova in: H.D. White & K.W. McCain, "Visualizing a discipline: An author co-citation analysis of information science, 1972-1995" *Journal of the American Society for Information Science* 1998, 49(4): 327-56. Due limiti di questa metodologia: 1) la base dati dell'ISI indicizza quasi sempre, per ogni citazione, solo il primo autore, che non sempre corrisponde al principale responsabile del contenuto intellettuale del documento; 2) è sempre possibile che un autore appartenga contemporaneamente a più di un'area specialistica.

Semplice dal punto di vista tecnico, l'analisi delle co-citazioni presenta invece notevoli difficoltà nell'interpretazione grafica dei risultati ottenuti. Davanti ad un fenomeno complesso, che coinvolge una molteplicità di variabili, servono infatti strumenti utili a ridurre il numero delle variabili necessarie a descriverlo, allo scopo di costruire un modello matematico plausibile e facile da visualizzare (quindi manipolare). Strumenti di questo tipo dipendono a loro volta dall'applicazione di tecniche matematiche, derivate per lo più dalla statistica multivariata, quali l'analisi dei cluster, l'analisi fattoriale, lo scaling multidimensionale. Si tratta di tecniche molto comuni nel data mining, cioè nei procedimenti utili a scoprire, in grandi volumi di dati, modelli e correlazioni latenti a vantaggio di manager e operatori della business intelligence. Nella rappresentazione di un fenomeno descritto da una molteplicità di variabili, i metodi di statistica multivariata consentono di ridurre il numero delle variabili in gioco, in modo da facilitare la classificazione dei dati sulla base delle loro affinità interne. Nel caso specifico, davanti alla popolazione delle coppie di documenti co-citati, l'analisi dei cluster traduce le differenze quantitative tra frequenze di co-citazione in gruppi (cluster) di documenti affini. A questo punto, se si vuole ottenere una mappa, serve un'ulteriore procedura per rappresentare, in uno spazio a due o tre dimensioni, lo schema di "prossimità" o "similarità" dei documenti clusterizzati. Nei primi esperimenti di mappatura della scienza, Henry Small e Belver Griffith utilizzarono lo scaling multidimensionale, una tecnica nata nel contesto della psicofisica e della psicomètria, il cui principio di funzionamento può essere illustrato ricorrendo ad un esempio geografico.¹⁸³ Data una mappa dell'Italia, sarebbe facile, conoscendo la riduzione di scala, tradurre le distanze chilometriche tra Milano, Bologna e Bari in una tabella che metta in relazione, a due a due, le tre città. Si consideri ora la situazione

¹⁸³ Un esempio analogo è sviluppato in: H. Kreuzman, "A co-citation analysis of representative authors in philosophy: Examining the relationship between epistemologists and philosophers of science" *Scientometrics* 2001, 51(3): 525-39.

inversa. Conosciamo le distanze tra le città, non c'è la mappa, quindi non sappiamo nulla della loro posizione relativa nello spazio. È ovvio che in questo secondo caso si può comunque ipotizzare una localizzazione spaziale dei punti che rispetti le misure delle distanze reciproche riportate nella tabella, ma tale operazione diventa progressivamente più complicata man mano che il numero delle città da collocare aumenta. Se poi le distanze in tabella sono calcolate tenendo conto della sfericità della terra, allora il grado di infedeltà della mappa bidimensionale aumenta in proporzione, come sanno bene i cartografi. Lo scaling multidimensionale fa proprio questo: parte da una tabella di valori numerici, che nel caso dei cluster di documenti co-citati elenca le frequenze di co-citazione, e traccia dei punti sulla mappa in modo da rappresentare il più fedelmente possibile i valori della tabella.

L'analogia geografica funziona ovviamente fino ad un certo punto. Nel mondo dell'informazione le distanze da misurare sono concettuali, non fisiche, ma i bibliometri accettano il postulato che le distanze concettuali siano esprimibili in termini di differenze quantitative tra legami bibliografici. Al termine dell'analisi, quindi, una classe formata da gruppi di documenti con alto tasso di co-citazione si traduce graficamente in una serie di punti contigui. Lo spazio che racchiude questi punti è un buon candidato a rappresentare una specialità scientifica o un fronte di ricerca in via di formazione.

L'esperimento compiuto da Small e Griffith su un campione di documenti estratti dall'edizione 1972 dello *Science Citation Index* dava risultati confortanti, mostrando che: 1) è relativamente facile, mediante lo scaling multidimensionale, identificare e rappresentare graficamente in uno spazio a due dimensioni aree specialistiche come la fisica delle particelle o la fisica nucleare, dove l'accordo tra studiosi sui documenti e concetti chiave della disciplina è molto forte; 2) esiste una corrispondenza di fondo, sul piano cognitivo, tra la distribuzione dei legami bibliografici e le parole chiave dei titoli degli articoli clusterizzati; 3) è più complicato, ma comunque possibile,

costruire una mappa soddisfacente delle specialità biomediche, dove alcuni documenti molto citati e co-citati, di solito lavori che illustrano tecniche e metodi largamente condivisi, tagliano trasversalmente diverse sottoaree.¹⁸⁴

La mappa può essere distillata, mediante un procedimento ricorsivo, applicando l'analisi delle co-citazioni non direttamente ai documenti ma ai cluster di documenti, poi ai cluster di cluster e così via, fino ad ottenere livelli di aggregazione sempre più alti, corrispondenti ai “super-documenti” di ogni disciplina.¹⁸⁵ Allineate in una serie temporale poi, la singole mappe di un dominio disciplinare diventano i fotogrammi di una sorta di proiezione cinematografica dei mutamenti nella struttura cognitiva del dominio. Il ruolo di attori principali del film spetta ancora ai documenti più citati, la cui sostituzione è interpretata come un indice di trasformazione intellettuale. L'ISI prese spunto dal lavoro dei due studiosi per sviluppare, negli anni '70 e '80, un programma di clustering annuale dello *Science Citation Index* che culminò nel progetto, poi abbandonato, di redigere un atlante della scienza mondiale. L'atlante era destinato a realizzare un sistema di recupero dell'informazione scientifica di tipo spaziale-deduttivo, sul modello dei sistemi geografici, potenziato da una trama ipertestuale fatta di nodi e link paragonabile all'architettura del Memex di Bush: «in tale sistema - scrive Small - l'utente prima osserva la mappa complessiva della scienza, seleziona un nodo e guarda la mappa della sottoarea, forse una disciplina scientifica. I nodi di questa sottoarea possono anche essere considerati mappe di sottodiscipline o specialità. L'utente può scorrere l'albero delle mappe nidificate fino a raggiungere il livello del documento, per il recupero di specifiche pubblicazioni [...] Chiaramente un tale sistema è qualcosa in più di un *retrieval system* bibliografico; è più simile ad un'enciclopedia della scienza

¹⁸⁴ H.G. Small & B.C. Griffith, “The structure of scientific literature. I: Identifying and graphing specialties” *Science Studies* 1974, 4: 17-40, pp. 35-40. L'argomento è ripreso e approfondito in: H.G. Small; B.C. Griffith; J.A. Stonehill et al., “The structure of scientific literature. II: Toward a macro- and microstructure for science” *Science Studies* 1974, 4: 339-65.

integrata con una struttura bibliografica, forse una realizzazione della combinazione *World Brain/Memex* prefigurata da Garfield». ¹⁸⁶ Si può anche chiedere di più ad una mappa del genere. Le citazioni non legano semplicemente i documenti, legano le idee e le argomentazioni contenute nei documenti. Queste connessioni sono trasversali, si dispiegano oltre i confini tra discipline e campi di ricerca ed in molti casi - sottolinea Small - la trasversalità, l'esistenza di concetti comuni è un valore positivo, incide direttamente sui criteri adoperati dagli scienziati per confermare la validità delle ipotesi. Se infatti un concetto originato in un campo risulta applicabile anche in un secondo campo, tale estendibilità funge da convalida esterna per quel concetto. Lo status di una scoperta nella biologia molecolare, per esempio, può essere rinforzato o indebolito da altre scoperte nel campo della fisica o della chimica e viceversa. Small si appoggia così alle tesi di Edward Wilson sull'armonia (*consilience*) nella scienza, ¹⁸⁷ nonché alla teoria della coerenza esplicativa (*explanatory coherence*) di Paul Thagard, ¹⁸⁸ per rivendicare la capacità delle mappe di co-citazioni di rivelare la struttura che tiene unite, al di là dei linguaggi e delle pratiche settoriali, le reti argomentative delle teorie scientifiche. ¹⁸⁹

Speculare sulle enormi potenzialità di una mappa globale della scienza non bastava comunque ad oscurare i problemi tecnici e concettuali dell'impresa. Il procedimento iniziale andava rettificato perché la mappa presentava diverse imperfezioni. Non era sempre possibile, a differenza di quanto sostenuto da Price, riprodurre in una cornice spaziale a due dimensioni le complesse relazioni tra documenti co-citati e specialità scientifiche: i cluster ben delineati

¹⁸⁵ H. Small & E. Garfield, "The geography of science: Disciplinary and national mapping" *Journal of the American Society for Information Science* 1985, 11(4): 147-59.

¹⁸⁶ H. Small; E. Sweeney; E. Grenlee, "Clustering the Science Citation Index using co-citations. 2. Mapping science" *Scientometrics* 1985, 8(5-6): 321-40, p. 339.

¹⁸⁷ E.O. Wilson, *L'armonia meravigliosa: dalla biologia alla religione, la nuova unità della conoscenza* (Milano: Mondadori, 2000).

¹⁸⁸ P. Thagard, *Rivoluzioni concettuali. Le teorie scientifiche alla prova dell'intelligenza artificiale* (Milano: Guerini e Associati, 1994).

¹⁸⁹ H. Small, "Citations and consilience in science" *Scientometrics* 1998, 43(1): 143-48.

di documenti affini coesistevano con macro-cluster dalla struttura amorfa, la rappresentazione delle diverse regioni disciplinari era nettamente sbilanciata nella direzione delle discipline biomediche, abbondavano documenti e cluster “isolati”. In certi casi, le incongruenze apparivano tali da far presagire una distanza infinita tra i punti: «è come se – ammettono Small e Griffith - avessimo disegnato una mappa geografica e scoperto che ogni cosa è pressoché in ordine, ma che Albany ha una distanza infinita da Wilmington, sebbene entrambe siano precisamente posizionate in rapporto sia a New York che a Filadelfia».¹⁹⁰

Le difficoltà dipendevano in gran parte dalla selezione del set iniziale di dati e dai criteri di identificazione delle reti di co-citazioni. Esistono diversi metodi per formare i cluster e la scelta di un metodo influenza in maniera decisiva i risultati. Ci si accorse, ad esempio, che nel determinare la soglia minima di co-citazioni di un documento candidato al *clustering*, era necessario introdurre dei coefficienti di normalizzazione per evitare che l'esistenza di documenti troppo citati e co-citati (tipicamente gli articoli metodologici in ambito biomedico) penalizzasse le aree di ricerca più piccole.¹⁹¹ Divenne subito chiaro inoltre che il peso delle citazioni varia a seconda delle discipline: i riferimenti bibliografici provenienti da aree di ricerca con un basso volume di pubblicazioni e con schemi di *referencing* orientati verso bibliografie brevi (la matematica o l'ingegneria) hanno verosimilmente un peso maggiore di quelli derivati da articoli di discipline in cui si pubblica molto e si cita in abbondanza (l'area biomedica).¹⁹² Furono introdotti allora sistemi più

¹⁹⁰ H.G. Small; B.C. Griffith; J.A. Stonehill et al., “The structure...” cit., p. 363.

¹⁹¹ La frequenza minima di co-citazione è stata così normalizzata: 1) in rapporto al numero di citazioni ricevute dai singoli documenti co-citati e alla media di tali citazioni (coefficienti di Jaccard e Tanimoto); 2) in rapporto al numero di citazioni ricevute dal documento meno citato all'interno della coppia; 3) in rapporto ad un valore che risulta dalla combinazione dei due valori precedenti (*CoCit-Score*); 4) sostituendo il conteggio delle co-citazioni con un coefficiente di correlazione (coefficiente di Pearson); 5) sostituendo il conteggio delle co-citazioni con l'analisi fattoriale basata su una matrice di correlazione. Per una valutazione comparativa di tali metodi: M. Gmür, “Co-citation analysis and the search for invisible colleges: A methodological evaluation” *Scientometrics* 2003, 57(1): 27-57.

¹⁹² H. Small & E. Sweeney, “Clustering the Science Citation Index using co-citations. 1. A comparison of methods” *Scientometrics* 1985, 7(3-6): 391-409, pp. 391-393.

complessi di elaborazione dei file annuali dello *Science Citation Index*, basati su stratagemmi quali il conteggio frazionario delle citazioni (il valore di una citazione è frazionato in base al numero di riferimenti bibliografici contenuti nell'articolo citante) e i livelli variabili di co-citazione (il livello di co-citazione necessario alla creazione di un cluster varia durante il processo analitico in accordo ad una regola prestabilita).

A partire dalla fine degli anni '80, le critiche indirizzate alle mappe di co-citazioni, sia sul versante metodologico che su quello dei possibili usi strategici, ne hanno rallentato il cammino.¹⁹³ Le attenzioni dei bibliometri si sono concentrate allora soprattutto sui metodi di *co-word analysis*, per almeno due motivi: le parole si possono estrarre da qualunque tipo documenti, non solo gli articoli di riviste dell'ISI, e tendono ad avere, nelle consuetudini linguistiche di una comunità specializzata di studiosi, un uso più prevedibile e costante nel tempo rispetto alle citazioni bibliografiche.¹⁹⁴

La manipolazione delle citazioni bibliografiche non poteva nascondere il fatto che, sul piano teorico, le mappe bibliometriche erano solo una vaga approssimazione della realtà,¹⁹⁵ e che non bastava sovrapporre le cartografie annuali derivate dallo *Science Citation Index* per far emergere un filo logico unitario nella sequenza di istantanee scattate al paesaggio della scienza. I fronti di ricerca sembravano variare da un anno all'altro, nei cluster entravano di continuo documenti nuovi mentre altri ne uscivano bruscamente e la parziale continuità di alcuni elementi su larga scala, come la posizione relativa

¹⁹³ D.O. Edge, "Why I am not a cocitationist" *Society for the Social Studies of Science Newsletter* 1977, 2: 13-19; D. Hicks, "Limitations of co-citation analysis as a tool for science policy" *Social Studies of Science* 1987, 17: 295-316.

¹⁹⁴ Sulle difficoltà di questo approccio causate dalla fluidità del linguaggio scientifico vedi: L. Leydesdorff, "Why words and co-words cannot map the development of the sciences" *Journal of the American Society for Information Science* 1997, 48(5): 418-27.

¹⁹⁵ Sui limiti del concetto bibliometrico di paradigma: «It is clear that any use of citation data can only be an approximation to Kuhn's original multifaceted notion of paradigm. Quite possibly, however, highly cited documents are as close to his exemplars ("exemplary past achievements") as information science can get. He had stated that exemplars would include "... at least some of the technical problem-solutions found in the periodical literature ...". The best that we can expect from bibliometric models is a partial and imperfect reflection, as Griffith (1979) would say, a "faulty mirror" of science»: H.G. Small, "Paradigms, citations and

delle discipline, era contrastata ai livelli più bassi da una notevole fluidità della base documentaria.

Malgrado i limiti evidenziati, l'analisi delle co-citazioni è stata, dalla metà degli anni '70, una metodologia molto diffusa per studiare la struttura socio-cognitiva della comunicazione scientifica ed alcuni studi documentano l'esistenza di un certo grado di corrispondenza tra i risultati dell'analisi co-citazionale e la percezione della struttura di una disciplina scientifica da parte dei suoi membri.¹⁹⁶ In anni recenti, tecniche di mappatura più raffinate hanno affiancato lo scaling multidimensionale nella visualizzazione delle strutture latenti dei dati bibliografici,¹⁹⁷ mentre il perfezionamento degli ipertesti e di interfacce utente di tipo grafico, quali HTML e applets Java, ha reso le mappe interattive e tridimensionali, quindi più facili da navigare e perlustrare, adattandole al recupero di informazioni specifiche. Ne è derivato un nuovo impulso allo studio delle tecniche di visualizzazione dei domini conoscitivi (*knowledge domain visualization*) soprattutto in applicazioni destinate all'*information retrieval*.¹⁹⁸ All'opposto, il vecchio sogno di Price di fabbricare *war maps* del paesaggio scientifico a supporto delle strategie politiche dei burocrati non si è realizzato. I politici, al pari degli scienziati, non hanno preso

maps of science: a personal history" *Journal of the American Society for Information Science and Technology* 2003, 54(5): 394-99, p. 396.

¹⁹⁶ Ad esempio: K.W. McCain, "Cocited author mapping as a valid representation of intellectual structure" *Journal of the American Society for Information Science* 1986, 37: 111-22.

¹⁹⁷ Per una rassegna delle tecniche attuali di mappatura e visualizzazione vedi: K. Börner; C. Chen; K.W. Boyack, "Visualizing knowledge domains" *Annual Review of Information Science and Technology* 2003, 37: 179-255. Esempi di tali tecniche sono: la triangolazione geometrica, le *pathfinder networks*, le mappe basate su reti neurali artificiali (mappe di Kohonen (*self organizing maps*), l'*annealing* simulato (*simulated annealing*) e l'analisi semantica latente (*latent semantic analysis*). Il metodo della triangolazione è stato usato da Small in congiunzione con il software SCI-Map per costruire mappe computerizzate: H.G. Small, "Visualizing science by citation mapping" *Journal of the American Society for Information Science* 1999, 50: 799-813. Un'applicazione della tecnica di *pathfinder networks* si trova in: H.D. White, "Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists" *Journal of the American Society for Information Science and Technology* 2003, 54(5): 423-34. Sull'*annealing* simulato vedi: K.W. Boyack; B.N. Wylie; G.S. Dividson, "Domain visualization using VxInsight for science and technology management" *Journal of the American Society for Information Science and Technology* 2002, 53(9): 764-74. Un'applicazione dell'analisi semantica latente si trova in: T.K. Landauer; D. Laham; M. Der, "From paragraph to graph: Latent semantic analysis for information visualization" *Proceedings of the National Academy of Sciences* 2004, 101(suppl 1): 5214-19. Un'applicazione scientometrica delle reti neurali artificiali, in particolare delle mappe di Kohonen si trova in: X. Polanco; C. Francois; J.C. Lamirel, "Using artificial neural networks for mapping of science and technology" *Scientometrics* 2001, 51(1): 267-92.

in grande considerazione le mappe bibliometriche giudicandole spesso astratte e di scarsa utilità pratica. La loro indifferenza ha indotto allora i cartografi a dare maggiore risalto ad alcuni elementi del paesaggio. Dal momento che il valore strategico di una mappa intellettuale si misura, in ultima analisi, dalla capacità di posizionare nello spazio cognitivo gli attori (università, istituti di ricerca, aziende), di quantificarne l'influenza rispetto al campo disciplinare e di monitorare le variazioni nel tempo delle loro performance, il passo successivo è stato quello di integrare le procedure descrittive di mappatura con quelle valutative di misurazione della qualità del lavoro scientifico.¹⁹⁹

¹⁹⁸ Vedi ad esempio: X. Lin, "Map displays for information retrieval" *Journal of the American Society for Information Science* 1997, 48(1): 40-54.

¹⁹⁹ «If we are to provide an instrument for monitoring and evaluating scientific research, within the science policy context, we should give much attention to the actors. They constitute the only element [...] that may be affected directly by policy. Every strategy or decision will apply to the actors, not to language, nor to the archive. This means that any mapping study in a science policy context should integrate information or linkages to actors, at any desired level of aggregation»: E. Noyons, "Science maps within a science policy context," in *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems*, edited by H.F. Moed, et al. (Dordrecht: Kluwer Academic, 2004): 237-55, p. 245. Sullo stesso argomento: E. Noyons; M. Luwel; H.F. Moed, "Combining map and citation analysis for evaluative bibliometric purpose. A bibliometric study on recent development in micro-electronics" *Journal of the American Society for Information Science and Technology* 1999, 50(2): 115-31; E. Noyons, *Bibliometric mapping as a science policy and research management tool*. Ph.D. Thesis (Leiden: Universiteit Leiden. Faculteit der Sociale Wetenschappen, 1999) <http://www.cwts.nl/cwts/Noyons_1999.pdf>.

Capitolo 5

Impact Factor e valutazione degli scienziati: la citazione bibliografica al servizio della politica della scienza

La scienza è fondata sul riconoscimento e sulla legittimazione dei risultati del singolo da parte di una comunità di individui che condividono un insieme di nozioni, pratiche, valori relativi al campo di indagine sul quale si esercita la loro competenza specifica. Uno scienziato è tale solo all'interno di una comunità di colleghi orientati alla risoluzione di una classe condivisa di problemi. In più, egli è uno scienziato bravo solo se qualcuno che ha l'autorità e i titoli per farlo gli dice che è bravo: nonostante la casistica pur significativa dei geni incompresi, il suo non è mai un dialogo esclusivo con la natura, ma un dibattito a più voci nel quale il parere dei colleghi già affermati è importante quanto e forse più della corretta applicazione di un metodo nella conduzione di un esperimento o nella risoluzione di un set di equazioni differenziali.

Il riconoscimento della qualità scientifica, che in sostanza equivale al riconoscimento dell'originalità di un contributo, funge sia da forza motivante per il ricercatore - tanto da indurre chi non lo ottiene a cambiare mestiere dirottando sull'insegnamento o sulla carriera amministrativa - sia, sul piano politico, da guida nell'allocazione delle risorse finanziarie e umane necessarie a far funzionare la macchina della scienza. Tale riconoscimento si esercita principalmente attraverso la valutazione della qualità delle pubblicazioni, la cui funzione cognitiva di comunicare nuove conoscenze è inscindibile da quella sociale di reclamare visibilità e priorità (sulle idee comunicate) per l'autore.

Il metodo tradizionale per la verifica della qualità della letteratura scientifica, nonché della validità dei progetti di ricerca candidati a ricevere dei finanziamenti, è la valutazione da parte di un gruppo di specialisti riconosciuti come esperti nella materia della pubblicazione o del progetto (*peer review*). Il sistema di *peer reviewing*, nato nell'Europa del XVII secolo sotto la spinta della specializzazione disciplinare, ha conosciuto una diffusione universale solo nel secondo dopoguerra, dopo una storia dai contorni irregolari, frammentaria, in gran parte ancora tutta da scrivere.²⁰⁰ Tale sistema ha rappresentato sinora una componente essenziale del carattere collettivo, cumulativo e auto-correttivo dell'impresa scientifica. Attraverso il giudizio competente dei pari infatti «molti scienziati possono costruire sul lavoro di altri con un certo grado di sicurezza».²⁰¹ L'attendibilità della *peer review* presuppone ovviamente che la procedura di verifica sia codificata e impersonale, che esista una convergenza spontanea e quasi necessaria tra i giudizi valutativi di diversi esperti su una determinata opera o progetto di ricerca e che tali giudizi, in ultima analisi, siano l'esito prevedibile di una missione durante la quale i giudici si disincarnano come uomini per ricomporsi come puri spiriti razionali giudicanti alla luce delle regole del metodo scientifico. Da diverso tempo nondimeno si è fatta strada la consapevolezza, tra gli stessi scienziati, che questa distillazione spirituale non avviene: il sistema dei *referees* è viziato da distorsioni, pregiudizi, conflitti di interesse, errori di valutazione ampiamente documentati nella letteratura critica.²⁰² I sociologi si sono affrettati a spiegare come ciò sia possibile alla luce delle dinamiche sociali spesso conflittuali che governano i campi scientifici, i filosofi hanno fatto altrettanto smontando il mito del metodo

²⁰⁰ Sulla storia del *peer reviewing*: J.C. Burnham, "The evolution of editorial peer review" *JAMA* 1990, 263(10): 1323-32; D.A. Kronick, "Peer review in 18th-century scientific journalism" *JAMA* 1990, 263(10): 1321-2.

²⁰¹ R.K. Merton & H. Zuckerman, "Modelli istituzionalizzati di valutazione nella scienza (1971)," in *La sociologia della scienza. Indagini teoriche ed empiriche*, edited by N.W. Storer (Milano: Franco Angeli, 1981): 569-609, p. 604.

unico dalla cui applicazione univoca deriverebbero le regole di condotta e valutazione dei prodotti della ricerca scientifica.

Tra le critiche, una in particolare sembra fungere da cornice intellettuale per tutte le altre e si segnala per la profondità con cui destabilizza la fiducia nella capacità del sistema di tutelare il cammino progressivo della scienza. Secondo Thomas Kuhn, gli scienziati che operano nel quadro di un paradigma scientifico consolidato mantengono un atteggiamento rigido e conservatore verso il corpo delle teorie scientifiche correnti e tendono a scoraggiare idee e ipotesi alternative. Questa resistenza al cambiamento, che si manifesta anche attraverso il controllo dei canali di diffusione della letteratura scientifica, è una tappa essenziale del percorso irregolare della scienza: ogniqualvolta la portata e la precisione delle conoscenze paradigmatiche diventano tali da far emergere difficoltà tecniche, anomalie comprensibili solo alla luce di teorie paradigmatiche nuove e radicalmente differenti, il terreno è pronto per una rivoluzione scientifica. Negli intervalli tra una rivoluzione e l'altra invece l'attività di ricerca è scienza normale, pura routine, soluzione di rompicapo (*puzzle solving*). «La scienza normale non ha per scopo quello di trovare novità di fatto o teoriche e, quando ha successo, non ne trova nessuna».²⁰³ Una *peer review* condotta in un periodo di scienza normale tende di fatto, secondo questa prospettiva, ad oscurare l'introduzione di novità teoriche significative formando una sorta di «legge di ferro dell'oligarchia disciplinare».²⁰⁴

La crisi della *peer review* ha avuto, tra i suoi effetti, la ricerca di metodi di valutazione alternativi, il più possibile sganciati da considerazioni soggettive e giudizi contingenti di merito. È in questo contesto che l'impatto delle pubblicazioni in termini di numero di citazioni accumulate ha cominciato a ricevere un'attenzione inedita da politici e manager della scienza, condizionati

²⁰² Per una panoramica nelle scienze biomediche vedi: I. Godlee & T. Jefferson (Ed.), *Peer Review in Health Sciences* (London: BMJ Books, 2003).

²⁰³ T. Kuhn, *La struttura delle rivoluzioni scientifiche* (Torino: Einaudi, 1978), p. 75.

²⁰⁴ T.J. Scheff, *Peer review: an iron law of disciplines?*, 2002
<<http://www.soc.ucsb.edu/faculty/scheff/23.html>>.

evidentemente dal fatto che un indice numerico è comodo, di immediata comprensione e di facile applicazione, ed esistono già i repertori dal quale attingerlo o ricavarlo con poco sforzo, cioè i *citation indexes* dell'ISI e prodotti derivati.

Usi e abusi degli indici citazionali nella valutazione degli scienziati hanno preso due strade: la prima è una scorciatoia e consiste nell'associare ad un articolo (e al suo autore) lo stesso numero che esprime, in termini statistici, l'impatto della rivista su cui è pubblicato; la seconda è una via più complessa, in gran parte ancora da costruire o perfezionare, che, partendo dalle banche dati dell'ISI, si concentra sull'analisi delle citazioni al livello del singolo autore, del gruppo, dell'istituzione di ricerca, e integra questa indagine con una serie di considerazioni di metodo tese a limitare o a precisare la portata e le condizioni di applicabilità dei calcoli effettuati.

5.1. La scorciatoia: l'Impact Factor delle riviste

L'*Impact Factor (IF)* delle riviste, ideato nei primi anni '60 da Eugene Garfield e Irving H. Sher per facilitare la selezione dei periodici da includere nello *Science Citation Index*, è contenuto in una pubblicazione annuale dell'ISI, il *Journal Citation Reports (JCR)*, nelle due edizioni per le scienze naturali e per le scienze sociali. Filtrando i dati dello *Science Citation Index* e del *Social Sciences Citation Index*, il *JCR* classifica annualmente le riviste sulla base di diversi parametri, il principale dei quali, o perlomeno quello più discusso, è proprio l'*IF*.

L'*IF* è una misura della frequenza con cui l'articolo medio («the average article») di una rivista è citato, in un dato anno, dagli articoli pubblicati nei due anni precedenti sul gruppo di riviste selezionate dall'ISI per l'indicizzazione. Il numero che esprime l'*IF* è il risultato di un rapporto tra citazioni ricevute e articoli recenti citabili. Questo rapporto è calcolato mediante una semplice divisione aritmetica tra un numeratore e un denominatore: il numeratore è il numero delle citazioni ricevute dagli articoli usciti sulla rivista nei due anni precedenti a quello per il quale si effettua il calcolo. Il denominatore è il numero complessivo di articoli pubblicati lungo lo stesso periodo di tempo sulla stessa rivista. Se nel 2004 gli articoli usciti nel 2002 e nel 2003 sulla rivista **X** sono stati citati rispettivamente 100 e 150 volte, e se il totale degli articoli pubblicati su quella rivista nel 2002 e nel 2003 è 70, allora l'*IF* della rivista **X** è dato da:

$$\mathbf{IF(X)} = \frac{\mathbf{100 + 150}}{\mathbf{70}} = \mathbf{3,57}$$

Alcune funzioni avanzate del *JCR* permettono di affinare questo coefficiente, per esempio di visualizzare e manipolare la percentuale delle autocitazioni, di ricavare un *IF* unificato per le riviste che, nel corso degli anni, hanno subito cambiamenti di titolo, scissioni, fusioni, oppure di costruire, per una data rivista, un grafico di tendenza risultante da una proiezione dell'indice su cinque anni.

Altri parametri di classificazione, correlati con il precedente, sono: 1) l'indice di immediatezza (*immediacy index*), calcolato dividendo il numero di citazioni che la rivista riceve in un anno per il numero di articoli pubblicati in quello stesso anno, interpretabile come una stima della velocità con cui l'articolo medio della rivista viene citato;²⁰⁵ 2) la semivita delle citazioni (*cited half-life*), cioè il numero di anni di pubblicazione della rivista che rendono conto del 50% delle citazioni ricevute nell'anno corrente, interpretabile come una stima dell'età media degli articoli citati. Entrambi i parametri non hanno di per sé un significato qualitativo e dei banali cambiamenti di formato o strategia editoriale possono alterarli in modo decisivo. Una rivista di rapida comunicazione, per esempio un *letters journal*, viene citata più velocemente rispetto ad un *full paper journal*, ma per un periodo inferiore di tempo, avrà pertanto un alto indice d'immediatezza ed una bassa semivita. In corrispondenza, l'*IF* della prima sarà più alto dal momento che viene calcolato solo sugli ultimi due anni di vita del periodico.

L'*IF* è presentato dall'ISI come uno strumento «per valutare l'importanza relativa di una rivista, specialmente se comparata alle altre riviste della stessa area».²⁰⁶ Sotto tale veste, esso ha rappresentato per molte biblioteche un supporto alle strategie di selezione e scarto di materiale periodico.²⁰⁷ Il

²⁰⁵ Ad esempio, l'*Immediacy Index* per una rivista X che nel 2004 ha pubblicato 50 articoli ed è stata citata 10 volte è dato da: $10/50 = 0,2$.

²⁰⁶ Thomson ISI, *ISI Journal Citation Reports*, 2005

<<http://isi17.isiknowledge.com/portal.cgi?DestApp=JCR&Func=Frame>>.

²⁰⁷ Una rassegna selettiva di studi sull'argomento si trova in: K.G. Altmann & G.E. Gorman, "Usage, citation analysis and costs as indicators for journal deselection and cancellation: a selective literature review" *Australian Library Review* 1996, 13(4): 379-92.

problema è nato quando, nel contesto di una generale revisione dei metodi tradizionali di controllo della qualità scientifica, ha cominciato ad essere usato in maniera estensiva da amministratori e politici per valutare il lavoro di singoli scienziati e istituzioni: più alto l'*IF* delle riviste in cui compaiono le pubblicazioni elencate nei curricula, maggiori le possibilità di ottenere una promozione, una cattedra, un finanziamento. In Europa soprattutto è invalsa l'abitudine di associare all'impatto, quindi ai meriti di un autore, ciò che originariamente designava solo l'impatto di un periodico. La Finlandia ha canonizzato sul piano legislativo la dipendenza dei finanziamenti accademici dall'*IF* delle riviste. Le università tedesche si appoggiano regolarmente al *JCR* per lo stesso scopo, mentre in Italia l'uso spregiudicato dell'*IF* nelle procedure di valutazione è una prerogativa del settore biomedico, dove paradossalmente il termine e il concetto di "impatto" hanno un significato profondo sulle condizioni materiali di vita dei cittadini.²⁰⁸

Editori e direttori di riviste, dal canto loro, hanno intuito l'utilità commerciale dell'indice di Garfield, sicché capita spesso di leggere editoriali dai titoli trionfalistici espressi nella forma: *Journal X receives Impact Factor*, *Journal X Impact Factor sets another record*, *Journal X Impact Factor increases by n%*. Non diversamente dall'audience televisivo, stimato dalla quantità di persone che si calcola siano raggiunte da un messaggio o da un programma, a prescindere dal gradimento effettivo ad essi riservato, l'acquisizione o l'aumento percentuale del punteggio nella classifica del *JCR* è diventato il preludio al successo editoriale, ad una maggiore visibilità commerciale e, in

²⁰⁸ L'*IF* è stato usato dal Ministero della Salute per assegnare il 40% dei finanziamenti alla ricerca agli Istituti di Ricovero e Cura per il triennio 2002-2004: Italia. Ministero della Salute. Direzione Generale Ricerca Sanitaria e Vigilanza Enti, *Circolare. Oggetto: Ricerca corrente 2002, 2003, 2004 - Acquisizione elementi ai fini della ripartizione*, 2002 <http://www.ministerosalute.it/resources/static/legis2002/Circolare_RC.pdf>. Telethon usa l'*IF* per valutare l'output delle ricerche finanziate ed ha chiesto a Thomson ISI di identificare le citazioni dei singoli articoli patrocinati: Comitato Telethon Fondazione Onlus, *Bilancio di missione al 30 giugno 2004*, 2004 <<http://www.telethon.it/telethon/allegati%5Cmissione2004.pdf>>.

taluni casi, ad una più lucrosa vendita degli spazi pubblicitari, con il conseguente aumento del costo delle riviste per l'utente finale.²⁰⁹

Contro questo metodo approssimativo di valutazione si è levato, negli ultimi anni, un coro unanime di voci. Recensirle tutte è impossibile, forse inutile per la ripetitività di molti argomenti. Si possono invece sintetizzare i principali motivi di contestazione con la necessaria avvertenza che, come normalmente accade quando la posta in palio va oltre la sterile contesa metodologica, i critici hanno spesso creato o rimodellato di sana pianta le argomentazioni dell'avversario per poterle smontare meglio alla luce delle proprie.²¹⁰

1) L'*impact factor* è una misura derivata da un semplice conteggio di citazioni, a prescindere dal contenuto e dal tenore delle stesse, dal contesto in cui sono state prodotte, quindi dai motivi per i quali un testo è stato citato, e da eventuali usi impropri, come nel caso delle citazioni volutamente ignorate per motivi personali o politici, delle citazioni rubate o aggiunte come condimento finale al testo, delle autocitazioni, delle citazioni fraudolente o sbagliate. Come in tutte le approssimazioni statistiche, è difficile dire quale sia l'incidenza, in una disciplina, in una rivista o nel singolo articolo, delle citazioni che non hanno un significato cognitivo esplicito e che pertanto non implicano un'apprezzamento qualitativo dell'autore citato da parte dell'autore citante (questo punto sarà discusso in maniera più approfondita nel capitolo 7).

2) La misura è effettuata su un campione di riviste selezionate da un ente commerciale, Thomson ISI, secondo criteri di qualità metodologicamente

²⁰⁹ Le analogie con il sistema dell'audience sono notevoli e si possono riscontrare, ad esempio, leggendo: A. Grasso & M. Scaglioni, *Che cos'è la televisione* (Milano: Garzanti, 2004), parte terza, pp. 393-477.

²¹⁰ Per le critiche all'*IF* vedi, ad esempio: P.O. Seglen, "Citation rates and journal impact factors are not suitable for evaluation of research" *Acta Orthopaedica Scandinavica* 1998, 69(3): 224-9; F. Hecht; B.K. Hecht; A.A. Sandberg, "The journal "impact factor": a misnamed, misleading, misused measure" *Cancer Genetics and Cytogenetics* 1998, 104: 77-81; D. Adam, "The counting house" *Nature* 2002, 415(6873): 726-9.

discutibili che non tengono in adeguata considerazione la dinamica effettiva della comunicazione scientifica e il giudizio degli esperti nei diversi settori disciplinari, riservando inoltre scarsa attenzione alla letteratura pubblicata in riviste non statunitensi e in lingue diverse dall'inglese. L'ISI peraltro non indicizza i riferimenti bibliografici delle monografie che, oltre a rappresentare l'esito più rilevante della ricerca in diversi settori delle scienze umane e sociali, nelle stesse scienze esatte rivestono spesso il ruolo di sintesi autorevoli delle conoscenze acquisite e delle teorie ritenute valide.

3) Il risultato della misura è un punteggio associato ad un oggetto inesistente, l'"articolo medio". In concreto, su una stessa rivista, esistono articoli buoni e articoli meno buoni, articoli con forte impatto e articoli con impatto nullo o quasi nullo. Calcolando la frequenza di citazione dei singoli articoli di una rivista, si osserva infatti che ognuno contribuisce in maniera diversa dagli altri al coefficiente d'impatto. La statistica al livello della riviste oscura tali differenze individuali.²¹¹ L'ascesa dell'*IF* nel cielo della politica ha generato peraltro un circolo vizioso: a) i ricercatori ambiscono a pubblicare nelle riviste a forte impatto, quindi b) le riviste a forte impatto possono selezionare i lavori migliori esercitando un drastico controllo di qualità su un numero maggiore di aspiranti autori, quindi c) i lavori usciti sulle riviste a forte impatto, di qualità mediamente elevata, sono più citati di quelli pubblicati su riviste a impatto medio-basso.

4) L'*IF* è condizionato da fattori che non hanno nulla a che vedere con la qualità dei contenuti. Un primo gruppo di fattori di disturbo è costituito dai limiti tecnici che dipendono dall'inaffidabilità della fonte, dai rischi dell'indicizzazione automatica e dalla necessità di contenere la spesa legata

²¹¹ Su questo punto ha insistito in particolare: P.O. Seglen, "Why the impact factor of journals should not be used for evaluating research" *BMJ* 1997, 314(7079): 498-502; P.O. Seglen, "The skewness of science" *Journal of the American Society for Information Science* 1999, 43(9): 628-38.

alle procedure manuali di controllo e correzione, per esempio: a) errori nell'attribuzione della paternità intellettuale dei lavori causati da inesattezze nei riferimenti bibliografici, da refusi tipografici, da ambiguità dovute ad omonimie o all'abitudine di molti autori di usare denominazioni diverse per gli stessi istituti; b) errori nelle procedure manuali di identificazione dei *citable items* di una rivista; c) errori nel conteggio delle citazioni di articoli firmati da consorzi (clamoroso il caso del basso numero di citazioni rilevato per l'articolo del International Human Genome Sequencing Consortium che nel 2001 annunciava la divulgazione di una *draft sequence* del genoma umano).²¹² Un secondo gruppo di difficoltà nasce dalle cattive abitudini degli editori, come quella di gonfiare l'indice delle proprie riviste selezionando contributi che citano preferibilmente articoli pubblicati sulle stesse riviste o articoli ben inseriti in una rete di citazioni reciproche. Un terzo limite è di natura metodologica più generale e rivela la dipendenza del calcolo dell'*IF* da fattori quali: a) il numero totale di articoli pubblicati annualmente da una rivista; b) la densità citazionale, cioè il numero medio di riferimenti bibliografici per articolo; c) l'argomento, l'area di ricerca in cui gli articoli si collocano, da cui dipende la semivita delle citazioni (*cited half-life*); d) la tipologia di articoli. Vediamo in che modo ciò accade. a) Il numero di articoli pubblicati annualmente da una rivista incide direttamente sul denominatore del rapporto di *IF* e sulle variazioni annuali dell'indice. b) La densità citazionale incide invece sul numeratore del rapporto e contribuisce a spiegare la grande variabilità di impatto tra le riviste di discipline diverse: negli articoli di matematica, ad esempio, si trova in media un numero di citazioni nettamente più basso rispetto agli articoli di biologia. Analogamente c) l'area tematica e la semivita delle citazioni condizionano il risultato del calcolo: in campo biomedico, ad esempio, le riviste che si occupano di ricerca fondamentale, come quelle di genetica molecolare, hanno un coefficiente d'impatto

²¹² "Errors in citation statistics" *Nature* 2002, 415: 101. L'articolo dello International Human Genome

nettamente superiore rispetto alle riviste di medicina clinica, sebbene queste ultime siano quotidianamente molto usate per potenziare la diagnosi e la terapia. Una delle cause del divario è il fatto che gli articoli pubblicati sulle prime sono "consumati" (e citati) più velocemente rispetto a quelli pubblicati sulle seconde (il che non implica perdita di validità scientifica per ciò che non viene più citato), quindi un numero maggiore di citazioni tende a concentrarsi sugli anni più recenti, in particolare sugli ultimi due che contano per il calcolo dell'*IF*. Anche d) la tipologia delle pubblicazioni ha un peso: le riviste di impostazione generalista o con molti articoli di contenuto metodologico sono generalmente più citate di quelle ultraspecializzate e con articoli di taglio sperimentale. La stessa sorte fortunata tocca alle riviste orientate in prevalenza verso la pubblicazione di articoli di sintesi (rassegne, *reviews*). Pur non apportando alcun contributo originale, tali riviste tendono ad essere molto citate perché i ricercatori usano le rassegne, soprattutto quelle scritte da colleghi prestigiosi, come surrogato della letteratura precedente.

5) Il numeratore e il denominatore della formula di *IF* sono incongruenti perché nel denominatore non sono contate delle unità che invece contribuiscono ad aumentare il valore del numeratore. L'ISI infatti non tiene conto, a livello di conteggio degli articoli pubblicati (denominatore), dei contributi che non siano classificati come saggi originali (*research articles*) o rassegne (*reviews*), ad esempio lettere, editoriali, abstract di convegni. Questi materiali però possono essere citati ed entrano di fatto nel conteggio globale delle citazioni (numeratore), alterando in modo significativo l'*IF* per le riviste che in un anno pubblicano molti contributi minori.

Molte di queste critiche erano state largamente previste da Garfield e dai suoi collaboratori, i quali hanno più volte richiamato l'attenzione sul fatto che l'uso

dell'*IF* per valutare gli scienziati anziché le riviste è improprio.²¹³ La pubblicazione su una rivista con alto *IF* porta sicuramente un documento all'attenzione di un'audience più ampia, ma l'impatto del documento, di cui il numero delle citazioni ricevute è solo un indice parziale, dipende dalla sua qualità intrinseca, non da quella della rivista, e l'accertamento della qualità è inscindibile dal giudizio dei pari. I dati citazionali «non sono in alcun modo pensati per sostituire la *peer review*»,²¹⁴ e per poter essere correttamente interpretati vanno normalizzati in modo da tenere in considerazione, oltre alla semivita delle citazioni (*cited half-life*) e alla densità citazionale, «variabili come il campo, o disciplina, e le pratiche citazionali».²¹⁵

5.2. La via lunga: costruzione e applicazione degli indicatori scientometrici

La scienza è un'attività complessa e multidimensionale, irriducibile ad una scala di valutazione fondata su un solo parametro o una sola tipologia di pubblicazioni. Un ricercatore universitario non spende tutto il tempo a scrivere: insegna, progetta, sperimenta, amministra, divulga. Quando scrive inoltre non scrive solo articoli di riviste, ma anche libri, rapporti e manuali tecnici, relazioni di convegni, software. Gli indicatori scientometrici di ultima generazione tendono pertanto a rispecchiare la molteplicità di output

²¹³ Ad esempio in: E. Garfield, "The impact factor and using it correctly" *Der Unfallchirurg* 1998, 48(2): 413 <[http://www.garfield.library.upenn.edu/papers/derunfallchirurg_v101\(6\)p413y1998english.html](http://www.garfield.library.upenn.edu/papers/derunfallchirurg_v101(6)p413y1998english.html)> e E. Garfield, "Journal impact factor: a brief review" *Canadian Medical Association Journal* 1999, 161(8): 979-80 <<http://www.garfield.library.upenn.edu/papers/journalimpactCMAJ1999.pdf>>.

²¹⁴ «You should not depend solely on citation data in your journal evaluations. Citation data are not meant to replace informed peer review. Additionally, careful attention should be paid to the many conditions that can influence citation rates such as language, journal history and format, publication schedule, and subject specialty»: Thomson ISI, *Using Journal Citation Reports® Wisely*, 2005 <<http://isi17.isiknowledge.com/portal.cgi?DestApp=JCR&Func=Frame>>. Ancora più esplicito è E. Garfield, "Citation impact depends upon the paper, not the journal! Don't count on 'citation by association'!" in *Essays of an Information Scientist 1962-1973*, 1 (Philadelphia: ISI Press, 1977): 452-53.

dell'attività di ricerca preservando al tempo stesso la possibilità di confrontare le singole unità a livello locale, nazionale e internazionale. La conciliazione di queste due opposte esigenze, multidimensionalità e comparabilità, passa necessariamente attraverso modelli probabilistici di analisi e semplificazione delle situazioni reali.²¹⁶ D'altro lato, data la centralità dell'analisi delle citazioni per la scientometria e la carenza, al di fuori dell'ISI, di database contenenti una quantità di informazioni strutturate sufficienti a coprire le diverse tipologie di output e la totalità della documentazione prodotta dalle singole istituzioni, si tende ad accettare l'ipotesi che i dati citazionali dell'ISI siano un campione rappresentativo del complesso dei risultati scientifici conseguiti da un singolo, da un gruppo o da un'istituzione. Per questo si usano largamente le citazioni bibliografiche nella costruzione degli indicatori di performance.²¹⁷ Un indicatore, non bisogna dimenticarlo, è qualcosa di fondamentalmente diverso dai dati che si usano per costruirlo: è un oggetto "derivato" da un'operazione matematica (il più delle volte aritmetica) sui dati e serve a misurare un aspetto o una dimensione di un fenomeno originariamente complesso e multidimensionale di cui non si conoscono in anticipo le leggi di sviluppo.²¹⁸ La misura è effettuata in relazione ad un sistema di riferimento convenzionale e colui che la effettua ha in mente un preciso significato da attribuirgli. Per esempio: il numero di citazioni ricevute da una pubblicazione lungo un certo lasso di tempo è un dato; il rapporto tra il numero di citazioni

²¹⁵ E. Garfield, "Journal impact factor: a brief review" *Canadian Medical Association Journal* 1999, 161(8): 979-80 <<http://www.garfield.library.upenn.edu/papers/journalimpactCMAJ1999.pdf>>, p. 979.

²¹⁶ Due esempi di tali modelli sono il DEA (*Data Envelopment Analysis*) e il POSCOR (*Partial Order SCORing*). Quest'ultima è una procedura per assegnare un ordine ad una serie di oggetti quando più di una variabile (le diverse misure di *performance* scientifica) è utilizzata simultaneamente come criterio di ordinamento. Vedi: P.S. Nagpaul & S. Roy, "Constructing a multi-objective measure of research performance" *Scientometrics* 2003, 56(3): 383-402. Per un approccio econometrico alle misure di multi-output vedi: A. Bonaccorsi & C. Daraio, "Econometric approaches to the analysis of productivity of R&D systems," in *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems*, edited by H.F. Moed, et al. (Dordrecht: Kluwer Academic, 2004): 51-74.

²¹⁷ Questa ipotesi, come prevedibile, si è rivelata parzialmente fondata per le scienze naturali, in particolare per quelle biomediche, molto meno per le scienze umane e sociali: P. Bourke & L. Butler, "Publication types, citation rates and evaluation" *Scientometrics* 1996, 37(3): 473-94.

²¹⁸ A.F.J. van Raan, "Measuring science..." cit., in particolare pp. 21-22; H. Zeisel, "Difficulties in indicator construction: Notes and queries," in *Toward a Metric of Science: the Advent of Science Indicators*, edited by Y. Elkana, et al. (New York: John Wiley and Sons, 1979): 253-58.

ricevute dalle pubblicazioni di un gruppo di ricerca in un certo settore e il numero di citazioni ricevute dalla totalità delle pubblicazioni nella stesso settore è un indicatore. È implicita nella costruzione di tale indicatore l'idea che la misura rifletta l'impatto scientifico internazionale del gruppo di ricerca e che l'impatto internazionale sia un sinonimo o quantomeno un sintomo della qualità scientifica.

Il concetto di “qualità” di una pubblicazione scientifica è alquanto ambiguo, indica una proprietà difficile da definire o tradurre in una sequenza di istruzioni adatte a distinguere ciò che la possiede in un certo grado da ciò che la possiede in un grado inferiore o non la possiede affatto. Qualità può essere sinonimo di utilità (e di fatto lo è in molti articoli di contenuto medico-clinico) oppure di originalità e potere esplicativo delle idee scientifiche esposte (la qualità per antonomasia della “scienza pura”), ma può anche avere un significato più debole ed alludere all'accuratezza delle tecniche e dei metodi di elaborazione dei dati (la qualità “metodologica” di tanti lavori non originali o puramente corroborativi rispetto a teorie esistenti) o semplicemente all'eleganza del formalismo e dei modelli matematici utilizzati (la qualità “estetica” dei teoremi matematici e di molte teorie fisiche contemporanee). La verifica di queste manifestazioni della qualità è in gran parte materia di giudizio soggettivo o di accordo intersoggettivo tra scienziati chiamati a pronunciarsi sul valore di una pubblicazione.

Gli indicatori scientometrici basati sull'analisi delle citazioni non operano a questo livello, il loro campo d'azione non si esercita direttamente sui contenuti; piuttosto, se parlano di qualcosa, parlano dell'efficacia con cui quei contenuti sono comunicati ad un gruppo di scienziati operanti nello stesso settore d'indagine di chi li ha prodotti. Tale efficacia comunicativa si manifesta attraverso l'“impatto” o “influenza” esercitata sul loro lavoro. “Qualità”, “impatto”, “influenza” sono entità concettualmente diverse. Non è affatto scontato che un documento di qualità abbia un impatto, così come non

è ovvio che un documento letto, recepito e citato da un autore ne influenzi in modo decisivo il comportamento e le scelte successive.²¹⁹ Quello di "influenza", come insegnano gli studi di critica letteraria, è un concetto pericoloso. Non si può parlare di influenza di un autore **A** su un autore **B** senza far intervenire un terzo polo **X** costituito dalla letteratura esistente e dall'universo paradigmatico di teorie, concetti, metodologie condivise da una comunità di studiosi in un periodo storico determinato.²²⁰ **A** può decidere di citare **B**, e non magari **C** che per primo ha introdotto il concetto richiamato dalla citazione, semplicemente perché ignora l'esistenza di **C** oppure per strizzare l'occholino a **B** e cautelarsi nel caso che **B**, un giorno, si trovi dall'altra parte della cattedra a decidere delle sorti accademiche di **A**, o ancora per ostentare la propria conoscenza dell'esistenza e del valore di **B**. L'influenza, in poche parole, non può essere implementata in un algoritmo.²²¹ I bibliometri esibiscono in compenso due tipi di prove dell'equivalenza tra qualità e impatto citazionale. Da un lato, cercano di correlare i risultati dell'analisi delle citazioni con misure indipendenti di performance individuale (peer review, produzione di nuovi farmaci).²²² Si rivolgono poi alla storia delle scienze del '900 e osservano che gli scienziati più blasonati, quelli che hanno concluso qualcosa di importante vedendo riconosciuta l'eccellenza delle loro acquisizioni sotto forma di premi o onoreficenze (il premio Nobel su tutti), hanno pubblicato articoli di gran lunga più citati rispetto alla media.²²³ Non

²¹⁹ Già nel 1964 Garfield avvertiva che: «High citations counts reflect impact but may or may not reflect intrinsic worth. The data obtained from citation analysis are always relative rather than absolute»: E. Garfield; I.H. Sher; R.J. Thorpie, *The Use of Citation Data...cit.*, p. 1.

²²⁰ U. Eco, *Sulla letteratura...cit.*, pp. 128-146.

²²¹ D.O. Edge, "Quantitative measures of communication in science: a critical review" *History of science* 1979, 17: 102-34.

²²² Per esempio: L.I. Meho & D.H. Sonnenwald, "Citation ranking versus peer evaluation of senior faculty research performance: A case study of Kurdish scholarship" *Journal of the American Society for Information Science* 2000, 51(2): 123-28; M.B. Albert; D. Avery; F. Narin et al., "Direct validation of citation counts as indicators of industrially important patents" *Research Policy* 1991, 20(3): 251-59; M.E.D. Koenig, "Bibliometric indicators versus expert opinion in assessing research performance" *Journal of the American Society for Information Science* 1983, 34(2): 136-45; J.A. Virgo, "A statistical procedure for evaluating the importance of scientific papers" *Library Quarterly* 1977, 47: 415-30.

²²³ E. Garfield, "Do Nobel Prize winners write citation classics?," in *Essays of an Information Scientist 1986, Towards Scientography*, 9 (Philadelphia: ISI Press, 1988): 182-87.

solo: l'alto tasso di citazioni appare indipendente dalla visibilità acquisita dopo il conferimento del premio, quindi, in qualche modo, risulta direttamente correlato al "significato" del loro lavoro.²²⁴ Se si accettano le conseguenze teoriche di tale situazione e se, come fanno i citazionisti, si considera l'impatto un elemento essenziale della qualità o più modestamente un aspetto della qualità scientifica, allora vale la pena interrogarsi sull'utilità degli indicatori bibliometrici nella valutazione della scienza.²²⁵

Negli ultimi anni, il dibattito sui sistemi e gli standard di valutazione ha prodotto un volume impressionante di letteratura, nella quale trovano posto tesi e soluzioni molto diverse tra loro. Ad un estremo si collocano i difensori dell'*ancien régime* i quali, pur riconoscendo la necessità di un cambiamento, condannano dopo un processo sommario l'uso valutativo degli indicatori bibliometrici, liquidati come fundamentalmente inutili, in quanto «non c'è niente di così affidabile come un accurato processo di *peer review*».²²⁶

Affermazioni così drastiche sono abbastanza frequenti, soprattutto tra direttori editoriali e autori che lamentano la scarsa visibilità imposta da una multinazionale della cultura alle proprie creature e si vedono ingiustamente penalizzati da una laconica hit parade di titoli. All'estremo opposto si trova chi ha riposto una tale fiducia nelle analisi bibliometriche da ritenere che una loro applicazione meccanica possa condurre ad automatizzare le procedure di reclutamento dei docenti universitari.²²⁷

²²⁴ Vedi esperimento condotto dai Cole sul tasso di citazioni dei premi Nobel in fisica tra il 1965 e il 1965: S. Cole & J.R. Cole, "Scientific output and recognition: a study in the operation of the reward system in science" *American Sociological Review* 1967, 32(3): 377-90.

²²⁵ «This means that basic quality is a necessary, but not a sufficient condition to make an impact. According to this view, one aspect of successful research performance is that researchers are active in presenting their research findings to colleagues researchers. In fact, we consider this activity as an aspect of scientific quality in a more broader sense. Scientific quality thus defined includes basic quality as well as the extent to which researchers successfully perform "public relations" activities»: H.F. Moed; W.J.M. Burger; J.G. Frankfort et al., *op. cit.*, pp. 134-135.

²²⁶ S. Bloch & G. Walter, "The Impact Factor: time for change" *The Australian and New Zealand Journal of Psychiatry*. 2001, 35(5): 563-68, p. 567.

²²⁷ Ad esempio C. Nicolini; S. Vakula; M.I. Balla et al., "Can the assignment of university chairs be automated?" *Scientometrics* 1995, 32: 93-107, da confrontare con l'esame critico condotto in P. Vinkler, "Some aspects of the evaluation of scientific and related performances of individuals" *Scientometrics* 1995, 32(2): 109-16.

Su un registro più moderato e metodologicamente più solido, altri autori hanno approfondito, sia dal punto di vista concettuale che operativo, i limiti e le condizioni di applicabilità dell'*IF*, delle sue possibili alternative e più in generale degli indicatori bibliometrici nel controllo di qualità della ricerca scientifica.

5.2.1. Il conteggio delle citazioni al livello dell'individuo

Un'alternativa sembra a portata di mano, persino ovvia: si contano i voti del candidato di turno, gli spettatori del talk show, i dischi venduti dalla rockstar, i lettori del best-seller, perché non contare le citazioni guadagnate dalle pubblicazioni del singolo scienziato nel corso della sua carriera? Il numero di citazioni accumulate in un certo intervallo di tempo dai lavori di uno scienziato è stato considerato «uno dei migliori indicatori scientometrici disponibili per la valutazione»,²²⁸ soprattutto se calcolato, come l'*IF*, in rapporto al numero complessivo dei lavori pubblicati in quel settore.

Per favorire i conteggi individuali, nel 2001 l'ISI ha dotato il portale *Web of Science* di uno strumento analitico chiamato *Essential Science Indicators* che, rielaborando i dati dei *citation indexes*, estrae in pochi passaggi «gli autori più citati, le istituzioni più citate, le nazioni più citate, le riviste più citate» all'interno di 22 campi scientifici predefiniti.²²⁹ Gli scienziati più citati di ciascuna categoria lungo un intervallo di tempo predefinito (per ora solo il periodo 1981-1999), in poche parole l'élite citazionale del mondo, godono inoltre di una vetrina d'eccezione nel sito ad accesso libero *ISIHighlyCited.com* dove, per ciascun autore, sono fornite informazioni biografiche e bibliografiche, curriculum professionale, onoreficenze ricevute,

²²⁸P. Vinkler, "Some aspects of the evaluation of scientific and related performances of individuals" *Scientometrics* 1995, 32(2): 109-16, p. 111.

membership e posti di responsabilità ricoperti.²³⁰ Chiunque può scoprire così in pochi passaggi quanto dista dall'olimpio dei *citation classics*.

In realtà un indicatore di performance al livello dei singoli articoli e scienziati solleva più problemi di quanti non ne risolva.

Garfield ha ammesso esplicitamente che i dati citazionali al livello dell'individuo sono controversi, «facilmente fraintesi o inconsapevolmente manipolati per obiettivi impropri», e che pertanto l'analisi delle citazioni è un'attività estremamente complessa, inscindibile da un'accurata e soggettiva procedura di *peer review*, in quanto «coinvolge un'interpretazione profonda di ciascun libro e articolo prodotto dal candidato».²³¹ Quando si restringe il campo di osservazione e raccolta dei dati dalla rivista al singolo documento e al singolo scienziato, inevitabilmente si perdono i vantaggi legati all'omogeneità e alle dimensioni del campione da cui ricavare le distribuzioni probabilistiche, e ci si trova a fare i conti con l'ambiguità dell'individuo, di ciò che rende l'individuo unico e per certi aspetti inclassificabile. Sarebbe assurdo, ad esempio, dare giudizi qualitativamente divergenti su candidati i cui lavori differiscono per una decina di citazioni ricevute. Le differenze non sono statisticamente significative, anche perché, nel caso di documenti prodotti da candidati giovani, è verosimile che non sia trascorso il tempo sufficiente per essere citati.²³² La probabilità di essere citati dipende infatti, oltre che dagli stili citazionali delle diverse discipline, dal tasso di crescita della letteratura e dal numero medio di citazioni per articolo caratteristici di un settore di ricerca: è evidente che un maggiore numero di pubblicazioni implica

²²⁹ Thomson ISI, *ISI Essential Science Indicators™ Tutorial Version 1*, 2001

<<http://www.thomsonisi.com/tutorials/esi/>>.

²³⁰ <<http://isihighlycited.com/>>.

²³¹ E. Garfield, "How to use citation analysis for faculty evaluations, and when is it relevant? Part 1," in *Essays of an Information Scientist 1983*, 6 (Philadelphia: ISI Press, 1984): 354-62, pp. 354-355.

²³² E' possibile anche in questo caso ricorrere a modelli matematici che, partendo dalle convenzioni citazionali e dal tasso di crescita della letteratura in un'area disciplinare, consentano di predire con una certa approssimazione la percentuale di citazioni ottenibili da un documento nell'arco di 40 anni. Vedi: N.L. Geller & J.S. De Cani, "Lifetime-citation rates to compare scientists' work" *Social Science Research* 1978, 7(4): 345-65; E. Garfield, "Lifetime citation rates," in *Essays of an Information Scientist 1979-1980*, 4 (Philadelphia: ISI Press, 1981): 355-58.

un maggiore numero di riferimenti bibliografici e che più si cita più si può essere citati.²³³ La velocità con cui un documento viene citato dopo la pubblicazione ed il suo tasso di invecchiamento, cioè il tempo necessario a cadere nell'oblio, sono a loro volta variabili dipendenti dalle consuetudini di un campo di ricerca e possono cambiare, all'interno dello stesso campo, nel corso del tempo. La tabella di *cited half life* del *JCR*, in tal caso, può aiutare a stabilire se quel contributo è invecchiato più rapidamente della media degli articoli della rivista su cui è uscito, ma invecchiamento non significa obsolescenza o perdita di validità, perché può darsi che l'argomento trattato sia semplicemente meno "alla moda" di altri. L'impatto di un lavoro individuale, peraltro, può manifestarsi nel breve e nel lungo periodo: nel breve periodo può indicare il grado di coinvolgimento di uno scienziato nel fronte attivo della ricerca in un dato settore, nel lungo periodo può essere la spia del livello di "durevolezza" di un contributo e precludere alla sua inclusione nel corpo delle conoscenze acquisite.²³⁴

Un'analisi attendibile al livello del singolo autore richiederebbe inoltre che, una volta stabilito "quante volte" egli è stato citato e "da chi", e ricostruita la rete delle connessioni tra gli appartenenti al medesimo campo di ricerca, si identificasse la sua collocazione precisa nella rete, "perché" è stato citato. Per chiarire questo punto bisogna risalire alla fonte delle citazioni ed effettuare un'analisi del contenuto e del contesto in cui compaiono: l'analisi del contesto rivela quale aspetto del lavoro del candidato è stato citato e quale grado di consenso esiste circa il significato ad esso attribuito, l'analisi del contenuto rivela l'intenzione della citazione, ad esempio se l'aspetto in questione è stato richiamato per essere contestato o confermato.²³⁵

²³³ P. Vinkler, "Relationships between the rate of scientific development and citations. The chance for citedness model" *Scientometrics* 1996, 35(3): 375-86.

²³⁴ H.F. Moed; W.J.M. Burger; J.G. Frankfort et al., "The use of bibliometric data for the measurement of university research performance" *Research Policy* 1985, 14: 131-49, pp. 133-134.

²³⁵ E. Garfield, "How to use citation analysis for faculty evaluations, and when is it relevant? Part 2," in *Essays of an Information Scientist* 1983, 6 (Philadelphia: ISI Press, 1984): 363-72, p. 363.

D'altra parte, il silenzio e la mancata citazione non sono necessariamente sinonimo di mediocrit . Un contributo pu  essere ignorato perch  la sua originalit  lo rende in anticipo sui tempi, come   accaduto, nella seconda met  dell'800, alle ricerche di Mendel sull'ereditariet  dei caratteri dominanti e recessivi, ma pu  essere ignorato, come suggeriva Merton, anche per la ragione opposta, cio    talmente importante e integrato in un corpo di conoscenze acquisite che si tende a darlo per scontato, ad "obliterarlo". Nel primo caso la giustizia citazionale postuma non rende alcun servizio al genio incompreso, nel secondo semplicemente i conteggi risultano falsati.

In conclusione, le citazioni di un documento non riflettono il suo valore assoluto n  la sua utilit  pratica, e non dicono nulla sulle abilit  didattiche o manageriali dell'autore, spesso utili quanto e forse pi  delle prime per il potenziamento di un settore di ricerca; piuttosto riflettono, con un certo grado di approssimazione, l'uso della pubblicazione da parte degli altri scienziati che si occupano di argomenti affini. Ne consegue che un impiego plausibile dell'analisi delle citazioni pu  avvenire, anzich  nella valutazione diretta del candidato, al livello della scelta e della valutazione degli stessi revisori: la mappa dei documenti e degli autori citati dal candidato nonch  degli autori che hanno citato i suoi lavori e di quelli che hanno citato i citanti, oppure i *cluster* di autori ricavabili applicando la tecnica delle co-citazioni, rappresentano un valido supporto all'identificazione del "collegio invisibile" cui il candidato appartiene, quindi degli studiosi pi  qualificati per giudicare il suo valore e degli studiosi coi quali comparare la qualit  della sua produzione.

5.2.2. Il conteggio delle citazioni al livello dell'istituzione

Alla luce delle critiche accumulate e delle difficolt  metodologiche dei conteggi individuali, l'*IF* delle riviste   stato ripreso, approfondito, smontato e rimontato nel quadro di modelli matematici pi  o meno raffinati. Ne   derivata

una serie di varianti e coefficienti di normalizzazione dell'indice di Garfield: *subfield factor*, *normalized impact factor*, *research production unit (rpu)*, *discipline impact factor (dif)*, *standardized citation score (scs)*, *characteristics score and scale (cs)*.²³⁶

La metodologia bibliometrica sviluppata all'Università di Leida da Anthony F.J. van Raan, per esempio, scavalca alcuni dei problemi tecnici e concettuali visti nel paragrafo precedente scegliendo le unità di valutazione ad un livello di aggregazione superiore a quello dell'individuo, precisamente al livello dell'istituzione (facoltà, dipartimento, istituto) e del gruppo di ricerca.

L'analisi delle citazioni si applica in questo caso al corpus di pubblicazioni prodotte lungo un certo numero di anni dai membri dell'istituzione o del gruppo di ricerca. L'impatto sulla comunità scientifica internazionale, considerato il principale criterio di eccellenza per chi svolge attività scientifica, è misurato da un *internationally standardized impact indicator*. La misura è ricavata comparando il tasso medio di citazioni per pubblicazione del corpus sotto esame con il tasso medio di citazioni di tutti gli articoli pubblicati, su scala mondiale, nelle stesse riviste in cui sono usciti i lavori dei membri del gruppo.²³⁷ Si ottiene così una stima di quanto l'impatto misurato sia al di sopra o al di sotto della media internazionale. L'analisi dei cluster di co-occorrenze di parole chiave consente in un secondo momento di perfezionare la misura "cartografando" le sottoaree disciplinari in cui si esercita maggiormente l'impatto citazionale dell'istituto o del gruppo di ricerca.

Si può discutere se l'indicatore proposto da van Raan misuri effettivamente ciò che intende misurare, cioè l'impatto internazionale di un'istituzione, e soprattutto se, nel valutare la performance di comunità locali di studiosi, la

²³⁶ Una rassegna di tali indicatori si trova in: P. Vinkler, "An attempt of surveying and classifying bibliometric indicators for scientometric purposes" *Scientometrics* 1988, 13: 239-59.

²³⁷ A.F.J. van Raan, "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* (Medford, NJ: Information Today Inc, 2000): 301-19, in particolare pp. 315-317 (Appendice).

misura debba necessariamente puntare così in alto. In ogni caso, l'esempio di Leida mostra che, nel dibattito scientometrico recente, il controllo di qualità dipende dalla costruzione di indicatori di performance "relativi", cioè indicatori che permettano di confrontare il numero di citazioni ricevute dalle pubblicazioni prodotte da gruppi di scienziati (non scienziati singoli) con uno standard di riferimento convenzionalmente scelto come "assoluto", ad esempio il numero di citazioni "attese" o probabili per le riviste o per il campo di ricerca cui le pubblicazioni appartengono.²³⁸ Non basta contare il tasso di citazioni e registrarne l'andamento lungo quattro, dieci o più anni, ci vuole una scala di misura per poter dire "quanto" un risultato sia "basso", "medio" o "alto", ed è su questo terreno che si incontrano seri problemi di metodo, sia perché non esiste una regola per stabilire quali differenze siano da considerare statisticamente significative e quali invece da attribuire al caso, sia perché le potenziali unità di misura da usare per il confronto, come il numero delle citazioni "attese" per rivista, dipendono a loro volta da calcoli effettuabili sulle basi dati dell'ISI e dalle relative limitazioni. Il punto di arrivo rischia così di coincidere con il punto di partenza del quale si volevano evitare i difetti.²³⁹

5.3. Le citazioni dei brevetti tra scienza, tecnologia e giurisprudenza

In un'economia globale e *knowledge-based* la competizione decisiva tra imprese non si esercita, come nei secoli passati, sul controllo dei mercati e delle materie prime, ma sulla proprietà intellettuale delle idee, in particolare di

²³⁸ P. Vinkler, "Model for quantitative selection of relative scientometric impact indicators" *Scientometrics* 1996, 36(2): 223-36, p. 224.

²³⁹ Su questo punto vedi: H.F. Moed; W.J.M. Burger; J.G. Frankfort et al., "The use of bibliometric data for the measurement of university research performance" *Research Policy* 1985, 14: 131-49, pp. 137-139.

quelle idee che si materializzano in tecniche e oggetti di immediato (o auspicato) valore commerciale.

Il brevetto è un titolo giuridico che riconosce ad un inventore il monopolio temporaneo di sfruttamento su un'invenzione - una macchina, un prodotto, un processo, una tecnica - dotata di alcuni requisiti essenziali quali la novità, la non ovvietà rispetto allo stato della tecnica esistente e l'applicabilità industriale. In un brevetto la componente tecnico-scientifica, quella economica e quella giuridico-territoriale sono inseparabili prima, durante e dopo la stesura del documento: l'inventore deve redigere la domanda di brevetto in una forma standardizzata, legalmente riconosciuta, e deve sottoporla agli uffici brevetti di tutti i paesi nei quali intende far valere i propri diritti.

Verso la metà degli anni '80, le tradizionali statistiche econometriche, mirate a quantificare l'innovazione tecnologica attraverso il conteggio dei brevetti, sono state affiancate dall'analisi delle citazioni. Questo passaggio, dettato dall'estensione delle tecniche bibliometriche alla documentazione tecnologica, ha posto il problema di chiarire le analogie e le diversità tra la composizione bibliografica di un brevetto e quella del classico articolo di rivista.

A differenza di un articolo scientifico, il brevetto non crea o non aspira a creare nuova conoscenza, ma utilizza una parte delle nozioni tecnico-scientifiche esistenti per produrre un oggetto tecnologicamente innovativo, provvisto di una qualche utilità sociale e riproducibile su scala industriale.

Una distinzione così netta è abbastanza plausibile e familiare. La si ritrova, ad esempio, in tutti i progetti educativi che, a partire dalle scuole inferiori, insegnano la scienza con esempi pratici tratti dagli oggetti della vita quotidiana. Il suo inconveniente però è quello di dare un'idea eccessivamente semplicistica delle complesse interrelazioni tra scienza, tecnologia ed economia che si materializzano nella letteratura brevettuale. Il limite principale consiste nel suggerire un rapporto lineare di dipendenza

dell'economia rispetto alla tecnologia e della tecnologia rispetto alla scienza: lo scienziato fa ricerca ed elabora teorie che reggono alla prova dell'esperienza, comunica poi al mondo i risultati che, una volta diventati di pubblico dominio, vengono codificati sotto forma di manuali e programmi di insegnamento superiori; le nozioni scientifiche così disponibili sono manipolate dall'inventore per progettare un'invenzione brevettabile che, una volta fabbricata, viene immessa sul mercato e contribuisce allo sviluppo dell'economia.

Questo schema, che in sostanza riduce la tecnologia a scienza applicata, fa da contrappunto alla tesi, risalente ad Aristotele, secondo cui scienza e tecnologia sono due forme di conoscenza radicalmente diverse: puramente speculativa e disinteressata la prima, pratica e condizionata da interessi socio-economici la seconda. Entrambe le concezioni trascurano aspetti essenziali del lavoro scientifico e tecnologico, ad esempio il fatto che esistono e sono esistite tecnologie capaci di svilupparsi quasi indipendentemente dal sapere speculativo, che molti settori della scienza contemporanea incorporano, sin dalla fase della raccolta e della registrazione dei dati empirici necessari alla costruzione delle ipotesi esplicative, una forte componente tecnologica, e che la scienza non è meno sensibile della tecnica a influenze sociali ed economiche nella selezione dei suoi obiettivi.

Esaminata attraverso la lente d'ingrandimento del brevetto, la trama di analogie e differenze tra scienza e tecnologia si arricchisce di ulteriori elementi.

Francis Narin, pioniere della bibliometria dei brevetti, ha mostrato che, in settori come le biotecnologie, non esiste uno scarto temporale tra i cicli di *referencing* della scienza e della tecnologia: il tempo che ci mette un articolo scientifico a citare un altro articolo scientifico è virtualmente lo stesso impiegato da un brevetto per citare un articolo scientifico o un altro

brevetto.²⁴⁰ Ciò dipende in gran parte dalla sovrapposizione materiale delle due attività. «Negli Stati Uniti - scrive Narin - l'inventore lavora all'Università o in un laboratorio governativo la mattina, presta servizio o fa da consulente per una società privata nel pomeriggio, e l'intervallo di tempo tra la ricerca accademica e l'attività inventiva privata è il pranzo».²⁴¹

Anche nella tecnologia, come nella scienza, il ciclo di vita delle voci bibliografiche varia a seconda dei settori: un brevetto di elettronica cita normalmente brevetti più recenti di quanto non faccia uno di tecnologie navali, così come un articolo di biochimica contiene riferimenti bibliografici più recenti di uno di acustica. Anche nella tecnologia inoltre si tende a citare maggiormente documenti della stessa area geografica e, in misura più rilevante che nella letteratura periodica, la distribuzione delle citazioni tra brevetti di diverso calibro è asimmetrica: per ogni area applicativa, esiste un numero assai limitato di brevetti molto citati e di riconosciuta importanza tra gli addetti ai lavori a fronte di un numero molto maggiore di brevetti di impatto scarso o nullo.

Malgrado le affinità, la struttura bibliografica di un brevetto è molto diversa da quella di un articolo scientifico e tale diversità impone cautela nella generalizzazione delle tecniche bibliometriche applicate alla letteratura scientifica.

Un brevetto è costituito da tre sezioni fondamentali: 1) un frontespizio contenente i dati bibliografici e amministrativi necessari all'identificazione del documento (titolo, abstract, numeri di classificazione, nome e indirizzo dell'inventore e della ditta assegnataria, data e numero della domanda); 2) la descrizione dell'invenzione, corredata da eventuali illustrazioni, con le specifiche tecniche e i riferimenti bibliografici forniti dall'inventore; 3) le rivendicazioni (*claims*), cioè l'elenco dettagliato delle caratteristiche specifiche dell'invenzione per le quali si chiede tutela giuridica.

²⁴⁰ F. Narin & E. Noma, "Is technology becoming science?" *Scientometrics* 1985, 7: 369-81.

Nella stesura della seconda e della terza sezione risulta spesso decisivo l'intervento di un consulente in materia di brevetti (*patent attorney* o *patent agent*). Laddove è prevista una procedura di verifica della sussistenza dei requisiti di brevettabilità, come negli Stati Uniti, in Giappone e presso l'European Patent Office, il brevetto contiene (spesso sul frontespizio) una lista aggiuntiva di riferimenti bibliografici alla letteratura brevettuale o tecnico-scientifica, cioè al complesso dei documenti che formano lo stato dell'arte. I riferimenti supplementari sono forniti da un esperto della materia, un esaminatore (*patent examiner*), con l'obiettivo di determinare in maniera esplicita l'ampiezza dei diritti da tutelare, quindi di precisare e limitare le rivendicazioni di novità e utilità del brevetto.

Le citazioni contenute nei brevetti sono pertanto il risultato di un processo sociale nel quale si trovano coinvolti diversi attori: l'inventore, che fornisce la materia prima del documento sotto forma di descrizioni, disegni, specifiche tecniche; il consulente brevettuale, che, in virtù di una conoscenza approfondita del contesto giuridico, assiste l'inventore nella redazione del testo e nella scelta dei riferimenti bibliografici appropriati; l'esaminatore, che deve effettuare una ricerca della letteratura esistente e attingere alle proprie conoscenze personali per verificare ed eventualmente ampliare la sezione bibliografica con riferimenti ignorati o dimenticati dall'inventore.

La diversa provenienza e finalità delle citazioni condiziona il tipo di informazioni ricavabili dall'analisi bibliometrica.

Le citazioni contenute nelle specifiche tecniche, difficili da estrarre e indicizzare, documentano la base teorica di un'invenzione o almeno ciò che l'inventore, spinto dal desiderio di amplificare la novità della proposta, cerca di far apparire come base teorica. La loro è una funzione retorica simile a quella delle citazioni accademiche.

²⁴¹ F. Narin, "Patent bibliometrics" *Scientometrics* 1994, 30: 147-55, pp. 150-151.

Le citazioni degli esaminatori invece, in linea di principio, hanno una correlazione più marcata con il background scientifico, tecnologico e giuridico di un'invenzione e, nell'ipotesi che il processo di esame sia condotto in maniera approfondita e imparziale, puntano alla rilevanza tecnologica "oggettiva" dei documenti citati anziché al loro impatto.²⁴² In un articolo scientifico non esiste alcuna forma di contraddittorio interno paragonabile all'intervento di un esaminatore. Un'analogia formale può stabilirsi quando il *referee* di una rivista rimanda all'autore il manoscritto esaminato con la raccomandazione di letture supplementari. Si tratta però di un'analogia apparente. Una volta incorporate nella struttura del testo, le nuove citazioni sono indistinguibili dalle vecchie ed ogni traccia dell'interpolazione si perde allorché la versione ufficiale viene pubblicata. Per quanto paradossale possa apparire, un documento destinato allo sfruttamento di interessi commerciali incorpora un sistema di verifica dell'oggettività bibliografica estraneo agli articoli di riviste, veicoli dell'oggettività scientifica (ma non bibliografica) per eccellenza.

Negli Stati Uniti, dove l'inventore ha l'obbligo giuridico di esibire lo stato dell'arte nella materia specifica dell'invenzione, molte delle citazioni prodotte dall'esaminatore coincidono con quelle dell'inventore, il quale tende a cautelarsi dall'accusa di insufficienza bibliografica citando anche documenti che abbiano una rilevanza minima rispetto al contenuto del brevetto. In tal caso il *search report* dell'esaminatore ha uno scopo più ampio rispetto alla verifica di brevettabilità ed assomiglia piuttosto ad una ricerca documentaria dei presupposti tecnico-scientifici dell'invenzione.²⁴³ Presso l'European Patent Office, invece, non sussistendo alcun obbligo di documentazione per

²⁴² M. Meyer, "What is special about patent citations? Differences between scientific and patent citations" *Scientometrics* 2000, 49(1): 93-123.

²⁴³ Sulla struttura del *search report* di un brevetto e sulle differenze tra il sistema americano, quello giapponese e quello europeo vedi: J. Michel & B. Bettels, "Patent citation analysis. A closer look at the basic input data from patent search reports" *Scientometrics* 2001, 51(1): 185-201.

l'inventore, vige un uso parsimonioso dei dati bibliografici, i quali sono prodotti in massima parte dall'esaminatore.

I bibliometri discordano sul valore da attribuire alle citazioni dei *patent examiners*, giudicate spesso inappropriate e fuorvianti,²⁴⁴ ed alcuni pensano addirittura che dovrebbero essere escluse dall'analisi statistica.²⁴⁵ In ogni caso, enumerate in un *patent citation index* e collegate ai brevetti di provenienza, esse formano delle chiavi di ricerca alternative rispetto ai codici di classificazione, artificiosi e perennemente in ritardo sulle dinamiche di mercato che orientano l'evoluzione tecnologica.²⁴⁶ Secondo Garfield le citazioni «indicizzano il brevetto in una maniera che sarebbe impossibile per un indicizzatore e permettono al *citation index* di dispiegarsi attraverso sottoclassi di letteratura brevettuale in modo analogo a come, nella letteratura periodica, si dispiega attraverso discipline e specialità scientifiche».²⁴⁷ Contate e classificate nel quadro di un'analisi statistica, effettuata normalmente su database commerciali con l'ausilio di software specifici come il *Derwent Analytics* dell'ISI-Thompson, le citazioni dei brevetti sono state usate per sviluppare indicatori di solidità o potere tecnologico su scala locale, nazionale o internazionale.²⁴⁸ L'analisi coinvolge diverse aree sensibili della pianificazione finanziaria e politica: l'identificazione dei settori tecnologici trainanti e dei ricercatori di punta in un settore; la ricerca dei brevetti "forti" detenuti da imprese concorrenti nell'ottica di una politica di acquisizione e fusione aziendale; lo studio delle performance tecnologiche a livello nazionale

²⁴⁴ U. Schmoch, "Tracing the knowledge transfer from science to technology as reflected in patent indicators" *Scientometrics* 1993, 26(1): 193-211; P. Collins & S. Wyatt, "Citations in patents to the basic research literature" *research Policy* 1988, 17(2): 65-74.

²⁴⁵ P. Vinkler, "The origin and features of information referenced in pharmaceutical patents" *Scientometrics* 1994, 30(1): 283-302.

²⁴⁶ M. Barbieri & G. Conti, "La ricerca documentale dei brevetti" *Biblioteche Oggi* 2005, 23(3): 15-18.

²⁴⁷ «...references that index the patent in a way that an indexer could not and permit the citation index to reach across subclasses of the patent literature in the same way that it reaches across disciplines and specialties in the journal literature.»: E. Garfield, *Citation Indexing: its Theory and Application...cit.*, p. 38.

²⁴⁸ Sugli usi dell'analisi citazionale dei brevetti vedi: C. Oppenheim, "Do patent citations count?," in *The Web of Knowledge: a Festschrift in Honor of Eugene Garfield* (Medford, NJ: Information Today Inc, 2000): 405-32; M.M.S. Karki, "Patent citation analysis: A policy analysis tool" *World Patent Information* 1997, 19(4): 269-72.

e aziendale; l'interpretazione dei legami tra aree di ricerca e sviluppo, tra aziende concorrenti, tra tecnologie e scienza di base. La percentuale di citazioni verso la letteratura scientifica, ad esempio, è diventato un indice della correlazione tra tecnologia e scienza di base utile a valutare, su base comparativa, quanto *high-tech* sia un'azienda di alta tecnologia.

Analogamente la percentuale di citazioni sedimentate su un brevetto o su una classe di brevetti è stata interpretata come un indice di competitività tecnologica in funzione di scopi diversi: fungere da supporto decisionale per le strategie di acquisizione, vendita o scambio di licenze; misurare il valore di mercato del portafoglio brevetti e degli investimenti di un'azienda nel settore della ricerca;²⁴⁹ prevedere l'entità dei guadagni derivanti dalle tecnologie brevettate.²⁵⁰

Una volta immessa nel circuito bibliometrico, la citazione bibliografica muove capitali, persone, mercati, e contribuisce a formare la base empirica di scelte e decisioni operative fondamentali per la sopravvivenza e il successo economico di un'azienda: «i database di brevetti sono Biblioteche di Alessandria virtuali di informazioni che, combinate con nuovi strumenti automatizzati di data-mining e visualizzazione, si trasformano in strumenti di potente intelligenza competitiva che le imprese possono usare traendone grande vantaggio».²⁵¹

Anche in questo caso però, come nell'analisi classica delle citazioni, si tenda a dare per scontato che il sistema di produzione e controllo dei dati bibliografici funzioni in modo uniforme e trasparente, e si accetta in maniera più o meno pacifica il principio di equivalenza tra numero di citazioni e qualità, in virtù del quale «un brevetto molto citato - cioè un brevetto che compare come

²⁴⁹ B.H. Hall; A. Jaffe; M. Trajtenberg, *Market value and patent citations*, 2002

<http://emlab.berkeley.edu/users/bhhall/papers/HallJaffeTrajtenberg_RJEjan04.pdf>.

²⁵⁰ B.N. Sampat & A.A. Ziedonis, "Patent citations and the economic value of patents. A preliminary assessment," in *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems*, edited by H.F. Moed, et al. (Dordrecht: Kluwer Academic, 2004): 277-97.

riferimento bibliografico in molti brevetti pubblicati successivamente - contiene un avanzamento tecnologico di particolare rilievo». ²⁵² Tale presupposto ha il pregio di far funzionare la macchina alimentandone il serbatoio con dati che si presume grezzi e omogenei, ma la sua solidità è minacciata tanto dall'inefficienza dei sistemi di verifica dei requisiti di brevettabilità, ²⁵³ quanto dalla nostra ignoranza circa i ruoli rivestiti dalla citazione bibliografica nella costruzione del discorso scientifico e tecnologico. ²⁵⁴

²⁵¹ K.G. Rivette & D. Kline, *Rembrandts in the Attic: Unlocking the Hidden Value of Patents* (Boston: Harvard Business School Press, 2000), p. 28.

²⁵² F. Narin & D. Olivastro, "Technology indicators based on patents and patent citations," in *Handbook of Quantitative Studies of Science and Technology*, edited by A.F.J. Van Raan (Amsterdam: Elsevier Science Publishers, 1988): 465-507, p. 475.

²⁵³ Una critica pesante all'efficienza del sistema americano di esame dei brevetti e all'utilità della *patent citation analysis* si trova in: G. Aharonian, "U.S. patent examination system is intellectually corrupt" *Patnews* 2000(1 may) <<http://www.questel.orbit.com/Piug/piugl2000/0523.html>>.

²⁵⁴ C. Oppenheim, "Do patent citations..." cit., in particolare pp. 421-423.

Capitolo 6

Misurare la comunicazione scientifica nel XXI secolo: dalla bibliometria alla webmetrica

L'avvento del World Wide Web ha reso il sogno di Garfield più concreto della realtà. Nella rete mondiale degli ipertesti, non solo il riferimento bibliografico è una delle possibili forme assunte da un hyperlink all'interno di un articolo scientifico in formato digitale, ma è il Web stesso ad esibire una struttura citazionale, con link tra pagine web formalmente simili alle citazioni bibliografiche ("sitazioni", *sitations*).²⁵⁵ La doppia natura della citazione ipertestuale, che da un lato mantiene salda la collocazione tradizionale nella sezione bibliografica del saggio accademico, dall'altro si libera da ogni costrizione discorsiva e stilistica per coinvolgere l'unità organizzativa di tutto lo spazio web, spiega i due percorsi intrapresi dall'analisi delle citazioni in ambiente digitale: primo, la progettazione di sistemi automatici di indicizzazione delle citazioni contenute negli *e-journals* e negli archivi aperti estranei ai circuiti ufficiali di *abstracting* e *indexing* (paragrafo 6.1.); secondo, l'applicazione dei metodi bibliometrici allo studio della rete di hyperlink su cui si fonda la struttura del Web (paragrafo 6.2.).

²⁵⁵ Il termine *sitation* è stato introdotto, nel primo numero della rivista *Cybermetrics*, da: R. Rousseau, "Sitations: an exploratory study" *Cybermetrics* 1997, 1(1) <<http://www.cindoc.csic.es/cybermetrics/vol1iss1.html>>.

6.1. La citazione negli e-journals e negli archivi aperti

Negli ultimi dieci anni, i grandi editori e aggregatori di database hanno aderito alla radicale trasformazione dei modi di produzione e circolazione della letteratura scientifica innescati dalle reti telematiche caricando nei loro server una massa imponente di documentazione scientifica in formato elettronico. Continuiamo allora a giocare al gioco della citazione secondo la regola che impone l'equivalenza tra impatto, qualità e numero di citazioni e chiediamoci in che modo si può stimare la qualità di tali materiali.

Tecnicamente l'impatto di una rivista elettronica estranea al circuito dell'ISI si può quantificare in diversi modi: 1) contando il numero di download degli articoli dal server che li ospita; 2) intervistando o sottoponendo un questionario ad un campione significativo di utenti, per esempio i ricercatori di un dipartimento universitario; 3) contando, attraverso un motore di ricerca, il numero di link al sito della rivista; 4) contando, come fanno gli indici dell'ISI, le citazioni bibliografiche ricevute dagli articoli della rivista.

La prima soluzione sfrutta la capacità dei calcolatori di immagazzinare enormi quantità di informazioni relative al traffico di rete: il monitoraggio dei file di log delle sessioni di collegamento a Internet è una metodologia diffusa di stima della popolarità di una risorsa, nella quale il criterio scelto per l'identificazione delle singole sessioni, notoriamente problematico per la natura *stateless* delle connessioni web, determina la qualità dell'analisi quantitativa.²⁵⁶ La seconda è una soluzione collaudata nel campo delle scienze sociali: a differenza della precedente può contare su una base empirica meno ampia e necessita di un faticoso lavoro preliminare di raccolta e

²⁵⁶ Il classico metodo del *timeout* è stato affiancato, di recente, da sistemi più raffinati, come i modelli statistici del linguaggio a n-grammi mutuati dalle tecniche di riconoscimento vocale, il cui obiettivo è quello di predire e identificare sequenze di *log entries* riconducibili ad uno stesso argomento: X. Huang; F. Peng; A. An et al., "Dynamic web log session identification with statistical language models" *Journal of the American Society for Information Science and Technology* 2004, 55(14): 1290-303.

classificazione manuale dei dati. Entrambe le metodologie hanno dei vantaggi, ma oscurano il fatto che l'analogia link/citazione propende per una soluzione diversa del problema, nella quale, in linea con il progetto di Garfield, non è il giudizio formulato da un singolo o la sessione di ricerca individuale, bensì la rete delle relazioni bibliografiche tra documenti a decidere del valore del singolo contributo. Se il Web è assimilabile ad un gigantesco *citation index*, allora deve essere possibile, come per gli articoli di riviste, ricavare dai link ipertestuali informazioni attendibili circa la visibilità e l'impatto di un documento. La terza soluzione opera in questa direzione, ma ha dato esiti discordanti nel confronto con l'*impact factor* canonico delle riviste: misurando il numero di link ai siti web di un campione eterogeneo di *e-journals* la correlazione tra le due misure è scarsa,²⁵⁷ mentre ritagliando il campione su un set di riviste specializzate in una sola disciplina la correlazione esiste ed appare significativa.²⁵⁸ L'ultima è la via di *CiteBase* e *CiteSeer*: sistemi automatici di indicizzazione della letteratura scientifica predisposti al conteggio individuale delle citazioni di documenti disseminati negli archivi ad accesso libero.

6.1.1. Citazioni e Open Access

«Immaginate un database bibliografico e citazionale universale che leghi ogni lavoro scientifico scritto - non importa in quale forma - a tutti i lavori che cita

²⁵⁷ S.P. Harter & C.E. Ford, "Web-based analysis of E-journal impact: Approaches, problems, and issues" *Scientometrics* 2000, 51(13): 1159-76.

²⁵⁸ L. Vaughan & K. Hysen, "Relationship between links to journal Web sites and impact factors" *Aslib Proceedings* 2002, 54(6): 356-61; L. Vaughan & M. Thelwall, "Scholarly use on the Web: What are the key inducers of links to journal Web sites?" *Journal of the American Society for Information Science* 2003, 54(1): 29-38. Un preprint del secondo articolo è disponibile all'indirizzo:

<http://www.scit.wlv.ac.uk/~cm1993/papers/2003_what_are_the_key_inducers_of_links_to_journal_web_sites.pdf>. Una correlazione tra le diverse misure di impatto è stata trovata anche confrontando, per gli articoli di un set di riviste in ambito LIS, le citazioni desunte dai repertori ISI con quelle trovate, utilizzando Google come motore di ricerca, nelle sezioni bibliografiche di saggi e articoli presenti sul Web: L. Vaughan & D. Shaw, "Bibliographic and Web citations: what is the difference?" *Journal of the American Society for Information Science and Technology* 2003, 54(14): 1313-22.

ed a tutti quelli che lo citano. Immaginate che questo database sia disponibile gratis su Internet e venga aggiornato quotidianamente con tutti i nuovi documenti pubblicati [...] Un database del genere cambierebbe sostanzialmente il modo in cui gli studiosi localizzano e si mantengono aggiornati sui lavori altrui. A sua volta, ciò condizionerebbe il modo in cui essi pubblicano i loro lavori alla luce di una maggiore visibilità della ricerca, oltre i confini geografici, e di una maggiore capacità di dimostrare la qualità dei contributi attraverso l'analisi delle citazioni. In breve, un database universale di citazioni sarebbe un importante catalizzatore nella riforma della comunicazione scientifica».²⁵⁹

Quando, nel febbraio 1997, Robert Cameron lanciò la sua proposta di una rete universale di documenti collegati attraverso citazioni bibliografiche, il mondo della ricerca era pronto a sperimentare i vantaggi della libera circolazione dell'informazione scientifica: diverse riviste *peer reviewed* e archivi disciplinari erano da tempo gratuitamente accessibili su Internet; nel maggio dello stesso anno nasceva la rete informativa *RePEc - Research Papers in Economics*; il 26 giugno la National Library of Medicine di Bethesda annunciava la disponibilità gratuita del database bibliografico *Medline* attraverso il sito *PubMed*, ed il 19 agosto Stevan Harnad, che tre anni prima aveva proposto per la prima volta l'idea dell'auto-archiviazione delle pubblicazioni scientifiche, lanciava il server di *CogPrints* per le scienze cognitive.

La rivoluzione scatenata da Internet nelle forme di organizzazione cooperativa della ricerca e nella diffusione dei suoi risultati si stava materializzando nella creazione di canali di circolazione della letteratura scientifica alternativi rispetto alla rivista commerciale *peer reviewed*. Il movimento Open Access (OA), decollato tra la fine degli anni '90 e i primi del 2000, ha elaborato da allora una serie di modelli teorici ed operativi a sostegno della libera

²⁵⁹ R.D. Cameron, "A universal citation database as a catalyst for reform in scholarly communication" *First*

diffusione delle pubblicazioni all'interno di *OA Journals* o di archivi istituzionali e disciplinari di e-print.²⁶⁰ Da qualche anno, un obiettivo strategico del movimento è dimostrare che l'impatto citazionale di un lavoro aumenta se l'autore lo rende disponibile sul Web nel più breve tempo possibile ed in forma gratuita. Dopo la pubblicazione, nel 2001, di un supercitato articolo di Steve Lawrence nel quale si dimostrava che la disponibilità online gratuita delle relazioni dei convegni di informatica ne incrementa la probabilità di citazione,²⁶¹ la correlazione tra impatto citazionale e OA è diventata un tema centrale di discussione, al punto da incoraggiare la compilazione e l'aggiornamento di un'apposita bibliografia online nell'ambito di *OpCit Project*.²⁶² Garfield ha replicato che le cifre di Lawrence non sono generalizzabili perché la letteratura congressuale è un mezzo di comunicazione essenziale per l'informatica, ma non per altre discipline. Inoltre, l'aumento del numero potenziale di lettori causato dall'accesso libero non implica necessariamente un maggiore impatto citazionale perché, in fin dei conti, tocca sempre alla popolazione dei ricercatori di punta («relevant researchers») nelle singole discipline giudicare dell'utilità e citabilità di un documento.²⁶³

Intanto Stevan Harnad e Tim Brody hanno inaugurato un progetto di ricerca mirato ad accertare, su un campione di 14 milioni di articoli indicizzati nei repertori dell'ISI, i vantaggi del modello OA. La loro metodologia è semplice:

Monday 1997, 2(4) <http://www.firstmonday.dk/issues/issue2_4/cameron/index.html>.

²⁶⁰ La letteratura sul movimento OA è in costante aumento, come documenta la pubblicazione di una recente bibliografia tematica: C.W. Bailey, *Open Access Bibliography* (Washington, DC: Association of Research Library, 2005) <<http://info.lib.uh.edu/cwb/oab.pdf>>. Sul tema del rapporto tra OA e controllo di qualità della ricerca vedi: A. De Robbio, "Auto-archiviazione per la ricerca: problemi aperti e sviluppi futuri" *Bibliotime* 2003, 4(3) <<http://eprints.rclis.org/archive/00000180/03/OAI-20maggio2003.pdf>>; A.M. Tamaro, *Qualità delle pubblicazioni scientifiche ed open access*, 2004 <<http://hdl.handle.net/1889/56>>.

²⁶¹ S. Lawrence, "Free online availability substantially increases a paper's impact" *Nature* 2001, 411(6837): 521. L'articolo è più noto col titolo del preprint "Online or invisible?" accessibile all'URL: <<http://citeseer.ist.psu.edu/online-nature01/>>.

²⁶² OpCit Project, *The effect of open access and downloads ('hits') on citation impact: a bibliography of studies*, 2005 <<http://opcit.eprints.org/oacitation-biblio.html>>.

²⁶³ «Undoubtedly the web will increase apparent readership of literature, but that will not necessarily change the population of relevant researchers who are in a position to cite particular studies»: E. Garfield, "Re: How to compare research impact of toll- vs. open-access research" *American Scientist Open Access Forum* (Apr 15, 2004) <<http://www.ecs.soton.ac.uk/~harnad/Hypermail/Amsci/3708.html>>.

comparare le citazioni individuali ricevute da articoli OA con quelle di articoli non OA apparsi sulla stessa rivista commerciale. I risultati parziali, per ora limitati alla fisica, danno ragione a Lawrence: gli articoli auto-archiviati dagli autori e resi gratuitamente disponibili su e-print server hanno un impatto più alto di quelli venduti.²⁶⁴ Una conclusione analoga si ricava da un test di Kristin Antelman su un campione di riviste ISI di scienze politiche, filosofia, matematica e ingegneria.²⁶⁵

L'ISI non è rimasta indifferente a questi sviluppi. Nonostante la sua definizione di OA sia strettamente commerciale, contemplando quasi esclusivamente riviste di consolidata forza editoriale «il cui contenuto più recente sia accessibile gratis online»,²⁶⁶ tra aprile e ottobre 2004 hanno visto la luce due studi sull'impatto citazionale dei periodici ad accesso libero indicizzati in *ISI Web of Science*. Il primo, concentrato su 148 riviste di scienze naturali incluse nel *Journal Citation Reports 2002*, mostra che, tranne alcune notevoli eccezioni, l'*impact factor* dei titoli OA è variabile a seconda delle categorie, complessivamente però inferiore a quello di riviste distribuite attraverso i canali convenzionali, mentre il principale beneficio che ci si aspetterebbe dalla rapida disponibilità online, cioè la tendenza ad essere citati più velocemente, è minimo, quasi nullo: gli scienziati, in breve, non usano gli articoli OA prima di quelli non OA, a dispetto della loro immediata disponibilità.²⁶⁷ Il secondo studio, condotto su 219 periodici OA inclusi nel

²⁶⁴ S. Harnad & T. Brody, "Comparing the impact of open access (oa) vs. non-oa articles in the same journals" *D-Lib Magazine* 2004, 10(6) <<http://www.dlib.org/dlib/june04/harnad/06harnad.html>>.

²⁶⁵ K. Antelman, "Do Open Access articles have a greater research impact?" *College & Research Libraries* 2004, 65(5): 372-82 <http://eprints.rclis.org/archive/00002309/01/do_open_access_CRL.pdf>.

²⁶⁶ «...the most rigorous definition of OA - that the most recent content is freely available online»: J. Pringle, "Thomson Scientific finds new opportunities in Open Access" *KnowledgeLink Newsletter from Thomson Scientific* 2004 <<http://www.thomsonscientific.com/media/newsletterpdfs/2004-11/open-access.pdf>>.

²⁶⁷ «Open Access journals certainly have the potential to garner an even greater readership. Except where technology does not permit, anyone in the world will have access to them. Yet how does this additional exposure affect the importance and influence of the journal, as measured by citations? To date, no clear effect has been observed. Though there is some suggestion in aggregate of a slightly more rapid accumulation of citations, this effect is, so far, minimal. The wide distribution of these OA journals has not yet been shown to have any appreciable effect on their appearance in lists of cited references in other journals»: J. Testa & M.E. McVeigh, *The impact of open access journals. A citation study from Thomson ISI*, 2004 <<http://www.isinet.com/media/presentrep/acropdf/impact-oa-journals.pdf>>, p. 10.

Journal Citation Reports 2003, rettifica in parte le conclusioni precedenti: un numero significativo di riviste OA di fisica, ingegneria, matematica ha un *impact factor* vicino al top delle rispettive categorie; nelle stesse discipline inoltre, come pure in medicina, «gli articoli recenti nei periodici OA ricevono una percentuale complessiva di citazioni più alta rispetto agli articoli recenti nei periodici ad accesso tradizionale [...] questo effetto si estende agli articoli dei due o tre anni precedenti».²⁶⁸

L'ISI, stando alle dichiarazioni dei suoi portavoce, continuerà a monitorare il fenomeno OA e ad includere nei *citation indexes* le riviste che, a prescindere dal modello di distribuzione, ottemperano ai rigidi criteri di selezione citazionale imposta alle proprie creature dagli anni '60 in poi. L'azienda non considera il modello OA una minaccia alla propria missione ed il motivo è facile da intuire: fino a quando il nucleo della letteratura che conta continuerà ad addensarsi attorno a poche riviste fondamentali ed i criteri di stratificazione sociale delle comunità scientifiche continueranno a mantenere uno stretto legame con il sistema editoriale di tipo commerciale, gli indici di citazioni hanno il posto assicurato. «L'*Open Access* - scrive Marie McVeigh per conto dell'ISI - sta cambiando il modo di distribuzione dei materiali elettronici. Ma la modalità di accesso incide solo sul percorso che il ricercatore compie per acquisire la letteratura, non altera la natura della sua interazione con quei materiali una volta localizzati. La letteratura che un autore reputa più significativa, scientificamente consistente e rilevante per il suo lavoro, non viene solo acquisita, viene letta e assimilata. Un sottoinsieme degli articoli ritenuti validi esercita poi una certa influenza sul percorso di ricerca. Infine, gli articoli più influenti e rilevanti vengono citati».²⁶⁹ Nonostante le battaglie

²⁶⁸ «In Physics, Engineering & Mathematics, and in Medicine, recent articles in OA journals receive a higher percentage of total citations than recent articles in traditional access journals. Interestingly, this effect extends through articles that are two or three years old, effectively covering the period of analysis of both key journal performance metrics, Immediacy Index and Journal Impact Factor»: M.E. McVeigh, *Open Access journals in the ISI citation databases: Analysis of impact factors and citation patterns. A citation study from Thomson Scientific*, 2004 <<http://www.isinet.com/media/presentrep/essayspdf/openaccesscitations2.pdf>>, p. 9.

²⁶⁹ *Ibid.*, p. 16.

di Stevan Harnad e Paul Ginsparg, il modello OA non ha prodotto sinora un sistema di controllo della qualità capace di rivaleggiare con la *peer review* nella creazione di un set di *core journals* ad accesso libero. I bibliometri ammettono senza problemi che l'accesso gratuito in tempo reale e le pratiche di *open peer commentary* potenziano la comunicazione scientifica e ne accentuano il carattere collaborativo, ma riconoscono anche che la reputazione dei ricercatori resta per ora saldamente legata allo status delle riviste tradizionali. «Quasi niente nell'impresa scientifica - afferma Anthony van Raan - può competere con l'importanza della pubblicazione nei *top-journals*».²⁷⁰

Intanto la mole crescente di documentazione archiviata nei server OA pone problemi tecnici di controllo bibliografico e verifica della qualità. Per fronteggiarli servono software capaci di fare, velocemente e su scala ancora più grande, quello che l'ISI ha fatto per più di mezzo secolo con le riviste di carta, cioè localizzare i documenti rilevanti, archivarli, indicizzarne i riferimenti bibliografici e leggere il foglio alla rovescia per analizzare e contare le citazioni. A questo servono *CiteBase* e *CiteSeer*.

6.1.2. CiteBase, OACI

CiteBase,²⁷¹ annunciato nel dicembre 2001, è un sistema di indicizzazione automatica di un certo numero di archivi di e-print, tra cui *arXiv.org*, *PubMed Central*, *Biomedcentral* e *Cogprints*, allineati con la filosofia OA. Messo a punto da Tim Brody e dal suo team dell'Università di Southampton, progettato inizialmente come motore di ricerca dei metadati Dublin Core per i record di *arXiv.org*, esso rappresenta per certi versi il culmine della *Open Archives*

²⁷⁰ A.F.J. van Raan, "Bibliometrics and Internet: Some observations and expectations" *Scientometrics* 2001, 50(1): 59-63, p. 62.

²⁷¹ <<http://citebase.eprints.org/cgi-bin/search>>.

Initiative (1999), orientata alle soluzioni tecniche per l'interoperabilità degli archivi, e del progetto *OpCit - Open Citation Project* (1999-2002), che aveva come obiettivo quello di creare, all'interno dell'archivio di e-print per la fisica *arXiv.org*, un hyperlink tra riferimenti bibliografici e full-text dei documenti citati. Il software di *CiteBase* analizza i riferimenti bibliografici di tutte le pubblicazioni ospitate nelle banche dati da cui attinge e, ogniqualvolta il riferimento corrisponde al full text di un articolo presente negli stessi archivi, crea un link con esso. Navigando tra i link si possono fare una serie di operazioni familiari agli utenti degli indici di citazioni: 1) recuperare tutti gli articoli citati da un articolo di partenza oppure tutti gli articoli che citano un articolo di partenza; 2) classificare i record trovati nel corso di una ricerca in ordine decrescente di numero di citazioni ricevute e visualizzare i *Top 5 Articles* che citano una data pubblicazione; 3) recuperare tutti gli articoli co-citati con un dato articolo e i *Top 5 Articles Cocited*; 4) visualizzare un grafico dell'andamento delle citazioni e del numero di download degli articoli nel tempo. È possibile inoltre elaborare una tabella di correlazione tra numero di citazioni ricevute e numero approssimativo di download di un articolo. Un parametro da interpretare, secondo Tim Brody, come verifica rudimentale dell'ipotesi che l'uso di un articolo implichi la sua citazione e viceversa. Lo stesso Brody avverte comunque che questa misura va letta con grande cautela per due motivi: copre solo gli accessi sul mirror inglese di *arXiv.org* dal 1999 e, al pari di altre informazioni desunte dai web log, può essere falsata da diversi fattori.²⁷²

Sulla via indicata da *CiteBase*, un gruppo di lavoro internazionale, costituito nel 2004 presso il CERN di Ginevra, ha inaugurato il progetto *OACI - Open Archive Citation Index*, volto a creare un indice multidisciplinare

²⁷² «A new potential measure of on-line impact, not available in the on-paper era, is usage, in the form of "hits." This measure is noisy (it can be inflated by automated web-crawlers, short-changed by intermediate caches, abused by deliberate self-hits from authors, and it can be indiscriminating between nonspecific site-browsing and item-specific reading) yet it seems to have some signal-value too, partly correlated with and partly independent of citation impact»: <<http://citebase.eprints.org/help/coverage.php#citations>>.

automatizzato della letteratura scientifica ad accesso libero (e non solo) che si proietti oltre i limiti materiali e concettuali dei database dell'ISI per abbracciare documentazione di varia provenienza e tipologia, nell'ottica di una generale revisione del ruolo delle citazioni bibliografiche nella verifica della qualità scientifica.²⁷³

6.1.3. CiteSeer: Dallo Science Citation Index al Web Citation Index

CiteSeer (in precedenza si chiamava *ResearchIndex*) è un sistema automatico di gestione di una digital library messo a punto, alla fine degli anni '90, da un gruppo di ricercatori del NEC Research Institute di Princeton.²⁷⁴ Il sistema indicizza le pubblicazioni scientifiche di ambito accademico, non necessariamente ad accesso gratuito, disseminate nei nodi della rete Internet. La localizzazione dei documenti, che possono presentarsi anche in formati diversi da HTML, ad esempio Postscript e pdf, avviene sia mediante un software che sfrutta i motori di ricerca, l'euristica e il web crawling, sia con metodi empirici quali la perlustrazione di archivi già esistenti, l'accordo con gli editori o la trasmissione diretta da parte degli autori.²⁷⁵ Per ciascun documento un software di indicizzazione automatica localizza la sezione dei riferimenti bibliografici, estrae i record delle singole citazioni e identifica il testo che si riferisce ai titoli citati, cioè il contesto della citazione. Lo stesso sistema raggruppa e rende ricercabili, dopo averne normalizzato la forma, tutti i riferimenti bibliografici che puntano al medesimo documento generando, per le pubblicazioni di un dato autore, un grafico della distribuzione nel tempo

²⁷³ Ho trovato l'informazione sul progetto OACI in: P. Gargiulo, "La comunicazione scientifica aperta della Sapienza" *Workshop sugli Open Archives*, Università di Roma La Sapienza, 10 giugno 2004 <<http://bids.citicord.uniroma1.it/padis/gargiulo10giugno04.pps>>.

²⁷⁴ <<http://citeseer.ist.psu.edu/>>.

²⁷⁵ La presentazione di CiteSeer da parte dei suoi creatori si trova in: S. Lawrence; C.L. Giles; K. Bollacker, "Digital libraries and autonomous citation indexing" *IEEE Computer* 1999, 32(6): 67-71 <<http://citeseer.ist.psu.edu/aci-computer/aci-computer99.html>>.

delle citazioni ricevute. Si può intuire come un'architettura di questo tipo faciliti enormemente il confezionamento di valutazioni bibliometriche al volo, ad esempio il calcolo del numero di citazioni ricevute da un articolo, distinte dalle autocitazioni. Con *CiteSeer* inoltre si possono correlare documenti sulla base delle citazioni condivise e si può fare, in poco tempo e senza intervento manuale, quello che alcuni bibliometri hanno considerato uno dei possibili antidoti all'ambiguità delle pratiche citazionali: esaminare il contesto in cui un documento viene citato per chiarire la natura del riferimento, l'uso che l'autore ha inteso fare del testo citato.

Le affinità elettive tra il software di indicizzazione automatica delle citazioni del NEC e il progetto editoriale dell'ISI sono tali da far apparire le vicende successive una storia già scritta.

CiteSeer, disponibile online gratuitamente in versione dimostrativa, ha viaggiato da solo per qualche tempo raccogliendo consensi e sponsor prestigiosi a livello internazionale (National Science Foundation, Microsoft Research, NASA), fino a quando, il 25 febbraio 2004, Thomson ISI di Philadelphia e NEC Laboratories America di Princeton hanno annunciato un accordo commerciale per la creazione di un indice di citazioni multidisciplinare delle risorse accademiche accessibili via Web.²⁷⁶ Il nuovo *Web Citation Index*, annunciato per il 2005, sfrutterà la tecnologia *CiteSeer* ed affiancherà gli indici tradizionali all'interno della piattaforma *ISI Web of Knowledge*, coronando la strategia commerciale riservata dall'ISI al fenomeno OA: seguirne attentamente gli sviluppi, quando possibile inglobarne dei frammenti nella macchina citazionale, soprattutto assicurare ai propri clienti il pieno controllo bibliografico sulle sue diramazioni e, in prospettiva, fornire gli strumenti per valutarne la qualità in termini di *impact factor*. Dopotutto, recita

²⁷⁶ Thomson ISI, *Thomson ISI and NEC team up to Index Web-based scholarship*, 2004 <<http://www.isinet.com/press/2004/8217120>>. Nel frattempo Lee Giles, uno dei responsabili del progetto del NEC, è migrato al Penn State's Smeal College of Business, dove ha impiegato la tecnologia di *CiteSeer*, coperta per ora da una licenza non commerciale, per realizzare *SMEALSearch*

il manifesto della *BOAI*, ciò che resta del copyright nel mondo OA è pur sempre il diritto degli autori a tutelare l'integrità del proprio lavoro e «ad essere adeguatamente riconosciuti e citati».²⁷⁷

CiteSeer rappresenta probabilmente il culmine del progetto enciclopedico di Garfield: una tecnologia per il controllo totale sulla documentazione di qualità prodotta dai "collegi sempre meno invisibili" di scienziati che usano la rete Internet per divulgare le loro ricerche. L'ISI ha trasformato così il principale punto debole del movimento OA, cioè la difficoltà a creare interfacce unificate per interrogare gli archivi di e-print, in un punto di forza del proprio progetto editoriale: la realizzazione, in un contesto governato dalla logica qualitativa degli indici di citazioni, di un portale per la ricerca unificata di tutta la letteratura accademica mondiale, indipendentemente dal modo di distribuzione e commercializzazione. Un ricercatore è soddisfatto se scarica gratuitamente un articolo dalla rete, il che presuppone comunque che lo abbia trovato, ma è ancora più contento se, davanti ad un'offerta più ampia di quella cui era abituato in precedenza, riesce ad avere un qualche tipo di conferma del "peso" scientifico dell'articolo verificandone il pedigree citazionale. La gratuità online non aumenta di per sé l'impatto di un documento se un sistema efficiente di recupero dell'informazione non permette di trovarlo e al tempo stesso di trasmettergli quell'"aura" di autorevolezza che, dai *Current Contents* in poi, i database dell'ISI hanno depositato sulle poche riviste internazionali giudicate fondamentali. Prima di prendere in considerazione l'articolo scaricato, in poche parole, il nostro ricercatore deve trovarlo e per trovarlo ha bisogno di un punto di partenza solido che gli trasmetta anche un qualche tipo di conferma del fatto che vale la pena citarlo. Nel momento in cui se ne rende

(<<http://smealsearch2.psu.edu/index.html>>) ed *eBizSearch* (<<http://gunther.smeal.psu.edu/>>), due sistemi di indicizzazione automatica della letteratura scientifica di argomento commerciale disseminata nel Web.

²⁷⁷ «The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited»: *BOAI, Budapest Open Access Initiative, 2002* <<http://www.soros.org/openaccess/read.shtml>>.

conto, egli acconsente a giocare al gioco della citazione secondo le regole stabilite dall'ISI.

6.2. La citazione come hyperlink e le metriche del Web

Un sito o una pagina web che ospita un link verso un altro sito o un'altra pagina web è, dal punto di vista formale, identico ad un documento che ne cita un altro mediante un riferimento bibliografico: un link rappresenta, con buona approssimazione, la versione high-tech della citazione bibliografica.²⁷⁸

L'analogia è addirittura sbilanciata a favore del primo: un link ipertestuale infatti è "più citazione", cita "meglio" di quanto non faccia un riferimento bibliografico, nel senso che in un ipertesto la citazione non è una componente tipograficamente e concettualmente marginale del documento, bensì la norma principale che regola la costruzione e l'organizzazione del testo. L'analogia però ha anche un limite plateale: il link non paga debiti intellettuali. Persino in un sito accademico è raro trovare link che abbiano un significato bibliografico: si linka per facilitare la navigazione di un sito, per segnalare risorse correlate, per offrire un servizio, talvolta (specie nei siti personali) per nessun motivo.²⁷⁹

Nella seconda metà degli anni '90, si sono sviluppati diversi fronti di ricerca interessati allo studio della struttura del Web mediante tecniche quantitative e modelli matematici applicati agli hyperlink.²⁸⁰ Il Web è stato riconosciuto contemporaneamente come un sistema complesso e una chiave per decifrare la

²⁷⁸ «Linking on the web is essentially the high-tech digital version of the intellectual act of citing – except for the purely financially motivated and sociopath links»: P. Jacsó, "Savvy searching. Citation searching" *Online Information Review* 2004, 28(6): 454-60, p. 454.

²⁷⁹ M. Thelwall, "What is this link doing here? Beginning a finegrained process of identifying reasons for academic hyperlink creation" *Information Research* 2003, 8(3) <<http://informationr.net/ir/8-3/paper151.html>>.

²⁸⁰ Una rassegna aggiornata al 2003 dei diversi metodi analitici si trova in: H.W. Park & M. Thelwall, "Hyperlink analyses of the World Wide Web: A review" *Journal of Computer-Mediated Communication* 2003, 8(4) <<http://www.ascusc.org/jcmc/vol8/issue4/park.html>>.

complessità di altri sistemi naturali e sociali. La sua struttura, indagata con strumenti concettuali mutuati dalla teoria dei grafi, è stata assimilata a quella di un grafo orientato, nel quale i nodi (pagine web) sono le unità di base, gli archi che li congiungono (hyperlink) rappresentano le interazioni tra i nodi. Sviluppare un modello stocastico accurato e realistico del grafo del Web, misurarne l'ampiezza e rilevare il comportamento degli utenti che lo attraversano, sono divenuti obiettivi sensibili sia sul piano culturale che commerciale.

I motori di ricerca di ultima generazione, per esempio, sfruttano la correlazione tra il grafo del Web ed il significato del testo delle pagine e combinano l'analisi dei contenuti con l'analisi degli hyperlink per potenziare gli algoritmi di crawling e di ranking.²⁸¹ L'assunto è semplice: così come, nell'analisi classica delle citazioni, è probabile che un documento citato abbia un'affinità di contenuti col documento citante, è altrettanto verosimile che una pagina web linkata sia concettualmente affine alla pagina che effettua il link. Quindi si può ricavare il significato di una pagina guardando alle altre pagine che puntano ad essa e la topologia dei link ha una connessione diretta con tale significato: minore il numero di link che separano una pagina da un'altra considerata rilevante, maggiore la probabilità che anche la prima sia rilevante. L'applicazione più nota di questa regola è *PageRank*, l'algoritmo di ranking di Google, che ordina le pagine corrispondenti ai criteri di ricerca immessi dall'utente mediante uno schema che conferisce un peso maggiore alle pagine più linkate e, tra queste, privilegia le pagine linkate da siti a loro volta molto linkati: la qualità di una pagina risulta così dalla somma della qualità delle pagine che si collegano ad essa.²⁸² Prima di *PageRank*, il *Clever Project* dell'IBM, nato come estensione dell'algoritmo *HITS (Hypertext-Induced Topic*

²⁸¹ Sulla correlazione tra link ipertestuali e contenuto semantico e lessicale delle pagine web: F. Menczer, "Lexical and semantic clustering by web links" *Journal of the American Society for Information Science and Technology* 2004, 55(14): 1261-69.

²⁸² S. Brin & L. Page, "The anatomy of a large-scale hypertextual Web search engine" *Computer Networks and ISDN Systems* 1998, 30(1-7): 107-17.

Search) di Jon Kleinberg, applicava l'analisi dei link alle tecniche di retrieval per identificare, in relazione a determinati argomenti, le pagine autorevoli (*authorities*) e i connettori (*hubs*): una pagina autorevole è quella che riceve molti link da hub di qualità, una specie di *citation classic* virtuale; inversamente un hub di qualità è una pagina che contiene molti link verso pagine autorevoli.²⁸³

La ricerca informetrica sugli hyperlink ha preso tre direzioni principali, i cui domini si intersecano in diversi punti: 1) la webmetrica (*webometrics*), nata nell'alveo della scienza dell'informazione, che indaga la struttura e le proprietà del Web con i metodi della bibliometria; 2) l'analisi delle reti di hyperlink (*hyperlink network analysis*), discendente diretta dall'analisi delle reti sociali (*social network analysis*), che interpreta le connessioni tra siti web come simboli tecnologici dei legami sociali tra individui, gruppi, organizzazioni, nazioni; 3) lo studio, sviluppato nell'ambito della fisica statistica, di Internet e del Web come modelli di reti complesse (*complex networks*).

Solo la prima area ha una parentela esplicita con la tradizione bibliometrica di analisi delle citazioni, ma la sovrapposizione con le altre due è evidente nel comune interesse per la nozione di rete, che da Price in poi ha sfidato i bibliometri a studiare la trama delle connessioni tra documenti e individui ad un livello di astrazione più alto rispetto a quello delle singole unità.

6.2.1. Webmetrica

Quali tecniche si possono usare per conoscere l'impatto di un'area del Web? In che misura l'impatto, tradizionalmente vincolato all'assimilazione di un testo strutturato (l'articolo scientifico) da parte di una comunità identificabile

²⁸³ S. Chakrabarti; B. Dom; D. Gibson et al., "Hypersearching the web" *Scientific American* 1999(June): 54-

di persone (gli specialisti di una disciplina), può essere calcolato, in un ambiente così anarchico come la rete Internet, per un oggetto dai contorni così sfumati come una risorsa web? E in che misura l'impatto cybermetrico si riflette nella trama degli hyperlink? Domande familiari ai frequentatori della letteratura bibliometrica, ma che ora assillano anche i geometri della rete.²⁸⁴ La webmetrica è nata per certi versi sulle spalle di un gigante - la tradizione quasi cinquantennale di studi scientometrici basati sul conteggio e l'analisi delle citazioni - con l'ambizione di guardare più lontano del gigante per capire in che misura sia possibile estrarre informazioni utili allo studio della comunicazione scientifica da un oggetto ambiguo e multiforme come un hyperlink. La strada era in gran parte tracciata, bastava sostituire le unità informative dei bibliometri - l'articolo, il libro, la citazione - con pagine, siti web e link dalla provenienza controllata, essenzialmente pagine e siti di università, enti di ricerca, singoli scienziati. Un link in uscita da una pagina web (*outlink*) è formalmente simile ad un riferimento bibliografico; un link che punta ad essa dall'esterno (*inlink*) equivale ad una citazione; due link che puntano simultaneamente allo stesso sito (*co-linking*) equivalgono ad una co-citazione; un link per la navigazione interna di un sito o per rinviare alle diverse sezioni di una pagina (*selflink*) è simile ad un'autocitazione; la somma dei link indirizzati verso un sito è un grezzo equivalente cybermetrico dell'*impact factor* di Garfield. L'estensione di metodi informetrici al WWW - hanno scritto Almind e Ingwersen - «permette all'analisi di svolgersi quasi nello stesso modo di quella tradizionale dei database di citazioni»,²⁸⁵ con il

60

²⁸⁴ Due rassegne approfondite degli studi webmetrici pubblicati dalla seconda metà degli anni '90 in poi sono: M. Thelwall; L. Vaughan; L. Björneborn, "Webometrics" *Annual Review of Information Science and Technology* 2005, 39: 81-135; C.L. Borgman & J. Furner, "Scholarly communication and bibliometrics," in *Annual Review of Information Science and Technology*, edited by B. Cronin (Medford, NJ: Information Today, 2002). Un preprint del lavoro di Borgman e Furner è gratuitamente accessibile all'indirizzo: <<http://polaris.gseis.ucla.edu/jfurner/arist02.pdf>>.

²⁸⁵ T. Almind & P. Ingwersen, "Informetric analysis of the World Wide Web: Methodological approaches to "Webometrics"" *Journal of Documentation* 1997, 53: 404-26, p. 404. Un approccio analogo a quello di Almind e Ingwersen è stato proposto, nello stesso periodo, in: J.M. Rodríguez Gairín, "Valorando el impacto

vantaggio che la rete di citazioni contiene ora anche i full text, quindi è possibile risalire immediatamente dal riferimento bibliografico al testo/contesto che ne illustra il significato per l'autore.

Nella forma più rudimentale, la misura della visibilità di un sito o la classificazione dei siti di un'area del Web in base al numero dei link ricevuti è stata effettuata, in piena sintonia con il *JCR* dell'ISI, partendo da liste di hyperlink compilate manualmente o selezionate in funzione della reputazione dei siti di partenza.²⁸⁶ Un'iniziativa più ambiziosa ha condotto invece ad elaborare un indice numerico, una specie di equivalente cybermetrico dell'*IF* di Garfield, utile per la valutazione comparativa dei siti web. Il calcolo del *Web Impact Factor (WIF)* sfrutta il potenziale conoscitivo dei motori di ricerca come AltaVista e AlltheWeb, i quali consentono, mediante il comando *link:*, di richiamare e quantificare i link diretti verso un sito di partenza. In una delle formulazioni possibili, il *WIF* di un sito è calcolato come rapporto tra il numero di pagine web che dall'esterno puntano un link verso il sito (esclusi quindi i rinvii interni tra le pagine) e il numero di pagine web di cui esso è composto.²⁸⁷ Le difficoltà concettuali e metodologiche legate a questa misura sono tuttavia ancora maggiori di quelle dell'*IF* di Garfield. La più seria riguarda l'incapacità dei motori di ricerca generalisti di restituire un quadro esauriente ed accurato degli hyperlink diretti verso un'area specifica del Web. Per questo motivo alcuni autori hanno insistito sulla necessità di costruire dei robot specializzati, un espediente già sperimentato, ad esempio, per il calcolo

de la información en Internet: Alavista, el 'Citation Index' de la Red" *Revista Española de Documentación Científica* 1997, 20(2): 175-81.

²⁸⁶ Cui ad esempio utilizza i link presenti sulle home page di 25 top medical schools degli Stati Uniti per stilare una classifica dei siti di medicina più citati: L. Cui, "Rating health web sites using the principles of citation analysis: a bibliometric approach" *Journal of Medical Internet Research* 1999, 1(1): E4 <<http://www.jmir.org/1999/1/e4/>>.

²⁸⁷ P. Ingwersen, "The calculation of Web impact factors" *Journal of Documentation* 1998, 54(2): 236-43, p. 237. Sulle diverse formulazioni del *WIF* vedi: X. Li, "A review of the development and application of the Web impact factor" *Online Information Review* 2003, 27(6): 407-17, pp. 413-414.

dell'impatto di un certo numero di siti accademici britannici.²⁸⁸ In secondo luogo, sia il numeratore che il denominatore del rapporto sono difficili da definire con precisione: non sempre è chiaro cosa debba essere inteso come link esterno, soprattutto nel caso di siti diversi facenti capo ad una medesima istituzione (il link dal sito di un'università a quello di un dipartimento affiliato è esterno o interno?); analogamente le pagine web sommate al denominatore possono includere un alto numero di unità non pertinenti (le migliaia di pagine di statistiche dei web log); inoltre le pagine web non hanno un formato univoco, una pagina HTML suddivisa in frame può includere numerosi file ed uno stesso documento, per esempio un libro online, può presentarsi alternativamente come un'unica pagina o frammentato in 100 pagine per comodità di lettura, quindi può pesare nel calcolo 1 o 100. Ecco perché, in una delle varianti del *WIF* applicata alla valutazione delle risorse accademiche, il numero delle pagine web al denominatore è sostituito da un valore più stabile: il numero dei membri dell'istituzione responsabile del sito.²⁸⁹

La pagina web non è una buona unità di analisi anche per un altro motivo: nell'*IF* tradizionale, se una rivista **A** cita due articoli di una rivista **B**, il fattore d'impatto della rivista **B** aumenta di 2 unità, mentre nel *WIF*, se una pagina web esterna **A** contiene due link a pagine diverse del sito **Y**, il fattore d'impatto aumenta di una sola unità. Per ovviare a questo inconveniente sono stati sperimentati dei modelli alternativi di selezione del campione di partenza (*ADM* o *Alternative Document Models*), nei quali si filtra il numero e la qualità dei documenti o dei link esterni da conteggiare scegliendo di volta in volta un diverso livello di aggregazione (la pagina, la directory, il dominio, il sito).²⁹⁰

²⁸⁸ J. Bar-Ilan, "Data collection methods on the Web for informetric purposes: a review and analysis" *Scientometrics* 2001, 50(1): 7-32. Un esperimento di *academic Web crawler* si trova in M. Thelwall, "A Web crawler design for data mining" *Journal of Information Science* 2001, 27(5): 319-25.

²⁸⁹ M. Thelwall, "Extracting macroscopic information from web links" *Journal of the American Society for Information Science and Technology* 2001, 52(13): 1157-68.

²⁹⁰ M. Thelwall, "Conceptualising documentation on the Web: an evaluation of different heuristic-based models for counting links between university Web sites" *Journal of the American Society for Information*

In analogia con gli indicatori basati sul conteggio delle citazioni, la scala appropriata per dare un significato qualitativo ai conteggi di link è stata individuata al livello dell'istituzione, cioè dei siti delle singole università. Si è cercato così di trovare una correlazione tra il numero dei link ottenuti da un sito e misure indipendenti di performance accademica.²⁹¹ I risultati ottenuti rappresentano un terreno di confronto e riflessione per stimare la reputazione e la visibilità di una risorsa web istituzionale, ma non hanno in alcun modo un significato paragonabile a quello dell'impatto citazionale degli articoli scientifici: le pagine web universitarie sono in massima parte prive di contenuti scientifici, le loro finalità informative sono di tipo generale e le motivazioni per inglobarle in un link differiscono radicalmente da quelle per cui si cita un articolo.²⁹²

La webmetrica ha provato a ripercorrere le stesse strade della bibliometria e dell'analisi classica delle citazioni anche in altre direzioni: 1) lo studio del contenuto e del contesto delle pagine in cui i nomi di singoli studiosi vengono “invocati” sul Web, da interpretare come un indice parziale della loro notorietà, oppure come un modo per far emergere percorsi comunicativi informali normalmente trascurati;²⁹³ 2) lo studio di cluster di documenti co-citati, formati da coppie di pagine linkate contemporaneamente da una terza pagina, con l'obiettivo di visualizzare la rete di correlazioni tra istituzioni

Science and Technology 2002, 53(12): 995-1005. In un'elaborazione successiva, gli stessi modelli sono stati perfezionati facendo in modo che i link multipli indirizzati ad una data risorsa da pagine diverse dello stesso sito universitario contassero come un singolo link: M. Thelwall & D. Wilkinson, “Three target document range metrics for university web sites” *Journal of the American Society for Information Science and Technology* 2003, 54(6): 489-96.

²⁹¹ A.G. Smith & M. Thelwall, “Web impact factors for Australasian universities” *Scientometrics* 2002, 54(3): 363-80; M. Thelwall & R. Tang, “Disciplinary and linguistic considerations for academic Web linking: An exploratory hyperlink mediated study with Mainland China and Taiwan” *Scientometrics* 2003, 58(1): 153-79.

²⁹² D. Wilkinson; G. Harries; M. Thelwall et al., “Motivations for academic web site interlinking: evidence for the Web as a novel source of information on informal scholarly communication” *Journal of Information Science* 2003, 29(1): 59-66.

²⁹³ B. Cronin; H. Snyder; H. Rosenbaum et al., “Invoked on the Web” *Journal of the American Society for Information Science* 1998, 49(14): 1319-28 e J. Bar-Ilan, “The mathematician, Paul Erdos (1913-1996) in the eyes of the Internet” *Scientometrics* 1998, 43(2): 257-67.

universitarie,²⁹⁴ di estrarre gruppi di *core pages* di interesse disciplinare,²⁹⁵ di identificare *web communities* aggregate attorno ad un tema specifico,²⁹⁶ di perfezionare gli algoritmi per la ricerca di pagine correlate ad una pagina di partenza,²⁹⁷ di sperimentare sistemi semi-automatici di propagazione dei metadati per l'indicizzazione delle pagine web;²⁹⁸ 3) l'analisi delle coppie di parole chiave adoperate dagli utenti dei motori di ricerca, utile a potenziare gli algoritmi di ranking,²⁹⁹ e l'analisi delle co-occorrenze di termini in domini nazionali e di primo livello, finalizzata ad osservare, nello specchio deformato della rete Internet, le molteplici connessioni tra governi, università, industrie.³⁰⁰

Nonostante gli entusiasmi iniziali per l'estensione delle metriche tradizionali alle reti di ipertesti, gli informetrici sanno bene che l'equivalenza link/citazione su cui riposano le fondamenta della loro indagine è debole e presenta diverse limitazioni.³⁰¹ Innanzitutto la materia prima, la pagina web, non ha una struttura formale paragonabile a quella dell'articolo scientifico: spesso manca l'autore e, non esistendo un criterio univoco di applicazione, l'uso dei codici, in particolare dei tag HTML, è soggettivo e può variare da pagina a pagina. La variabilità di formati è un riflesso diretto della libertà

²⁹⁴ X. Polanco; M.A. Boudourides; D. Besagni et al., *Clustering and mapping web sites. For displaying implicit associations and visualizing networks.*, 2001

<http://www.math.upatras.gr/~mboudour/articles/web_clustering&mapping.pdf20>.

²⁹⁵ R.R. Larson, "Bibliometrics of the World Wide Web: an exploratory analysis of the intellectual structure of cyberspace," in *Global Complexity: Information, Chaos and Control. Proceedings of the 59th ASIS Annual Meeting, Baltimore, MD, October 21-24 1996*, edited by S. Hardin (Medford, NJ: Information Today Inc, 1996).

²⁹⁶ R. Kumar; P. Raghauam; S. Rajagopalan et al., "Trawling the Web for emerging cyber-communities" *Computer Networks* 1999, 31(11): 1481-93 <<http://www8.org/w8-papers/4a-search-mining/trauling/trauling.html>>.

²⁹⁷ J. Dean & M.R. Henzinger, "Finding related pages in the World Wide Web," in *Proceedings of the 8th International World Wide Web Conference. May 1999* (1999): 389-401.

²⁹⁸ C. Prime-Claverie & M. Beigbeder, "Transposition of the cocitation method with a view to classifying web pages" *Journal of the American Society for Information Science and Technology* 2004, 55(14): 1282-89.

²⁹⁹ N.C.M. Ross & D. Wolfram, "End user searching on the Internet: An analysis of term pair topics submitted to the Excite search engine" *Journal of the American Society for Information Science* 2000, 51(10): 949-58.

³⁰⁰ L. Leydesdorff & M. Curran, "Mapping university-industry-government relations on the Internet: the construction of indicators for a knowledge-based economy" *Cybermetrics* 2000, 4(1, paper 2) <<http://www.cindoc.csic.es/cybermetrics/articles/v4i1p2.html>>.

³⁰¹ Vedi ad esempio: L. Egghe, "New informetric aspects of the Internet: some reflections - many problems" *Journal of Information Science* 2000, 26(5): 329-35.

espressiva al livello dei contenuti: fatta eccezione per il sottoinsieme delle riviste e degli archivi di e-print allineati con gli schemi tradizionali della *peer review*, la porzione rimanente del cyberspazio è popolata da isole di documentazione in gran parte sottratte ad ogni tipo di controllo di qualità, nelle quali chiunque può pubblicare (e linkare) qualunque cosa. Inoltre, a differenza degli indici di citazioni, il Web è un oggetto in costante movimento, un fiume nel quale è impossibile bagnarsi due volte nelle stesse acque, con pagine che quotidianamente scompaiono, cambiano nome, indirizzo, struttura, contenuti, quindi può essere studiato solo in modo asincrono, su un campione temporalmente e qualitativamente sfasato rispetto all'originale. Un aspetto significativo di tale fluidità coinvolge la dimensione temporale della scrittura e della composizione dei testi: nei periodici tradizionali, un documento **A** che cita un documento **B** è stato scritto quasi sempre dopo di esso, **B** non poteva prevedere l'avvento di **A** e non poteva a sua volta esserne influenzato, tranne ovviamente nel caso di ritardi editoriali o nel caso in cui i due autori, lavorando nella stessa area e appartenendo al medesimo “collegio invisibile”, fossero a conoscenza di versioni preliminari dei rispettivi contributi; nel cyberspazio invece sono frequenti link reciproci tra pagine web scritte o aggiornate in segmenti temporali diversi, per cui il documento **B** linkato da **A** può a sua volta adattarsi al discendente accomodandone il link nella propria struttura.

Variabilità di stili e contenuti, instabilità, carenza di filtri qualitativi implicano che la raccolta dei dati di partenza è il nodo cruciale dell'impresa: nella webmetrica, a differenza delle altre due aree di ricerca delineate nei paragrafi seguenti, lo sviluppo di nuove ipotesi e strumenti analitici è inscindibile dalla riflessione sui limiti di affidabilità e rappresentatività dei dati raccolti e sulla validità delle tecniche usate per interrogarli. Nelle analisi basate sui *citation indexes* tradizionali i dati - la totalità dei riferimenti bibliografici relativi agli articoli delle riviste indicizzate dall'ISI - sono già confezionati e pronti per

l'uso. Per quanto se ne possa contestare il criterio iniziale di selezione, essi offrono un punto di partenza solido, identico per tutti coloro che vi si applicano con il proprio arsenale di tecniche bibliometriche. La webmetrica invece lavora su dati derivati da operazioni ripetute di *re-engineering*, dati costruiti, filtrati e ripuliti per l'occasione dal singolo ricercatore. Non esiste un metodo per analizzare tutta la popolazione, la totalità del Web, nè per estrarre un campione di studio realmente casuale. Le ricerche quantitative su argomenti specifici devono allora ricorrere all'unico espediente lecito, cioè l'interrogazione di uno o più motori di ricerca. Ma è una strategia inaffidabile sotto tanti punti di vista e continuerà ad esserlo almeno fino a quando il "Web semantico" di Tim Berners Lee non si sarà tradotto in applicazioni su larga scala, rendendo più espliciti e trasparenti al trattamento automatico i contenuti e le connessioni tra documenti. I motori di ricerca hanno infatti una copertura limitata dal punto di vista tematico, geografico e linguistico ed i loro algoritmi di crawling e ranking, spesso oscuri e proprietari, rendono incerta la composizione del campione di partenza. Sono inoltre tecnicamente incapaci a rendere conto del "Web invisibile", una terra di nessuno sconfinata dove abitano le pagine soggette ai criteri di esclusione dei crawler, le pagine schermate da permessi di accesso ed i milioni di pagine generate dinamicamente dai server ogniqualevolta rispondono all'interrogazione di un database. Infine la possibilità di costruire serie temporali, uno dei punti di forza degli indici di citazioni, è limitata quasi per intero ai dati ricavabili da *The Internet Archive Wayback Machine*,³⁰² che archivia una copia di ogni versione delle pagine web indicizzate. Questa barriera imposta ad ogni archeologia del Web è il sintomo più evidente del fatto che la citazione ipertestuale vive, per lo più, una vita fatta di eterno presente.

³⁰² <<http://www.archive.org/>>.

6.2.2. Analisi delle reti di hyperlink (Hyperlink Network Analysis)

Un navigatore che, su *Amazon.com*, compra un libro, poi scopre che altri acquirenti hanno fatto la stessa scelta ed hanno comprato anche "questo" o "quell'altro" libro, intrattiene con gli altri acquirenti una forma peculiare di relazione sociale mediata dagli hyperlink.³⁰³ Un webmaster che decide di linkare un sito di e-commerce comunica all'esterno una certa convinzione circa la sua affidabilità, è lecito perciò considerare il numero di link destinati a quel sito come un indicatore sociale del grado di fiducia accordato alla ditta che ne è responsabile.³⁰⁴ I siti di organizzazioni collegate ad un movimento sociale o coinvolte nel dibattito su un determinato argomento effettuano una selezione accurata delle risorse web in funzione delle tesi e dei principi rivendicati. Così facendo creano, attraverso la rete dei link preferenziali, una rappresentazione simbolica della struttura del movimento o delle alleanze strategiche tra i partecipanti al dibattito.³⁰⁵ I siti che distribuiscono link verso obiettivi nazionali o internazionali lo fanno secondo criteri variabili: seguendo la direzione geografica delle connessioni in un campione abbastanza ampio, si può vedere se il divario tra centro e periferia del mondo ipotizzato dalla *World*

³⁰³ V. Krebs, "Working in the connected world book network" *IHRIM - International Association for Human Resource Information Management Journal* 2000, 4(1): 87-90.

³⁰⁴ E. Davenport & B. Cronin, "The citation network as a prototype for representing trust in virtual environments," in *The Web of Knowledge: a Festschrift in Honor of Eugene Garfield*, edited by B. Cronin, et al. (Medford, NJ: Information Today Inc, 2000): 517-34; J.W. Palmer; J.P. Bailey; S. Faraj, "The role of intermediaries in the development of trust on the WWW: the use and prominence of trusted third parties and privacy statements" *Journal of Computer-Mediated Communication* 2000, 5(3)
<<http://www.ascusc.org/jcmc/vol5/issue3/palmer.html>>.

³⁰⁵ Halavais e Garrido, per esempio, hanno mappato le reti di hyperlink tra i siti di supporto del movimento zapatista: A. Halavais & M. Garrido, "Mapping networks of support for the Zapatista movement," in *Cyberactivism: Online activism in theory and practice*, edited by M. McCaughy, et al. (London: Routledge, 2003). Rogers e Marres hanno fatto la stessa cosa per i siti istituzionali e aziendali coinvolti nel dibattito sui cambiamenti climatici R. Rogers & N. Marres, "Landscaping climate change: a mapping technique for understanding science and technology debates on the World Wide Web" *Public Understanding of Science* 2000, 9: 141-63. Adamic ha studiato la distribuzione dei link nei siti correlati a tematiche quali l'aborto e gli UFO: L. Adamic, *The small world web. Proceedings of the 3rd European Conference of Research and Advanced Technology for Digital Libraries, ECDL, 1999*
<<http://www.hpl.hp.com/research/idl/papers/smallworld/smallworldpaper.html>>. Bae e Choi hanno cercato di catturare la struttura delle reti di link tra i siti di 402 organizzazioni non governative per la difesa dei diritti umani: S. Bae & J.H. Choi, "Cyberlinks between human rights NGOs: A network analysis" *58th Annual National Meeting of the Midwest Political Science Association*, Chicago, 2000, April.

System Theory trova espressione anche nella gerarchia dei link.³⁰⁶ I siti delle università che contengono molti link ai siti di altre università e istituzioni di ricerca creano una rete informale di relazioni: si può ipotizzare che l'orientamento dei link rifletta le interazioni socio-cognitive tra gli scienziati che lavorano in quelle università.³⁰⁷

Tutti questi esempi, riferiti a progetti di ricerca reali, mostrano che il grafo del Web non è solo un'astrazione matematica. Ogni giorno milioni di utenti nel mondo, dei quali solo una piccolissima percentuale ha interessi in qualche modo riconducibili alla ricerca scientifica, lo attraversano materialmente per acquistare beni, per svago, per cultura personale, per il puro piacere della navigazione, ed il tipo di incontri che essi fanno è condizionato dal labirinto delle strade prefabbricate che collegano i nodi. Le strade, i link, sono costruzioni tecnologiche, ma nel momento in cui qualcuno le attraversa acquistano un altro significato. Una connessione tra due computer infatti è anche, necessariamente, un legame di una qualche natura tra i rispettivi utilizzatori: «quando una rete di computer connette persone o organizzazioni, è una rete sociale».³⁰⁸ Un link tra siti web può essere interpretato allora come un canale comunicativo tra i rispettivi autori, il segno di una forma peculiare di relazione sociale, nella quale due nodi della rete - due individui o due organizzazioni - attivano uno scambio di beni tangibili (beni commerciali) o immateriali (informazione) in uno spazio geografico decentrato (le pagine di un sito web possono risiedere su migliaia di server disseminati in tutto il mondo).

³⁰⁶ A. Halavais, "National borders on the World Wide Web" *New Media & Society* 2000, 2(1): 7-28; S. Brunn & M. Dodge, "Mapping the 'Worlds' of the World-Wide Web: (Re)structuring global commerce through hyperlinks" *American Behavioral Scientist* 2001, 44(10): 1717-39; G.A. Barnett; B. Chon; H.W. Park et al., "An examination of international Internet flows: An autopoietic model" *Annual Conference of International Communication Association*, Washington, D.C., 2001, May.

³⁰⁷ Comparando le reti di hyperlink nelle università di 10 paesi asiatici con le reti dei co-autori di articoli scientifici appartenenti alle stesse istituzioni, Park ha trovato una correlazione tra le due strutture confermando l'ipotesi che la comunicazione scientifica si rifletta nella rete di hyperlink: H.W. Park, "E-science and Hyperlink Network Analysis: Collaborative communication through hyperlinking" *Conference of the Netherlands School of Communications Research*, Utrecht, 2002, November.

L'analisi delle reti di hyperlink, discendente diretta dell'analisi delle reti sociali (*social network analysis*), cerca di interpretare la struttura comunicativa del Web, di capire in che misura le connessioni online riflettano o siano influenzate da connessioni offline tra attori sociali.³⁰⁹ Un tipico obiettivo dell'analisi, paragonabile alla ricerca tradizionale dei soggetti che godono di prestigio sociale all'interno di un gruppo, è indagare se esistano o meno, in una rete, dei nodi centrali, ovvero quei siti web che, spesso in virtù della visibilità dell'autore o dell'istituzione da cui derivano, fungono da hub e controllori della comunicazione per gli altri nodi. La misura della centralità di un nodo è effettuata, come nell'analisi classica delle citazioni, contando il numero di link che riceve dagli altri nodi (*indegree centrality*) o che si dipartono da esso (*outdegree centrality*). La centralità inoltre ha un aspetto di "vicinanza" (*closeness centrality*), nel caso del nodo che contiene il percorso più breve per tutti gli altri del gruppo, ed un aspetto legato all'"essere-in-mezzo" (*in-betweenness*), riferito alla frequenza con cui un nodo si trova collocato in mezzo a coppie di altri nodi rappresentando un potenziale fulcro per la loro comunicazione.

Le misurazioni si basano su un repertorio consolidato di modelli e tecniche matematiche: l'analisi dei cluster, lo scaling multidimensionale, i modelli "a blocco" (*block modelling*), i concetti di centralità (*centrality*), centralizzazione (*centralization*) e densità (*density*). Il grado di centralizzazione di una rete di link è la misura della sua tendenza a svilupparsi attorno ai siti centrali, mentre la densità riflette la diversa distribuzione dei link attorno ai singoli nodi.

Grazie all'analisi dei cluster, lo stesso strumento usato dai co-citazionisti per costruire mappe della scienza, si possono classificare i siti in funzione del

³⁰⁸ L. Garton; C. Haythornthwaite; B. Wellman, "Studying online social networks" *Journal of Computer-Mediated Communication* 1997, 3(1) <<http://www.ascusc.org/jcmc/vol3/issue1/garton.html>>.

³⁰⁹ Rassegne di problemi, metodi e risultati della Hyperlink Network Analysis si trovano in: H.W. Park, "Hyperlink Network Analysis: A new method for the study of social structure on the Web" *Connections* 2003, 25(1): 49-61 <<http://www.sfu.ca/%7Einsna/Connections-Web/Volume25-1/7.Hyperlink.pdf>>; M. Thelwall, *Link Analysis. An Information Science Approach* (Amsterdam: Elsevier, 2004)

numero e della distribuzione dei link che ne determinano l'interconnessione diretta o indiretta, mentre i modelli "a blocco" consentono di associare nello stesso blocco siti che presentano schemi di connettività simili.

L'analisi degli hyperlink ha generato una serie di ricerche sperimentali nel campo del commercio elettronico, dei movimenti sociali, della comunicazione interpersonale o interistituzionale. Nel complesso, tali studi confermano che il Web è lo specchio deformato di un mondo variopinto di interessi sociali ed economici e, proprio come nell'universo delle citazioni bibliografiche, la rete di strade che lo attraversano può essere "cartografata" per cogliere gli interessi e le motivazioni di chi vi transita o potrebbe transitarvi.

6.2.3. Analisi delle reti complesse (*Complex Network Analysis*)

Dopo la nascita di Internet ci si è accorti che le reti erano ovunque, bastava aprire gli occhi e prenderne coscienza. Nelle mani di informatici e fisici teorici con ambizioni filosofiche, la rete delle reti mondiali di computer ha perso ogni riferimento alle relazioni spaziali e al contenuto dei documenti ed è diventata un caso particolare della nozione più generale di reti complesse (*complex networks*), mediante la quale possono essere studiate e rappresentate in termini matematici svariate configurazioni naturali, tecnologiche e sociali, dalla rete dei circuiti elettrici a quella dei neuroni del cervello, delle reazioni biochimiche tra le cellule, delle parole del linguaggio. «È in atto una rivoluzione - ha scritto uno dei fautori della nuova scienza delle reti - dove scienziati di ogni disciplina scoprono che la complessità ha un'architettura ben precisa. Siamo arrivati a capire l'importanza delle reti [...] la struttura e la

costruzione di grafi e reti sono la chiave per comprendere il mondo complesso che ci circonda»³¹⁰

La topologia della rete, cioè il numero e la distribuzione delle connessioni tra i nodi che la compongono, è il fulcro del sistema. Alla fine degli anni '50, allorché i matematici ungheresi Erdos e Rényi diedero una sistemazione teorica coerente alla teoria dei grafi ispirata da Eulero, era prevalente la convinzione che l'unico modo per descrivere sistemi sociali e naturali complessi fosse quello di equipararli a dei grafi casuali, nei quali ogni nodo ha, rispetto agli altri, le stesse probabilità di acquisire un link. Secondo il modello del grafo casuale, aumentando progressivamente le dimensioni della rete, quasi tutti i nodi avrebbero alla fine lo stesso numero di link conformandosi ad un particolare tipo di distribuzione probabilistica nota come distribuzione di Poisson. Le ricerche successive sull'organizzazione delle reti complesse in campo tecnologico, naturale e sociale, hanno smentito tale previsione: le reti non sono agglomerati casuali di nodi e link, ma tendono ad auto-organizzarsi in accordo con certe leggi che scandiscono il ritmo della complessità.

La scoperta forse più sorprendente è quella che il mondo delle reti complesse è un "mondo piccolo" (*small world*), dove il numero medio di link sufficienti a collegare due nodi qualunque è basso se commisurato al numero totale dei nodi e l'interazione tra due nodi vicini è possibile tanto quanto l'interazione a lunga distanza. Nella rete che funge da modello e laboratorio per tutte le altre, cioè il World Wide Web, due documenti qualunque si trovano, secondo questa ipotesi, ad una distanza media di 19 click l'uno dall'altro.³¹¹

La riduzione dei confini del mondo in una rete complessa è un effetto statistico della sua architettura globale. In un mondo piccolo infatti i nodi non

³¹⁰ A.L. Barabási, *Link: la nuova scienza delle reti* (Torino: Einaudi, 2004), p. 8 e p. 14. Una rassegna degli studi sulle reti complesse dal punto di vista della meccanica statistica si trova in: A. Scharnhorst, "Complex networks and the Web: insights from nonlinear physics" *Journal of Computer-Mediated Communication* 2003, 8(4) <<http://jcmc.indiana.edu/vol8/issue4/scharnhorst.html>>.

si sparpagliano alla cieca, ma formano dei gruppi o cluster di unità fortemente interconnesse nei quali è sempre possibile, per una qualunque coppia di nodi appartenenti a cluster diversi, stabilire una connessione in un numero limitato di passaggi.³¹² Studiando lo spazio web di 109 università del Regno Unito, ad esempio, Lennart Björneborn ha scoperto che la comparsa di fenomeni legati ai piccoli mondi nasce, in tale contesto, dalla complementarità di due forme di contrazione delle distanze, una al livello degli argomenti (*topic drift*), l'altra al livello della tipologia delle pagine (*genre drift*): i siti interconnessi per affinità di argomento includono una varietà di tipologie di pagine (home page, bookmark, pagine web personali e così via), consentono perciò un accorciamento delle relative distanze all'interno di ciascun cluster tematico; analogamente alcune tipologie di pagine, come i bookmark personali o istituzionali, tendono a includere link a contenuti di argomento eterogeneo, fungono quindi da ponte tra diversi cluster tematici.³¹³

Il Web, le reti neurali del nematode *C. elegans*, la rete di distribuzione dell'energia elettrica degli Stati Uniti occidentali, la rete di collaborazione tra gli attori di Hollywood, la rete delle citazioni di Price, Garfield, Small e quella degli autori di articoli scientifici sono tutti esempi di mondi piccoli intessuti di reti fortemente clusterizzate. I cluster però non nascono da un'aggregazione casuale dei link. L'analisi della topologia dei link in un campione di pagine web rivela infatti un'asimmetria familiare ai bibliometri: pochi siti sembrano attrarre un numero insolitamente alto di connessioni, a fronte di un numero molto maggiore di siti poco connessi e di un numero significativo di siti totalmente scollegati. Così come, nel mondo della documentazione scientifica, pochi autori e poche riviste sono sufficienti a rendere conto della maggior

³¹¹ R. Albert; H. Jeong; A.L. Barabási, "Diameter of the World Wide Web" *Nature* 1999, 401: 130-31 <<http://www.nd.edu/~networks/Papers/401130A0.pdf>>.

³¹² D.J. Watts & S.H. Strogatz, "Collective dynamics of "small world" networks" *Nature* 1998, 393(6684): 440-42.

³¹³ L. Björneborn, *Small-world link structures across an academic web space: a library and information science approach*. Ph.d. Dissertation (Copenhagen: Royal School of Library and Information Science Department of Information Studies, 2004) <<http://www.db.dk/lb/phd/phd-thesis.pdf>>.

parte di ciò che viene letto e citato dalle comunità di specialisti, la rete degli ipertesti è dominata da un certo numero di supernodi o connettori (hub), collegati ad un numero enorme di altri nodi e appartenenti simultaneamente a molti grandi cluster, una sorta di *citation classics* ipermediali che a loro volta citano più di tutti gli altri. Accorciando le distanze tra i nodi, gli hub determinano il comportamento, la stabilità strutturale e la tolleranza del sistema ad attacchi ed errori. Una rete complessa così organizzata è una rete ad invarianza di scala (*scale free network*), in cui non c'è un nodo caratteristico (il nodo medio), nè un picco nel grafico che rappresenta la distribuzione dei link tra i nodi, bensì una gerarchia continua che si distende dai numerosi piccoli nodi con pochi link ai più sporadici hub ricchi di connessioni.³¹⁴ L'opinione diffusa di un cyberspazio democratico ed egualitario, nel quale chiunque può dire la sua sperando di essere visto, si è rivelata perciò una pura utopia: essere online non significa automaticamente essere connessi. Il meccanismo che provoca questo collasso strutturale non è ancora chiaro, ma la sua analogia con il modello del *success breeds success* di Price e Merton appare evidente nell'accumulo preferenziale dei link responsabile della formazione di hub.³¹⁵

Intanto il quadro si è arricchito di ulteriori dettagli. Nel 1999 un gruppo di ricercatori di Altavista, Compaq e IBM ha esplorato lo schema di connettività di oltre 200 milioni di pagine web.³¹⁶ Ne ha ricavato un disegno ancora più intricato della struttura macroscopica in cui sono inserite. Il modello restituisce un Web con una curiosa geometria a forma di papillon (*bow-tie*), nel quale il 28% circa del totale forma un nucleo centrale di pagine fortemente interconnesse (*Strongly Connected Components*), a cui è facile arrivare e da

³¹⁴ A.L. Barabási, *Link...cit.*, pp. 73-88; R. Albert & A.L. Barabási, "Statistical mechanics of complex networks" *Reviews of Modern Physics* 2002, 74(1): 47-97.

³¹⁵ Una delle spiegazioni proposte tiene conto di fattori quali il grado di *fitness* e la competitività dei nodi nel contesto di modelli matematici derivati dalla fisica quantistica: G. Bianconi & A.L. Barabási, "Competition and multiscaling in evolving networks" *Europhysics Letters* 2001, 54: 436-42. Altri modelli esplicativi sono citati in: M. Thelwall; L. Vaughan; L. Björneborn, "Webometrics"...cit., p. 117.

cui è facile uscire per andare altrove. Poi ci sono due gruppi di nodi, ciascuno comprendente un 21% di pagine, con una connettività sbilanciata in sensi opposti: le pagine in entrata (*IN Pages*) hanno molti link verso quelle del nucleo ma ne ricevono pochissimi (si pensi alle pagine web personali), mentre le pagine in uscita (*OUT Pages*) ricevono molti link dal nucleo ma sono quasi prive di collegamenti verso l'esterno (si pensi ai siti aziendali che cercano di imbrigliare il navigatore senza farlo uscire). Un ulteriore 21% di pagine forma "viticci" (*Tendrils*) collaterali da cui si raggiungono pagine di entrata o di uscita senza passare dal nucleo. Infine vi è un 8% di pagine totalmente scollegate dal resto (*Disconnected Pages*). Il modello mostra anche che il mondo piccolo del Web non è poi così piccolo, visto che la connettività è condizionata dalla struttura globale ed esiste solo una probabilità del 24% di trovare un qualunque percorso capace di connettere due pagine prese a caso. Ciò che l'esperimento tuttavia conferma in pieno è la validità su larga scala delle leggi di potenza. La distribuzione dei link segue, in poche parole, la stessa legge che Pareto aveva rilevato nella ripartizione dei redditi tra fasce di popolazione e Lotka, Bradford, Zipf avevano estratto dall'analisi della documentazione scientifica e delle parole del linguaggio: l'eccellenza webmetrica, non diversamente da quella bibliometrica, sembra concentrarsi in un numero limitato di nodi.

Il fatto che una topologia simile a quella appena delineata sia emersa, su scala diversa, anche da studi relativi ai *webspaces* di sistemi universitari nazionali e di singole nazioni, ha suggerito l'ipotesi che le reti complesse condividano alcune importanti proprietà con gli oggetti frattali e che questo sia uno dei sintomi più evidenti della loro capacità di auto-organizzarsi.³¹⁷

³¹⁶ A. Broder; R. Kumar; F. Maghoul et al., "Graph structure in the Web" *9th International World Wide Web Conference*, May 2000 <<http://www.almaden.ibm.com/cs/k53/www9.final>>.

³¹⁷ M. Thelwall & D. Wilkinson, "Graph structure in three national academic Webs: Power laws with anomalies" *Journal of the American Society for Information Science and Technology* 2003, 54(8): 706-12; R. Baeza-Yates & C. Castillo, "Relating Web characteristics with link based Web page ranking" *Proceedings of String Processing and Information Retrieval*, Laguna San Rafael, Chile, November 2001.

Capitolo 7

La citazione come artificio retorico e le critiche al modello normativo

Vale la pena analizzare, contare, confrontare le citazioni? Cosa dice una citazione bibliografica? Come varia il suo significato in funzione del testo che la sorregge? Soprattutto: a cosa (e a chi) serve citare?

Dalla metà degli anni '70, l'urgenza di dare un fondamento teorico agli studi scientometrici e di convalidarne l'uso politico in sede di valutazione dei meriti individuali e collettivi ha spinto ricercatori di varia estrazione a rendere espliciti i motivi per i quali gli autori citano e ad inquadrare in uno schema concettuale plausibile la variabilità degli stili e delle pratiche citazionali.

Un'indagine di questo tipo poteva prendere almeno due direzioni: concentrarsi sulla citazione come processo privato, quindi sulla psicologia del citante e sulla fenomenologia dei comportamenti citazionali, oppure, in maniera più pragmatica, tentare una classificazione tipologica dei riferimenti bibliografici e del loro significato in funzione del contesto che li sorregge.³¹⁸

³¹⁸ Una rassegna sulla *citation context analysis* si trova in: H.G. Small, "Citation context analysis," in *Progress in Communication Sciences*, 3, edited by B.J. Dervin, et al. (Norwood, NJ: Ablex, 1982): 287-310.

7.1. Motivi per citare

Intervistando 26 ricercatori di varie discipline scientifiche e umanistiche dell'Università dello Iowa, Terrence Brooks ha registrato la variazione dei loro motivi per citare su sette scale precompilate. I motivi proposti erano: 1) documentare il proprio livello di aggiornamento (*currency scale*); 2) distribuire crediti negativi, quindi criticare, correggere, sconfessare tesi e risultati altrui (*negative credit*); 3) documentare informazioni di tipo “operativo”, cioè concetti, tecniche, strumenti, risultati (*operational information*); 4) persuadere i lettori della validità dei propri argomenti (*persuasiveness*); 5) distribuire crediti positivi, il classico “pagare un debito” (*positive credit*); 6) suggerire ulteriori letture di approfondimento o utili per la comprensione del contesto (*reader alert*); 7) mostrare il proprio consenso con la comunità degli specialisti, citare i documenti che si pensa siano ritenuti meritevoli di citazione da parte dei colleghi e potenziali lettori (*social consensus*).³¹⁹

La scala più quotata, sia per gli scienziati che per gli umanisti, si è rivelata quella della persuasione: la maggior parte degli intervistati ammetteva di fare uno delle note per convincere l'audience virtuale della validità delle tesi sostenute. Un risultato abbastanza sorprendente se si pensa che filosofi e storici della scienza hanno speso quintali di erudizione per mostrare che gli scienziati del passato costruivano teorie di successo essenzialmente grazie ad artifici retorico-discorsivi dei quali non pare fossero consapevoli.³²⁰

³¹⁹ T.A. Brooks, “Private acts and public objects: an investigation of citer motivations” *Journal of the American Society for Information Science* 1985, 36(4): 223-29, p. 226.

³²⁰ Il caso più clamoroso è quello di Galileo Galilei nella ricostruzione di Maurice Finocchiaro: M.A. Finocchiaro, *Galileo and the Art of Reasoning: Rhetorical Foundations of Logic and Scientific Method* (Dordrecht-Boston-London: Reidel, 1980). Da punti di vista e con obiettivi diversi il ruolo della retorica nell'argomentazione scientifica si trova discusso in: P.K. Feyerabend, *Contro il metodo: abbozzo di una teoria anarchica della conoscenza*, 2 ed. (Milano: Feltrinelli, 2003); M. Pera, *Scienza e retorica* (Bari: Laterza, 1991).

Garfield non ha mai attribuito un grosso peso allo studio dei motivi per i quali un autore cita. In un'occasione tuttavia egli ha elencato alcuni di tali motivi, in aggiunta a quello generico di documentare un'affermazione fatta nel testo:³²¹

- 1) omaggiare i pionieri;
- 2) accreditare lavori simili o correlati (omaggio ai pari);
- 3) identificare metodologie e tecniche;
- 4) suggerire letture di contesto o inquadramento;
- 5) correggere il proprio lavoro;
- 6) correggere il lavoro altrui;
- 7) criticare lavori precedenti;
- 8) giustificare delle asserzioni;
- 9) segnalare pubblicazioni imminenti;
- 10) pubblicizzare lavori poco conosciuti;
- 11) autenticare dati e classi di fatti;
- 12) identificare le pubblicazioni originali nelle quali è stato discusso un concetto o un'idea;
- 13) identificare le pubblicazioni che descrivono un concetto eponimo (morbo di Parkinson, malattia di Hodgkin);
- 14) negare la validità del lavoro o delle idee altrui;
- 15) contestare le pretese di priorità altrui.

Un'analisi in profondità delle citazioni - sostenevano Moravcsick e Murugesan nel 1975 - non era stata condotta fino ad allora sia per l'età relativamente giovane dei *citation indexes*, sia soprattutto perché gli autori di tali studi erano in genere ricercatori (sociologi e bibliotecari) non sufficientemente equipaggiati per comprendere il contenuto tecnico-scientifico delle pubblicazioni esaminate. Ciò poneva dei limiti alle conclusioni che essi

³²¹ E. Garfield, "Can citation indexing be automated?," in *Essays of an Information Scientist 1962-1973*, 1

potevano trarre «perché molte delle sottigliezze delle citazioni sono connesse con la qualità del documento citato ed il contesto in cui la citazione è fatta».³²²

I due studiosi indagarono allora, mediante un'analisi contenutistica, le funzioni e la qualità dei riferimenti bibliografici in un campione di 30 articoli di fisica delle alte energie pubblicati su *Physical Review* dal 1968 al 1972 chiedendosi, per ogni citazione, a quale delle seguenti categorie dicotomiche potesse essere assegnata: 1) Concettuale o operativa: significa un concetto, una teoria oppure una tecnica, un dispositivo strumentale?; 2) Organica o superficiale: è necessaria per la comprensione del documento citante (o per la costruzione del suo contenuto) oppure è soltanto un generico riconoscimento di altri lavori compiuti nella stessa area? 3) Evolutiva o giustappositiva: è il fondamento delle conclusioni del testo citante o si riferisce a soluzioni alternative?; 4) Confermativa o negativa: è confermata o smentita dal testo citante?

Le classi 1) e 4) stimano unicamente il grado di connessione tra documenti, mentre le classi 2) e 3) implicano una valutazione qualitativa dei lavori citati.

Il risultato più significativo fu che il 41% delle citazioni esaminate erano superficiali (*perfunctory*), inserite per il solo fatto che "funzionano comunque", ed il 31% ridondanti, cioè aggiunte, senza una reale necessità, semplicemente con lo scopo di non scontentare nessuno. Chiunque - concludevano gli autori - potrebbe incrementare il proprio prestigio citazionale scrivendo lavori mediocri su argomenti "di moda" che saranno poi citati per puro dovere di cronaca, quindi l'utilità delle citazioni per la valutazione è quantomeno dubbia.³²³ Le citazioni superficiali comunque non

(Philadelphia: ISI Press, 1977): 84-90, p. 85.

³²² M.J. Moravcsik & P. Murugesan, "Some results on the quality and function of citations" *Social Studies of Science* 1975, 5: 86-92, p. 87. I due autori sperimentarono la classificazione elaborata sulle citazioni di articoli appartenenti a due settori della fisica teorica (teoria della superconduttività di Bardeen-Cooper-Schrieffer e teoria della non conservazione della parità) per studiare l'effetto delle rivoluzioni scientifiche sulle pratiche citazionali degli scienziati: M.J. Moravcsik & P. Murugesan, "Citation patterns in scientific revolutions" *Scientometrics* 1979, 1(2): 161-69. Un ulteriore test empirico dello schema si trova in: V. Cano, "Citation behavior: classification, utility, and location" *Journal of the American Society for Information Science* 1989, 40(4): 284-90.

³²³ «A large fraction of the references are perfunctory. This raises serious doubts about the use of citations as a quality measure, since it is then quite possible for somebody or some group to chalk up high citation counts

sono totalmente prive di valore perché, come suggerito da Belver Griffith, segnalano al sociologo l'esistenza di un gruppo di documenti riconosciuti e ben integrati nell'arsenale metodologico e concettuale di una comunità di specialisti.³²⁴

Una classificazione simile alla precedente fu sperimentata, su un ulteriore campione di 43 articoli di fisica delle alte energie, da Chubin e Moitra, i quali approfondirono lo schema iniziale dei comportamenti citazionali con tre variabili aggiuntive: la forma delle pubblicazioni (lettera, articolo di ricerca), la tipologia di ricerca documentata (lavoro sperimentale o teorico) e la modalità di distribuzione (la rivista su cui i lavori sono pubblicati). Scoprirono così che i fisici citano quasi sempre per affermare anziché per negare (un buon argomento a favore del conteggio delle citazioni), che nelle *physical letters* si tende, più che negli articoli *in extenso*, a citare lavori realmente "essenziali" per i risultati conseguiti e che, ribaltando la prospettiva, gli articoli dei *letter-journals* sono citati più velocemente e più diffusamente rispetto alle versioni definitive delle pubblicazioni che preannunciano. Alla lunga però sono queste ultime a far progredire la disciplina trasformandosi in conoscenza certificata e continuando a vivere, nella migliore delle ipotesi, una vita ricca di citazioni.³²⁵ Dalla tabella riassuntiva dei risultati si ricava inoltre che il 20% delle citazioni erano superficiali ed il 32% supplementari, cioè contenevano informazioni addizionali non strettamente pertinenti con le tesi discusse.³²⁶

Davanti alle difficoltà concettuali dell'analisi delle citazioni, l'ambizione dei bibliometri era perfezionare le tecniche di conteggio attribuendo un "peso" diverso alle citazioni a seconda delle funzioni esercitate nel testo. L'analisi del contesto e del contenuto rappresentava perciò «un'estensione logica delle

by simply writing barely publishable papers on fashionable subjects which will then be cited as perfunctory, 'also ran' references»: M.J. Moravcsik & P. Murugesan, "Some results..." cit., p. 91.

³²⁴ L'osservazione di Belver Griffith è riportata in: M.J. Moravcsik, *Current Contents - This Week's Citation Classic* 1985(48, December 2): 18, p. 18.

³²⁵ D.E. Chubin & S.D. Moitra, "Content analysis of references: adjunct or alternative to citation counting?" *Social Studies of Science* 1975, 5: 423-41.

³²⁶ *Ibid.*, p. 429.

misure sempre più raffinate di output scientifico: numero di scienziati, numero di autori scientifici, numero di pubblicazioni, numero di citazioni, e ora, numero di citazioni “rilevanti”». ³²⁷ Ma l'analisi generava un paradosso: se le citazioni sono gli elementi basilari di un linguaggio altamente strutturato per la trasmissione e la manipolazione di concetti scientifici (Garfield, Small), prese una alla volta e proiettate sullo sfondo del testo che ne reclama la presenza, esse tradiscono una molteplicità di significati difficilmente compatibile con la fiducia riposta nell'utilità del loro conteggio.

Il dubbio ha accompagnato l'uso dei *citation indexes* sin dagli inizi, ed è un dubbio radicale, una vera e propria crisi d'identità della citazione bibliografica e della sua missione intellettuale. Già verso la metà degli anni '60 era affiorato il sospetto che, a differenza di quanto pensava Merton, solo occasionalmente le pratiche citazionali riflettano l'attenzione degli autori verso il riconoscimento e la puntuale ricostruzione dei meriti di colleghi e precursori. «Quante volte - si chiedeva Norman Kaplan nel 1965 - i lavori altrui sono citati senza essere letti attentamente? Quante volte le citazioni sono semplicemente copiate dalla bibliografia del lavoro di qualcun'altro senza leggere o dare credito alla persona che ha condotto la ricerca bibliografica originale? Quante volte le citazioni sono aggiunte all'articolo finito a seguito di un ripensamento o per allestire la vetrina?» ³²⁸ Gli usi impropri possono essere molteplici. Uno scienziato bravo può decidere di dispensare crediti ai suoi colleghi citandone le pubblicazioni per puro sfoggio di modestia istituzionale, così come uno scienziato mediocre può scaricare in parte o totalmente la responsabilità di errori o omissioni sui testi autorevoli che sceglie di citare, mentre uno scienziato disonesto può citare il lavoro non pubblicato di un collega sulle cui intuizioni ha costruito una pubblicazione di successo: in tal caso, per il collega defraudato, il risarcimento bibliografico è

³²⁷ M.J. Moravcsik, *Current Contents - This Week's Citation Classic* 1985(48, December 2): 18, p. 18.

³²⁸ N. Kaplan, “The norms of citation behavior: prolegomena to the footnote” *American Documentation* 1965, 16(3): 179-84, p. 181.

motivo di ben scarsa soddisfazione sottraendogli persino l'opportunità di una causa di plagio.³²⁹ Nella misura in cui le citazioni possono essere imprecise, incomplete o persino disoneste, le norme della loro produzione appaiono molto meno stabili di quanto pretendono Merton e i bibliometri.

Si può anche andare oltre e sostenere che la citazione incorpora il documento citato in un'argomentazione, funge pertanto da dispositivo retorico a tutti gli effetti, il cui fine ultimo è lo stesso della sapiente organizzazione formale conferita dall'autore al saggio scientifico: persuadere la comunità scientifica della validità (novità, importanza, verità) delle tesi sostenute.³³⁰

Nella sua forma ortodossa, l'articolo scientifico ha una struttura molto lineare. Un tipico *research article* pubblicato su una rivista biomedica, ad esempio, è organizzato in cinque sezioni canoniche:³³¹

- 1) un'introduzione, nella quale si espongono gli obiettivi dello studio e si ricostruisce lo stato dell'arte, citando la maggior parte dei lavori precedenti sull'argomento;
- 2) la sezione dei metodi, che descrive le metodologie di raccolta e trattamento dei dati sperimentali;
- 3) la sezione dei risultati, nella quale, oscurando ogni questione relativa al significato dei dati presentati, si viene messi di fronte ad una quantità talvolta impressionante di informazioni fattuali;
- 4) la discussione, nella quale finalmente si disserta sul significato dei dati sperimentali, spesso riferendosi ai lavori di altri autori per corroborarli o

³²⁹ J. Bensam, "The aesthetics and politics of footnoting" *International Journal of Politics, Culture and Society* 1988, 1(3): 443-70.

³³⁰ Questa la tesi sostenuta, tra gli altri, da: J.R. Ravetz, *Scientific Knowledge and its Social Problems* (Oxford: Clarendon Press, 1971); G.N. Gilbert, "Referencing as persuasion" *Social Studies of Science* 1977, 7: 113-22. Due rassegne di studi sull'argomento si trovano in: M.H. MacRoberts & B.R. MacRoberts, "Problems of citation analysis: a critical review" *Journal of the American Society for Information Science* 1989, 40(5): 342-49; M. Liu, "The complexities of citation practice: A review of citation studies" *Journal of Documentation* 1993, 49(4): 370-408.

³³¹ Questa presentazione semplificata della struttura di un articolo scientifico si basa sul breve ma illuminante: P.B. Medawar, "Is the scientific paper fraudulent? Yes; it misrepresents scientific thought" *Saturday Review* 1964(1 August): 42-43.

contestarli, e si fa finta di chiedersi se i risultati sono significativi conoscendo già la risposta, cioè sempre e solo: “sì, sono significativi, in un modo o nell’altro”.

5) la bibliografia, contenente una lista disincarnata dei titoli citati nel testo.

Con l'avvento dei *citation indexes* la sezione bibliografica, sulla cui lettura ottica l'ISI ha costruito le proprie fortune, ha visto amplificati i propri meriti. Il prezzo da pagare per la celebrità nondimeno è stato alto in termini di solitudine e omologazione, entrambe condizioni necessarie a garantire la riproducibilità tecnica dei dati citazionali nel circuito degli apparati bibliografici: da un lato, soprattutto nelle scienze naturali e sociali, isolamento e separazione progressiva della citazione da ogni elemento testuale con cui poteva confondersi, in primis dalle note discorsive a piè di pagina; dall'altro, assoggettamento alle norme redazionali rigorose dei manuali di stile, nati tra la fine del XIX e gli inizi del XX secolo e gradualmente adottati come standard di riferimento dagli editori delle riviste (*Harvard Style, MLA Style, Vancouver Style, APA Style*). Gli indici di Garfield non sono stati i diretti responsabili di questo processo, ma con il loro velato disprezzo per stili citazionali *no machine-readable* ne hanno senza dubbio tratto vantaggio contribuendo ad accelerarlo.³³² Per molte riviste scientifiche infatti è storicamente documentabile la transizione da uno stile di citazione basato su un uso estensivo delle note a piè di pagina ad uno stile che associa, ad un numero tra parentesi (*Vancouver Style*) o alla segnalazione abbreviata autore-

³³² «References interspersed throughout the text or split between the text and footnotes, as well as footnotes that contain multiple references, can be the toughest part of the editing job. Most often found in social sciences journals, these types of references require extensive editing notation to identify, integrate, and complete them, and may require the help of a professional translator if they involve non-English citations»: E. Garfield, *Citation Indexing: its Theory and Application...cit.*, pp. 25-26. A proposito della selezione delle riviste per i *Current Contents*: «But format considerations are important in choosing between those non-“core” journals that may or may not warrant coverage, We must consider whether a journal’s contents page is photoreproducible, if the article titles are sufficiently descriptive, and if the titles can be scanned easily»: E. Garfield, “Journal selection for *Current Contents*: editorial merit vs. political pressure,” in *Essays of an Information Scientist 1985. Ghostwriting and other Essays*, 8 (Philadelphia: ISI Press, 1986): 96-104, pp. 102-103.

data nel testo (*Harvard Style*), una lista di *full references* in coda al manoscritto. Un sistema del genere è tecnicamente compatibile con la cattura e la memorizzazione digitale automatica dei record bibliografici, ma ha anche una funzione più subdola. Nell'ambito delle scienze sociali esso ha rappresentato, a partire dagli anni '30 del '900, una delle strategie testuali messe in atto dagli editori per marcare i documenti con i segni appropriati di obiettività scientifica.³³³ Nel manuale di stile dell'American Psychological Association (APA), per esempio, le correlazioni tra l'*Harvard Style*,³³⁴ le convenzioni stilistiche utili alla preparazione dei manoscritti per la pubblicazione e i valori fondamentali dell'epistemologia disciplinare sono evidenti. Per risultare adatto alla pubblicazione, il resoconto sperimentale (*empirical report*) su cui si baserà l'articolo finale deve essere rimodellato nelle cinque sezioni canoniche (introduzione, metodi, risultati, discussione, bibliografia), secondo uno schema che riduce lo studio dei fenomeni psicologici ad una sequenza lineare di operazioni di natura sostanzialmente empirica: 1) passare in rassegna lo stato dell'arte nel settore specifico in cui si colloca l'esperimento; 2) raccogliere e analizzare nuovi dati; 3) correlare le conclusioni raggiunte con le teorie psicologiche correnti. L'articolo deve essere preferibilmente redatto in uno stile oggettivo e impersonale; non deve citare integralmente i frammenti di testi altrui ma limitarsi a parafrasarli o sintetizzarli, deve adoperare un linguaggio prudente nella presentazione dei risultati e delle conclusioni ed evitare animosità e attacchi personali nelle questioni controverse.³³⁵

³³³ D. Gregory, "Editorial" *Environment and Planning. D: Society and Space* 1990, 8(1): 1-6. L'editoriale di Gregory è stampato due volte con due stili citazionali diversi per illustrare visivamente la differenza argomentata nel testo: nel primo caso le note a piè di pagina contengono la segnalazione bibliografica completa dei testi citati, nel secondo sono conformi allo *Harvard Style*.

³³⁴ American Psychological Association, *Publication manual of the American Psychological Association*, 5 ed. (Washington, DC: APA, 2001).

³³⁵ L'interpretazione epistemologica del manuale dell'APA qui discussa si trova in: R. Madigan; S. Johnson; P. Linton, "The language of psychology. APA Style as epistemology" *American Psychologist* 1995, 50(6): 428-36; R. Madigan; P. Linton; S. Johnson, "APA Style: Quo vadis?" *American Psychologist* 1996, 51(6): 653-55; D. Vipond, "Problems with a monolithic APA style" *American Psychologist* 1996, 51(6): 653.

L'esclusione delle note discorsive a favore dello *Harvard Style* è perfettamente coerente con questo modello e con i valori epistemologici che incarna. Le conseguenze dell'esclusione sono diverse: neutralizza e mette sullo stesso piano tutte le citazioni sacrificando alla linearità dell'argomentazione ogni irregolarità di percorso; impedisce all'autore (e al lettore) di qualificare o contestualizzare l'uso di ciascuna fonte; esclude la possibilità di arricchire l'esposizione con linee parallele di discussione e appiattisce il testo sulla funzione primaria di riferire "fatti", "informazioni", nell'ingenua convinzione che il linguaggio sia soltanto uno strumento per trasferire in maniera non ambigua informazioni dall'autore al lettore.³³⁶

Uno schema argomentativo di questo tipo dice poco o nulla su come è nata l'idea scientifica alla base della pubblicazione e dice ancora meno sugli stratagemmi messi in atto per far combaciare l'ipotesi iniziale con i dati osservativi. Piuttosto esso racconta una favola ben congegnata: c'è una mente vuota e priva di pregiudizi, quella dello scienziato, nella quale si versano come in una brocca i dati dell'esperienza sensoriale raccolti nel laboratorio; la teoria scientifica nasce dalla generalizzazione delle osservazioni imparziali compiute in laboratorio e condensate in proposizioni semplici, ognuna delle quali descrive una situazione di fatto empiricamente controllabile. È il mito dell'induzione di Bacone e Stuart Mill, l'idea cioè che un semplice accumulo di esperienze individuali sia sufficiente tanto a scoprire quanto a giustificare la validità di nuove proposizioni scientifiche. La filosofia della scienza del '900, da Bertrand Russell e Karl Popper in poi, ha disintegrato la logica induttiva,

³³⁶ «If at one time an author was able to use footnotes to make editorial asides, stepping out of his or her role as a 'neutral' reporter to address the reader more directly or informally, now such notes are discouraged. Neutral as this change may seem, it has several effects. Perhaps most obviously, it virtually precludes the publication of works that rely heavily on archival sources; in that way it skews publication away from historical subjects and benefits contemporary works and works of secondary scholarship. It also makes it easier for readers to 'mine' works, since they no longer need go through references one by one. But more important, it neutralizes citations, making each of equal weight, thus making it much more difficult for an author to qualify his or her conclusions, or his or her references»: M.R. Curry, "On the possibility of ethics in geography: writing, citing and the construction of intellectual property" *Progress in Human Geography* 1991, 15(2): 125-47, p. 133. Osservazioni analoghe si trovano in: J. Hartley, "What do we know about footnotes? Opinions and data" *Journal of Information Science* 1999, 25: 205-12.

mostrando che non esiste osservazione neutrale, non esiste esperienza senza pregiudizi così come non esiste esperimento di laboratorio senza un insieme di ipotesi e attese che condizionano la scelta dei metodi e dei criteri di valutazione dei risultati ottenuti.

Sviluppata sino in fondo, la tesi che considera la citazione bibliografica uno degli espedienti retorici utilizzati dagli scienziati per rafforzare e rendere accettabili le ipotesi contenute nel saggio accademico è in sintonia con gli esiti di una parte della sociologia contemporanea, in particolare con le posizioni sostenute dagli epigoni del "programma forte" della scuola di Edinburgo (David Bloor, Barry Barnes), dai seguaci del "programma empirico del relativismo" del gruppo di Bath (Harry Collins, Trevor Pinch), dagli etnometodologi (Harold Garfinkel, Bruno Latour, Steve Woolgar, Karin Knorr-Cetina) e dai fautori dell'*actor-network theory* (Bruno Latour, Michel Callon).³³⁷ Sebbene da prospettive e con esiti molto diversi, questi autori hanno prodotto, mediante *case studies* storiografici o attraverso l'osservazione diretta della vita di laboratorio, una serie di evidenze tese a mostrare che la scatola nera della scienza, davanti alla quale Merton si è fermato senza aprirla, contiene l'immagine riflessa degli scienziati stessi e dei loro interessi sociali piuttosto che quella del mondo esterno. Confrontando ciò che gli scienziati fanno e dicono realmente nella vita di laboratorio con ciò che gli scienziati scrivono o dicono di aver fatto nei resoconti ufficiali, in primo luogo negli articoli scientifici, i sociologi di ultima generazione hanno cercato di demolire il mito del laboratorio come bottega per il confezionamento della verità assoluta, neutrale, disinteressata. Dal confronto emerge infatti una differenza radicale tra il registro dei resoconti formali - il *research article* pubblicato sulla rivista ufficiale - e quello dell'attività pratica: nel primo il mondo fisico sembra agire e parlare da solo, lo stile è impersonale, la metodologia esposta

³³⁷ Per un inquadramento critico delle tesi dei costruttivisti vedi: P. Bourdieu, *Il mestiere di scienziato: corso al Collège de France 2000-2001* (Milano: Feltrinelli, 2003); M. Bucchi, *Scienza e società* (Bologna: Il Mulino, 2002).

in termini di formule generali ricalcate sui manuali correnti, e manca ogni riferimento della dipendenza delle osservazioni dalla speculazione teorica; nel secondo invece regna la pura contingenza, la scelta e la soluzione dei problemi tradiscono l'applicazione di un senso pratico (*practical skills*) e di un repertorio di trucchi del mestiere estranei ad ogni forma di codifica. I *Dictionaries of Useful Research Phrases* trovati da Nigel Gilbert e Michael Mulkay durante le frequentazioni dei laboratori di biochimica sono un esempio eloquente, per quanto umoristico, della incompatibilità tra i due domini: nel registro formale si scrive "È noto da tanto tempo che ..." per dire semplicemente "Non ho perso tempo a controllare i riferimenti bibliografici...", oppure "Sono mostrati i risultati tipici ..." per dire "Si mostrano solo i risultati positivi ...", oppure "Di grande importanza teorica e pratica ..." per dire "Importante per me ...", o ancora "Non è stato possibile trovare risposta definitiva a tali quesiti ..." per dire "L'esperimento non ha funzionato, ma ho pensato bene di ricavarne comunque una pubblicazione".³³⁸

In poche parole, nel passaggio dall'esperienza di laboratorio al resoconto formale si parte dai dati empirici, dai risultati sperimentali, e si "aggiusta" l'esposizione fino a far quadrare i conti. Nessuno si accorgerà del trucco perché l'organizzazione formale dello *scientific paper*, complice anche la citazione bibliografica, è la scena di un delitto perfetto, dove le zone d'ombra sono occultate dalla trasparenza formale dell'argomentazione. Che poi quella teoria risulti accettabile o meno alla comunità dei pari è una questione di strategia "militare", di alleanze intra- ed extra-testuali, di giochi di potere e contrattazioni informali svolte in larga misura secondo la logica della "combriccola". L'immagine della scienza che ne deriva è quella di un'attività sociale che non intrattiene alcun rapporto privilegiato con la verità, un gioco linguistico simile a quelli giocati nella letteratura, nell'arte, nella giurisprudenza, con regole convenzionali dettate da fazioni locali di ricercatori

³³⁸ G.N. Gilbert & M. Mulkay, *Opening Pandora's Box: a Sociological Analysis of Scientists' Discourse*

rissosi e per nulla disinteressati. La vita di laboratorio si riduce, nelle formulazioni più estreme, ad una mera attività semiologica, circolazione di iscrizioni (Derrida), enunciati (Foucault), testi che raccontano le cose non come si sono effettivamente svolte ma come dovrebbero svolgersi secondo le norme ideali del protocollo scientifico.³³⁹ Il ruolo della citazione bibliografica, in tale contesto, appare in netto contrasto con quello che gli attribuiscono Garfield, Merton, Small: la citazione non testimonia un trasferimento di idee bensì un commercio di interessi sociali, la sua missione è inibire le critiche, schiacciare gli avversari col peso dell'autorità, persuadere l'audience degli addetti ai lavori della validità di un testo elencando tutti i potenziali alleati dell'autore. Un articolo senza citazioni - scrive Latour - «è come un bambino che vaga da solo in una città che non conosce».³⁴⁰

Le critiche hanno indotto i bibliometri ad affinare le teorie sulla natura delle citazioni bibliografiche. Le citazioni sono diventate di conseguenza oggetti sempre più sfuggenti, multidimensionali, situati nel punto d'intersezione di sistemi diversi e spesso in conflitto tra loro: il sistema sociale delle ricompense, il sistema linguistico delle tecniche retoriche di costruzione dei testi, il sistema delle pubblicazioni e della comunicazione scientifica ufficiale.³⁴¹

(Cambridge: Cambridge University Press, 1984), pp. 176-177.

³³⁹ Knorr Cetina ha analizzato gli stadi successivi attraversati, in un laboratorio, da un *draft* per giungere alla versione finale destinata alla pubblicazione, mostrando come le strategie di de-personalizzazione del resoconto intervengano sin dalle prime fasi di stesura del testo di un articolo scientifico. Vedi: K.D. Knorr-Cetina & M. Mulkay (Ed.), *Science Observed. Perspectives in the Social Study of Science* (London: Sage, 1983).

³⁴⁰ «Un documento diventa scientifico quando le sue asserzioni cessano di essere isolate e quando il numero di persone coinvolte nella sua pubblicazione sono molte ed esplicitamente indicate nel testo. [...] Le citazioni persuadono non solo con il "prestigio" o con il "bluff". E' di nuovo una questione di numeri. Un articolo senza citazioni è come un bambino che vaga da solo in una città che non conosce: è isolato, perso, in balia degli eventi. Al contrario, attaccare un articolo infarcito di citazioni, significa che lo scettico deve indebolire ognuno degli articoli, o almeno sarà atterrito all'idea di doverlo fare»: B. Latour, *La scienza in azione. Introduzione alla sociologia della scienza* (Torino: edizioni di Comunità, 1998), pp. 42-43. Sulla teoria della citazione di Latour vedi: T. Luukkonen, "Why has Latour's theory of citations been ignored by the bibliometric community? Discussion of sociological interpretations of citation analysis" *Scientometrics* 1997, 38: 27-37.

³⁴¹ S.E. Cozzens, "What do citations count? The rhetoric-first model" *Scientometrics* 1989, 15: 437-47; L. Leydesdorff, "Theories of citation?" *Scientometrics* 1998, 43(1): 5-25 <Preprint in: <http://users.fmg.uva.nl/lleydesdorff/citation/>>; M. Callon; J. Law; A. Rip, "Qualitative scientometrics," in *Mapping the Dynamics of Science and Technology*, edited by M. Callon, et al. (London: The Macmillan

Una teoria soddisfacente della citazione bibliografica non esiste e qualcuno sostiene che non ce n'è affatto bisogno.³⁴² Le tesi dei sociologi costruttivisti sono comunque un antidoto efficace contro la fiducia eccessiva e acritica nel significato dei dati bibliometrici, mentre la loro estremizzazione, spingendo ad identificare la realtà fisica con il prodotto di una costruzione sociale, impedisce di distinguere l'attività scientifica ed i relativi prodotti, comprese le citazioni bibliografiche, da altre forme di cultura ed espressione creativa. In campo scientifico, difficile non ammetterlo, esistono delle norme che orientano l'attività dei ricercatori verso una comprensione dei fenomeni naturali e sociali diversa da quella dell'arte. Il fatto che tali norme siano flessibili o socialmente condizionate, che non vengano seguite sempre e ovunque e che il sistema ufficiale di comunicazione scientifica le rifletta in maniera distorta e idealizzata non implica la loro negazione assoluta. Comunque se ne valuti la presa sul comportamento effettivo degli scienziati, le norme forniscono un punto di riferimento convenzionale ed un insieme di aspettative contro il quale misurare le deviazioni. Ecco perché, prima di intraprendere una conversazione con un sociologo di indirizzo costruttivista, Belver Griffith era solito prescrivere all'interlocutore una dose pesante di Merton: «take two Mertons and see me in the morning».³⁴³

7.2. Vita e morte dell'autore scientifico

La destinazione finale di ogni conteggio delle citazioni a fini politico-amministrativi è l'autore. Gli scientometri contano pubblicazioni e citazioni di documenti prodotti da “qualcuno” - un individuo, un gruppo, un'istituzione -

Press, 1986): 103-23; T. Luukkonen, *Citations in the rhetorical, reward, and communication systems of science*. Ph.D. Thesis (Tampere: University of Tampere, 1990).

³⁴² A.F.J. van Raan, “In matters of quantitative studies of science the fault of theorists is offering too little and asking too much” *Scientometrics* 1998, 43(1): 421-28.

³⁴³ Citato in: H.G. Small, “On the shoulders of Robert Merton...cit., p. 72.

con l'obiettivo di elaborare indicatori di performance utili a stabilire se quel “qualcuno” è più o meno meritevole di “qualcun altro” avendo esercitato un impatto maggiore sulla comunità dei ricercatori. Il sistema accademico di ricompense e riconoscimenti, cliente privilegiato della scientometria, si basa a sua volta sul presupposto che sia sempre possibile identificare e assegnare la responsabilità intellettuale individuale di un lavoro scientifico. Tale presupposto è perfettamente coerente sia con il modello classico (mertoniano) di distribuzione dei crediti scientifici, sia con la dottrina liberale del diritto d'autore che, dagli inizi del XVIII secolo, ha ingabbiato la creatività espressiva nella categoria della proprietà privata plasmando la figura dell'autore titolare dei diritti (moralì ed economici) di proprietà intellettuale. A partire dagli anni '50, per motivazioni e con strategie diverse, la monarchia dell'autore è stata destabilizzata su vari fronti. Filosofi e teorici della letteratura hanno cominciato a demolire qualunque pretesa di scoprire l'origine soggettiva dei testi letterari e di spiegare un'opera d'arte attraverso l'interiorità dello scrittore: «un testo - scriveva Roland Barthes nel 1968 - non consiste in una serie di parole esprimenti un significato unico, in un certo senso teologico (che sarebbe il messaggio dell'Autore-Dio), ma è uno spazio a più dimensioni, in cui si congiungono e si oppongono svariate scritture, nessuna delle quali è originale: il testo è un tessuto di citazioni, provenienti da i più diversi settori della cultura»³⁴⁴ Oggi questa tesi non gode più di molta fortuna e la critica ne ha smascherato le debolezze,³⁴⁵ ma la crisi d'identità dell'autore che essa ratificava non era un fatto isolato. Negli stessi anni, all'interno della fisica delle alte energie e delle scienze biomediche,³⁴⁶ l'autore, il "qualcuno" di cui la scientometria vuole misurare l'impatto, il ricercatore meritevole che

³⁴⁴R. Barthes, “La morte dell'autore,” in *Il brusio della lingua. Saggi critici IV* (Torino: Einaudi, 1988): 51-54, p. 54. Nel 1969 Foucault tracciava le linee di un'indagine archeologica sull'origine dell'entità "autore" in: M. Foucault, “Che cos'è un autore?,” in *Scritti letterari* (Milano: Feltrinelli, 1984): 1-21.

³⁴⁵C. Benedetti, *L'ombra lunga dell'autore. Indagine su una figura cancellata* (Milano: Feltrinelli, 1999).

³⁴⁶Per la fisica delle alte energie vedi, ad esempio: A.R. Pickering & P.W. Trower, “Sociological problems of high-energy physics” *Nature* 1985, 318: 243-45; per la biomedicina: J.T. King, “How many neurosurgeons

l'università vuole premiare, il soggetto giuridico che detiene i diritti di proprietà intellettuale, era sempre meno un'unità indivisibile responsabile dell'intero contenuto intellettuale di una pubblicazione e sempre più il coefficiente frazionario di un valore associato ad una classe di co-autori che poteva contare da poche unità a centinaia e persino migliaia di individui. Il fenomeno della paternità intellettuale multipla (*multiauthorship*) è, dal secondo dopoguerra in poi, un carattere distintivo della *Big Science*, della scienza che, alimentata dai soldi e dalle costose attrezzature della grande industria, progettava conquiste come lo sbarco sulla luna, la fissione dell'atomo, la mappatura del genoma umano. Imprese del genere erano concepibili solo all'interno di un modello organizzativo di rigida divisione del lavoro intellettuale, al cospetto del quale impallidiva il mito romantico dell'uomo di genio che produce idee, le mette alla prova dell'esperienza in un laboratorio domestico e pubblica i risultati per rendere beneficio ai colleghi e all'umanità intera. Le collaborazioni internazionali in occasione dei costosi esperimenti di fisica delle particelle e dei trial clinici su larga scala impongono una suddivisione dei compiti in tanti obiettivi, che vengono assegnati alle varie università o istituti di ricerca, i quali a loro volta li frantumano in micro-obiettivi assegnati a giovani ricercatori collocati nei gradini più bassi della scala gerarchica oppure a tecnici di laboratorio e clinici privi, in taluni casi, della minima consapevolezza circa le finalità teoriche del lavoro.³⁴⁷ Ciascuno di loro si aspetta sicuramente una quota di riconoscimento nella misura in cui ha contribuito a rendere possibile l'esperimento, ma la loro partecipazione in attività come l'ideazione, la progettazione, la revisione critica e la stesura finale del lavoro non è affatto scontata. Difficilmente quindi il loro apporto settoriale potrebbe essere chiamato in causa se, per caso, l'impostazione metodologica dell'articolo finale risultasse errata oppure se, davanti ad

does It take to write a research article? Authorship proliferation in neurosurgical research" *Neurosurgery* 2000, 47(2): 435-40.

un'accusa di frode, gli autori dell'articolo fossero chiamati in giudizio per difenderne l'integrità scientifica. Nemmeno l'atto che tradizionalmente connota l'"aura" della paternità intellettuale, quello della scrittura, può servire a ricomporre l'unità perduta perché nella scienza contemporanea «essere un autore non necessariamente significa essere uno scrittore»:³⁴⁸ gli articoli nascono spesso da forme di collaborazione massiva tra ricercatori di provenienza e attitudini settoriali diverse, la maggioranza dei quali riveste un ruolo insignificante nella redazione materiale del testo finale.

All'autore scientifico contemporaneo resta pertanto, in molti casi, una responsabilità intellettuale diluita, decentrata, il cui valore ai fini dell'analisi citazionale e del calcolo dell'*impact factor* non può essere semplicemente ricavato dalla posizione occupata in una lista: il primo nome elencato non corrisponde necessariamente ad una quota significativa di responsabilità ed è nota la consuetudine di riservare al direttore del dipartimento o della struttura, il cui ruolo effettivo in un progetto di ricerca può essere decisivo ma anche superficiale e puramente onorifico, l'ultima posizione nella lista dei coautori. Per questo motivo alcuni critici hanno proposto di abolire il concetto di "autore" per sostituirlo con quelli di "contributore" e "garante",³⁴⁹ mentre altri, in maniera più spicciola, hanno ipotizzato un conteggio frazionario della paternità intellettuale,³⁵⁰ oppure una rilevazione empirica delle singole quote di responsabilità per mezzo di questionari.³⁵¹

Manager della scienza e direttori editoriali sono perfettamente consapevoli delle difficoltà concettuali ed operative collegate al nuovo statuto della

³⁴⁷ M. Biagioli & P. Galison (Ed.), *Scientific Authorship: Credit and Intellectual Property in Science* (New York: Routledge, 2003).

³⁴⁸ B. Cronin, "Hyperauthorship: a postmodern perversion or evidence of a structural shift in scholarly communication practices?" *Journal of the American Society for Information Science and Technology* 2001, 52(7): 558-69, p. 567.

³⁴⁹ D. Rennie; V. Yank; L. Emanuel, "When authorship fails: a proposal to make contributors accountable" *JAMA* 1997, 287(7): 579-85.

³⁵⁰ G.B. Folly; B. Hajtman; J.I. Nagy et al., "Some metodological problems in ranking scientists by citation analysis" *Scientometrics* 1981, 3(2): 135-47.

³⁵¹ P. Vinkler, "Research contribution, authorship and team cooperativeness" *Scientometrics* 1993, 26: 213-30.

paternità intellettuale, ma hanno inquadrato il problema esclusivamente sotto l'aspetto amministrativo, tralasciando la definizione di protocolli di *authorship* adeguati alla nuova realtà. Di conseguenza, nei canali ufficiali della comunicazione scientifica, non solo l'autore è vivo e vegeto, nel pieno delle sue facoltà creative, ma la valutazione di un contributo resta indissolubilmente legata a quella dei nomi individuali elencati nella *byline*. Persino la finzione catalografica di un autore corporativo non trova spazio nella scienza, dove la responsabilità «è assoluta come la verità perché, al pari del credito scientifico, verità e responsabilità non possono essere quantificate».³⁵²

La stessa equivalenza è stabilita dagli editori. «Il credito autoriale - recitano le linee guida della International Committee of Medical Journal Editors (ICMJE) - dovrebbe basarsi su: 1) contributi sostanziali all'ideazione e progettazione, all'acquisizione o all'analisi e interpretazione dei dati; 2) stesura di una bozza dell'articolo o revisione critica del suo contenuto intellettuale rilevante; 3) approvazione della versione finale destinata alla pubblicazione [...] Quando il lavoro è condotto da un gruppo esteso e policentrico, il gruppo dovrebbe identificare gli individui che accettano la diretta responsabilità del manoscritto. Tali individui dovrebbero soddisfare pienamente i criteri per il credito autoriale appena definiti [...] Tutti i responsabili di contributi che non soddisfano i criteri per il credito autoriale dovrebbero essere elencati in una sezione di riconoscimenti».³⁵³

Se si prendono per buone queste linee guida, si deve anche apprezzarne la vena cinematografica: come in un film, che relega nei titoli di coda la lista dei tecnici che hanno preso parte alla realizzazione materiale delle singole scene, un articolo scientifico dovrebbe eclissare in una generica sezione di *acknowledgments*, cui è negata persino la dignità di chiave di ricerca nei database in linea, tutti coloro che, pur formando un anello della catena delle

³⁵² M. Biagioli, "The instability of authorship: credibility and responsibility in contemporary biomedicine" *The FASEB Journal* 1998, 12: 3-16, p. 6.

condizioni materiali di possibilità del lavoro, non hanno il controllo totale del processo. Una soluzione pragmatica e nemmeno tanto originale: qualcosa di simile accadeva nella filosofia sperimentale del XVII secolo, allorché tecnici di laboratorio, artigiani, ingegneri e fabbricanti di strumenti, pur collaborando attivamente alle dinamiche di produzione e convalida della conoscenza empirica, godevano di credibilità (quindi di visibilità) scientifica nulla a causa del loro basso status sociale.³⁵⁴

³⁵³ International Committee of Medical Journal Editors, *Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication*, 2004 <<http://www.icmje.org/>>.

³⁵⁴ S. Shapin, "The invisible technician" *American Scientist* 1989, 77(November-December): 554-63; M. Biagioli, *Galileo, courtier: the practice of science in the culture of absolutism* (Chicago: University of Chicago Press, 1993).

Conclusioni

La disponibilità, a partire dalla metà degli anni '60, di un *citation index* interdisciplinare è stata un'occasione unica per diverse categorie di professionisti: per gli *information scientists* desiderosi di sperimentare forme di indicizzazione automatica dei documenti alternative a quelle basate sulle parole; per storici e sociologi che cercavano indizi concreti delle dinamiche di formazione delle comunità di studiosi disciplinari e dei percorsi accidentati seguiti dalle idee e dalle teorie scientifiche nella loro evoluzione storica; per politici e manager della scienza impegnati a rivestire le proprie decisioni con argomenti di razionalità ed efficienza economica.

L'indicizzazione delle citazioni ha riproposto il mito dell'enciclopedismo rinascimentale in forma di compendio, una «visione integrata della letteratura scientifica non delimitata da confini disciplinari»,³⁵⁵ nella quale però, con il supporto delle leggi "troppo universali" della bibliometria, si sottintende che la letteratura scientifica è stratificata come la comunità di coloro che la producono. Non serve raccogliere tutto, è sufficiente limitarsi al "meglio" perché l'eccellenza è confinata in poche menti che vivono e lavorano in poche istituzioni e consegnano i frutti del proprio ingegno a pochi periodici internazionali.

Si può rifiutare tout court questa premessa liquidandola come «un problema di potere» e far notare, per esempio, che guarda caso i pochi periodici predestinati sono in gran parte gli stessi sui quali si esercita il controllo delle multinazionali dell'editoria,³⁵⁶ aziende che espropriano la comunità scientifica

³⁵⁵ «...an integrated view of the scientific literature that is unrestricted by disciplinary boundaries»: E. Garfield, *Citation Indexing: its Theory and Application...cit.*, p. 19.

³⁵⁶ A. Figà-Talamanca, "Uso e abuso delle banche dati nella valutazione delle riviste e della qualità scientifica: un problema di potere," in *Partecipare la scienza*, edited by A. Valente, et al. (Roma: Biblink,

dei risultati del proprio lavoro per poi rivenderli agli stessi scienziati e alle biblioteche soffocate dai costi sempre più esorbitanti degli abbonamenti. Questo però è solo un aspetto del problema, i quesiti che accompagnano gli usi extra-bibliografici delle banche dati restano: c'è un modo per valutare la qualità scientifica più affidabile della tradizionale *peer review* o un sistema per migliorare la *peer review* che non sia tanto ingenuo da far affidamento su una riforma morale degli individui? In che misura la qualità si identifica con l'impatto e in che misura l'impatto può essere studiato con metodi bibliometrici? Chi decide i criteri, gli standard e il modo in cui vanno applicati? Chi giudica i giudicanti? In che misura la produzione scientifica in tempi di "scienza normale" ha bisogno del regime di oligopolio editoriale per far emergere differenze qualitative altrimenti inattingibili ma necessarie a preservare la stratificazione sociale delle comunità scientifiche? In che misura gli indicatori bibliometrici giocano pro o contro un sistema di relazioni di potere e interessi precostituiti?

È evidente che un soggetto ordinario come la citazione bibliografica non è il candidato ideale per rispondere a domande che coinvolgono lo statuto epistemologico delle discipline scientifiche. Il percorso seguito dalla citazione per arrivare a recitare su un palcoscenico così prestigioso è comunque molto istruttivo, anche perché l'incapacità di rispondere a tali interrogativi non ha pregiudicato lo sviluppo di soluzioni tecniche efficaci in contesti diversi e più limitati.

Abbiamo così trovato la citazione bibliografica in prima fila quando, in seguito alla rivoluzione comunicativa innescata dalle reti telematiche, la disponibilità in linea di una massa enorme di documentazione in formato digitale ha posto il problema di facilitarne la reperibilità attraverso sistemi automatici di indicizzazione. Produttori e aggregatori di database e riviste elettroniche hanno sfruttato ampiamente la capacità delle citazioni

bibliografiche di legare in modo rapido ed economico due documenti di contenuto affine. Un'operazione che ha potenziato la reperibilità e al tempo stesso la commerciabilità dell'informazione digitale.

La citazione è solo un piccolo ingranaggio del sistema, ma ha indicato chiaramente la via da percorrere, l'idea cioè che il controllo bibliografico universale non può essere ottenuto attraverso una catalogazione manuale. Piuttosto è il documento stesso che, nell'atto di nascita, deve contenere le istruzioni necessarie per essere cercato e trovato da qualcuno, dal suo lettore. In questa direzione, la costruzione automatica di archivi disciplinari e multidisciplinari integrati, esemplificata da CiteSeer e CiteBase, porta a compimento una mutazione genetica decisiva: il documento "cofanetto", al cui interno sono contenuti i nessi tematici e bibliografici con le altre unità documentarie, viene rimpiazzato dal documento "in relazione", già materialmente collegato alla rete di tutti gli altri documenti che hanno una qualche forma di affinità con esso (argomento, paternità intellettuale, rapporto citato-citante, citazioni condivise).

Le reti telematiche promettono anche qualcosa in più ai sociologi e agli storici della scienza: mentre nel circuito convenzionale della comunicazione scientifica, dominato dalla rivista *peer reviewed*, il lettore è messo di fronte al prodotto finito, alla scienza già fatta, ora si intravede l'opportunità di capire qualcosa in più circa la "scienza in azione", ovvero l'attività socio-cognitiva che precede la costruzione dei fatti scientifici.³⁵⁷ Internet infatti ha potenziato e reso tangibili molti dei percorsi informali seguiti dai "collegi invisibili" di scienziati per elaborare, trasmettere, organizzare, convalidare i risultati della ricerca. In un futuro che è già cominciato, i modelli di analisi e valutazione collaudati nella letteratura periodica sono destinati perciò ad estendersi a

Daniele Archibugi nello stesso volume (pp. 37-49).

³⁵⁷ M.J. Bossy, "The last of the litter : "Netometrics"" *Solaris* 1995, 2 <<http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d02/2bossy.html>>.

circuiti informali di comunicazione come siti web personali e istituzionali, mailing list, forum di discussione, blog, reti P2P di file-sharing.

Salire alla ribalta ha significato, per la citazione bibliografica, smettere di improvvisare e cominciare a recitare secondo un copione. Nel momento in cui è entrata in un gioco più grande di lei, essa ha visto moltiplicarsi tanto gli appelli verso un uso più responsabile degli apparati bibliografici quanto le condanne per usi impropri e fraudolenti. In compenso, le esigenze di fotoriproducibilità e la sottomissione ad un'epistemologia empirista ne hanno appiattito forma e capacità espressive. Esiliata dal corpo del testo, privata del supporto critico delle note discorsive, sacrificata alle istanze di linearità espositiva imposte dallo *Harvard Style*, il suo rapporto con il testo è diventato più equivoco, al punto da indurre i sociologi costruttivisti a ritenere che, in linea di principio, si può citare qualunque documento a supporto di una tesi così come, data una qualunque teoria, si può trovare un set appropriato di dati empirici che la confermano. Un'ambiguità pervasiva, che coinvolge inevitabilmente il testo circostante.

L'articolo scientifico infatti, fonte numero uno dei dati bibliometrici, è un dispositivo nel quale l'uso del linguaggio e di apparati ausiliari come le note e la bibliografia possono essere compresi pienamente solo in relazione a quattro contesti, ciascuno dei quali presenta un grado diverso di opacità: 1) l'oggetto di studio (la realtà naturale o storico-sociale); 2) la letteratura del campo di ricerca in cui il testo si colloca (la tradizione, il paradigma scientifico dominante); 3) il pubblico dei potenziali fruitori (la comunità disciplinare degli scienziati che operano nella stessa area); 4) la mente dell'autore. Le operazioni di quest'ultima, virtualmente inaccessibili, sono all'origine della mediazione tra le altre sfere da cui risulta il prodotto finito, l'articolo pubblicato, sezionato, contabilizzato dagli scientometri. I quattro contesti sono interdipendenti: la percezione che l'autore e l'audience hanno dell'oggetto di studio è plasmata in gran parte dalla lettura della medesima letteratura

disciplinare, la quale a sua volta fornisce le definizioni socialmente condivise dell'oggetto di studio; l'autore scientifico, a sua volta, non scrive nel vuoto o avendo in mente solo la corrispondenza del testo con l'oggetto di studio, ma scrive e sceglie le parole (e le citazioni) da un repertorio non infinito di combinazioni che trova nella letteratura assimilata e rivolgendosi a quello che egli percepisce come il pubblico dei potenziali lettori. «Un testo, in un certo senso, è una soluzione al problema di come fare un'affermazione in modo da prestare l'attenzione appropriata a tutti e quattro i contesti contemporaneamente. In maniera più esplicita, l'articolo è una risposta alla domanda: "Sullo sfondo della conoscenza accumulata nella disciplina, come posso presentare una rivendicazione originale su un fenomeno all'audience appropriata in maniera tanto convincente da modificarne di conseguenza il modo di pensare e il comportamento"?».³⁵⁸

Una citazione bibliografica non necessariamente implica l'impatto o l'influenza di un autore su un altro perché, lo abbiamo visto, ci sono sempre troppi e contrastanti motivi per citare o non citare. Questo non significa però che gli stili citazionali degli scienziati siano anarchici. Il fatto che, interrogando un *citation index*, si trovino risultati spesso pertinenti è un sintomo della relativa stabilità del contenuto cognitivo delle citazioni (le citazioni come simboli di concetti di Small) nonché del grado di adesione degli scienziati alla norma che impone di riconoscere e ricompensare, con la moneta spicciola della citazione, il contributo di un predecessore (le citazioni atomi di *peer recognition* di Merton).

Un suggerimento ad inquadrare meglio il problema viene dall'abbondanza di avverbi frequentativi che si usano di solito per caratterizzare, in un senso o nell'altro, le abitudini citazionali degli scienziati: "spesso" gli scienziati citano per riconoscere, "frequentemente" citano per omaggiare, "talvolta" citano in malafede e così via. Sono avverbi cui si ricorre quando, in una linea

³⁵⁸ C. Bazerman, "What written knowledge does: Three examples of academic discourse" *Philosophy of the*

argomentativa, non si vuole o non si è in grado di stabilire "quanto", "quante volte" una cosa è così anziché in un altro modo. Ma è proprio qui il punto. "Quante volte" è così non può essere stabilito enumerando uno ad uno i casi singoli, perlomeno non in questo caso: bisognerebbe leggere tutti gli articoli, collegare ogni citazione alla frase che la giustifica e, nell'ipotesi che il senso della frase sia perfettamente trasparente, confrontare ciascuna coppia citazione/contesto con un modello ideale nel quale si cita il documento giusto al posto giusto, quindi trarre le debite conclusioni in merito alla natura normativa o costruttiva della singola citazione. Uno scenario vicino più ai racconti di Borges che alle reali potenzialità della scientometria, la cui unica via di scampo è quella di prendere un campione il più possibile casuale di citazioni, studiarne le relazioni quantitative e confidare nella solidità statistica dei calcoli effettuati. Presa da sola, una citazione non ha nulla da dire, l'unico modo per darle un senso è moltiplicarne le ricorrenze, raggrupparla con altre citazioni, tessere una rete.³⁵⁹ La possibilità di usare le citazioni per studiare il progresso della scienza e le dinamiche della comunicazione scientifica dipende quindi, in ultima analisi, da considerazioni di natura stocastica: selezionando accuratamente il campione e aumentando il volume dei dati di partenza, aumenta, su base probabilistica, la precisione e l'attendibilità dei risultati. Garfield stesso e i suoi collaboratori lo hanno sottolineato più volte: gli assunti dell'analisi citazionale sono "deboli". Il primo e più importante può essere espresso nella forma: "Se un documento **X** cita un documento **Y** è più probabile che il contenuto di **X** sia correlato a quello di **Y** piuttosto che a quello di un documento non citato".³⁶⁰

Le stesse considerazioni "deboli" valgono per le misure di impatto e qualità del lavoro degli scienziati desunte dal conteggio delle citazioni. A parte gli

Social Sciences 1981, 11: 361-87, pp. 363-364.

³⁵⁹ P. Wouters, "The signs of science" *Scientometrics* 1998, 41(1-2): 225-41.

³⁶⁰ B.C. Griffith; C.M. Drott; H. Small, "On the use of citations in studying scientific achievement and communication," in *Essays of an Information Scientist*, 3, edited by E. Garfield (Philadelphia: ISI Press, 1980): 234-46, p. 234. Il saggio era stato pubblicato originariamente sui *Current Contents* nel 1977.

slogan propagandistici iniziali, che esaltavano il potenziale impiego dello *Science Citation Index* nell'amministrazione della scienza oltre che nella ricerca bibliografica, Garfield ed i suoi epigoni sono stati tra i primi ad avvertire dei pericoli e dei vizi metodologici insiti in un uso troppo spregiudicato, nella valutazione dei ricercatori, di indicatori rudimentali di qualità come l'*impact factor* delle riviste o il conteggio puro e semplice delle citazioni ricevute dal singolo scienziato. Nello stesso tempo però, l'intero impianto teorico ed organizzativo dei *citation indexes* poggia su premesse che hanno un forte connotato valutativo e che lo stesso conteggio delle citazioni, unitamente alle leggi iperboliche della bibliometria (Bradford, Lotka, Zipf), al modello storiografico di Derek Price e alla sociologia funzionalista di Robert Merton, contribuisce ad avallare: da quando esiste il sistema disciplinare della scienza, un piccolo numero di scienziati ed un limitato numero di riviste scientifiche sono sufficienti a rendere conto della maggior parte di ciò che viene pubblicato e della stragrande maggioranza di ciò che viene citato in ciascuna area disciplinare. I fronti di ricerca attivi si sviluppano attorno a "collegi invisibili" composti da gruppi ristretti di ricercatori che l'analisi dei cluster di co-citazioni permette di identificare, mappare e rendere visibili, quindi amministrativamente e politicamente controllabili. La storia citazionale della scienza, del resto, lo ribadisce: se, da un lato, è vero che molte pubblicazioni supercite provengono da scienziati di media levatura, è anche vero che i premi Nobel e le personalità di spicco sul piano scientifico a livello internazionale pubblicano con maggiore frequenza e le loro pubblicazioni sono citate molto più della media. Questo principio ha trovato un'eco persino nell'universo in espansione del World Wide Web, che i fisici statistici hanno scoperto essere in fin dei conti un mondo piccolo, con la maggior parte dei link addensati attorno a pochi nodi o hub.

L'architettura del portale *ISI Web of Knowledge*, che può essere considerata la versione postmoderna dello *Science Citation Index*, riflette a pieno la struttura

elitaria della scienza propagandata dagli indici di citazioni. Al suo interno la ricerca bibliografica è solo il primo di una serie di passi destinati a produrre un output valutativo. Le unità bibliografiche che lo compongono sono battezzate con un indice scientometrico, di conseguenza si cercano articoli di riviste, relazioni di convegni, brevetti e nell'arco della stessa sessione si apprende la loro posizione relativa in una classifica di *most highly cited*. Le critiche feroci che molti osservatori hanno indirizzato alle citazioni come indicatore di qualità scientifica puntano alla debolezza metodologica della bibliometria valutativa e alla insostenibilità delle premesse teoriche su cui si fonda l'analisi delle citazioni. A prescindere dalla loro attendibilità, gli attacchi frontali azzerano la riflessione sugli indici bibliometrici e sul loro contributo alla revisione delle procedure di *peer reviewing* e dei criteri di reclutamento e promozione dei ricercatori. Il rifiuto totale degli indicatori bibliometrici rischia così di legittimare lo stato di cose esistenti, lasciando inalterate le dinamiche locali di gestione e redistribuzione del potere accademico, spesso fondate su reti informali di relazioni e accordi personali tra i "veterani". In campo umanistico, dove i limiti di applicabilità delle valutazioni bibliometriche sono ancora più evidenti, logiche di questo tipo si materializzano in repertori di pubblicazioni "ad hoc" di livello mediocre, tipicamente autofinanziate (*vanity press*), la cui storia citazionale piena di silenzi parlerebbe molto di più se vi fossero repertori adeguati ad accertarne, seppure in forma debole e probabilistica, l'impatto cognitivo. Si può decidere di non giocare al gioco della citazione, ma si deve anche essere consapevoli che il gioco ha delle regole più complesse di quelle che frettolosamente si vuole demolire e che giocando si possono far emergere aspetti inediti dell'organizzazione (e della dis-organizzazione) del lavoro scientifico.

Bibliografia

- Adair, W.C. (1955) "Citation indexes for scientific literature?" *American Documentation* 6: 31-32
<<http://www.garfield.library.upenn.edu/papers/adaircitationindexesforscientificliterature1955.html>>.
- Adam, D. (2002) "The counting house" *Nature* 415(6873): 726-9.
- Adamic, L. (1999) *The small world web. Proceedings of the 3rd European Conference of Research and Advanced Technology for Digital Libraries, ECDL*
<<http://www.hpl.hp.com/research/idl/papers/smallworld/smallworldpaper.html>>.
- Aharonian, G. (2000) "U.S. patent examination system is intellectually corrupt" *Patnews* (1 may)
<<http://www.questel.orbit.com/Piug/piugl2000/0523.html>>.
- Albert, M.B.; Avery, D.; Narin, F.; McAllister, P.R. (1991) "Direct validation of citation counts as indicators of industrially important patents" *Research Policy* 20(3): 251-59.
- Albert, R.; Barabási, A.L. (2002) "Statistical mechanics of complex networks" *Reviews of Modern Physics* 74(1): 47-97.
- Albert, R.; Jeong, H.; Barabási, A.L. (1999) "Diameter of the World Wide Web" *Nature* 401: 130-31
<<http://www.nd.edu/~networks/Papers/401130A0.pdf>>.
- Almind, T.; Ingwersen, P. (1997) "Informetric analysis of the World Wide Web: Methodological approaches to "Webometrics"" *Journal of Documentation* 53: 404-26.
- Altmann, K.G.; Gorman, G.E. (1996) "Usage, citation analysis and costs as indicators for journal deselection and cancellation: a selective literature review" *Australian Library Review* 13(4): 379-92.
- American Psychological Association (2001), ed. 5. *Publication manual of the American Psychological Association* (Washington, DC: APA).
- Antelman, K. (2004) "Do Open Access articles have a greater research impact?" *College & Research Libraries* 65(5): 372-82 <http://eprints.rclis.org/archive/00002309/01/do_open_access_CRL.pdf>.
- Bae, S.; Choi, J.H. (2000) "Cyberlinks between human rights NGOs: A network analysis" *58th Annual National Meeting of the Midwest Political Science Association*, Chicago, 2000, April.
- Baeza-Yates, R.; Castillo, C. (2001) "Relating Web characteristics with link based Web page ranking" *Proceedings of String Processing and Information Retrieval*, Laguna San Rafael, Chile, November 2001.
- Bailey, C.W. (2005) *Open Access Bibliography* (Washington, DC: Association of Research Library)
<<http://info.lib.uh.edu/cwb/oab.pdf>>.
- Barabási, A.L. (2004) *Link: la nuova scienza delle reti* (Torino: Einaudi).
- Barbieri, M.; Conti, G. (2005) "La ricerca documentale dei brevetti" *Biblioteche Oggi* 23(3): 15-18.
- Bar-Hillel, Y. (1951) "The present state of research on mechanical translation" *American Documentation* 2(4): 229-37.
- Bar-Ilan, J. (2001) "Data collection methods on the Web for informetric purposes: a review and analysis" *Scientometrics* 50(1): 7-32.
- (1998) "The mathematician, Paul Erdos (1913-1996) in the eyes of the Internet" *Scientometrics* 43(2): 257-67.
- Barnett, G.A.; Chon, B.; Park, H.W.; Rosen, D. (2001) "An examination of international Internet flows: An autopoietic model" *Annual Conference of International Communication Association*, Washington, D.C., 2001, May.
- Barthes, R. (1988) "La morte dell'autore" in *Il brusio della lingua. Saggi critici IV* (Torino: Einaudi): 51-54.
- Bazerman, C. (1991) *Le origini della scrittura scientifica: come è nata e come funziona l'argomentazione del saggio sperimentale* (Ancona; Bologna: Transeuropa).
- (1981) "What written knowledge does: Three examples of academic discourse" *Philosophy of the Social Sciences* 11: 361-87.
- Benedetti, C. (1999) *L'ombra lunga dell'autore. Indagine su una figura cancellata* (Milano: Feltrinelli).
- Bensam, J. (1988) "The aesthetics and politics of footnoting" *International Journal of Politics, Culture and Society* 1(3): 443-70.
- Bensman, S.J. (2001) "Urquhart's and Garfield's Laws: The British controversy over their validity" *Journal of the American Society for Information Science and Technology* 52(9): 714-24.
- Berkencotter, C.; Huckin, T.N. (1993) "You are what you cite: Novelty and intertextuality in a biologist's experimental article" in *Professional Communication: The Social Perspective*, edited by N.R. Blyler & C. Thralls (Newbury Park, CA: Sage): 107-27.

- Bernal, J.D. (1983) "Review of the *Science Citation Index*" in *Essays of an Information Scientist 1981-1982*, 5 (Philadelphia: ISI Press): 521-23 <<http://www.garfield.library.upenn.edu/essays/v5p511y1981-82.pdf>>.
- (1967) *The Social Function of Science* (Cambridge, MA: M.I.T. Press).
- Biagioli, M. (1993) *Galileo, courtier: the practice of science in the culture of absolutism* (Chicago: University of Chicago Press).
- (1998) "The instability of authorship: credibility and responsibility in contemporary biomedicine" *The FASEB Journal* 12: 3-16.
- Biagioli, M.; Galison, P. (Ed.) (2003) *Scientific Authorship: Credit and Intellectual Property in Science* (New York: Routledge).
- Bianconi, G.; Barabási, A.L. (2001) "Competition and multiscaling in evolving networks" *Europhysics Letters* 54: 436-42.
- Björneborn, L. (2004) *Small-world link structures across an academic web space: a library and information science approach*. Ph.d. Dissertation (Copenhagen: Royal School of Library and Information Science Department of Information Studies) <<http://www.db.dk/lb/phd/phd-thesis.pdf>>.
- Björneborn, L.; Ingwersen, P. (2004) "Toward a basic framework for webometrics" *Journal of the American Society for Information Science and Technology* 55(14): 1216-27.
- Bloch, S.; Walter, G. (2001) "The Impact Factor: time for change" *The Australian and New Zealand Journal of Psychiatry*. 35(5): 563-68.
- BOAI (2002) *Budapest Open Access Initiative* <<http://www.soros.org/openaccess/read.shtml>>.
- Bonaccorsi, A.; Daraio, C. (2004) "Econometric approaches to the analysis of productivity of R&D systems" in *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems*, edited by H.F. Moed, W. Glänzel, et al. (Dordrecht: Kluwer Academic): 51-74.
- Bookstein, A. (1990) "Informetric distributions, Part I: Unified Overview" *Journal of the American Society for Information Science* 41(5): 368-75.
- (1990) "Informetric distributions, Part II: Resilience to ambiguity" *Journal of the American Society for Information Science* 41(5): 376-86.
- Borgman, C.L.; Furner, J. (2002) "Scholarly communication and bibliometrics" in *Annual Review of Information Science and Technology*, edited by B. Cronin (Medford, NJ: Information Today).
- Börner, K.; Chen, C.; Boyack, K.W. (2003) "Visualizing knowledge domains" *Annual Review of Information Science and Technology* 37: 179-255.
- Bossy, M.J. (1995) "The last of the litter : "Netometrics"" *Solaris 2* <<http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d02/2bossy.html>>.
- Bourdieu, P. (2003) *Il mestiere di scienziato: corso al Collège de France 2000-2001* (Milano: Feltrinelli).
- Bourke, P.; Butler, L. (1996) "Publication types, citation rates and evaluation" *Scientometrics* 37(3): 473-94.
- Bovi, M. (2004) *Anche Mozart copiava. Cover, somiglianze, plagi e cloni* (Milano: Auditorium).
- Boyack, K.W.; Wylie, B.N.; Davidson, G.S. (2002) "Domain visualization using VxInsight for science and technology management" *Journal of the American Society for Information Science and Technology* 53(9): 764-74.
- Bradford, S.C. (1950) *Documentation* (Washington, D.C.: Public Affairs Press).
- (1934) "Sources of information on specific subjects" *Engineering* 137: 85-86.
- Brand, A. (2001) "CrossRef turns one" *D-Lib Magazine* 7(5) <<http://www.dlib.org/dlib/may01/brand/05brand.html>>.
- Brenno, E.; Fava, G.A.; Guardabasso, V. (2002) *La ricerca scientifica nelle università italiane. Una analisi delle citazioni della banca dati ISI* (Roma: CRUI).
- Brin, S.; Page, L. (1998) "The anatomy of a large-scale hypertextual Web search engine" *Computer Networks and ISDN Systems* 30(1-7): 107-17.
- Broadus, R.N. (1987) "Early approaches to bibliometrics" *Journal of the American Society for Information Science* 38(2): 127-29.
- Broder, A.; Kumar, R.; Maghoul, F.; Raghavan, P.; Rajagopalan, S.; Stata, R.; Tomkins, A.; Wiener, J. (2000) "Graph structure in the Web" *9th International World Wide Web Conference*, May 2000 <<http://www.almaden.ibm.com/cs/k53/www9.final>>.
- Brookes, B.C. (1990) "Biblio-, sciento-, infor-metrics??? What are we talking about" in *Informetrics 89/90: Selection of papers submitted for the Second International Conference on Bibliometrics, Scientometrics and Informetrics*, edited by L. Egghe & R. Rousseau (Amsterdam: Elsevier Science): 31-43.
- (1977) "Theory of the Bradford law" *Journal of Documentation* 33(3): 173-250.
- Brooks, T.A. (1985) "Private acts and public objects: an investigation of citer motivations" *Journal of the American Society for Information Science* 36(4): 223-29.
- Brunn, S.; Dodge, M. (2001) "Mapping the 'Worlds' of the World-Wide Web: (Re)structuring global commerce through hyperlinks" *American Behavioral Scientist* 44(10): 1717-39.

- Bucchi, M. (2002) *Scienza e società* (Bologna: Il Mulino).
- Burnham, J.C. (1990) "The evolution of editorial peer review" *JAMA* 263(10): 1323-32.
- Callon, M.; Law, J.; Rip, A. (1986) "Qualitative scientometrics" in *Mapping the Dynamics of Science and Technology*, edited by M. Callon, J. Law, et al. (London: The Macmillan Press): 103-23.
- Cameron, R.D. (1997) "A universal citation database as a catalyst for reform in scholarly communication" *First Monday* 2(4) <http://www.firstmonday.dk/issues/issue2_4/cameron/index.html>.
- Campbell, F. (1896) *The Theory of the National and International Bibliography: with Special Reference to the Introduction of System in the Record of Modern Literature* (London: Library Bureau).
- Cano, V. (1989) "Citation behavior: classification, utility, and location" *Journal of the American Society for Information Science* 40(4): 284-90.
- Carnap, R. (1971) *I fondamenti filosofici della fisica: Introduzione alla filosofia della scienza* (Milano: Il Saggiatore).
- Cason, H.; Lubotsky, M. (1936) "The influence and dependence of psychological journals on each other" *Psychological Bulletin* 33: 95-103.
- Cawkell, A.E. (1998) "Checking research progress on 'image retrieval by shape-matching' using the Web of Science" *Aslib Proceedings* 50(2): 27-31.
- (2000) "Methods of information retrieval using Web of Science: pulmonary hypertension as a subject example" *Journal of Information Science* 26(1): 66-70.
- Chakrabarti, S.; Dom, B.; Gibson, D.; Kleinberg, J.; Kumar, S.R.; Raghavan, P.; Rajagopalan, S.; Tomkins, A. (1999) "Hypersearching the web" *Scientific American* (June): 54-60.
- Chubin, D.E. (1983) *Sociology of Sciences: An Annotated Bibliography on Invisible Colleges, 1972-1981* (New York: Garland).
- Chubin, D.E.; Moitra, S.D. (1975) "Content analysis of references: adjunct or alternative to citation counting?" *Social Studies of Science* 5: 423-41.
- Cole, F.J.; Eales, N.B. (1917) "The history of comparative anatomy. Part I: A statistical analysis of the literature" *Science Progress* 11: 578-96.
- Cole, J.R. (2000) "A short history of the use of citations as a measure of the impact of scientific and scholarly work" in *The Web of Knowledge: a Festschrift in Honor of Eugene Garfield* (Medford, NJ: Information Today Inc): 281-300.
- Cole, J.R.; Cole, S. (1972) "The Ortega hypothesis" *Science* 178(4059): 368-75.
- (1973) *Social Stratification in Science* (Chicago: University of Chicago Press).
- Cole, S.; Cole, J.R. (1967) "Scientific output and recognition: a study in the operation of the reward system in science" *American Sociological Review* 32(3): 377-90.
- Cole, S.; Meyer, G.S. (1985) "Little Science Big Science revisited" *Scientometrics* 7(3-6): 443-58.
- Collins, P.; Wyatt, S. (1988) "Citations in patents to the basic research literature" *research Policy* 17(2): 65-74.
- Comba, V. (2002) *La valutazione delle pubblicazioni: dalla letteratura a stampa agli Open Archives* <<http://eprints.rclis.org/archive/00000095/01/valutazione.pdf>>.
- Comitato Telethon Fondazione Onlus (2004) *Bilancio di missione al 30 giugno 2004* <<http://www.telethon.it/telethon/allegati%5Cmissione2004.pdf>>.
- Courtial, J.P.; Callon, M. (1984) "Is indexing trustworthy? Classification of articles through co-word analysis" *Journal of Information Science* 9(2): 47-56.
- Cozzens, S.E. (1989) "What do citations count? The rhetoric-first model" *Scientometrics* 15: 437-47.
- Crane, D. (1972) *Invisible Colleges: Diffusion of Knowledge in Scientific Communication* (Chicago: University of Chicago Press).
- Cronin, B. (1984) *The Citation Process: the Role and Significance of Citations in Scientific Communications* (London: Taylor Graham).
- (2001) "Hyperauthorship: a postmodern perversion or evidence of a structural shift in scholarly communication practices?" *Journal of the American Society for Information Science and Technology* 52(7): 558-69.
- (2000) "Semiotics and evaluative bibliometrics" *Journal of Documentation* 56(4): 440-53.
- Cronin, B.; Snyder, H.; Rosenbaum, H.; Martinson, A.; Callahan, E. (1998) "Invoked on the Web" *Journal of the American Society for Information Science* 49(14): 1319-28.
- Cui, L. (1999) "Rating health web sites using the principles of citation analysis: a bibliometric approach" *Journal of Medical Internet Research* 1(1): E4 <<http://www.jmir.org/1999/1/e4/>>.
- Curry, M.R. (1991) "On the possibility of ethics in geography: writing, citing and the construction of intellectual property" *Progress in Human Geography* 15(2): 125-47.
- Davenport, E.; Cronin, B. (2000) "The citation network as a prototype for representing trust in virtual environments" in *The Web of Knowledge: a Festschrift in Honor of Eugene Garfield*, edited by B. Cronin & H.B. Atkins (Medford, NJ: Information Today Inc): 517-34.
- De Robbio, A. (2003) "Auto-archiviazione per la ricerca: problemi aperti e sviluppi futuri" *Bibliotime* 4(3) <<http://eprints.rclis.org/archive/00000180/03/OAI-20maggio2003.pdf>>.

- Dean, J.; Henzinger, M.R. (1999) "Finding related pages in the World Wide Web" in *Proceedings of the 8th International World Wide Web Conference. May 1999* (389-401
<<http://net.pku.edu.cn/~webg/papers/040309FindingRelatedpagesintheWeb.pdf>>.
- Dracos, A.; Cognetti, G. (1995) "La letteratura scientifica: indicatori bibliometrici e bibliografici quali criteri integrativi per la valutazione obiettiva di un'attività di ricerca" *Annali dell'Istituto Superiore di Sanità* 31(3): 381-90.
- Dumezil, G. (1989) *Feste romane* (Genova: Il Melangolo).
- Duncan, E.B. (1981) "Qualified citation indexing: its relevance to educational technology" in *Information Retrieval in Educational Technology. Proceedings of the First Symposium on Information Retrieval in Educational Technology, held at ETIC '81, Aberdeen, Scotland, 1st April, 1981*, edited by E.B. Duncan & R. McAleese (Aberdeen: University of Aberdeen): 70-79.
- Eco, U. (2001), ed. 12. *Come si fa una tesi di laurea. Le materie umanistiche* (Milano: Bompiani).
 ——— (1992) "Dell'impossibilità di costruire la carta dell'impero 1 a 1" in *Il secondo diario minimo* (Milano: Bompiani): 157-63.
 ——— (2003) *Dire quasi la stessa cosa. Esperienze di traduzione* (Milano: Bompiani).
 ——— (2002) *Sulla letteratura* (Milano: Bompiani).
- Edge, D.O. (1979) "Quantitative measures of communication in science: a critical review" *History of science* 17: 102-34.
 ——— (1977) "Why I am not a cocitazionist" *Society for the Social Studies of Science Newsletter* 2: 13-19.
- Egghe, L. (1985) "Consequences of Lotka's law for the law of Bradford" *Journal of Documentation* 41: 173-89.
 ——— (1995) "Extension of the general "success breeds success" principle to the case that items can have multiple sources" in *Proceedings of the 5th Biennial Conference of the International Society for Scientometrics and Informetrics*, edited by M. Koenig & A. Bookstein (Medford, N.J.: Learned Information): 147-56.
 ——— (2000) "New informetric aspects of the Internet: some reflections - many problems" *Journal of Information Science* 26(5): 329-35.
 ——— (2005) *Power Laws in the Information Production Process: Lotkaian Informetrics* (Amsterdam: Elsevier Academic Press).
- Egghe, L.; Rousseau, R. (1995) "Generalized success-breeds-success principle leading to time-dependent informetric distributions" *Journal of the American Society for Information Science* 46(6): 426-45.
 ——— (1990) *Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science* (Amsterdam: Elsevier).
- Elkana, Y.; Lederberg, J.; Merton, R.K.; Thackray, A.; Zuckerman, H. (Ed.) (1979) *Toward a Metric of Science: the Advent of Science Indicators* (New York: John Wiley and Sons).
- Fairthorne, R.A. (1969) "Empirical hyperbolic distributions (Bradford-Zipf-Mandelbrot) for bibliometric description and prediction" *Journal of Documentation* 25: 319-43.
- Fernandez-Cano, A.; Torralbo, M.; Vallejo, M. (2004) "Reconsidering Price's model of scientific growth: An overview" *Scientometrics* 61(3): 301-21.
- Feyerabend, P.K. (2003), ed. 2. *Contro il metodo: abbozzo di una teoria anarchica della conoscenza* (Milano: Feltrinelli).
- Figà-Talamanca, A. (2005) "Uso e abuso delle banche dati nella valutazione delle riviste e della qualità scientifica: un problema di potere" in *Partecipare la scienza*, edited by A. Valente & D. Luzi (Roma: Binklink): 27-35.
- Finocchiaro, M.A. (1980) *Galileo and the Art of Reasoning: Rhetorical Foundations of Logic and Scientific Method* (Dordrecht-Boston-London: Reidel).
- Folly, G.B.; Hajtman, B.; Nagy, J.I.; Ruff, I. (1981) "Some metodological problems in ranking scientists by citation analysis" *Scientometrics* 3(2): 135-47.
- Fortunati, V. (1988) "Citazione" *Quaderni di Filologia Germanica della Facoltà di Lettere e Filosofia dell'Università di Bologna* 4: 7-13.
 ——— (2002) "Intertestualità e citazione fra Modernismo e Postmodernismo" *Leitmotiv* 2: 87-96
 <<http://www.ledonline.it/leitmotiv/allegati/leitmotiv020208.pdf>>.
- Foucault, M. (1984) "Che cos'è un autore?" in *Scritti letterari* (Milano: Feltrinelli): 1-21.
- Frost, C. (1979) "The use of citations in literary research: preliminary classifications of citation functions" *Library Quarterly* 49(4): 399-414.
- Fujii, A.; Ishikawa, T. (2000) "A novelty-based evaluation method for information retrieval" in *Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC-2000), Athens, Greece 31.5.-2.6.2000* (Paris: European Language Resources Association): 1637-41
 <http://arxiv.org/PS_cache/cs/pdf/0011/0011002.pdf>.
- Furner, J. (2001) *Scholarly Communication and Bibliometrics* <<http://polaris.gseis.ucla.edu/jfurner/03-04/208/208res.html>>.

- Fussler, H.H. (1949) "Characteristics of the research literature used by chemists and physicists in the United States" *Library Quarterly* 19: 19-35.
- Garfield, E. (1981) "Bradford's Law and related statistical patterns" in *Essays of an Information Scientist 1979-1980*, 4 (Philadelphia: ISI Press): 476-83
<<http://www.garfield.library.upenn.edu/essays/v4p476y1979-80.pdf>>.
- (1977) "Can citation indexing be automated?" in *Essays of an Information Scientist 1962-1973*, 1 (Philadelphia: ISI Press): 84-90 <<http://www.garfield.library.upenn.edu/essays/V1p084y1962-73.pdf>>.
- (2002) "Citation consciousness. Interview with Eugene Garfield, Chairman Emeritus of I.S.I., Philadelphia, U.S.A." *Password* (June): 22-25
<<http://www.garfield.library.upenn.edu/papers/passwordinterview062002.pdf>>.
- (1977) "Citation impact depends upon the paper, not the journal! Don't count on 'citation by association!'" in *Essays of an Information Scientist 1962-1973*, 1 (Philadelphia: ISI Press): 452-53
<<http://www.garfield.library.upenn.edu/essays/V1p452y1962-73.pdf>>.
- (1956) "Citation indexes - new paths to scientific knowledge" *The Chemical Bulletin* 43(4): 11
<<http://www.garfield.library.upenn.edu/papers/31.html>>.
- (1983) "Citation Indexes for science: a new dimension in documentation through association of ideas" in *Essays of an Information Scientist*, 6 (Philadelphia: ISI Press): 468-71
<<http://www.garfield.library.upenn.edu/essays/v6p468y1983.pdf>>.
- (1955) "Citation Indexes for science: a new dimension in documentation through association of ideas" *Science* 122(3159): 108-11.
- (1977) "Citation indexes in sociological and historical research" in *Essays of an Information Scientist 1962-1973*, 1 (Philadelphia: ISI Press): 43-46
<<http://www.garfield.library.upenn.edu/essays/V1p043y1962-73.pdf>>.
- (1983) *Citation Indexing: its Theory and Application in Science, Technology, and Humanities* (Philadelphia: ISI Press).
- (1994) *Citation-based and descriptor-based search strategies*
<<http://www.isinet.com/essays/useofcitationdatabases/3.html>>.
- (1993) "Co-citation analysis of the scientific literature: Henry Small on mapping the collective mind of science" in *Essays of an Information Scientist 1992-1993. Of Nobel Class, Women in Science, Citation Classics and other Essays*, 15 (Philadelphia: ISI Press): 293-94
<<http://www.garfield.library.upenn.edu/essays/v15p293y1992-93.pdf>>.
- (1977) "Current Contents - Ninth Anniversary" in *Essays of an Information Scientist 1962-1973*, 1 (Philadelphia: ISI Press): 12-15 <<http://www.garfield.library.upenn.edu/essays/V1p012y1962-73.pdf>>.
- (1988) "Do Nobel Prize winners write citation classics?" in *Essays of an Information Scientist 1986, Towards Scientography*, 9 (Philadelphia: ISI Press): 182-87
<<http://www.garfield.library.upenn.edu/essays/v9p182y1986.pdf>>.
- (2001) "From computational linguistics to algorithmic historiography" *Lazerow Lecture held in conjunction with panel on "Knowledge and Language: Building Large-Scale Knowledge Bases for Intelligent Applications"*, University of Pittsburgh, September 2001
<<http://www.garfield.library.upenn.edu/papers/pittsburgh92001.pdf>>.
- (2003) "Historiographic mapping of knowledge domains literature" *Arthur M. Sackler Colloquium on "Mapping Knowledge Domains"*, Beckman Center of the National Academy of Sciences, Irvine, CA, May 9-11, 2003 <<http://www.garfield.library.upenn.edu/papers/knowledgedomain2003.pdf>>.
- (1977) "Historiographs, librarianship, and the history of science" in *Essays of an Information Scientist 1974-1976*, 2 (Philadelphia: ISI Press): 136-50
<<http://garfield.library.upenn.edu/essays/v2p136y1974-76.pdf>>.
- (1996) "How can impact factors be improved?" *BMJ* 313(7054): 411-3
<[http://www.garfield.library.upenn.edu/papers/bmj313\(7054\)p411y1996.html](http://www.garfield.library.upenn.edu/papers/bmj313(7054)p411y1996.html)>.
- (1981) "How do we select journals for *Current Contents*?" in *Essays of an Information Scientist 1979-1980*, 4 (Philadelphia: ISI Press): 309-12
<<http://garfield.library.upenn.edu/essays/v4p309y1979-80.pdf>>.
- (1984) "How to use citation analysis for faculty evaluations, and when is it relevant? Part 1" in *Essays of an Information Scientist 1983*, 6 (Philadelphia: ISI Press): 354-62.
- (1984) "How to use citation analysis for faculty evaluations, and when is it relevant? Part 2" in *Essays of an Information Scientist 1983*, 6 (Philadelphia: ISI Press): 363-72.
- (1998) "The impact factor and using it correctly" *Der Unfallchirurg* 48(2): 413
<[http://www.garfield.library.upenn.edu/papers/derunfallchirurg_v101\(6\)p413y1998english.html](http://www.garfield.library.upenn.edu/papers/derunfallchirurg_v101(6)p413y1998english.html)>.
- (1991) "In truth, the 'flood' of scientific literature is only a myth" *The Scientist* 5(17)
<[http://www.garfield.library.upenn.edu/commentaries/tsv05\(17\)y19910902.html](http://www.garfield.library.upenn.edu/commentaries/tsv05(17)y19910902.html)>.

- (2002) "The ISI-VINITI Connection - Remarks on the 50th anniversary of VINITI" in *Proceedings of the 6th International Conference of Information Society, Intelligent Information Processing and Information Technologies. October 16-18, 2002* (Moscow: 409-11 <<http://www.garfield.library.upenn.edu/papers/viniti101502.html>>).
- (1983) "J.D. Bernal—The sage of Cambridge. 4S Award memorializes his contributions to the social studies of science" in *Essays of an Information Scientist 1981-1982*, 5 (Philadelphia: ISI Press): 511-20 <<http://www.garfield.library.upenn.edu/essays/v5p511y1981-82.pdf>>.
- (1999) "Journal impact factor: a brief review" *Canadian Medical Association Journal* 161(8): 979-80 <<http://www.garfield.library.upenn.edu/papers/journalimpactCMAJ1999.pdf>>.
- (1986) "Journal selection for *Current Contents*: editorial merit vs. political pressure" in *Essays of an Information Scientist 1985. Ghostwriting and other Essays*, 8 (Philadelphia: ISI Press): 96-104 <<http://garfield.library.upenn.edu/essays/v8p096y1985.pdf>>.
- (1981) "Lifetime citation rates" in *Essays of an Information Scientist 1979-1980*, 4 (Philadelphia: ISI Press): 355-58 <<http://www.garfield.library.upenn.edu/essays/v4p355y1979-80.pdf>>.
- (1977) "The mystery of the transposed journal lists wherein Bradford's law of scattering is generalized according to Garfield's law of concentration" in *Essays of an Information Scientist 1962-1973*, 1 (Philadelphia: ISI Press): 222-23 <<http://www.garfield.library.upenn.edu/essays/V1p222y1962-73.pdf>>.
- (1977) "The mysteries of ISI's International OATS Library Service Revealed (Again)!" in *Essays of an Information Scientist 1974-1976*, 2 (Philadelphia: ISI Press): 55-57 <<http://www.garfield.library.upenn.edu/essays/v2p055y1974-76.pdf>>.
- (1976) *The Science Citation Index and ISI's Journal Citation Reports: their implications for journal editors. Presented at: Third General Assembly of the European Association of Editors of Biological Periodicals. Paris, 10—12 May 1976* <<http://www.garfield.library.upenn.edu/papers/255.html>>.
- (1996) "The significant scientific literature appears in a small core of journals" *The Scientist* 10(17): 13-16 <[http://www.garfield.library.upenn.edu/commentaries/tsv10\(17\)p13y090296.html](http://www.garfield.library.upenn.edu/commentaries/tsv10(17)p13y090296.html)>.
- (1959) "A unified index to science" in *Proceedings of the International Conference on Scientific Information*, 1 (Washington, D.C.: National Academy of Sciences - National Research Council): 461-74 <http://books.nap.edu/html/sci_inf_1959/461-474.pdf>.
- Garfield, E.; Pudovkin, A.I.; Istomin, V.S. (2002) "Algorithmic citation-linked historiography - Mapping the literature of science" *ASIS&T 2002: Information, Connections and Community. 65th Annual Meeting of ASIST*, Philadelphia, PA, November 18-21, 2002 <<http://garfield.library.upenn.edu/papers/asis2002/asis2002presentation.html>>.
- (2003) "Why do we need algorithmic historiography?" *Journal of the American Society for Information Science and Technology* 54(5): 400-12 <[http://www.garfield.library.upenn.edu/papers/jasist54\(5\)400y2003.pdf](http://www.garfield.library.upenn.edu/papers/jasist54(5)400y2003.pdf)>.
- Garfield, E.; Sher, I.H.; Thorpie, R.J. (1964) *The Use of Citation Data in Writing the History of Science* (Philadelphia: Institute for Scientific Information) <<http://www.garfield.library.upenn.edu/papers/useofcitdatawritinghistofsci.pdf>>.
- Gargiulo, P. (2004) "La comunicazione scientifica aperta della Sapienza" *Workshop sugli Open Archives*, Università di Roma La Sapienza, 10 giugno 2004 <<http://bids.citicord.uniroma1.it/padis/gargiulo10giugno04.pps>>.
- Garton, L.; Haythornthwaite, C.; Wellman, B. (1997) "Studying online social networks" *Journal of Computer-Mediated Communication* 3(1) <<http://www.ascusc.org/jcmc/vol3/issue1/garton.html>>.
- Geller, N.L.; De Cani, J.S. (1978) "Lifetime-citation rates to compare scientists' work" *Social Science Research* 7(4): 345-65.
- Gilbert, G.N. (1977) "Referencing as persuasion" *Social Studies of Science* 7: 113-22.
- Gilbert, G.N.; Mulkay, M. (1984) *Opening Pandora's Box: a Sociological Analysis of Scientists' Discourse* (Cambridge: Cambridge University Press).
- Glänzel, W. (2003) *Bibliometrics as a Research Field. A Course on Theory and Application of Bibliometric Indicators* <http://www.norslis.net/2004/Bib_Module_KUL.pdf>.
- Glänzel, W.; Schubert, A. (1985) "Price distribution. An exact formulation of Price's square root law" *Scientometrics* 7(3-6): 211-19.
- Gmür, M. (2003) "Co-citation analysis and the search for invisible colleges: A methodological evaluation" *Scientometrics* 57(1): 27-57.
- Godin, B. (2000) *Outline for a History of Science Measurement* <http://www.csiic.ca/PDF/Godin_1.pdf>.
- Godlee, I.; Jefferson, T. (Ed.) (2003) *Peer Review in Health Sciences* (London: BMJ Books).
- Google: *A hacker's best friend* (2005) <<http://www.awarenetwork.org/home/paris2k/P2KCEGOOGLE.pdf>>.
- Gosnell, C.F. (1944) "Obsolescence of books in college libraries" *College & Research Libraries* 5: 115-25.
- Grafton, A. (1994) "The footnote from De Thou to Ranke" *History and Theory Theme Issue* 33(4): 53-76.
- (2000) *La nota a piè di pagina: una storia curiosa* (Milano: Sylvestre Bonnard).

- Granovsky, Y.V. (2001) "Is it possible to measure science? V. V. Nalimov's research in scientometrics" *Scientometrics* 52(2): 127-50.
- Grasso, A.; Scaglioni, M. (2004) *Che cos'è la televisione* (Milano: Garzanti).
- Gregory, D. (1990) "Editorial" *Environment and Planning, D: Society and Space* 8(1): 1-6.
- Griffith, B.C.; Drott, C.M.; Small, H. (1980) "On the use of citations in studying scientific achievement and communication" in *Essays of an Information Scientist*, 3, edited by E. Garfield (Philadelphia: ISI Press): 234-46 <<http://www.garfield.library.upenn.edu/essays/v3p234y1977-78.pdf>>.
- Gross, P.L.K.; Gross, E.M. (1927) "College libraries and chemical education" *Science* 66(1713): 385-89.
- Gurjeva, L.G. (1992) *Early Soviet Scientometrics and Scientometricians*. Thesis for the degree of MSc in Science Dynamics (Universiteit van Amsterdam).
- Hahn, H.; Neurath, O.; Carnap, R. (1979) *La concezione scientifica del mondo: il Circolo di Vienna* (Roma; Bari: Laterza).
- Hahn, R. (1980) *A bibliography of quantitative studies on science and its history* (Berkeley: office for History of Science and Technology).
- Halavais, A. (2000) "National borders on the World Wide Web" *New Media & Society* 2(1): 7-28.
- Halavais, A.; Garrido, M. (2003) "Mapping networks of support for the Zapatista movement" in *Cyberactivism: Online activism in theory and practice*, edited by M. McCaughy & M.D. Ayers (London: Routledge).
- Hall, B.H.; Jaffe, A.; Trajtenberg, M. (2002) *Market value and patent citations* <http://emlab.berkeley.edu/users/bhhall/papers/HallJaffeTrajtenberg_RJEjan04.pdf>.
- Harley, J.B. (1988) "Maps, knowledge and power" in *The Iconography of Landscape: essays on the Symbolic Representation, Design and Use of Past Environments*, edited by D. Cosgrove & S. Daniels (Cambridge: Cambridge University Press): 277-312.
- Harnad, S.; Brody, T. (2004) "Comparing the impact of open access (oa) vs. non-oa articles in the same journals" *D-Lib Magazine* 10(6) <<http://www.dlib.org/dlib/june04/harnad/06harnad.html>>.
- Harter, S.P.; Ford, C.E. (2000) "Web-based analysis of E-journal impact: Approaches, problems, and issues" *Scientometrics* 51(13): 1159-76.
- Harter, S.P.; Nisonger, T.E.; Weng, A. (1993) "Semantic relationships between cited and citing articles in library and information science journals" *Journal of the American Society for Information Science* 44(9): 543 - 52.
- Hartley, J. (1999) "What do we know about footnotes? Opinions and data" *Journal of Information Science* 25: 205-12.
- He, Y.; Hui, S.C.; Fong, A.C.M. (2003) "Citation-based retrieval for scholarly publications" *Intelligent Systems, IEEE* 18(2): 58-65.
- Hecht, F.; Hecht, B.K.; Sandberg, A.A. (1998) "The journal "impact factor": a misnamed, misleading, misused measure" *Cancer Genetics and Cytogenetics* 104: 77-81.
- Heifetz, S.R. (1999) "Blue in the face: the Bluebook, the Bar exam, and the paradox of our legal culture" *Rutgers Law Review* 51: 695-711.
- Henkle, H.H. (1938) "The periodical literature of biochemistry" *Bulletin of the Medical Library Association* 27: 139-47.
- Hertz, D.H. (1987) "Bibliometrics, history of the development of ideas" in *Encyclopedia of Library and Information Science*, 42, Supplement 7 (New York: M. Dekker): 144-211.
- Hicks, D. (1987) "Limitations of co-citation analysis as a tool for science policy" *Social Studies of Science* 17: 295-316.
- Hjerpe, R. (1980) *A Bibliography of Bibliometrics and Citation Indexing & Analysis* (Stockholm: Royal Institute of Technology).
- Holton, G. (1979) "Can science be measured?" in *Toward a Metric of Science: the Advent of Science Indicators*, edited by Y. Elkana, J. Lederberg, et al. (New York: John Wiley and Sons): 39-68.
- Hood, W.W.; Wilson, C.S. (2001) "The literature of Bibliometrics, Scientometrics, and Informetrics" *Scientometrics* 52(2): 291-314.
- Huang, X.; Peng, F.; An, A.; Schuurmans, D. (2004) "Dynamic web log session identification with statistical language models" *Journal of the American Society for Information Science and Technology* 55(14): 1290-303.
- Hui, S.C.; Fong, A.C.M. (2004) "Document retrieval from a citation database using conceptual clustering and co-word analysis" *Online Information Review* 28(1): 22-32.
- Hulme, E.W. (1923) *Statistical Bibliography in Relation to the Growth of Modern Civilization* (London: Grafton).
- Hummon, N.P.; Doreian, P. (1989) "Connectivity in a citation network: the development of DNA theory" *Social Networks* 11: 39-63 <<http://www.garfield.library.upenn.edu/papers/hummondoreian1989.pdf>>.
- Ingwersen, P. (1998) "The calculation of Web impact factors" *Journal of Documentation* 54(2): 236-43.

- (1996) "Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory" *Journal of Documentation* 52(1): 3-50.
- (1992) *Information Retrieval Interaction* (London: Taylor Graham) <http://www.db.dk/pi/iri/files/Ingwersen_IRI.pdf>.
- International Committee of Medical Journal Editors (2004) *Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication* <<http://www.icmje.org/>>.
- Italia. Ministero della Salute. Direzione Generale Ricerca Sanitaria e Vigilanza Enti (2002) *Circolare. Oggetto: Ricerca corrente 2002, 2003, 2004 - Acquisizione elementi ai fini della ripartizione* <http://www.ministerosalute.it/resources/static/legis2002/Circolare_RC.pdf>.
- Jacsó, P. (2004) "Citation-enhanced indexing/abstracting databases" *Online Information Review* 28(3): 235-38.
- (2004) "Savvy searching. Citation searching" *Online Information Review* 28(6): 454-60.
- Kaplan, N. (1965) "The norms of citation behavior: prolegomena to the footnote" *American Documentation* 16(3): 179-84.
- Karki, M.M.S. (1997) "Patent citation analysis: A policy analysis tool" *World Patent Information* 19(4): 269-72.
- Kessler, M.M. (1963) "Bibliographic coupling between scientific papers" *American Documentation* 14(January-February): 10-25.
- (1963) "Bibliographic coupling extended in time: ten case histories" *Information Storage & Retrieval* 1: 169-87.
- King, J.T. (2000) "How many neurosurgeons does it take to write a research article? Authorship proliferation in neurosurgical research" *Neurosurgery* 47(2): 435-40.
- Knorr-Cetina, K.D.; Mulkay, M. (Ed.) (1983) *Science Observed. Perspectives in the Social Study of Science* (London: Sage).
- Koenig, M.E.D. (1983) "Bibliometric indicators versus expert opinion in assessing research performance" *Journal of the American Society for Information Science* 34(2): 136-45.
- Krebs, V. (2000) "Working in the connected world book network" *IHRIM - International Association for Human Resource Information Management Journal* 4(1): 87-90.
- Kretschmer, H.; Rousseau, R. (2001) "Author inflation leads to a breakdown of Lotka's law" *Journal of the American Society for Information Science and Technology* 52(8): 610-14.
- Kreuzman, H. (2001) "A co-citation analysis of representative authors in philosophy: Examining the relationship between epistemologists and philosophers of science" *Scientometrics* 51(3): 525-39.
- Kronick, D.A. (1962) *A History of Scientific and Technical Periodicals* (New York: Scarecrow Press).
- (1990) "Peer review in 18th-century scientific journalism" *JAMA* 263(10): 1321-2.
- Kuhn, T. (1978) *La struttura delle rivoluzioni scientifiche* (Torino: Einaudi).
- (1978) "Poscritto 1969" in *La struttura delle rivoluzioni scientifiche* (Torino: Einaudi): 209-51.
- Kuhn, T.S. (1985) "Nuove riflessioni sui paradigmi" in *La tensione essenziale. Cambiamenti e continuità nella scienza* (Torino: Einaudi): 321-50.
- Kumar, R.; Raghauam, P.; Rajagopalan, S.; Tomkins, A. (1999) "Trawling the Web for emerging cyber-communities" *Computer Networks* 31(11): 1481-93 <<http://www8.org/w8-papers/4a-search-mining/trawling/trawling.html>>.
- Landauer, T.K.; Laham, D.; Der, M. (2004) "From paragraph to graph: Latent semantic analysis for information visualization" *Proceedings of the National Academy of Sciences* 101(suppl 1): 5214-19.
- Lange, C. (2005) *L'angolo aureo* <<http://www.sectioaurea.com/sectioaurea/angoloaureo.htm>>.
- Larsen, B. (2004) *References and Citations in Automatic Indexing and Retrieval Systems - Experiments with the Boomerang Effect*. PhD dissertation (Copenhagen: Department of Information Studies, Royal School of Library and Information Science) <<http://www.db.dk/blar/dissertation>>.
- Larson, R.R. (1996) "Bibliometrics of the World Wide Web: an exploratory analysis of the intellectual structure of cyberspace" in *Global Complexity: Information, Chaos and Control. Proceedings of the 59th ASIS Annual Meeting, Baltimore, MD, October 21-24 1996*, edited by S. Hardin (Medford, NJ: Information Today Inc) <<http://sherlock.berkeley.edu/asis96/asis96.html>>.
- Latour, B. (1998) *La scienza in azione. Introduzione alla sociologia della scienza* (Torino: edizioni di Comunità).
- Lawrence, S. (2001) "Free online availability substantially increases a paper's impact" *Nature* 411(6837): 521.
- Lawrence, S.; Giles, C.L.; Bollacker, K. (1999) "Digital libraries and autonomous citation indexing" *IEEE Computer* 32(6): 67-71 <<http://citeseer.ist.psu.edu/aci-computer/aci-computer99.html>>.
- Leach, E.R. (1981) *Cultura e comunicazione: la logica della connessione simbolica : un'introduzione all'uso dell'analisi strutturale nell'antropologia sociale* (Milano: Franco Angeli).
- Leydesdorff, L. (2004) *The Evaluation of Research and the Scientometric Research Program: Historical Evolution and Redefinitions of the Relationship* <<http://users.fmg.uva.nl/lleydesdorff/sss04/>>.

- (1998) "Theories of citation?" *Scientometrics* 43(1): 5-25 <Preprint in: <http://users.fmg.uva.nl/lleydesdorff/citation/>>.
- (1997) "Why words and co-words cannot map the development of the sciences" *Journal of the American Society for Information Science* 48(5): 418-27.
- Leydesdorff, L.; Curran, M. (2000) "Mapping university-industry-government relations on the Internet: the construction of indicators for a knowledge-based economy" *Cybermetrics* 4(1, paper 2) <<http://www.cindoc.csic.es/cybermetrics/articles/v4i1p2.html>>.
- Li, X. (2003) "A review of the development and application of the Web impact factor" *Online Information Review* 27(6): 407-17.
- Lin, X. (1997) "Map displays for information retrieval" *Journal of the American Society for Information Science* 48(1): 40-54.
- Lipetz, B.-A. (1965) "Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators" *American Documentation* 16(2): 81-90.
- Liu, M. (1993) "The complexities of citation practice: A review of citation studies" *Journal of Documentation* 49(4): 370-408.
- Lotka, A.J. (1926) "Statistics - The frequency distribution of scientific productivity" *Journal of Washington Academy of Sciences* 16(12): 317-25.
- 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: 265-75.
- Luukkonen, T. (1990) *Citations in the rhetorical, reward, and communication systems of science*. Ph.D. Thesis (Tampere: University of Tampere).
- (1997) "Why has Latour's theory of citations been ignored by the bibliometric community? Discussion of sociological interpretations of citation analysis" *Scientometrics* 38: 27-37.
- MacRoberts, M.H.; MacRoberts, B.R. (1989) "Problems of citation analysis: a critical review" *Journal of the American Society for Information Science* 40(5): 342-49.
- Madigan, R.; Johnson, S.; Linton, P. (1995) "The language of psychology. APA Style as epistemology" *American Psychologist* 50(6): 428-36.
- Madigan, R.; Linton, P.; Johnson, S. (1996) "APA Style: Quo vadis?" *American Psychologist* 51(6): 653-55.
- Malvano, A. (2003) *Voci da lontano. Schumann e l'arte della citazione* (Torino: EDT).
- Martin, B.R.; Irvine, J. (1983) "Assessing basic research. Some partial indicators of scientific progress in radio astronomy" *Research Policy* 12: 61-90.
- Martyn, J. (1965) "An examination of citation indexes" *Aslib Proceedings* 17(6): 184-96.
- Mastermann, M. (1979) "La natura di un paradigma" in *Critica e crescita della conoscenza*, edited by I. Lakatos & A. Musgrave (Milano: Feltrinelli): 129-63.
- McCain, K.W. (1986) "Cocited author mapping as a valid representation of intellectual structure" *Journal of the American Society for Information Science* 37: 111-22.
- (1989) "Descriptor and citation retrieval in the medical behavioral sciences literature: Retrieval overlaps and novelty distribution" *Journal of the American Society for Information Science* 40(2): 110-14.
- McVeigh, M.E. (2004) *Open Access journals in the ISI citation databases: Analysis of impact factors and citation patterns. A citation study from Thomson Scientific* <<http://www.isinet.com/media/presentrep/essayspdf/openaccesscitations2.pdf>>.
- Medawar, P.B. (1964) "Is the scientific paper fraudulent? Yes; it misrepresents scientific thought" *Saturday Review* (1 August): 42-43.
- Meho, L.I.; Sonnenwald, D.H. (2000) "Citation ranking versus peer evaluation of senior faculty research performance: A case study of Kurdish scholarship" *Journal of the American Society for Information Science* 51(2): 123-28.
- Menczer, F. (2004) "Lexical and semantic clustering by web links" *Journal of the American Society for Information Science and Technology* 55(14): 1261-69.
- Merton, R.K. (1981) "La priorità nella scoperta scientifica (1957)" in *La sociologia della scienza. Indagini teoriche ed empiriche*, edited by N.W. Storer (Milano: Franco Angeli): 371-414.
- (1981) "La struttura normativa della scienza (1942)" in *La sociologia della scienza. Indagini teoriche ed empiriche*, edited by N.W. Storer (Milano: Franco Angeli): 349-62.
- (1988) "The Matthew effect in science, 2: cumulative advantage and the symbolism of intellectual property" *ISIS* 79(299): 606-23.
- (1981) "Scoperte singole e multiple nella scienza (1961)" in *La sociologia della scienza. Indagini teoriche ed empiriche*, edited by N.W. Storer (Milano: Franco Angeli): 436-66.
- Merton, R.K.; Zuckerman, H. (1981) "Modelli istituzionalizzati di valutazione nella scienza (1971)" in *La sociologia della scienza. Indagini teoriche ed empiriche*, edited by N.W. Storer (Milano: Franco Angeli): 569-609.
- Meyer, M. (2000) "What is special about patent citations? Differences between scientific and patent citations" *Scientometrics* 49(1): 93-123.

- Michel, J.; Bettels, B. (2001) "Patent citation analysis. A closer look at the basic input data from patent search reports" *Scientometrics* 51(1): 185-201.
- Moed, H.F. (2002) "Towards research performance in the humanities" *Library Trends* 50(3): 498-520.
- Moed, H.F.; Burger, W.J.M.; Frankfort, J.G.; Van Raan, A.F.J. (1985) "The use of bibliometric data for the measurement of university research performance" *Research Policy* 14: 131-49.
- Moed, H.F.; Glänzel, W.; Schmoch, U. (Ed.) (2004) *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems* (Dordrecht: Kluwer Academic).
- Momigliano, A. (1984) "Le regole del gioco nello studio della storia antica" in *Sui fondamenti della storia antica* (Torino: Einaudi): 477-86.
- Moravcsik, M.J. (1985) *Current Contents - This Week's Citation Classic* (48, December 2): 18.
- Moravcsik, M.J.; Murugesan, P. (1979) "Citation patterns in scientific revolutions" *Scientometrics* 1(2): 161-69.
- (1975) "Some results on the quality and function of citations" *Social Studies of Science* 5: 86-92.
- Muddiman, D. (2003) "Red information scientist: the information career of J.D. Bernal" *Journal of Documentation* 59(4): 387-409.
- Murphy, L.J. (1973) "Lotka's law in the humanities?" *Journal of the American Society for Information Science* 24: 461-62.
- Nagpaul, P.S.; Roy, S. (2003) "Constructing a multi-objective measure of research performance" *Scientometrics* 56(3): 383-402.
- Nalimov, V.V.; Mulchenko, B.M. (1971) *Measurement of Science. Study of the Development of Science as an Information Process* (Washington, DC: Foreign Technology Division, US Air Force System Command) <<http://garfield.library.upenn.edu/nalimovmeasurementofscience/book.pdf>>.
- Narin, F. (1976) *Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity* (Cherry Hill, NJ: Computer Horizons).
- (1994) "Patent bibliometrics" *Scientometrics* 30: 147-55.
- Narin, F.; Noma, E. (1985) "Is technology becoming science?" *Scientometrics* 7: 369-81.
- Narin, F.; Olivastro, D. (1988) "Technology indicators based on patents and patent citations" in *Handbook of Quantitative Studies of Science and Technology*, edited by A.F.J. Van Raan (Amsterdam: Elsevier Science Publishers): 465-507.
- Negri, A. (1996) *Ludici disincanti. Forme e strategie del cinema postmoderno* (Roma: Bulzoni).
- Newby, G.B.; Greenberg, J.; Jones, P. (2003) "Open source software development and Lotka's law: bibliometric patterns in programming" *Journal of the American Society for Information Science and Technology* 54(2): 169-78.
- Nicholls, P.T. (1988) "Price's square root law: empirical validity and relation to Lotka's law" *Information Processing & Management* 24(4): 469-77.
- Nicolini, C.; Vakula, S.; Balla, M.I.; Gandini, E. (1995) "Can the assignment of university chairs be automated?" *Scientometrics* 32: 93-107.
- Nimis, S. (1984) "Fussnoten: das Fundament der Wissenschaft" *Arethusa* 17: 105-34.
- Noyons, E. (1999) *Bibliometric mapping as a science policy and research management tool*. Ph.D. Thesis (Leiden: Universiteit Leiden. Faculteit der Sociale Wetenschappen) <http://www.cwts.nl/cwts/Noyons_1999.pdf>.
- (2004) "Science maps within a science policy context" in *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems*, edited by H.F. Moed, W. Glänzel, et al. (Dordrecht: Kluwer Academic): 237-55.
- Noyons, E.; Luwel, M.; Moed, H.F. (1999) "Combining map and citation analysis for evaluative bibliometric purpose. A bibliometric study on recent development in micro-electronics" *Journal of the American Society for Information Science and Technology* 50(2): 115-31.
- O'Connor, D.O.; Voos, H. (1981) "Empirical laws, theory construction and bibliometrics" *Library Trends* 30(1): 9-20.
- Ogden, P. (1993) "'Mastering the lawless science of our law': a story of legal citation indexes" *Law Library Journal* 85: 1-47.
- OpCit Project (2005) *The effect of open access and downloads ('hits') on citation impact: a bibliography of studies* <<http://opcit.eprints.org/oacitation-biblio.html>>.
- Oppenheim, C. (2000) "Do patent citations count?" in *The Web of Knowledge: a Festschrift in Honor of Eugene Garfield* (Medford, NJ: Information Today Inc): 405-32.
- Osareh, F. (1996) "Bibliometrics, citation analysis and co-citation analysis: A review of literature I" *Libri: International Library Review* 46: 149-58.
- (1996) "Bibliometrics, citation analysis and co-citation analysis: A review of literature II" *Libri: International Library Review* 46: 217-25.
- Osservatorio per la Valutazione del Sistema Universitario (1999) *Proposta per un programma di valutazione della produzione scientifica nelle università. Programma VPS (DOC 3/99)* (Roma).

- Padielli, L.; Perotti, M.; Tuzi, F. (2005) "Criteri ed indicatori per la valutazione degli istituti del CNR" in *Partecipare la scienza*, edited by A. Valente & D. Luzi (Roma: Biblink): 113-32.
- Palmer, J.W.; Bailey, J.P.; Faraj, S. (2000) "The role of intermediaries in the development of trust on the WWW: the use and prominence of trusted third parties and privacy statements" *Journal of Computer-Mediated Communication* 5(3) <<http://www.ascusc.org/jcmc/vol5/issue3/palmer.html>>.
- Pao, M.L. (1993) "Term and citation retrieval - a field study" *Information Processing & Management* 29(1): 95-112.
- Pao, M.L.; Worthen, D.B. (1989) "Retrieval effectiveness by semantic and citation searching" *Journal of the American Society for Information Science* 40(4): 226-35.
- Park, H.W. (2002) "E-science and Hyperlink Network Analysis: Collaborative communication through hyperlinking" *Conference of the Netherlands School of Communications Research*, Utrecht, 2002, November.
- (2003) "Hyperlink Network Analysis: A new method for the study of social structure on the Web" *Connections* 25(1): 49-61 <<http://www.sfu.ca/%7Einsna/Connections-Web/Volume25-1/7.Hyperlink.pdf>>.
- Park, H.W.; Thelwall, M. (2003) "Hyperlink analyses of the World Wide Web: A review" *Journal of Computer-Mediated Communication* 8(4) <<http://www.ascusc.org/jcmc/vol8/issue4/park.html>>.
- Pera, M. (1991) *Scienza e retorica* (Bari: Laterza).
- Peters, H.P.F.; Braam, R.R.; van Raan, A.F.J. (1995) "Cognitive resemblance and citation relations in chemical engineering publications" *Journal of the American Society for Information Science* 46(1): 9-22.
- Peters, H.P.F.; van Raan, A.F.J. (1991) "Structuring scientific activities by co-author analysis: an exercise on a university faculty level" *Scientometrics* 20(1): 235-55.
- Pickering, A.R.; Trower, P.W. (1985) "Sociological problems of high-energy physics" *Nature* 318: 243-45.
- Polanco, X. (1995) "Aux sources de la scientométrie" *Solaris 2* <<http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d02/2polanco1.html>>.
- Polanco, X.; Boudourides, M.A.; Besagni, D.; Roche, I. (2001) *Clustering and mapping web sites. For displaying implicit associations and visualizing networks.* <http://www.math.upatras.gr/~mboudour/articles/web_clustering&mapping.pdf>.
- Polanco, X.; Francois, C.; Lamirel, J.C. (2001) "Using artificial neural networks for mapping of science and technology" *Scientometrics* 51(1): 267-92.
- Popper, K.R. (1970) *Logica della scoperta scientifica: Il carattere autocorrettivo della scienza* (Torino: Einaudi).
- Potter, W.G. (1981) "Lotka's law revisited" *Library Trends* 30(1): 21-39.
- Price, D.J.d.S. (1976) "A general theory of bibliometric and other cumulative advantage processes" *Journal of the American Society for Information Science* 27: 292-306.
- (1986) *Little Science, Big Science...and Beyond* (New York: Columbia University Press).
- (1965) "Networks of scientific papers" *Science* 149: 510-15.
- (1951) "Quantitative measures of the development of science" *Archives Internationales d'Histoire des Sciences* 14: 85-93.
- (1967) *Sociologia della creatività scientifica* (Milano: Bompiani).
- Prime-Claverie, C.; Beigbeder, M. (2004) "Transposition of the cocitation method with a view to classifying web pages" *Journal of the American Society for Information Science and Technology* 55(14): 1282-89.
- Pringle, J. (2004) "Thomson Scientific finds new opportunities in Open Access" *KnowledgeLink Newsletter from Thomson Scientific* <<http://www.thomsonscientific.com/media/newsletterpdfs/2004-11/open-access.pdf>>.
- Pritchard, A. (1972) "Bibliometrics and information transfer" *Research in Librarianship* 4: 37-46.
- (1969) "Statistical bibliography or bibliometrics?" *Journal of Documentation* 25(4): 348-49.
- Radhakrishnan, T.; Kernizan, R. (1979) "Lotka's law and computer science literature" *Journal of the American Society for Information Science* 30(1): 51-54.
- Ranganathan, S.R. (1949) *Aslib Proceedings* 1(2): 102-03.
- (1969) "Librametry and its scope" in *7th Documentation Research Training Centre. Annual Seminars I.* (Bangalore: DTRC).
- Rapoport, A. (1978) "Rank-size relations" in *International Encyclopedia of Statistics*, edited by W. Kruskal & J. Tanur (New York: The Free Press): 847-54.
- Ravetz, J.R. (1971) *Scientific Knowledge and its Social Problems* (Oxford: Clarendon Press).
- Rennie, D.; Yank, V.; Emanuel, L. (1997) "When authorship fails: a proposal to make contributors accountable" *JAMA* 287(7): 579-85.
- Revelli, C. (2002) *Citazione bibliografica* (Roma: Associazione Italiana Biblioteche).
- Riess, P. (1984) *Towards a Theory of the Footnote* (Berlin: De Gruyter).

- Rivette, K.G.; Kline, D. (2000) *Rembrandts in the Attic: Unlocking the Hidden Value of Patents* (Boston: Harvard Business School Press).
- Rodríguez Gairín, J.M. (1997) "Valorando el impacto de la información en Internet: Alavista, el 'Citation Index' de la Red" *Revista Española de Documentación Científica* 20(2): 175-81.
- Rogers, R.; Marres, N. (2000) "Landscaping climate change: a mapping technique for understanding science and technology debates on the World Wide Web" *Public Understanding of Science* 9: 141-63.
- Ross, N.C.M.; Wolfram, D. (2000) "End user searching on the Internet: An analysis of term pair topics submitted to the Excite search engine" *Journal of the American Society for Information Science* 51(10): 949-58.
- Rousseau, R. (1997) "Sitations: an exploratory study" *Cybermetrics* 1(1) <<http://www.cindoc.csic.es/cybermetrics/vol1iss1.html>>.
- Salton, G. (1971) "Automatic indexing using bibliographic citations" *Journal of Documentation* 27(2): 98-110.
- Sampat, B.N.; Ziedonis, A.A. (2004) "Patent citations and the economic value of patents. A preliminary assessment" in *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems*, edited by H.F. Moed, W. Glänzel, et al. (Dordrecht: Kluwer Academic): 277-97.
- Scharnhorst, A. (2003) "Complex networks and the Web: insights from nonlinear physics" *Journal of Computer-Mediated Communication* 8(4) <<http://jcmc.indiana.edu/vol8/issue4/scharnhorst.html>>.
- Scheff, T.J. (2002) *Peer review: an iron law of disciplines?* <<http://www.soc.ucsb.edu/faculty/scheff/23.html>>.
- Schmoch, U. (1993) "Tracing the knowledge transfer from science to technology as reflected in patent indicators" *Scientometrics* 26(1): 193-211.
- Schorr, A.E. (1974) "Lotka's law and library science" *RQ. Reference Quarterly* 14(1): 32-33.
- (1975) "Lotka's law and map librarianship" *Journal of the American Society for Information Science* 26: 189-90.
- (1975) "Lotka's law and the history of legal medicine" *Research in Librarianship* 30: 205-09.
- Seglen, P.O. (1998) "Citation rates and journal impact factors are not suitable for evaluation of research" *Acta Orthopaedica Scandinavica* 69(3): 224-9.
- (1999) "The skewness of science" *Journal of the American Society for Information Science* 43(9): 628-38.
- (1997) "Why the impact factor of journals should not be used for evaluating research" *BMJ* 314(7079): 498-502.
- Sellen, M.K. (1993) *Bibliometrics: an Annotated Bibliography, 1970-1990* (New York: G.K. Hall).
- Sengupta, I.N. (1992) "Bibliometrics, Informetrics, Scientometrics and Librametrics: An Overview" *Libri: International Library Review* 42(2): 75-98.
- Serrai, A. (1984) *Dai "loci communes" alla bibliometria* (Roma: Bulzoni).
- Shapere, D. (1964) "The Structure of Scientific Revolutions" *Philosophical Review* 73: 383-94.
- Shapin, S. (1989) "The invisible technician" *American Scientist* 77(November-December): 554-63.
- Shapiro, F.R. (1992) "Origins of bibliometrics, citation indexing, and citation analysis: the neglected legal literature" *Journal of the American Society for Information Science* 43(5): 337-39.
- Small, H. (1998) "Citations and consilience in science" *Scientometrics* 43(1): 143-48.
- Small, H.; Garfield, E. (1985) "The geography of science: Disciplinary and national mapping" *Journal of the American Society for Information Science* 11(4): 147-59.
- Small, H.; Sweeney, E. (1985) "Clustering the Science Citation Index using co-citations. 1. A comparison of methods" *Scientometrics* 7(3-6): 391-409.
- Small, H.; Sweeney, E.; Grenlee, E. (1985) "Clustering the Science Citation Index using co-citations. 2. Mapping science" *Scientometrics* 8(5-6): 321-40.
- Small, H.G. (1982) "Citation context analysis" in *Progress in Communication Sciences*, 3, edited by B.J. Dervin & M.J. Voigt (Norwood, NJ: Ablex): 287-310.
- (1978) "Cited documents as concept symbols" *Social Studies of Science* 8(3): 327-40.
- (1973) "Co-citation in the scientific literature: a new measure of the relationship between two documents" *Journal of the American Society for Information Science* 24: 265-69.
- (2004) "On the shoulders of Robert Merton: towards a normative theory of citation" *Scientometrics* 60(1): 71-79.
- (2003) "Paradigms, citations and maps of science: a personal history" *Journal of the American Society for Information Science and Technology* 54(5): 394-99.
- (1999) "Visualizing science by citation mapping" *Journal of the American Society for Information Science* 50: 799-813.
- Small, H.G.; Griffith, B.C. (1974) "The structure of scientific literature. I: Identifying and graphing specialties" *Science Studies* 4: 17-40.

- Small, H.G.; Griffith, B.C.; Stonehill, J.A.; Dey, S. (1974) "The structure of scientific literature. II: Toward a macro- and microstructure for science" *Science Studies* 4: 339-65.
- Smith, A.G.; Thelwall, M. (2002) "Web impact factors for Australasian universities" *Scientometrics* 54(3): 363-80.
- "Special issue: Visualizing Scientific Paradigms" (2003) *Journal of the American Society for Information Science and Technology* 54(5).
- Spinak, E. (1998) "Indicadores cientometricos" *Ciência da informação* 27(2): 141-48.
- Steinbach, H.B. (1964) "The quest for certainty: Science citation index" *Science* 145: 142-43.
- Stowe, R.C. (1986) *Annotated bibliography of publications dealing with qualitative and quantitative indicators of the quality of science* (Cambridge, Massachusetts: Harvard University).
- Szigei, H. (2001) *The ISI Web of Knowledge™ Platform: Current and Future Directions* <<http://www.isinet.com/media/presentrep/essayspdf/wokplat.pdf/>>.
- Tague-Sutcliffe, J. (1992) "An introduction to informetrics" *Information Processing & Management* 28(1): 1-3.
- Tague-Sutcliffe, J.; Beheshti, J.; Reespotter, L. (1981) "The law of exponential growth. Evidence, implications and forecasts" *Library Trends* 30(1): 125-45.
- Tammaro, A.M. (2001) "Qualità della comunicazione scientifica" *Biblioteche Oggi* (7): 104-07.
- (2001) "Qualità della comunicazione scientifica 2" *Biblioteche Oggi* (8): 74-78.
- (2004) *Qualità delle pubblicazioni scientifiche ed open access* <<http://hdl.handle.net/1889/56/>>.
- Testa, J.; McVeigh, M.E. (2004) *The impact of open access journals. A citation study from Thomson ISI* <<http://www.isinet.com/media/presentrep/acropdf/impact-oa-journals.pdf/>>.
- Thackray, A. (1979) "Measurement in the Historiography of Science" in *Toward a Metric of Science: the Advent of Science Indicators*, edited by Y. Elkana, J. Lederberg, et al. (New York: John Wiley and Sons): 11-30.
- Thagard, P. (1994) *Rivoluzioni concettuali. Le teorie scientifiche alla prova dell'intelligenza artificiale* (Milano: Guerini e Associati).
- Thelwall, M. (2002) "Conceptualising documentation on the Web: an evaluation of different heuristic-based models for counting links between university Web sites" *Journal of the American Society for Information Science and Technology* 53(12): 995-1005.
- (2001) "Extracting macroscopic information from web links" *Journal of the American Society for Information Science and Technology* 52(13): 1157-68.
- (2004) *Link Analysis. An Information Science Approach* (Amsterdam: Elsevier) <<http://linkanalysis.wlv.ac.uk/index.html/>>.
- (2001) "A Web crawler design for data mining" *Journal of Information Science* 27(5): 319-25.
- (2003) "What is this link doing here? Beginning a finegrained process of identifying reasons for academic hyperlink creation" *Information Research* 8(3) <<http://informationr.net/ir/8-3/paper151.html/>>.
- Thelwall, M.; Tang, R. (2003) "Disciplinary and linguistic considerations for academic Web linking: An exploratory hyperlink mediated study with Mainland China and Taiwan" *Scientometrics* 58(1): 153-79.
- Thelwall, M.; Vaughan, L.; Björneborn, L. (2005) "Webometrics" *Annual Review of Information Science and Technology* 39: 81-135.
- Thelwall, M.; Wilkinson, D. (2003) "Graph structure in three national academic Webs: Power laws with anomalies" *Journal of the American Society for Information Science and Technology* 54(8): 706-12.
- (2003) "Three target document range metrics for university web sites" *Journal of the American Society for Information Science and Technology* 54(6): 489-96.
- Thomson ISI (2005) *Century of Science™* <<http://www.isinet.com/media/presentrep/facts/centuryofscience.pdf/>>.
- (2001) *ISI Essential Science Indicators™ Tutorial Version 1* <<http://www.thomsonisi.com/tutorials/esi/>>.
- (2005) *ISI Journal Citation Reports* <<http://isi17.isiknowledge.com/portal.cgi?DestApp=JCR&Func=Frame>>.
- (2005) *The ISI® database: The journal selection process* <<http://www.isinet.com/essays/selectionofmaterialforcoverage/199701.html/>>.
- (2004) *Thomson ISI and NEC team up to Index Web-based scholarship* <<http://www.isinet.com/press/2004/8217120/>>.
- (2005) *Using Journal Citation Reports® Wisely* <<http://isi17.isiknowledge.com/portal.cgi?DestApp=JCR&Func=Frame>>.
- Tuzi, F. (2005) "The scientific specialization of the italian regions" *Scientometrics* 62(1): 87-111.
- Ugolini, D.; Casilli, C. (2003) "The visibility of italian journals" *Scientometrics* 56(3): 345-55.
- Valente, A.; Luzi, D. (Ed.) (2005) *Partecipare la scienza* (Roma: Biblink).

- van Raan, A.F.J. (2001) "Bibliometrics and Internet: Some observations and expectations" *Scientometrics* 50(1): 59-63.
- (1998) "In matters of quantitative studies of science the fault of theorists is offering too little and asking too much" *Scientometrics* 43(1): 421-28.
- (2004) "Measuring science. Capita selecta of current main issues" in *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems*, edited by H.F. Moed, W. Glänzel, et al. (Dordrecht: Kluwer Academic): 19-50.
- (2000) "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* (Medford, NJ: Information Today Inc): 301-19.
- Vaughan, L.; Hysen, K. (2002) "Relationship between links to journal Web sites and impact factors" *Aslib Proceedings* 54(6): 356-61.
- Vaughan, L.; Shaw, D. (2003) "Bibliographic and Web citations: what is the difference?" *Journal of the American Society for Information Science and Technology* 54(14): 1313-22.
- Vaughan, L.; Thelwall, M. (2003) "Scholarly use on the Web: What are the key inducers of links to journal Web sites?" *Journal of the American Society for Information Science* 54(1): 29-38.
- Vinkler, P. (1988) "An attempt of surveying and classifying bibliometric indicators for scientometric purposes" *Scientometrics* 13: 239-59.
- (1996) "Model for quantitative selection of relative scientometric impact indicators" *Scientometrics* 36(2): 223-36.
- (1994) "The origin and features of information referenced in pharmaceutical patents" *Scientometrics* 30(1): 283-302.
- (1996) "Relationships between the rate of scientific development and citations. The chance for citedness model" *Scientometrics* 35(3): 375-86.
- (1993) "Research contribution, authorship and team cooperativeness" *Scientometrics* 26: 213-30.
- (1995) "Some aspects of the evaluation of scientific and related performances of individuals" *Scientometrics* 32(2): 109-16.
- Vipond, D. (1996) "Problems with a monolithic APA style" *American Psychologist* 51(6): 653.
- Virgo, J.A. (1977) "A statistical procedure for evaluating the importance of scientific papers" *Library Quarterly* 47: 415-30.
- Vlachy, J. (1978) "Frequency distributions of scientific performance: A bibliography of Lotka's law and related phenomena" *Scientometrics* 1: 109-30.
- Voos, H. (1974) "Lotka and information science" *Journal of the American Society for Information Science* 25(4): 270-72.
- Watts, D.J.; Strogatz, S.H. (1998) "Collective dynamics of "small world" networks" *Nature* 393(6684): 440-42.
- Weinberg, B.H. (1974) "Bibliographic coupling: a review" *Information Storage & Retrieval* 10: 189-96.
- (1997) "The earliest hebrew citation indexes" *Journal of the American Society for Information Science* 48: 318-30.
- (1999) "Indexes and religion: reflections on research in the history of indexes" *Indexer* 21(3): 111-18.
- Weinstock, M. (1971) "Citation indexes" in *Encyclopaedia of Library and Information Science*, 5 (New York: Marcel Dekker): 16-40 <<http://www.garfield.library.upenn.edu/essays/v6p548y1983.pdf>>.
- White, H.D. (2003) "Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists" *Journal of the American Society for Information Science and Technology* 54(5): 423-34.
- White, H.D.; Griffith, B.C. (1981) "Author cocitation: A literature measure on intellectual structure" *Journal of the American Society for Information Science* 32: 163-71.
- White, H.D.; McCain, K.W. (1998) "Visualizing a discipline: An author co-citation analysis of information science, 1972-1995" *Journal of the American Society for Information Science* 49(4): 327-56.
- Wilkinson, D.; Harries, G.; Thelwall, M.; Price, E. (2003) "Motivations for academic web site interlinking: evidence for the Web as a novel source of information on informal scholarly communication" *Journal of Information Science* 29(1): 59-66.
- Wilkinson, E.A. (1972) "The ambiguity of Bradford's law" *Journal of Documentation* 28(2): 122-30.
- Wilson, E.O. (2000) *L'armonia meravigliosa: dalla biologia alla religione, la nuova unità della conoscenza* (Milano: Mondadori).
- Wouters, P. (1999) *The Citation Culture*. PhD Thesis (University of Amsterdam)
<<http://www.garfield.library.upenn.edu/wouters/wouters.pdf>>.
- (1998) "The signs of science" *Scientometrics* 41(1-2): 225-41.
- Wyllys, R.E. (1981) "Empirical and theoretical bases of Zipf's law" *Library Trends* 30(1): 53-64.
- Zeisel, H. (1979) "Difficulties in indicator construction: Notes and queries" in *Toward a Metric of Science: the Advent of Science Indicators*, edited by Y. Elkana, J. Lederberg, et al. (New York: John Wiley and Sons): 253-58.

- Zipf, G.K. (1949) *Human Behavior and the Principle of least Effort. An Introduction to Human Ecology* (Reading, Mass: Addison Wesley Press).
- (1936) *The Psycho-Biology of Language: An Introduction to Dinamic Philology* (London: Routledge).